

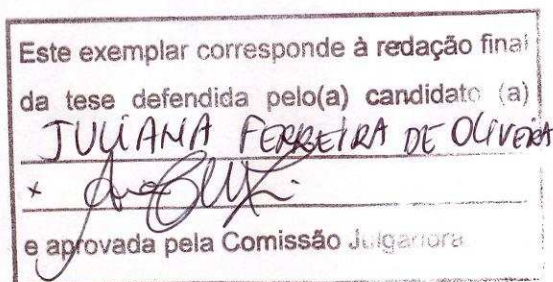
UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



JULIANA FERREIRA DE OLIVEIRA

**“ESTUDOS ESTRUTURAIS E FUNCIONAIS DE PROTEÍNAS
DA FAMÍLIA SBDS COM ÊNFASE NAS ORTÓLOGAS DE
Trypanosoma cruzi E HUMANA”**



Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Biologia Funcional e Molecular, na área de Bioquímica.

Orientadora: Profa. Dra. Ana Carolina de Mattos Zeri

Co-Orientador: Prof. Dr. Nilson Ivo Tonin Zanchin

Campinas, 2009

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

OL42e

Oliveira, Juliana Ferreira de
Estudos estruturais e funcionais de proteínas da família
SBDS com ênfase nas ortólogas de *Trypanosoma cruzi* e
humana / Juliana Ferreira de Oliveira. – Campinas, SP:
[s.n.], 2009.

Orientadores: Ana Carolina de Mattos Zeri, Nilson Ivo
Tonin Zanchin.

Tese (doutorado) – Universidade Estadual de
Campinas, Instituto de Biologia.

1. Proteína SBDS. 2. Ribossomos - Biossíntese. 3.
Trypanosoma cruzi. 4. Ressonância magnética nuclear.
5. Proteínas - Estrutura. I. Zeri, Ana Carolina de Mattos.
II. Zanchin, Nilson Ivo Tonin. III. Universidade Estadual
de Campinas. Instituto de Biologia. IV. Título.

(rcdt/ib)

Título em inglês: Structural and functional analysis of the SBDS protein family.

Palavras-chave em inglês: SBDS protein; Ribosomes - Biosynthesis; *Trypanosoma cruzi*;
Nuclear magnetic resonance; Proteins – Structure.

Área de concentração: Bioquímica.

Titulação: Doutora em Biologia Funcional e Molecular.

Banca examinadora: Ana Carolina de Mattos Zeri, Fabio Ceneviva Lacerda Almeida, Ronaldo
Alves Pinto Nagem, Luis Eduardo Soares Netto, Jörg Kobarg.

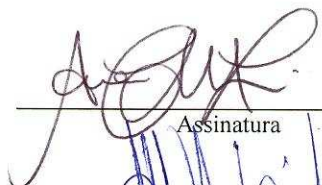
Data da defesa: 20/08/2009.

Programa de Pós-Graduação: Biologia Funcional e Molecular.

Campinas, 20 de agosto de 2009

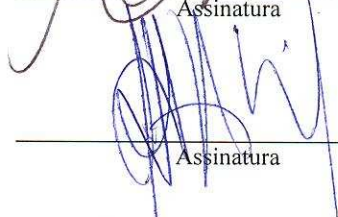
BANCA EXAMINADORA

Profa. Dra. Ana Carolina de Mattos Zeri (Orientadora)



Assinatura

Prof. Dr. Fabio Ceneviva Lacerda Almeida



Assinatura

Prof. Dr. Ronaldo Alves Pinto Nagem



Assinatura

Prof. Dr. Luis Eduardo Soares Netto



Assinatura

Prof. Dr. Jorg Kobarg



Assinatura

Profa. Dra. Carmen Veríssima Ferreira

Assinatura

Profa. Dra. Anete Pereira de Souza

Assinatura

Profa. Dra. Cláudio Chrysostomo Werneck

Assinatura

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, pelo auxílio financeiro concedido a este projeto.

Ao Laboratório Nacional de Luz Síncrotron, LNLS, pela infra-estrutura oferecida.

Ao Nilson Zanchin, pela co-orientação e pela oportunidade de realizar este projeto de doutorado.

À Beatriz Guimarães, sempre muito paciente, solícita e atenciosa, pela sua orientação.

À Ana Carolina Zeri, por me aceitar como aluna no decorrer do doutorado e pela paciência e orientação.

Ao Mauricio Sforça, pela co-orientação em RMN. Agradeço sua paciência e todos os seus ensinamentos.

Aos Profs. Marco Krieger e Beatriz Castilho, por me receberem em seus laboratórios, pela atenção e disponibilidade.

Aos pesquisadores do IBMP Christian Probst, Daniela Pavoni e Alejandro Correa e a todos os alunos que me auxiliaram durante meus “estágios” no IBMP.

Às Profas. Carla Oliveira e Tharin Blumenschein, pela colaboração nesse projeto, pelas sugestões e críticas oferecidas.

Aos Profs. Jorg Kobarg, Claudio Werneck e Mario Murakami, pela participação no exame de qualificação e avaliação deste trabalho.

Aos membros da banca - Profs. Fabio Almeida, Ronaldo Nagem, Luis Eduardo Netto e Jorg Kobarg - pela disposição em participar da avaliação final deste trabalho.

Ao programa de Pós-graduação em Biologia Funcional e Molecular da UNICAMP.

Aos pesquisadores, funcionários e alunos do LNLS, pelo suporte técnico, convivência e troca de experiências.

Aos amigos do grupo LBM, por terem me recebido muito bem e pelos ensinamentos: Bia, Cedric, Celisa, Flavia, Ju, Leandro, Pati, Romenia, Sandra e Thais.

Aos amigos do grupo de cristalografia, pela ótima convivência: Amanda, Camila, Carla, Carol, Fabio, Karen, Kelly, Marcos, Mario, Nadia, Pri, Tati, Vanessa.

Aos novos amigos do grupo de RMN: Alexandre, Camila, Fabio e Jorge.

Aos amigos do CeBiME, pela troca de experiência, sugestões e críticas: Aline, Ana Olívia, André, Andrés, Daniel MT, Daniel Lanza, Eugenia, Guga, Gustavo, Julio, Leticia, Luli, Marcos, Margareth, Marina, Naty, Renata, Rosi, Tiago, Veruska.

Às queridas amigas Adriana, Andreia, Celisa, Elaine, Givanil, Tereza, e Zildene, pelo auxílio prático e teórico, sugestões e críticas e pela ótima convivência. Agradeço por serem sempre muito solícitas.

Aos meus queridos amigos químicos Alessandra, Amadeu, Helder e Samanta.

Ao meu companheiro e sempre amigo Fabio. Agradeço por estar sempre ao meu lado, não me permitindo desistir em momento algum.

À minha família, que sempre me apoiou e incentivou muito. Principalmente aos meus pais, por me passarem os valores necessários para eu chegar até aqui.

Por fim, agradeço a Deus por tudo.

ÍNDICE

LISTA DE TABELAS	viii
LISTA DE FIGURAS	ix
LISTA DE ABREVIações	xi
RESUMO	xiv
ABSTRACT	xvi
CAPÍTULO 1. INTRODUÇÃO	1
1.1. Biossíntese de ribossomos	1
1.2. Fatores envolvidos na biossíntese de ribossomos	2
1.3. A doença SDS e a proteína SBDS	6
1.4. Objetivos	12
CAPÍTULO 2. ESTUDOS ESTRUTURAIS DE PROTEÍNAS DA FAMÍLIA SBDS ..	14
2.1. Materiais e Métodos	15
2.1.1. Construção dos vetores de expressão	15
2.1.2. Expressão das proteínas	16
2.1.3. Purificação das proteínas e remoção da cauda de histidina	17
2.1.4. Cromatografia de exclusão molecular analítica	18
2.1.5. Proteólise limitada	19
2.1.6. Ensaio para análise de agregação	19
2.1.7. Análises de Dicroísmo Circular (CD)	20
2.1.8. Ensaios de cristalização	21
2.1.9. Experimentos de Ressonância Magnética Nuclear (RMN)	22
2.2. Resultados	26
2.2.1. SBDS de <i>Pyrococcus abyssi</i> : PaUPF0023	26
2.2.2. SBDS de <i>Saccharomyces cerevisiae</i> : YLR022c	29
2.2.3. SBDS de <i>Trypanosoma cruzi</i> : TcSBDS	31
2.2.4. SBDS de <i>Homo sapiens</i> : HsSBDS	33
2.3. Discussão	50
CAPÍTULO 3. ANÁLISE ESTRUTURAL E FUNCIONAL DAS PROTEÍNAS SBDS DE <i>Trypanosoma cruzi</i> (TcSBDS) E <i>Homo sapiens</i> (HsSBDS)	53

3.1. ARTIGO I - Characterization of the Trypanosoma cruzi ortholog of the SBDS protein reveals an intrinsically disordered extended C-terminal region showing RNA-interacting activity	54
3.2. ARTIGO II - Structure, dynamics and RNA interaction analysis of the human SBDS protein	64
CAPÍTULO 4. CONCLUSÕES E PERSPECTIVAS	90
REFERÊNCIAS BIBLIOGRÁFICAS	92

LISTA DE TABELAS

Tabela 1.1: Mutações no gene <i>SBDS</i> relacionadas à doença SDS	11
Tabela 2.1: Oligonucleotídeos utilizados para produção dos clones de expressão ..	15
Tabela 2.2: Região correspondente às diferentes construções da HsSBDS.....	16
Tabela 2.3: Aditivos utilizados na tentativa de diminuir a agregação da proteína TcSBDS truncada.....	19
Tabela 2.4: Tampões e concentrações utilizados nas medidas de CD	20
Tabela 2.5: Descrição dos ensaios de cristalização realizados com as diferentes proteínas.....	21
Tabela 2.6: Descrição das amostras da HsSBDS utilizadas para os experimentos de RMN	23
Tabela 2.7: Análise da TcSBDS truncada em diferentes pHs por experimentos de DLS e espectroscopia de absorção de luz ultravioleta	32

LISTA DE FIGURAS

Figura 1.1: Esquema do processamento do pré-rRNA em <i>Saccharomyces cerevisiae</i>	3
Figura 1.2: Representação de alguns dos fatores envolvidos no processo de maturação e exportação das subunidades ribossomais 40S e 60S.....	4
Figura 1.3: Alinhamento das sequências de aminoácidos de membros representativos da família SBDS.....	8
Figura 1.4: Estrutura da proteína SBDS de <i>Archaeoglobus fulgidus</i> (AfSBDS).....	9
Figura 2.1: Análises de dicroísmo circular da PaUPF0023.....	27
Figura 2.2: Cristais da proteína PaUPF0023.....	27
Figura 2.3: Purificação e ensaios de cristalização da YLR022c.....	30
Figura 2.4: Análises de dicroísmo circular da YLR022c.....	30
Figura 2.5: Purificação da proteína TcSBDS truncada.....	32
Figura 2.6: Purificação da proteína SBDS humana.....	34
Figura 2.7: Cromatografia de exclusão molecular analítica da proteína HsSBDS.....	35
Figura 2.8: Análises de dicroísmo circular da HsSBDS.....	35
Figura 2.9: Micro-cristais obtidos durante ensaios de cristalização com a proteína SBDS humana.....	36
Figura 2.10: Proteólise limitada da HsSBDS.....	37
Figura 2.11: Espectro de massa da HsSBDS após proteólise limitada com tripsina.....	37
Figura 2.12: Purificação dos domínios da HsSBDS.....	38
Figura 2.13: Ilustração dos experimentos de tripla ressonância coletados para o assinalamento sequencial da proteína HsSBDS.....	40
Figura 2.14: Assinalamento sequencial da proteína HsSBDS utilizando-se o programa Smartnotebook.....	42
Figura 2.15: Chemical Shift Index (CSI) para os átomos C α da proteína HsSBDS..	43
Figura 2.16: Ilustração da transferência de magnetização envolvida no experimento de HCCH-TOCSY.....	44
Figura 2.17: Assinalamento da cadeia lateral do aminoácido isoleucina 10 da HsSBDS.....	45
Figura 2.18: Experimentos de troca hidrogênio/deutério (H/D) da HsSBDS.....	47

Figura 2.19: NOEs característicos de estruturas secundárias.....	48
Figura 2.20: Etapas do cálculo da estrutura da proteína HsSBDS	49

LISTA DE ABREVIACÕES

3'ETS	Espaçador externo 3' (<i>External Transcribed Spacer 3'</i>)
5'ETS	Espaçador externo 5' (<i>External Transcribed Spacer 3'</i>)
AfSBDS	Proteína SBDS de <i>Archaeoglobus fulgidus</i>
BMRB	Banco de dados de RMN (<i>Biological Magnetic Resonance Data Bank</i>)
C-terminal	Extremidade carboxi terminal da proteína
CD	Dicroísmo circular (<i>Circular Dichroism</i>)
cDNA	DNA complementar
CHH	Hipoplasia cartilagem-cabelo (<i>Cartilage-Hair Hypoplasia</i>)
CNS	Programa computacional para determinação estrutural a partir de dados de cristalografia e RMN (<i>Crystallography & NMR System</i>)
COSY	Espectroscopia de correlação homonuclear (<i>CORrelation SpectroscopY</i>)
CYANA	Programa computacional para cálculo semi-automatizado de estrutura de proteínas por RMN (<i>Combined assignment and dYnamics Algorithm for NMR Applications</i>)
DBA	Anemia Diamond-Blackfan (<i>Diamond-Blackfan Anemia</i>)
DC	Disqueratose congênita (<i>Dyskeratosis Congenita</i>)
DLS	Espalhamento dinâmico de luz (<i>Dynamic Light Scattering</i>)
DNA	Ácido desoxiribonucleico (<i>Deoxyribonucleic Acid</i>)
DTT	1,4-ditiotreitol
DYANA	Programa computacional para cálculo de estrutura de proteínas por RMN (<i>DYnamics Algorithm for NMR Applications</i>)
EMSA	Ensaio de interação proteína-RNA (<i>Electrophoretic mobility shift assay</i>)
FYSH	Domínio N-terminal da SBDS (<i>Fungal, Yhr087wp, Shwachman</i>)
HSQC	Espectroscopia de correlação heteronuclear (<i>Heteronuclear Single Quantum Coherence</i>)
HsSBDS	Proteína SBDS de <i>Homo sapiens</i>
IPTG	Isopropil β -D-1-tiogalactopiranosídeo
ITS1	Espaçador interno 1 (<i>internal transcribed spacer 1</i>)
ITS2	Espaçador interno 2 (<i>internal transcribed spacer 2</i>)
LB	Meio de cultura de bactérias "Luria Bertani"

mRNA	RNA mensageiro
mthSBDS	Proteína SBDS de <i>Methanothermobacter thermautotrophicus</i>
NOE	Efeito Nuclear Overhauser (<i>Nuclear Overhauser Effect</i>)
NOESY	Espectroscopia do Efeito Nuclear Overhauser (<i>Nuclear Overhauser Effect Spectroscopy</i>)
N-terminal	Extremidade amino terminal da proteína
OD ₆₀₀	Densidade ótica a 600 nanômetros
PaUPF0023	Proteína SBDS de <i>Pyrococcus abyssi</i>
PCR	Reação em cadeia da polimerase (<i>Polymerase Chain Reaction</i>)
PDB	Banco de dados de estrutura de proteínas (<i>Protein Data Bank</i>)
pH	Potencial hidrogeniônico
PMSF	Fluoreto de fenilmetilsulfonila
RMN	Ressonância Magnética Nuclear (<i>Nuclear Magnetic Resonance, NMR</i>)
RNA	Ácido Ribonucleico (<i>Ribonucleic Acid</i>)
RNaseP	Ribonuclease P
RNP	Ribonucleoproteína
RRM	Domínio de reconhecimento ao RNA (<i>RNA Recognition Motif</i>)
rRNA	RNA ribossomal
SBDS	Proteína associada à síndrome Shwachman-Bodian-Diamond (<i>Shwachman-Bodian-Diamond Syndrome</i>)
<i>SBDS</i>	Gene que codifica a proteína SBDS
<i>SBDSP</i>	Pseudogene de <i>SBDS</i>
<i>Sdo1p</i>	Proteína SBDS de <i>Saccharomyces cerevisiae</i>
SDS	Síndrome Shwachman-Diamond (<i>Shwachman-Diamond Syndrome</i>)
SDS-PAGE	Eletroforese em gel de poliacrilamida com dodecil sulfato de sódio
snoRNA	Pequeno RNA nucleolar (<i>small nucleolar RNA</i>)
snoRNP	Pequena partícula ribonucleoprotéica nucleolar (<i>small nucleolar ribonucleoprotein</i>)
TcSBDS	Proteína SBDS de <i>Trypanosoma cruzi</i>
TCTP	Proteína associada a tumor, regulada em nível de tradução (<i>translationally controlled tumor-associated protein</i>)
TEV	Protease do vírus do tabaco (<i>Tobacco Etch Virus protease</i>)
T _m	Temperatura média de transição

TOCSY	Espectroscopia de correlação homonuclear total (<i>T</i> otal <i>C</i> orrelation <i>S</i> pectroscop <i>Y</i>)
tRNA	RNA transportador
YLR022c	Proteína SBDS de <i>Saccharomyces cerevisiae</i>

RESUMO

Proteínas da família SBDS (*Shwachman-Bodian-Diamond Syndrome*) ocorrem largamente na natureza e são bastante conservadas, apresentando ortólogas em Archaea e eucariotos. Estudos de análises genômica e biofísica tem relacionado a SBDS com o metabolismo de RNA e biosíntese de ribossomos. O gene ortólogo da SBDS de Archaea está localizado em um operon conservado que contém genes do processamento de RNA; estudos de perfil de expressão gênica tem agrupado o gene da proteína SBDS de *Saccharomyces cerevisiae*, *Sdo1p*, com fatores do processamento de rRNA e estudos de análise proteômica identificaram a interação da proteína *Sdo1p* com fatores da biossíntese de ribossomos; ortólogas de planta contém um C-terminal estendido apresentando motivo de ligação a RNA. Mutações identificadas no gene *SBDS* tem sido relacionadas com a síndrome Shwachman-Diamond (SDS), uma doença caracterizada por insuficiência exócrina pancreática e disfunção na medula óssea, cujos pacientes apresentam grandes chances de desenvolver leucemia. SDS representa, portanto, um importante modelo para entender os processos envolvidos no desenvolvimento da leucemia. O objetivo principal desse trabalho consistiu na caracterização estrutural e funcional de proteínas da família SBDS. Foram realizados ensaios de cristalização com ortólogas da SBDS de Archaea, levedura, tripanossoma e humana. A SBDS de *Pyrococcus abyssi* foi cristalizada, porém os cristais difrataram a baixa resolução (3,50 Å). A caracterização da SBDS ortóloga de *Trypanosoma cruzi* (TcSBDS) mostrou que esta proteína contém uma região C-terminal estendida. Ensaio de proteólise limitada, dicroísmo circular e espectroscopia por Ressonância Magnética Nuclear (RMN) indicaram que a região adicional da TcSBDS se comporta como um fragmento de proteína intrinsicamente desenovelado, responsável pela interação da TcSBDS com RNA, verificada por ensaios de *Electrophoretic Mobility Shift Assay* (EMSA). Também foi realizada a determinação da estrutura da ortóloga humana (HsSBDS) em solução por espectroscopia de RMN. A proteína HsSBDS é composta de três domínios bem estruturados, apresentando mobilidade conformacional entre os domínios N-terminal e central. Experimentos de titulação de RNA, novamente utilizando-se RMN, possibilitaram a confirmação da interação direta da SBDS humana com RNA. A região de ligação ao RNA foi identificada no N-terminal da

proteína, região bastante conservada na família e considerada o principal alvo das mutações relacionadas à doença SDS.

ABSTRACT

The Shwachman-Bodian-Diamond syndrome (SBDS) protein family occurs widely in nature and is highly conserved, with orthologues in Archaea and eukaryotes. Genomic and biophysical studies have suggested involvement of this protein in RNA metabolism and in ribosome biogenesis. Archaeal SBDS orthologue genes are located within highly conserved operons that include RNA-processing genes; transcriptional profiling analysis has clustered the yeast ortholog protein *Sdo1p* with rRNA processing factors and proteomic analysis have identified potential interactions between Sdo1p and ribosome biogenesis factors; several plant SBDS orthologues contain extended C-terminal region with putative RNA binding motif. Mutations in the *SBDS* gene are associated to the Shwachman-Diamond syndrome (SDS), a rare multisystem disorder characterized by exocrine pancreatic insufficiency, bone marrow dysfunction, and an increased risk of acute myeloid leukemia. SDS therefore represents an extremely useful model for understanding leukaemogenesis. The objective of the present work was the structural and functional characterization of the SBDS protein family. SBDS orthologues from Archaea, yeast, trypanosomatid and human were assayed for crystallization. The Archaeal SBDS orthologue, PaUPF0023 in *Pyrococcus abyssi*, was crystallized, but the crystals diffracted to a relatively low resolution (3.50 Å). Characterization of the *Trypanosoma cruzi* SBDS ortholog (TcSBDS) by using limited proteolysis, circular dichroism and NMR analyses indicated that the C-terminal additional region of TcSBDS behaves as a natively unfolded protein segment, responsible for TcSBDS-RNA interaction activity in electrophoretic mobility shift assays. We have also determined the solution structure and backbone dynamics of the human SBDS protein using NMR spectroscopy. The overall structure of human SBDS comprises three well-folded domains with conformational exchange in the linker between the N-terminal and the central domains. RNA titration experiments using NMR spectroscopy provide evidence that human SBDS interacts with RNA via the N-terminal domain, a conserved region in the SBDS family and the most frequent target for SDS-associated mutations.

CAPÍTULO 1 - INTRODUÇÃO

1.1. BIODISSÍNTESE DE RIBOSSOMOS

Os ribossomos são complexos macromoleculares responsáveis por realizar a síntese protéica dentro das células. Duas subunidades de tamanhos diferentes compõem os ribossomos, cada uma contendo pelo menos uma molécula de RNA ribossômico (rRNA) e uma grande quantidade de proteínas ribossomais (Gesteland *et al.*, 1999). Os ribossomos de procariotos e eucariotos são bastante semelhantes na sua forma, mas em geral, os ribossomos eucariotos são maiores e mais complexos (Wool, 1979). Nos ribossomos de procariotos, 70S, a subunidade maior (50S) contém dois rRNAs, o 5S e o 23S, e 34 proteínas; a subunidade menor (30S) consiste de apenas um rRNA (16S) associado a 23 proteínas. Já nos ribossomos de eucariotos, 80S, na subunidade maior são encontrados três rRNAs (5S, 25S/28S e 5.8S) e 46 proteínas e na subunidade menor (40S) um rRNA (18S) e 32 proteínas (Lecompte *et al.*, 2002).

Muitas moléculas de RNA de procariotos e praticamente todas as moléculas de RNA de eucariotos são sintetizadas como moléculas precursoras que sofrem processamento e modificação após a transcrição até adquirirem a estrutura final madura. O processamento do rRNA inclui clivagens do pré-rRNA, conversão de uridinas a pseudouridinas e metilação de nucleotídeos e bases. No caso dos organismos eucariotos, a transcrição, o processamento e a maturação dos RNAs ocorrem no nucléolo da célula, havendo um organizado sistema de transporte para exportar os RNAs maduros e as subunidades ribossomais para o citoplasma (Alberts *et al.*, 2002; Gesteland *et al.*, 1999).

A biogênese dos ribossomos é um processo bastante complexo. A maquinaria de síntese ribossomal é conservada em eucariotos e encontra-se bem caracterizada na levedura *Saccharomyces cerevisiae*. Neste organismo, três dos quatro rRNA são transcritos pela RNA polimerase I, na forma de um precursor (35S), que contém as seqüências dos rRNAs 18S, 5.8S e 25S flanqueadas por regiões espaçadoras nas extremidades 5' e 3', denominadas 5'ETS e 3'ETS (ETS, *external transcribed spacer*) e intercaladas por duas seqüências espaçadoras internas, a ITS1 e ITS2 (ITS, *internal transcribed spacer*). Durante o processo de maturação (Figura 1.1),

essas seqüências espaçadoras são removidas por uma seqüência coordenada de reações endo- e exonucleolíticas e nucleosídeos específicos sofrem metilação das bases ou riboses, além de certas uridinas serem isomerizadas à pseudo-uridinas (Tollervey & Kiss, 1997; Venema & Tollervey, 1999). O rRNA 5S é transcrito separadamente pela RNA polimerase III e o transcrito primário sofre maturação independente. Sua extremidade 3' é rapidamente clivada por exoribonucleases para gerar o rRNA maduro 5S, que será reunido ao pré-60S ainda no nucléolo (Fromont-Racine *et al.*, 2003). À medida que o pré-rRNA é processado ocorre sua associação com as proteínas ribossomais e, após a montagem completa da subunidade, esta é transportada para o citoplasma.

