Solution structure and characterization of the DNA binding activity of the B3BP Smr domain

Tammo Diercks[1,2], Eiso AB[1,3], Mark A. Daniels[4], Rob N. de Jong[5], Rogier Besseling, Robert Kaptein and Gert E. Folkers[6].

Bijvoet Centre for Biomolecular Research, Utrecht University, Faculty of Chemistry, Dept. NMR Spectroscopy, Padualaan 8, 3584 CH Utrecht, The Netherlands

[1] Both authors contributed equally to this work.

[2] Current address: CiC bioGUNE, Parque Tecnologico de Bizkaia, Ed. 800, 48160 Derio, Spain

[3] Current address: Leiden University, Leiden Institute of Chemistry, Einsteinweg 55, 2333 CC Leiden, The Netherlands

[4] Current address: Amsterdam University, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.

[5]Current address: Genmab B.V., Yalelaan 60, 3584 CM Utrecht, The Netherlands

[6] To whom correspondence should be addressed (g.e.folkers@uu.nl).

Running title: Structure and DNA binding of the B3BP-Smr domain

Abstract

The MutS1 protein recognizes unpaired bases and initiates mismatch repair Mismatch repair which is essential for high fidelity DNA replication. The homologous MutS2 protein does not contribute to mismatch repair, but suppresses homologous recombination. MutS2 lacks the damage recognition domain of MutS1, but contains an additional C-terminal extension: the small mutS related (Smr) domain. This domain, present in both prokaryotes and eukaryotes, was reported to bind to DNA and possess nicking endonuclease activity. We here determine the solution structure of the Smr domain of the Bcl3 binding protein (B3BP), also known as Nedd-4 binding protein 2 (N4BP2), a protein with unknown function, that lacks other domains present in MutS proteins.

The Smr domain adopts a two-layer α-β sandwich fold, with structural similarity to the C-terminal domain of IF3, the R3H domain and the N-terminal domain of DNase I. The most conserved residues are located in three loops that show distinct sequence identity for prokaryotic and eukaryotic Smr domains. NMR titration experiments and DNA binding studies using B3BP-Smr domain mutants suggested that this most conserved loop regions participates in DNA binding to single-double strand DNA junctions. Based on the observed DNA-binding-induced multimerization, the structural similarity with both subdomains of DNase I and the experimentally identified DNA-binding surface we propose a model for DNA recognition by the Smr domain.

Keywords:

DNA repair, IF3C-fold, NMR, DNA binding domain, DNase I

Introduction

The small MutS related (Smr) domain has previously been identified as the highly conserved C-terminal domain of MutS2 proteins [1]. MutS2 represent a subfamily of MutS homologues [2], named after the DNA mismatch recognition and binding component of the ternary *E.coli* MutSLH complex involved in mismatch repair (MMR). MutS recognizes ds- DNA mismatches ranging from single mispaired to extended loops of unpaired bases [3]. MutH acts as a nicking endonuclease that exclusively targets the damaged unmethylated DNA daughter strand, while MutL provides the linking scaffold between MutH and MutS [4]. While MutS and MutL are widely conserved, the endonuclease MutH has no eukaryotic homolog, instead eukaryotes often contain a MutS2 family member [5].

Like MutS1, MutS2 proteins recognize transitions between single and double stranded DNA that occur during recombination, e.g. at replication forks or Holliday junction [6; 7]. While MutS1 proteins recognize and remove heterologous mismatches, MutS2 proteins might interfere with both homo- and homeologous recombination (HR). Thereby, MutS2 proteins can regulate the rearrangement of endogenous DNA in meiotic crossing-over and chromosome segregation, as well as the incorporation of exogenous DNA [6; 7].

This functional difference between MutS1 and MutS2 is reflected in their distinct domain architecture [2] (Figure 1a). Both share high sequence conservation only for three central domains: the dimerization domain (III), DNA binding domain (IV) and an ATPase domain (V). The smaller MutS2 proteins, however, lacks the N-terminal mismatch recognition domain (I) and the connector domain (II)[8; 9]. In addition, MutS2 has the highly conserved Smr domain located at the C-terminus connected to the conserved core via a putative linker region[4]. This arrangement was suggested to emulate the structure of the bacterial ternary MutSLH complex, where the linker and Smr domain would play the role of the MutL and MutH components, respectively, implying a MutH-like nicking endonuclease activity for Smr [4; 5]. This hypothetical biochemical function was first verified for the C-terminal Smr domain of the human BCL-3 binding protein, B3BP, that converts supercoiled plasmid DNA into nicked open circular DNA, confirming nicking endonuclease activity [10]. DNA binding and incision was also shown for the prokaryotic Smr domain

in *Thermus thermophilus* MutS2 [11]. The "resolving endonuclease" repair activity [4] of the Smr domain in MutS2 proteins can then explain their cellular function, namely the reversion of DNA strand exchange reactions that initiate HR, as shown for *Helicobacter pylori* HpMutS2 [6; 7].

In humans the Smr domain is only present in the B3BP/N4BP2 protein, which was isolated in a yeast two-hybrid screen by its ability to interact with the E3 ubiquitin ligase Nedd-4 [12]. Although the exact function for this protein remains unclear, it was postulated to be involved in transcription, recombination or DNA repair [10]. The recent suggestion that this protein might contribute to sporadic nasopharyngeal carcinoma in the Southern Chinese population underscores its importance[13]. Furthermore, the proteins shown to interact with B3BP such as Nedd4 and Bcl3 are frequently linked to various cancer types[14; 15].

In this study we present the NMR structure of the Smr domain, the most conserved domain of B3BP. We confirm DNA binding to mixed single/double strand DNA sequences, both NMR chemical shift changes upon DNA addition and DNA binding experiments with B3BP-Smr mutants show that the most conserved residues, located in loop regions, form a contiguous, exposed DNA binding surface. Based on these interaction studies and structural homology with DNase I, we propose a model for DNA binding by the B3BP Smr domain.

Results and Discussion

The Smr domain was identified through BLAST searches using the MutS2 specific C-terminal domain[1], that was subsequently shown to be present in bacteria and eukaryotes[10]. In the bacterial kingdom this domain can be present either in isolation or in combination with the MutS core domains III-V (Figure 1a). Sequence analysis further revealed significant differences between the two lineages. An amino acid sequence of a representative subset of the eukaryotic Smr domains is shown in Figure 1b. To determine the solution structure of the Smr domain we expressed the Smr domain of B3BP (1688-1770) or an N-terminally extended domain (1657-1770). Only with the latter we were able to obtain sufficient soluble protein for structural analysis.

Since our Smr domain contains an N-terminal extension in comparison with the previously characterized Smr domain of B3BP, we first confirmed that this protein retains the ability to nick supercoiled DNA (Supplementary Figure 1a) [10]. Surprisingly, linear DNA was observed at elevated Smr concentrations, suggesting additional (endo)nuclease activity. To exclude that the observed catalysis was mediated by impurities, a bacterial culture containing an empty vector was expressed and purified in parallel with the Smr domain protein this control sample was not catalytically active (Supplementary Figure 1a). The metal co-factors magnesium or manganese were required for the nicking endonuclease activity (Supplementary Figure 1b), while barium, cadmium and zinc failed to support this reaction (data not shown). The temperature optimum for this reaction (Supplementary Figure 1c) is in good agreement with the observed temperature dependent unfolding of the Smr domain as determined by thermofluor analysis[16] (Supplementary Figure 1d). These data indicate that the N-terminally extended B3BP Smr domain is functionally active.

The B3BP-Smr domain folds as an α-β two-layer sandwich

The NMR structure of the B3BP-Smr domain (1688–1770) reveals a classical α2β4 sandwich structure with a βαβαββ succession of 4 β strands and 2 α helices. As shown in Figure 1c, all four β-strands form one slightly twisted β-sheet, where strands β_1 (1690–1692), β_2 (1724–1728) and β_3 (1756–1761) run parallel,

with the antiparallel β_4 (1764–1768) inserting between β_2 and β_3. The α-helices pack against one side of the β-sheet, with the longer, slightly bent helix α_1 (1698–1719) running along strands β_1 and β_2, while helix α_2 (1742 – 1753) stacks against strands β_3 and β_4. The helices diverge towards their C-termini at an angle of ca. 30°.

A prominent structural feature is the extended β_2–α_2 loop L_3 (1729–1741) protruding at the bottom of the $\alpha 2\beta 4$ sandwich opposite of both termini of the Smr domain (Figure 1c, indicated in green). This loop is in close contact with the adjacent β_1–α_1 loop L_1 (1693–1697) and short β_3–β_4 loop L_5 (1761–1764). While the overall domain structure appears very rigid and well defined (Table 1), a superposition of the 28 lowest-energy NMR structures (Figure 1c) reveals that only L_3 scatters significantly around R1731–V1739, with increasing amplitude towards its tip near S1735. NMR data (e.g., line-broadening and doubling of signals) indicate that this lack of structural definition is caused by substantial local motions and conformational heterogeneity in L_3. Similar observations for D1692, H1694 and G1695 also reveal flexibility for loop L_1, suggesting that the local structure there reflects a conformational average. Contrarily, the C-terminus of B3BP-Smr appears tightly fixed, where the free K1770 carboxylate forms a salt bridge with the ammonium group of K1717 (end of helix α_1), and the preceding L1769 and M1768 are hydrogen-bonded by the R1756 guanidinium group (in strand β_3).

Recently, the group of S. Yokoyama also determined the NMR solution structure of B3BP Smr domain (**2D9I**). Despite vast overall agreement (backbone RMSD 1.6/2.4Å, without/with L_3), their structure shows a smaller β-sheet with rather frayed edges, where strand β_3 is poorly defined, and confined to only F1757 and S1758. Contrarily, the extended loop L_3 appears more ordered, with some propensity for a short helix at its tip (1732 – 1736).

The N-terminal extension present in our Smr-domain protein is absent in the Yokoyama structure and apparently dispensable for Smr domain folding. The helix α_0 found N-terminal to B3BP-Smr is absent in virtually all structural homologs of Smr (see below), and probably is an artefact of the truncation of the predicted helix-turn-helix fold of the preceding Pfam domain DUF1771[17]. We therefore excluded the N-terminal DUF1771 fragment from our discussion of the B3BP-Smr structure.

