The Shwachman-Bodian-Diamond Syndrome Protein Family Is Involved in RNA Metabolism*

Alexei Savchenko, Nevan Krogan, John R. Cort, Elena Evdokimova, Jocelyne M. Lew, Adelinda A. Yee, Luis Sánchez-Pulido, Miguel A. Andrade, Alexey Bochkarev, James D. Watson, Michael A. Kennedy, Jack Greenblatt, Timothy Hughes, Cheryl H. Arrowsmith, Johanna M. Rommens, and Aled M. Edwards

From the "Ontario Center for Structural Proteomics, the "Banting and Best Department of Medical Research, the "Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L6, Canada, the "Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1A8, Canada, the "Northeast Structural Genomics Consortium, and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, the "Division of Molecular and Structural Biology, Ontario Cancer Institute, Toronto, Ontario M5G 2M9, Canada, the "Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2M9, Canada, the "Protein Design Group, Centro Nacional de Biotecnologia (CNB-CSIC), Cantoblanco, E-28049 Madrid, Spain, the "Bioinformatics Group, Ontario Genomics Innovation Centre, Ottawa Health Research Institute, Ottawa, Ontario K1H 8L6, Canada, the "Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190, the "European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD United Kingdom, and the "Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada.

Received for publication, December 22, 2004, and in revised form, January 24, 2005
Published, JBC Papers in Press, February 8, 2005, DOI 10.1074/jbc.M414421200

A combination of structural, biochemical, and genetic studies in model organisms was used to infer a cellular role for the human protein (SBDS) responsible for Shwachman-Bodian-Diamond syndrome. The crystal structure of the SBDS homologue in *Archaeoglobus fulgidus*, AF0491, revealed a three domain protein. The N-terminal domain, which harbors the majority of disease-linked mutations, has a novel three-dimensional fold. The central domain has the common winged helix-turn-helix motif, and the C-terminal domain shares structural homology with known RNA-binding domains. Proteomic analysis of the SBDS sequence homologue in *Saccharomyces cerevisiae*, YLR022C, revealed an association with over 20 proteins involved in ribosome biosynthesis. NMR structural genomics revealed another yeast protein, YHR087W, to be a structural homologue of the AF0491 N-terminal domain. Sequence analysis confirmed them as distant sequence homologues, therefore related by divergent evolution. Synthetic genetic array analysis of YHR087W revealed genetic interactions with proteins involved in RNA and rRNA processing including Mdm20/Nat3, Nsr1, and Npl3. Our observations, taken together with previous reports, support the conclusion that SBDS and its homologues play a role in RNA metabolism.

---

* This work was supported by Genome Canada, the Ontario Research and Development Challenge Fund, and the National Institutes of Health Protein Structure Initiative (Grants P50-GM62413-02 to the NE Structural Genomics Consortium and P50-GM62414 to the Midwest Center for Structural Genomics). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1P9Q and 1NYN) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

NMR data (code 5695) on chemical shifts have been deposited in the BioMagResBank, Madison, WI.

Holds the Banbury Chair of Medical Research. To whom correspondence should be addressed: Banting and Best Dept. of Medical Research, C. H. Best Institute, University of Toronto, 12 College St., Toronto, ON M5G 1L6, Canada. Tel.: 416-946-3436; Fax: 416-946-0588; E-mail: aled.edwards@utoronto.ca.

---


d The abbreviations used are: SBD, Shwachman-Bodian-Diamond; TAP, tandem affinity purification; SGA, synthetic genetic array; RM, RNA-recognition motif; AF0491-N, AF0491 N-terminal domain; GAR, glycine/arginine-rich repeats; wHTH, winged helix-turn-helix; r.m.s.d., root mean square deviation; EST, expressed sequence tag; PDB, Protein Data Bank; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight.
three blocks of sequence conservation in most organisms, including humans, several plant SBDS homologues contain a fourth C-terminal region that is predicted to be an RNA binding domain (1).

In an effort to determine the molecular and cellular functions of SBDS, we initiated a series of studies in model organisms. As a part of our structural genomics project, we used x-ray crystallography to determine a high resolution structure of the Archaeal SBDS homologue, AF0491 in *Archaeoglobus fulgidus*, and NMR to discover an unanticipated structural homologue of the N-terminal domain of SBDS in *Saccharomyces cerevisiae* (YHR087W). The biochemical and genetic links to RNA metabolism for both the *S. cerevisiae* SBDS sequence homologue, YLR022C, and the more distantly related homologue, YHR087W, suggest a common functional role for this new family.

### EXPERIMENTAL PROCEDURES

**Cloning, Purification, and Crystallization of AF0491**—The AF0491 gene was subcloned, expressed, and its product purified and screened for crystallization as described previously (9). Crystals for x-ray diffraction data collection were obtained from hanging drop vapor diffusion conditions containing 5 μl of Se-Met derivative of the protein plus 2 μl of 0.2 M sodium acetate, 0.1M sodium cacodylate at pH 6.4, and 20% polyethylene glycol 3350 over 2–5 days at 21 °C. The crystals were flash-frozen with crystallization buffer completed with 17% glycerol.