1.2. FATORES ENVOLVIDOS NA BIODSSÍNTESE DE RIBOSSOMOS

A simples adição de proteínas ribossomais ao rRNA não resulta em subunidades ribossomais funcionais. Ensaios *in vitro* com sistemas bacterianos demonstraram que a montagem das subunidades ribossomais a partir de proteínas individuais e rRNAs ocorre naturalmente, porém esse processo requer energia (Nierhaus, 1991). *In vivo*, fatores adicionais que diminuem a energia de ativação das reações envolvidas na montagem dos complexos proteína-rRNA são necessários, como helicases e moléculas chaperonas (Nierhaus, 1991; Maki *et al.*, 2002). Em eucariotos, o processo é ainda mais complexo, pois além das proteínas constituintes das subunidades ribossomais, mais de 170 fatores se associam transientemente aos pré-rRNAs e são necessários para a efetiva maturação dos ribossomos (Figura 1.2). Esses fatores incluem componentes da maquinaria transcricional da RNA polimerase I, pequenos complexos de ribonucleoproteínas nucleolares (snoRNPs) envolvidos na modificação de rRNA, helicases, nucleases, fatores envolvidos no transporte nucléolo-citoplasma, dentre outros (Fatica & Tollervey, 2002). Esses fatores interagem com o pré-rRNA de maneira ordenada e podem se localizar no nucléolo, núcleo ou citoplasma. Já está bem estabelecido, por exemplo, que a ligação do complexo U3 snoRNP ao pré-rRNA é essencial para que ocorram as primeiras clivagens do pré-RNA e a montagem do 40S (Venema & Tollervey, 1999).

extremidade 5'ETS nos sítios A₀ e A₁ (gerando os pré-rRNAs 33S e 32S, respectivamente) e então clivado no sítio A₂ do ITS1, produzindo os pré-rRNAs 20S e 27SA₂. As etapas seguintes de processamento do rRNA 20S ocorrem no citoplasma, através da clivagem no sítio D para gerar o rRNA 18S maduro, enquanto o processamento do pré-rRNA 27SA₂ continua no núcleo. O 27SA₂ é transformado em 5.8S e 25S por dois mecanismos alternativos: cerca de 85% é clivado no sítio A₃ do ITS1, seguido de digestão por exonucleases no sítio B_{1S}, e 15% é clivado diretamente no sítio B_{1L}. A clivagem no sítio B₂ ocorre concomitantemente com a clivagem no sítio B₁. As duas formas do 27SB (27SB_S e 27SB_L) são processadas nos sítios C1 e C2, liberando o rRNA maduro 25S e o pré-rRNA 7S_S ou 7S_L, que após digestão por exonucleases no sítio E gera o rRNA 5.8S_S ou 5.8S_L.

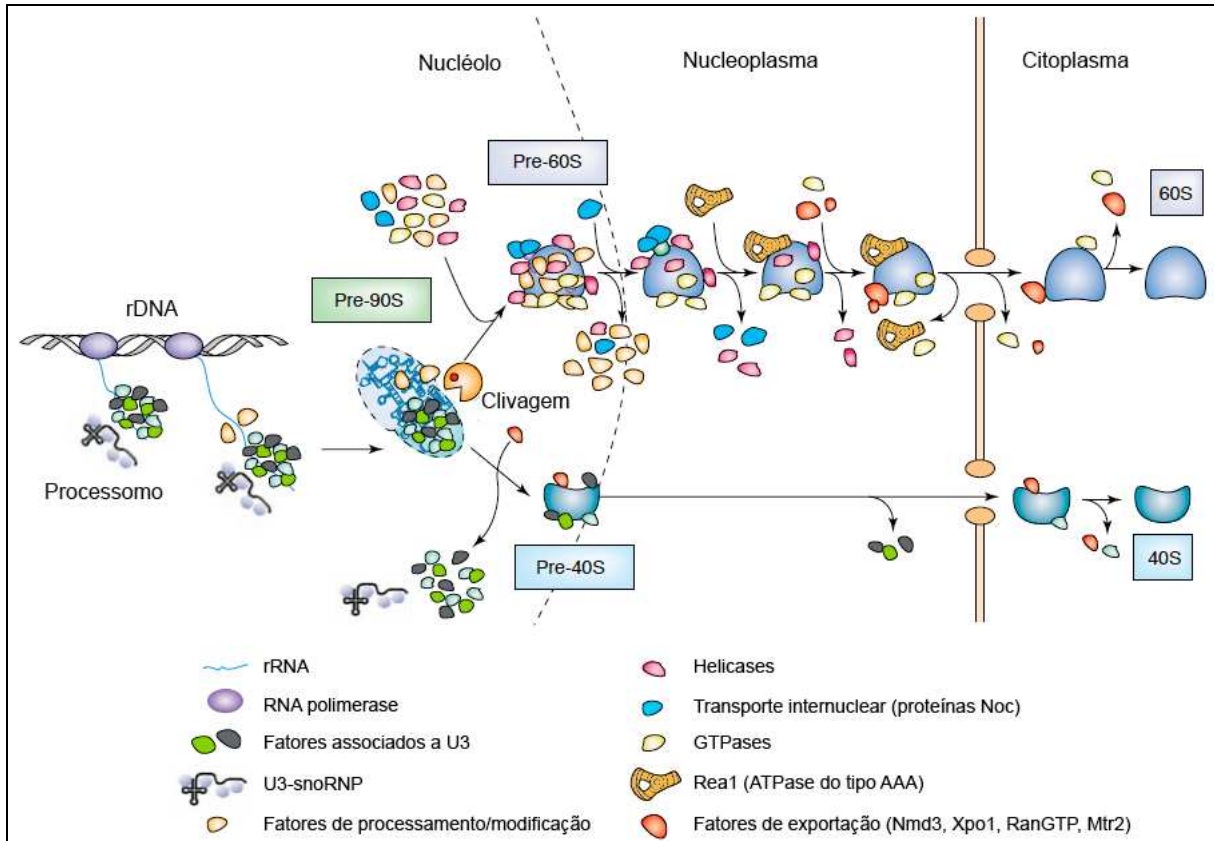


Figura 1.2: Representação de alguns dos fatores envolvidos no processo de maturação e exportação das subunidades ribossomais 40S e 60S (figura adaptada de Tschochner & Hurt, 2003).

A isomerização de uridinas a pseudo-uridinas e a metilação das riboses são as principais modificações de nucleotídeos encontradas nos rRNAs: aproximadamente 100 sítios de cada uma dessas modificações ocorrem nos rRNA humanos, enquanto esse número diminui para 50 em *S. cerevisiae*; nos ribossomos de *E. coli* existem apenas quatro nucleotídeos metilados em ribose e dez pseudouridilações (Maden, 1990; Rozenski *et al.*, 1999). Diferentemente do sistema bacteriano, no qual as modificações são realizadas por enzimas específicas, a metilação e a

pseudouridilação em eucariotos é mediada por pequenos RNAs nucleolares, o snoRNA C/D box e o snoRNA H/ACA box, respectivamente. Esses RNAs são encontrados em pequenas RNPs em complexo com proteínas integrais específicas: Nop1, Nop56, Nop58/Nop5 e Snu13 para o snoRNA C/D box e Cbf5, Gar1, Nhp2 e Nop10 para o snoRNA H/ACA box (Tollervey & Kiss, 1997; Kiss, 2001, Fromont-Racine *et al.*, 2003). Ortólogos do snoRNA C/D box tem sido caracterizados em Archaea (Dennis *et al.*, 2001).

Juntamente com as proteínas ribossomais e os fatores transientes envolvidos na biossíntese dos ribossomos, o rRNA 35S forma um grande complexo ribonucleoprotéico, o 90S. Dentro da partícula pré-ribossômica 90S, o pré-rRNA 35S é rapidamente convertido nos precursores das subunidades ribossomais 40S e 60S. A partícula pré-40S é então processada no citoplasma, enquanto a maturação do pré-60S continua no núcleo antes de sua exportação para o citoplasma (Fromont-Racine *et al.*, 2003).

Os fatores não-ribossomais associados à partícula pré-60S podem ser agrupados em cinco diferentes classes (Figura 1.2): (1) fatores associados com o pré-60S apenas no nucléolo, relacionados com os eventos iniciais de modificação do pré-rRNA (exemplos Cbf5, Ssf1, Ssf2); (2) fatores que se ligam ao pré-60S no nucléolo e o acompanham até os poros nucleares; (3) fatores necessários para o transporte intranuclear (exemplo proteínas Noc); (4) fatores que se ligam ao pré-60S nos estágios finais do nucleoplasma e são liberados antes da exportação para o citoplasma (ex. Rea1); e (5) fatores que realizam o transporte núcleo-citoplasma (ex. Nmd3, Xpo1, Mtr2). Além desses, alguns fatores associados ao pré-60S o acompanham desde o núcleo até o citoplasma, como exemplo Nug1, Nug2, Nog1 e Kre35 (Tschochner & Hurt, 2003).

O processo de maturação da subunidade 40S é mais simples, mas seu mecanismo de exportação para o citoplasma ainda é pouco conhecido. O precursor da partícula 40S contém dois tipos de fatores, aqueles já presentes no pré-ribossomo 90S (exemplos Enp1p, Dim1p, Hrr25p e Rrp12p), e aqueles que são adicionados posteriormente à partícula pré-40S (exemplo Nob1p, Rio2p e Tsr1p) (Tschochner & Hurt, 2003).

A maioria dos genes que codificam fatores envolvidos na biossíntese de ribossomos são essenciais para a célula ou causam consideráveis defeitos no

crescimento celular quando mutados, indicando que eles apresentam um papel fundamental para o crescimento e divisão celular. Problemas na biossíntese dos ribossomos tem sido relacionados com algumas síndromes humanas associadas a disfunção na medula óssea e predisposição a desenvolver certos tipos de câncer. Essas doenças são: disqueratose congênita (DC), hipoplasia cartilagem-cabelo (CHH), anemia Diamond-Blackfan (DBA) e síndrome Shwachman-Diamond (SDS) (Liu & Ellis, 2006).

1.3. A DOENÇA SDS E A PROTEÍNA SBDS

A síndrome Shwachman-Diamond, SDS (OMIM 260400), é uma doença autossômica recessiva rara (1:76000), cujas características clínicas incluem insuficiência pancreática exócrina, disfunção hematológica e anormalidades no esqueleto. SDS é a segunda causa mais comum de insuficiência exócrina pancreática congênita, perdendo apenas para fibrose cística. Disfunção na medula óssea ocorre em praticamente todos os pacientes com SDS, que apresentam considerável predisposição a desenvolver problemas hematológicos, como mielodisplasia ou leucemia mielóide aguda (Shwachman *et al.*, 1964; Bodian *et al.*, 1964; Boocock *et al.*, 2003; Woloszynek *et al.*, 2004; Shammass *et al.*, 2005). SDS representa, portanto, um importante modelo para entender os mecanismos moleculares envolvidos na progressão da leucemia.

O gene Shwachman-Bodian-Diamond, *SBDS*, foi recentemente identificado e localizado no cromossomo 7, próximo ao centrômero. Também foi encontrado um pseudogene adjacente, *SBDSP*, apresentando alto grau de identidade sequencial com *SBDS* (97%), mas contendo deleções e trocas de nucleotídeos que impedem a produção de proteína funcional. A maioria dos pacientes com a síndrome SDS apresentam mutações no gene *SBDS* resultantes de evento de conversão gênica com seu pseudogene (Boocock *et al.*, 2003).

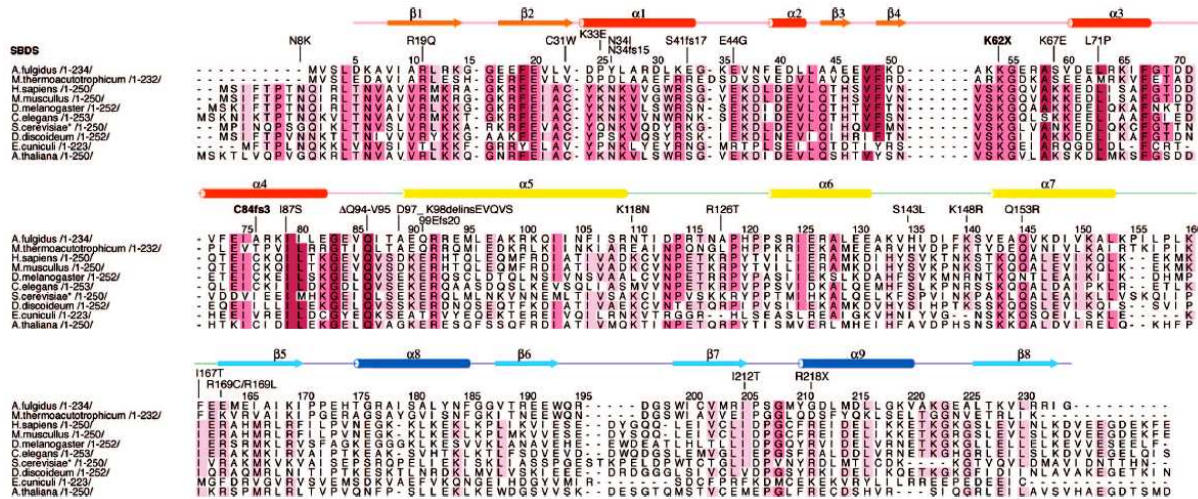
O gene *SBDS* codifica uma proteína de 250 aminoácidos cuja seqüência não apresenta similaridade a nenhum domínio de função conhecida (Boocock *et al.*, 2003). Embora não exista informação adequada para definir a função exata da SBDS, evidências mostram que a síndrome pode ser causada por deficiência em algum aspecto do metabolismo de RNA, que é essencial para o desenvolvimento do

pâncreas exócrino, hematopoiese e formação óssea (Boocock *et al.*, 2003). Análises de perfil transcricional tem mostrado que “knock-down” de SBDS afeta a expressão de importantes genes envolvidos em desenvolvimento cerebral, morfogênese óssea, proliferação e diferenciação de células sanguíneas e adesão celular (Hesling *et al.*, 2007). A manifestação da doença pode refletir perda ou perturbação de uma função celular crítica para o desenvolvimento das células pancreáticas, mielóides e de formação óssea. Os sintomas associados e as complicações do defeito da medula óssea podem resultar não apenas da perda da função da SBDS mas também serem conseqüências pleiotrópicas de um processo aberrante (Boocock *et al.*, 2003).

A proteína SBDS é um membro de uma família de proteínas altamente conservadas com ortólogas em diversos organismos, desde planta, Archaea e levedura até animais vertebrados (Figura 1.3), sugerindo que seu papel bioquímico seja fundamental e conservado. No momento em que este projeto foi iniciado, algumas linhas indiretas de evidência indicavam que SBDS apresentava função no metabolismo de RNA: (1) Estudos de perfil de expressão gênica através de micro-arranjo de DNA agruparam o gene ortólogo de *Saccharomyces cerevisiae*, codificado pela ORF YLR022c, com genes que codificam enzimas do processamento de RNA; (2) Ortólogos de Archaea estão localizados em operons conservados que contêm genes do processamento de RNA, incluindo homólogos do exossomo eucarioto e subunidades do complexo RNaseP; (3) O gene ortólogo de *Arabidopsis thaliana* e seqüências parciais de *Populus balsamifera* e *Gossypium arboreum* predizem um C-terminal adicional com domínio de ligação a RNA (Boocock *et al.*, 2003).

Posteriormente, estudos de análise proteômica realizados com a SBDS de *Saccharomyces cerevisiae*, YLR022c, revelaram a associação desta com mais de 20 proteínas envolvidas na biossíntese de ribossomos, muitas das quais componentes da partícula 60S (Savchenko *et al.*, 2005). Recentes estudos realizados com a ortóloga de levedura, também denominada *Sdo1p*, tem demonstrado que esta proteína é necessária para a liberação e reciclagem de Tif6 da partícula pré-60S, um fator nucleolar cuja liberação do pré-60S é fundamental para a maturação e ativação dos ribossomos (Menne *et al.*, 2007). Ensaio de duplo-híbrido tem associado a proteína *Sdo1p* com os fatores de processamento de rRNA Nip7p, Prp43p e

Spp382p, com componentes das subunidades ribossomais Rpl3p e Asc1p e com o fator de alongação da tradução Eft2p (Luz *et al.*, 2009).



de camundongo apresenta alta expressão durante o rápido crescimento e proliferação celular (Zhang *et al.*, 2006).

Com base nesses estudos funcionais, acredita-se que a proteína SBDS apresenta um importante papel no metabolismo de rRNA e biogênese ribossomal. Deve-se ressaltar ainda que a identificação da associação da proteína SBDS com diversos parceiros, envolvidos em diferentes etapas do processo de maturação dos ribossomos, sugere a ligação da SBDS às subunidades pré-ribossomais ainda no nucléolo, permanecendo esta associada ao complexo durante todo o seu processamento e mesmo após o seu transporte para o citoplasma.

Em 2005, a estrutura tridimensional da ortóloga da SBDS de *Archaeoglobus fulgidus* (AfSBDS) foi resolvida por cristalografia de raios X, a 1,9 Å de resolução (Shammas *et al.*, 2005). Foi verificado que essa proteína contém três domínios arranjados em forma de V, o domínio N-terminal (domínio I), o domínio central (domínio II) e o domínio C-terminal (domínio III), formando uma cavidade central na interface entre os três domínios (Figura 1.4). O domínio I apresenta um enovelamento novo composto de α -hélices e fitas β . Foi identificada similaridade estrutural entre este domínio e uma proteína de levedura de um único domínio, Yhr087wp, também envolvida no metabolismo de RNA. Como consequência dessa similaridade estrutural, o domínio I da AfSBDS foi chamado de domínio FYSH (Eungal, Yhr087wp, Shwachman). O domínio II consiste de um feixe de três hélices e o domínio III contém o enovelamento do tipo ferredoxina, característico de proteínas que ligam RNA.

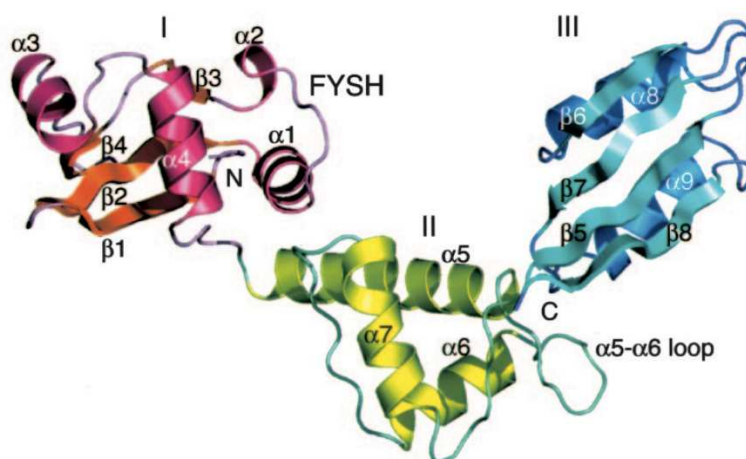


Figura 1.4: Estrutura da proteína SBDS de *Archaeoglobus fulgidus* (AfSBDS). A estrutura, resolvida por cristalografia de raios X, mostra que a proteína contém três domínios globulares (Shammas *et al.*, 2005; Savchenko *et al.*, 2005).

Recentemente, foi determinada a estrutura de mais uma ortóloga de Archaea da SBDS, de *Methanothermobacter thermautotrophicus* (mthSBDS) (Leong *et al.*, 2009). A estrutura também foi resolvida por cristalografia de raios X, a 1,75 Å de resolução, e foram verificadas duas moléculas da proteína na unidade assimétrica, apresentando diferenças conformacionais entre si nas regiões entre os domínios, indicando que a proteína é capaz de realizar ajustes rotacionais dos domínios I e III com relação ao domínio central.

Foram caracterizadas mais de 30 mutações no gene que codifica a proteína SBDS relacionadas com a doença humana SDS, conforme indicado na Tabela 1.1 (Boocock *et al.*, 2003; Nakashima *et al.*, 2004; Shammass *et al.*, 2005, Erdos *et al.*, 2006). As mutações se concentram principalmente no domínio FYSH, indicando que este domínio deve ser importante para a função da proteína (Shammass *et al.*, 2005). Análises de complementação genética realizadas com a ortóloga de levedura têm demonstrado que os domínios FYSH e central apresentam um papel essencial *in vivo*, enquanto o domínio C-terminal é dispensável para a complementação da função da SBDS em levedura (Shammass *et al.*, 2005). Através do mesmo tipo de estudo, foi verificado que o domínio central da proteína SBDS de levedura, diferentemente do domínio FYSH, não pode ser substituído pela sequência de outros ortólogos eucariotos de SBDS, sugerindo que esse domínio confere a especificidade para cada espécie (Boocock *et al.*, 2006).

Com base na estrutura tridimensional da proteína AfSBDS e considerando a conservação das seqüências de aminoácidos das proteínas humana e de Archaea (24% de identidade e 48% de similaridade), Shammass e colaboradores (2005) analisaram a estabilidade de algumas proteínas mutantes da SBDS de *H. sapiens* e *S. cerevisiae* e propuseram a divisão das mutações relacionadas à doença humana em três grupos: (1) mutações que geram proteínas truncadas, resultando na perda da função da proteína (ex. K62X, C84fs3); (2) mutações em resíduos do core hidrofóbico da proteína, que são preditas desestabilizar seu enovelamento (ex. C31W, L71P, I87S); (3) mutações em resíduos da superfície da proteína, que alteram sua superfície eletrostática (ex. R19Q, K33E, N34I, E44G, K67E, K118N, N121T, R126T, S143L, K148R, Q153R, R169C, R169L). Enquanto mutantes do segundo grupo não puderam ser expressos, sugerindo que essas mutações perturbam seu enovelamento global, mutantes do grupo 3 não apresentaram

alterações no enovelamento e estabilidade da proteína, indicando que esse tipo de mutação deve interferir na interação da proteína com outras moléculas, necessária para a realização da sua função.

Tabela 1.1: Mutações no gene *SBDS* relacionadas à doença SDS.

Mutação	Aminoácido mutado	Tipo de mutação	Referência
c.24C>A	N8K	Substituição	Boocock <i>et al.</i> , 2003
c.56G>A	R19Q	Substituição	Shammas <i>et al.</i> , 2005
c.93C>G	C31W	Substituição	Shammas <i>et al.</i> , 2005
c.95A>G	Y32C	Substituição	Nicolis <i>et al.</i> , 2005
c.97A>G	K33E	Substituição	Shammas <i>et al.</i> , 2005
c.96-97insA	N34fsX15	Inserção; frameshift	Boocock <i>et al.</i> , 2003
c.101A>T	N34I	Substituição	Nicolis <i>et al.</i> , 2005
c.119delG	S41fsX17	Deleção; frameshift	Boocock <i>et al.</i> , 2003
c.123delC	S41fsX17	Deleção; frameshift	Shammas <i>et al.</i> , 2005
c.183-184TA>CT	K62X	Substituição; <i>stop codon</i>	Boocock <i>et al.</i> , 2003
c.131A>G	E44G	Substituição	Boocock <i>et al.</i> , 2003
c.199A>G	K67E	Substituição	Boocock <i>et al.</i> , 2003
c.212T>C	L71P	Substituição	Shammas <i>et al.</i> , 2005
c.258+2T>C	C84fsX3	Substituição; frameshift	Boocock <i>et al.</i> , 2003
c.258+374-459+250del	I87-Q153del	Deleção	Costa <i>et al.</i> , 2007
c.260T>G	I87S	Substituição	Boocock <i>et al.</i> , 2003
c.279-284delTCAAGT	Q94-V95del	Deleção	Shammas <i>et al.</i> , 2005
c.291-293TAA del insAGTTCAAGTATC	D97-K98 del insEVQVS	Deleção; inserção	Boocock <i>et al.</i> , 2003
c.292-295delAAAG	E99fsX20	Deleção; frameshift	Nakashima <i>et al.</i> , 2004
c.307-308delCA	Q103fsX8	Deleção; frameshift	Nicolis <i>et al.</i> , 2005
c.354A>C	K118N	Substituição	Shammas <i>et al.</i> , 2005
c.362A>C	N121T	Substituição	Erdos <i>et al.</i> , 2006
c.377G>C	R126T	Substituição	Boocock <i>et al.</i> , 2003
c.428C>G	S143W	Substituição	Taneichi <i>et al.</i> , 2006
c.428C>T+443A>G	S143L+K148R	Substituição	Shammas <i>et al.</i> , 2005
c.458A>G	Q153R	Substituição	Shammas <i>et al.</i> , 2005
c.505C>T	R169C	Substituição	Boocock <i>et al.</i> , 2003
c.506G>T	R169L	Substituição	Shammas <i>et al.</i> , 2005
c.523C>T	R175W	Substituição	Erdos <i>et al.</i> , 2006
c.635T>C	I212T	Substituição	Boocock <i>et al.</i> , 2003
c.652C>T	R218X	Substituição; <i>stop codon</i>	Nicolis <i>et al.</i> , 2005

O alinhamento da sequência de aminoácidos da SBDS de 159 espécies permitiu a identificação de um C-terminal estendido nas ortólogas de plantas e protistas, representando um quarto domínio adicional nessas proteínas (Boocock *et al.*, 2006). Na região adicional de plantas, é predito conter um motivo de ligação a RNA do tipo C2H2 “zinc finger” (Boocock *et al.*, 2003). Já no domínio estendido da SBDS de protistas, que varia de 199 a 269 resíduos, rico em glicinas e aminoácidos carregados, não se identifica a presença de nenhum motivo estrutural conhecido. Ensaio de complementação utilizando-se uma quimera da SBDS de *S. cerevisiae* contendo o C-terminal estendido da ortóloga de *Arabidopsis thaliana* mostraram que a região adicional de plantas é compatível com a função da proteína de levedura (Boocock *et al.*, 2006).