Residue and charge conservation cluster in the $L_{1,3,5}$ loop region

While the Smr surface is largely non-conserved (Figure 2b), the compact α2β4 sandwich structure of the B3BP Smr domain is held together by a network of buried and generally conserved hydrophobic sidechains (Figure 2a). Remarkably, the most conserved residues within the eukaryotic Smr domains are not structurally important, but mainly present in the loops L_1, L_3 and L_5 (Figure 2d). These surface exposed residues reside on the same side of the structure (Figure 2e) and form a continuous, positively charged surface (Figure 2f) including the fully conserved histidines H1694, H1734 and the partially conserved R1731 and K1761, overall the B3BP-Smr protein is highly basic, with a calculated pK_I of 9.6.

The prominent loop L_1 appears tightly fixed to the α2β4 core involving interactions within the conserved ^{1692}DLHGΦxΦxEA^{1701} motif (Φ: hydrophobic residue) Hydrophobic interactions between L1693, L1696 (both in L_1) and the small A1701 (in helix α_1), as well as hydrogen bonds between the backbone H^N of H1697 (not conserved) and sidechain carboxylate of E1700 (in helix α_1), and between the sidechain $H^{\delta 1}$ of H1694 and sidechain carboxylate of D1692 (C-terminal in strand β_1) stabilize this loop.

Most notably, NMR data show that only H1694 and α_2 H1753 are doubly protonated. Apparently, the local environment increases their basicity and assists in stabilizing the positive charge. Consequently the charged sidechains of D1692 and H1694 can form a salt bridge, possibly explaining their full conservation among eukaryotic Smr domains. The conserved G1695 appears essential for the turn in this tightly tethered loop L_1 by allowing unusual ϕ and φ angles (ca. 90° and -20°), respectively.

In contrast with L_1, the other prominent loop L_3, is rather flexible and only tethered at its N-terminus. The characteristic ^{1730}G(R)GxHS^{1735} motif at the tip of loop L_3 did not participate in local structure or stabilizing interactions, suggesting functional reasons for the high conservation of these residues. Both ends of the extended loop L_3 feature significantly conserved positive charges in R1731, R1741, and K1743 at the beginning of helix α_2.

Refined sequence alignment reveals differences between pro- and eukaryotic Smr

A refined, structure based sequence alignment reveals some remarkable differences between eukaryotic and prokaryotic Smr domains (Figure 3). Most prominently, the extended loop L_3 (1729-1741 in

B3BP-Smr) with its signatory G(R)GxHS motif in eukaryotic Smr is shorter in prokaryotes, where the histidine is swapped to form a conserved HG(K)G(T)G motif instead. This histidine has recently been shown to be required for catalyzing the nicking endonuclease reaction by MutS2[11]. Prokaryotic Smr domains, however, feature an extended loop L_5 with a characteristic GGxG motif and conserved N-terminal alanine (Figure 3). The prominent L_1 motif DLHG in eukaryotic Smr domains is less conserved in prokaryotes, where the histidine is commonly replaced by arginine, while the aspartate may be substituted by other nucleophilic residues. Although experimental support is needed, these findings suggest that pro- and eukaryotic Smr domains might use distinct sequence motives and mechanisms for catalysis, and recognize different substrate sequences.

Smr domain belongs to an evolutionary conserved fold

A DALI search [18], using the B3BP-Smr structure both with and without N-terminal extension returned more than 500 structurally related proteins (Z scores > 2.0). Structural homology was generally restricted to the α2β4-core of the Smr domain. Most of the closest related structures are implicated in nucleic acid binding, and several possess nuclease activity. They were classified as mixed α-β two layer sandwiches, belonging to the translation initiation factor IF3-like fold or topology according to SCOP[19] and CATH[20]. The B3BP-Smr domain structure indeed shows large structural similarities with the C-terminal domain of IF3[21] (**1TIG**, Z=6.0). The IF3C domain was shown to bind to the ribosome involving a complex pattern of protein-RNA interactions with various IF3C regions[22]. Both length and sequence identity of the regions involved in ribosome binding are distinct in B3BP-Smr.

Various structurally homologous proteins implicated in nucleic acid binding contain functionally relevant features around the most conserved Smr $L_{1,3,5}$ region which might help to understand the proposed DNA binding and catalytic function[10] of the Smr domain in more detail. The highest structural similarity is found for the E.coli YhhP protein of the SirA family (**1DCJ**, Z=7.4 with an RMSD of 1.8A over 64 B3BP residues[23]) , that was implicated in cell division and putatively RNA binding. The most conserved surface residues cluster in the equivalent of the $L_{1,3,5}$ loop region in Smr, but differ greatly from the local Smr sequence (Figure 4). The N-terminal subdomain of the ribosomal S8 protein[24] (**1SEI**, Z=5.9) shows high structural similarity with Smr. Its RNA binding is partially mediated by the equivalent of the most conserved

Smr $L_{1,3,5}$ loop region [25]. However, the primary, specific interaction with RNA is located in the structurally dissimilar C-terminal subdomain of S8 [26]. The YhbY structure[27] (**1LN4**, Z=5.0) contains its most conserved residues in an additional N-terminal helix (oriented differently from B3BP-Smr helix α0) and a loop corresponding to L_1 (Figure 4). Ostheimer et al. [27] proposed an RNA binding motif formed by the L_1 GxxG motif and its surrounding conserved basic surface. Furthermore substantial structural homology was detected with isolated R3H domain structures (e.g. **1MSZ**, Z=4.9 [28])., which have been implicated in single strand nucleic acid binding[29]. Yet, regions proposed to be required for this interaction are absent and the location and type of conserved residues differ greatly between R3H and Smr domains [28; 29]. Finally the R3H structure lacks the equivalent of strand β_1 and the conspicuous loop L_1 (Figure 4).

Another prominent structural homolog is DNase I (**2DNJ** [30], **3DNI** [31], Z=4.5). Most of the residues contacting the DNA are found in the N-terminal domain and again cluster in the equivalent of the largely unstructured $L_{1,3,5}$ loop region in Smr[30; 31] (Figure 5a). In comparison, L_3 is shorter while L_1 (with an intermediate helical turn) and L_5 (with most DNA contacts) are extended. Binding to DNA phosphate groups and in the minor groove mostly occurs through hydrogen bonds (also water-mediated) via both accessible backbone (preferably glycine) and polar sidechain groups. Such residue types are also found generally conserved in the corresponding $L_{1,3,5}$ loop region of Smr.

Characterization of the DNA binding domain

The analysis of structural homologs strongly suggests that Smr domain binds nucleic acids, in line with previous results on DNA binding and incision[5; 10]. We therefore performed electrophoretic mobility shift assays using several probes. We detected weak B3BP-Smr domain-DNA complex formation to single stranded DNA that dissociated during electrophoresis, and binding to double stranded DNA that could not be saturated even at elevated protein concentrations (data not shown). The strongest binding was observed using DNA containing single-double strand transitions, as in a Holliday junction, 20mer hairpin, or bubble-forming DNA probe (a probe composed of two ds DNA stems separated by two opposing 10 bp ss DNA strands) (Figure 5c). A negative control with an empty vector expressed and purified in parallel indicated that the observed complex was indeed formed by B3BP-Smr. DNA-binding was not caused by the N-

terminal His-tag either, since a GST-B3BP-Smr fusion protein likewise produced a (slower migrating) complex (Figure 5d). Addition of EDTA or $MgCl_2$ had no effect on DNA binding (data not shown).

The binding affinities of His-B3BP-Smr and GST-B3BP-Smr for bubble10 DNA are comparable (apparent dissociation constant of 3.1 ± 0.7 μM). Both form higher order complexes upon addition of excess Smr domain, but differ in the behaviour of their DNA complexes in EMSA experiments. The higher order His-B3BP-Smr DNA complex dissociates during electrophoresis, evidenced by smearing of the protein-DNA complex, while the GST-B3BP-Smr DNA complex appears more prominent (Figure 5d, data not shown), possibly stabilized by GST-mediated dimerization (see e.g. Maru *et al* [32]). To further characterize the identity of the B3BP-Smr DNA complex, we made a GST-B3BP DNA complex where the putative dimeric complex is most prominent. Upon addition of increasing amounts of thrombin, that cleaves the GST-tag from the GST-B3BP-Smr fusion protein, we observed an intermediary, probably heterodimeric GST-Smr/Smr DNA complex (Figure 5d). After complete GST removal by thrombin treatment, a complex is formed with equal mobility as His-B3BP DNA complex. Interestingly analytical size exclusion chromatography demonstrated that B3BP-Smr domain is monomeric in solution (data not shown), suggesting that DNA binding induces dimerization.

B3BP-Smr binds DNA primarily in the conserved $L_{1,3,5}$ loop region

To map the DNA binding site on the B3BP-Smr structure, we monitored chemical shift changes of backbone amide signals in the ^{15}N-HSQC upon addition of the hairpin or bubble10 DNA probes; both yielded essentially the same results (Figure 5e). The strongest shifting signals are found around the $\alpha_2\beta_3$ cleft between outer strand β_3 and helix α_2, with one cluster comprising the C-terminus of α_2 and N-terminus of β_3, and a second cluster comprising loop L_5 with adjacent strands β_3 and β_4. Residues within and near loops L_1 and L_3 also shift, although to smaller extent. Unfortunately, signal intensities of several loop region residues fell below the detection limit (e.g., D1692, H1734 and S1735), such that this conspicuous region could not be covered completely by this method. Still, our NMR data shows that DNA binding clearly involves the conserved $L_{1,3,5}$ loop region of Smr, but strongly impacts the $\alpha_2\beta_3$ cleft as well. As some of the

most shifted amide resonances locate at the edges of the β-sheet, a local conformational change upon DNA binding can not be excluded.

To corroborate the DNA binding interface suggested by our NMR data, we created a number of B3BP-Smr domains with point mutations in surface exposed residues. We targeted the conserved motifs in loop L_1 (D1692K and H1694E) and L_3 (G1732P, H1734E and S1735D), as well as conserved charged residues around L_3 (R1731E, R1741E, K1743E) and in helix α_1 (E1700K). As controls, we exchanged poorly conserved residues near loop L_1 (H1697E, D1699K), and residues outside the $L_{1,3,5}$ *loop region* (K1722E in loop L_2, R1756E in strand β_3 near loop L_4).

