Structure, Dynamics, and RNA Interaction Analysis of the Human SBDS Protein

Juliana Ferreira de Oliveira[1], Maurício L. Sforça[1], Tharin M. A. Blumenschein[2], Mauricio B. Goldfeder[3], Beatriz G. Guimarães[1], Carla Columbano Oliveira[3], Nilson I. T. Zanchin[1] and Ana-Carolina Zeri[1]*

[1]*Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS Rua Giuseppe Maximo Scolfaro 10000, PO Box 6192, CEP 13083-970 Campinas, SP, Brazil*

[2]*School of Chemistry, University of East Anglia, Norwich NR4 7TJ, UK*

[3]*Department of Biochemistry, Institute of Chemistry, University of São Paulo, 05508-000 São Paulo, SP, Brazil*

Edited by M. F. Summers

Shwachman–Bodian–Diamond syndrome is an autosomal recessive genetic syndrome with pleiotropic phenotypes, including pancreatic deficiencies, bone marrow dysfunctions with increased risk of myelodysplasia or leukemia, and skeletal abnormalities. This syndrome has been associated with mutations in the *SBDS* gene, which encodes a conserved protein showing orthologs in Archaea and eukaryotes. The Shwachman–Bodian–Diamond syndrome pleiotropic phenotypes may be an indication of different cell type requirements for a fully functional SBDS protein. RNA-binding activity has been predicted for archaeal and yeast SBDS orthologs, with the latter also being implicated in ribosome biogenesis. However, full-length SBDS orthologs function in a species-specific manner, indicating that the knowledge obtained from model systems may be of limited use in understanding major unresolved issues regarding SBDS function, namely, the effect of mutations in human SBDS on its biochemical function and the specificity of RNA interaction. We determined the solution structure and backbone dynamics of the human SBDS protein and describe its RNA binding site using NMR spectroscopy. Similarly to the crystal structures of Archaea, the overall structure of human SBDS comprises three well-folded domains. However, significant conformational exchange was observed in NMR dynamics experiments for the flexible linker between the N-terminal domain and the central domain, and these experiments also reflect the relative motions of the domains. RNA titrations monitored by heteronuclear correlation experiments and chemical shift mapping analysis identified a classic RNA binding site at the N-terminal FYSH (*f*ungal, *Y*hr087wp, *Sh*wachman) domain that concentrates most of the mutations described for the human SBDS.

© 2010 Elsevier Ltd. All rights reserved.

*Corresponding author. E-mail address: aczeri@lnls.br.
Present address: B. G. Guimarães, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 Gif-sur-Yvette Cedex, France.

Abbreviations used: SDS, Shwachman–Bodian–Diamond syndrome; rRNA, ribosomal RNA; AfSBDS, *Archaeoglobus fulgidus* SBDS ortholog; PDB, Protein Data Bank; mthSBDS, *Methanothermobacter thermautotrophicus* SBDS ortholog; PaSBDS, *Pyrococcus abyssi* SBDS ortholog; HsSBDS, *Homo sapiens* SBDS ortholog; SHHP, strand helix hairpin; RRM, RNA recognition motif; NOE, nuclear Overhauser enhancement; HSQC, heteronuclear single quantum coherence; NOESY, NOE spectroscopy; TOCSY, total correlated spectroscopy.

Introduction

Currently, five human genetic syndromes, which include Diamond–Blackfan anemia, cartilage–hair hypoplasia, Treacher Collins syndrome, dyskeratosis congenita, and Shwachman–Bodian–Diamond syndrome (SDS), are associated with loss-of-function mutations in genes encoding proteins involved in ribosome biogenesis.[1] Mutations in *SBDS* gene were reported for 90% of patients with SDS (OMIM 260400). Most SDS-associated mutations are located at the N-terminal domain and seem to result from recombination between the *SBDS* gene and a pseudogene (*SBDSP*) sharing a 97% identity.[2] SDS is an autosomal recessive disorder characterized by hematological dysfunction, pancreatic exocrine insufficiency, and skeletal abnormalities, usually manifesting clinically in infancy or early childhood.[3–5] SDS is also associated with a high propensity for malignant myeloid transformation into myelodysplasia and leukemia.[6] Consistent with an essential function, targeted disruption of the *SBDS* gene leads to early embryonic lethality in mice,[7] and no homozygote for null mutation has been identified, suggesting that complete absence of the SBDS protein is lethal.

In addition to the human genetic syndrome, the importance of the *SBDS* gene for cell function is supported also by high sequence conservation (Fig. 1) and wide distribution among species, including Archaea and all eukaryotes. From a group of 159 genomes, 155 contain a single copy of the *SBDS* gene per haploid genome.[8] Regarding primary structure conservation, orthologs of SBDS can be separated into two major groups according to the extent of the C-terminus. The first group, which possesses SBDS proteins of approximately 250 amino acid residues displaying a three-domain architecture, includes Archaea, animals, and fungi.[8] The second group, which contains SBDS orthologs showing extended C-terminal regions ranging from 100 to 280 additional residues, includes plants and protists.[8] The C-terminal extension seems to play a role in RNA interaction, since several plant SBDS orthologs contain putative RNA-binding U1-type zinc fingers in the extended C-terminal region[8] and the C-terminal extension of the *Trypanosoma cruzi* SBDS ortholog behaves as a natively unfolded protein segment that mediates interaction with RNA.[9]

The implication of SBDS in ribosome biogenesis has started with genome analyses that identified archaeal SBDS orthologs within highly conserved operons that contain RNA-processing genes,[10] and transcriptional profiling has described the clustering of the yeast ortholog *SDO1* with ribosomal RNA (rRNA) processing factors.[11,12] Recent interaction studies using affinity capture, followed by mass spectrometry analysis and yeast two-hybrid assays, revealed an association of both yeast and human SBDS with ribosome structural components and ribosome biogenesis factors.[13–15] In yeast, Sdo1p is required for the accurate synthesis and nuclear export of 60S ribosome subunits.[16] Sdo1p was also implicated in the release and recycling of the nucleolar shuttling factor Tif6p from pre-60S ribosomes.[16] This is a key step in the translational activation of 60S ribosomes that involves also elongation-factor-like 1 to facilitate the release of Tif6p from late cytoplasmic pre-60S subunits.[16] As for the human SBDS, further evidence of its involvement with ribosome biosynthesis includes its nucleolar localization[17] and the global decrease in rRNA synthesis reported for cells derived from SDS patients and for human skin fibroblast knockdown for SBDS.[18] However, the phenotype resulting from SBDS downregulation may vary depending on the cell line. For example, downregulation of SBDS in HEK293 cells did not seem to significantly affect rRNA synthesis, although these cells showed an altered expression of several critical genes both at the transcriptional level and at the translational level.[19]

Structural information has also been used to infer SBDS molecular function. The structures of the *Archaeoglobus fulgidus* SBDS ortholog (AfSBDS) [Protein Data Bank (PDB) code 1T95[20]] and, more recently, of the *Methanothermobacter thermautotrophicus* SBDS ortholog (mthSBDS; PDB code 2WBM[21]) were determined by X-ray crystallography, revealing a three-domain architecture in which the N-terminal domain [termed FYSH (*f*ungal, *Y*hr087wp, *Sh*wachman)] displays a novel mixed α/β-fold and harbors the largest number of mutations related to the syndrome. The two archaeal proteins are very similar in sequence and structure, with 50% sequence identity and low RMSD between equivalent domains. The crystals of mthSBDS[21] had two molecules in the asymmetric unit and show an intrinsic interdomain flexibility. In addition, the SBDS N-terminal FYSH domain was shown to be interchangeable between several eukaryotic species.[20,22] Interestingly, the N-terminal domain shows structural identity to the yeast protein encoded by the gene *YHR087W*, which is implicated in RNA metabolism but whose exact function has not yet been determined. The central domain consists of a three-helix bundle, and the C-terminal domain with a ferredoxin-like fold sharing structural homology with known RNA binding domains was proposed to mediate RNA interaction.[20,22] Indeed, the archaeal *Pyrococcus abyssi* SBDS ortholog (PaSBDS) and yeast Sdo1p were shown to interact directly with RNA in electrophoresis mobility shift assays (Luz *et al.*[14]; J. S. Luz and C.C.O., personal communication). It is important to point out that Sdo1p does not discriminate between the different RNA sequences from the yeast pre-rRNAs and rRNAs tested *in vitro*. However, both Sdo1p and PaSBDS show preference for interaction with polyA RNA relative to the homopolymers polyC, polyU, and polyG (Luz *et al.*[14]; J. S. Luz and C.C.O., personal communication). Nevertheless, these results do not answer the question on whether SBDS orthologs bind specific cognate RNAs *in vivo*. A study based on yeast complementation assays,
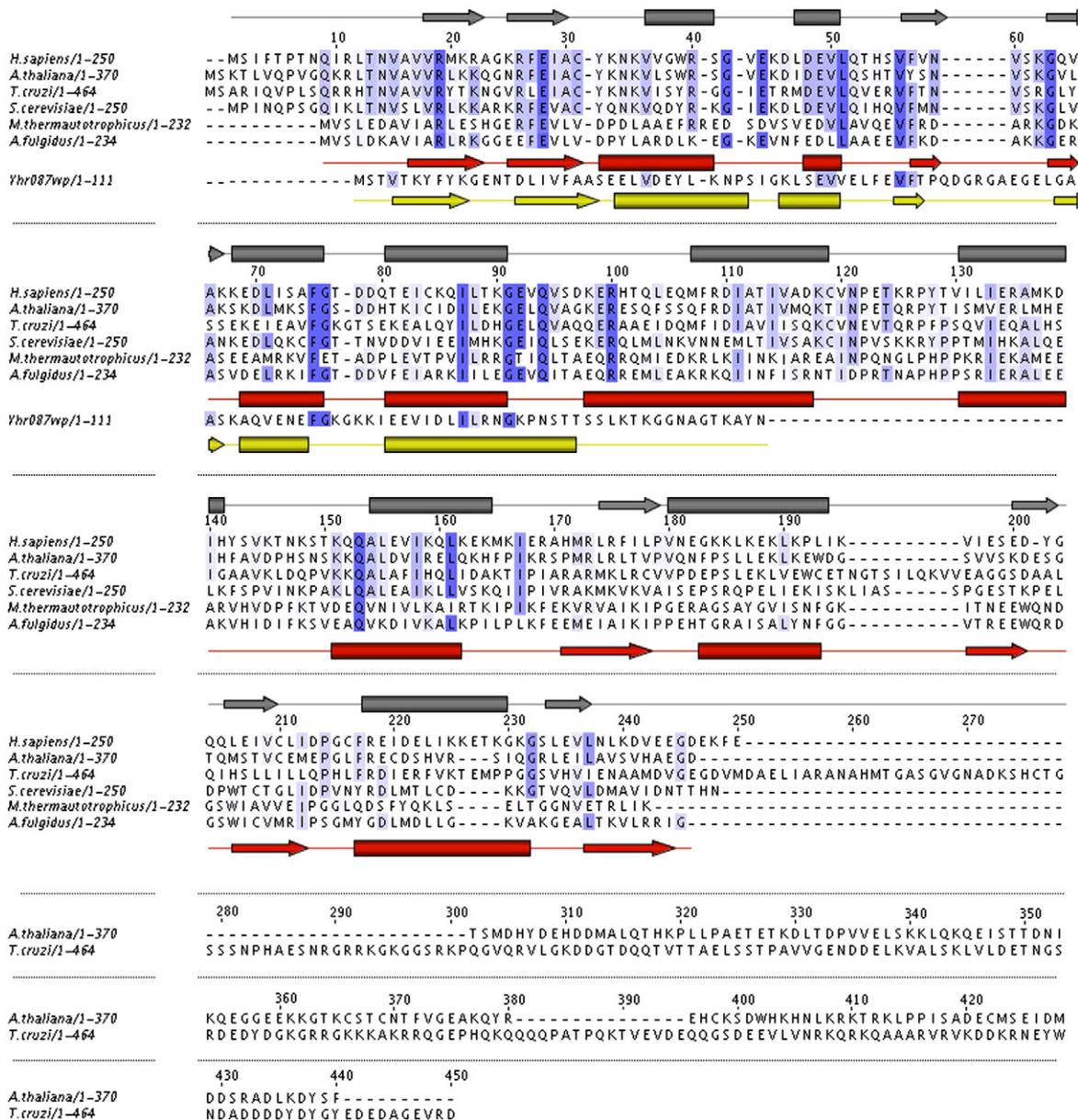
Fig. 1. Sequence alignment of SBDS orthologs and Yhr087wp protein. The sequences of the SBDS orthologs are derived from the following: human (*H. sapiens*), plant (*Ara. thaliana*), trypanosomatid (*T. cruzi*), yeast (*S. cerevisiae*), and Archaea (*A. fulgidus* and *M. thermautotrophicus*). Yhr087wp is derived from *S. cerevisiae*. Conserved residues are highlighted in blue. The alignment was generated by ClustalW. Secondary structure elements for human SBDS, AfSBDS, and Yhr087wp are shown in gray, red, and yellow, respectively. The numbering corresponds to the amino acid sequence for human SBDS.