Embora os dados demonstrem que a proteína SBDS esteja relacionada à biogênese ribossomal, o papel específico desta proteína nesse processo e seus parceiros de interação ainda permanecem desconhecidos. Muitos fatores que atuam de forma transiente na biossíntese de ribossomos tem sido identificados e fatores adicionais ainda devem ser descobertos. SBDS deve ser um desses fatores, essencial em Archaea e eucariotos e dispensável para bactérias (Savchenko *et al.*, 2005).

1.4. OBJETIVOS

Este trabalho buscou a caracterização estrutural e funcional de proteínas da família SBDS, visando contribuir para o entendimento do mecanismo molecular envolvido no seu funcionamento e analisar as possíveis interações dessa proteína com outras moléculas. O projeto teve como objetivo inicial a aplicação da técnica de cristalografia de raios X para o estudo estrutural das proteínas SBDS de *Pyrococcus abyssi* (PaUPF0023), *Saccharomyces cerevisiae* (YLR022c), *Trypanosoma cruzi* (TcSBDS) e *Homo sapiens* (HsSBDS), mas, devido à dificuldade na obtenção de cristais das proteínas de interesse, a técnica de Ressonância Magnética Nuclear (RMN) foi empregada para a resolução estrutural da proteína SBDS humana.

Neste contexto, os objetivos específicos do projeto consistiram de:

- Identificação da sequência de aminoácidos da proteína TcSBDS.

- Clonagem, expressão, purificação e ensaios de cristalização das ortólogas PaUPF0023, YLR022c, TcSBDS e HsSBDS.
- Análises de dicroísmo circular e estabilidade térmica das proteínas PaUPF0023, YLR022c e HsSBDS.
- Clonagem, expressão, purificação e ensaios de cristalização de domínios da HsSBDS.
- Clonagem, expressão e purificação da região C-terminal da TcSBDS.
- Estudo estrutural do C-terminal da TcSBDS por dicroísmo circular, proteólise limitada e RMN.
- Estudos de interação da TcSBDS com RNA.
- Estudos funcionais com a TcSBDS, de associação aos polissomos e rRNA.
- Determinação da estrutura tridimensional da SBDS humana por RMN.
- Estudos de dinâmica da proteína SBDS humana por RMN.
- Análise da interação da SBDS humana com RNA por RMN.

CAPÍTULO 2

ESTUDOS ESTRUTURAIS DE PROTEÍNAS DA FAMÍLIA SBDS

Neste capítulo encontram-se descritos a expressão, a purificação e os ensaios de cristalização realizados com as proteínas ortólogas da SBDS de *Pyrococcus abyssi* (PaUPF0023), *Saccharomyces cerevisiae* (YLR022C), *Trypanosoma cruzi* (TcSBDS) e *Homo sapiens* (HsSBDS), na tentativa de resolução estrutural de membros dessa família de proteínas por cristalografia de raios X. Embora o objetivo de resolução da estrutura das proteínas acima mencionadas por cristalografia de raios X não tenha sido alcançado, grande esforço foi investido nesse processo e considera-se importante o registro desses dados para consultas futuras. O processo de determinação estrutural da proteína SBDS humana por Ressonância Magnética Nuclear também encontra-se descrito nesse capítulo, com os resultados finais mostrados no Capítulo 3.

2.1. MATERIAIS E MÉTODOS

2.1.1. Construção dos vetores de expressão

Os vetores de expressão pET28a/PaUPF0023 e pET-TEV/HsSBDS foram fornecidos, respectivamente, pela Dra. Carla Oliveira do Departamento de Bioquímica da USP, São Paulo, e pelo Dr. Nilson Zanchin do Laboratório Nacional de Luz Síncrotron, co-orientador deste projeto. Os cDNAs da YLR022c e da TcSBDS truncada (resíduos 1 a 263) foram amplificados por PCR a partir do DNA genômico de levedura e de *T. cruzi*, respectivamente, utilizando-se os oligonucleotídeos listados na Tabela 2.1. As sequências de nucleotídeos sublinhadas correspondem aos sítios de restrição utilizados. Os produtos de PCR foram inseridos no vetor de expressão pET-TEV (Carneiro *et al.*, 2006). Os números de acesso (NCBI) das proteínas ortólogas PaUPF0023, YLR022c, TcSBDS e HsSBDS são, respectivamente, NP 126299, NP 013122, EU715774 e NP 057122. Também foram feitas clonagens de domínios da proteína SBDS humana separadamente, conforme indicado na Tabela 2.2. O gene correspondente a cada domínio foi amplificado por PCR a partir de DNA genômico humano, utilizando-se os oligonucleotídeos descritos na Tabela 2.1, e inseridos no vetor pET-TEV.

Tabela 2.1: Oligonucleotídeos utilizados para produção dos clones de expressão.

Oligonucleotídeo	Seqüência	Sítio de Restrição
5' YLR022c	5' GATCC <u>C</u> ATATGCCTATCAATCAACCGTCGGG 3'	NdeI
3' YLR022c	5' GTGT <u>C</u> TCGAGGATATGTATG 3'	XhoI
5' TcSBDS	5' ATGCC <u>C</u> ATATGAGTGCCCGTATTCAGGTTCC 3'	NdeI
3' TcSBDS	5' <u>G</u> GATCCTCAAACATCACCTTCGCCCACATC 3'	BamHI
5' Domínio 1 HsSBDS	5' GCCC <u>C</u> ATATGTCGATCTTCACCCCCACCAAC 3'	NdeI
3' Domínio 1 HsSBDS	5' TTGGGATC <u>C</u> TCATTTAGTCAAAATCTGCTT 3'	BamHI
5' Domínio 2 HsSBDS	5' GTTCATATGTCAGATAAAGAAAGACACACA 3'	NdeI
3' Domínio 2 HsSBDS	5' CATGGATCCACGTCATATCTTCATTTTCTC 3'	BamHI
5' Domínio 3 HsSBDS	5' ATGCATATGGAACGTGCTCACATGAGGCTT 3'	NdeI
3' Domínio 3 HsSBDS	5' TTCGGATCCCTCTCATCCTTCTTCTACATC 3'	BamHI
5' Domínio 1+2 HsSBDS	5' GCCC <u>C</u> ATATGTCGATCTTCACCCCCACCAAC 3'	NdeI
3' Domínio 1+2 HsSBDS	5' CATGGATCCACGTCATATCTTCATTTTCTC 3'	BamHI
5' Domínio 2+3 HsSBDS	5' GTTCATATGTCAGATAAAGAAAGACACACA 3'	NdeI
3' Domínio 2+3 HsSBDS	5' GATGGATCCGTGTCATTCAAATTTCTCATC 3'	BamHI

Tabela 2.2: Região correspondente às diferentes construções da HsSBDS.

Domínio	Seqüência de aminoácidos
1	1 a 90
2	96 a 167
3	168 a 244
1 + 2	1 a 167
2 + 3	96 a 250

A confirmação da construção correta dos vetores de expressão contendo os genes de interesse foi realizada através de digestão com as respectivas enzimas de restrição e por seqüenciamento de DNA utilizando-se o sistema ABI PRISM 377 Sequence Analyser (Applied Biosystems), de acordo com as instruções do fabricante.

2.1.2. Expressão das proteínas

Para a expressão das proteínas de interesse foi utilizada a cepa de *E. coli* BL21(DE3)*SlyD*⁻ (Yan *et al.* 2001). As células transformadas com os respectivos vetores de expressão foram inoculadas em meio LB contendo canamicina (50 µg/ml). As condições de indução foram otimizadas para cada caso. As células foram crescidas a 37°C até OD₆₀₀ entre 0,8 e 1,0, quando foi feita a indução com 0,5 mM de IPTG por 4 horas a 37°C para a PaUPF0023 e a 25°C para a YLR022c. A indução da proteína TcSBDS truncada foi realizada a 16°C por 16 horas. No caso da HsSBDS, a indução foi feita com 10 mM de lactose por 16 horas a 25°C. Os domínios 1+2 e 2+3 da HsSBDS foram induzidos a 18°C com 0,5 mM IPTG durante 16 horas. Após indução, as células foram coletadas por centrifugação (6000 rpm, 10 min a 4°C) e ressuspendidas em tampão A contendo 50 mM Tris-HCl pH 7,2 e 500 mM NaCl (para a HsSBDS foi adicionado também 1 mM β-mercaptoetanol). Após adição de PMSF (1 mM), o extrato foi tratado com lisozima (50 µg/ml) por 30 minutos, sonificado e o sobrenadante foi separado da fração insolúvel através de centrifugação (13.000 rpm, 30 min, a 4°C). A fração solúvel foi precipitada com sulfato de estreptomicina (2 % m/v) por 1 hora a 4°C, com agitação, centrifugada novamente e então o sobrenadante foi submetido a métodos cromatográficos para purificação das proteínas de interesse.

2.1.3. Purificação das proteínas e remoção da cauda de histidina

- **PaUPF0023:** a proteína SBDS de *P. abyssi* foi purificada através de cromatografia de afinidade a níquel (Ni-NTA Qiagen). O extrato foi incubado com a resina por 1 hora a 4°C e transferido para a coluna BioSpin (Bio-Rad). Após lavagem da resina com tampão A (mesmo da preparação do extrato), foram realizadas eluições com o mesmo tampão contendo concentrações crescentes de imidazol (20, 50, 100, 200 e 500 mM). A cauda de histidina presente no N-terminal da proteína PaUPF0023 foi clivada através de proteólise parcial com tripsina. A clivagem foi feita à temperatura ambiente por 30 minutos, com agitação, utilizando-se 0,2% m/m de protease e a reação foi interrompida através da adição de 1 mM de PMSF. Então, uma segunda etapa de purificação envolveu cromatografia de troca iônica em coluna Mono-S HR 5/5 de 1 mL (Amersham Biosciences), utilizando-se o sistema ÄKTA-FPLC (Amersham Biosciences). O tampão A utilizado para essa purificação continha 10 mM acetato de sódio pH 5,2 e a eluição foi realizada com tampão B contendo 10 mM acetato de sódio pH 5,2 e 1 M NaCl.

- **YLR022c:** a ortóloga de levedura foi purificada através de cromatografia de afinidade a níquel (Ni-NTA Qiagen), seguindo o mesmo procedimento utilizado para a PaUPF0023. A cauda de histidina da YLR022c foi digerida com a protease TEV, a 25°C por 8 horas, com TEV a 2% m/m, no seguinte tampão: 50 mM Tris-HCl pH 8,0; 0,05 mM EDTA; 1 mM DTT e 50 mM NaCl. Uma segunda etapa de purificação foi realizada através de cromatografia de troca iônica em coluna HiTrap SP HP de 1 mL (Amersham Biosciences), utilizando-se o sistema ÄKTA-FPLC, com tampão A contendo 10 mM acetato de sódio pH 5,2; 10 mM NaCl e 1 mM DTT e tampão B igual, com a adição de 1 M NaCl.

- **TcSBDS truncada:** a proteína TcSBDS truncada foi purificada através de cromatografia de afinidade utilizando-se uma coluna HiTrap Chelating HP de 1 mL e o sistema ÄKTA-FPLC, com tampão A igual ao de preparação do extrato. O tampão B consistiu do tampão A na presença de 300 mM imidazol. Não foi feita a clivagem da cauda de histidina desta proteína. A amostra proveniente da afinidade foi dialisada contra tampão contendo 25 mM acetato de sódio pH 5,2.

- **HsSBDS:** foi realizada a purificação da proteína humana através de cromatografia de afinidade utilizando-se uma coluna HiTrap Chelating HP de 5 mL, com tampão A contendo 50 mM Tris-HCl pH 7,2; 500 mM NaCl e 10 mM β -mercaptoetanol. O tampão B consistiu do tampão A na presença de 300 mM imidazol. A clivagem da cauda de histidina foi realizada com a protease TEV, a 25°C por 24 horas, no seguinte tampão: 50 mM Tris-HCl pH 8,0; 0,5 mM EDTA; 50 mM NaCl e 1 mM DTT. Foi utilizada a TEV a 15% (m/m) em relação à proteína. Uma nova etapa de purificação, utilizando-se a coluna HiTrap Heparina HP de 5 mL (Amersham Biosciences), foi realizada, com tampão A contendo 50 mM Tris-HCl pH 7,2; 20 mM NaCl e 1 mM DTT e tampão B com 50 mM Tris-HCl pH 7,2; 1 M NaCl; 1 mM DTT.

- **HsSBDS Domínio 1+2:** foi realizada a purificação do domínio 1+2 da HsSBDS através de coluna de afinidade HiTrap Chelating HP de 1 mL, utilizando-se o sistema ÄKTA-FPLC. O tampão A utilizado continha 50 mM fosfato de sódio pH 7,2; 50 mM NaCl e 0,5 mM PMSF e o tampão B consistiu do Tampão A acrescido de 1 M imidazol. As frações eluídas da coluna de afinidade foram reunidas e dialisadas contra tampão contendo 50 mM Tris HCl pH 7,2; 50 mM NaCl e 0,5 mM PMSF, sem a clivagem da cauda de histidina.

- **HsSBDS Domínio 2+3:** o domínio 2+3 da HsSBDS foi purificado através de coluna de afinidade HiTrap Chelating HP de 1 mL, com os mesmos tampões utilizados para o domínio 1+2, acrescidos de 5% glicerol. Uma segunda etapa de purificação foi realizada, utilizando-se a coluna HiTrap Heparina HP de 5 mL, com tampão A contendo 50 mM Tris-HCl pH 7,2 e 50 mM NaCl e o tampão B acrescido de 1 M NaCl. A cauda de histidina desse domínio não foi clivada.

2.1.4. Cromatografia de exclusão molecular analítica

Ensaio de cromatografia de exclusão molecular foram realizados com a proteína HsSBDS, utilizando-se a coluna Superdex 200 10/30 (Amersham Biosciences) e o sistema ÄKTA-FPLC. Foram utilizados como padrões de massa molecular as seguintes proteínas: Ribonuclease A (13,7 kDa; 3 mg/mL), Ovalbumina (43 kDa; 4 mg/mL), Conalbumina (75 kDa; 3 mg/mL) e Ferritina (440 kDa; 0,3 mg/mL). A diluição dos padrões moleculares e as corridas cromatográficas foram

feitas com tampão contendo 50 mM fosfato de sódio pH 7,2; 20 mM NaCl e 1 mM DTT.

2.1.5. Proteólise limitada

Experimentos de proteólise limitada foram realizados com a proteína HsSBDS utilizando-se as proteases tripsina, quimiotripsina, papaína, proteinase K, subtilisina e termolisina. Os experimentos foram realizados à temperatura ambiente em tampão contendo 50 mM Tris-HCl pH 7,2; 20 mM NaCl e 1 mM DTT. Foi utilizado 1% m/m de protease com relação à proteína e a reação foi interrompida através da adição de 1 mM de PMSF. A concentração de proteína utilizada foi de 1 mg/mL e foram tiradas alíquotas após 0, 15, 30, 60 e 120 minutos de clivagem. A amostra digerida com tripsina por 2 horas foi submetida a análise por Espectrometria de Massas com ionização por *Electrospray* (ESI-MS).

2.1.6. Ensaio para análise de agregação

A proteína TcSBDS truncada foi submetida a ensaios para detecção e prevenção de agregação, conforme descrito por Bondos e Bicknell, 2003. Foram preparadas alíquotas de 300 μ L da proteína (0,2 mg/mL) em tampão contendo 50 mM Tris-HCl pH 8,5; 20 mM NaCl e 10 mM β -mercaptoetanol e a estas alíquotas foram adicionados os aditivos descritos na Tabela 2.3, nas concentrações indicadas. Após incubação em gelo por 1 hora e 30 minutos, a proteína solúvel foi separada da fração agregada utilizando-se um concentrador Centricon (Millipore) de 100 kDa, com centrifugação de 1000 g por 15 minutos a 4°C. Os agregados retidos na membrana foram ressuspensos em 50 μ L de tampão contendo 20 mM Tris-HCl pH 8 e 8M uréia. As frações solúvel e agregada foram aplicadas em gel SDS-PAGE 12,5% para análise (20 μ L cada).

Tabela 2.3: Aditivos utilizados na tentativa de diminuir a agregação da proteína TcSBDS truncada.

0,2 M MgSO ₄	0,5 M Glicose	0,75 M Uréia	5 mM DTT
0,5 M KCl	0,5% Nonidet P-40	1% Glicina	0,1 M CaCl ₂

2.1.7. Análises de Dicroísmo Circular (CD)

Os experimentos de CD foram realizados em um espectropolarímetro J-810 (JASCO), acoplado a um sistema interno de controle de temperatura (Peltier Type Control System PFD 425S – JASCO). As cubetas de quartzo utilizadas possuíam caminho ótico de 0,1 cm e as medidas foram realizadas em temperatura constante de 20°C. Os programas utilizados para registro e tratamento dos dados foram o Spectra Manager® (JASCO) e o Origin® (Microcal), respectivamente. Espectros cumulativos de 4 coletas (velocidade de coleta = 100 nm/min e tempo de resposta do detector de 4 segundos) foram normalizados em elipticidade molar residual (EMR ou $[\theta]$), após subtração do espectro do respectivo tampão (acumulado por 4 vezes), pela aplicação da seguinte equação:

$$\text{EMR} = (\theta \times \text{MM}) / (10 \times c \times l \times n)$$

onde θ é a elipticidade (em mDeg), MM é a massa molar (em g/mol), c é a concentração da proteína (em mg/mL), l é o caminho ótico (cm) e n é o número de resíduos de aminoácidos da proteína. A porcentagem de estrutura secundária foi calculada com o auxílio do programa “CDNN Deconvolution”. As medidas foram feitas com as proteínas a aproximadamente 10 μM , nos tampões descritos na Tabela 2.4.

Tabela 2.4: Tampões e concentrações utilizados nas medidas de CD.

Proteína	Tampão	Concentração
PaUPF0023	10 mM acetato de sódio pH 5,2	7,2 μM
YLR022c	10 mM acetato de sódio pH 5,2; 5% glicerol; 1 mM DTT	9,5 μM
HsSBDS	50 mM Tris-HCl pH 7,2; 5 mM NaCl; 5% glicerol; 1 mM DTT	9,3 μM

Também foram realizados experimentos de desnaturação térmica. O aumento de temperatura (1°C/min) foi produzido pelo sistema interno de controle de temperatura descrito acima e espectros de CD foram coletados em intervalos de temperatura de 5°C. As curvas de desnaturação térmica foram obtidas por monitoramento do valor da elipticidade a 222 nm durante o aquecimento de 15 a 95°C, em intervalos de temperatura de 0,5°C. A temperatura média de transição, T_m , foi estimada a partir do ajuste da primeira derivada da curva de desnaturação por uma gaussiana através do programa Origin® (Microcal). As medidas de

desnaturação térmica foram realizadas com as proteínas nas condições indicadas na Tabela 2.4.

2.1.8. Ensaios de cristalização

Para os ensaios de cristalização, as proteínas de interesse foram concentradas por centrifugação, utilizando-se o concentrador Amicon Ultra (Millipore), em diferentes tampões, conforme indicado na Tabela 2.5. Os ensaios de cristalização foram realizados a 18°C, através do método de difusão de vapor, utilizando-se os seguintes kits de cristalização: Crystal Screen (Hampton Research), Wizard (Emerald BioSystems), JBScreen (Jena Biosciences), Index (Hampton Research), SaltRX (Hampton Research), PACT (Nextal/Qiagen), JCSG (Nextal/Qiagen) e Precipitant Sinergy (Emerald BioSystems).

Tabela 2.5: Descrição dos ensaios de cristalização realizados com as diferentes proteínas.

Proteína	Tampão de cristalização	Concentração da proteína	Kits testados
PaUPF0023	10 mM acetato de sódio pH 5,2	10 mg/mL	Crystal Screen, Wizard, JBScreen
YLR022c	50 mM Tris-HCl pH 7,2; 10 mM NaCl; 1 mM DTT	12 mg/mL	Crystal Screen, Wizard, JBScreen, Index, SaltRX
TcSBDS truncada	50 mM acetato de sódio pH 5,2; 20 mM NaCl; 10 mM β -mercaptoetanol	8 mg/mL	Crystal Screen, Wizard, PACT, JCSG, SaltRX e Precipitant Sinergy
TcSBDS truncada	50 mM acetato de sódio pH 5,2; 200 mM NaCl; 10 mM β -mercaptoetanol	6 mg/mL	Crystal Screen, Wizard
HsSBDS	50 mM Tris-HCl pH 7,2; 20 mM NaCl; 1 mM DTT	12 mg/mL	Crystal Screen, Wizard, JBScreen, Index, SaltRX, PACT, JCSG e Precipitant Sinergy
His ₆ SBDS*	50 mM Tris-HCl pH 7,2; 20 mM NaCl e 1 mM DTT	11 mg/mL	Crystal Screen, Wizard, PACT, JCSG, SaltRX e Precipitant Sinergy
HsSBDS Domínio 1+2	50 mM tris-HCl pH 7,2; 50 mM NaCl; 0,5 mM PMSF	5 mg/mL	Crystal Screen, Wizard
HsSBDS Domínio 2+3	50 mM tris-HCl pH 7,2 e 50 mM NaCl	10 mg/mL	Crystal Screen, Wizard, JCSG, SaltRX e Precipitant Sinergy

* HsSBDS contendo cauda de histidina.

Os ensaios de cristalização das proteínas PaUPF0023, YLR022c, HsSBDS e do domínio 1+2 da HsSBDS foram realizados manualmente, utilizando-se 1,5 μ L de amostra de proteína e a mesma quantidade de solução do poço nas gotas de cristalização (do tipo gota pendurada ou *hanging drop*) e o reservatório contendo 300 μ L de solução. Com as proteínas TcSBDS truncada, His₆SBDS e o domínio 2+3 da HsSBDS, os ensaios de cristalização foram feitos automaticamente, utilizando-se o robô Honeybee 961 (Genomic Solutions), com gotas do tipo sentada (*sitting drop*), contendo 0,5 μ L de amostra de proteína e a mesma quantidade de solução do poço e o reservatório com 80 μ L de solução.

2.1.9. Experimentos de Ressonância Magnética Nuclear (RMN)

Amostras da proteína HsSBDS para estudos de RMN

Para utilização da metodologia de resolução de estruturas por espectroscopia de RMN para proteínas com mais de 100 aminoácidos faz-se necessária a utilização de marcação com isótopos estáveis de carbono (isótopo ¹³C) e nitrogênio (¹⁵N), além de deutério (²H) no caso de proteínas maiores de 20 kDa. A presença destes isótopos incorporados à cadeia polipeptídica permite a realização de uma série de experimentos de correlação entre ¹H, ¹³C e ¹⁵N, denominados experimentos de ressonância tripla, desenvolvidos inicialmente pelo grupo do Dr. Ad Bax (NIH) nos anos 90 e descritos com detalhe por Sattler e colaboradores (1999).