**Experimental Diffraction and Structure Determination**—The protein was crystallized in the P1 space group with the unit cell parameters of a = 33.848, b = 44.189, c = 61.296, β = 92.575°, γ = 109.771°. The crystals contained one independent molecule per unit cell. A 2.0-Å three-wavelength MAD diffraction experiment from selenomethionine-containing crystal was collected and then processed with the HKL2000 package (10). Five selenium sites (of a total of six) were found by direct methods using the SnB program (11). The experimental phases were calculated and solvent flattened by SHARP (12). Of a total of 234 amino acids, 197 were autotraced and partially docked in the sequence by ARP/wARP (13). The remaining amino acids were manually built into the experimental map using the program O (14). The structure was refined, and the water substructure generated in CNS (15). The final model contains all amino acids in an interval of 2–234 (except for a disordered loop comprising residues 115–117) and 145 water molecules.

**Calculation and Analysis of YHR087W Structural Ensemble**—NOE distance restraints had uniform lower bounds of 1.8 Å and upper bounds of 2.8, 3.2, 4.0, and 5.0 Å. Hydrogen bond restraints were derived from amide proton D$_2$O exchange data. Amide 1H-15N HSQC cross-peaks still present 30 min after dissolution of a lyophilized sample in D$_2$O were given bounds of 1.8–2.5 Å for the N-O distance, provided preliminary structural ensembles clearly indicated the correct acceptor atom. Dihedral restraints for phi were derived from the BINHA experiment (20) and had bounds of −55 ± 30 degrees for helical residues with J < 5 Hz and −120 ± 50 degrees for extended residues with J > 7.5 Hz. Dihedral restraints for psi had values of −47 ± 30 for helical residues and 140 ± 50 for extended residues. Psi restraints were added only for residues in helices and β-sheets and only after consideration of amidic-α NOE ratios (H$_N$-H$_α$, versus H$_N$-H$_Ni$), α-carbon chemical shifts, and the evident secondary structure propensities in preliminary ensembles of structures.

Structures were calculated with NIH-Xplor (21) using distance geometry and simulated annealing. The routines dg_sub, embed, dgsa, and dgsa were used as provided except that in dgsa, an initial temperature of 2,000 K was used with 30,000 high temperature steps and 200,000 cooling steps. Sum averaging was used for methyl groups and methylene proton pairs.

### TABLE I

**X-ray data collection and refinement statistics**

| Data set | Wavelength (Å) | Resolution (Å) | No. of reflections measured | No. of independent reflections | R-factor (overall/outer shell) | l/σ (overall/outer shell) | Completeness (%) | Experimental phase calculation (SHARP) | Resolution (Å) | No. of reflections | Figure of merit (acentric) |
|----------|----------------|----------------|---------------------------|-------------------------------|-------------------------------|--------------------------|-----------------|----------------------------------|----------------|---------------------|--------------------------|
|          | 1 (peak)       | 0.9793         | 2 (reflection)            | 0.9795                        | 2.0                           | 74.886                   | 35,721          | 24/23.3                         | 2.6/16.6        | 20–2.0              | 20.8/20               |
|          | 3 (high energy)| 0.8537         |                           |                               |                               |                          |                 |                                  |                 |                     |                          |

$^a$ R(+1) and R(−1) are processed as independent reflections.

---

SBDS Is Involved in RNA Metabolism

The YHR087W gene was PCR-amplified from genomic DNA and inserted into a pET15b (Novagen) vector. This construct yields the protein with an N-terminal His$_6$ tag and thrombin cut site. The protein expression and purification method are described by Yee et al. (17). The His$_6$ tag was not cleaved, leaving an extra 20 residues at the N terminus (MGSSHHHHTTSSGGVPRGSH). NMR samples of ~1.5 mM uniformly $^{13}$C-$^{15}$N-labeled protein were prepared in 10 mM NaOAc, 300 mM NaCl, 10 mM dithiothreitol, 10 μM Zn, 1 mM benzamidine, 1× inhibitor mixture (Roche Applied Science), and 0.01% (v/v) NaN$_3$ in 10% v/v 2H$_2$O/H$_2$O at pH 5.0. This buffer was not optimized to determine the importance of any components. Samples were placed in 5-mm Shigemi susceptibility matched NMR tubes.

**NMR Spectroscopy**—NMR data were collected at 25 °C using standard triple resonance pulse sequences (18). Backbone and side chain correlation experiments (HNCACB, CBCACONH, HNCO, CBCA-COCAHA, CCC-TOCSY-NH, HCC-TOCSY-NH, and HCHC-TOCSY) were acquired on a Varian Unity 600 spectrometer. $^1$H-$^1$C HSQC, three-dimensional $^1$H-$^1$C-$^1$C HMOC-NOESY-HMOC (125-ms mixing time), HHNA, and aromatic ring side chain correlation experiments (HBCBGC-DHD/C-CEHE-aro) were acquired on a Varian Inova 600 spectrometer. $^1$H-$^1$C HSQC and three-dimensional $^1$C-$^1$NOESY (120 ms mixing time) experiments were acquired on a Varian Inova 750 spectrometer. Amide proton exchange was monitored by acquiring $^1$H-$^1$N HSQC spectra following dissolution of a lyophilized sample in D$_2$O. Stereospecific Leu and Val side chain assignments were obtained from a $^1$H-$^1$C HSQC experiment recorded on a sample that had been produced in minimal growth medium containing 10% [U-$^{13}$C]glucose and 90% unenriched glucose (19). The raw data and pulse sequences used have been deposited in the BioMagResBank (Madison, WI). Data were processed with Felix (MSI) and analyzed with Felix and Sparky (www.cgl.ucsf.edu/home/sparky).

