Structural and Mutational Analysis of the SBDS Protein Family

INSIGHT INTO THE LEUKEMIA-ASSOCIATED SHWACHMAN-DIAMOND SYNDROME

Received for publication, December 29, 2004, and in revised form, January 31, 2005
Published, JBC Papers in Press, February 8, 2005, DOI 10.1074/jbc.M414656200

Camille Shammas‡§, Tobias F. Menne‡§, Christine Hilkenko‡, Stephen R. Michell‡, Beatriz Goyenechea‡, Graeme R. B. Boocock‡, Peter R. Durie***¶, Johanna M. Rommens**, and Alan J. Warren**††††

From the ‡Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, United Kingdom, the §§Department of Hematology, University of Cambridge, Cambridge CB2 2XY, United Kingdom, the Departments of ¶Molecular and Medical Genetics and ¶¶Pediatrics, University of Toronto, Toronto, Ontario M5S 1A8, Canada, and the |Program in Genetics and Genomic Biology and **Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

Shwachman-Diamond Syndrome (SDS) is an autosomal recessive disorder characterized by bone marrow failure with significant predisposition to the development of poor prognosis myelodysplasia and leukemia, exocrine pancreatic failure and metaphyseal chondrodysplasia. Although the SBDS gene mutated in this disorder is highly conserved in Archaea and all eukaryotes, the function is unknown. To interpret the molecular consequences of SDS-associated mutations, we have solved the crystal structure of the Archaeoglobus fulgidus SBDS protein orthologue at a resolution of 1.9 Å, revealing a three domain architecture. The N-terminal (FYSH) domain is the most frequent target for disease mutations and contains a novel mixed αβ-fold identical to the single domain yeast protein Ylr022cp that is implicated in RNA metabolism. The central domain consists of a three-helical bundle, whereas the C-terminal domain has a ferredoxin-like fold. By genetic complementation analysis of the essential Saccharomyces cerevisiae SBDS orthologue YLR022C, we demonstrate an essential role in vivo for the FYSH domain and the central three-helical bundle. We further show that the common SDS-related K62X truncation is non-functional. Most SDS-related missense mutations that alter surface epitopes do not impair YLR022C function, but mutations affecting residues buried in the hydrophobic core of the FYSH domain severely impair or abrogate complementation. These data are consistent with absence of homozygosity for the common K62X truncation mutation in individuals with SDS, indicating that the SDS disease phenotype is a consequence of expression of hypomorphic SBDS alleles and that complete loss of SBDS function is likely to be lethal.

Shwachman-Diamond syndrome (SDS, OMIM 260400) is a rare (frequency of 1:76000) autosomal recessive disorder with clinical features that include hematological dysfunction, pancreatic exocrine insufficiency and skeletal abnormalities (1–3). SDS patients are significantly predisposed to the development of hematological abnormalities, including cytopenia of one or more lineages, myelodysplasia (MDS) and acute leukemia (4). The cumulative risk of MDS/acute leukemia among SDS patients was recently reported as 18.8 and 36.1% at 20 and 30 years, respectively (5). SDS therefore represents an important model for understanding the genetic determinants underlying the multistep progression to leukemia.

SDS-associated mutations were recently described in a gene designated SBDS (Shwachman-Bodian-Diamond syndrome) (6) that encodes a member of a highly conserved protein family of unknown function with orthologues in diverse species including Archaea, plants, and eukaryotes but not Eubacteria (Pfam P000023, Ref. 7). Indirect evidence supports the hypothesis that the SBDS protein family may function in RNA metabolism. Thus, the archaeal SBDS orthologues are located in highly conserved operons containing RNA-processing genes that include orthologues of the eukaryotic exosome and RNase P complex subunits (8) and at least three plant SBDS orthologues contain extended C termini with putative RNA binding U1-type zinc fingers. Haploid spores deleted for the Saccharomyces cerevisiae SBDS orthologue YLR022C were reported to be inviable (9). Although the Ylr022cp protein has been clustered functionally with RNA-processing enzymes and ribosomal RNA-processing factors in microarray expression analyses (10, 11), it has also been reported to bind the phospholipids PI(4,5)P2 and PI4P in vitro (12). GFP-tagged Ylr022cp was distributed in both the nucleus and cytoplasm (13).