DNA binding experiments using *bubble10* DNA (Figure 6a,b) show that most significant reduction in affinity (< 30% of *wt*) is caused by the mutations S1735D, R1741E and K1743E (all in loop L_3), K1722E (in L_2) and R1756E (in L_4). The L_3 mutations R1731E and H1734E reduced binding mildly (40 – 70% of *wt*). DNA binding is not markedly affected by the L_1 mutations D1692K, H1694E, H1697E and E1700K. Interestingly, D1699K in L_1 and G1732P in L_3 enhanced DNA binding compared to *wt*. From our mutation studies we may conclude that DNA binding in the $L_{1,3,5}$ *loop* region is largely controlled by loop L_3 residues. The conserved residues in the nearby loop L_1 appear rather irrelevant for DNA binding, but might have catalytic functions. Both NMR titration and mutation studies, however, also implicate other regions, most notably the $\alpha_2\beta_3$ cleft, in DNA binding as well.

The unexpected effects of DNA addition on the $\alpha_2\beta_3$ cleft, revealed by significant shifts in the ^{15}N-HSQC spectra (Figure 5e), are contrasting with the low residue conservation in this region and may have various reasons. For instance, DNA could bind non-specifically to the strongly positively charged $\alpha_2\beta_3$ cleft region, as also proposed for the structurally homologous R3H domains (see above); this explanation is supported by the greatly reduced DNA binding observed for the charge inverted R1756E mutant in strand β_3. Yet, in contrast to the other mutants, the ^{15}N-HSQC spectrum for R1756E also showed chemical shift changes farther away from the mutation site, including the L_3 loop.

Alternatively, the $\alpha_2\beta_3$ cleft region could be the interface for DNA-mediated dimerization that we have indeed observed in our EMSA experiments (Figure 5d). Also, the prokaryotic Smr domain of *Thermus thermophilus* MutS2 [11] was shown to dimerize, like many endonucleases, and dimerization of the

structurally homologous Alba protein (**1H0X,** *Z*=4.5) (Figure 4) does in fact take place via the corresponding cleft region [33]. The lack of significant residue conservation in the $\alpha_2\beta_3$ cleft region is, however, difficult to reconcile with a crucial functional role for this region, be it direct by DNA binding or indirect via dimerization.

Structural similarity with nucleic acid binding domains suggests models for DNA binding by Smr

Our NMR titration and mutation studies show that the $L_{1,3,5}$ loop region, and particularly loop L_3, are involved in DNA binding by B3BP-Smr. This finding is corroborated by the distinct local residue conservation, and by comparison with structural homologs (Figure 4). Using the structures of B3BP homologues complexed with DNA, we attempted to model DNA binding by Smr domains.

DNase I is composed of N- and C-terminal domains with substantial structural similarity, forming an intrinsic pseudo-dimer via the β-sheets of both domains[31]. B3BP-Smr shows significant structural homology to both domains (Figure 5b), with more similarity to the primarily DNA-binding N-terminal domain (Z=4.%) than to the catalytic C-terminal domain (Z=2.8). Figure5a shows the best-fit superposition of B3BP-Smr onto the N-terminal domain of DNA-bound DNase I (**2DNJ**) [30]. Smr could contact DNA with its $L_{1,3,5}$ loop region in an analogous manner, but the length of the loops and residue identity differ between both proteins. For instance, DNase I residues N74 and S75, both implied in protein-DNA interaction and located in the longer loop L_5 equivalent, are absent in B3BP-Smr. Instead, the adjacent loop L_3 is much longer in eukaryotic Smr and contains the highly conserved S1735 that might substitute for S75 in DNase I. Eukaryotic Smr loop L_3 also shows a high propensity for positive charges (R1731, R1741, K1743), where the partially conserved R1731 has a structural counterpart in R41 of DNase I. Furthermore R9, located in L_1 of DNase I, makes water-mediated base contacts and has a conserved positive charge, the doubly protonated H1694, at the corresponding position in eukaryotic Smr domain.[30; 31; 34](Figure 5a).

The C-terminal domain of DNase which is primarily active in catalysis, may serve as an instructive model to identify the putative endonucleolytic residues in B3BP Smr. The conserved H252, in conjunction with D212 or D251, acts as the base deprotonating the attacking water molecule, while H134, assisted by

E78 (located in the N-terminal domain), acts as general acid that protonates the leaving O3' of the cleaved ribose [34]. In a best-fit superposition with the C-terminal domain of DNase I (Supplementary Figure 2), the fully conserved loop L_1 residue H1694 in (eukaryotic) Smr domains could come spatially close to H134, and H1734 in the very flexible loop L_3 close to H252. In this scenario, the role of assisting carboxylate groups could be provided by the fully conserved D1692 as partner to H1694 (analogous to the E78 – H134 dyad in Dnase I), and the highly conserved E1700 as partner to H1734 (analogous to the D212–H252 or D251–H252 dyad). As delineated before, we indeed found a significantly increased basicity (*cf* protonation) of H1694, and experimental evidence for a D1692–H1694 sidechain interaction. Unfortunately, the surprisingly low endonuclease activity of wt B3BP Smr prohibited sound experimental corroboration of these results by mutagenesis. This low activity could be caused by the absence of other contributing B3BP subdomains or unidentified cofactors; furthermore, our DNA templates chosen for the cleavage assays may not be appropriate targets for B3BP in vivo.

While more detailed studies will be needed to elucidate the detailed mechanism of DNA binding and catalysis, our studies indicate that largely conserved charged residues within the $L_{1,3,5}$ region contribute to DNA binding, and show resemblance to the contacts made by DNase I in complex with DNA. The fact that residues involved in DNase I activity have structural counterparts in the DNA-bound dimeric B3BP-Smr model, suggests a similar mechanism for catalysis.

In conclusion, the determination B3BP-Smr domain structure and experimental data on DNA binding suggests that one of the functions of this domain is the recognition of distorted DNA sequences with stretches of unpaired bases, underscored by its ability to bind to Holliday junction, bubble and hairpin substrates. This agrees with the proposed role for MutS2 in the inhibition of homologous or homeologous recombination.

Sample preparation

PCR, Cloning and Functional Validation

The B3BP Smr domain expression constructs were obtained by PCR amplification of cDNA obtained from a pool of RNA isolated from various human cell lines [35]. The PCR products were cloned into pLICHIS, a His-tag containing pET15B derived expression vector via, ligation independent cloning[36] . The core B3BP Smr domain (1688–1770) with an N-terminal His-tag was poorly expressed and precipitated at concentrations above 100µM. We therefore expressed an N-terminally extended construct (1657-1770). This protein was highly soluble and a final concentration of 1 mM was obtained in 50 mM sodium phosphate buffer (pH 6.0) and 150 mM NaCl.

Mutations

Expression constructs for the mutant B3BP Smr proteins were created by a double mutant PCR reaction described by Barik et al [37]. This PCR product was subsequently cloned into pLICHIS using our previously described Enzyme Free Cloning strategy[36]. All (mutant) expression constructs were verified by sequencing.

Expression and Purification

Recombinant protein expression and isotope labeling was performed in the *Escherichia coli* strain BL21 Rosetta2 (Novagen) essentially as described before[35; 36]. Induction was performed with 0.5mM IPTG at 18°C for 16 hours. The cell pellet was resuspended in 10 ml lysisbuffer [50mM HPO_4 pH8.0, 300mM NaCl, 20mM imidazole, 1 mM β-mercaptoethanol, 0.2% Triton X100, 0.2 mM PMSF] containing 100 µl EDTA-free protease inhibitor cocktail (Sigma). Following resuspension, 10 mg lysozyme were added per liter culture. After two freeze/thaw cycles and sonication, the sample was cleared by centrifugation (30000g) and applied on a Poros 20 metal chelate column (PerSeptive Biosystems). The His-tagged protein was eluted with elution buffer [50mM HPO_4 pH8.0, 450mM NaCl, 750mM imidazole], and following buffer-exchange to gel-filtration buffer [50mM HPO_4 pH8.0, 450mM NaCl], the protein was applied onto a Sephadex 75 gel-filtration column (GE HealthCare). The purified protein was concentrated (Amicon) and the buffer

exchanged to either NMR buffer [50mM HPO_4 pH6.0, 150mM NaCl, 5%D_2O, 0.02% NaN_3], containing 2µl complete protease inhibitor cocktail (Roche) for 500 µl sample or to assay buffer [50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM dithiothreitol, 0.02% NaN_3].

Protein Quantification

The protein concentration in the NMR samples was determined by UV absorbance at 280 nm using the calculated extinction coefficient. The protein concentration for purified mutants was first determined using a modified Bradford assay (Bio Rad) followed by normalization based on coomassie-stained SDS-Page, and referenced against samples with known amounts of proteins.

Nicking endonuclease assay and DNA binding

EMSA

All electrophoretic mobility shift assays (EMSA) were carried out as described before using the indicated radiolabeled probes[38] in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1.0 mM DTT, 7.5% glycerol, and 10 mM $MgCl_2$ in the presence of the indicated amount of B3BP Smr domain protein. Samples were incubated for 30 minutes on ice and loaded on a 8 % non-denaturing polyacrylamide gels buffered with 0.5x TBE at 4^{o}C for 2 to 3 hours. The gels were dried, exposed overnight to a phosphor-imager screen, and visualized using a Personal FX Phosphor Imager (Bio-Rad). Quantification was performed as described before[38].

Nicking endonuclease activity assay

The indicated amounts of recombinant B3BP Smr were added to assay buffer [50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$ and 1 mM dithiothreitol] containing 150 ng of cesium chloride gradient purified (pTKLuc[39] , or Maxi-prep (Qiagen) purified pET15B (Novagen) supercoiled circular plasmid DNA. The reaction mixture (10µl) was incubated for 2 hours at 37°C and the reaction was stopped by adding 10x concentrated stop buffer [10mM Tris pH8.0, 50%glycerol, 0.25% Bromphenol Blue, 0.25% Xylene Cyanol, 250mM EDTA]. The samples were separated on 0.8% agarose gel, stained by ethidium bromide and visualized under UV light.