using domain deletion and interspecies chimeric proteins, has proposed independent functions for the three domains of SBDS, although the full-length SBDS orthologs function in a species-specific manner.[8] These analyses have shown that the SBDS N-terminal FYSH domain can be interchanged between several eukaryotic species and that the C-terminus is dispensable for the complementation of a yeast deletion strain.[8] The fact that the C-terminus is dispensable raises doubts as to whether RNA interaction is actually mediated by this domain.

Despite all of the studies described above, the molecular function of the SBDS protein remains elusive. Additional open questions regard the

understanding of the RNA interaction mechanism, determining whether SBDS recognizes a specific RNA sequence *in vivo* and how mutations in the SBDS protein affect its function, leading to SDS. With this in mind, we sought to investigate the mechanism of human SBDS interaction with RNA and to correlate RNA-interacting residues with the mutations described in the *SBDS* gene of SDS patients.[20,22] We have used solution NMR techniques, which are highly suitable for this purpose. Initially, we determined the structure of the human SBDS protein, which was shown to be highly similar to the structures of *A. fulgidus* and *M. thermautotrophicus*, with three well-folded domains, although it is

important to point out that the N-terminal and central domains are connected by a flexible linker. The dynamic behavior of the protein backbone was investigated by relaxation measurements, and multidimensional heteronuclear NMR spectroscopy was used to study the interaction of SBDS with RNA oligonucleotides. Contrary to the prediction that RNA interaction might be mediated by the C-terminal domain,[20,22] the analyses performed in this work show that RNA interaction takes place via the N-terminal FYSH domain. The results presented here allowed us to identify the amino acid residues involved in RNA binding, providing the first three-dimensional insight into the mechanism of SBDS–RNA interaction.

Results

Overall structure of the human SBDS protein

We obtained an overall assignment of 85% of hydrogens and heavy atoms. The NMR structure of SBDS was determined using the program CYANA 2.1, and the 40 lowest-energy structures were further refined using CNS. An ensemble of the 20 lowest-energy conformers with the most favorable geometrical parameters was selected for analysis, and the structure calculation summary is given in Table 1.

**Table 1.** Experimental restraints and structural statistics for human SBDS

Distance restraints	
All	2673
Intraresidue	617
Sequential	921
Medium range ($I+1 < i < I+4$)	556
Long range ($i > I+5$)	522
Hydrogen bonds	57
Residual target function (Å^2)	8.49 ± 1.12
Residual constraints violations	
Distance violation >0.2 Å	0
Angle violation >5 Å	0
Van der Waals violation >0.5 Å	0
Average pairwise	
RMSD N-terminal domain (residues 9–95)	
Backbone atoms (Å)	1.20 ± 0.21
All heavy atoms (Å)	2.19 ± 0.33
RMSD central domain (residues 107–145 and 151–167)[a]	
Backbone atoms (Å)	1.91 ± 1.06
All heavy atoms (Å)	3.09 ± 1.21
RMSD C-terminal domain (residues 173–236)	
Backbone atoms (Å)	0.86 ± 0.08
All heavy atoms (Å)	1.65 ± 0.12
RMSD for secondary structure residues[b]	
Backbone atoms (Å)	0.45 ± 0.27
All heavy atoms (Å)	1.43 ± 0.52
Ramachandran plot	
Most favored region (%)	75.4
Additionally allowed region (%)	23.6
Generously allowed region (%)	1.0
Disallowed region (%)	0

[a] Regions with conformational exchange are excluded.
[b] Secondary structure residues: 18–22, 25–29, 36–41, 47–50, 56–60, 63–66, 68–75, 80–90, 107–118, 130–140, 154–164, 174–179, 180–193, 200–202, 205–209, 217–229, and 233–236.

The SBDS NMR structure contains three independent domains: the N-terminal domain (residues 1–95) consists of four α-helices and four β-strands in the sequential arrangement β1–β2–α1–α2–β3–β4–α3–α4, a strand helix hairpin (SHHP) domain;[23] a flexible linker (residues 96–106) connects the N-terminal domain to the central domain (residues 107–167), which forms a three-helical bundle (α5–α6–α7); and a short flexible linker (residues 168–172) connects the central domain to the C-terminal domain (residues 173–250), comprising a four-stranded anti-parallel β-sheet with two α-helices packing against the sheet's concave surface (β5–α8–β6–β7–α9–β8), in an RNA recognition motif (RRM) fold. The structure ensemble is depicted in Fig. 2a, where the structured domains of the 20 lowest-energy structures are independently superposed. A ribbon model of a representative structure with the flexible loop (yellow) is shown in Fig. 2c. A few interdomain nuclear Overhauser enhancements (NOEs) could be assigned, mostly between the central domain and the C-terminal domain, consistent with the higher flexibility seen for the N-terminal domain relative to the remaining parts of the protein. A few medium-sized NOE signals were identified between residues at the N-terminal domain and residues at the central domain, as shown in detail in Fig. 2d. The most striking signals are those between the side chain of Lys151 (in the central domain) and the C^α hydrogens of Ser41 and Gly42 in the N-terminal domain. This indicates that the relative motions between the domains are not completely random but somewhat constrained, as shown in Fig. 2b, where seven low-energy structures are superposed via the central domain, suggesting that the domains do not partition their time uniformly among all regions of conformational space. NOEs can be observed between flexible regions, provided they spend a significant amount of time close to each other in space. The homologous residues in *M. thermautotrophicus* and *A. fulgidus* are in the same orientation in the crystal structures.

A subset of residues could not be assigned with the triple-resonance strategy devised by Sattler *et al.*, and some lie on an 11-residue fragment (Ser96-Gln106) within the loop connecting domains 1 and 2 (Fig. 2c, highlighted in yellow).[24] Missing peaks in ^{15}N heteronuclear single quantum coherence (HSQC) experiments are characteristic of regions in the polypeptide chain undergoing conformational exchange on an intermediate NMR timescale,[25] and a search for the alternative origins of extreme line broadening for this region excluded other explanations. Published NMR studies support the conformation exchange hypothesis,[26] as well as the relaxation parameters measured for the protein, also in agreement with published results.[27] The plot of the transverse relaxation rate R_2 *versus* the residue number, as shown below, presents elevated values for residues next to the loop region, suggesting that those sites experience chemical exchange on microsecond-to-millisecond timescales.[27] The corresponding residues in the yeast protein Yrh087w, also
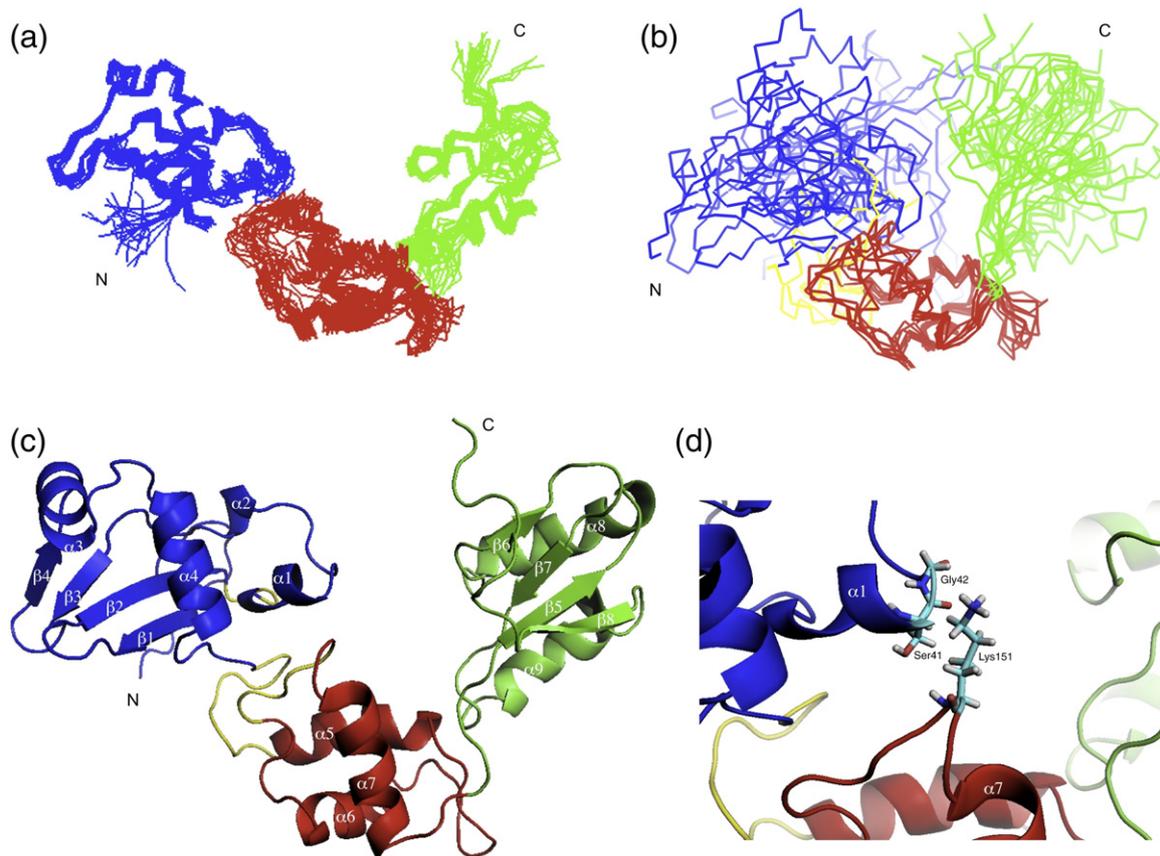
Fig. 2. Structure of human SBDS protein. (a) The 20 lowest-energy structures with folded domains independently superposed and the flexible loop left out for clarity of presentation. (b) Superposition of seven representative low-energy structures by the central domain illustrating the broad range of interdomain motions. A few NOE signals confirm the lower spread of movements between the central domain and the C-terminal domain, and dynamics data indicate apparent correlation times consistent with an almost independent and well-folded N-terminal portion. (c) Ribbon representation of the structure (third lowest-energy conformer, chosen for ease of visualization and comparison with X-ray structures), with the unassigned loop connecting the N-terminal and central domains shown in yellow. N-terminal domain, blue; central domain, red; C-terminal domain, green. (d) Zoom of the region between the N-terminal domain and the central domain showing residues that give rise to medium-intensity NOE signals.