In SDS, recurring mutations in the SBDS gene arise from gene conversion caused by recombination between SBDS and a pseudogene copy that lies in a 305-kb paralogous duplicon located 5.8 Mb distally (6). Converted gene segments consistently include at least one of two pseudogene-like sequence changes that are predicted to result in SBDS protein truncation. Specifically, the SBDS dinucleotide mutation 183–184TA→CT introduces an in-frame stop codon (K62X) whereas the 258 + 2T→C mutation is predicted to result in premature truncation.
calculation of the SBDS protein by frameshift (C84fs3) (6). At least 90% of affected individuals carry one common conversion mutation, whereas 50% are compound heterozygotes with respect to the K62X mutation were identified, suggesting that this may be a lethal mutation. Alleles from affected individuals who do not have the common conversion mutations carry additional missense mutations in the coding region of the SBDS gene (14, 15).

To interpret the molecular consequences of SDS-related mutations and to begin to address the conserved function of the SBDS protein, we have determined the structure of the A. fulgidus SBDS orthologue AF0491 (herein called AfsSBDS). The AfsSBDS structure represents a paradigm for the SBDS protein family in view of the striking evolutionary amino acid sequence conservation. The AfsSBDS structure has a three domain architecture: an N-terminal domain with a mixed three-helical bundle and a C-terminal ferredoxin-like domain. We show that the N-terminal domain is the most frequent target for disease mutations and is structurally related to the SBDS orthologue AF0491 (AfSBDS) were expressed with an N-terminal His6 tag in either phic affinity resin (Qiagen) and Resource S ion-exchange (Amersham Biosciences), concentrated to 10 mg/ml in buffer A (20 mM citrate, pH 5.3, 300 mM NaCl, 2 mM dithiothreitol, 1 mM EDTA) and stored at 21 °C. His6-tagged human SBDS, residues 1–250 and yeast His6-Yhr022cp, residues 1–250 were expressed and purified as above and dialyzed into 20 mM Hepes, pH 7.0, 1 mM dithiothreitol, 1 mM EDTA.

Cryocrystallography—Crystals of SeMet-substituted His6-AfsSBDS protein were grown as hanging drops against 1 ml of reservoir solution (19–20% w/v polyethylene glycol 8000, 100 mM CHES buffer pH 9.5, 2 mM dithiothreitol) at 21 °C, using 1 µl of 10 mg/ml protein solution and 1 µl of crystallization solution. Crystals grew to typical dimensions of 300 × 100 × 50 µm over 3 days, and were cryocooled in liquid nitrogen following transfer to a cryobuffer comprising reservoir solution supplemented with 20% (v/v) glycerol.

**Data Collection, Phasing, and Model Refinement—**Data were collected at 100 K at the ESRF beamline ID29 in Grenoble using an ADSC Q210 CCD detector. The SeMet-substituted AfsSBDS protein crystal was grown in space group P1 with cell parameters a = 33.68 Å, b = 44.44 Å, c = 54.72 Å, α = 75.85°, β = 65.61°, γ = 69.49° and a solvent content of 55.6%. Data sets were collected at three different wavelengths, A1 = 0.97929 Å, λ2 = 0.97929 Å, and λ3 = 0.891971 Å, corresponding to peak, inflection, and remote points in the fluorescence spectrum of the SeMet-substituted crystal respectively. 360° of data were collected at each wavelength in 1° oscillations from a single crystal. Data were processed using MOSFLM (17) and scaled using SCALA (18). SHARP (19) was used to locate five of the six possible selenium atoms in the asymmetric unit to generate phases. The phases were calculated by SHARP and improved by density modification, histogram matching, and solvent flattening using DM (20). The phases from DM were used as the starting point for automated chain tracing and model building using ARP/wARP (21). Manual model building was performed using the program O (22), and the structure refined using Refmac5 (23). All methionine residues were modeled and refined as SeMet. Maximum likelihood residues restrained using Hendrickson-Lattmann coefficients from the SHARP output were used throughout refinement against the peak wavelength data. Table I lists the statistics for data collection and phase refinement. The structure has been refined to a free R factor (Rfree) of 24.9% and shows good stereochemistry with no residues in disallowed regions of the Ramachandran plot. No density was seen for the N-terminal four amino acid residues of AfsSBDS or the His6 tag.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification—**Residues 1–234 of the A. fulgidus SBDS orthologue AF0491 (AfsSBDS) were expressed with an N-terminal His6 tag in either Escherichia coli C41 (DE3) or the methionine auxotroph B834 (DE3) for selenomethionine (SeMet) incorporation as described (16). The protein was purified using Ni2+ nitrilotriacetic acid affinity resin (Qiagen) and Resource S ion-exchange (Amersham Biosciences), concentrated to 10 mg/ml in buffer A (20 mM citrate, pH 5.3, 300 mM NaCl, 2 mM dithiothreitol, 1 mM EDTA) and stored at 21 °C. His6-tagged human SBDS, residues 1–250 and yeast His6-Yhr022cp, residues 1–250 were expressed and purified as above and dialyzed into 20 mM Hepes, pH 7.0, 1 mM dithiothreitol, 1 mM EDTA.