NMR measurements

All NMR experiments were run on BRUKER AVANCE spectrometers operating at 700, 750 and 900 MHz using a 1 mM sample of [U-^{13}C,^{15}N] doubly isotope-labeled B3BP Smr domain protein in NMR buffer (see above); a corresponding [U-^{15}N] singly isotope-labeled sample was used for all NMR experiments that did not require ^{13}C labeling. A prior temperature test series of ^{15}N-HSQC spectra established maximal signal intensities around 308 K (from an optimum between T_2 relaxation reduction and increased H^N / H_2O exchange), all spectra were recorded at 305 K., NMR data was collected according to our standard protocol[40] using a series of standard 3D triple resonance experiments (reviewed in[41; 42]). NOE distance constraints for the structure elucidation were derived from high-resolution 2D H,H- and H[^{15}N-suppressed], H-NOESY and a set of 3D H,NH-, H,CH-, [H]C,NH- and [H]C,CH-[HSQC]-NOESY-HSQC spectra [43].

The protonation states of the histidine imidazole rings were derived from a 2D ^{15}N-HMQC tuned for evolving the small $^2J_{HN}$ coupling between $H^{\varepsilon 2}$-$N^{\varepsilon 2}$ or $H^{\varepsilon 1}$-$N^{\delta 1}$ [44].

Structure Calculation

Automated NOE assignment and structure calculations were performed using the CANDID [45] module of CYANA2.1 [46]. 10 CANDID runs with different random seeds were performed in order to prevent accidental convergence. Distance restraints were used if they were present in the final cycle of at least half of the 10 runs. Unassigned resonances that were unambiguously involved in NOE contacts were represented by appropriate PROXY residues during the structure calculations [47]. Dihedral angle restraints were calculated using TALOS [48]. Water refinement was performed using CNS [49] according to the RECOORD protocol [50]. Structures were validated using WHAT IF [51] and PROCHECK [52].

Accession Number

Structure coordinates have been deposited in the Protein Data Bank with accession number **2VKC.**

Acknowledgements

This work was supported by the EU SPINE (QLG2-CT-2002-00988) and SPINE2-complexes (Contract number 031220) Grants and by the NWO Horizon Breakthrough Project (050-71-629).

Figure legends

Figure1 Solution structure of the human B3BP Smr domain. (a) Schematic domain organization of *Escherichia coli* MutS1 (UniProt Database entry: **P23909**) and various proteins containing the Smr domain: *Thermotoga maritima* Muts2 (**Q9X105**), Eschericha coli YFCN (**P0A8B2**), human N4BP2/ B3BP (**Q86UW6**), *Saccharomyces cerevisiae* YP199 (**Q08954**). DUF refers to Pfam[17] domain of unknown function DUF 1771. (b). Multiple sequence alignment of a representative set of eukaryotic Smr domain proteins, **Q86UW6**: human N4BP2, **Q9UTP4**: *Saccharomyces pombe* YLL3, **Q08954**: *Saccharomyces cerevisiae* YP199, **O64843**: *Arabidopsis thaliana* At2g26280, **O74840**: *Saccharomyces pombe* YCY3, **A2QVU0**: *Aspergillus niger* An11g02630 , **O60961**: *Leishmania major* LMJ_0021. Secondary structure elements are indicated by color coding: α-helices (red), β-strand (cyan), loop L_3 (green). Sequence similarity is shown by box shading. More similar residues are indicated by darker box shading. Amino acid numbering for B3BP is indicated above the sequence. (c) Backbone trace of an ensemble of the 28 lowest energy NMR structures deposited in the Protein Data Bank (**2VKC**), color coding as in b. Amino- (N) and Carboxyl (C)-terminal ends are indicated (d). Ribbon plot of the lowest energy structure with the various structure elements and loops indicated. All pictures were created using Pymol (www.pymol .sourceforge.net).

Figure 2 Conserved residues contribute to hydrophobic packing and form a consecutive surface exposed patch. (a) Ribbon representation showing the most conserved residues (in stick representation) that contribute to the formation of the hydrophobic core of the protein., for clarity hydrogens are not displayed. (b) Surface representation in the same orientation, with the most conserved surface exposed residues annotated. (c) Surface representation coloured by electrostatic potential, with surface exposed charged residues indicated. (d) Ribbon representation showing the most conserved surface exposed residues. (e)

Surface representation showing a patch of highly conserved exposed residues clustering mostly in loops L_1 and L_3. (f) Surface representation coloured by electrostatic potential, with surface exposed charged residues indicated. Panels a,b,d (side chains only) and e are coloured according to sequence conservation, determined with Consurf [53], using the multiple sequence alignment of all eukaryotic Smr domain sequences according to Pfam[17] as input, presented using the default colouring scheme (red: most conserved; cyan: non conserved). Panels c and e: electrostatic surface potential calculated using APBS[54], coulored blue for positive, red for negative potential.

Figure 3. Multiple sequence alignment of a subset of prokaryotic and eukaryotic Smr domains. Residue numbering and secondary structure of B3BP Smr are given above. Shading code for residue conservation within all non-redundant eukaryotic (44) or prokaryotic (116) Smr sequences contained in the SMART[55]database, as described in figure 1, insertion is indicated by a red triangle. Most distantly related Smr domain sequences were selected for both the pro- and eukaryotes subgroup, aligned separately using ClustalW[56] with subsequent manual refinement using G1720 as anchor point and adjusting the length of helix α1. Beyond G1720, automatic alignment of pro- and eukaryotic Smr diverges significantly. Secondary structure predictions for prokaryotic Smr domains using JPred [57](shown below the prokaryotic alignment), and residues G1738 and L1750 as further anchor points to guide the alignment against eukaryotic Smr domain structure by adjusting the length of the predicted loops. In between the two alignments residue numbering is indicated according to the position within the two alignments.
(B) Ribbon representation of the B3BP-Smr domain with sequence conservation calculated using Consurf [53], and plotted using the default coloring scheme as described in figure 2. Sidechains for the most conserved residues in stick representation, and numbered according to the alignment position numbering (C) Homology model was created for the Desulfovibrio vulgaris Smr domain using the Swiss-modeling server [58] with the alignment depicted above as input. Sequence conservation was determined and plotted as above using default settings of Consurf.

Figure 4 Structural similarities with other two layer α-β sandwich folds. (a) Manually adjusted, structure-based alignment showing a representative set of the most similar structural homologs present in the PDB, as identified using DALI searches[18] with the B3BP Smr core domain (1691-1770) as input. Residues most highly conserved among Smr domains or among its structural homologs are colored in red or blue respectively. For DNase I (**2DNJ**) the residues shown to interact with DNA are highlighted in blue. Conserved residues among all structural homologs are indicated as gray shaded boxes. The secondary structure elements of B3BP-Smr domain are indicated above the sequence. (b). Ribbon representation (orientation as in Figure 2d) of B3BP-Smr domain overlaid on the selected structural homologs (indicated below the structure). Regions with structural similarity according to DALI are coloured red (B3BP-Smr) or green (structural homolog), dissimilar regions are shown in gray, conserved residues are shown in stick representation and coloured red (Smr) and blue (structural homolog). (c) As in b but oriented for better view of the most conserved residues in the structural homolog. For clarity, only the last two digits of B3BP residue numbers are shown.

Figure 5 Structural similarities provide insight into the putative DNA binding mechanism of B3BP-Smr domain. (a) Ribbon representation (as in Figure 4b) of the structural similarity between B3BP-Smr domain and DNase I (**2DNJ**) in two orthogonal orientations, with the DNase I residues required for DNA binding shown as blue sticks. Conserved Smr domain residues that come in close proximity of the DNA (bright-orange) are presented as red sticks. For clarity, only the last two digits of B3BP residue numbers are shown. (b) Structural similarity between B3BP-Smr domain and the N-and C-terminal repeat domains of DNase I. Structurally similar regions according to DALI are colored in yellow for DNase I. The structurally similar regions of the B3BP-Smr domain with the N- and C-terminal repeat of DNase I are colored in green and magenta respectively. The N and C termini are indicated in black for DNase I and magenta for B3BP-Smr respectively. (c) Electrophoretic mobility shift assay (EMSA) using a 30 bases ssDNA sequence, and a 30 base-pair dsDNA sequence, a 30 base pair bubble substrate with 10 unpaired bases in the presence of 0, 0.63, 1.25, 2.5, 5, 10, 20 μM B3BP-Smr. Neg. refers to a control sample obtained from an empty expression vector by simultaneous expression and purification with the B3BP-Smr domain protein, using identical

methods. (d) The left panel shows binding of 10 μM (+) His-tagged (H) or GST-tagged (G) B3BP-Smr domain, or an equivalent amount of control His or GST protein without B3BP-Smr domain (-) bound to the bubble substrate. Open and closed circles refer to the most probable DNA complex of His-tagged and GST-tagged B3BP-Smr domain respectively. The right panel shows 1.5 μM GST-B3BP-Smr bound to a bubble substrate (closed circles), incubated with increasing amount of thrombin (Thr) resulting in the cleavage of the GST-moiety from GST-B3BP-Smr, leading to the formation of a heterodimeric complex at intermediary thrombin concentrations, and a faster migrating fully cleaved complex at the highest thrombin concentrations (open circles). The asterisk refers to a non-specific complex. (e). Ribbon representation showing the chemical shift perturbation of the amide resonances upon addition of a 2 fold excess of bubble substrate. Unaffected residues (cutoff = 0.035 ppm, averaged shift change) are shown in light gray, strongly affected residues in red, unassigned residues are shown in dark gray. Left panel in the same orientation as Figure 2d, middle panel oriented as Figure 4a.

Figure 6 Identification of the DNA binding surface by site-directed mutagenesis. (a) The upper panel shows a representative EMSA experiment as performed in Figure 5c, using wild type B3BP-Smr (wt) and the indicated mutants. The lower panel shows average and standard deviation of the quantification of these binding experiments where the fraction of bound DNA is plotted as a function of the B3BP Smr domain protein concentration (μM). (b) For all binding experiments, the apparent K_d was calculated using a non-linear regression curve fitting program. Average K_d and standard deviation was determined for wild type and all mutants. The relative binding affinity is determined by dividing the determined apparent K_d of the wild type with the apparent K_d of the mutant, where the binding affinity of wild type is set at 1. For clarity only the last two digits of the mutated residue numbers are shown. (c) Coomassie stained SDS-PAGE of the protein samples used for the DNA binding studies. (d) Surface representation where the influence of the mutation on the DNA binding capability is plotted on the surface of the B3BP-Smr domain structure. No or positive effect on binding affinity: white, most effect residues in red. Insets show the corresponding ribbon representations with the mutated residues indicated as sticks, all orientations as in Figure 5e.