**Calculation and Analysis of YHR087W Structural Ensemble**—NOE distance restraints had uniform lower bounds of 1.8 Å and upper bounds of 2.8, 3.2, 4.0, and 5.0 Å. Hydrogen bond restraints were derived from amide proton D$_2$O exchange data. Amide $^1$H-$^1$N HSQC cross-peaks still present 30 min after dissolution of a lyophilized sample in D$_2$O were given bounds of 1.8–2.5 Å for the N-O distance, provided preliminary structural ensembles clearly indicated the correct acceptor atom. Dihedral restraints for phi were derived from the BINHA experiment (20) and had bounds of −55 ± 30 degrees for helical residues with J < 5 Hz and −120 ± 50 degrees for extended residues with J > 7.5 Hz. Dihedral restraints for psi had values of −47 ± 30 for helical residues and 140 ± 50 for extended residues. Psi restraints were added only for residues in helices and β-sheets and only after consideration of amidic-α NOE ratios (H$_N$-H$_α$, versus H$_N$-H$_Ni$), α-carbon chemical shifts, and the evident secondary structure propensities in preliminary ensembles of structures.

Structures were calculated with NIH-Xplor (21) using distance geometry and simulated annealing. The routines dg_sub, embed, dgsa, and dgsa were used as provided except that in dgsa, an initial temperature of 2,000 K was used with 30,000 high temperature steps and 200,000 cooling steps. Sum averaging was used for methyl groups and methylene proton pairs.
SBDS Is Involved in RNA Metabolism

19215

The structural ensemble was analyzed with PROCHECK-NMR (22). Surface electrostatic features of the protein were examined using GRASP (23). Structure similarity searches using Dali (24) and VAST (25) were conducted using the first structure from the ensemble as a representative structure for similarity searching. Structural statistics are presented in Table II.

Sequence Analysis of the Conserved N-terminal Region of the SBDS Family—The alignment of the orthologues of human SBDS was used to search for distant protein families via intermediate searches (26) using global hidden Markov model profiles (using hmmssearch of HMMer; hmmer.wustl.edu/) (27). To improve the profile quality we followed two approaches: first, BLAST searches against unfinished genomes (28), to identify the SBDS family with an E-value of 0.018. Statistically significant E-values connected all the sequences containing the N-terminal conserved region of the SBDS family scored against the non-redundant protein databases. The alignment in Fig. 3 was produced with HMMer (27) and Belvu (www.sanger.ac.uk/Software/Pfam/help/belvu_setup.shtml), and the phylogenetic tree was produced using ClustalW (31).

Purification of TAP-tagged Proteins and Mass Spectrometry—TAP-tagged (tandem affinity purification) (32) proteins were purified from extracts of yeast cells as previously described (33) on IgG and calmodulin columns. The purified proteins were separated by SDS-PAGE on gels containing 10% polyacrylamide, and the proteins were visualized by silver staining. Protein bands were digested with trypsin and peptide samples were spotted onto a target plate with a matrix of hydroxycinnamic acid (Fluka). MALDI-TOF mass spectrometry analysis was conducted using a Reflex IV instrument (Bruker Daltonics, Billerica, MA) in positive ion reflectron mode. Tandem mass spectrometry on a Finnigan LCQ-Deca instrument was carried out as previously described (33) on an aliquot of each preparation that was digested with trypsin in solution after elution of the tagged protein complex from the columns. The purified proteins were separated by SDS-PAGE on gels containing 10% polyacrylamide, and the proteins were visualized by silver staining. Protein bands were digested with trypsin and peptide samples were spotted onto a target plate with a matrix of hydroxycinnamic acid (Fluka). MALDI-TOF mass spectrometry analysis was conducted using a Reflex IV instrument (Bruker Daltonics, Billerica, MA) in positive ion reflectron mode. Tandem mass spectrometry on a Finnigan LCQ-Deca instrument was carried out as previously described (33) on an aliquot of each preparation that was digested with trypsin in solution after elution of the tagged protein complex from the columns.

Synthetic Genetic Array Analysis (SGA)—SGA analysis was carried out as previously described (34), except with a miniarray of 383 deletion strains.