**Cryocrystallography—**Crystals of SeMet-substituted His6-AfsSBDS protein were grown as hanging drops against 1 ml of reservoir solution (19–20% w/v polyethylene glycol 8000, 100 mM CHES buffer pH 9.5, 2 mM dithiothreitol) at 21 °C, using 1 µl of 10 mg/ml protein solution and 1 µl of crystallization solution. Crystals grew to typical dimensions of 300 × 100 × 50 µm over 3 days, and were cryocooled in liquid nitrogen following transfer to a cryobuffer comprising reservoir solution supplemented with 20% (v/v) glycerol.

**Data Collection, Phasing, and Model Refinement—**Data were collected at 100 K at the ESRF beamline ID29 in Grenoble using an ADSC Q210 CCD detector. The SeMet-substituted AfsSBDS protein crystal was grown in space group P1 with cell parameters a = 33.68 Å, b = 44.44 Å, c = 54.72 Å, α = 75.85°, β = 65.61°, γ = 69.49° and a solvent content of 55.6%. Data sets were collected at three different wavelengths, A1 = 0.97929 Å, λ2 = 0.97929 Å, and λ3 = 0.891971 Å, corresponding to peak, inflection, and remote points in the fluorescence spectrum of the SeMet-substituted crystal respectively. 360° of data were collected at each wavelength in 1° oscillations from a single crystal. Data were processed using MOSFLM (17) and scaled using SCALA (18). SHARP (19) was used to locate five of the six possible selenium atoms in the asymmetric unit to generate phases. The phases were calculated by SHARP and improved by density modification, histogram matching, and solvent flattening using DM (20). The phases from DM were used as the starting point for automated chain tracing and model building using ARP/wARP (21). Manual model building was performed using the program O (22), and the structure refined using Refmac5 (23). All methionine residues were modeled and refined as SeMet. Maximum likelihood residues restrained using Hendrickson-Lattmann coefficients from the SHARP output were used throughout refinement against the peak wavelength data. Table I lists the statistics for data collection and phase refinement. The structure has been refined to a free R factor (Rfree) of 24.9% and shows good stereochemistry with no residues in disallowed regions of the Ramachandran plot. No density was seen for the N-terminal four amino acid residues of AfsSBDS or the His6 tag.

**Yeast Plasmids, Strains, and Media—**The sequence encoding yeast Yhr022cp-His6 was cloned into pYC2/CT (CEN6/ARS1/UR43) (In-vitrogen) to generate pYC2(Gal10::YL022CPHis6). The YLR022Cp open reading frame (ORF) was PCR-amplified with 500 bp of 5' and 400 bp of 3' genomic sequence and cloned into plasmids pRS114 (ARS/CEN/
TRP to generate pRS314[YLR022C]. NdeI and EcoR1 restriction sites were introduced at the 5’- and 3’-ends, respectively, of the YLR022C ORF to generate pRS314[YLR022CNE]. Mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions. The human SBDS, Dictyostelium discoideum SBDS and AfSBDS ORFs were PCR-cloned as NdeI-EcoR1 fragments into pRS314[YLR022CNE] to generate pRS314[YLR022CNE_HSBDS], pRS314[YLR022CNE_DSBDS], and pRS314[YLR022CNE_AfSBDS]. The diploid yeast strain 20519D (W303; Mat a/a; ura3-1::his3-11/12; trp1::ura3-1::his3-11/12; leu2-3::can1-100) was purchased from Euroscarf. Genomic de-
leation of YLR022C in the 20519D strain was confirmed by sequencing. 20519D was transformed with plasmid pYC2[Gal10::YLR022CHis6]). sporulated and tetrads dissected to give haploid strain H1 (YLR022C::kanMX4/YC2[Gal10::YLR022CHis6]). Yeast cells were grown in YPD rich medium (1% w/v yeast extract, 2% w/v Bacto Peptone, and 2% w/v glucose). In some experiments, glucose was replaced with 2% galactose (YPG). For tetrad dissections, a Singer MSM micro-manipulator was used. Yeast cells were transformed by the lithium acetate method (24).