As amostras da proteína HsSBDS foram preparadas em meio mínimo M9 (Li *et al.*, 2002), variando-se a composição dos isótopos estáveis presentes no meio de cultura, pela utilização de cloreto de amônio enriquecido em ¹⁵N e/ou glicose enriquecida com ¹³C (Tabela 2.6). Para viabilizar o assinalamento, também foi feita a marcação da proteína com deutério, após condicionamento do crescimento das bactérias em meio contendo D₂O (Li *et al.*, 2002). As induções em meio mínimo foram realizadas com 0,5 mM IPTG por 4 horas a 37°C e a preparação do extrato seguiu conforme descrito no item 2.1.2. A purificação da proteína marcada isotopicamente foi feita igual ao método descrito para a proteína HsSBDS sem marcação isotópica (item 2.1.3). Os experimentos de RMN foram realizados com a proteína concentrada a 0,35 mM em tampão contendo 50 mM fosfato de sódio pH 7,2; 20 mM NaCl e 1 mM DTT.

Tabela 2.6: Descrição das amostras da HsSBDS utilizadas para os experimentos de RMN.

Marcação isotópica	Rendimento da proteína por litro de cultura (mg)	Número de amostras utilizadas*	
		Determinação estrutural	Ensaio de interação
^{15}N	13	1	8
^{15}N , ^{13}C	11	3	-
^{15}N , ^{13}C , ^2H	6	1	-

*Os experimentos foram feitos com a proteína a aproximadamente 10 mg/mL.

Determinação da estrutura da HsSBDS por RMN

Os experimentos de ressonância tripla fazem parte da rotina de experimentos disponíveis ao operador do equipamento e várias combinações de experimentos são necessárias para a resolução da estrutura de uma proteína. Experimentos adicionais de relaxação são realizados para a investigação do comportamento dinâmico da molécula, isto é, seus parâmetros de movimento e mudanças conformacionais.

Para a determinação de uma estrutura por RMN, inicialmente são coletados experimentos de correlação entre hidrogênios e nitrogênios, do tipo HSQC (Heteronuclear Single Quantum Coherence) (Bodenhausen & Ruben, 1980). Neste espectro tem-se um mapa em duas dimensões com picos que correspondem aos pares NH das amidas da cadeia principal da proteína (e algumas cadeias laterais específicas). Como exemplo podem ser citados os espectros de ^{15}N -HSQC de duas proteínas deste trabalho: da HsSBDS (Capítulo 3, Artigo 2, Figura Suplementar 1), característico de proteínas bem enoveladas, em que os picos apresentam grande dispersão em ambas as dimensões; e do domínio C-terminal adicional da TcSBDS (Capítulo 3, Artigo 1, Figura 4), com pequena dispersão na dimensão do hidrogênio, característico de proteínas desenoveladas.

A maioria dos experimentos subsequentes é derivada do experimento de HSQC, com a adição de outras correlações, com os carbonos alfa e seus hidrogênios, com os carbonos das carbonilas, carbonos beta e outros átomos das cadeias laterais. Estes experimentos de múltiplas dimensões fornecem informações que permitem o assinalamento de grande parte dos sinais experimentais aos correspondentes átomos da proteína. Desses experimentos obtêm-se a confirmação da sequência de aminoácidos e sua disposição em elementos de estrutura secundária, pela análise de pares de experimentos do tipo “*out-and-back*” (Sattler, 1999).

Outro importante experimento para a determinação estrutural por RMN é o NOESY (Nuclear Overhauser Effect Spectroscopy) e suas derivações (Wuthrich, 1986). Neste experimento, os sinais fornecem informação sobre a proximidade espacial entre núcleos, para todos os hidrogênios a menos de 5 Å entre si. Estes espectros são bastante complexos, mas assim como os experimentos de ressonância tripla, se prestam a análises semi-automáticas. No entanto, a intervenção do operador é imprescindível para o sucesso do estudo estrutural.

Os espectros de RMN da proteína HsSBDS foram adquiridos a 20°C no espectrômetro Varian Inova de 600 MHz do Laboratório Nacional de Luz Síncrotron (LNLS), com campo magnético de 14,1 T, operando na frequência de 599,883 MHz para ^1H ; 60,786 MHz para ^{15}N ; e 150,866 MHz para ^{13}C ; com sonda criogênica de ressonância tripla. Para resolução estrutural dessa proteína foram realizados os seguintes experimentos de RMN:

- ^{15}N -HSQC, coletado com janela de 8000 Hz, com 2048 pontos para ^1H e 2430 Hz e 128 incrementos para ^{15}N , utilizando-se a técnica TROSY (Pervushin *et al.*, 1997).

- HNCA, HN(CO)CA, HNCOC, HN(CA)CO, CBCA(CO)NH e HNCACB, HBHA(CBCACO)NH, para assinalamento das ressonâncias da cadeia principal da proteína. Adicionalmente, experimentos HNCACB e HN(CO)CACB, com a amostra contendo $^{15}\text{N}/^{13}\text{C}/^2\text{H}$, foram coletados no espectrômetro de 800 MHz da Universidade de East Anglia, em Norwich, Inglaterra, por nossa colaboradora Dra. Tharin Blumenschein. Todos os experimentos utilizaram a técnica TROSY.

- ^{13}C -HSQC, HCCH-TOCSY e HCCH-COSY, para atribuição das ressonâncias das cadeias laterais. Previamente a estes experimentos, a amostra foi liofilizada e ressuspensa em D_2O .

- Experimentos de troca hidrogênio-deutério (H/D) para estudos de acessibilidade ao solvente. Novamente, a amostra da HsSBDS foi liofilizada e ressuspensa em D_2O , e espectros de ^{15}N -HSQC foram adquiridos após 2h30m; 5h; 7h30m; 10h; 12h30m; 15h; 17h30m; 60h; 95h; 158h; 220h.

- ^{15}N -NOESY-HSQC e ^{13}C -NOESY-HSQC, utilizando-se um tempo de mistura de 80 ms, para atribuição dos sinais de NOE.

Todos os dados foram processados utilizando-se o aplicativo NMRPipe (Delaglio *et al.*, 1995) e os espectros foram visualizados através do programa NMRView5 (Johnson & Blevins, 1994).

As estruturas foram inicialmente calculadas utilizando-se o programa DYANA (Güntert *et al.*, 1997, versão 1.5), a partir das restrições de distâncias obtidas dos NOEs assinalados do N-terminal da proteína. A cada ciclo de cálculo eram geradas 20 estruturas (com 8000 steps) e as 10 melhores eram analisadas quanto a violações de distância. Posteriormente, utilizou-se o programa CYANA (Güntert, 2004, versão 2.1) para atribuição automatizada dos NOEs da proteína inteira, a partir da lista inicial atribuída e da estrutura da homóloga de *A. fulgidus* (PDB 1T95). Nesta etapa, foram utilizadas 100 estruturas (com 10000 steps), considerando-se os dados das 20 melhores. Com esta lista adicional de NOEs obtida automaticamente, continuou-se a atribuição manual dos NOEs e o refinamento da estrutura utilizando-se o CYANA, no qual a cada ciclo eram geradas 40 estruturas. As estruturas obtidas eram visualizadas através do programa PyMOL (versão 0.98, www.pymol.org). A minimização final da energia das estruturas obtidas foi feita com o auxílio do programa CNS (Brunger *et al.*, 1998).

Caracterização dinâmica da HsSBDS por RMN

A mobilidade atômica de uma proteína pode ser estimada a partir dos tempos de relaxação dos sinais de RMN no espectro HSQC. As propriedades dinâmicas de alguns átomos, no caso a dupla NH, são estimadas através da modelagem dos dados experimentais, que refletem a dispersão de energia na amostra. Utilizando-se o formalismo de densidade espectral (Farrow *et al.*, 1995), perturbações podem ser descritas a nível nuclear por campos magnéticos flutuantes. Assim, diferenças de mobilidade podem ser estimadas ao nível de cada aminoácido, na faixa de pico a mili-segundos. A metodologia utilizada para as medidas de relaxação da proteína HsSBDS encontram-se descritas no Capítulo 3, Artigo 2, Sessão Materias e Métodos.

2.2. RESULTADOS

2.2.1. SBDS de *Pyrococcus abyssi*: PaUPF0023

A expressão da proteína SBDS de *P. abyssi* foi realizada em *E. coli* utilizando-se o vetor pET28a/PaUPF0023. Essa ortóloga da SBDS, em fusão com polihistidina, foi abundantemente expressa na sua forma solúvel e inicialmente purificada por cromatografia de afinidade. Após proteólise parcial com tripsina para remoção da cauda de histidina, a amostra foi submetida a uma segunda etapa de purificação, utilizando-se cromatografia de troca iônica. Análises de dicroísmo circular (CD) foram realizadas com a proteína para a caracterização da sua estrutura secundária (Figura 2.1A). O espectro desta proteína à temperatura ambiente apresentou-se semelhante ao espectro de proteínas ricas em α -hélices (Rodger e Norden, 1997), com picos negativos em 208 e 222 nm. A predição dos tipos de estruturas secundárias presentes nesta proteína, calculada com o auxílio do programa “CDNN Deconvolution”, indica realmente a predominância de α -hélice (43%), mas também presume a existência de grande quantidade de regiões desestruturadas (“random coils”, 27%).

Experimentos de desnaturação térmica utilizando-se dicroísmo circular foram realizados para investigar a estabilidade da PaUPF0023. Para isso, espectros de CD da proteína foram coletados durante o aumento progressivo da temperatura (Figura 2.1A). Foram verificadas apenas pequenas alterações no espectro original da proteína (à temperatura ambiente) com o aumento da temperatura, indicando que a proteína perde pouco de seu conteúdo de estrutura secundária mesmo após aquecimento a 95°C (monitorado pelo sinal do espectro a 222 nm, Figura 2.1B). Essa alta estabilidade térmica da proteína é justificada pelo fato de pertencer a um organismo termófilo, *P. abyssi*, que vive em altas temperaturas e cujo desenovelamento de suas proteínas ocasionado por aumento de temperatura inviabilizaria sua sobrevivência.

Uma vez verificada a presença de estrutura secundária na PaUPF0023 e analisada sua estabilidade, foram realizados os ensaios de cristalização com essa proteína. Foram obtidos cristais em várias condições de cristalização testadas, contendo diferentes concentrações de fosfato de amônio na solução do reservatório

(Figura 2.2). Os cristais se formaram dentro de 3 dias e cresceram rapidamente, atingindo seu tamanho máximo (aproximadamente 1 mm) em menos de 24 horas a partir de seu surgimento. Os cristais foram submetidos a experimentos de difração de raios X na linha de luz de cristalografia de proteínas MX1 do Laboratório Nacional de Luz Síncrotron. Apesar do tamanho e morfologia dos cristais, os dados de difração obtidos apresentaram resolução máxima de 4,0 Å, indicando desordem na rede cristalina.

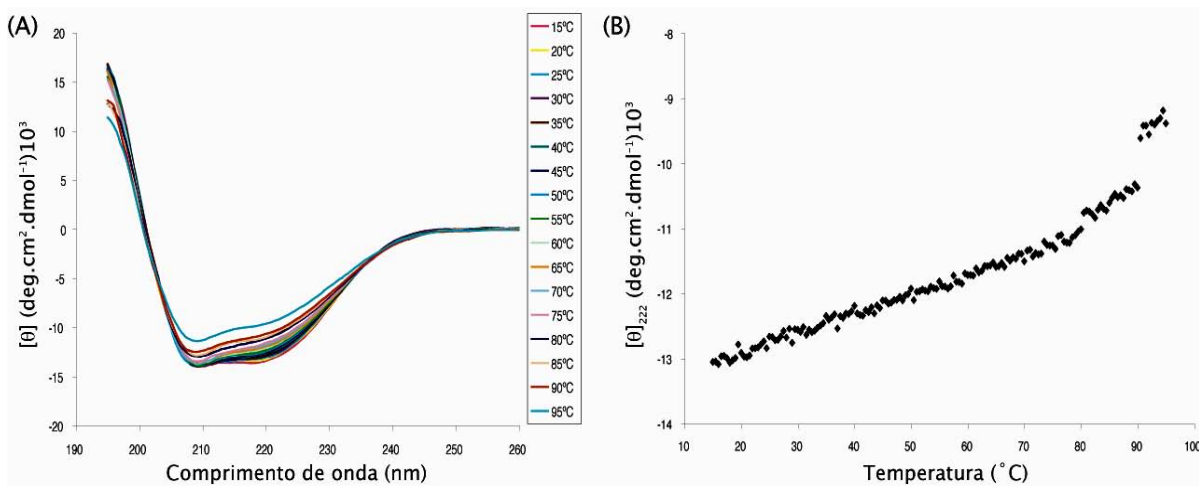


Figura 2.1: Análises de dicroísmo circular da PaUPF0023. (A) Espectros medidos durante o aumento da temperatura, de 15°C a 95°C, com intervalos de 5°C. (B) Ensaio de desnaturação térmica, monitorado através da elipicidade molar a 222 nm com a variação da temperatura de 15 a 95°C.

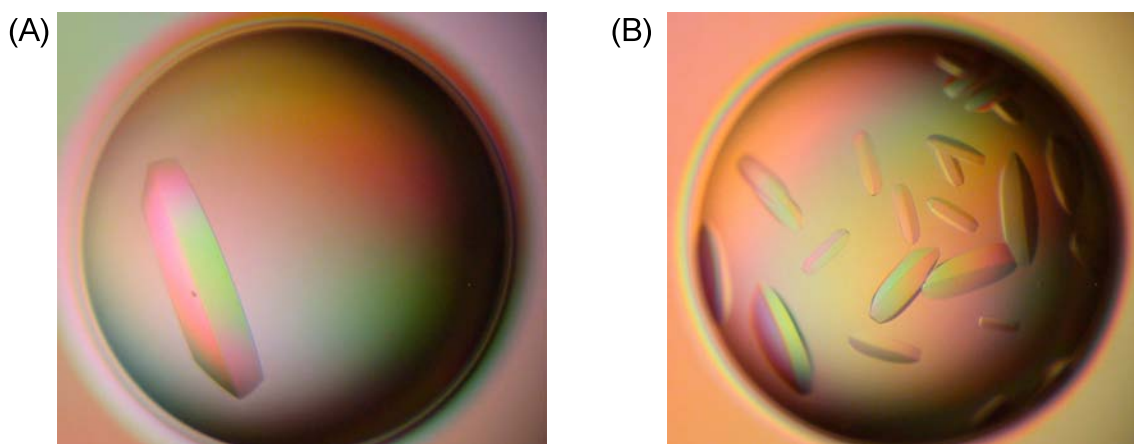


Figura 2.2: Cristais da proteína PaUPF0023. Condições de cristalização contendo (A) 0,4 M e (B) 0,5 M de fosfato de amônio.

Com o intuito de melhorar a qualidade dos cristais de proteína para obtenção de dados com melhor resolução, foram realizadas modificações nos protocolos de

preparo da amostra e de cristalização da proteína, além da utilização de diversas técnicas de melhoramento de cristais após sua formação.

Durante o preparo da amostra, com a finalidade de torná-la mais homogênea, a clivagem da proteína foi realizada com trombina ao invés de tripsina. Também foi realizado o aquecimento da amostra de proteína a 40°C, 60°C, 80°C e 95°C por um curto período de tempo (10 minutos) antes da realização dos ensaios de cristalização, pois o aquecimento induz à desnaturação e precipitação de moléculas de proteínas parcialmente enoveladas, que podem então ser removidas do sistema através de centrifugação (Newman, 2006). Não foi verificada melhora nos dados de difração dos cristais obtidos após clivagem com trombina; o aquecimento da amostra a 95°C previamente aos ensaios de cristalização resultou em ligeira melhora na difração dos cristais.

Durante os experimentos de cristalização, além da utilização de diferentes aditivos na gota de cristalização que poderiam influenciar no processo de formação dos cristais, alguns ensaios visando diminuir a velocidade de crescimento do cristal também foram testados: utilização de menor concentração de proteína na gota de cristalização; acréscimo de óleo sobre a solução do reservatório; realização de ensaios a 4°C e 12°C; desacoplamento da nucleação e crescimento do cristal (Saridakis & Chayen, 2000); adição de agarose na gota de cristalização (Biertümpfel *et al*, 2002). A adição de óleo na solução do poço e os ensaios com aditivos resultaram em uma ligeira melhora no padrão de difração dos cristais.

Adicionalmente, foram testados os procedimentos de *annealing*, desidratação (*soaking*) e *seeding* na tentativa de melhoria da qualidade dos cristais. No processo de annealing (Harp *et al*, 1998), o cristal foi retirado do feixe de nitrogênio, após ter sido resfriado, e colocado novamente durante a coleta de dados, a fim de sofrer um “choque térmico”. Já a desidratação do cristal foi realizada tanto na própria placa de cristalização, transferindo o cristal para uma gota com maior concentração de precipitante (Fu *et al*, 1999), como na coleta de dados, deixando que o cristal permanecesse um período maior na solução crioprotetora (Abergel, 2004). Ambos os procedimentos de annealing e soaking resultaram em piora na difração ou mesmo congelamento dos cristais de proteína. Já a técnica de seeding (Bergfors, 2003), realizada através da trituração de um cristal em “sementes” (seeds) e a

transferência dessas sementes para uma nova gota de cristalização, permitiu certa melhora no padrão de difração dos cristais.

Ao todo, foram testados aproximadamente 200 cristais de proteína na linha de luz de cristalografia do LNLS, entretanto, apesar dos diversos métodos empregados na tentativa de melhoria dos cristais, o limite de difração dos mesmos permaneceu em torno de 4 Å de resolução. Decidiu-se, então, investir na tentativa de cristalização das demais ortólogas.

2.2.2. SBDS de *Saccharomyces cerevisiae*: YLR022c

O cDNA que codifica a proteína YLR022c foi clonado no vetor pET-TEV e sua expressão em *E. coli* resultou em grande quantidade de proteína solúvel. Esta proteína, contendo polihistidinas em seu N-terminal, foi purificada por cromatografia de afinidade seguida de coluna de troca iônica (Figura 2.3A). A cauda de histidina presente em seu N-terminal foi removida utilizando-se a protease TEV e ensaios de cristalização foram realizados extensivamente com esta proteína. Em nenhuma das condições iniciais de cristalização testadas foi observada a formação de cristais. As seguintes condições nas quais foram encontrados precipitados não-amorfos, promissores para a possível formação de cristais, foram refinadas variando-se o pH e a concentração de sal da solução do reservatório (Figura 2.3B):

- Wizard 1 - 21 (20% m/v PEG 8000; HEPES pH 7,2);
- Index 1 - 47 (28% m/v PEG MME 2000; 0,1 M Bis-Tris-HCl pH 6,5);
- JBS 1 - D6 (30% m/v PEG 3000; 0,1 M Tris-HCl pH 8,5; 0,2 M Sulfato de Lítio);
- JBS 3 - D6 (32% m/v PEG 4000; 0,1 M Tris-HCl pH 8,5; 0,8 M Cloreto de Lítio).

Entretanto, em nenhum dos refinamentos realizados foi verificada a formação de cristais de proteína, novamente impedindo o estudo cristalográfico desta homóloga da SBDS.

Apesar da dificuldade encontrada na cristalização da proteína YLR022c, análises por dicroísmo circular indicaram a presença de grande quantidade de estruturas secundárias do tipo α -hélice na proteína (Figura 2.4A), igualmente à sua homóloga PaUPF0023. A predição de estrutura secundária demonstra que a

proteína deve possuir aproximadamente 30% de α -hélice em sua estrutura e quase a mesma quantidade de folha β , mas também apresenta grande quantidade de estrutura desenovelada ou “random coil” (26%). A análise da estabilidade térmica da YLR022c por dicroísmo circular (Figura 2.4B) sugere que, embora a proteína perca gradativamente seus elementos de estrutura secundária com o aumento da temperatura, os espectros obtidos em temperaturas elevadas conservam mais da metade do sinal inicial, indicando que essa proteína conserva regiões estruturadas mesmo a 95°C.

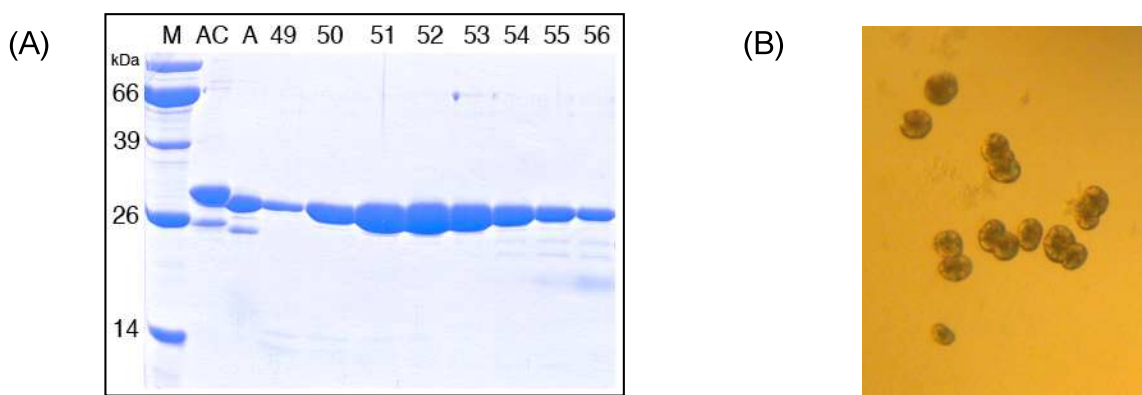


Figura 2.3: Purificação e cristalização da YLR022c. (A) Análise por SDS-PAGE (12,5%) das frações da purificação por cromatografia de troca iônica. M – marcador de peso molecular; AC – amostra antes de clivar com a protease TEV; A – amostra aplicada na coluna, após clivagem da cauda de histidina; 49 a 56, 60 e 61 – frações durante eluição. (B) Precipitado cristalino obtido no refinamento da condição JBS 3 – D6, durante os ensaios de cristalização da proteína YLR022c.

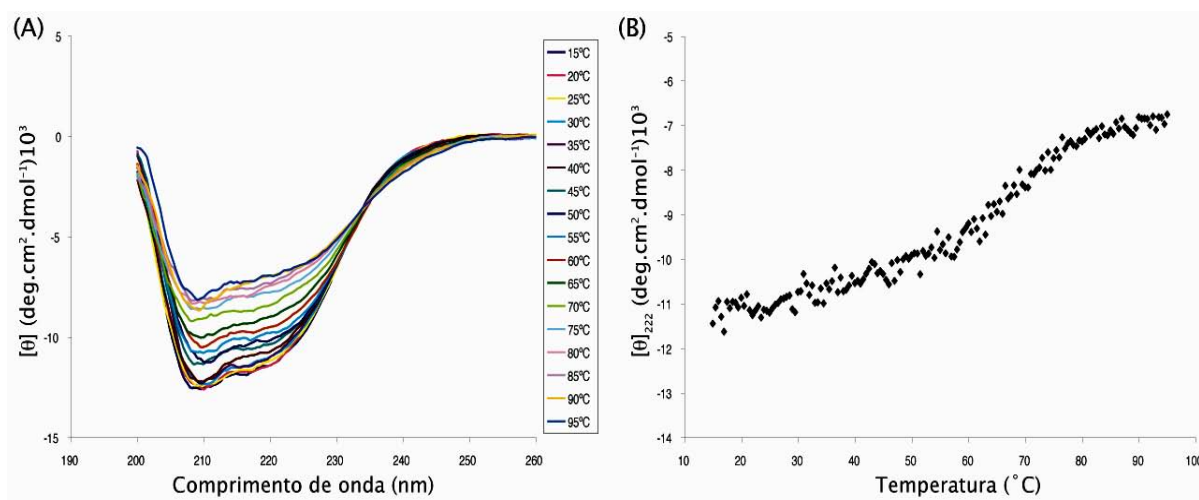


Figura 2.4: Análises de dicroísmo circular da YLR022c. (A) Espectros coletados durante o aumento da temperatura de 15°C a 95°C. (B) Ensaio de estabilidade térmica, analisado através da variação da elipicidade molar a 222 nm com o aumento da temperatura.