studied by NMR[22] (PDB code 1NYN), were not assigned. This C-terminal region of Yrh087w shows very little similarity with the corresponding linker in *Homo sapiens* SBDS ortholog (HsSBDS).

Comparison of human SBDS to AfSBDS, mthSBDS, and Yhr087w structures

The first member of the SBDS protein family whose structure has been determined is the archaeal SBDS ortholog from *A. fulgidus*. Analysis of the sequence conservation for SBDS protein family orthologs (Fig. 1) reveals a high evolutionary conservation between human and *A. fulgidus* SBDS proteins (24% identity and 48% similarity).[20] In agreement with these highly conserved sequences, the structural comparison reveals that the overall folds are very similar, as can be seen for the separated domains in Fig. 3a–c. The major difference between human SBDS (in green) and AfSBDS (blue) is the connecting loop between the N-terminal domain and the central domain. While the geometry

of this loop is largely undefined in the solution structure of human SBDS, the geometry described in the X-ray crystal structure of AfSBDS might represent a conformer favored by crystal packing. The superposition of the C^α atoms for the individual N-terminal, central, and C-terminal domains of the human SBDS solution structure and the AfSBDS crystal structure resulted, respectively, in RMSDs of 1.6, 1.7, and 2.1 Å. A total of 70, 52, and 61 C^α were superposed for each domain using the software SuperPose.[28] The topologies of the domains are highly similar, with small differences in the lengths of some of the secondary structure elements. For the N-terminal domain of the lowest-energy structure, β-strands $\beta1$ and $\beta2$ and helix $\alpha1$ are slightly shorter, and strands $\beta3$ and $\beta4$ are longer for human SBDS, with a small helical segment between them. A comparison between HsSBDS and mthSBDS yields similar results, with RMSDs for the independently superposed domains being similar to those calculated for AfSBDS (1.7, 1.6, and 2.4 Å for the N-terminal, central, and C-terminal domains, respectively), and
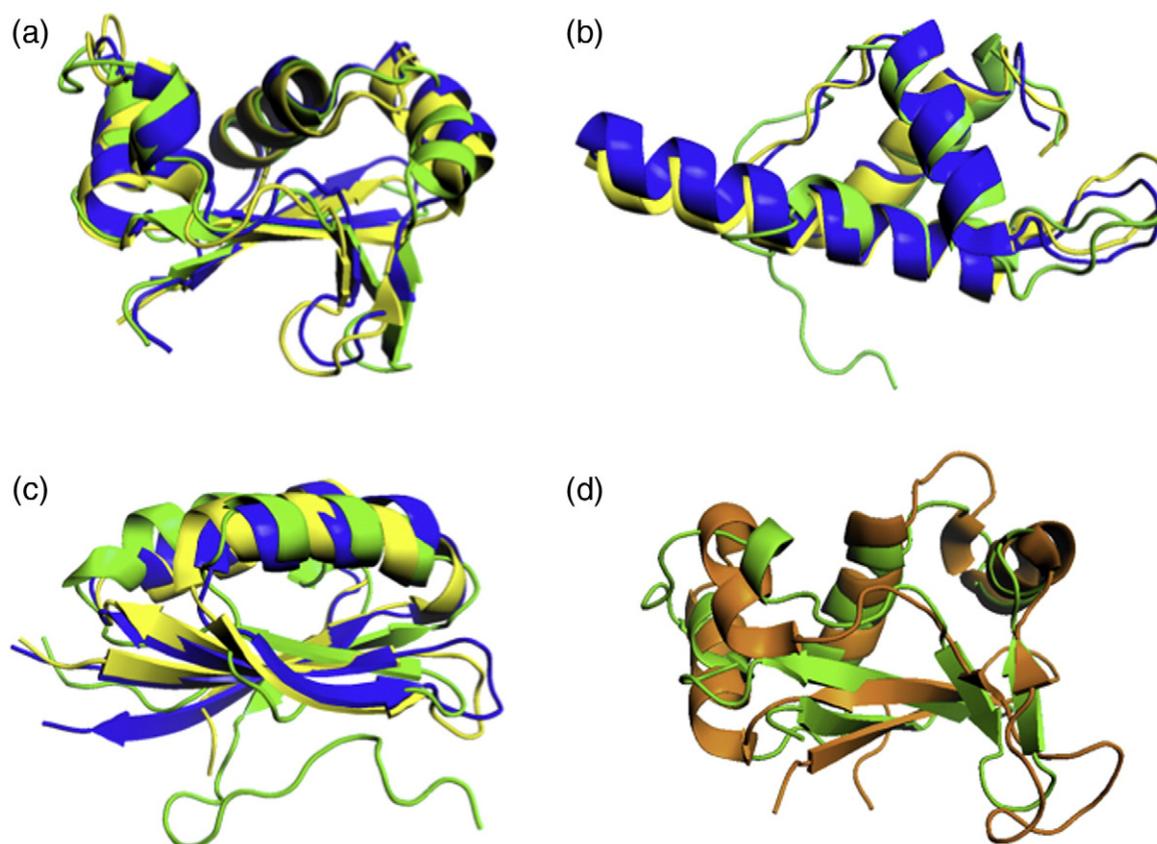
Fig. 3. Structural comparison of the domains of human SBDS with orthologs from Archaea and with yeast Yhr087w. (a–c) Structure superposition of the N-terminal domain (a), central domain (b), and C-terminal domain (c) of human SBDS (PDB code 2KDO) (green) with the respective domains of *A. fulgidus* SBDS (PDB code 1T95; blue) and *M. thermautotrophicus* (PDB code 2WBM; yellow). (d) Structure superposition of the N-terminal domain of human SBDS (green) with the yeast single-domain Yhr087w (PDB code 1NYN; orange). Figures were generated using PyMOL.

can be seen in Fig. 3a–c, where the mthSBDS domains are depicted in yellow. The superposition was performed with the software SuperPose,[28] considering 72, 53, and 56 C^α for each domain. Shammas *et al.* described the NMR structure of the *Saccharomyces cerevisiae* single-domain protein Yhr087w, which shows structural similarity to the N-terminus of AfSBDS.[20] Alignment of 67 C^α of Yhr087wp and the N-terminal domain of human SBDS (Fig. 3d) results in an RMSD of 2.0 Å. The major difference between these structures is a longer loop between strands β3 and β4 in the Yhr087w structure.

Backbone dynamics and spectral density function analysis

The backbone dynamics of SBDS, described by the relaxation rates depicted in Fig. 4, indicates that the three domains of human SBDS have different behaviors, but all have well-folded characteristics (with NOE values around 0.7) for measurements performed at 600 MHz. The most striking feature of the plots of relaxation rates is the difference in dynamic properties between the N-terminal domain and the remaining domains. While the central and C-terminal domains behave like one entity, relaxa-

tion times are significantly different from the N-terminal domain. The missing resonances for the residues in the loop connecting these domains reflect these interdomain movements, and the superposition of the structures in Fig. 2b shows that even though the domains are rigid, there are significant interdomain movements in solution.