References

1. Moreira, D. & Philippe, H. (1999). Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. *Trends in Biochemical Sciences* **24**, 298-300.
2. Eisen, J. A. (1998). A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res* **26**, 4291-300.
3. Joshi, A. & Rao, B. J. (2001). MutS recognition: multiple mismatches and sequence context effects. *J Biosci* **26**, 595-606.
4. Malik, H. S. & Henikoff, S. (2000). Dual recognition-incision enzymes might be involved in mismatch repair and meiosis. *Trends in Biochemical Sciences* **25**, 414-418.
5. Fukui, K., Masui, R. & Kuramitsu, S. (2004). Thermus thermophilus MutS2, a MutS paralogue, possesses an endonuclease activity promoted by MutL. *J Biochem (Tokyo)* **135**, 375-84.
6. Pinto, A. V., Mathieu, A., Marsin, S., Veaute, X., Ielpi, L., Labigne, A. & Radicella, J. P. (2005). Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Mol Cell* **17**, 113-20.
7. Kang, J., Huang, S. & Blaser, M. J. (2005). Structural and functional divergence of MutS2 from bacterial MutS1 and eukaryotic MSH4-MSH5 homologs. *J Bacteriol* **187**, 3528-37.
8. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N. & Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* **407**, 711-7.
9. Obmolova, G., Ban, C., Hsieh, P. & Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-10.
10. Watanabe, N., Wachi, S. & Fujita, T. (2003). Identification and characterization of BCL-3-binding protein: implications for transcription and DNA repair or recombination. *J Biol Chem* **278**, 26102-10.
11. Fukui, K., Kosaka, H., Kuramitsu, S. & Masui, R. (2007). Nuclease activity of the MutS homologue MutS2 from Thermus thermophilus is confined to the Smr domain. *Nucleic Acids Res*.
12. Murillas, R., Simms, K. S., Hatakeyama, S., Weissman, A. M. & Kuehn, M. R. (2002). Identification of developmentally expressed proteins that functionally interact with Nedd4 ubiquitin ligase. *Journal of Biological Chemistry* **277**, 2897-2907.
13. Zheng, M. Z., Qin, H. D., Yu, X. J., Zhang, R. H., Chen, L. Z., Feng, Q. S. & Zeng, Y. X. (2007). Haplotype of gene Nedd4 binding protein 2 associated with sporadic nasopharyngeal carcinoma in the Southern Chinese population. *J Transl Med* **5**, 36.
14. Chen, C. & Matesic, L. E. (2007). The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev* **26**, 587-604.
15. Robinson, H. M., Taylor, K. E., Jalali, G. R., Cheung, K. L., Harrison, C. J. & Moorman, A. V. (2004). t(14;19)(q32;q13): a recurrent translocation in B-cell precursor acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **39**, 88-92.
16. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289-98.
17. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. & Bateman, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247-51.
18. Holm, L. & Sander, C. (1996). The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res* **24**, 206-9.
19. Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2008). Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* **36**, D419-25.
20. Greene, L. H., Lewis, T. E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., Pearl, F., Nambudiry, R., Reid, A., Sillitoe, I., Yeats, C., Thornton, J. M. & Orengo, C. A. (2007). The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res* **35**, D291-7.

21. Biou, V., Shu, F. & Ramakrishnan, V. (1995). X-ray crystallography shows that translational initiation factor IF3 consists of two compact alpha/beta domains linked by an alpha-helix. *Embo J* **14**, 4056-64.
22. Sette, M., Spurio, R., van Tilborg, P., Gualerzi, C. O. & Boelens, R. (1999). Identification of the ribosome binding sites of translation initiation factor IF3 by multidimensional heteronuclear NMR spectroscopy. *Rna* **5**, 82-92.
23. Katoh, E., Hatta, T., Shindo, H., Ishii, Y., Yamada, H., Mizuno, T. & Yamazaki, T. (2000). High precision NMR structure of YhhP, a novel Escherichia coli protein implicated in cell division. *J Mol Biol* **304**, 219-29.
24. Davies, C., Ramakrishnan, V. & White, S. W. (1996). Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from Bacillus stearothermophilus at 1.9 A resolution. *Structure* **4**, 1093-104.
25. Merianos, H. J., Wang, J. & Moore, P. B. (2006). The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA* **10**, 954-964.
26. Tishchencko, S., Nikulin, A., Fomenkova, N., Nevskaya, N., Nikonov, O., Dumas, P., Moine, H., Ehresmann, B., Ehresmann, C., Piendl, W., Lamzin, V., Garber, M. & Nikonov, S. (2001). Detailed analysis ofRNA-protein interaction within the ribosomal protein S8-rRNA complex from the archaeon Methanococcus jannaschii. *Journal of Molecular Biology* **311**, 311-324.
27. Ostheimer, G. J., Barkan, A. & Matthews, B. W. (2002). Crystal structure of E. coli YhbY: a representative of a novel class of RNA binding proteins. *Structure* **10**, 1593-601.
28. Liepinsh, E., Leonchiks, A., Sharipo, A., Guignard, L. & Otting, G. (2003). Solution structure of the R3H domain from human Smubp-2. *J Mol Biol* **326**, 217-23.
29. Grishin, N. V. (1998). The R3H motif: a domain that binds single-stranded nucleic acids. *Trends Biochem Sci* **23**, 329-30.
30. Lahm, A. & Suck, D. (1991). DNase I-induced DNA conformation. 2 A structure of a DNase I-octamer complex. *J Mol Biol* **222**, 645-67.
31. Oefner, C. & Suck, D. (1986). Crystallographic refinement and structure of DNase I at 2 A resolution. *J Mol Biol* **192**, 605-32.
32. Maru, Y., Afar, D. E., Witte, O. N. & Shibuya, M. (1996). The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. *J Biol Chem* **271**, 15353-7.
33. Wardleworth, B. N., Russell, R. J. M., Bell, S. D., Taylor, G. L. & White, M. F. (2002). Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO Journal* **21**, 4654-4662.
34. Weston, S. A., Lahm, A. & Suck, D. (1992). X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 A resolution. *J Mol Biol* **226**, 1237-56.
35. Folkers, G. E., van Buuren, B. N. & Kaptein, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. *J Struct Funct Genomics* **5**, 119-31.
36. de Jong, R. N., Daniels, M. A., Kaptein, R. & Folkers, G. E. (2006). Enzyme free cloning for high throughput gene cloning and expression. *J Struct Funct Genomics* **7**, 109-18.
37. Barik, S. (1995). Site-directed mutagenesis by double polymerase chain reaction. *Mol Biotechnol* **3**, 1-7.
38. de Jong, R. N., Truffault, V., Diercks, T., Ab, E., Daniels, M. A., Kaptein, R. & Folkers, G. E. (2008). Structure and DNA binding of the human Rtf1 Plus3 domain. *Structure* **16**, 149-59.
39. Folkers, G. E., van der Burg, B. & van der Saag, P. T. (1998). Promoter architecture, cofactors, and orphan receptors contribute to cell-specific activation of the retinoic acid receptor beta2 promoter. *J Biol Chem* **273**, 32200-12.
40. Ab, E., Atkinson, A. R., Banci, L., Bertini, I., Ciofi-Baffoni, S., Brunner, K., Diercks, T., Dotsch, V., Engelke, F., Folkers, G. E., Griesinger, C., Gronwald, W., Gunther, U., Habeck, M., de Jong, R. N., Kalbitzer, H. R., Kieffer, B., Leeflang, B. R., Loss, S., Luchinat, C., Marquardsen, T., Moskau, D., Neidig, K. P., Nilges, M., Piccioli, M., Pierattelli, R., Rieping, W., Schippmann, T., Schwalbe, H., Trave, G., Trenner, J., Wohnert, J., Zweckstetter, M. & Kaptein, R. (2006). NMR in the SPINE Structural Proteomics project. *Acta Crystallogr D Biol Crystallogr* **62**, 1150-61.

41. Oschkinat, H., Müller, T. & Dieckmann, T. (1994). Protein structure determination with three- and four-dimensional NMR spectroscopy. *Angewandte Chemie International Edition in English* **33**, 277-293.
42. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experimets for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93-158.
43. Diercks, T., Coles, M. & Kessler, H. (1999). An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. *Journal of Biomolecular NMR* **15**, 177-180.
44. Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. (1993). Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci* **2**, 543-58.
45. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-227.
46. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-298.
47. AB, E., Pugh, D. J., Kaptein, R., Boelens, R. & Bonvin, A. M. (2006). Direct use of unassigned resonances in NMR structure calculations with proxy residues. *Journal of the American Chemical Society* **128**, 7566-7571.
48. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302.
49. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1997–2001). Crystallography & NMR System (CNS) 1.1 edit. Yale University, New Haven, CT.
50. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R. & Bonvin, A. M. (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662-672.
51. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics* **8**, 52-56.
52. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**, 477-486.
53. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302.
54. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41.
55. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. & Bork, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**, D257-60.
56. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.
57. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892-3.
58. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.