2.2.3. SBDS de *Trypanosoma cruzi*: TcSBDS

O gene correspondente à proteína TcSBDS foi amplificado e inserido no vetor pET-TEV para expressão desta ortóloga. Foi identificada a presença de uma região C-terminal adicional nessa proteína quando comparada à maioria das ortólogas (resíduos 264 a 464). A caracterização estrutural deste C-terminal estendido foi realizada (resultados descritos no capítulo 3) e neste estudo foi verificado que o domínio adicional da proteína TcSBDS é intrinsecamente desenhado. Regiões desestruturadas dificultam o empacotamento regular de moléculas de proteína e consequentemente sua cristalização. Por este motivo, resolveu-se realizar os ensaios de cristalização com a proteína TcSBDS truncada, contendo os resíduos 1 a 263. O limite C-terminal da proteína truncada foi determinado com base no alinhamento sequencial das ortólogas que não contém a região C-terminal estendida.

O cDNA correspondente à TcSBDS truncada foi clonado e a proteína foi expressa em *E. coli*, contendo polihistidinas em seu N-terminal. Embora grande parte da proteína truncada tenha sido expressa na fração insolúvel, foi possível obter quantidade razoável de proteína solúvel. Sua purificação foi realizada através de cromatografia de afinidade (Figura 2.5) e a análise por SDS-PAGE indicou alta pureza da proteína. Porém, através de ensaio de espectroscopia no ultravioleta (UV) para a determinação da concentração da proteína, verificou-se alta absorbância da amostra a 260 nm, provavelmente devido à presença de grande quantidade de ácidos nucleicos. Foram feitos alguns procedimentos para tentar eliminar esses ácidos nucleicos da amostra, como: adição de glicerol na preparação do extrato; tratamento do extrato protéico com DNase e RNase; precipitação do extrato com sulfato de estreptomicina; cromatografia de troca aniônica, com colunas carregadas positivamente que interagem com o ácido nucleico; expressão da proteína como corpos de inclusão. Porém nenhum procedimento testado mostrou-se eficiente para tal propósito.

Embora a possibilidade de realização de ensaios de cristalização desta proteína na presença do ácido nucleico tenha sido considerada, uma vez que este poderia ser importante para a estabilização da proteína, análises por espalhamento dinâmico de luz (DLS) demonstraram que a amostra apresentava-se polidispersa e

com raio de giro aproximado de 15 nm, indicando um agregado de quase 2.000 kDa. Na tentativa de eliminar essa agregação foram utilizados diferentes tipos de aditivos, baseado em metodologia descrita por Bondos e Bicknell, 2003, mas nenhum dos aditivos testados foi eficiente para inibição da agregação. Além disso, foi testado o efeito da variação do pH da solução da proteína. Em pH 5,2 foi verificada diminuição da absorbância em 260 nm por espectroscopia e a presença de monômero da proteína através de experimentos de DLS (Tabela 2.7). Entretanto, aproximadamente 30% da proteína precipitou neste pH (conforme indicado na última coluna da Tabela 2.7).

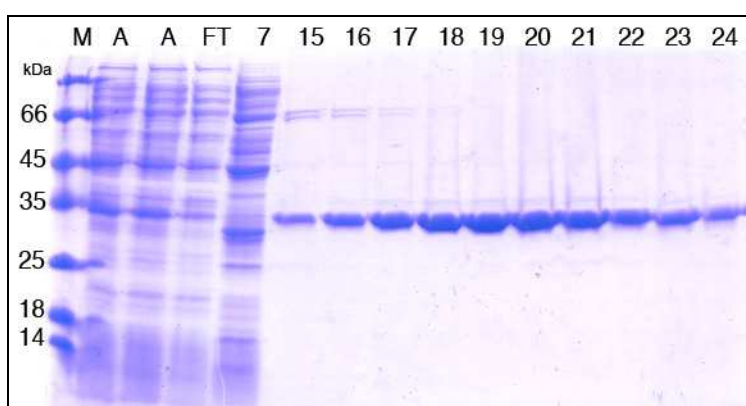


Figura 2.5: Purificação da proteína TcSBDS truncada. Análise por SDS-PAGE (12,5%) das frações da cromatografia de afinidade obtidas durante eluição. M – marcador de peso molecular; A – amostra aplicada na coluna (os dois poços correspondem a dois volumes idênticos de culturas de células utilizados para expressão); FT - flow-through; 7, 15 a 24 – frações do cromatograma durante eluição.

Tabela 2.7: Análise da TcSBDS truncada em diferentes pHs por experimentos de DLS e espectroscopia de absorção de luz ultravioleta.

pH	Raio (nm)	% polidispersividade	Massa molecular (kDa)	Concentração (mg/mL)
5,2	2,6	34,6	32	0,28
7,2	14,9	62	1863	0,42
8,5	16,2	53	2278	0,43

Apesar da perda de grande parte da proteína em pH 5,2, ensaios de cristalização foram realizados com a TcSBDS truncada neste pH, uma vez verificada a ausência de agregados na amostra e considerando-se que a diminuição da concentração neste pH poderia indicar a precipitação da fração agregada de proteína/ácido nucléico. Logo após a confecção dos experimentos de cristalização,

foi verificada a formação de precipitado na maioria das gotas das placas preparadas, além da precipitação da proteína que restou dos ensaios (mantida em gelo durante todo o tempo), indicando instabilidade da proteína na condição em que se encontrava. As placas de cristalização foram visualizadas e percebeu-se que, dentre as condições apresentando gota clara (sem precipitado), a maioria continha alta concentração de sal. Na tentativa de estabilizar a proteína, resolveu-se aumentar a concentração de sal na amostra e diminuir a concentração de proteína. Porém, novamente foi verificada a formação de precipitado nas gotas durante a realização de novos ensaios de cristalização. Inclusive, a própria amostra de proteína precipitou no tubo “eppendorf” enquanto as primeiras placas eram montadas automaticamente pelo robô de cristalização, o que inviabilizou a continuação dos experimentos. Portanto, o estudo cristalográfico desta ortóloga de *T. cruzi* foi impossibilitado devido à instabilidade da proteína truncada. Estudos funcionais realizados com a TcSBDS e a caracterização estrutural de seu C-terminal adicional encontram-se descritos no capítulo 3.

2.2.4. SBDS de *Homo sapiens*: HsSBDS

O vetor pET-TEV/HsSBDS foi utilizado para a expressão da proteína SBDS humana em *E. coli*. Grande quantidade de proteína foi obtida na fração solúvel, inicialmente purificada por cromatografia de afinidade a níquel. A análise das frações desta purificação por SDS-PAGE permitiu a identificação da proteína em dois picos distintos do cromatograma. Experimentos de DLS foram realizados com amostras dos diferentes picos e foi verificado que um pico correspondia à proteína na sua forma monomérica e o outro, oligomérica. A fração monomérica foi submetida a uma segunda etapa de purificação, através de coluna de troca iônica. Antes disso, parte da amostra foi submetida à proteólise para remoção da cauda de histidina, utilizando-se a protease TEV (Figura 2.6). Análises recentes de comparação de estruturas resolvidas com e sem “His-tag” tem demonstrado que caudas adicionais podem não interferir na estrutura da proteína e no empacotamento das moléculas e, em alguns casos, até mesmo ajudar na sua cristalização (Carson *et al.*, 2007). Dessa forma, decidiu-se utilizar amostras da proteína com e sem cauda de histidina para ensaios de cristalização.

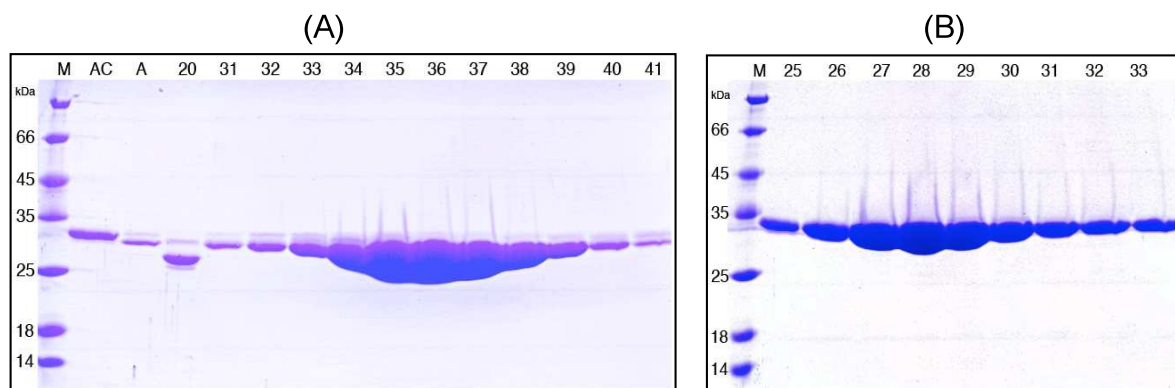


Figura 2.6: Purificação da proteína SBDS humana. SDS-PAGE (12,5%) contendo as frações eluídas da cromatografia de troca iônica (HiTrap Heparina HP). (A) Proteína sem cauda de histidina e (B) proteína com cauda de histidina. M – marcador de peso molecular; AC – amostra antes de clivar com TEV; A – amostra aplicada na coluna, após clivagem da cauda; 20, 31 a 41 e 25 a 33 - frações do cromatograma durante eluição.

Análises por DLS foram realizadas com a proteína HsSBDS purificada, na presença e ausência da cauda de histidina. Foi verificado que a proteína encontrava-se monomérica (com raio de giro de ~ 3 nm, representando uma proteína globular de aproximadamente 40 kDa) e monodispersa (com 14% de polidispersividade). Para confirmar os dados de DLS, foi feita uma cromatografia de exclusão molecular analítica, com a proteína sem cauda de histidina (Figura 2.7). O valor estimado para a massa molecular da HsSBDS, obtido através da comparação do volume de retenção desta com o de proteínas com massas moleculares conhecidas (padrões moleculares), foi de 32 kDa, valor próximo à sua massa teórica (28.957 Da). Portanto, a proteína se comporta realmente como um monômero em solução.

Experimentos de CD foram feitos com a proteína sem cauda de histidina. Similarmente às proteínas PaUPF0023 e YLR022c, esta proteína apresentou um espectro de CD correspondente ao de proteínas que contêm principalmente α -hélices (Figura 2.8A). A predição de estrutura secundária da HsSBDS a partir do programa “CDNN Deconvolution” sugere a predominância de estruturas desenoveladas (“random coil”, 36%), com α -hélice em segundo lugar (28%). Experimentos de desnaturação térmica mostraram grande alteração no espectro da proteína com o aumento da temperatura (Figura 2.8B), indicando que a proteína perde praticamente toda a sua estrutura secundária em temperaturas elevadas. Por esse motivo, foi possível calcular a temperatura média de transição (T_m) dessa proteína, na qual 50% das moléculas estão desenoveladas: 70,6°C.

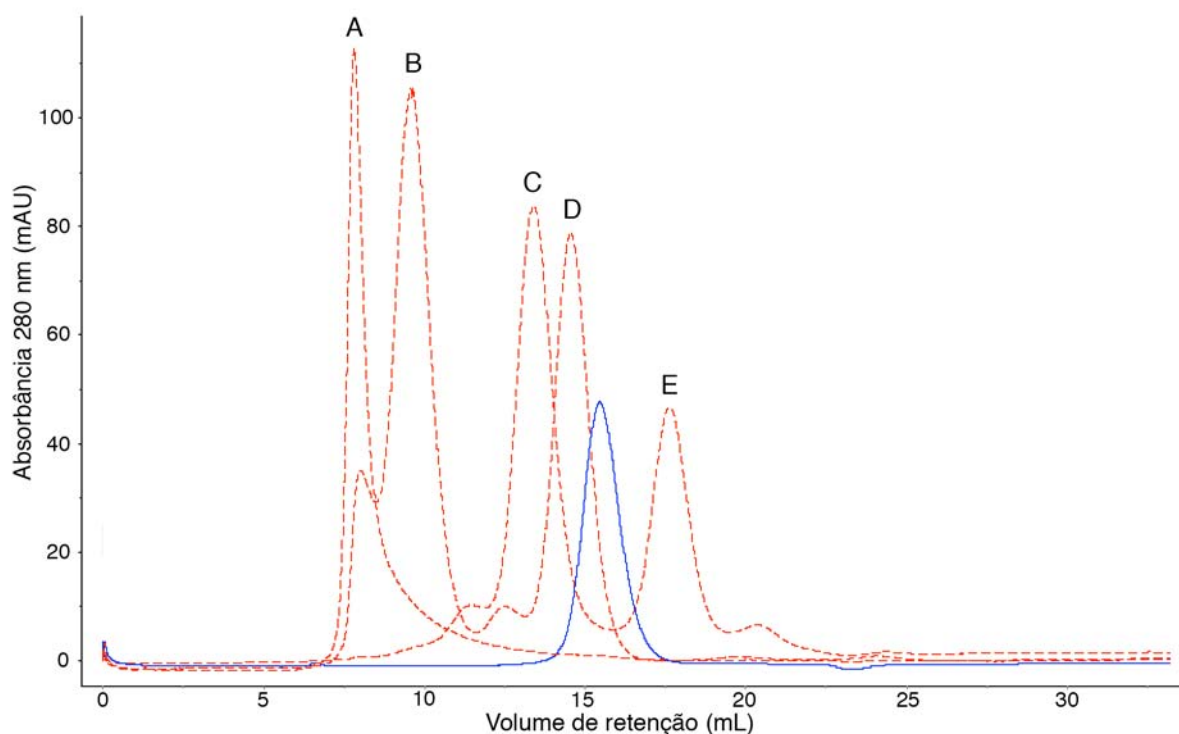


Figura 2.7: Cromatografia de exclusão molecular analítica da proteína HsSBDS. Perfil cromatográfico da HsSBDS (azul) e de padrões moleculares (vermelho), realizado em coluna Superdex 200 10/30. A – Blue dextran (2000 kDa); B – Ferritina (440 kDa); C – Conalbumina (75 kDa); D – Ovalbumina (43 kDa); E – Ribonuclease A (13,7 kDa).

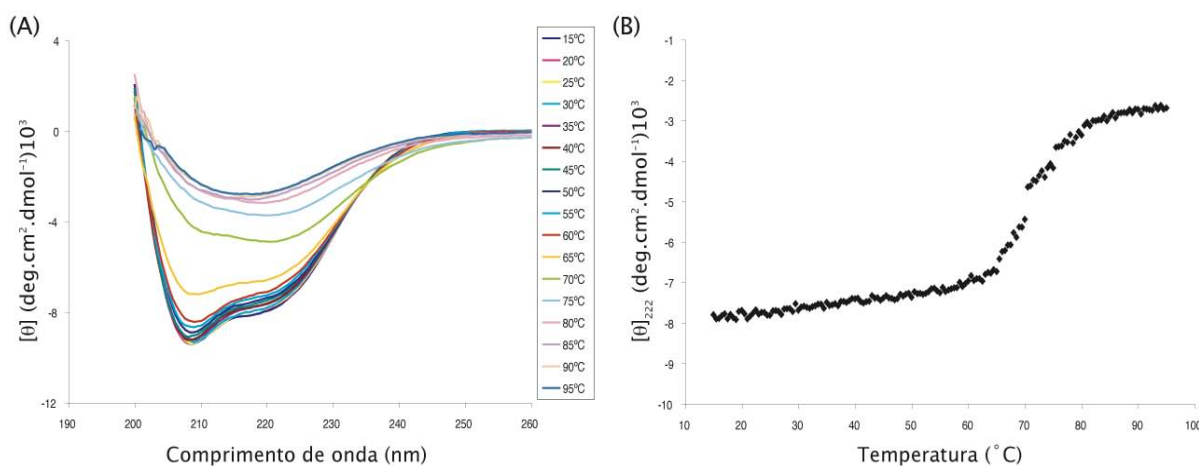


Figura 2.8: Análises de dicroísmo circular da HsSBDS. (A) Espectros coletados de 15°C a 95°C. (B) Desnaturação térmica da proteína, acompanhada pela variação da elipcidade molar a 222 nm com o aumento da temperatura.

Os ensaios iniciais de cristalização realizados com a proteína sem cauda de histidina permitiram a identificação de algumas condições interessantes para refinamento (Figura 2.9):

- Crystal Screen 2 - 2 (0,5 M NaCl; 0,01 M CTAB; 0,01 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$);
- Wizard 1 – 26 (10% PEG 3000; CHES pH 9,5);
- SaltRX1 – 34 (2,2 M ácido málico pH 7,0; 0,1 M Bis-Tris propano pH 7,0).

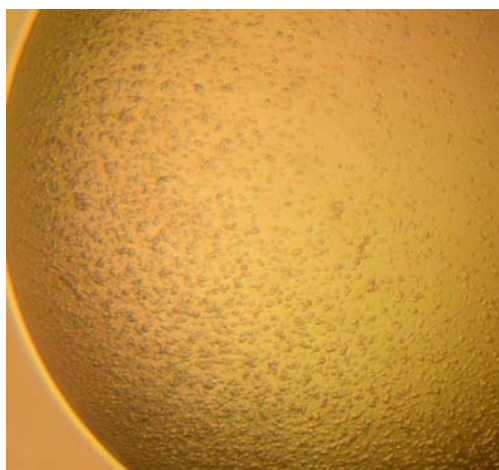


Figura 2.9: Micro-cristais obtidos durante ensaios de cristalização com a proteína SBDS humana, observados na condição 34 do kit SaltRX1, feito com a proteína sem cauda de histidina.

As extensivas tentativas de refinamento das condições iniciais de cristalização encontradas não apresentaram resultados promissores. Diante disso, foram realizados ensaios de proteólise limitada com esta proteína. O método de proteólise limitada tem sido descrito como uma ferramenta eficiente para remoção de regiões flexíveis ou desordenadas de proteínas (Brooun *et al.*, 2007), ou seja, de regiões que interferem no empacotamento das moléculas, podendo inviabilizar sua cristalização. A combinação de proteólise limitada com espectrometria de massas tem sido uma metodologia rápida e sensível para analisar os fragmentos gerados durante a proteólise, o que permite identificar e remover regiões flexíveis da proteína de interesse (Renzone *et al.*, 2007). Foram realizados experimentos utilizando-se as seguintes proteases: tripsina, quimiotripsina, papaína, proteinase K, subtilisina e termolisina (Figura 2.10).

Os ensaios realizados com tripsina, quimiotripsina, proteinase K e subtilisina demonstraram fragmentos estáveis de aproximadamente 20 kDa (indicados por setas na Figura 2.10). A amostra digerida com tripsina por 2 horas foi submetida a análise por espectrometria de massas com ionização por *Electrospray* (ESI-MS), que permitiu a identificação do domínio resistente a proteólise limitada, de massa 17494.0 Da (Figura 2.11). Com base na sequência da proteína, essa massa

corresponde aos resíduos 101 a 250 (massa teórica de 17494.5 Da), indicando que a tripsina removeu o N-terminal da proteína (correspondente ao domínio I na Figura 1.3).

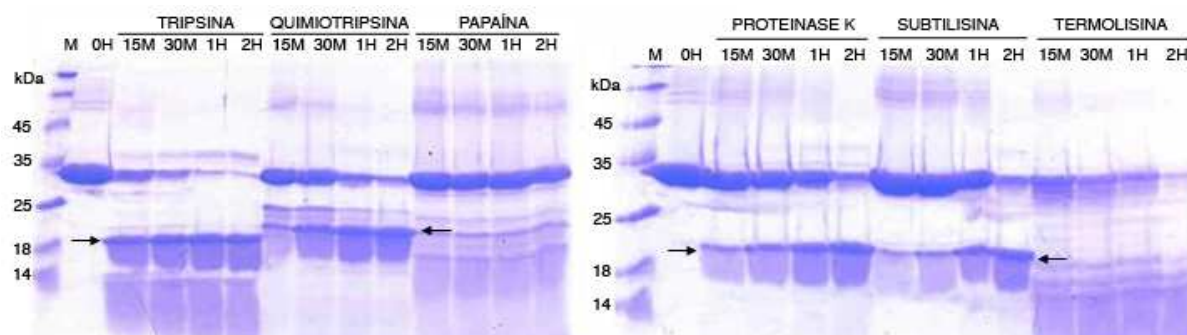


Figura 2.10: Proteólise limitada da HsSBDS. Análise por SDS-PAGE (15%) das frações obtidas durante ensaio com as proteases tripsina, quimiotripsina, papaína, proteinase K, subtilisina e termolisina. M – marcador de peso molecular; 0H – amostra coletada antes da adição da protease; 15M e 30M – amostra coletada após 15 e 30 minutos de clivagem; 1H e 2H – amostra coletada após 1 e 2 horas de clivagem. As setas indicam os domínios resistentes a proteólise.

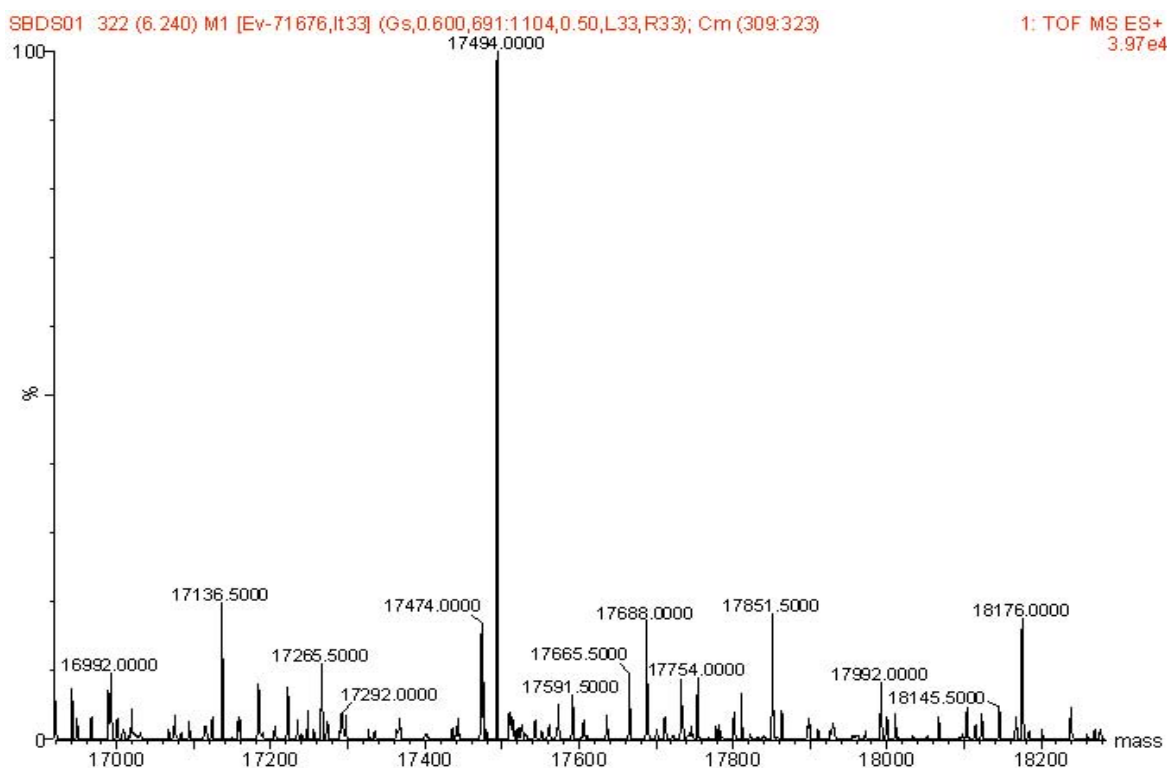


Figura 2.11: Espectro de massa da HsSBDS após proteólise limitada com tripsina. O domínio resistente a proteólise contém massa de 17494.0 Da, correspondente aos resíduos 101-250 da proteína.

O domínio da HsSBDS resistente à proteólise limitada foi submetido a ensaios de cristalização, após sua purificação por cromatografia de troca iônica, porém não foram verificadas condições promissoras para a formação de cristal que pudessem ser refinadas. Provavelmente, a presença de fragmentos de diferentes tamanhos na amostra de proteína, resultantes da proteólise limitada, interferiu na formação de cristais ou precipitado cristalino. Como alternativa, foram planejadas construções de clones contendo os domínios da proteína separadamente e pares desses domínios (conforme indicado na Tabela 2.2). Tais construções foram baseadas nos dados de espectrometria de massas e alinhamento com a estrutura resolvida da AfSBDS.

Após clonagem dos domínios, foi possível expressar em grande quantidade, de forma solúvel, os domínios 2, 1+2 e 2+3. Os domínios 1+2 e 2+3 foram purificados para ensaios de cristalização. A purificação do domínio 1+2 consistiu de apenas uma etapa, através de cromatografia de afinidade, pois grande parte dos contaminantes restantes na amostra precipitaram durante diálise contra tampão de cristalização (Figura 2.12A). Para a purificação do domínio 2+3 foi necessário, além da cromatografia de afinidade, uma segunda etapa realizada através de cromatografia de troca iônica (Figura 2.12B).