Reduced spectral density mapping analysis[29] evaluates the spectral density function *J* at three different frequencies for each magnetic field: $J(0)$, increased by slow motions on the millisecond-to-microsecond timescale and decreased by fast internal motions on the picosecond-to-nanosecond timescale; $J(\omega_N)$, at the ^{15}N frequency, correlated with fast movements; and $J(0.87\omega_H)$, correlated with fast internal motions such as bond vector variations. In Fig. 5, the $J(0)$ frequency plot indicates increased slow movements for the second and third domains, while the N-terminal domain shows an increase in fast motions with respect to the other domains. The values for $J(0.87\omega_H)$, which indicate internal movements in the subnanosecond timescale, are very similar for the three domains, confirming their well-folded character. The plot of $J(\omega_N)$ shows the largest difference between the N-terminal domain and the central and C-terminal domains, and shows that it
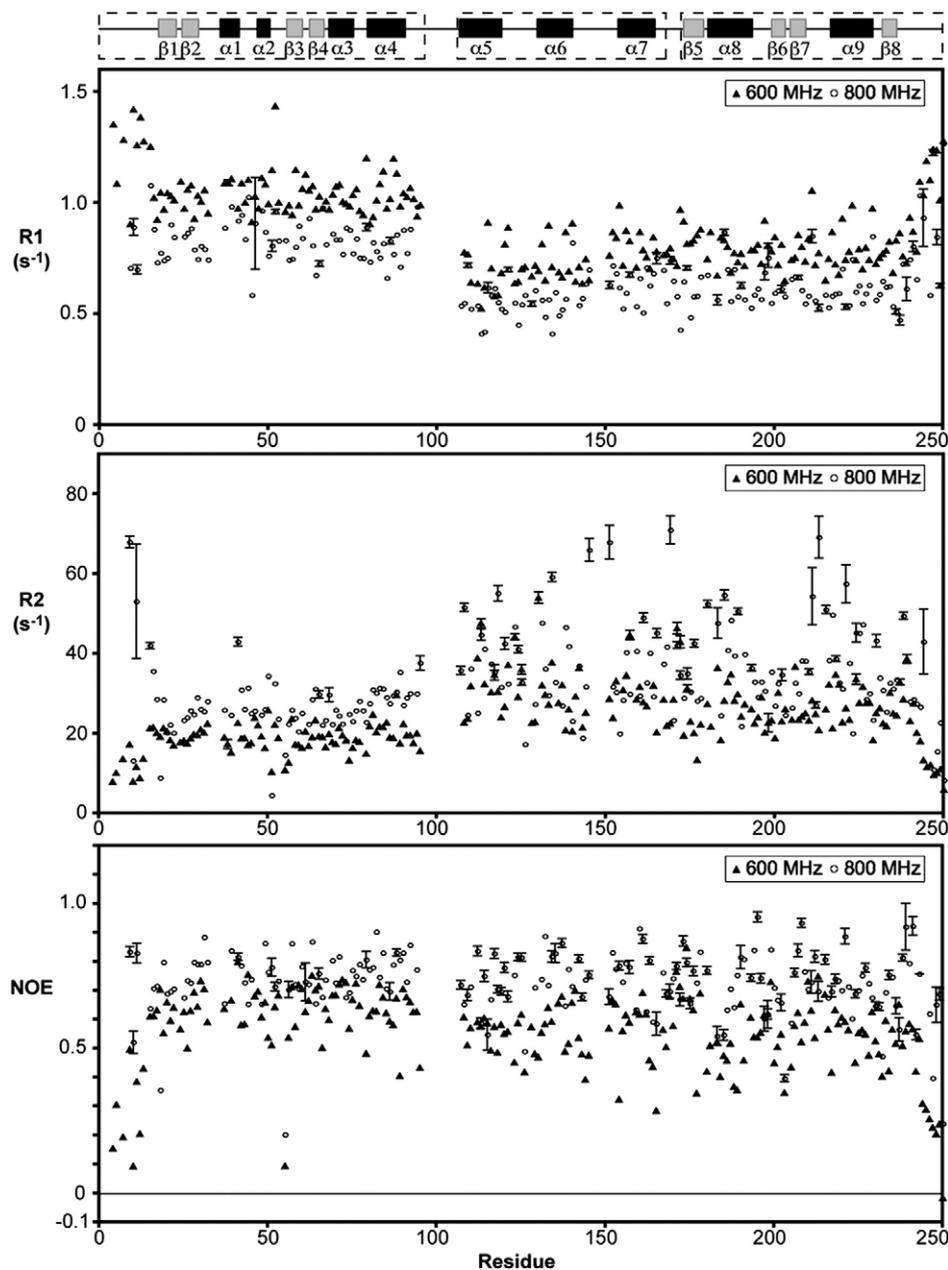
Fig. 4. Relaxation parameters *versus* the residue number. Top to bottom: R_1, R_2, and heteronuclear ^{1}H–^{15}N NOE derived from data collected at 600 and 800 MHz, plotted against the residue number. Residues for which no result is shown correspond either to prolines or to missing/overlapped cross-peaks that could not be analyzed quantitatively. Domain limits are shown by dashed boxes around the sequence representation. Error bars are smaller than the point marker and are not visible for some points.

moves independently of the other two. Apparent rotational correlation times can be calculated for the different regions of the molecule, according to the formula:[30]

$$\tau_m = \omega_N^{-1} \left(\frac{J(0) - J(\omega_N)}{J(\omega_N)} \right)^{1/2}$$

where ω_N is the Larmor frequency for ^{15}N in the magnetic field used. With application of the formula to individual residues, two distinct regions can be seen in SBDS: the N-terminal domain, with residues 15–95 presenting apparent rotational correlation times of 13.7 ± 1.4 ns at 600 MHz and 13.5 ± 1.9 at 800 MHz; and the central and C-terminal domains, with residues 107–242 presenting apparent rotational correlation times of 19.0 ± 3.9 ns at 600 MHz and 18.5 ± 3.8 ns at 800 MHz. The first 10 residues and the last 8 residues of SBDS show decreased apparent rotational correlation times, as is normally the case with the N-terminus and the C-terminus of proteins.

The larger standard deviation for the apparent rotational correlation times for the central and C-terminal domains is probably due to anisotropic
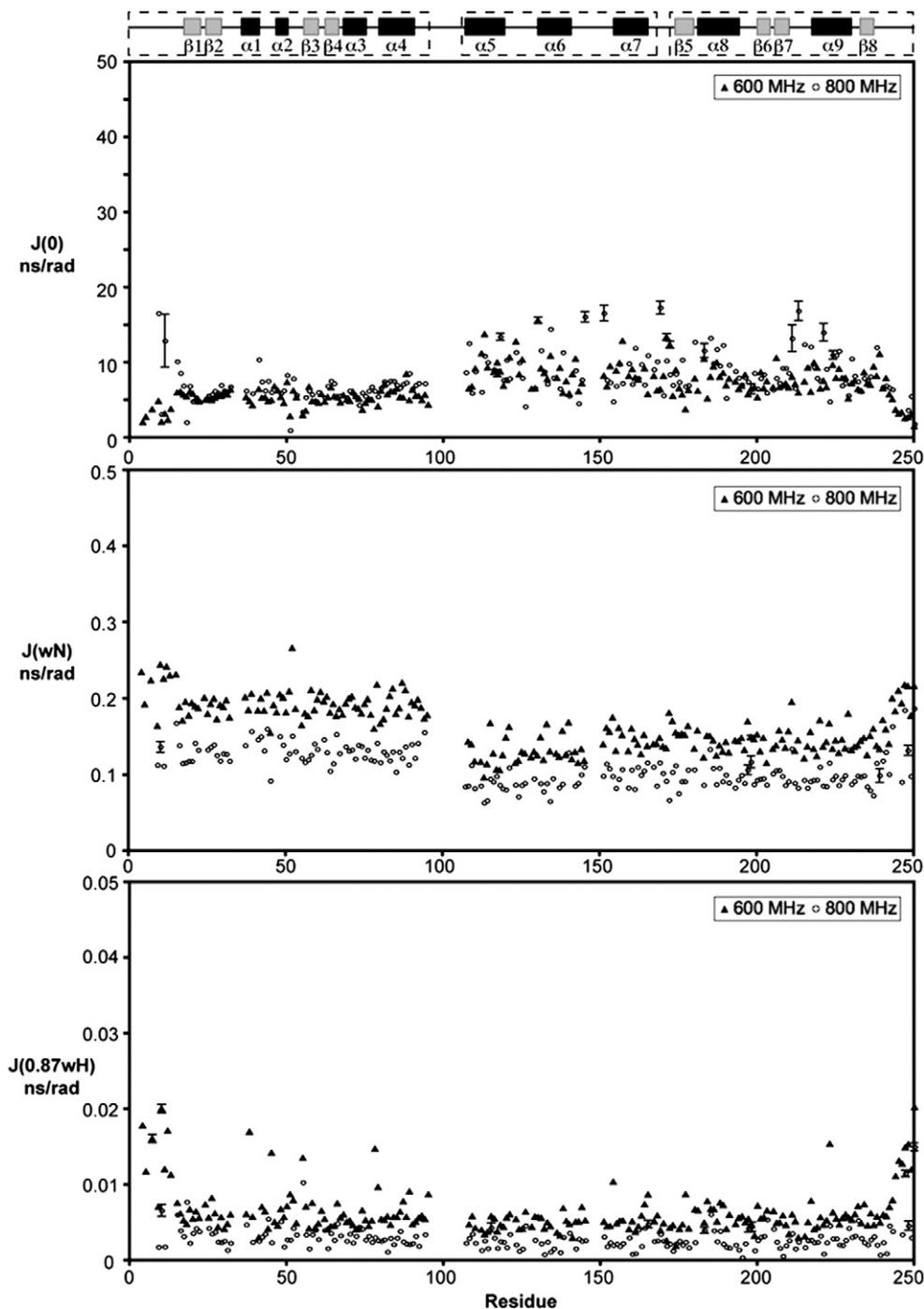
Fig. 5. Plots of reduced spectral density mapping data for human SBDS at 600 and 800 MHz. Domain limits are shown by dashed boxes around the sequence representation. Error bars are smaller than the point marker and are not visible for some points.

movements. Behaving as a single unit, the central and C-terminal domains form an elongated structure, which will tumble anisotropically in the magnetic field, leading to different apparent rotational correlation times, depending on the average orientation of each NH bond vector in the magnetic field. The N-terminal domain, on the other hand, forms a compact, virtually globular structure, changing orientation independently of the rest of the molecule in a more isotropic way.

Characterization of SBDS–RNA interaction

In order to selectively monitor SBDS during the formation of the protein–RNA complex, we used uniformly ^{15}N-labeled SBDS and recorded two-dimensional ^{15}N HSQC spectra of the protein before and after the progressive addition of small aliquots of unlabeled RNA to the NMR tube containing a constant amount of protein. Different characteristic behaviors of the human SBDS N–H cross-peaks

could be identified in the course of the titration. Figure 6a shows the full spectra; in Fig. 6b, we highlight a portion of the ^{15}N HSQC spectra of RNA/SBDS mixtures at four different points through the titration corresponding to molar ratios of 0:1, 0.25:1, 0.5:1, and 2:1. This region can be used to illustrate the various types of behavior identified. When the protein interacts with a ligand, there are changes in chemical shifts for backbone residues, indicating modifications in the chemical environment. The shift changes can be grouped into three regimes: fast, intermediate, and slow, depending on the relationship between the chemical shift difference and the rate of exchange between the free state and the bound state. If the exchange rate is higher than the chemical shift difference, a single peak appears at a position between the chemical shift of the free form and the chemical shift of the bound form. If the exchange rate is lower than the chemical shift difference, two peaks are observed (for the free state and the bound state). When the exchange rate is comparable with the chemical shift difference (the intermediate case), the peaks become broadened and may be unobservable.[25] In the SBDS–RNA complex spectra, all the resonances that change significantly are in intermediate-to-fast exchange rates. There is an overall broadening of cross-peaks with increasing amounts of RNA concentration, but some residues show a more pronounced broadening and a concomitant intensity decrease (e.g., residue Ile72 in Fig. 6d), and a few cross-peaks undergo changes in chemical shift (e.g., residue Arg26 in Fig. 6e). Most cross-peaks are essentially unperturbed even at the highest RNA concentrations, as shown in Fig. 6a. This behavior is consistent with binding reactions in which interconversion between the RNA-bound form and the free form of the protein occurs in the intermediate-to-fast NMR exchange regime, ranging from microseconds to milliseconds.[31] The experiments with DNA showed no significant perturbation of the resonances upon titration, and the titration with RNA-polyA$_{(10)}$ and RNA-polyA$_{(15)}$ (data not shown) showed perturbations in chemical shift frequencies for residues in the same region as for the RNA with mixed nucleotide sequence, but no line broadening was observed, an indication of a weaker interaction. The titration experiments offered an estimative for the dissociation constant K_d, suggesting that it is lower in the nonrepetitive sequence than in the case of RNA-polyA, indicating a higher affinity.