Table 1.a) Overview of restraints used in the NMR structure calculation

NOE-based distance restraints	
Intraresidue (|i - j| = 0)	771
Sequential (|i - j| < 1)	824
Medium range (1 < |i - j| < 5)	680
Long range (|i - j| > 5)	823
Total	3098
Dihedral restraints phi + psi (TALOS)	70 + 70
H-bond restraints	16 + 16
Distance violations > 0.2 Å	0
Dihedral violations > 5°	0

Table 1.b) Validation statistics for the ensemble of calculated 28 lowest-energy structures

Residues	**1666 – 1770 (all)**	**1686 – 1770 (Smr domain)**	**1686 – 1730 1740 – 1770 (NOT loop L_3)**
RMSD [Å] from pairwise alignment of backbone atoms	1.47 ± 0.37	1.25 ± 0.34	0.53 ± 0.09
RMSD [Å] from pairwise alignment of heavy atoms	2.08 ± 0.38	1.90 ± 0.37	1.05 ± 0.12
PROCHECK: Distribution of residues [%] in φ/ψ–space (Ramachandran plot)			
Most favoured regions	89.47 ± 1.81	86.90 ± 2.41	92.09 ± 2.25
Allowed regions	8.88 ± 2.20	11.00 ± 2.74	6.95 ± 2.19
Generously allowed regions	0.94 ± 0.88	1.20 ± 1.12	0.46 ± 0.74
Disallowed regions	0.71 ± 0.80	0.90 ± 1.02	0.51 ± 0.87
WHATCHECK: Structure Z-scores			
1st generation packing quality	0.12 ± 0.24	-0.19 ± 0.19	0.70 ± 0.18
2nd generation packing quality	0.26 ± 0.32	0.32 ± 0.31	1.11 ± 0.30
Ramachandran plot appearance	-2.47 ± 0.42	-2.54 ± 0.43	-2.16 ± 0.56
ζ_1/ζ_2 rotamer normality	-1.05 ± 0.41	-0.69 ± 0.45	-0.68 ± 0.50
Backbone conformation	-3.16 ± 0.37	-3.41 ± 0.38	-1.67 ± 0.33
WHATCHECK: RMS Z-scores			
Bond lengths	0.78 ± 0.03	0.81 ± 0.03	0.80 ± 0.03
Bond angles	0.82 ± 0.03	0.84 ± 0.03	0.82 ± 0.03
Omega angle restraints	0.71 ± 0.05	0.74 ± 0.06	0.73 ± 0.06
Side chain planarity	0.81 ± 0.10	0.82 ± 0.11	0.85 ± 0.12
Improper dihedral distribution	0.88 ± 0.04	0.88 ± 0.04	0.88 ± 0.04
Inside/Outside distribution	1.04 ± 0.02	1.04 ± 0.02	1.02 ± 0.02
WHATCHECK: Counts			
Number of bumps per 100 residues	6.87 ± 1.61	7.90 ± 1.92	7.75 ± 1.94
Unsatisfied buried hydrogen donors	9.61 ± 1.95	10.04 ± 1.91	8.25 ± 1.58
Unsatisfied buried hydrogen acceptors	0.54 ± 0.74	0.14 ± 0.45	0.07 ± 0.26

Dihedral restraints were generated by TALOS on the basis of backbone atom chemical shifts.

Hydrogen bonds were used during CANDID runs, not during water refinement. Each Hydrogen bond was defined with two distance restraints: 1.8 Å $\leq$ d(HN-O) $\leq$ 2.3 and 2.8 Å $\leq$ d(N-O) $\leq$ 3.3 Å.

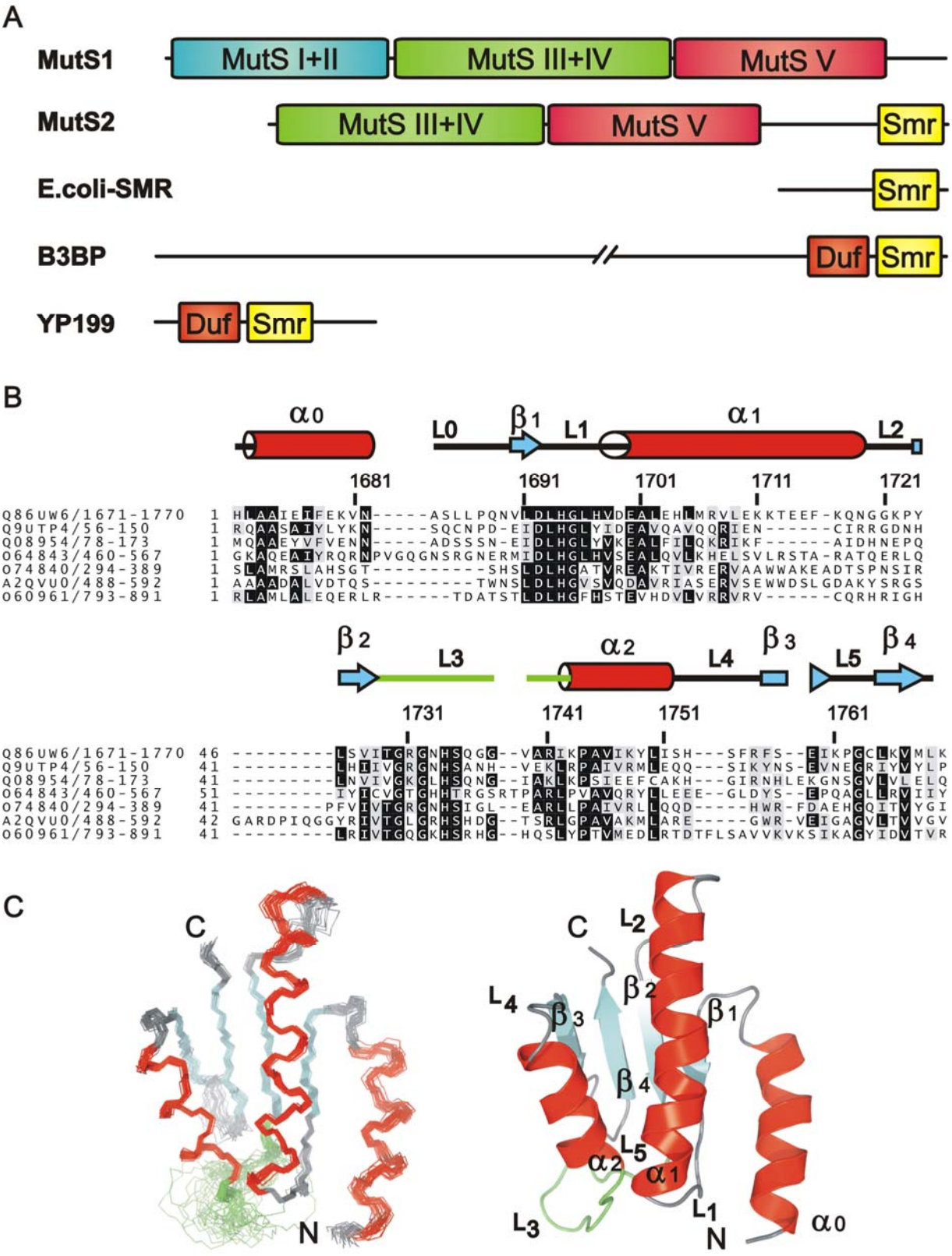
A
MutS1
MutS I+II
MutS III+IV
MutS V
MutS2
MutS III+IV
MutS V
Smr
E.coli-SMR
Smr
B3BP
Duf
Smr
YP199
Duf
Smr
B
α0
L0
β1
L1
α1
L2
1681
1691
1701
1711
1721
Q86UW6/1671-1770 1 HLAAIETFEKVN-----ASLLPQNVLDLHGLHVDEALEHLMRVLEKKTEEF-KQNGGKPY
Q9UTP4/56-150 1 RQAASAIYLYKN-----SQCNPD-EIDLHGLYIDEAVQAVQQRIEN------CIRRGDNH
Q08954/78-173 1 MQAAEYVFVENN-----ADSSSN-EIDLHGLYVKEALFILQKRIKF------AIDHNEPQ
O64843/460-567 1 GKAQEAIYRQRNPVGQGNSRGNERMIDLHGLHVSEALQVLKHELSVLRSTA-RATQERLQ
O74840/294-389 1 SLAMRSLAHSGT----------SHSLDLHGATVREAKTIVRERVAAWWAKEADTSPNSIR
A2QVU0/488-592 1 AAAADALVDTQS---------TWNSLDLHGVSVQDAVRIASERVSEWWDSLGDAKYSRGS
O60961/793-891 1 RLAMLALEQERLR------TDATSTLDLHGFHSTEVHDVLVRRVRV------CQRHRIGH
β2
L3
α2
L4
β3
L5
β4
1731
1741
1751
1761
Q86UW6/1671-1770 46 ---------LSVITGRGNHSQGG--VARIKPAVIKYLISH---SFRFS--EIKPGCLKVMLK
Q9UTP4/56-150 41 ---------LHIIVGRGNHSANH--VEKLRPAIVRMLEQQ---SIKYNS-EVNEGRIYVYLP
Q08954/78-173 41 ---------LNVIVGKGLHSQNG--IAKLKPSIEEFCAKH---GIRNHLEKGNSGVLVLELQ
O64843/460-567 51 ---------IYICVGTGHHTRGSRTPARLPVAVQRYLLEEE--GLDYS--EPQAGLLRVIIY
O74840/294-389 41 --------PFVIVTGRGNHSIGL--EARLLPAIVRLLQQD-----HWR-FDAEHGQITVYGI
A2QVU0/488-592 42 GARDPIQGGYRIVTGLGRHSHDG--TSRLGPAVAKMLARE-----GWR-VEIGAGVLTVVGV
O60961/793-891 41 ---------LRIVTGQGKHSRHG--HQSLYPTVMEDLRTDTFLSAVVKVKSIKAGYIDVTVR
C
C
N
C
L2
L4
β3
β2
β1
β4
L5
α2
α1
L3
L1
N
α0