**Western Blotting and Antibodies**—Purified full-length His6-Ylr022cp protein was used to immunize rabbits. Polyclonal antisera were affinity-purified according to manufacturer’s instructions (AminoLink Plus Immobilization kit, Pierce). Affinity-purified polyclonal antibody (1: 2000) was incubated overnight and visualized with horseradish peroxidase (HRP)-conjugated donkey anti-rabbit antibody (Amersham Biosciences). Tubulin was visualized using a mouse anti-tubulin monoclonal antibody (Sigma) and an HRP-conjugated sheep anti-mouse antiserum (Amersham Biosciences).

**Flow Cytometry**—Cells were stained with propidium iodide as described (25). Flow cytometry was performed with the FACSCalibur (BD Biosciences), and data analyzed using Cell Quest software.

**Guadinidium Unfolding Assays**—Two protein stock solutions of 1 µM concentration were made in buffer containing 150 mM NaCl, 50 mM Tris, pH 7.4, and 0.5 mM DTE with either 0 or 7.5 M guanidinium hydrochloride (MP Biochemicals Inc.). Circular dichroism measurements were recorded at 222 nm using an AVIV CD spectrophotometer in a quartz cuvette (1-cm path length), final volume 1.8 ml, at 25 °C. A minimum of ten traces were averaged per data point and fit to a linear least-squares alignment of Yhr087wp and domain I of AfSBDS using Superpose (18). (18).

**RESULTS AND DISCUSSION**

**Overall Structure of the A. fulgidus SBDS Orthologue**—We have determined the structure of the A. fulgidus SBDS orthologue AF0491 (NP_068937). The protein was expressed and purified from E. coli and crystallized in space group P1. The 1.9-Å structure was determined by x-ray diffraction using multiple anomalous dispersion data collected from a single crystal generated from SeMet-substituted AFSBDS protein. The crystallographic data are summarized in Table I. The current model includes residues 5–234 of AFSBDS and has been refined to a crystallographic R factor of 21.1% (Rfree = 24.9%) using data between 52.7 and 1.9 Å.

AFSBDS has a three domain architecture (Fig. 1A): the N-terminal domain (I, residues 5–87), central domain (II, residues 88–161), and the C-terminal domain (III, residues 162–234) are arranged in a V-shape of overall dimensions 59 × 73 Å. A central cavity, ~10 Å deep and 24 Å long, is formed at the interface between the three domains.

**FYSH Domain**—Domain I (Asp5–Ile87) (Fig. 1, A and B) has a mixed αβ topology, comprising four β-strands and four α-helices arranged as a three stranded anti-parallel β-sheet (β1-β2-β3-β4) broken by the insertion of a hydrogen-bonded water molecule between strands β3 and β4. Helices a3 and a4 are packed against the concave surface of the β-sheet perpendicular to the long axis of the β-strands, whereas helices a1 and a2 together with the intervening loop line the face of the cleft between domains I and III.

We noted close structural homology between the N-terminal domain of AFSBDS and a single domain protein from S. cerevisiae denoted Yhr087wp (AAB68927.1) whose structure was solved using NMR spectroscopy by the Northeast Structural Genomics Consortium (PDB code, 1NYN) (27). The YHR087W ORF lies within the NAM8-GAR1 intergenic region of S. cerevisiae and encodes a 111 amino acid protein belonging to a small family whose distribution is restricted to fungi. Least squares alignment of Yhr087wp and domain I of AFSBDS using Superpose (18) gives an r.m.s. deviation of 1.47 Å for 59 residues.