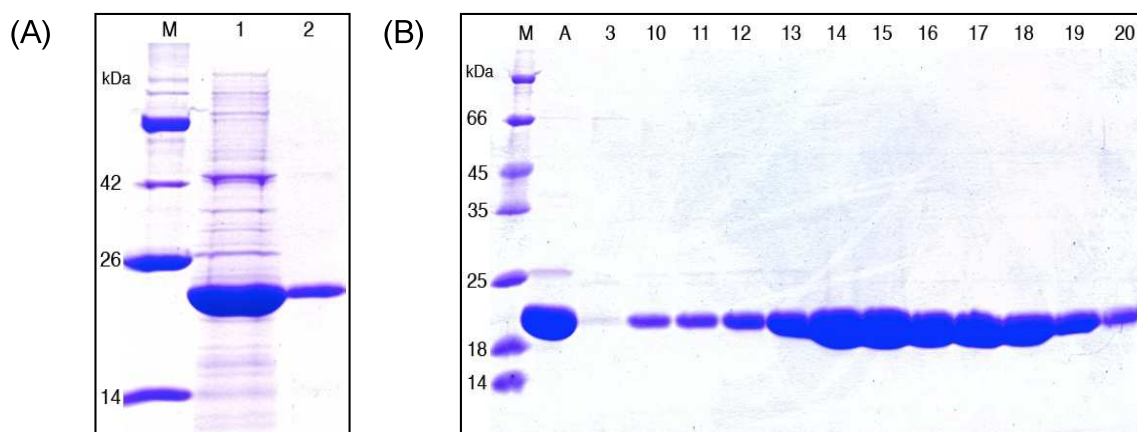


Figura 2.12: Purificação dos domínios da HsSBDS. (A) Análise por SDS-PAGE (12,5%) de amostras do domínio 1+2 da HsSBDS: (1) após purificação por cromatografia de afinidade e (2) após diálise contra tampão de cristalização. (B) Análise por SDS-PAGE de frações do domínio 2+3 da HsSBDS após purificação por cromatografia de troca iônica (HiTrap Heparina HP). M – marcador de peso molecular; A – amostra aplicada na coluna; 3, 10 a 20 – frações durante eluição.

Nos ensaios de cristalização realizados com os domínios 1+2 foram encontradas esferulitas em duas condições testadas, ambas contendo 100 mM imidazol pH 8,0 e 200 mM acetato de zinco, uma na presença de 2,5 M NaCl e a outra contendo 20% 1,4-butanodiol. Com relação ao domínio 2+3, verificou-se a presença de microcristais na condição contendo 100 mM Tris-HCl pH 8,5 e 2,4 M fosfato de amônio. A estas condições foi adicionado o corante Izit Crystal Dye (Hampton Research), próprio para cristal de proteína, e foi verificada a coloração azul nesses pequenos cristais, indicando serem formados por proteína. Tais condições de cristalização foram refinadas na tentativa de obtenção de cristais apropriados para experimentos de difração, porém sem sucesso. Paralelamente aos experimentos de cristalização dos domínios da HsSBDS, foi dado início aos experimentos visando a resolução da estrutura da proteína SBDS humana por Ressonância Magnética Nuclear (RMN).

Ressonância Magnética Nuclear da HsSBDS

A determinação da estrutura de uma proteína por RMN inicia-se com o assinalamento dos espectros, o que significa associar cada ressonância, ou pico no espectro, a um núcleo específico da proteína de interesse. Esse assinalamento é dividido em duas etapas, o assinalamento sequencial da cadeia principal da proteína e o assinalamento das cadeias laterais dos aminoácidos. Para o assinalamento sequencial da HsSBDS, foram coletados os seguintes experimentos de tripla ressonância no espectrômetro de 600 MHz do LNLS, com a proteína duplamente marcada ($^{15}\text{N}/^{13}\text{C}$): HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CBCACO)NH. A nomenclatura desses experimentos reflete a transferência da magnetização que ocorre em cada um deles, sendo que os núcleos envolvidos na transferência da magnetização formam o nome do experimento. Por exemplo, o experimento que transfere a magnetização do próton da amida (H_N) via nitrogênio da amida (N) para o carbono da carbonila (CO) do resíduo anterior é chamado de HNCO (Sattler *et al.*, 1999). Na Figura 2.13 estão ilustradas as transferências de magnetização que ocorrem em cada um dos experimentos utilizados.

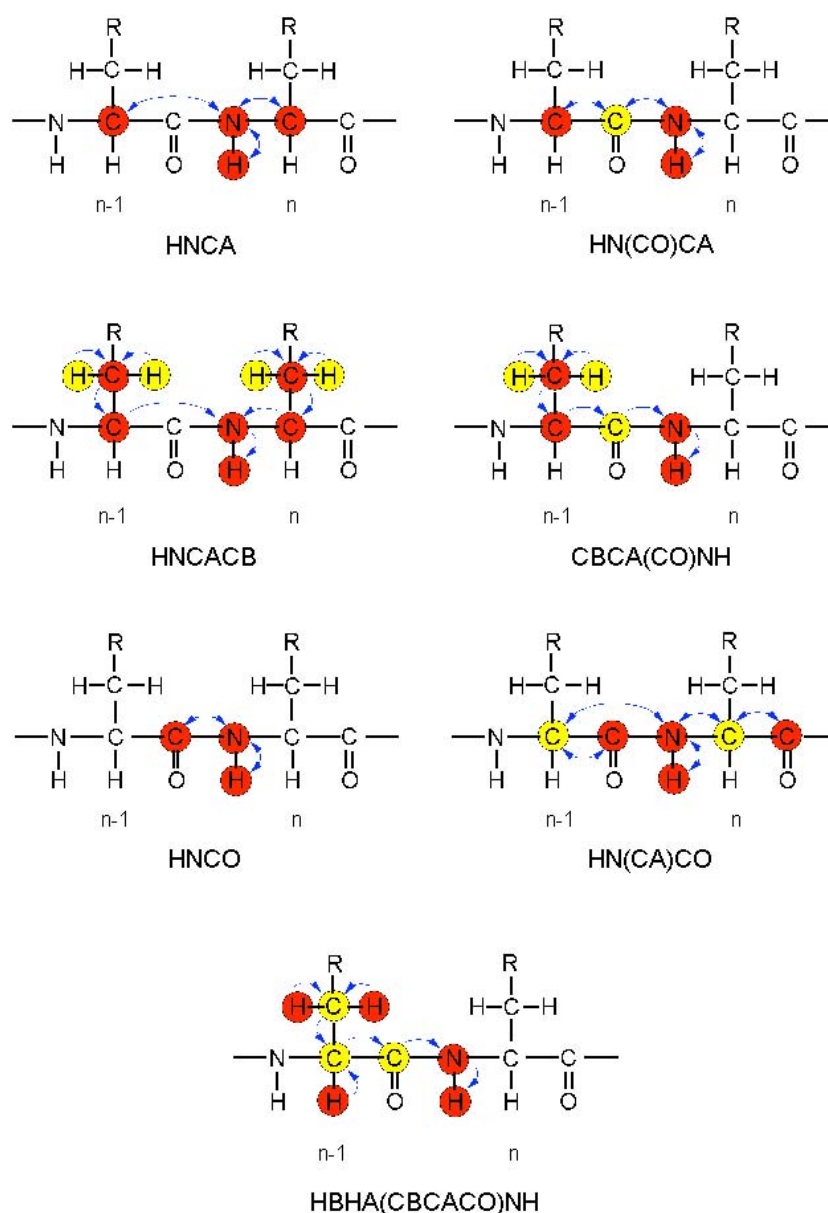


Figura 2.13: Ilustração dos experimentos de tripla ressonância coletados para o assinalamento sequencial da proteína HsSBDS. A transferência da magnetização que ocorre em cada um dos experimentos está indicada por setas. Os átomos em vermelho são os núcleos observados nos experimentos e os átomos em amarelo apenas estão envolvidos na transferência da magnetização (figuras retiradas do site da Escola de Cristalografia da Universidade de Londres: <http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/home.htm>).

Como a proteína HsSBDS, de 250 resíduos, é relativamente grande para a técnica de RMN, foi constatada a ausência de alguns sinais e a sobreposição de picos nos espectros obtidos, fazendo-se necessária a marcação da proteína com deutério, para aumentar a intensidade dos sinais, e a realização de experimentos em um espectrômetro de maior campo magnético, para obtenção de maior resolução espectral. Após deuteração da proteína, foi estabelecida uma colaboração com a

Dra. Tharin Blumenschein da Universidade de East Anglia (*School of Chemical Sciences and Pharmacy, University of East Anglia*), em Norwich, Inglaterra. A amostra da HsSBDS marcada com $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ foi enviada para Dra. Tharin e por ela foram realizados os experimentos HNCACB e HN(CO)CACB no espectrômetro de 800 MHz. O experimento HNCACB com a proteína deuterada também foi coletado no espectrômetro de 600 MHz do LNLS. Verificou-se uma melhora significativa na intensidade dos sinais dos espectros com a proteína deuterada. Já o aumento na resolução dos espectros obtidos no equipamento de 800 MHz, considerado pouco expressivo quando comparado aos espectros obtidos no 600 MHz, auxiliou no processo de diferenciação entre picos sobrepostos.

Utilizando-se os espectros tridimensionais obtidos (a 600 e 800 MHz), foi possível atribuir o deslocamento químico dos seguintes átomos de cada resíduo de aminoácido da proteína: HN_i , N_i , $\text{C}\alpha_i$, $\text{C}\beta_i$, CO_i , $\text{H}\alpha_{i-1}$, $\text{C}\alpha_{i-1}$, $\text{C}\beta_{i-1}$, CO_{i-1} , onde i é o resíduo em questão e $i-1$ é o resíduo anterior. Conhecendo-se os deslocamentos químicos de $\text{C}\alpha$, $\text{C}\beta$ e CO do resíduo i e de seu antecessor, considerando-se os valores médios de deslocamento químico de cada resíduo (obtido no *Biological Magnetic Resonance Data Bank* - BMRB) e sabendo-se a sequência de aminoácidos da proteína, foi realizada a atribuição sequencial dos resíduos da HsSBDS. Esse assinalamento sequencial foi realizado com o auxílio do programa Smartnotebook (Slupsky *et al.*, 2003), conforme demonstração na Figura 2.14.

A proteína HsSBDS utilizada para os experimentos de RMN continha, ao todo, 252 resíduos (considerando-se 2 aminoácidos adicionais no N-terminal, restantes do sítio de clivagem da protease TEV, após clivagem da cauda de histidina). Excluindo-se o resíduo N-terminal e as 6 prolinas presentes na sequência da proteína, que não apresentam pico no espectro de ^{15}N -HSQC, existiam 245 aminoácidos possíveis de serem assinalados neste espectro. Ao final do processo de assinalamento, foram identificados 213 resíduos, ou seja, 87% da cadeia principal da proteína. Restaram no espectro de ^{15}N -HSQC 8 picos que não foram atribuídos nem conectados na sequência assinalada, sendo que destes, 7 desapareciam em alguns espectros, ou seja, devem corresponder a resíduos que estão em regiões da proteína apresentando mobilidade conformacional, igualmente aos demais 24 resíduos que não aparecem no espectro. O espectro de ^{15}N -HSQC da HsSBDS assinalado encontra-se no Capítulo 3, Artigo 2, Figura Suplementar 1.

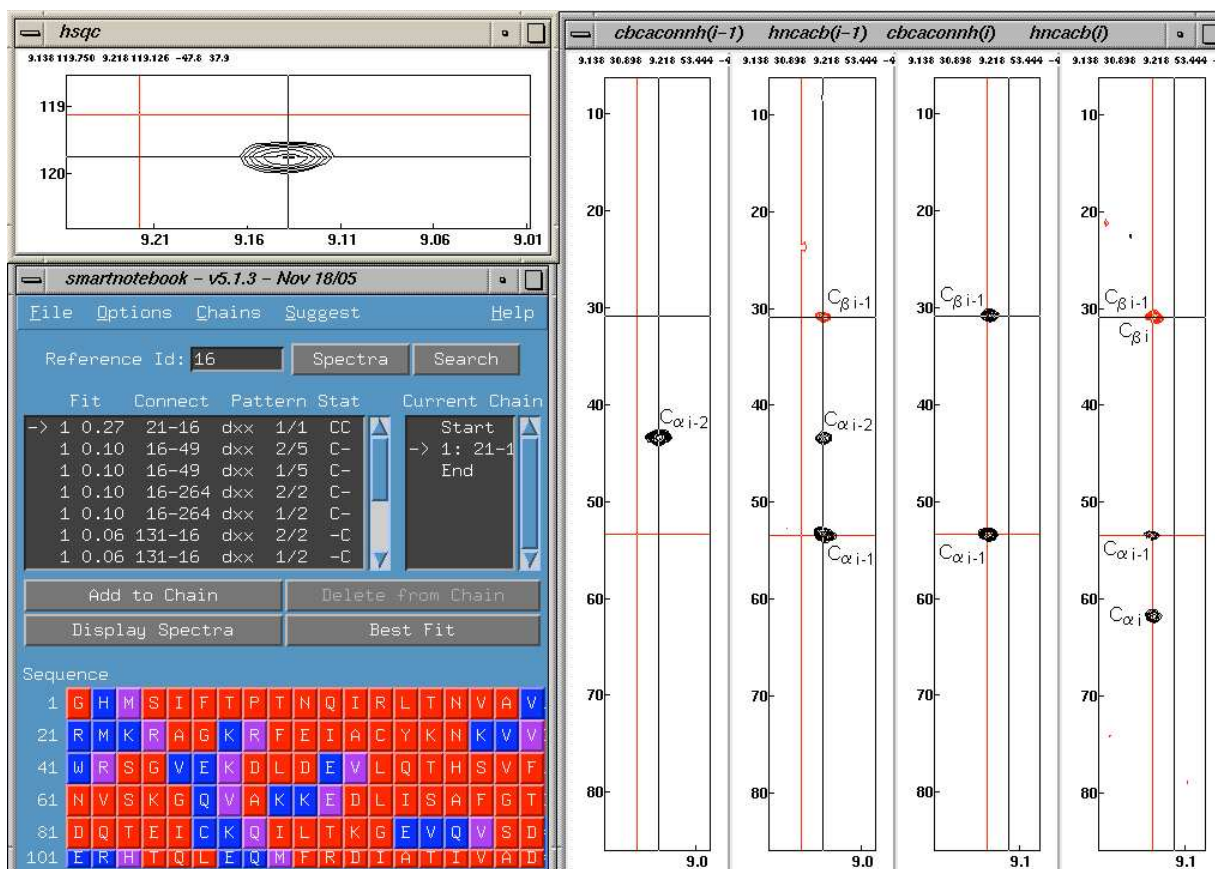


Figura 2.14: Assinalamento sequencial da proteína HsSBDS utilizando-se o programa Smartnotebook. No canto superior esquerdo encontra-se o espectro de ^{15}N -HSQC, ampliado no pico i de interesse (número 16 neste exemplo), conforme indicado na janela do Smartnotebook (azul) em “Reference Id”. Nessa janela, são sugeridos os picos do espectro de ^{15}N -HSQC possivelmente relacionados sequencialmente com o pico i (coluna Connect), indicando o “fit” de cada conexão. Ao selecionar uma das opções - neste exemplo a primeira linha (21-16) - na janela da direita são mostradas fatias alinhadas do plano $\text{H}-^{13}\text{C}$ dos espectros CBCA(CO)NH do resíduo $i-1$, HNCACB do resíduo $i-1$, CBCA(CO)NH do resíduo i , HNCACB do resíduo i , respectivamente. Nos espectros do tipo HNCACB são encontrados os picos correspondentes aos átomos $\text{C}\alpha$ e $\text{C}\beta$ dos resíduos i e $i-1$ e nos espectros CBCA(CO)NH estão os picos apenas do $\text{C}\alpha$ e $\text{C}\beta$ do resíduo $i-1$, conforme indicado na figura. Realizando esse procedimento, foi possível determinar que o pico número 21 encontra-se anteriormente ao 16 na sequência da proteína e, através dos valores dos deslocamentos químicos dos $\text{C}\alpha$ e $\text{C}\beta$, foi possível correlacioná-los com os resíduos na sequência da proteína: 16 corresponde a Val93; 21 é Glu92. Anterior ao pico 21 existe uma glicina, pois não apresenta $\text{C}\beta$ (é a Gly91). E assim por diante.

Uma vez determinado o deslocamento químico dos átomos da cadeia principal da HsSBDS, foi possível estimar os elementos de estrutura secundária presentes na proteína, pois existe uma grande dependência entre o valor de deslocamento químico e o tipo de estrutura secundária para cada um dos 20 aminoácidos. O “Chemical Shift Index” (CSI) é um método utilizado para estabelecer a estrutura secundária da proteínas baseando-se nas diferenças de deslocamento

químico com relação a valores tabelados, medidos para aminoácidos em estruturas desenoveladas (Wishart *et al.*, 1992; Wishart & Sykes, 1994). Nesse método, quando o valor do deslocamento químico medido é maior que o valor especificado no CSI, este é considerado como 1; se o valor medido for menor, é considerado -1; e quando encontra-se dentro da faixa esperada, vale 0. Uma região de α -hélice é definida quando são encontrados sequencialmente 4 ou mais valores -1 para o H α (ou 1 para o C α /CO) e fita β existe quando 3 ou mais valores 1 para o H α (ou -1 para o C α /CO) são encontrados sequencialmente. Utilizando-se o programa NRMView5 foram calculados os elementos de estrutura secundária presentes na HsSBDS (Figura 2.15).

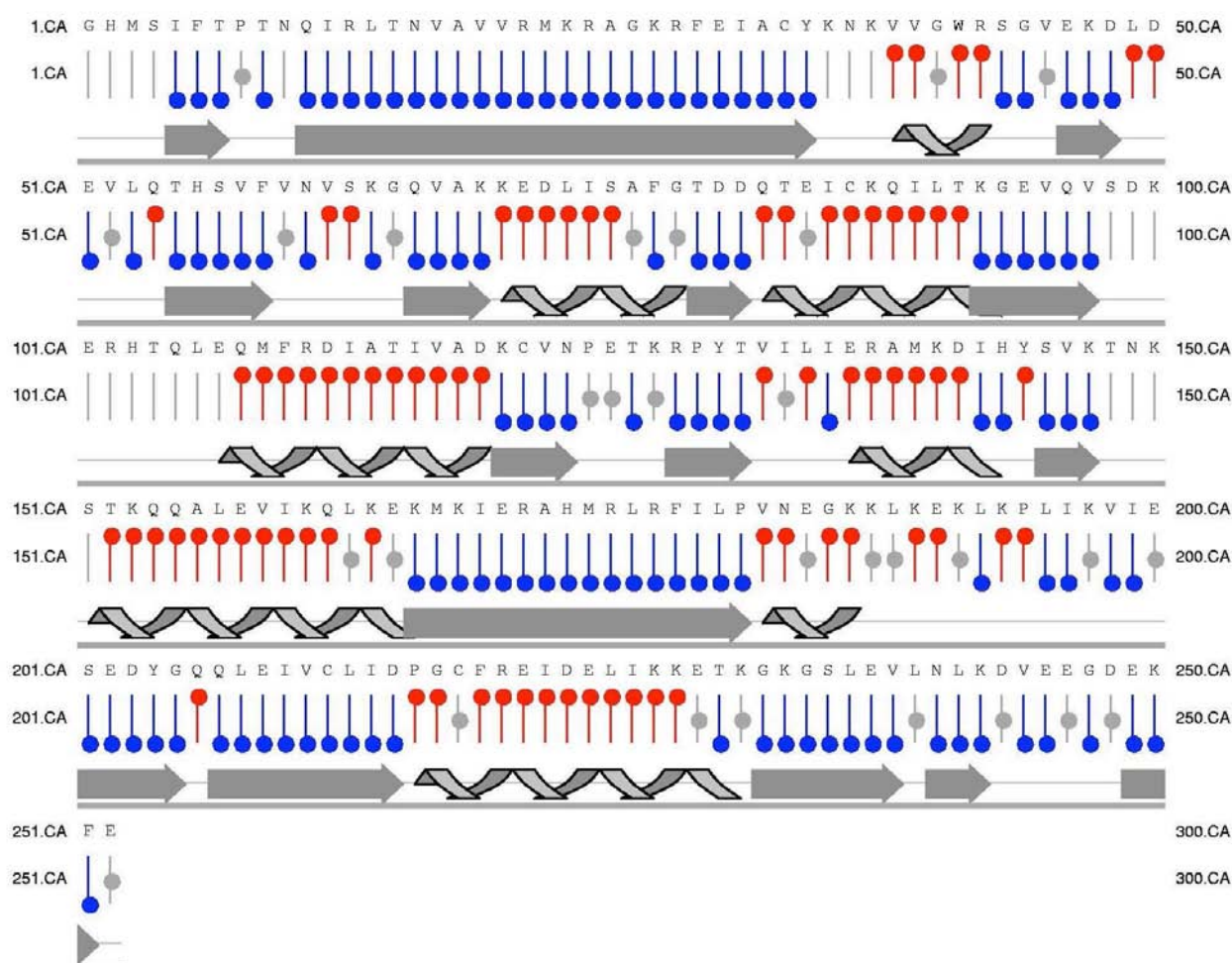


Figura 2.15: Chemical Shift Index (CSI) para os átomos C α da proteína HsSBDS. Em azul estão indicados os valores negativos e em vermelho os valores positivos. O CSI igual a 0 é mostrado em cinza. As setas indicam folha β e as espirais, α -hélices.

Após finalização do assinalamento sequencial da proteína HsSBDS, iniciou-se a etapa de assinalamento das cadeias laterais dos aminoácidos alifáticos. Para isso, foram coletados os experimentos ^{13}C -HSQC, HCCH-COSY e HCCH-TOCSY. No experimento 2D ^{13}C -HSQC são observadas as correlações de todos os ^1H ligados a ^{13}C e os experimentos 3D HCCH-COSY e HCCH-TOCSY são basicamente experimentos 2D de COSY e TOCSY resolvidos em uma terceira dimensão adicional contendo o ^{13}C . Enquanto no experimento de HCCH-COSY são observados apenas átomos relacionados por até 3 ligações químicas, o HCCH-TOCSY correlaciona todos os ^1H alifáticos da cadeia lateral de cada resíduo (Figura 2.16). Considerando-se os valores conhecidos de $\text{C}\alpha$, $\text{C}\beta$ e $\text{H}\alpha$ de cada aminoácido, obtidos na etapa de assinalamento sequencial, e utilizando-se os espectros acima descritos, foi possível obter os demais ^1H e ^{13}C da cadeia lateral dos resíduos alifáticos, conforme demonstrado na Figura 2.17 para o aminoácido isoleucina 10 (Ile10) da proteína.

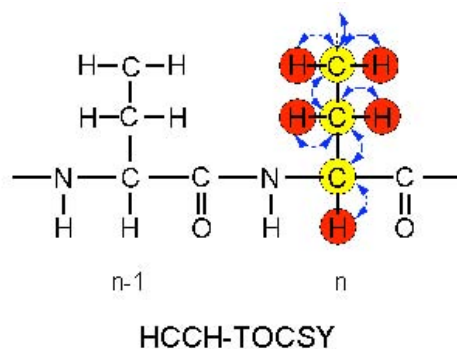


Figura 2.16: Ilustração da transferência de magnetização envolvida no experimento de HCCH-TOCSY. Os átomos em amarelo estão envolvidos na transferência da magnetização e os átomos em vermelho são os núcleos observados no experimento. As setas indicam a direção da transferência da magnetização (figura extraída do site <http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/home.htm>).

Os prótons das amidas da cadeia principal da proteína são facilmente trocados por prótons do solvente, sendo que a rapidez dessa troca depende da acessibilidade do aminoácido ao solvente. Portanto, a taxa de troca pode fornecer informações a respeito das regiões da proteína que estão expostas ou protegidas do solvente e/ou realizando ligações de hidrogênio, que por sua vez, são dados importantes para a determinação da estrutura secundária e terciária da proteína. Como o deutério não é visualizado nos experimentos tradicionais de RMN, utiliza-se água deuterada como solvente na realização de experimentos para análise da taxa

de troca dos prótons amídicos das proteínas (conhecidos por experimentos de troca hidrogênio/deutério ou H/D) e, à medida que ocorre a troca do próton pelo deutério, os picos correspondentes no espectro de ^{15}N -HSQC desaparecem.