RESULTS

Crystal Structure of AF0491, the Archaeal Homologue of SBDS—The SBDS protein is conserved in Archaea and Eukarya. A 2.0-Å structure of AF0491, the SBDS homologue from A. fulgidus, was solved from a single selenomethionine-containing crystal. The structure was refined to an R_factor/R_free of 21.9/27.0%. The data collection and refinement details are provided under “Experimental Procedures” and in Table I. AF0491 comprises three domains (Fig. 1). The AF0491 N-terminal domain (residues 1–86) is an α/β domain that adopts a novel fold. It consists of a five-stranded antiparallel β-sheet in which strands β2 and β3, as well as β4 and β5, are connected by pairs of helices. The secondary structural elements are thus arranged as follows: β1-β2-a1-α2-β3-loop-β4-a3-α4-β5 (Fig. 2B). On strand β3 on the exposed side of the β-sheet, there is an irregular two residue β-bulge, in which the hydrogen bonding pattern of the antiparallel strands is disrupted, and the normal backbone geometry is distorted to accommodate the two additional residues. There is a loop at one end of the β-sheet that bends in the direction of the bulge, forming a cupped incursion in the surface. It appears that this fold is new, as no similar structures were identified with Dali or VAST. Structure-based sequence alignments showed that this domain corresponds to residues 10–95 in human SBDS.

The middle domain of AF0491 (residues 87–160) is a winged helix-turn-helix, a common fold associated with DNA binding (35). In the context of the human SBDS, this domain corresponds to residues 98–169. The C-terminal domain of AF0491 (residues 161–234), which corresponds to residues 170–241 in the human protein, comprises a four-stranded β-sheet that is buttressed on one side by two helices. This is also a common fold, sharing structural homology with the RNA recognition motif (RRM).

More than 20 different mutations have been identified in SBDS syndrome patients (1, 5, 6) (Fig. 1A). Most of the mutations that alter surface residues are located in the N-terminal half of the protein, where many of the conserved residues are also located. An analysis of disease-linked mutations can be found in the report by Shammas et al. (6).

YHR087W and the AF0491 N-terminal Domain—As part of an ongoing structural genomics project in yeast, we used NMR spectroscopy to determine the structure of YHR087W from S. cerevisiae (Fig. 2A). This protein did not have a previously known structure or function; however, the YHR087W orthologue in S. pombe is expressed during sporulation and environmental stress (36, 37). The structure of YHR087W displayed striking similarity to that of the N-terminal domain of AF0491 (Fig. 2B), a similarity that was first recognized by Alexey Murzin. This was confirmed using pairwise comparisons: Dali returned an alignment of 79 residues with an r.m.s.d. of 2.1 Å for the CA atoms, whereas VAST reported 68 aligned residues with an r.m.s.d. of 2.3 Å for the CA atoms, whereas VAST reported 68 aligned residues with an r.m.s.d. of 2.3 Å for the CA atoms.

Consequently, we searched for sequence similarity between the two proteins using HMMer. Profiles of the N-terminal conserved region of the SBDS family scored against the YHR087W sequence with an E-value of 0.042. Reciprocally, the profile of the YHR087W sequence with its orthologues detected the SBDS family with an E-value of 0.018. Statistically significant E-values connected all the sequences contained in both families. None of these HMMer profile searches retrieved any new unrelated sequences and, as stated above, reciprocal searches produced convergent results. Thus, YHR087W is a distant sequence homologue of the AF0491 N-terminal domain and, by extension, of both the human SBDS and YLR022C, an SBDS sequence homologue that we identified in S. cerevisiae (Figs. 1A and 3). The runs of hydrophobic residues are in good agreement between both alignments, and there are also several conserved residues. The grouping by taxonomy is in agreement with the structure of the phylogenetic tree derived from the multiple sequence alignment, with all sequences in the branch of YHR087W being found exclusively in fungi (with the exception of the Zea mais

| Table II
| Structural statistics for the 20-member ensemble of structures of YHR087W |
| Distance restraints | Total | 665 |
| Interaresidue | 61 |
| Sequential | 161 |
| Medium range (1 < |j| < 5) | 160 |
| Long range | 283 |
| Hydrogen bonds | 225 |
| Dihedral angle restraints | Total | 73 |
| Corrected  | 38, 35 |
| Total restraints | 788 |
| Residues 4–90 | 9.6 |
| Backbone atoms (C, CA, N) | 0.95 |
| All heavy atoms | 1.48 |
| Residues 4–50 and 60–90 (loop excluded) | 0.66 |
| Backbone atoms | 1.15 |
| All heavy atoms | 1.59 |
| Residues 4–50, 60–71, 77–90 (secondary structure) | 1.08 |
| Backbone atoms | 2.62 |
| All heavy atoms | 2.7 |
| Ramachandran plot, residues 4–90 | 0.3 |
| Most favored region (%) | 70.7 |
| Additional allowed region (%) | 26.2 |
| Generously allowed region (%) | 2.7 |
| Unallowed region (%) | 0.3 |

A. Murzin, personal communication.
sequence, which could be either a true case of horizontal transfer or the result of a fungal contamination of the corresponding EST library).