The most significant difference between the two structures is a six-residue loop insertion between strands β3 and β4 of YHR087W relative to the AFSBDS structure (Fig. 2). In view of the structural homology between domain I of AFSBDS and Yhr087wp, and the restricted distribution of Yhr087wp homologues to fungi, we have denoted domain I of AFSBDS the FYSH (Fungal, Yhr087wp, Shwachman) domain.

Yhr087wp is non-essential (28) and localizes to both the nucleus and cytoplasm of yeast cells (29). It interacts with the serine/arginine-rich protein kinase Sky1p/Ymr216cp (30) and with Rho3p/Yil118wp (31) in high-throughput two-hybrid screens, and co-immunoprecipitates with the FLAG-tagged yeast ribosomal protein Rpp0p (32). Sky1p phosphorylates the shuttling mRNA export carrier protein Npl3 to promote efficient release of mRNA (33, 34). Physical interaction with Sky1p therefore supports a potential role for Yhr087wp in RNA metabolism.

**Central Domain**—The central domain II (Thr88–Phe161) consists of a compact three-helical bundle (α5–α7). Helices a5 and a6 are connected by a conserved proline-rich loop (α5-α6 loop) (Figs. 1A and 3). The protein that is most closely related struc-
turally to domain II is the C-terminal domain of *E. coli* RuvA that functions in Holliday junction recognition (35). Superposition of *AfsBDS* domain II with the RuvA C-terminal domain gives an r.m.s. deviation of 2.2 Å for 45 common Cα atoms.

**C-terminal Domain**—The C-terminal domain III (Glu162–Gly234) is composed of a four-stranded antiparallel β-sheet (β6-β7-β8-β9) with two α-helices (α6, α9) packing against the concave surface of the sheet (Fig. 1A). The β6β7β8β9 folding topology is typical of a ferredoxin-like fold as defined in the SCOP database (36). Of the large number of protein domains with a ferredoxin-like fold, the closest structural homologue to domain III of *AfsBDS* is domain V of *S. cerevisiae* elongation factor 2 (Efr2p/Ydr385wp) (PDB code, 1N0U, 1N0V) (37). The r.m.s. deviation is 2.36 Å over 63 Cα atoms despite no obvious homology between the protein sequencnes. Efr2p and its bacterial counterpart EF-G are members of the GTPase superfamily of proteins (38). Domain III of *AfsBDS* is also closely related structurally to a conserved *E. coli* protein of unknown function, YigZ (PDB code 1VI7) (39). One or more copies of the β6β7β8β9 ferredoxin-like fold are seen in a number of RNA-binding proteins, including heterogeneous nuclear ribonucleoproteins, proteins implicated in the regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (40). It is the most common RNA binding fold occurring in the translation system where it is referred to as the RNA binding domain or RNA recognition motif. The motif also appears in a few single-stranded DNA-binding proteins.

Fig. 3 shows a structure-based sequence alignment containing the secondary structure elements from both *AfsBDS* and Yhr087wp. The amino acid sequences of representative members of the conserved SBDS protein family and orthologues of Yhr087wp are indicated. As there is significant evolutionary sequence conservation (24% identity, 48% homology) between the human and *A. fulgidus* SBDS proteins, the *AfsBDS* structure represents a paradigm for the entire SBDS family. From the alignment, the predicted domain boundaries for the human SBDS protein are Met1–Val95 (FYSH domain), Ser96–Ile167 (domain II), and Glu168–Glu250 (domain III); for yeast Yhr022p, the predicted boundaries are Met1–Leu95 (FYSH domain), Ser96–Ile169 (domain II), and Val170–Asn250 (domain III).

**Electrostatic Surface Potential**—The electrostatic charge distribution on the *AfsBDS* protein surface was calculated using
the program DelPhi (41). One face of the protein consists of an extended positively charged surface that includes contributions from all three domains (Fig. 1C). On the opposite face of the protein, the FYSH domain is predominantly negatively charged, whereas domains II and III are mainly basic. Surface residues mutated in disease (Arg19, Lys118, Arg126, Lys148, Arg169) are predicted to map to an analogous extended basic surface of the human SBDS protein (Fig. 1C). Since nucleic acid-binding proteins often use basic residues to contact the sugar-phosphate backbone, the extended basic electrostatic surface potential would be consistent with such a role for AfSBDS.