Amino acid residues of human SBDS involved in RNA interaction and structural mapping of the RNA binding site

Amino acids that experience changes in their chemical environment upon RNA titration are identified by an analysis of relative changes in cross-peak intensities and frequencies for all the backbone amide NH signals between the spectrum of SBDS on its own and the spectrum of the SBDS/RNA mixture as shown in the histograms presented in Fig. 6. RNA-induced chemical shift changes were measured for all cross-peaks in the SBDS ^{15}N HSQC spectrum recorded with an RNA/SBDS molar ratio of 2:1. These values are given in histogram form in Fig. 6e, which combines the changes in ^{1}H and ^{15}N chemical shifts (in ppm) into the length of the vector joining the positions in the 0:1 and 2:1 RNA/SBDS spectra, with the ^{15}N values scaled by a factor of 10 to account for the difference in gyromagnetic ratios.[25,32,33] A small number of cross-peaks showed increased chemical shift changes when compared with the remaining residues, and the broken line in Fig. 6e indicates the 0.02-ppm threshold. This group includes Lys21, Arg26, Asn59, Gly63, Val93, and Val95.

Changes in peak intensities for all amino acids could be observed throughout a complete titration series, but the largest discrimination between specific and more global changes was observed at low RNA/SBDS molar ratios (0.25:1 RNA/SBDS). The histogram in Fig. 6d shows that all peaks that are more profoundly affected by the presence of 20-mer RNA correspond to residues located in the N-terminal domain of the protein: Ile29, Gln52, Phe57, Val58, Asn59, Ile72, and Ile83.

Mapping the perturbation results of the major cross-peak broadening and chemical shift changes observed in the titration with RNA onto the three-dimensional structure of SBDS (Fig. 6c) reveals that the most strongly perturbed residues are centered on the surface of the β-sheet formed by the two SHHP motifs on the N-terminal domain of SBDS. These results suggest that RNA makes its major contact with this region. It should be noted that NMR mapping implicates not only the surface of the actual contact site because secondary effects will also be observed. The broadening of amide resonances of hydrophobic residues suggests that internal contacts may be disturbed by the interaction with RNA, as it was already described for RNA–RRM domain interactions studied by NMR.[34] The electrostatic surface model of human SBDS (Fig. 7, upper panel, left) shows a large patch of positively charged residues overlapping the region containing the amino acid residues whose cross-peaks showed large (>0.02 ppm) chemical shift changes. A prediction of RNA binding residues using the Web server RNABindR[35] indicates some of the residues in this region (Arg22, Lys25, and Arg26) as potential RNA binding sites and does not predict any residues in the central or C-terminal domain. Taken together, these data provide strong evidence that SBDS interaction with RNA takes place via the N-terminal domain.

The electrostatic surface of the HsSBDS, AfSBDS, and mthSBDS models shown in Fig. 7, calculated and plotted with MOLMOL,[36] reveals differences in the distribution of basic residues, and the patch of positively charged residues is present in human SBDS in the region showing that interaction with RNA is absent in the AfSBDS and mthSBDS structures. The surface of the C-terminal region of human SBDS shows a large acidic patch. This may

be a site for protein–protein interaction even though it has an RRM-like fold. In contrast, the surfaces of AfSBDS and mthSBDS show an acidic patch on the N-terminal region.

Further evidence of a direct interaction of HsSBDS with RNA was obtained by performing electrophoretic mobility shift assays in which purified protein was incubated with ^{32}P-labeled RNA oligonucleotides, and complexes were separated on native polyacrylamide gels. The results show that HsSBDS binds RNA *in vitro*, with a higher affinity for polyAU RNA than for polyA RNA (Fig. 8). In these assays, HsSBDS was able to bind [32P]polyrAU, which was competed off by the
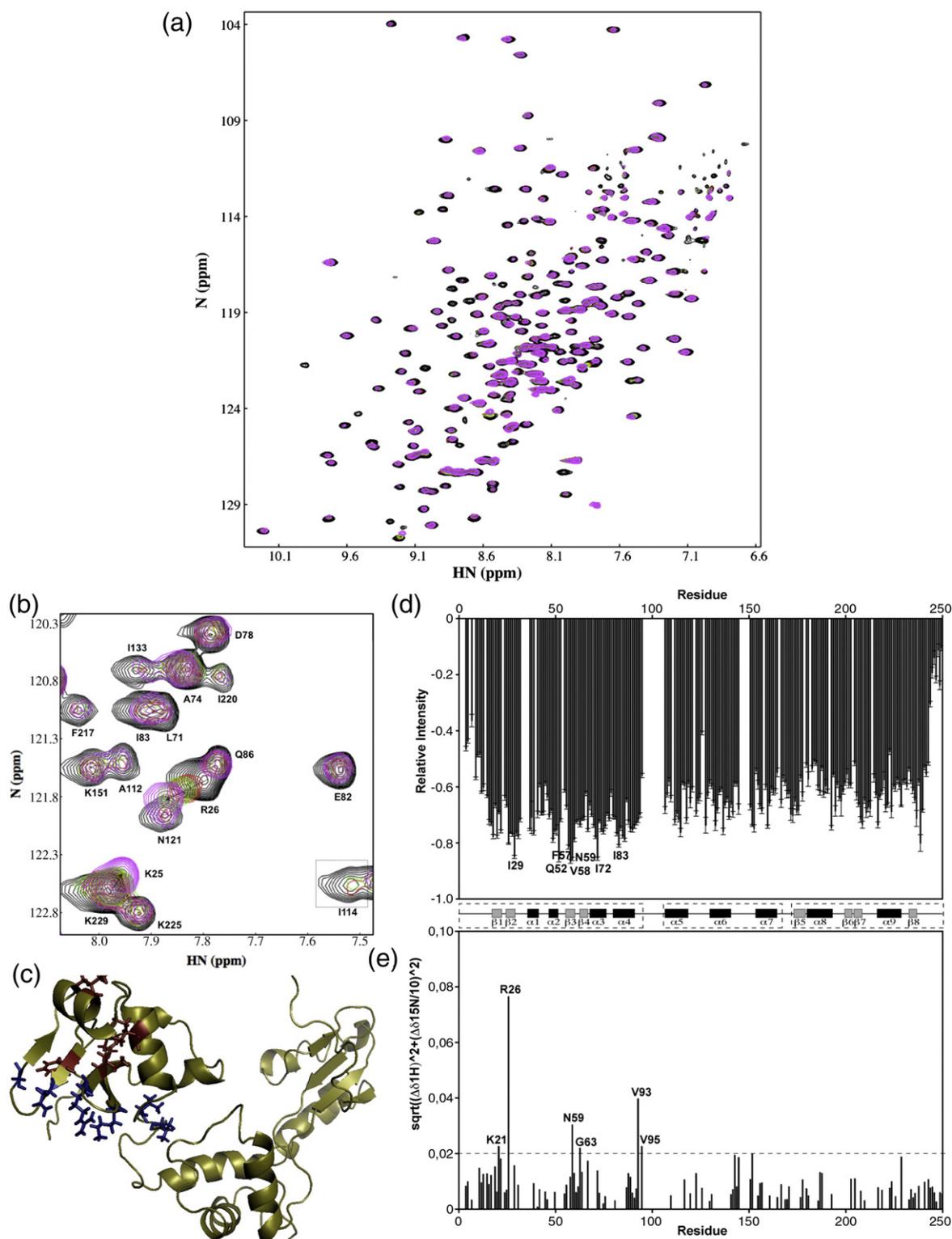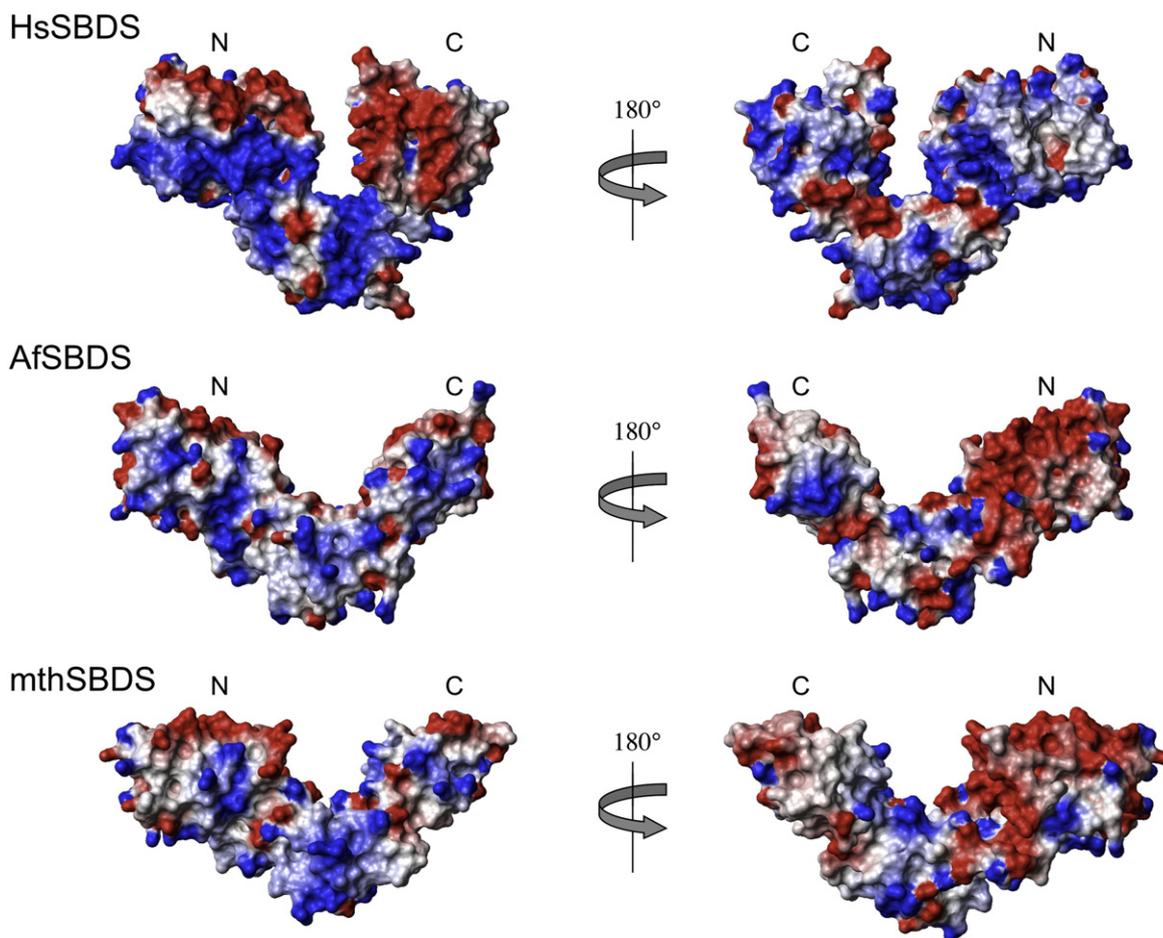

Fig. 6 (*legend on next page*)

HsSBDS

AfSBDS

mthSBDS

Fig. 7. Electrostatic surfaces of human SBDS, AfSBDS, and mthSBDS shown from blue (positive) to red (negative). Figures were generated using MOLMOL.

addition of cold polyrAU, but not by the addition of cold polyA.