The structural similarity between YHR087W and the AF0491 N-terminal domain (AF0491-N) extends over the entire structure, including conservation of the irregular $\beta$-bulge in $\beta3$ and the loop at one end of the $\beta$-sheet. Conservation of these structural details, particularly within such an unusual fold, likely indicates a shared function. There is a 9-residue loop inserted between $\beta3$ and $\beta4$ of YHR087W; this loop contains 3 residues in AF0491-N (Fig. 1A). In the NMR structure, this loop is structurally underdetermined because of a lack of resonance assignments and NOE data, most likely because of intermediate time scale conformational exchange.

The electrostatic properties of YHR087W and AF0491-N are also similar (Fig. 4). One side of both proteins, formed by the side of helix $\alpha2$, the irregular bulge region in strand $\beta3$ and the edge of $\beta4$ as well as by the N-terminal end of helix $\alpha4$ and the loop preceding it, is particularly rich in acidic residues and devoid of basic residues, whereas the other sides of the structures have a mixture of acidic, basic, and uncharged side chains. The conservation of electrostatic surface potential is explained by the presence of a number of conserved or identical residues. Among the conserved residues are two glutamates.

---

**Fig. 1. Structure of AF0491 from A. fulgidus.** A, sequence alignment of structural (YHR087W) and sequence homologues of human SBDS. SBDS mutations are indicated above the sequences with the most common mutations in red. Structurally aligned YHR087W residues are in uppercase letters and unaligned residues are lowercase. YHR087W residues in red or green are structurally aligned residues that are identical or conserved, respectively. B, stereo image of AF0491, showing its three domain structure.
which contribute to the acidic side of the structure, although additional acidic residues, which are not conserved, also contribute to this feature in both structures. Notably, there is a conserved basic region (Lys\textsuperscript{31}/Lys\textsuperscript{32}, YHR087W/AF0491-N), which, in AF0491, is located in the large cleft formed by the three domains. Also present are several conserved hydrophobic residues, probably involved in stabilization of the fold.

Functional Analysis of YLR022C and YHR087W, SBDS Homologues in S. cerevisiae—The biochemical and cellular functions of both YLR022C and YHR087W are unknown. Strains lacking the YLR022C gene are not viable, indicating it is an essential gene, whereas strains lacking YHR087W are viable (38). Thus, in an effort to infer the functions of YLR022C and YHR087W, and by inference those of SBDS, we turned to the genomic and proteomic analyses that have been used with such success in yeast.

YLR022C was tagged at the C terminus with a TAP tag by recombining a tagging cassette directly into the chromosomal YLR022C locus. The protein was then purified by tandem affinity chromatography (Fig. 5) and associated proteins were

---

**Fig. 2. Structural comparison of YHR087W and AF0491.** A, stereo image of 20 superimposed NMR structures of S. cerevisiae YHR087W. B, secondary structure of the AF0491 N-terminal domain.

**Fig. 3.** Multiple sequence alignment of representative sequences from the SBDS protein family (includes the human SBDS, swSBDS_HUMAN, and the A. fulgidus AF0491 sequence, sw:Y491_ARCFU), and representative sequences from the yeast protein family of S. cerevisiae YHR087W (sw:YHO7_YEAST). The sequences are named with their Swiss-Prot or spTrEMBL identifiers, and also, if necessary, with their common species name: Human, Homo sapiens; Fish, Brachydanio rerio; Fly, Drosophila melanogaster; Ara, Arabidopsis thaliana; Yeast, S. cerevisiae; Magna, Magnaporthe grisea; Neuc, Neurospora crassa; Aspe, Aspergillus nidulans; Pyrho, Pyrococcus horikoshii; Arcfu, Archaeoglobus fulgidus; Gibbe, Gibberella zeae; Zea, Z. mais; Hypo, Hypocrea jecorina. The prefix identifies the database from which the sequence originates: sw, Swiss-Prot; sp, spTrEMBL; est, consensus sequences manually reconstructed by assembling highly similar expressed sequence tags from a given species (conceptual translations) obtained from the EST database at NCBI (28); gb, genome BLAST server at NCBI (28). Branch coloring indicates the taxonomy of the sequence: blue, yeasts; green, plants; red, other eukaryotes; grey, Archaea. The color scheme of the alignment indicates average BLOSUM62 score (correlated to amino acid conservation) in each alignment column: cyan, greater than 3.5; light red, between 3.5 and 1.5; light green, between 1.5 and 0.8.
identified using LC/MS/MS mass spectrometry. YLR022C co-purified with small amounts of a large number of polypeptides, most of which are linked to the processing of rRNA and many of which are known components of the 60 S particle. Because ribosomal proteins are abundant, there is a concern that they might be adventitiously associated with YLR022C. However, we believe that this is unlikely because we have now purified more than 3,000 yeast proteins, and only a limited number have co-purified with the 60 S particle. Therefore, these data support a role for YLR022C in ribosomal processing.