Figure 1

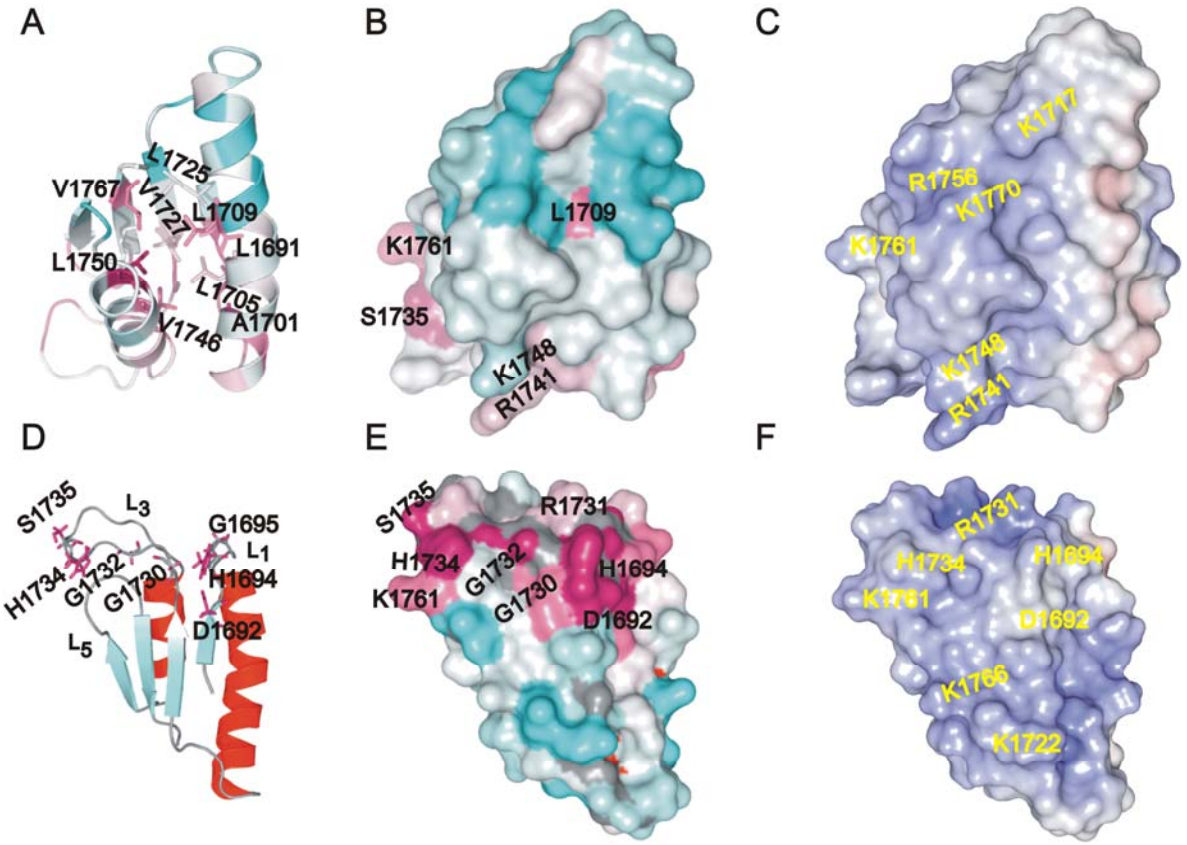
A
L1725
V1767
V1727
L1709
L1691
L1750
L1705
A1701
V1746
B
L1709
K1761
S1735
K1748
R1741
C
K1717
R1756
K1770
K1761
K1748
R1741
D
S1735
L3
G1695
L1
H1734
G1732
G1730
H1694
D1692
L5
E
S1735
R1731
H1734
G1732
H1694
K1761
G1730
D1692
F
R1731
H1734
H1694
K1761
D1692
K1766
K1722

Figure 2

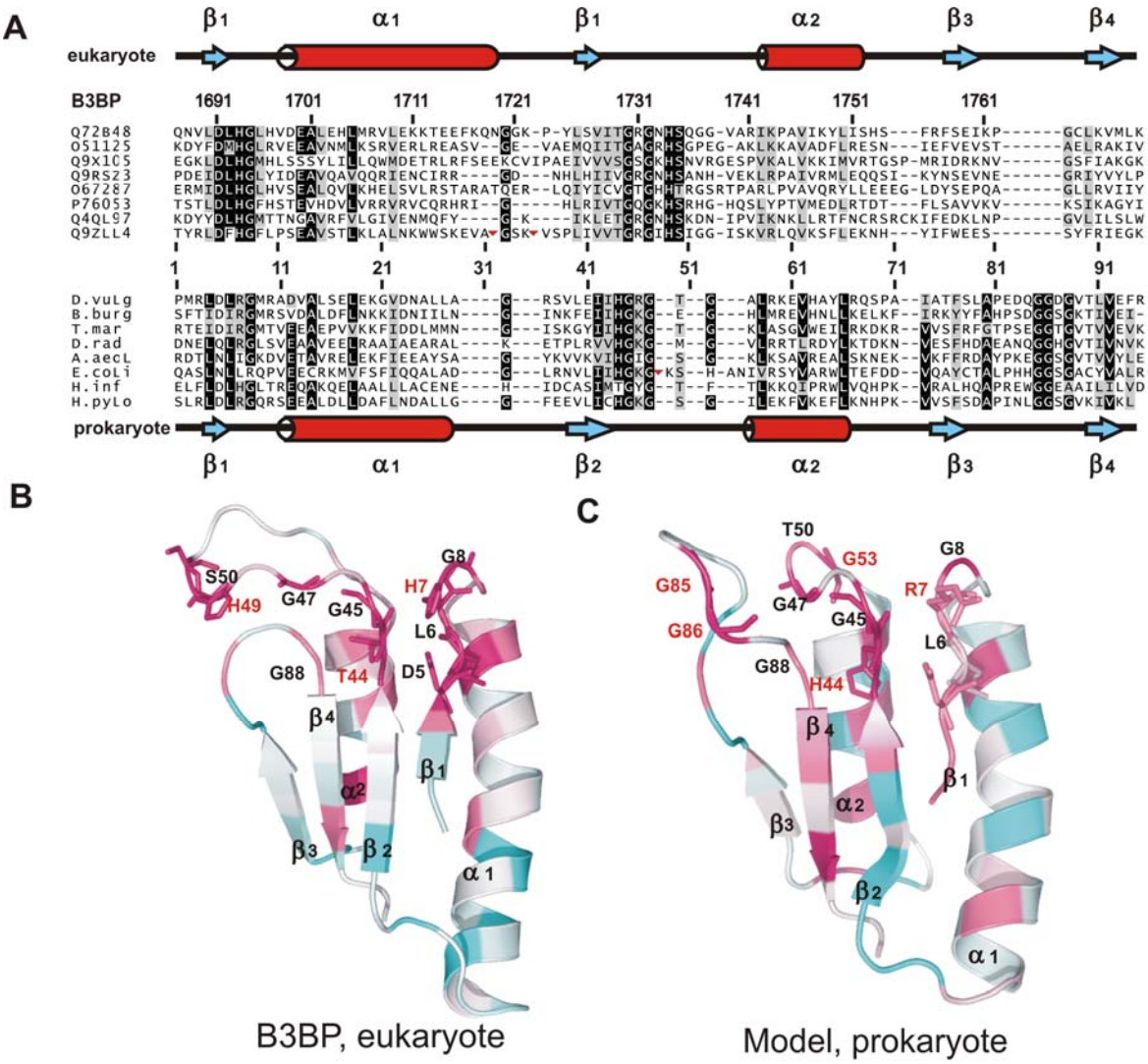
A
eukaryote
prokaryote
B3BP
B
C
B3BP, eukaryote
Model, prokaryote

Figure 3

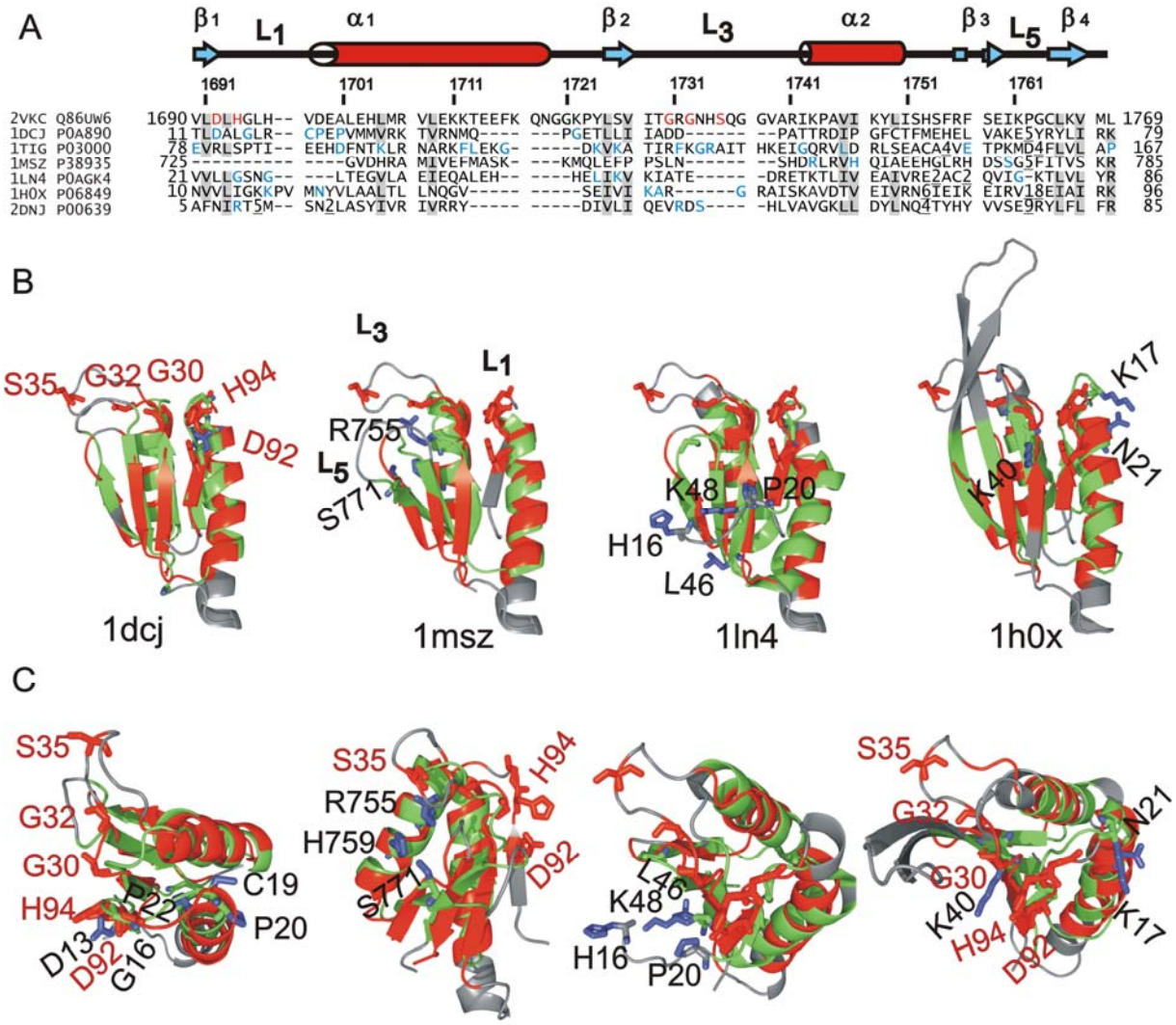
A
β1 L1 α1 β2 L3 α2 β3 L5 β4
1691 1701 1711 1721 1731 1741 1751 1761
2VKC Q86UW6 1690 VLDLHGLH-- VDEALEHLMR VLEKKTEEFK QNGGKPYLSV ITGRGNHSQG GVARIKPAVI KYLISHSFRF SEIKPGCLKV ML 1769
1DCJ P0A890 11 TLDALGLR-- CPEPVMMVRK TVRNMQ---- ---PGETLLI IADD------ --PATTRDIP GFCTFMEHEL VAKE5YRYLI RK 79
1TIG P03000 78 EVRLSPTI-- EEHDFNTKLR NARKFLEKG- -----DKVKA TIRFKGRAIT HKEIGQRVLD RLSEACA4VE TPKMD4FLVL AP 167
1MSZ P38935 725---------- ----GVDHRA MIVEFMASK- ---KMQLEFP PSLN------ --SHDRLRVH QIAEEHGLRH DSSG5FITVS KR 785
1LN4 P0AGK4 21 VVLLGSNG-- ---LTEGVLA EIEQALEH-- ----HELIKV KIATE----- -DRETKTLIV EAIVRE2AC2 QVIG-KTLVL YR 86
1H0X P06849 10 NVVLIGKKPV MNYVLAALTL LNQGV----- -----SEIVI KAR------G RAISKAVDTV EIVRN6IEIK EIRV18EIAI RK 96
2DNJ P00639 5 AFNIRT5M-- SN2LASYIVR IVRRY----- -----DIVLI QEVRDS---- -HLVAVGKLL DYLNQ4TYHY VVSE9RYLFL FR 85
B
S35 G32 G30 H94 D92
1dcj
L3 L1 R755 L5 S771
1msz
K48 P20 H16 L46
1ln4
K17 N21 K40
1h0x
C
S35 G32 G30 H94 C19 P22 P20 D13 D92 G16
S35 H94 R755 H759 D92 S771
L46 K48 H16 P20
S35 G32 N21 G30 K40 K17 H94 D92