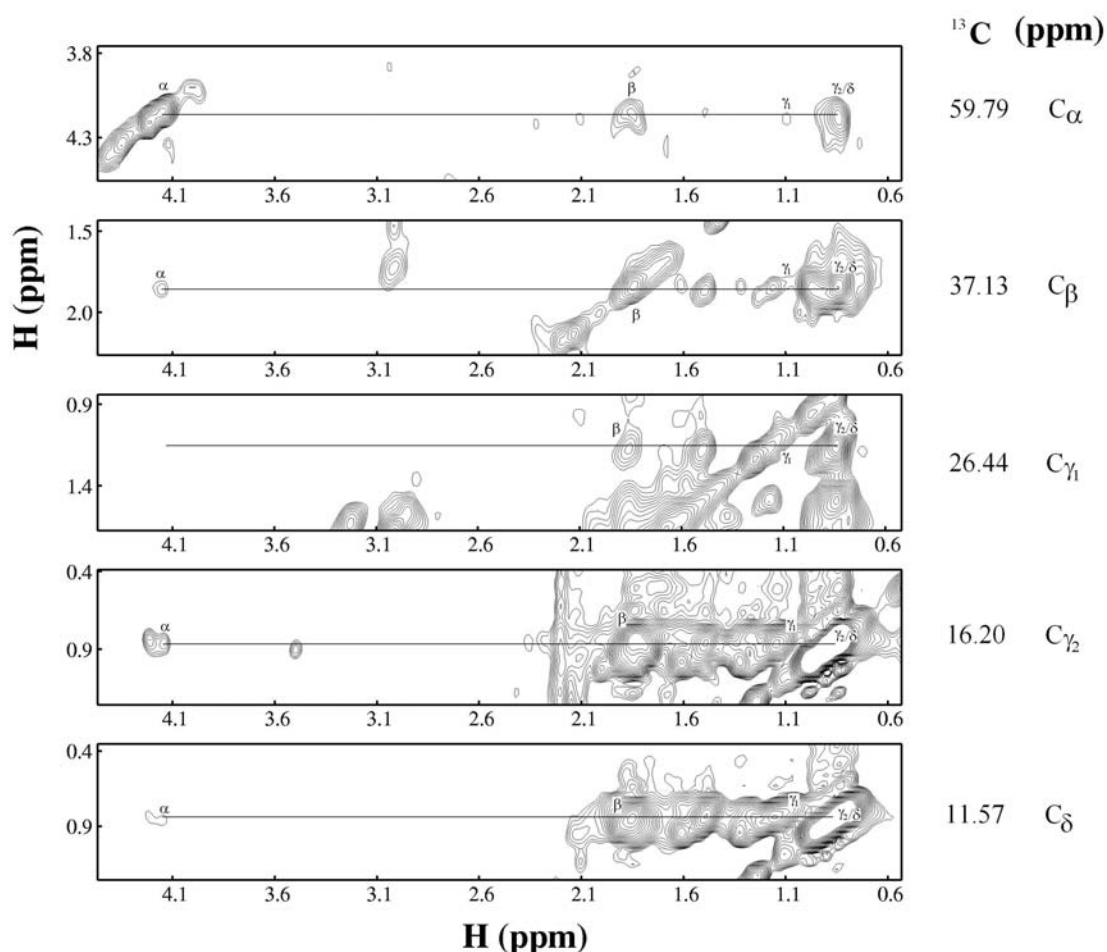


Figura 2.17: Assinalamento da cadeia lateral do aminoácido isoleucina 10 da HsSBDS. Seções do plano ^1H - ^1H do espectro de HCCH-TOCSY, demonstrando os picos correspondentes aos ^1H da cadeia lateral da Ile10. Cada plano do espectro corresponde a uma frequência de ^{13}C , indicada ao lado direito da figura. Os prótons α , β , γ_1 , γ_2 e δ estão indicados nos espectros.

Para a realização dos experimentos de troca H/D com a HsSBDS, a amostra de proteína foi liofilizada, ressuspensa em D_2O e, subsequentemente, foram coletados espectros de ^{15}N -HSQC em determinados intervalos de tempo. Os espectros obtidos após 2 horas e 30 minutos, 15 horas e 158 horas estão demonstrados na Figura 2.18. Os picos correspondentes aos aminoácidos Val18, Phe27, Ile29, Ala30, Val37, Val43, Val58, Leu71, Ser73, Phe75, Thr81, Glu82, Ile83, Cys84, Ile87, Leu88, Ala112, Ile114, Val115, Leu132, Ile133, Glu134, Ala154,

Val157, Leu161, Leu174, Arg175, Phe176, Ile177, Leu178, Leu190, Ile194, Glu207, Ile208, Val209, Cys210, Leu211, Ile212, Ile220, Leu223, Ile 224, Glu235, Phe249 e Glu250 permaneceram presentes no espectro de ^{15}N -HSQC após 2 horas e meia da adição de D_2O , indicando que esses aminoácidos estão envolvidos em algum tipo de interação e/ou protegidos do solvente. Segundo o CSI da proteína, a grande maioria desses resíduos está em regiões que potencialmente apresentam elementos de estrutura secundária, ou seja, realizam ligações de hidrogênio. As excessões correspondem aos resíduos Val43, Val58, Leu132, Ile133, Leu190 e Ile194, ou seja, todos aminoácidos hidrofóbicos que devem estar presentes no “core” hidrofóbico da proteína. Após 158 horas do início do experimento permaneceram no espectro apenas dois sinais, indicando que a estrutura da HsSBDS deve apresentar certa flexibilidade, pois existem proteínas, como por exemplo a proteína do capsídio do bacteriófago filamentosso fd, que apresentam sinais nos espectros de troca H/D mesmo após meses da adição de D_2O (Zeri *et al.*, 2003).

A principal informação estrutural para resolução tridimensional de proteínas em solução através da técnica de RMN encontra-se nas distâncias próton-próton, estimadas através da intensidade dos picos dos espectros contendo o Efeito Nuclear Overhauser, denominado NOE. Este efeito ocorre por acoplamento dipolar entre dois spins através do espaço, ou seja, átomos não-ligados nem perto na sequência de aminoácidos, mas próximos no espaço, são capazes de causar este efeito. Por este motivo, estes espectros fornecem importantes informações estruturais sobre a proteína, indicando os aminoácidos que estão próximos espacialmente e quão próximo seus átomos estão. A intensidade do pico no espectro contendo o Efeito Nuclear Overhauser pode ser classificada como forte, indicando distâncias de 1.8 a 2.7 Å; média, com distâncias de até 3.3 Å; ou fraca, de até 5.0 Å (Evans, 1996). As medidas de distância entre os átomos para a HsSBDS foram obtidas a partir da intensidade dos picos nos espectros de ^{15}N -NOESY-HSQC e ^{13}C -NOESY-HSQC, após calibração utilizando-se sinais de NOE de distâncias conhecidas, como por exemplo, elementos de estrutura secundária. Para atribuição dos NOEs não intra-residuais, foram levados em consideração os dados de CSI calculados para a proteína e a estrutura secundária da homóloga de *A. fulgidus*, considerando-se que existem NOEs característicos para os diferentes tipos de estruturas secundárias, conforme indicado na Figura 2.19.

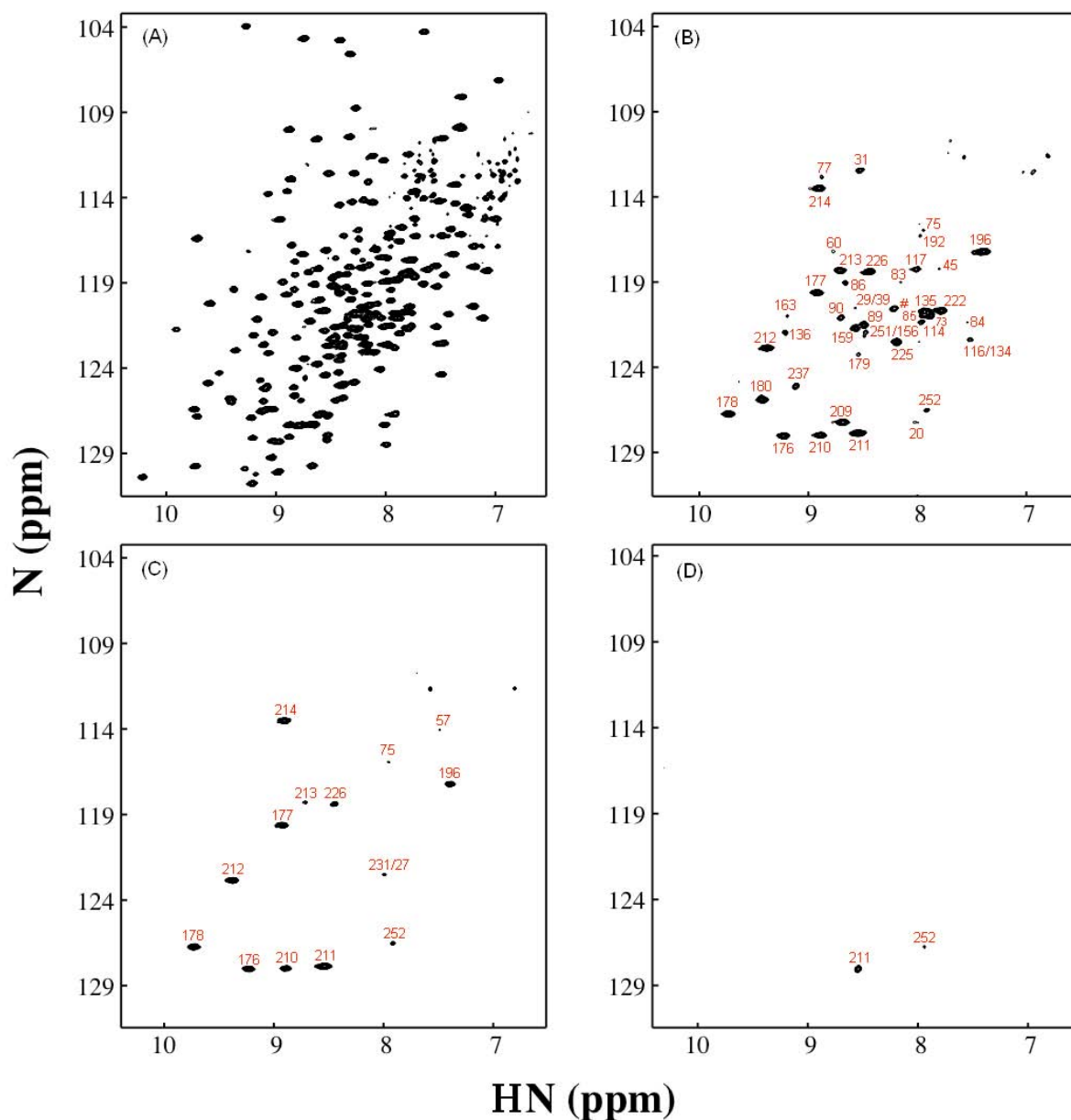


Figura 2.18: Experimentos de troca hidrogênio/deutério (H/D) da HsSBDS. Espectros de ^{15}N -HSQC da HsSBDS (A) na ausência e após (B) 2 horas e 30 minutos, (C) 15 horas e (D) 158 horas da adição de D_2O . Os picos que permaneceram nos espectros estão indicados pela numeração da sequência da proteína, considerando-se os dois aminoácidos iniciais que restam da cauda de histidina.

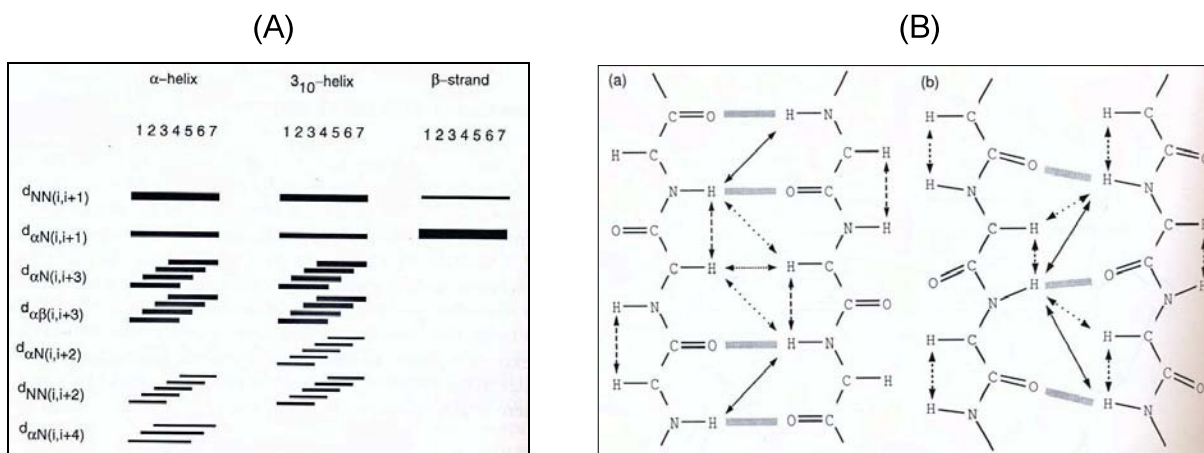


Figura 2.19: NOEs característicos de estruturas secundárias. (A) Padrões de NOEs de curta distância característicos de estruturas secundárias do tipo α -hélice, hélice 3_{10} e folha β , envolvendo prótons amídicos (N), H α (α) e H β (β), onde i representa o resíduo em questão, $i+1$ o resíduo seguinte e assim por diante. A espessura da linha indica a intensidade do NOE. (B) Esquema demonstrando NOEs encontrados em (a) folha β antiparalela e (b) folha β paralela, representados por setas. As faixas cinzas representam ligações de hidrogênio (figura extraída de Roberts, 1995).

Inicialmente, os cálculos estruturais foram realizados para o domínio N-terminal da HsSBDS, utilizando-se o software DYANA (*DYnamics Algorithm for NMR Applications*, versão 1.5) (Güntert *et al.*, 1997), a partir da lista inicial dos NOEs assinalados. A cada ciclo de cálculo eram geradas 20 estruturas e as 10 melhores avaliadas com relação às violações de distância para continuação da atribuição dos NOEs. Durante este processo, foi adquirido pelo LNLS o programa CYANA (*Combined assignment and dYnamics Algorithm for NMR Applications*, versão 2.1) (Güntert, 2004), que realiza o assinalamento semi-automatizado do espectro de NOESY. Utilizando-se a lista inicial de NOEs e considerando-se os dados da estrutura da homóloga de *A. fulgidus* resolvida por cristalografia (código PDB 1T95, Shammass *et al.*, 2005), este programa foi capaz de assinalar um número considerável de NOEs adicionais da proteína inteira, possibilitando iniciar os cálculos da estrutura total da proteína com este mesmo programa. Neste processo, a cada ciclo eram geradas 40 estruturas, que novamente eram avaliadas quanto às violações de distância para continuação da atribuição dos NOEs. Foram realizados vários ciclos de atribuição automática e manual de NOEs seguido de cálculos estruturais e algumas estruturas intermediárias obtidas durante este processo encontram-se na Figura 2.20. Este processo de refinamento é cíclico e exige a inspeção detalhada das violações existentes e dos dados de NOE, sendo necessário

adicionar novos NOEs e retirar NOEs possivelmente incorretos, a fim de eliminar o maior número possível de violações. Após conclusão do refinamento, foi feita a minimização de energia com o programa CNS (*Crystallography & NMR System*) (Brunger *et al.*, 1998), para finalização com as 20 melhores estruturas. Estas foram avaliadas quanto à sua qualidade para validação da estrutura final. A descrição detalhada da estrutura final da proteína HsSBDS encontra-se no Capítulo 3, Artigo 2.

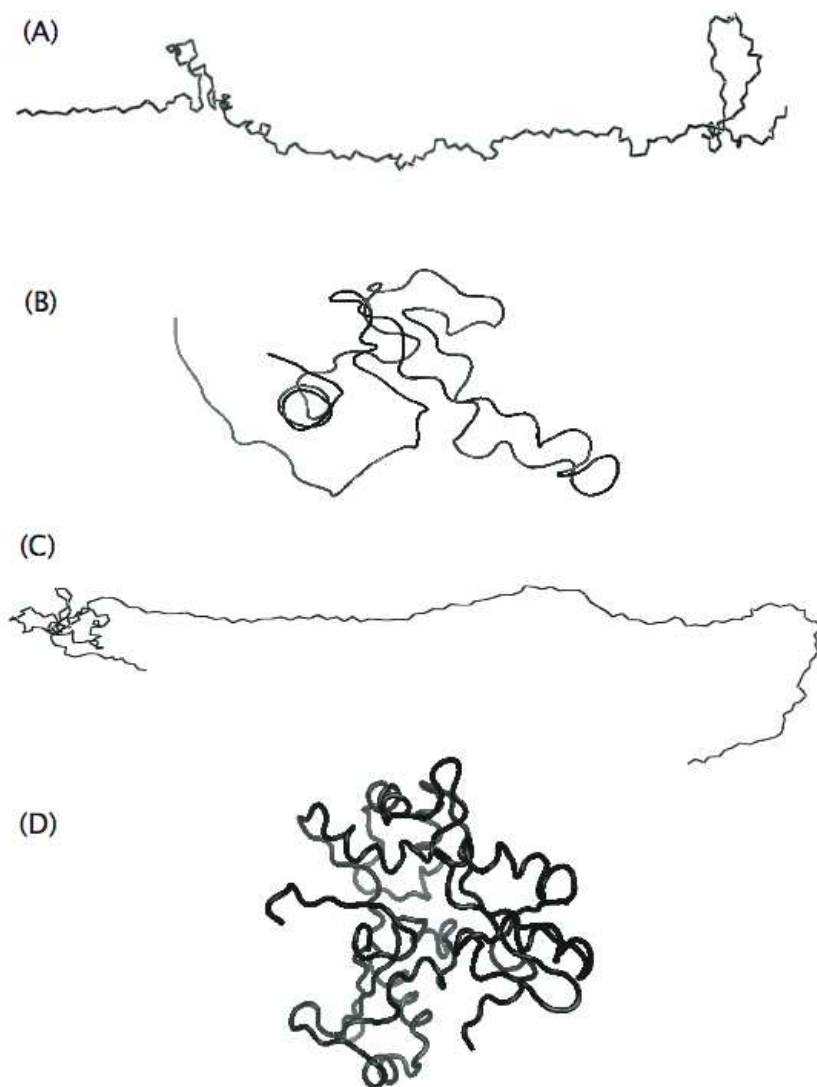


Figura 2.20: Etapas do cálculo da estrutura da proteína HsSBDS. O cálculo estrutural foi realizado a partir das restrições de distâncias obtidas por RMN, utilizando-se os programas DYANA e CYANA. Estrutura obtida (A) do N-terminal da HsSBDS no primeiro ciclo de cálculo com o programa DYANA; (B) do N-terminal da HsSBDS no 11º ciclo do refinamento com DYANA; (C) da proteína inteira no primeiro ciclo com CYANA; (D) da proteína inteira no 15º ciclo com CYANA.

2.3. DISCUSSÃO

Com o intuito de estudar a estrutura tridimensional de proteínas da família SBDS, experimentos de cristalização foram feitos com as ortólogas de *P. abyssi*, *S. cerevisiae*, *T. cruzi* e *H. sapiens*. No início deste trabalho, não havia descrição de nenhuma estrutura dessa família de proteínas, ainda pouco estudada. Cristais grandes da SBDS de *P. abyssi*, PaUPF0023, foram obtidos e submetidos a difração por raios X. Os padrões de difração resultantes apresentaram baixa resolução para estudos estruturais (~ 4 Å), indicando desordem da rede cristalina.

Análises de dicroísmo circular indicaram uma provável predominância de estruturas secundárias do tipo α -hélice na proteína PaUPF0023. A predição de estrutura secundária a partir de programa computacional sugeriu, além da presença de mais de 40% de α -hélice, um alto conteúdo de regiões desordenadas nessa proteína. Provavelmente, a desordem interna encontrada nos cristais é refletida em seu padrão de difração decorrente de pequenas variações na estrutura das moléculas de proteína dentro do cristal, resultantes da mobilidade dessas regiões pouco estruturadas. Essa hipótese é suportada pelo conhecimento posterior das estruturas das ortólogas de *A. fulgidus* (Shammas *et al.*, 2005; Savchenko *et al.*, 2005) e *M. thermotrophicus* (Leong *et al.*, 2009), que apresentam loops e regiões desestruturadas entre os três domínios que compõem essas proteínas e, principalmente, pela presença de duas moléculas de proteína na unidade assimétrica do cristal da mthSBDS com diferentes orientações relativas entre seus domínios, sugerindo que as proteínas dessa família são altamente flexíveis.

Os experimentos de dicroísmo circular realizados com as ortólogas de levedura e humana indicaram um menor conteúdo de α -hélice quando comparadas à PaUPF0023. Essa diminuição de α -hélices, além da presença de aproximadamente 30% de regiões desordenadas em suas estruturas, indicam que essas proteínas devem ser mais flexíveis do que a ortóloga de Archaea. Portanto, a dificuldade em cristalizar as proteínas SBDS de levedura e humana se deve, provavelmente, à alta flexibilidade dessas proteínas.

A alta estabilidade térmica da ortóloga de Archaea, verificada nos ensaios de desnaturação causada por aumento de temperatura, sugere a existência de diferenças estruturais entre esta e as demais proteínas estudadas. Essas diferenças

estruturais, também verificadas nos experimentos de dicroísmo circular, devem ter contribuído para a formação dos cristais da PaUPF0023, justificando o sucesso alcançado na cristalização de apenas uma proteína deste trabalho. Essa idéia é condizente com o fato de ambas as estruturas existentes de proteínas da família SBDS serem de Archaea.

Durante a purificação e o preparo da TcSBDS truncada para experimentos de cristalização, sua precipitação em diferentes condições indicou a instabilidade dessa proteína truncada. No processo de construção da proteína TcSBDS sem o C-terminal desnovelado, foi utilizado como parâmetro as sequências das proteínas ortólogas que não apresentavam o C-terminal estendido, portanto, essa construção da proteína equivale às sequências das ortólogas de Archaea, levedura e humana. Ensaio de proteólise limitada realizados com a TcSBDS (descritos no Capítulo 3) indicam que a proteína truncada é bastante sensível à degradação por protease (Capítulo 3, Artigo I, Figura 4). A comparação dos ensaios de proteólise limitada da TcSBDS truncada com a proteína humana (Figura 2.10) mostra que a primeira produziu um fragmento menor e menos estável do que a região correspondente aos domínios 2 e 3 da proteína humana, identificado como um domínio resistente à proteólise. Possivelmente, a proteína TcSBDS precisaria de seu C-terminal estendido (ou de um fragmento deste) para formar um domínio mais resistente. Ensaio de proteólise limitada realizados com a proteína TcSBDS inteira (Capítulo 3, Artigo I, Figura 4) mostraram a existência de um fragmento de aproximadamente 40 kDa mais estável do que o restante da proteína, porém essa região não resistiu à degradação por mais de 30 minutos, indicando que a proteína inteira deve ser flexível. A falta de um domínio compacto e estável na SBDS de *T. cruzi* impossibilitou a cristalização desta ortóloga.

Embora não tenha sido possível determinar a estrutura por cristalografia de raios X das proteínas da família SBDS estudadas neste trabalho, a técnica de Ressonância Magnética Nuclear permitiu o estudo estrutural da proteína SBDS humana. Esta proteína se demonstrou bastante estável nas condições exigidas pela técnica (pouco sal, temperatura ambiente, alta concentração, etc) e seu espectro de ¹⁵N-HSQC apresentou grande dispersão nos valores de deslocamento químico dos sinais. Tanto a estabilidade da proteína quanto a qualidade de seus espectros possibilitou o trabalho estrutural dessa proteína por RMN.