To explore whether the sequence and structural similarity between YHR087W and YLR022C corresponded to a functional similarity, we performed genetic and protein interaction studies with YHR087W. When YHR087W was tagged with a TAP tag and purified using tandem affinity chromatography, we did not detect any co-purifying proteins (data not shown). However, since YHR087W is a non-essential gene, we were able to explore genetic interactions between YHR087W and other yeast genes using SGA analysis; results were confirmed by tetrad analysis. A strain deleted for YHR087W was crossed with a miniarray of 383 other deletion strains, each lacking a protein implicated in RNA metabolism. Deletion of YHR087W caused marked synthetic lethality when combined with deletions of several other genes including those that encode the NatB complex (Mdm20/Nat3), which is required for the acetylation of ribosomal proteins (39), and Nar1, a nucleolar protein involved in the synthesis of 18 S rRNA and its 20 S precursor (40, 41). Nar1 has two RNA recognition domains and is a member of the GAR (glycine/arginine-rich repeats) family of proteins (42). YHR087W also interacted genetically with each of Npl3, Air2, and Yra2, which others have shown to interact with each other both by GST pull-down and two hybrid-based methods (43, 44). Npl3 is a protein involved in 18 S and 25 S rRNA processing, export of RNA from the nucleus, and import of proteins into the nucleus (45–48). Npl3 is also associated with the U1 snRNP and is predicted to have two RNA recognition domains (46, 49). Air2 is a RING-type zinc finger domain protein, and Yra2 is a protein that associates with RNP complexes (43, 44). Therefore, the genetic interactions of YHR087W support a role for the protein in RNA processing.

DISCUSSION

Shwachman-Bodian-Diamond syndrome is the second most common cause of pancreatic insufficiency in children, after cystic fibrosis (50). The syndrome is caused by mutations in the SBDS gene on chromosome 7 (1). Our studies of SBDS homologues in Archaea and yeast provide experimental evidence for a role for SBDS in RNA metabolism.

We have solved the crystal structure of AF0491, the SBDS homologue in A. fulgidus, revealing a three domain protein. The C-terminal domain comprises a four-stranded β-sheet with two helices on one side, a commonly occurring fold. Although the domain shares structural homology with the RRM, this is a common fold making it difficult to infer a function. RRMs are found in many RNA-binding proteins, including translation factors, poly(A)-binding proteins, and proteins associated with pre-mRNA and pre-rRNA processing (51). This domain has also been found in DNA-binding proteins and can be involved in protein-protein interactions (52–54).

The central domain of the protein adopts another common fold, the winged helix-turn-helix (wHTH). Proteins with an HTH domain are abundant in all Archaea, with the wHTH fold being the most common (55), and so it is difficult to infer a molecular function. Although HTH domains are widely used in DNA binding and have also been identified in RNA-binding proteins (56–58), a role in nucleic acid binding is not supported since the surface of AF0491 does not have the general basic character expected for such a function. Part of this domain may be involved in protein-protein interactions, as with K. lactis HSF, another wHTH protein. Unlike other wHTH proteins where the wings contact DNA, the wing in K. lactis HSF is involved in HSF dimerization (59).

The N-terminal domain of AF0491 is a novel αβ-fold composed of a five-stranded antiparallel β-sheet, with pairs of helices between strands β2 and β3, and strands β4 and β5. Most of the disease-linked mutations identified in SBD syndrome patients are located in this unusual domain, and genetic evidence suggests that, in humans, the presence of the N-terminal domain is required for viability. Interestingly, this new fold was found in another protein, YHR087W, whose structure was determined by NMR spectroscopy. Subsequent sequence similarity analysis demonstrated that the two proteins are indeed distant homologues. Structural comparisons revealed the striking conservation of details such as the unusual irregular β-bulge in strand β3, and electrostatic surface char-
SBDS Is Involved in RNA Metabolism

19219

acteristics. These findings prompted us to search for a shared function.

To elucidate the function of SBDS, we made use of the various experimental advantages of the yeast *S. cerevisiae*, and studied the SBDS structural and sequence homologues, YHR087W and YLR022C respectively. Our results link both YLR022C and YHR087W to RNA metabolism.

TAP-tagged YLR022C co-purified with numerous ribosomal proteins and proteins associated with rRNA processing. Protein complexes that are purified using the TAP protocol can also be probed for the presence of specific RNAs through hybridization to DNA microarrays. This process is particularly applicable to RNA processing enzymes, which are often ribonucleoprotein assemblies. Previously, the affinity-purified YLR022C complex was probed for potential co-purifying RNAs by using a dedicated DNA microarray (60). YLR022C was found to co-purify with snoRNAs and exhibited a profile that is similar to that of YHR040W (BCD1), an RNA-associating protein. snoRNPs, which contain these non-coding snoRNAs, have been implicated in the cleavage, modification, and folding of precursor rRNA substrates (61).

Our studies, which provide experimental evidence for a role for SBDS in RNA metabolism, support previous hypotheses which were based on bioinformatics studies. In Archaea, the orthologous gene is located in an operon that includes RNA-processing enzymes (7). Through a computational study, YLR022C was predicted to function in RNA processing (62). This was accomplished by analyzing protein-protein interaction networks and identifying biologically relevant functional groups. Also, SBDS RNA is expressed ubiquitously (1), which is consistent with a basic cellular function such as RNA processing. Finally, in some plants, the SBDS orthologues have a fourth domain that contains a putative RNA-binding motif (1).