**Impact of SDS-associated Mutations on Protein Stability**—By direct sequencing of SBDS alleles from individuals judged clinically to have SDS, fourteen novel disease-related mutations have been identified, including missense, frameshift, splice site mutations, and deletions (Table II). The availability of the three-dimensional x-ray structure of the AfSBDS protein together with the evolutionary conservation of amino acid sequence allows us to rationalize the consequences of mutations associated with SDS (6, 14, 15) by mapping the SDS-associated mutations onto the molecular surface of the AfSBDS protein (Fig. 4A). The positions of the mutations within the structure-based sequence alignment are indicated in Fig. 3. Strikingly, this analysis reveals that the FYSH domain is the most frequent target for SDS-associated mutations. On the basis of our structure, SDS-associated mutations can be categorized into three main groups: first, the common protein truncations, second rare missense mutations that are predicted to destabilize the fold of the SBDS Protein Structure
protein and third, rare missense mutations that are predicted to modify surface epitopes.

In the first group, the K62X mutation truncates the SBDS protein within the δ4-α3 loop of the FYSH domain, whereas the C84fs1 frameshift mutation is predicted to result in truncation within helix α4. These mutations might be expected to result in loss of protein function. The second group includes buried sites invariant residues in yeast Ylr022cp (N34I, E44G, R100A) or in human SBDS (E44G, K118N, E28A, R100A) that are predicted to significantly perturb protein function. The third group of mutations is predicted to affect surface epitopes do not indeed result in global unfolding of the protein (Fig. 4C). By contrast, we were unable to express sufficient quantities of the SBDS mutants C31W, L71P, or I87S to test in our assay, strongly suggesting that these mutations perturb the overall fold of the protein and are predicted to significantly perturb protein function in vivo. Although the number of samples tested is limited, these data support the impact of the mutations on protein stability in vitro predicted by our structural analysis.

Ylr022cp Is Required for G1 Cell Cycle Progression—To test the functional consequences of SDS-related mutations in vivo, we exploited the essential role of the S. cerevisiae SBDS orthologue Ylr022cp (9) to examine the ability of variant Ylr022c proteins to restore yeast cell growth in a complementation assay. We created a yeast strain (H1) in which YLR022C was deleted from the genome and cell viability maintained by a centromeric plasmid expressing YLR022C under the control of the GAL10 promoter (42) that can be induced in galactose-containing medium (YPD) and repressed in glucose-containing medium (YPD) (see “Experimental Procedures”). Growth of H1
or isogenic wild-type parental haploid (W303-1) cells was monitored by serial cell density ($\lambda_{agg}$) measurements. Following transfer of H1 cells to YPD medium, Ylr022cp became undetectable by immunoblotting after 2 h, with delayed onset of a progressive growth arrest from around 7 h (Fig. 5, A and B). Correlating with the growth arrest, flow cytometry demonstrated a progressive increase in the proportion of H1 cells with a 1N DNA content, indicative of delayed progression through $G_1$ phase of the cell cycle (Fig. 5C).

We first tested the ability of a human SBDS allele to complement the growth defect of H1 cells in YPD medium. Centromeric pRS314 plasmids (TRP) carrying either a wild-type yeast YLR022C or human SBDS allele under the control of the endogenous YLR022C promoter were constructed (see “Experimental Procedures”) and transformed into H1 cells. Because Ylr022cp is undetectable in H1 cells by immunoblotting 2 h after the shift to YPD medium, the growth phenotype of H1 cells in YPD medium is attributable solely to protein expression from the transformed pRS314 plasmids. Growth of transformed H1 cells was monitored at 30 °C in URA-TRP selective medium under restrictive conditions, and as shown in Table III, unlike the wild-type YLR022C allele, the human SBDS allele failed to rescue the growth defect of H1 cells in YPD medium. Therefore, all further genetic complementation assays were performed using variant YLR022C alleles transformed into H1 cells.