Discussion

Overall structure and domain dynamics

Evidence of the modular nature of the SBDS protein family was first suggested from the *A. fulgidus* SBDS crystal structure.[20] The three-domain architecture observed for archaeal and human SBDS orthologs can be extended to the remaining orthologs of the short variant group found in Archaea, animals, and fungi, which possesses SBDS proteins of approximately 250 amino acid residues.[8] The C-terminal extension found in the SBDS of plants and protists corresponds to a fourth domain that may be responsible for additional protein–RNA contacts.[8,9] Differently from the well-folded three-domain arrangement of the ~250-residue portion,

Fig. 6. Changes in the NMR spectra of SBDS upon RNA binding. (a) Full spectra and (b) section of the ^{1}H–^{15}N HSQC spectra of free SBDS (black) superimposed on a series of spectra of the protein bound to 20-mer RNA. The peaks in red, green, and pink represent spectra with RNA/SBDS ratios of 0.25:1; 0.5:1, and 2:1, respectively. (c) Ribbon diagram of the structure of human SBDS, highlighting amino acid residues involved in interaction with RNA. Residues that exhibited cross-peak broadening effects are shown in red; residues with RNA-induced changes in cross-peak chemical shifts are shown in blue. The drawing was made with the program PyMOL. Most of the residues undergoing changes can be mapped onto the β-sheet portion of the FYSH motif in the N-terminal domain of SBDS. (d) Plot of fractional changes in peak intensities for individual backbone amide resonances of SBDS upon addition of 20-mer RNA, with an RNA/SBDS stoichiometric ratio of 0.25:1. The position of each amino acid in the primary sequence of SBDS is indicated in the abscissa; the ordinate presents the relative peak intensity change $(V - V_0)/V_0$ of each assigned cross-peak caused by the addition of RNA, where V is the peak intensity measured in the presence of a given amount of RNA and V_0 is the peak intensity measured in the free protein. Blank slots correspond to residues for which the peak intensity could not be determined because of the absence of their signal (i.e., for Pro residue and missing cross-peaks). (e) The graph shows the scaled chemical shift differences between the HSQC spectra of the protein and the 2:1 RNA/protein mixture, with the most affected residues labeled. Scaling was performed according to the equation $\Delta\delta(^{15}N + {}^1H) = [(\Delta\delta^{15}N/10)^2 + (\Delta\delta^1H)^2]^{1/2}$,[25] and the broken line indicates the 0.02-ppm threshold.
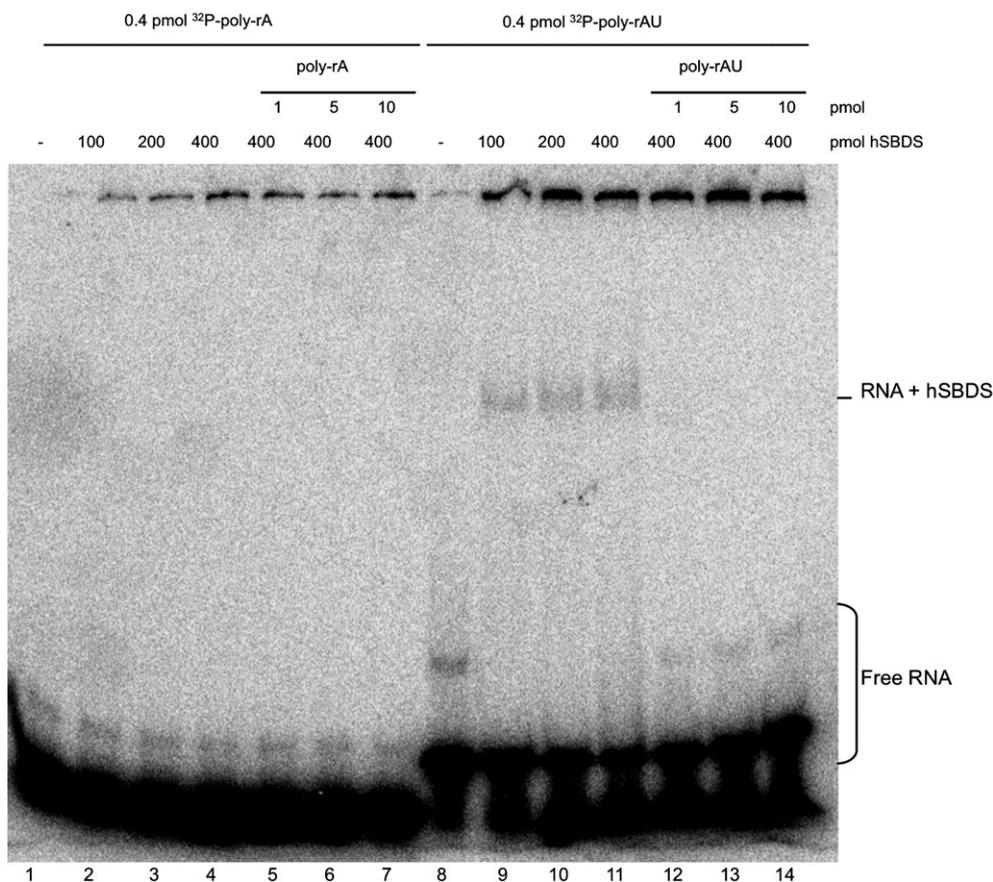
Fig. 8. Analysis of HsSBDS interaction with RNA oligonucleotides *in vitro*. Electrophoretic mobility shift assays with different radiolabeled RNA probes incubated with the indicated amounts of purified HsSBDS (100, 200, or 400 pmol). Protein was incubated with 0.4 pmol of either ^{32}P-labeled 14-mer polyrA or 21-mer polyrAU RNA oligonucleotides at 37 °C for 30 min. Competition was performed by adding unlabeled RNA oligonucleotides to the reaction. RNA/protein complexes were fractionated on 8% native polyacrylamide gels and visualized by phosphor imaging. (1) [32P]PolyrA alone; (2–4) increasing amounts of HsSBDS incubated with [32P]polyrA; (5–7) HsSBDS incubated with [32P]polyrA and increasing amounts of cold polyrA; (8) [32P]polyrAU alone; (9–11) increasing amounts of HsSBDS incubated with [32P]polyrAU; (12–14) HsSBDS incubated with [32P]polyrAU and increasing amounts of cold polyrAU. Bands corresponding to free RNA and RNA bound to HsSBDS are indicated on the right-hand side.

evidence from experimental analyses performed with the *T. cruzi* ortholog[9] and from theoretical predictions indicates that the C-terminal extension of the longer SBDS orthologs behaves as an intrinsically disordered protein domain. The modular nature of SBDS was supported by evidence from genetic complementation assays of a Δ*sdo1* strain. These assays have shown that, except for the closely related *Schizosaccharomyces pombe* ortholog, full-length SBDS proteins function in a species-specific manner,[8] although the N-terminal domain of Sdo1p can be exchanged with the equivalent domain of several eukaryotic species (*Homo sapiens, Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana*, and *Leishmania major*). The results described by Boocock *et al.* have also led to the conclusions that the central domain confers species specificity and that the C-terminal domain can be dispensable to complement this particular yeast strain.[8]

The major difference between the human SBDS structure and the AfSBDS structure involves the flexibility of the connecting loops, especially the 11-residue loop connecting the N-terminal domain and the central domain. In the solution structure of human SBDS, the geometry of this loop is largely undefined, whereas the geometry described for the X-ray crystal structure of AfSBDS might represent a conformer favored by crystal packing. The interdomain flexibility shown by the NMR data presented here and in the recently published archaeal protein mthSBDS (PDB code 2WBM) suggests that structural flexibility may be essential for the conformational adjustments of a modular protein that interacts simultaneously with different targets. Yeast and human SBDS proteins are indeed part of complexes containing proteins and RNA components of ribosomes and factors involved in biogenesis.[13–15] Direct protein–protein interaction has been demonstrated for yeast Sdo1p and Nip7p.[14] SBDS interacts also with RNA; therefore, we have to consider also the possibility that it plays the role of an adaptor mediating interaction between proteins and RNAs. Among the protein interactions identified so far that are common to human and yeast SBDS orthologs,

there are RPL3, Nip7p, and the RNA helicase Prp43p, and its human ortholog DDX21.[13,15,19]

Linker loop flexibility is a common feature found in modular proteins, especially in those involved in nucleic acid binding. For instance, sex-lethal and PABP proteins bind to single-stranded RNA using two RNA binding domains separated by a short 10-amino-acid to 12-amino-acid linker,[37,38] and hnRNPA1 binds to single-stranded DNA or RNA using two domains separated by a 17-amino-acid linker.[39] Nucleolin interacts with a stem loop located at the 5′ external transcribed sequence of the mammalian pre-rRNA using two domains of the RRM type connected by a short four-residue linker.[40] Similarly to human SBDS, the RNA binding domains of nucleolin show essentially the same tertiary structure both in the free protein and in complex with RNA, while the linker in nucleolin, which is flexible in the free protein, assumes an ordered and more stable conformation upon interaction with the RNA stem loop.[40] In the case of the polypyrimidine tract binding protein 1, which contains four RRM-type domains separated by three long linkers of 51, 91, and 23 residues, respectively, although the structures of individual domains are highly similar in both the free form and the RNA-bound form,[41] RRM domains 3 and 4 connected by the long flexible linker tumble independently in solution, having no fixed relative orientation.[42]

RNA interaction analysis

The C-terminal domain of human SBDS displays a ferredoxin-like fold typical of RRM motifs, which has earlier led to the hypothesis that it would be the domain mediating SBDS interaction with RNA.[20,22] However, no residue in the C-terminal domain shows significant chemical shift changes in the presence of RNA. This indicates that the C-terminal domain shows an extremely low affinity for RNA, and it may play a function in protein interaction. Different RRM domains interact with RNAs with different affinities[41] and, more importantly, RRM2 of the human polypyrimidine tract binding protein 1 has been implicated in protein–protein interaction.[43] Similarly, the human PACT protein is formed by three RRM domains, with RRM1 showing high RNA and with RRM2 showing intermediate binding affinities, whereas RRM3 does not interact with RNA but mediates protein–protein interaction.[44,45] Therefore, it is possible that the C-terminal domain of human SBDS might not be directly involved in RNA interaction.