Figure 4

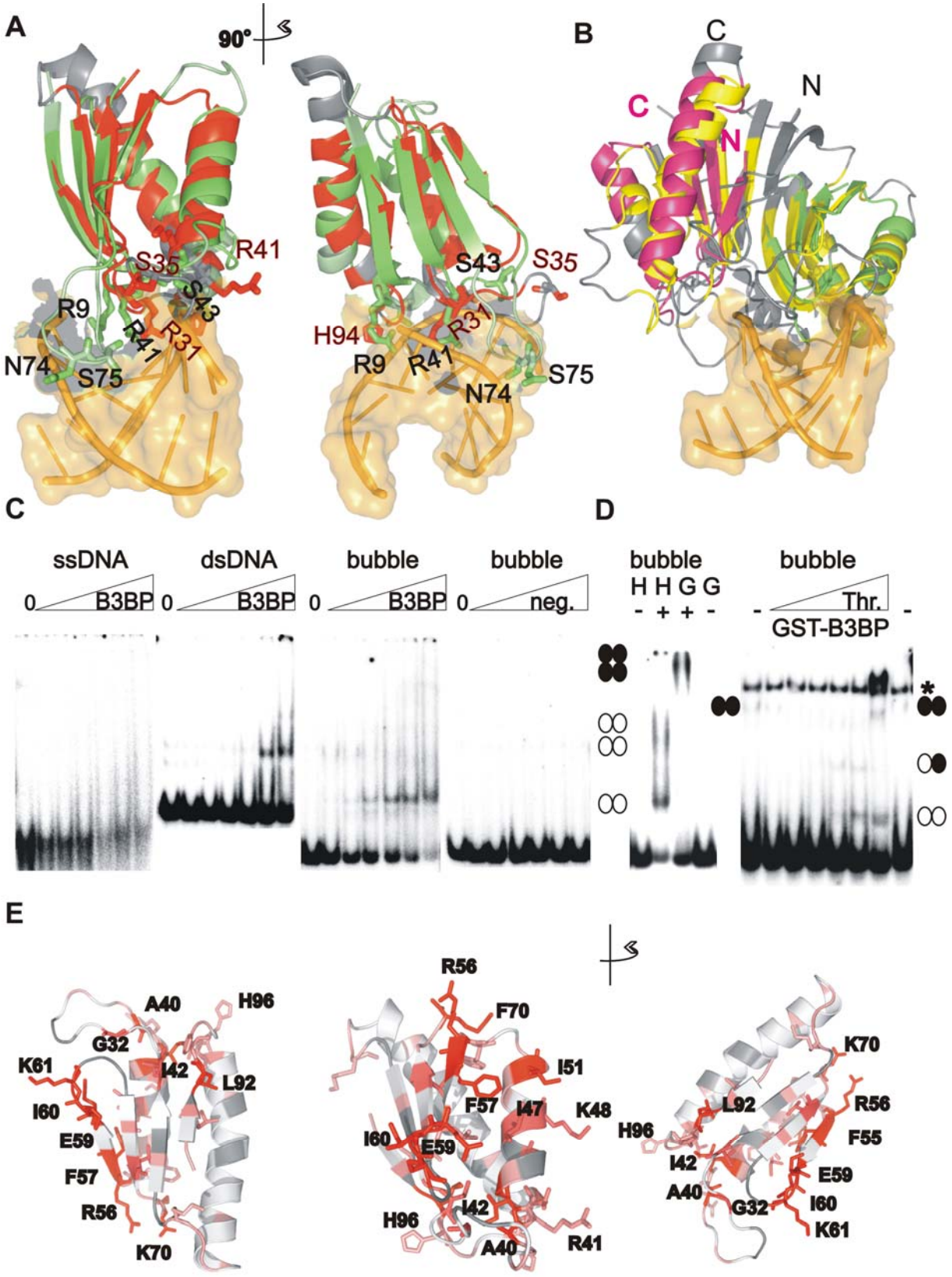
A
90°
R41
S35
S43
R9
R41
R31
N74
S75
S43
S35
H94
R31
R9
R41
N74
S75
B
C
N
C
N
C
D
ssDNA
dsDNA
bubble
bubble
0
B3BP
0
B3BP
0
B3BP
0
neg.
bubble
H H G G
- + + -
bubble
- Thr. -
GST-B3BP
*
E
A40
H96
G32
K61
I42
L92
I60
E59
F57
R56
K70
R56
F70
I51
F57
I47
K48
I60
E59
H96
I42
A40
R41
K70
L92
R56
H96
F55
I42
E59
A40
G32
I60
K61

Figure 5

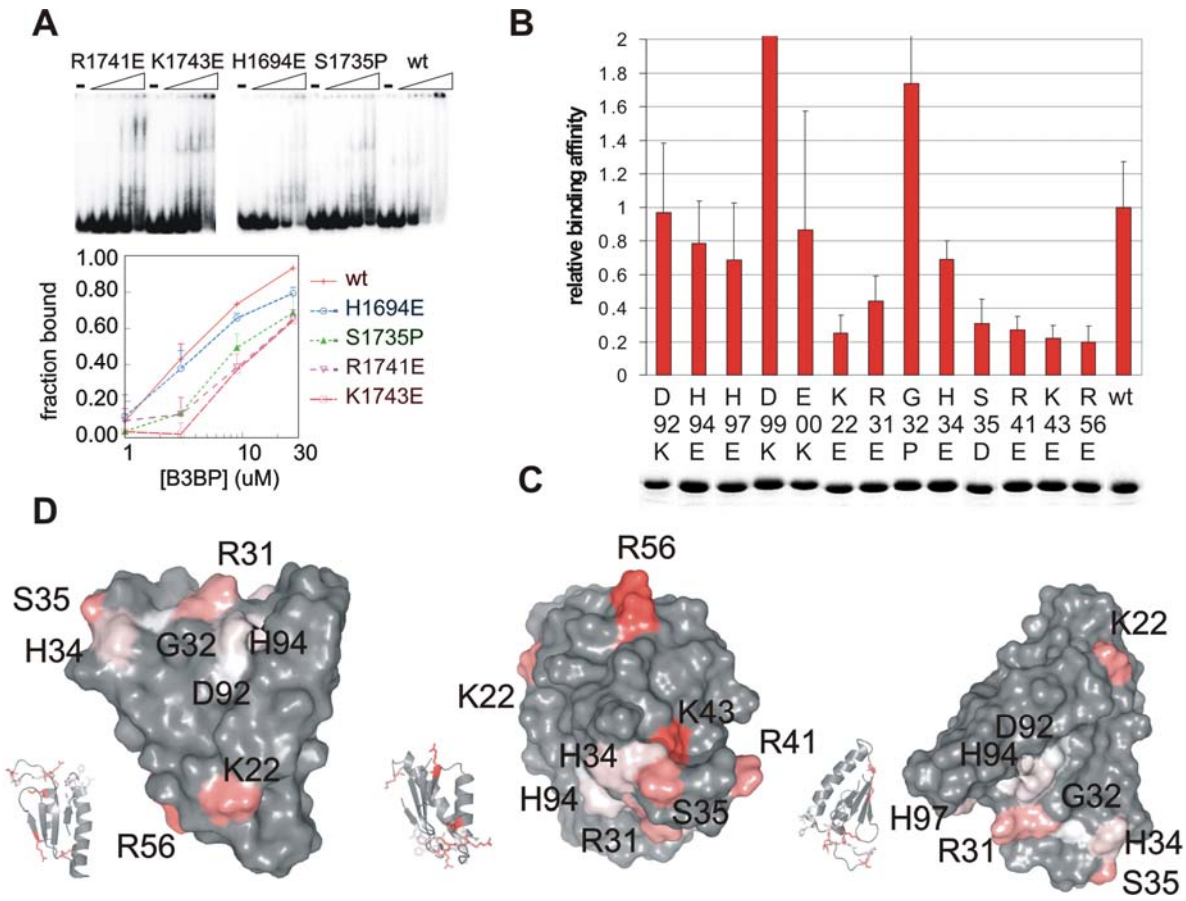
A
R1741E K1743E H1694E S1735P wt
fraction bound
[B3BP] (uM)
wt
H1694E
S1735P
R1741E
K1743E
B
relative binding affinity
D 92 K
H 94 E
H 97 E
D 99 K
E 00 K
K 22 E
R 31 E
G 32 P
H 34 E
S 35 D
R 41 E
K 43 E
R 56 E
wt
C
D
R31
S35
H34
G32
H94
D92
K22
R56
R56
K22
K43
R41
H34
H94
S35
R31
K22
D92
H94
G32
H97
R31
H34
S35

Figure 6