Although our data link the SBDS protein to ribosomal biogenesis, the specific role of SBDS in this pathway remains to be determined. Whereas the function of the ribosome is conserved in all living species, ribosomal biogenesis is fundamentally different in Bacteria compared with Archaea and Eukarya; many families of ribosomal genes are specific to Archaea and Eukarya and are absent from Bacteria (63). The observation that the SBDS gene is restricted to Archaea and Eukarya (1) suggests that SBDS plays a role that is specific for the process in these two kingdoms.

Bacterial ribosomes, though sharing many proteins with those in eukaryotes and Archaea, are assembled differently (64–67). Active prokaryotic ribosomes can be reconstituted in vitro using only the individual ribosomal components. Although this does not exclude the involvement of additional factors in vivo, it is evident that all the information needed to assemble an active prokaryotic ribosome is contained within the sequences of the ribosomal proteins and RNAs. By contrast, eukaryotic, and likely archaeal, ribosome biosynthesis requires the coordinated action of hundreds of accessory proteins, snoRNPs, and ribosomal proteins to produce the final assembly of over 70 ribosomal proteins with four rRNAs. In this pathway, three of these rRNAs (18 S, 25 S, and 5.8 S) are produced from a single 35 S transcript whereas the fourth 5 S rRNA is independently transcribed. Processing of the pre-rRNA occurs following association of ribosomal and non-ribosomal proteins with the pre-rRNAs, forming a pre-ribosomal particle. Pre-rRNA modifications needed to produce the mature rRNA include cleavage of the pre-rRNAs, conversion of uridine residues to pseudouridine, and nucleotide methylation. This pre-ribosomal particle then separates into 40 S and 60 S pre-subunits. Until the final maturation of the ribosome occurs in the cytoplasm.

Numerous trans-acting protein factors with a role in ribosome biogenesis have been identified in yeast, including rRNA-modifying enzymes, nucleases and putative RNA helicases, and it is likely that others have not yet been discovered (65). SBDS may be one such factor, necessary for eukaryotic, but not prokaryotic, ribosomal biosynthesis.

Acknowledgments—All NMR spectra were performed at the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the United States Department of Energy Office of Biological and Environmental Research, located at Pacific Northwest National Laboratory and operated by Battelle for the DOE. Diffraction data were collected at the Argonne National Laboratory (Beamline 19ID).

REFERENCES

1. Boocock, G. R., Morrison, J. A., Popovic, M., Richards, N., Ellis, L., Durie, P. R., and Rommens, J. M. (2003) *Nat. Genet.* 33, 97–101

2. Bodian, M., Sheldon, W., and Lightwood, R. (1984) *Acta Paediatr.* 53, 282–293

3. Shwachman, H., Diamond, L. K., Osik, F. A., and Khaw, K. T. (1964) *J. Pediatr.* 65, 645–663

4. Ginzburg, H., Shin, J., Ellis, L., Gobie, S., Morrison, J., Corey, M., Durie, P. R., and Rommens, J. M. (2000) *Am. J. Hum. Genet.* 66, 1413–1416

5. Nakashima, E., Mabuchi, A., Makita, Y., Masuno, M., Ohashi, H., Nishimura, O., Naoi, K., and Ikegawa, S. (2000) *Hum. Genet.* 114, 345–348

6. Shammas, C., Menne, T. F., Hildenken, C., Michell, S. R., Goyenechea, B., Boocock, G. R. B., Durie, P. R., Rommens, J. M., and Warren, A. J. (2005) *J. Biol. Chem.* 280, 19221–19229

7. Koonin, E. V., Wolf, Y. I., and Aravind, L. (2001) *Genome Res.* 11, 240–252

8. Wu, L. F., Hughes, T. R., Davierwala, A. P., Robinson, M. D., Stoughton, R., and Altschuler, S. J. (2002) *Nat. Genet.* 31, 253–265

9. Zhang, R. G., Kim, Y., Skarf, M. M., Braseley, S., Laskowski, R., Arrowmith, C., Edwards, A., Joachimich, A., and Savchenko, A. (2002) *J. Biol. Chem.* 277, 19183–19190

10. Owczynski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326

11. Weeks, C. M., and Miller, R. (1999) *J. Am. Crystallogr.* 35, 123–124

12. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* 276, 472–494

13. Lammzin, V. S., Perrakis, A., and Wilson, K. S. (2001) in *Int. Tables for Crystallography, Vol F: Crystallography of Biological Macromolecules* (Bomman, M. G., and Arnold, E., eds) pp. 720–722, Kluwer Academic Publishers, Dordrecht, The Netherlands

14. Jones, A. T., and Kjeldgaard, M. (1997) *Methods Enzymol.* 277, 173–208

15. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gron, P., Gross-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921

16. Collaborative Computational Project, Number 4. (1994) *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763

17. Yao, A., Chang, X., Pineda, A., A., Wu, B., Semesi, A., Lee, B., Ramot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C. H., Cort, J. R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L. P., Gehring, K., Kennedy, M. A., Edwards, A. M., and Arrowmith, C. H. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 1825–1830

18. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy, Principles and Practice*, pp. 411–528, Academic Press, San Diego, CA

19. Neri, D., Syzperci, T., Oting, G., Senn, H., and Wuthrich, K. (1989) *Biochemistry* 28, 7510–7516

20. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 7772–7777

21. Schwiers, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003) *Acta Crystallogr.* 59, 665–667

22. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003) *Protein NMR Spectroscopy, Principles and Practice*, pp. 411–528, Academic Press, San Diego, CA

23. Neri, D., Syzperci, T., Oting, G., Senn, H., and Wuthrich, K. (1989) *Biochemistry* 28, 7510–7516

24. Holm, L., and Sander, C. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 45, 957–961

25. Stroud, R. M., and Wuthrich, K. (1992) *Protein NMR Spectroscopy, Principles and Practice*, pp. 411–528, Academic Press, San Diego, CA

26. Holm, L., and Sander, C. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 45, 957–961

27. Stroud, R. M., and Wuthrich, K. (1992) *Protein NMR Spectroscopy, Principles and Practice*, pp. 411–528, Academic Press, San Diego, CA
SBDS Is Involved in RNA Metabolism

Chen, D., Tone, W. M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bahler, J. (2003) Mol. Biol. Cell 14, 214–229

Mata, J., Lyne, R., Burns, G., and Bahler, J. (2002) Nat. Genet. 32, 143–147

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lu, A., Manley, J., Myers, R. M., Mattanovich, D., Beach, D., Huang, E., and待提供...

Kondo, K., Kowalski, L. R., and Inouye, M. (1992) J. Biol. Chem. 267, 16252–16258

Zenklusen, D., Vinciguerra, P., Strahm, Y., and Stutz, F. (2001) Mol. Cell. Biol. 21, 4219–4232

Inoue, K., Mizuno, T., Wada, K., and Hagiwara, M. (2000) J. Biol. Chem. 275, 32793–32799

Russell, I. D., and Tollervey, D. (1992) J. Cell Biol. 119, 737–747

Singleton, D. R., Chen, S., Hitomi, M., Kamagai, C., and Tartakoff, A. M. (1995) J. Cell Sci. 108, 265–272

Bossie, M. A., DeHoratius, C., Barcelo, G., and Silver, P. (1992) Mol. Cell. Biol. 12, 3865–3871

Schlenstedt, G., Hurt, E., Doye, V., and Silver, P. A. (1993) J. Cell Biol. 123, 785–798

Gotzschalk, A., Tang, J., Puig, O., Salgado, J., Neuhauer, G., Colot, H. V., Mann, M., Sper thlin, B., Rosbash, M., Lührmann, R., and Fabrizio, P. (1998) RNA 4, 374–383

Drerup, Y., and Freedman, M. H. (2002) Br. J. Haematol. 118, 701–713

Birney, E., Kumar, S., and Krainer, A. R. (1993) Nucleic Acids Res. 21, 5803–5816

Newberry, E. P., Latifi, T., and Towler, D. A. (1999) Biochemistry 38, 10675–10690

DeAngelo, D. J., DeFalco, J., Rybacki, L., and Childs, G. (1995) Mol. Biol. Cell 15, 1254–1264

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Bhana, Z., Kramer, A., and Sattler, M. (2003) Mol. Cell 11, 965–976

Aravind, L., and Koonin, E. (1999) Nucleic Acids Res. 27, 4658–4670

Xing, Y., Guha Thakurta, D., and Draper, D. E. (1997) Nat. Struct. Biol. 4, 24–27

Schade, M., Turner, C. J., Lowenhaupt, K., Rich, A., and Herbert, A. (1999) EMBO J. 18, 470–479

Wolf, Y. I., Aravind, L., Grishin, N. V., and Koonin, E. V. (1999) Genome Res. 9, 689–710

Littlefield, O., and Nelson, H. C. (1999) Nat. Struct. Biol. 6, 464–470

Peng, W. T., Robinson, M. D., Mannmeh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W., Mitsakakis, N., Ryan, O. W., Datta, N., Ogjenc, V., Pal, C., Canadion, V., Richards, D., Beattie, B., Wu, L. F., Alschuler, S. J., Roweis, S., Frey, B. J., Emili, A., Grénbblatt, J. F., and Hughes, T. R. (2003) Cell 113, 919–933

Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J., and Dennis, P. P. (2003) Mol. Microbiol. 48, 617–629

Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., Li, G., and Chen, B. (2003) Nucleic Acids Res. 31, 2443–2450

Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002) Nucleic Acids Res. 30, 5382–5390

Tschochner, H., and Hurt, E. (2003) Trends Cell Biol. 13, 255–263

Kressler, D., Linder, P., and de La Cruz, J. (1999) Mol. Cell. Biol. 19, 7907–7912

Nierhaus, K. H. (1991) Biochimie (Paris) 73, 739–755

Venema, J., and Tollerey, D. (1999) Annu. Rev. Genet. 33, 261–311

Nicholas, A. (1992) GRASP: Graphical Representation and Analysis of Surface Properties, Columbia University, New York