The finding that the C-terminus is dispensable for the complementation of a yeast Δsdo1 strain[8] indicated that, if RNA interaction is essential for SBDS function, it should take place by a different domain. Indeed, the solution NMR analyses performed in this study identified chemical shift changes mainly in residues located at the N-terminal domain upon incubation of SBDS with RNA, and no residue in the C-terminal domain shows significant changes in the presence of RNA. Interestingly, our findings confirm with high accuracy a previous statement by Wessels *et al.*, who, based on the crystallographic and NMR structures of AfSBDS and yeast Yhr087w, respectively, proposed that the cleft between the two SHHP motifs in the N-terminal domain was the site for interaction with ribonucleoprotein complexes.[23]

The pattern of chemical shift changes in the N-terminal domain residues identified by RNA titration is consistent with electrostatic surface analysis, which revealed a patch of positively charged residues in the same region. Taken together, these results indicate that SBDS–RNA interaction may be favored, at least in part, by electrostatic interactions between positively charged residues on SBDS and negatively charged phosphates of the RNA backbone. Regarding stacking interactions that usually involve phenylalanine and tyrosine residues, it is not clear whether they can take place because NMR mapping implicates not only the surface of the actual contact site. The broadening of amide resonances of hydrophobic residues (Ile29, Gln52, Phe57, Val58, Asn59, Ile72, and Ile83) may be better interpreted as internal contacts that are disturbed by the interaction with RNA, as it was already described for RNA–RRM domain interactions studied by NMR.[34]

The yeast protein Yhr087w, homologous to the N-terminal region of SBDS, has been proposed to play a role in protein–protein interaction.[20,22] On the other hand, the N-terminal region of the *T. cruzi* ortholog TcSBDS[9] did not interact with RNA in electrophoresis mobility shift assays. Instead, TcSBDS–RNA interaction is mediated by the extra 200 residues in the unstructured C-terminal domain,[9] suggesting that these proteins evolved to assume slightly different functions while conserving the same overall fold and high sequence similarity. Possibly, the U1-type zinc finger domain found in plants and the intrinsically disordered RNA-binding C-terminus of Trypanosomatidae SBDS[9] mediate additional contacts with RNA, although the possibility that SBDS might play additional functions in RNA metabolism in these organisms cannot be ruled out yet based only on the set of information available to date.

Recent data published on the yeast ortholog of SBDS Sdo1p show that although it does not require any specific RNA sequence for binding, it has a higher affinity for longer polyA and polyAU RNAs in electrophoresis mobility shift assays.[14] Similar results were observed for the *P. abyssi* ortholog (J. S. Luz and C.C.O., personal communication). These results are corroborated by recently published data on archaeal mthSBDS that show no sequence specificity for binding.[21] Interestingly, no difference could be detected in our initial experiments comparing 10-residue *versus* 15-residue oligoadenylate sequences for interaction with the human protein by NMR. However, *in vitro* RNA binding assays shown here (Fig. 8) indicate that, similarly to yeast and archaeal orthologs, HsSBDS has a higher affinity for AU-rich RNA sequences. The actual RNA target sequence of human SBDS is not known, but the data

presented in this work show that it has a higher affinity for the RNA with a mixed nucleotide sequence than for the corresponding single-stranded DNA sequence and for the homopolymer RNA polyA. The residues perturbed are in the same region for both RNA sequences. Line broadening caused by the intermediate exchange regime is not sufficient for a precise calculation of the affinity constant for RNA, but it was possible to see an increase in binding strength for the nonrepetitive RNA sequence. The line broadening also hampered further dynamics and orientation studies with the protein–RNA complex aimed at mapping changes in domains, movements, and relative orientations.

Mutations in human SBDS and implication on RNA interaction

Most of the disease-related mutations in human SBDS[46] localize to the N-terminal domain and are in close proximity to the RNA binding region. The majority of the reported mutations occur in charged residues of the domain, which could perturb the RNA interaction by changing the basic character of the surface patch, as is the case for R19Q, K33E, and K67E. Other reported mutations, such as the L71P mutation, which causes loss of protein expression and is located next to a residue perturbed by RNA interaction (Ile72), could destabilize the structure. Based on the results shown here, except for the mutations that create premature stop codons, SDS could also be caused by mutations that alter SBDS interaction with RNA. We are starting an effort to map and compare the RNA interaction properties of other members of this protein family in order to gain insights into its function.

Materials and Methods

Sample preparation

SBDS cDNA was amplified by PCR from a human fetal brain cDNA library using the primers ONZ299 (5′ GGG CAT ATG TCG ATC TTC ACC CCC ACC 3′) and ONZ301 (5′ GGG GGA TCC TCA TTC AAA TTT CTC ATC TCC 3′) and inserted into the NdeI-BamHI restriction sites of plasmid pET-TEV[47] to generate pET-SBDS. This plasmid includes an N-terminal His tag, followed by a TEV protease cleavage site. pET-SBDS was transformed into BL21(DE3)*slyD⁻ Escherichia coli* expression cells and grown at 37 °C in M9 minimal medium supplemented with 50 μg/mL kanamycin, 4 g/L [^{13}C]glucose, and 1 g/L [^{15}N]ammonium chloride. Expression was induced at an OD_{600} of 0.8–1.0 for 4 h by adding IPTG to a final concentration of 0.5 mM. Triply labeled (^{15}N, ^{13}C, and ^{2}H) SBDS was prepared as described by Li *et al.*, with 12 h of induction at 30 °C.[48] The cells were harvested by centrifugation at 5000*g* for 10 min and suspended in buffer A [50 mM Tris–HCl (pH 7.2), 500 mM NaCl, 10 mM β-mercaptoethanol, and 1 mM PMSF]. After lysozyme treatment (50 mg/mL for 30 min on ice) and sonication, the total extract was subjected to streptomycin sulfate precipitation (2% wt/vol) for 1 h at 4 °C, with agitation.

The supernatant fraction was isolated by centrifugation (23,000*g*) for 30 min at 4 °C, and the His–SBDS protein was purified by metal-chelating affinity chromatography on a HiTrap Chelating HP column (GE Healthcare). The adsorbed protein was eluted with a 75-mL gradient of 0–50% buffer B (buffer A + 300 mM imidazole), followed by a 25-mL gradient of 50–100% buffer B. The histidine tag was removed by digestion with 15% wt/wt TEV protease at 25 °C for 24 h in buffer C [50 mM Tris–HCl (pH 8.0), 0.05 mM ethylenediaminetetraacetic acid, 1 mM 1,4-dithiothreitol (DTT), and 50 mM NaCl]. Two linker residues were left at the N-terminus after the cleavage. The cleaved SBDS protein was purified further on a HiTrap Heparin HP column (GE Healthcare) with buffer D as binding buffer [50 mM Tris–HCl (pH 7.2), 20 mM NaCl, and 1 mM DTT] and eluted with a 0–100% gradient of buffer D containing 1 M NaCl in 100 mL. Fractions containing pure SBDS were pooled and concentrated to 0.35 mM in a buffer containing 20 mM sodium phosphate (pH 7.2), 50 mM NaCl, and 1 mM DTT. SDS-PAGE, circular dichroism, and dynamic light scattering reassured protein purity and identity, and protein identity was confirmed by mass spectrometry, with a sequence coverage of 46% (data not shown).

NMR spectroscopy

NMR experiments for structure determination were performed at 293 K using a Varian Inova 600-MHz spectrometer equipped with a cryogenic probe and a Bruker Avance III 800-MHz spectrometer (at the University of East Anglia). The following experiments were recorded at 600 MHz: ^{15}N HSQC; ^{15}N-edited NOE spectroscopy (NOESY) (80 ms of mixing time, chosen after a build-up curve with other mixing time values); ^{15}N-edited total correlated spectroscopy (TOCSY); HNCA; HN(CO)CA; HNCACB; CBCA(CO)NH; HNCO; HN(CA)CO; HCCH-COSY; HCCH-TOCSY; CCH-TOCSY; ^{13}C HSQC; ^{15}C-edited NOESY (80 ms of mixing time); and HBHA (CO)NH.[24,49–51] Triply labeled (^{15}N, ^{13}C, and ^{2}H) SBDS was used to record the experiments HNCACB[50] and HN (CO)CACB[52] at 800 MHz. Dynamics experiments were conducted at both 600 and 800 MHz. NMR data were processed using the NMRPipe software[53] and analyzed using NMR View.[54]

Structure calculation

The structure of SBDS protein was calculated in a semi-automated iterative manner with the program CYANA version 2.1,[55] using 100 starting conformers. CYANA 2.1 protocol was applied to calibrate and assign NOE cross-peaks. After the first few rounds of automatic calculations, the NOESY spectra were analyzed again to identify additional cross-peaks consistent with the structural model and to correct misidentified NOEs. Slowly exchanging amides were identified by lyophilizing the protein from water and then dissolving it in 100% ^{2}H$_{2}$O; hydrogen-bond donors were identified by the presence of an amide peak in the ^{15}N HSQC recorded after 2 h 30 min. Hydrogen-bonding constraints were then added to the structure calculation protocol. The structures obtained were further refined by restrained minimization and molecular dynamics studies using CNS.[56] The 20 structures with the lowest target function were selected to represent the ensemble of SBDS structures. The quality of the structures was analyzed with PROCHECK-NMR,[57] and structural statistics are presented in Table 1.

Relaxation measurements

For backbone amide relaxation measurements, ^{15}N T_1, ^{15}N T_2, and heteronuclear NOE experiments were recorded in 600-MHz and 800-MHz spectrometers. T_1 relaxation delays were set to 10, 210, 410, 610, 810, 1010, 1210, 1410, 1710, 2010, and 2410 ms at 600 MHz, and to 20, 200, 500, 1000, 1500, 2000, 3000, 4000, and 6000 ms at 800 MHz. T_2 relaxation delays were set to 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, and 210 ms at 600 MHz, and to 17, 34, 68, 102, 136, 170, and 204 ms at 800 MHz. In all the experiments performed at 600 MHz, a relaxation delay of 3 s was used, while a relaxation delay of 8 s was used at 800 MHz. For heteronuclear NOE measurements performed at 600 MHz, a pair of spectra was recorded with and without proton saturation. Spectra recorded with proton saturation utilized a 5-s recycle delay, followed by a 15-s period of saturation, while spectra recorded in the absence of saturation employed a recycle delay of 15 s. For heteronuclear NOE measurements at 800 MHz, spectra with and without proton saturation were acquired in an interleaved manner, and the recycle or saturation delay was 8 s. In both cases, peak volumes were fitted to a single-exponential decay function using the program NMR View.

NMR titration

Since little is known about SBDS specificity for RNA sequences, we used a 20-mer oligoribonucleotide test sequence (RNA = 5′-UAA UAC GAC UCA CUA UAG GG-3′). A corresponding 20-mer single-stranded oligodeoxyribonucleotide (DNA = 5′-TAA TAC GAC TCA CTA TAG GG-3′) was used as control. For practical convenience, we used the sequence of the promoter of the T7 bacteriophage RNA polymerase for the oligonucleotides. Oligoribonucleotides composed of adenine base only with 10 (RNA-polyA$_{(10)}$) and 15 (RNA-polyA$_{(15)}$) residues were also used as controls. Synthetic oligonucleotides were purchased from IDT-Integrated DNA Technologies, Inc. The titrations with RNA or DNA oligonucleotides were followed by the recording of the two-dimensional ^{15}N HSQC spectra of [^{15}N]SBDS protein. Titrations were performed by direct addition of small aliquots (1–16.3 µL) of the 20-mer oligonucleotide (15 mM) to the NMR tube containing 350 µL of [^{15}N]SBDS (0.35 mM) to obtain titration points of oligonucleotide/SBDS at ratios of 0:1, 0.25:1, 0.5:1, 1:1, and 2:1. Throughout all NMR experiments, the temperature was maintained at 293 K. Binding of oligonucleotides was characterized by changes in protein ^{15}N HSQC signals, intensities, and chemical shift values as a function of the concentration of unlabeled nucleic acid.