**In Vivo Functional Consequences of SBDS-associated Mutations—**A comprehensive series of YLR022C variants was generated and transformed into H1 cells. Fig. 6 shows representative quantitative growth curves of transformed H1 cells under restrictive conditions and the results are also summarized in Table III. Variant YLR022C alleles were classified into four groups based on their ability to complement growth of H1 cells in YPD medium: full (+++) partial (+++), or no (-) growth complementation. Several disease-related mutations had severe effects on yeast cell growth: complete loss of function was associated with the K62X and S96X truncations and with the C31W, N34I, and L71P missense mutations, whereas severe growth impairment was associated with I87S (+). Partial loss of function was associated with K118N (++), but the majority of SBDS-related missense mutations tested showed no apparent growth defects in our assay (R19Q, E44G, K118N, K126A, S143L+K148R, Q153R, R169C). Surprisingly, with the exception of Arg$^{100}$ (++) alanine substitutions of evolutionarily invariant residues (Glu$^{28}$, Lys$^{42}$, Glu$^{84}$) or conserved charged residues exposed on the surface of the central three-helical bundle (Lys$^{98}$, Lys$^{148}$, Glu$^{156}$, Lys$^{159}$) showed no apparent growth defects.

Consistent with a critical role for the FYSH domain, the N-terminal deletion variant Δ1–95 was non-functional. Deletion of domain III (M174X) or the missense mutation R169P (but interestingly, not the disease-related mutation R169C), resulted in partial (+++) loss of function. Taken together, the Δ1–95 deletion and M174X truncation mutations indicate that the FYSH domain alone is insufficient to complement Ylr022cp function, and reveal that domains I and II together are essential for function.

Importantly, these data demonstrate that the disease-related Ylr022cp variants K62X and S96X are null mutations. S96X introduces a premature stop codon at the junction between the FYSH domain and domain II, analogous to the effect of the premature protein truncation by frameshift that is predicted for the common 258 + 2T → C (C84fs3) disease mutation. This indicates, therefore, that the most common predicted disease-associated SBDS truncations are null mutations. The FYSH domain hydrophobic core mutations (C31W, L71P, I87S) are particularly deleterious to Ylr022cp function, presumably as a consequence of impaired protein folding and consistent with our inability to express these variant proteins in vitro. As the N34I mutation does not appear to perturb global protein folding (Fig. 4C), the consequence of the introduction of a bulky hydrophobic Ile residue on the protein surface may be the reduction of protein solubility. The surface-exposed residues Lys$^{118}$ and Arg$^{100}$ may be critical for interaction with a potential partner protein or ligand, as the respective K118N and R100A missense mutations significantly affect Ylr022cp function in vivo in the absence of an effect on global folding (Figs. 4C and 6, Table III).

**Insight into the Pathogenesis of SDS—**Affected individuals from 50% of families are compound heterozygotes with respect to the 258 + 2T → C (C84fs3) splice site and the 183–184TA → CT (K62X) premature stop codon mutations (6). Of 158 families reported, no homozygotes for the 183–184TA → CT (K62X) premature stop mutation were identified (6). By contrast, seven families were found to be homozygous for the 258 + 2T → C (C84fs3) splice-site mutation. This strongly suggests that the 183–184TA → CT (K62X) premature stop is a null mutation, and that a homozygous K62X-null mutation is embryonic lethal. By contrast, the 258 + 2T → C splice-site mutation is an obligate hypomorph with respect to SBDS function. Consistent with these observations, our yeast genetic complementation analysis indeed demonstrates that the K62X truncation is a null mutation. However, the S96X truncation reveals that the highly conserved FYSH domain is insufficient to complement wild-type Ylr022cp function. This indicates that the hypomorphic nature of the C84fs3 allele is not the result of residual function associated with the predicted SBDS truncation fragment. Rather, we speculate that the GC dinucleotide donor splice site that arises as a consequence of the 258 + 2T → C mutation may result in inefficient low level splicing activity and SBDS protein expression. Alternative splicing may yield a transcript that results in a functional SBDS protein product. Interestingly, the majority of SDS-related missense mutations fail to abrogate Ylr022cp function. Taken together, these results indicate that the clinical phenotype of SDS is associated with expression of functionally hypomorphic SBDS alleles.

The structural features of the AfSBDS protein together with the homology to Yhr087wp described herein, are consistent
with the proposed role for the SBDS protein family in RNA metabolism (6), although we have insufficient data to further define the specific function at present. Our structural and genetic analysis reveals that the novel FYSH domain is frequent target for disease mutations and is required for protein function in vivo, suggesting that elucidation of the specific role of the FYSH domain may provide critical insights into the function of the SBDS protein family. It will be of interest to determine potential interactions (genetic/physical) common to both Yhr022cp and Yhr087wp to further elucidate the function of these novel proteins.

Coordinates—The coordinates have been deposited in the Protein Data Bank (accession code 1T95).

Acknowledgments—We thank G. Leonard for help with MAD data collection at synchrotron beamline ID29 at the ESRF, R. Williams for advice and assistance in data collection, J. Löwe for providing archaeal genomic DNA, Phil Evans for crystallographic assistance, J. Kilmartin advice and assistance in data collection, J. Löwe for providing archaeal collection at synchrotron beamline ID29 at the ESRF, R. Williams for genetic analysis reveals that the novel FYSH domain is a

REFERENCES

1. Shwachman, H., Diamond, L. K., Oski, F. A., and Khaw, K. T. (1964) J. Pediatri. 65, 465–663
2. Bodian, M., Sheldon, W., and Lightwood, R. (1964) Acta Pediaatria 53, 282–293
3. Ginzberg, H., Shin, J., Ellis, L., Morrison, J., Ip, W., Dror, Y., Freedman, M., Heitlinger, L. A., Belt, M. A., Corey, M., Rommens, J. M., and Durie, P. R. (1999) J. Pediatr. 135, 81–88
4. Smith, O. P., Hann, I. M., Chessells, J. M., Reeves, B. R., and Milla, P. (1996) J. Pediatri. 129, 247–250
5. Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Arnheim, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) Nature 402, 623–627
6. Io, T., Chiba, T., Ozawa, Y., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 4569–4574
7. Ho, Y., Grubauer, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Beutler, K., Yang, L., Wolting, C., Donaldson, I., Schondorf, D., Shewnawara, J. L., Taggart, J., Goudreau, M., Muskat, B., Alfarno, C., Dewar, D., Lin, Z., Michalikova, K., Willems, A. B., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, B. H., Hansen, L. H., Jessen, H. H., Jentoft, N., Nielsen, E., Oldham, M., Huneault, M., Huang, R., Aerts, J., Brem, S., Sturtevant, A. M., Jansen, R., Boeke, J., brushes, H., Boeke, J., Snyder, M., Philipsen, P., Davis, R. W., and Johnston, M. (2002) Science 295, 901–906
8. Wu, W., Hughes, T. R., Davigwala, A. P., Robinson, M. D., Stoughton, R., and Altschuler, S. J. (2002) Nat. Genet. 33, 97–101
9. Bateman, A., Birney, E., Cerutti, L., Durbin, R., Eddy, S. R., Griffith-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) Nucleic Acids Res. 30, 276–278
10. Koskin, E. V., Wolf, Y. I., and Aravind, L. (2001) Genome. Res. 11, 240–252
11. Winn, J. C., Shoemaker, D. D., Astromoff, A., Liang, H., Andersen, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., Chu, M., Connelly, C., Davis, K., Dietrich, F., Dow, S., W. E., El-Bakkoury, M., Foury, F., Friend, S. H., Gentilena, E., Giaever, G., Hegemann, J. H., Jones, T. A., Lobb, M., Liao, H., Liebundug, N., Lockhart, D. J., Lu, J., Lu, J., Matsuda, M., Maret, N., Meier, M., Mittman, M., Narasim, C., Rudolphi, C., Revuelta, J. L., Roberts, C. J., Ross-Macdonald, P., Scheuring, E., Snyder, M., Sokkho-Mahadeo, S., Strosa, R. K., Veronese, S., Voet, M., Volckaert, G., Ward, T. R., Ywosski, R., Yuen, G. S., Yuen, Y., Zimmermann, K., Philipsen, P., Johnston, M., and Davis, R. W. (1999) Science 285, 901–906