RNA binding assay

RNA binding assays were carried out with 0.4 pmol of ^{32}P 5′-labeled 14-mer polyrA and 21-mer polyrAU (5′-UUAUUAUUUAUUUAUUAUUUA-3′) oligoribonucleotides (IDT-Integrated DNA Technologies, Inc.). The assays were performed in 20 mM Tris–HCl (pH 8.0), 150 mM KOAc, 5 mM Mg(OAc)$_2$, 0.2% Triton X-100, 1 mM DTT, 1 mM PMSF, and 0.8 U of RNasin. Different amounts of protein were incubated with the substrate RNA in 20 µl at 37 °C for 30 min. The samples were resolved on 8% native polyacrylamide gels and visualized on a Phosphor-Imager (Molecular Dynamics).

Accession codes

The coordinates have been deposited at the PDB under accession code 2KDO, and experimental data have been deposited at the Biological Magnetic Resonance Data Bank under accession number 16119.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.12.039

References

1. Liu, J. M. & Ellis, S. R. (2006). Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood*, **107**, 4583–4588.
2. Boocock, G. R., Morrison, J. A., Popovic, M., Richards, N., Ellis, L., Durie, P. R. & Rommens, J. M. (2003). Mutations in SBDS are associated with Shwachman–Diamond syndrome. *Nat. Genet.* **33**, 97–101.
3. Bodian, M., Sheldon, W. & Lightwood, R. (1964). Congenital hypoplasia of the exocrine pancreas. *Acta Paediatr.* **53**, 282–293.
4. Shwachman, H., Diamond, L. K., Oski, F. A. & Khaw, K. T. (1964). The syndrome of pancreatic insufficiency and bone marrow dysfunction. *J. Pediatr.* **65**, 645–663.
5. Ginzberg, H., Shin, J., Ellis, L., Morrison, J., Ip, W., Dror, Y. *et al.* (1999). Shwachman syndrome: phenotypic manifestations of sibling sets and isolated cases in a large patient cohort are similar. *J. Pediatr.* **135**, 81–88.
6. Smith, O. P., Hann, I. M., Chessells, J. M., Reeves, B. R. & Milla, P. (1996). Haematological abnormalities in Shwachman–Diamond syndrome. *Br. J. Haematol.* **94**, 279–284.
7. Zhang, S., Shi, M., Hui, C. C. & Rommens, J. M. (2006). Loss of the mouse ortholog of the Shwachman–Diamond syndrome gene (Sbds) results in early embryonic lethality. *Mol. Cell. Biol.* **26**, 6656–6663.
8. Boocock, G. R., Marit, M. R. & Rommens, J. M. (2006). Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics*, **87**, 758–771.
9. deOliveira, J. F., Castilho, B. A., Sforça, M. L., Krieger, M. A., Zeri, A. C., Guimarães, B. G. & Zanchin, N. I. (2009). Characterization of the *Trypanosoma cruzi* ortholog of the SBDS protein reveals an intrinsically

disordered extended C-terminal region showing RNA-interacting activity. *Biochimie*, **91**, 475–483.

10. Koonin, E. V., Wolf, Y. I. & Aravind, L. (2001). Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11**, 240–252.

11. Wu, L. F., Hughes, T. R., Davierwala, A. P., Robinson, M. D., Stoughton, R. & Altschuler, S. J. (2002). Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nat. Genet.* **31**, 255–265.

12. Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q. *et al.* (2003). A panoramic view of yeast noncoding RNA processing. *Cell*, **113**, 919–933.

13. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A. *et al.* (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, **440**, 637–643.

14. Luz, J. S., Georg, R. C., Gomes, C. H., Machado-Santelli, G. M. & Oliveira, C. C. (2009). Sdo1p, the yeast orthologue of Shwachman–Bodian–Diamond syndrome protein, binds RNA and interacts with nuclear rRNA-processing factors. *Yeast*, **26**, 287–298.

15. Ball, H. L., Zhang, B., Riches, J., Gandhi, R., Li, J., Rommens, J. M. & Myers, J. S. (2009). SBDS is a multifunctional protein implicated in cellular stress responses. *Hum. Mol. Genet.* **18**, 3684–3695.

16. Menne, T. F., Goyenechea, B., Sanchez-Puig, N., Wong, C. C., Tonkin, L. M., Ancliff, P. J. *et al.* (2007). The Shwachman–Bodian–Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat. Genet.* **39**, 486–495.

17. Austin, K. M., Leary, R. J. & Shimamura, A. (2005). The Shwachman–Diamond SBDS protein localizes to the nucleolus. *Blood*, **106**, 1253–1258.

18. Ganapathi, K. A., Austin, K. M., Lee, C. S., Dias, A., Malsch, M. M., Reed, R. & Shimamura, A. (2007). The human Shwachman–Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood*, **110**, 1458–1465.

19. Hesling, C., Oliveira, C. C., Castilho, B. A. & Zanchin, N. I. (2007). The Shwachman–Bodian–Diamond syndrome associated protein interacts with HsNip7 and its down-regulation affects gene expression at the transcriptional and translational levels. *Exp. Cell Res.* **313**, 4180–4195.

20. Shammas, C., Menne, T. F., Hilcenko, C., Michell, S. R., Goyenechea, B., Boocock, G. R. *et al.* (2005). Structural and mutational analysis of the SBDS protein family. Insight into the leukemia-associated Shwachman–Diamond Syndrome. *J. Biol. Chem.* **280**, 19221–19229.

21. Ng, C. L., Waterman, D. G., Koonin, E. V., Walters, A. D., Chong, J. P., Isupov, M. N. *et al.* (2009). Conformational flexibility and molecular interactions of an archaeal homologue of the Shwachman–Bodian–Diamond syndrome protein. *BMC Struct. Biol.* **9**, 32.

22. Savchenko, A., Krogan, N., Cort, J. R., Evdokimova, E., Lew, J. M., Yee, A. A. *et al.* (2005). The Shwachman–Bodian–Diamond syndrome protein family is involved in RNA metabolism. *J. Biol. Chem.* **280**, 19213–19220.

23. Wessels, D., Srikantha, T., Yi, S., Kuhl, S., Aravind, L. & Soll, D. R. (2006). The Shwachman–Bodian–Diamond syndrome gene encodes an RNA-binding protein that localizes to the pseudopod of *Dictyostelium amoebae* during chemotaxis. *J. Cell Sci.* **119**, 370–379.

24. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.

25. Cavanagh, J., Fairbrother, W. J., III, Palmer, A. G., Skelton, N. J. & Rance, M. (2007). *Protein NMR Spectroscopy: Principles and Practice*, 2nd edit. Academic Press, Amsterdam.

26. Yao, S., Liu, M. S., Masters, S. L., Zhang, J. G., Babon, J. J., Nicola, N. A. *et al.* (2006). Dynamics of the SPRY domain-containing SOCS box protein 2: flexibility of key functional loops. *Protein Sci.* **15**, 2761–2772.

27. Seeliger, M. A., Spichty, M., Kelly, S. E., Bycroft, M., Freund, S. M., Karplus, M. & Itzhaki, L. S. (2005). Role of conformational heterogeneity in domain swapping and adapter function of the Cks proteins. *J. Biol. Chem.* **280**, 30448–30459.

28. Maiti, R., Domselaar, G. H. V., Zhang, H. & Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* **32**, W590–W594.

29. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR*, **6**, 153–162.

30. Bracken, C., Carr, P. A., Cavanagh, J. & Palmer, A. G., III (1999). Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.* **285**, 2133–2146.

31. Mott, H. R., Nietlispach, D., Hopkins, L. J., Mirey, G., Camonis, J. H. & Owen, D. (2003). Structure of the GTPase-binding domain of Sec5 and elucidation of its Ral binding site. *J. Biol. Chem.* **278**, 17053–17059.

32. Williamson, R. A., Carr, M. D., Frenkiel, T. A., Feeney, J. & Freedman, R. B. (1997). Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry*, **36**, 13882.

33. Yuan, X., Davydova, N., Conte, M. R., Curry, S. & Matthews, S. (2002). Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Res.* **30**, 456–462.

34. Bae, E., Reiter, N. J., Bingman, C. A., Kwan, S. S., Lee, D., Jr., Phillips , G. N. *et al.* (2007). Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J. Mol. Biol.* **367**, 1447–1458.

35. Terribilini, M., Sander, J. D., Lee, J. H., Zaback, P., Jernigan, R. L., Honavar, V. & Dobbs, D. (2007). RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.* **35**, W578–W584.

36. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.

37. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y. *et al.* (1999). Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature*, **398**, 579–585.

38. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly (A)-binding protein. *Cell*, **98**, 835–845.

39. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. & Xu, R.-M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115.

40. Allain, F. H., Bouvet, P., Dieckmann, T. & Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870–6881.

41. Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P. *et al.* (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054–2057.

42. Conte, M. R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S. & Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132–3141.

43. Rideau, A., Gooding, C., Simpson, P. J., Monie, T. P., Lorenz, M., Huttelmaier, S. *et al.* (2006). A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nat. Struct. Mol. Biol.* **13**, 839–848.

44. Patel, R. C. & Sen, G. C. (1998). PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.* **17**, 4379.

45. Huang, X., Hutchins, B. & Patel, R. C. (2002). The C-terminal, third conserved motif of the protein activator PACT plays an essential role in the activation of double-stranded-RNA-dependent protein kinase (PKR). *Biochem. J.* **366**, 175–186.

46. Costa, E. & Santos, R. (2008). Hematologically important mutations: Shwachman–Diamond syndrome. *Blood Cells Mol. Dis.* **40**, 183–184.

47. Carneiro, F. R., Silva, T. C., Alves, A. C., Haline-Vaz, T., Gozzo, F. C. & Zanchin, N. I. (2006). Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem. Biophys. Res. Commun.* **343**, 260–268.

48. Li, M. X., Corson, D. C. & Sykes, B. D. (2002). Structure determination by NMR. Isotope labeling. *Methods Mol. Biol.* **173**, 255–265.

49. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-senriched proteins. *J. Biomol. NMR*, **3**, 185–204.

50. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. B*, **101**, 201–205.

51. Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). ^{1}H–^{1}H correlation via isotropic mixing of ^{13}C magnetization: a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88**, 425–431.

52. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and sidechain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.

53. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

54. Johnson, B. A. & Blevins, R. A. (1994). NMR View: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

55. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378.

56. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54** , 905–921.

57. 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. *J. Biomol. NMR*, **8**, 477–486.