As duas principais dificuldades encontradas no estudo de proteínas maiores de 20-25 kDa pela técnica de RMN se devem: (1) à sobreposição dos sinais nos espectros, devido a grande quantidade de picos resultante do grande número de átomos na molécula; e (2) à baixa razão sinal/ruído do espectro (ou sensibilidade), como resultado do rápido decaimento do sinal de RMN, devido ao aumento da taxa de relaxação causada pelo lento tombamento de moléculas grandes em solução (Sattler & Simon, 2002). Para contornar esses problemas com relação à proteína HsSBDS, de aproximadamente 30 kDa, foi necessário realizar experimentos de RMN em dois campos magnéticos diferentes, de 600 MHz (com sonda criogênica) e 800 MHz, utilizar a técnica de TROSY (*Transverse Relaxation-Optimised Spectroscopy*) e marcar a proteína com deutério. A utilização de sonda criogênica e o aumento do campo magnético permitem o aumento considerável da resolução espectral. A técnica TROSY (Pervushin *et al.*, 1997) explora a interferência de dois diferentes mecanismos de relaxação (dipolo-dipolo, DD e *chemical shift anisotropy*, CSA) que afetam a largura das linhas dos sinais de RMN, selecionando as linhas mais estreitas e, portanto, possibilitando o aumento da sensibilidade para proteínas grandes. A deuteração da proteína permite o aumento da razão sinal/ruído do espectro de RMN devido ao aumento da supressão da difusão do spin e à diminuição da taxa de relaxação dos spins do ^{13}C e ^{15}N (Sattler & Fesik, 1996).

A análise dos espectros de tripla ressonância da proteína HsSBDS permitiram o assinalamento sequencial da cadeia principal da proteína e de suas cadeias laterais alifáticas e, através do cálculo do CSI, de experimentos de troca H/D e da atribuição inicial dos NOEs, foi possível identificar os elementos de estrutura secundária presentes na proteína. Os cálculos estruturais foram realizados com os programas DYANA e CYANA, utilizando-se dados de NOE assinalados manual e automaticamente. Em um processo cíclico de assinalamento de NOEs e cálculo da estrutura automatizado, passando pela análise das violações de distância presentes nas estruturas obtidas em cada etapa, foram realizados vários ciclos de refinamento que, após minimização de energia, resultou na estrutura da proteína SBDS humana, descrita e analisada em detalhe no Capítulo 3, Artigo2.

CAPÍTULO 3

Análise estrutural e funcional das proteínas SBDS de *Trypanosoma cruzi* (TcSBDS) e *Homo sapiens* (HsSBDS)

Neste capítulo são apresentadas as publicações resultantes deste trabalho de doutorado, que se referem à caracterização estrutural e funcional das proteínas SBDS de *Trypanosoma cruzi* e humana.

O primeiro artigo foi publicado na revista *Biochimie* e contém a caracterização da proteína TcSBDS, cuja sequência encontrava-se incorreta nos bancos de dados de proteínas. Neste trabalho, sua sequência correta foi identificada, depositada no banco de dados *Genbank* e análises estruturais e bioquímicas foram realizadas com essa proteína. Este foi o primeiro estudo realizado com uma ortóloga da SBDS que contém a região C-terminal estendida. Esse trabalho permitiu identificar que esse domínio adicional da SBDS de *T. cruzi* comporta-se como um fragmento de proteína intrinsecamente desenhado e é capaz de realizar interações com RNA.

O segundo artigo, submetido para a revista *Journal of Molecular Biology* (JMB), revela a estrutura da proteína SBDS humana, determinada por Ressonância Magnética Nuclear (RMN). Neste trabalho são analisadas as propriedades dinâmicas dessa proteína e, pela primeira vez, é estudada a interação direta entre essa proteína e RNA. Este estudo possibilitou a confirmação da existência de mobilidade conformacional entre os diferentes domínios da proteína e permitiu o mapeamento da interação da proteína com RNA, verificada na região N-terminal da proteína, onde se encontra a maioria das mutações relacionadas à doença humana SDS.

3.1. ARTIGO I

Characterization of the *Trypanosoma cruzi* ortholog of the SBDS protein reveals an intrinsically disordered extended C-terminal region showing RNA-interacting activity

Juliana F. de Oliveira, Beatriz A. Castilho, Mauricio L. Sforça, Marco Aurelio Krieger, Ana Carolina Zeri, Beatriz G. Guimarães e Nilson I. T. Zanchin



Research paper

Characterization of the *Trypanosoma cruzi* ortholog of the SBDS protein reveals an intrinsically disordered extended C-terminal region showing RNA-interacting activity

Juliana Ferreira de Oliveira^a, Beatriz A. Castilho^b, Mauricio L. Sforça^a, Marco Aurélio Krieger^c, Ana Carolina Zeri^a, Beatriz G. Guimarães^{a,*}, Nilson I.T. Zanchin^{a,*}

^a Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS, Rua Giuseppe Maximo Scolfaro 10000, PO Box 6192, CEP13083-970, Campinas SP, Brazil

^b Department of Microbiology, Immunology and Parasitology, Universidade Federal de São Paulo, Rua Botucatu 862, São Paulo, SP 04023-062, Brazil

^c Instituto de Biologia Molecular do Paraná, IBMP, Curitiba Paraná, Brazil

ARTICLE INFO

Article history:

Received 24 June 2008

Accepted 5 December 2008

Available online 16 December 2008

Keywords:

SBDS ortholog

Trypanosoma cruzi

RNA-protein interaction

Protein structure

ABSTRACT

The human *SBDS* gene and its yeast ortholog *SDO1* encode essential proteins that are involved in ribosome biosynthesis. *SDO1* has been implicated in recycling of the ribosomal biogenesis factor Tif6p from pre-66S particles as well as in translation activation of 60S ribosomes. The SBDS protein is highly conserved, containing approximately 250 amino acid residues in animals, fungi and *Archaea*, while SBDS orthologs of plants and a group of protists contain an extended C-terminal region. In this work, we describe the characterization of the *Trypanosoma cruzi* SBDS ortholog (TcSBDS). TcSBDS co-fractionates with polysomes in sucrose density gradients, which is consistent with a role in ribosome biosynthesis. We show that TcSBDS contains a C-terminal extension of 200 amino acids that displays the features of intrinsically disordered proteins as determined by proteolytic, circular dichroism and NMR analyses. Interestingly, the C-terminal extension is responsible for TcSBDS–RNA interaction activity in electrophoretic mobility shift assays. This finding suggests that Trypanosomatidae and possibly also other organisms containing SBDS with extended C-terminal regions have evolved an additional function for SBDS in ribosome biogenesis.

© 2008 Elsevier Masson SAS. All rights reserved.

1. Introduction

The SBDS gene was described as the locus altered in the majority of patients with the Shwachman–Bodian–Diamond Syndrome (SBDS), which is an autosomal recessive disease characterized by pleiotropic phenotypes including pancreatic insufficiency, skeletal abnormalities and predisposition to develop hematological abnormalities [1,2]. These deficiencies indicate that the SBDS protein plays an essential function. Consistently with this, knock out of the yeast ortholog (*SDO1*) and double knock out of the murine ortholog lead to lethality [3,4]. However, SBDS deficiency may have different implications depending on the cell type. Cells from SBDS patients and cells knocked down for SBDS are hypersensitive to actinomycin D and show reduced ribosomal RNA synthesis but do not show a discrete defect in pre-rRNA processing [5]. Knock down of SBDS in

HEK293 cells did not cause a detectable defect in pre-ribosomal RNA processing but affected gene expression at both the transcriptional and translational levels [6]. The human SBDS is a ubiquitously expressed nucleolar protein [7] and interacts with the ribosome biogenesis factor Nip7 [6]. Its yeast ortholog Sdo1p was initially found in complexes containing ribosomal RNA (rRNA) processing factors and 60S ribosomal proteins [3,8]. Subsequently, Sdo1p was shown to participate in the biogenesis of the 60S ribosome and to function in the release of the ribosomal biogenesis factor Tif6p from pre-66S particles as well as on translation activation of ribosomes [9]. Solution of the *Archaeoglobus fulgidus* SBDS ortholog structure revealed a three-domain architecture with the C-terminal domain showing closest structural similarity to the domain V of the yeast translation elongation factor 2 [3,10]. This finding, together with studies based on genetic interactions between various *SDO1*, *TIF6* and *EFL1* (elongation factor-like 1) mutant and null alleles [9] has provided further insights into the mechanism of Sdo1p function in the final steps of 60S ribosome maturation, which suggest that Sdo1p may bind to the pre-60S GTPase center and induce conformational changes that lead to dissociation of Tif6p from pre-60S particles [9,10].

* Corresponding authors. Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS, Rua Giuseppe Máximo Scolfaro 10000, PO Box 6192, CEP13083-970, Campinas SP, Brazil. Tel.: +55 19 3512 1113; fax: +55 19 3512 1004.

E-mail addresses: beatriz@lnls.br (B.G. Guimarães), zanchin@lnls.br (N.I.T. Zanchin).

Based on sequence analysis, it is possible to identify two major groups of SBDS orthologs. One group includes Archaea, animals and fungi which contain SBDS proteins of approximately 250 amino acid residues, while the second group that includes plants and protists (Fig. 1) contains SBDS orthologs possessing extended C-terminal regions with ~100–280 additional residues [11]. Similar three-dimensional structures are expected for the short SBDS orthologs and for the corresponding N-terminal region of the long SBDS orthologs. The extended C-terminal of protist and plant SBDS orthologs suggests that in these organisms SBDS plays additional roles. In fact, SBDS orthologs from a group of plants and protists contain a putative RNA-interacting motif (C2H2 zinc finger) in the extended C-terminal region indicating that it might mediate interaction with RNA targets [11]. However, the Trypanosomatidae SBDS orthologs do not contain this conserved zinc finger domain in the extended C-terminal region, raising new questions about what would be its role in the cell. In order to obtain information on the structure and function of the SBDS from Trypanosomatidae, we cloned the SBDS ortholog from *Trypanosoma cruzi* (TcSBDS) and characterized the recombinant protein. We show that the C-terminal region of TcSBDS behaves as a natively unfolded protein segment and mediates its interaction with RNA.

2. Materials and methods

2.1. Cloning of the *T. cruzi* SBDS ortholog and plasmid constructs

The coding sequence of the *T. cruzi* SBDS ortholog was amplified by PCR from genomic DNA isolated from the strain DM28C. The PCR product obtained using oligonucleotides TcSBDSF (5' ATG CCA TAT GAG TGC CCG TAT TCA GGT TCC 3') and ONZ398 (5' CAG GGA TCC TCA GTC TCG TAC TTC CCC TGC 3') amplified a DNA fragment encoding the full-length protein, the PCR product obtained using oligonucleotides TcSBDSF and ONZ428 (5' GGA TCC TCA AAC ATC ACC TTC GCC CAC ATC 3') encodes a truncated protein comprising residues 1–263 and the PCR product obtained using oligonucleotides ONZ429 (5' GCA TCA TAT GGA CGC CGA GCT TAT TGC GC 3') and ONZ398 encodes a truncated protein comprising residues 264–464. The PCR products were cloned into the pGEM-T PCR product cloning system (Promega) and submitted to DNA sequencing analysis. Subsequently, the PCR products were transferred from pGEM-T intermediate vectors to the expression vector pET-TEV [12] using the NdeI and BamHI restriction sites, generating plasmids pET-TcSBDS; pET-N-TcSBDS, and pET-C-TcSBDS. The cDNA of TcSBDS was synthesized using total *T. cruzi* DM28C RNA in reverse transcriptase reactions with an oligonucleotide RDEdT (oligo-dT₁₅ containing a BamHI restriction site) and the AMV Reverse Transcriptase (Promega). The cDNA products were used as templates for PCR reactions to amplify the full-length TcSBDS coding sequence, which was cloned into the pGEM-T cloning system (Promega) and submitted to DNA sequencing analysis.

2.2. Expression and purification of recombinant proteins

Histidine-tagged recombinant proteins were expressed in *Escherichia coli* BL21(DE3)*slyD*[−] transformed with plasmids pET-TcSBDS, pET-N-TcSBDS or pET-C-TcSBDS. *E. coli* cells were incubated in LB medium containing kanamycin (50 µg/mL) at 37 °C up to an OD₆₀₀ of ~1.0, when the cultures were shifted to 25 °C and induced with 0.5 mM IPTG for 2–4 h. In the case of N-TcSBDS, the induction was performed at 16 °C during 16–18 h. Following induction, the cells were harvested by centrifugation (5000g) for 10 min at 4 °C and suspended in buffer A (50 mM Tris-HCl pH 7.2, 500 mM NaCl, 1 mM PMSF). The cells were lysed by lysozyme treatment (50 µg/mL for 30 min on ice), followed by sonication and

the extract isolated by centrifugation (23,000g) for 30 min at 4 °C. Purification of the recombinant proteins was carried out by affinity chromatography on a 5 mL HiTrap Chelating HP column (GE Healthcare) pre-equilibrated in buffer A and 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 chromatography was performed on an ÄKTA-FPLC system (GE Healthcare), with a flow rate of 1 mL/min and fractions of 1.8 mL were collected. In the case of N-TcSBDS, the affinity chromatography was performed on a 1 mL HiTrap Chelating HP column (GE Healthcare) pre-equilibrated in buffer A and eluted with 50 mL gradient of 0–100% buffer B. C-TcSBDS was purified further by ion exchange chromatography on a 1 mL HiTrap Q-Sepharose HP column equilibrated with 50 mM Tris-HCl pH 8.5 and eluted with a 0–50% gradient of buffer containing 50 mM Tris-HCl pH 8.5 and 1 M NaCl in 35 mL. The chromatography was performed at a flow rate of 1 mL/min and fractions of 1.2 mL were collected.

2.3. Isotopic labeling of C-TcSBDS

E. coli BL21(DE3)*slyD*[−] cells carrying plasmid pET-C-TcSBDS were incubated in modified M9 minimal medium containing ¹⁵NH₄Cl as described by Li et al. [13]. Expression and purification of ¹⁵N-labeled C-TcSBDS were performed using the same procedure as described above for unlabeled C-TcSBDS. The histidine-tag of C-TcSBDS was removed by digestion with 7% w/w TEV protease at 25 °C for 24 h in buffer containing 50 mM Tris-HCl pH 8.0, 0.05 mM EDTA; 1 mM DTT and 50 mM NaCl. Subsequently, the reaction was subjected to a chromatography on a 1 mL HiTrap Chelating HP column (GE Healthcare) to remove the digested histidine tag.

2.4. Circular dichroism (CD) and prediction of unstructured regions

Far UV CD spectra were recorded on a Jasco-810 spectropolarimeter using a Peltier system PFD 425S for temperature control. CD spectra are average of five scans acquired at 20 °C using 1- or 2-mm path length cells at 0.5-nm intervals over the wave length range from 190 to 260 nm. Ellipticity is reported as the mean residual ellipticity $[\theta]$ (deg cm² dmol^{−1})10³. The TcSBDS and C-TcSBDS protein concentrations were 5 and 10 µM in 10 mM sodium phosphate buffer pH 7.2, respectively, and N-TcSBDS was at 5 µM in sodium acetate buffer pH 5.2 containing 20 mM NaCl. For the analysis of C-TcSBDS interaction with RNA, 5 µM of both C-TcSBDS and of an RNA poly-A₍₂₅₎ were incubated for 45 min at 20 °C and the CD spectra acquired as described above. C-TcSBDS at 5 µM was also analyzed by CD in the presence of increasing concentrations of 2,2,2-trifluoroethanol (TFE) (0, 5, 10, 20 and 30%) following 15 min incubation at 25 °C. Unstructured regions were predicted using PONDR (www.pondr.com, with VL-XT predictor) [14]. Mean net charge versus mean hydrophobicity was plotted as described by Uversky et al. [15].

2.5. Limited proteolysis assays

For limited proteolysis experiments, protein samples at 0.5 mg/mL in 50 mM sodium phosphate pH 7.2 and 20 mM NaCl were digested with 0.75 µg trypsin in 150 µL reaction volumes. At various times, samples were removed and the reaction stopped by adding 1 mM PMSF. The proteolysis products were analyzed on 12.5% SDS-PAGE gels and stained with coomassie blue.

2.6. ¹⁵N-HSQC NMR spectra acquisition

¹⁵N-HSQC NMR spectra were acquired on a Varian Inova 600 MHz NMR spectrometer, for a 0.72 mM recombinant [¹⁵N]C-TcSBDS

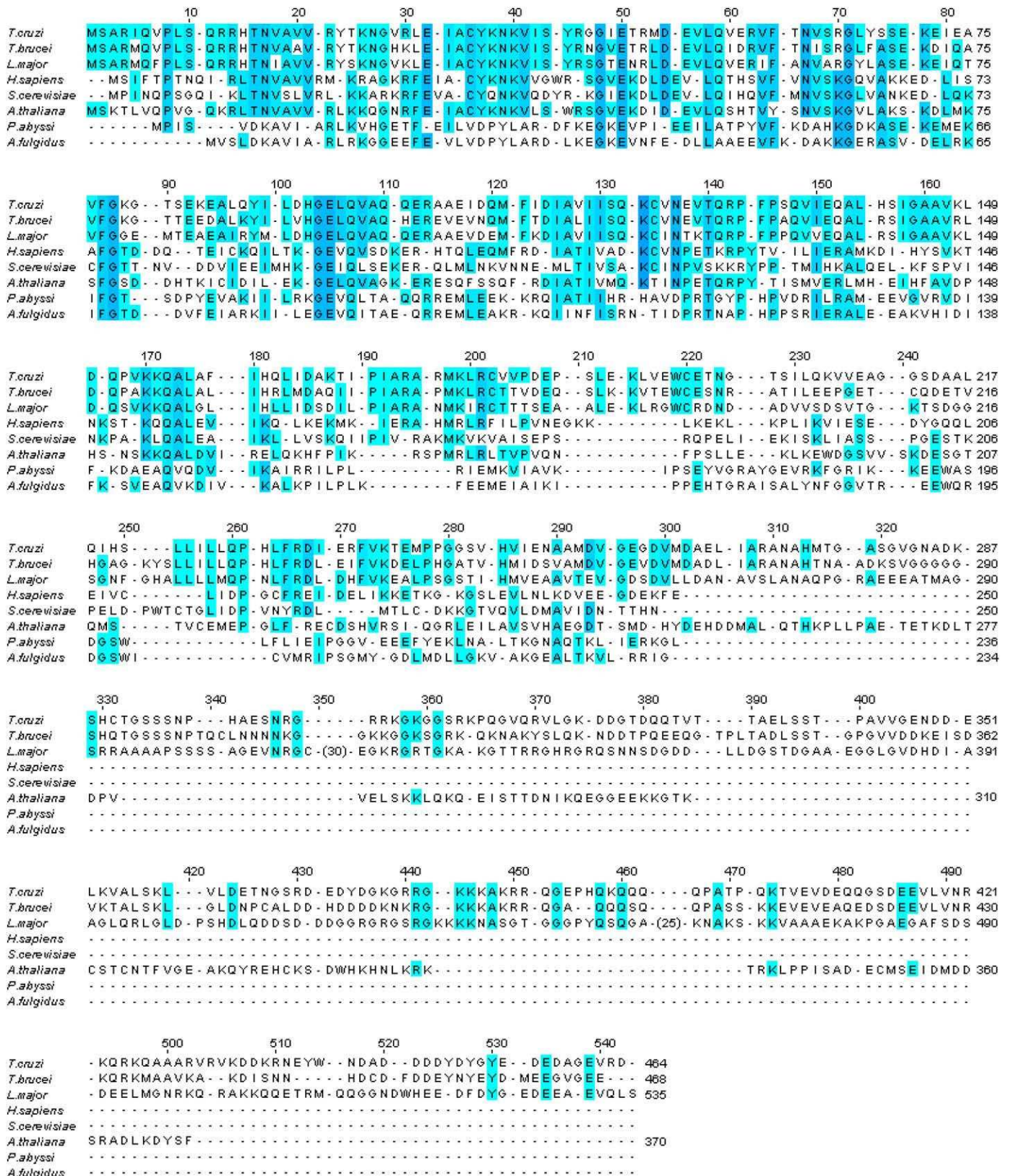


Fig. 1. Amino acid sequence alignment of SBDS orthologs. The sequences shown are from trypanosomatids (*T. cruzi*, *T. brucei* and *L. major*), human (*H. sapiens*), yeast (*S. cerevisiae*), plant (*A. thaliana*) and Archaea (*P. abyssi* and *A. fulgidus*). Conserved residues are highlighted.

sample in 50 mM sodium phosphate buffer pH 6.0 with 250 mM NaCl and 5% (v/v) D₂O, at 20 °C. Eight scans were acquired with a window of 8000 Hz and 1024 points for ¹H, and 2430 Hz and 256 points for the ¹⁵N dimension.

2.7. Electrophoretic mobility shift assays (EMSA)

20 pmol of the oligoribonucleotides or oligodeoxyribonucleotides used in these experiments were labeled with ³²P using 1 U of T4 PNK (Fermentas) and 20 μCi of [γ-³²P]-ATP. 0.24 pmol of [³²P]-labeled oligoribonucleotides or oligodeoxyribonucleotides were incubated in 20 μl reactions with different amounts of the indicated proteins in buffer containing 20 mM Tris-HCl pH 8.0; 5 mM Mg(OAc)₂; 150 mM KOAc; 0.2% Triton X-100; 1 mM DTT; 1 mM PMSF and 1 U RNaseOUT (Invitrogen). Subsequently, 2 μl of sample buffer (30 mM Tris-HCl pH 7.5; 40% sucrose; 0.2% w/v bromophenol blue) were added to the reactions. Complexes were resolved on 6% polyacrylamide gels using TBE buffer pH 8.0 for electrophoresis and visualized on a Storm 820 Scanner (Molecular Dynamics). The oligoribonucleotides used in these assays were an RNA poly-A₍₂₅₎ and RNA NZ5 (5' UUG AAU CAA CAG GUU CU 3') and the oligodeoxyribonucleotides used were a DNA poly-A₍₂₀₎ and a random DNA sequence (5' TAA TAC GAC TCA CTA TAG GG 3').

2.8. Sucrose density fractionation of polysomes

T. cruzi cell extracts for polysome analysis were prepared essentially as described for *Trypanosoma brucei* by Brecht and Parsons [16] except that cycloheximide was omitted and the cells were fixed with formaldehyde as described in Nielsen and co-workers [17]. Briefly, 3 × 10⁹ epimastigotes (*T. cruzi* Y strain) growing in 300 mL liver infusion tryptose medium containing 10% (v/v) fetal bovine serum at 28 °C were collected by centrifugation, washed in cold sodium phosphate buffer (PBS) pH 7.4 and suspended in PBS containing 1% v/v formaldehyde. The suspension was incubated on ice for 1 h, the cells harvested by centrifugation, washed in PBS and suspended in 750 μL of buffer containing 10 mM Tris-HCl pH 7.4, 300 mM KCl, 10 mM MgCl₂, and 1 μM DTT. Subsequently, 250 μL of buffer containing 10 mM Tris-HCl pH 7.4, 300 mM KCl, 10 mM MgCl₂, 1 μM DTT, 1.2% v/v Triton X-100 and 0.2 M sucrose were added to the cell suspension. Following lysis, the extract was centrifuged for 1 min at 18,000g and the supernatant was collected. 20 OD₂₆₀ units were loaded onto 7–47% sucrose gradients prepared in 10 mM Tris-HCl pH 7.4, 300 mM KCl, 10 mM MgCl₂. The gradients were centrifuged at 39,000 rpm on an SW41 rotor for 2 h and fractionated by using a Beckman gradient collector and a Bio-Rad fractionation system.

2.9. Co-immunoprecipitation assays and quantitative PCR analysis

Co-immunoprecipitation assays were performed with rabbit antisera raised for TcSBDS and the *T. cruzi* TCTP ortholog, which is not related to pre-rRNA processing and ribosome biogenesis as a negative control. 1 mL of each antiserum was incubated with 20 μL of Protein A-Sepharose beads (GE healthcare) for 3 h at 4 °C. The beads were washed three times with 1 mL of buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 25 mM EDTA and 0.5% (v/v) Nonidet P-40. Subsequently, the Protein A-Sepharose beads were incubated for 16 h at 4 °C with the 60S–80S ribosome fractions obtained by sucrose density gradient fractionation. Following washing using the conditions described above, the Protein A-Sepharose beads were divided into two fractions, which were used for RNA extraction and immunoblot with antisera for TcSBDS and TcRPL26.

RNA was extracted from co-immunoprecipitation reactions with the Trizol reagent (Invitrogen) and used for quantitative PCR analysis. The cDNAs were synthesized in reverse transcriptase reactions with oligonucleotides complementary to specific regions of the 5.8S and 28S rRNAs and of the internal transcribed spacer 2 (ITS2) and the M-MLV Reverse Transcriptase (Invitrogen), according to the supplier's instructions. 2 μL of each cDNA reaction were used in 25 μL quantitative PCR reactions containing 12.5 μL of the SYBR® GREEN PCR Master Mix (Applied Biosystems) and 12.5 pmol of the following

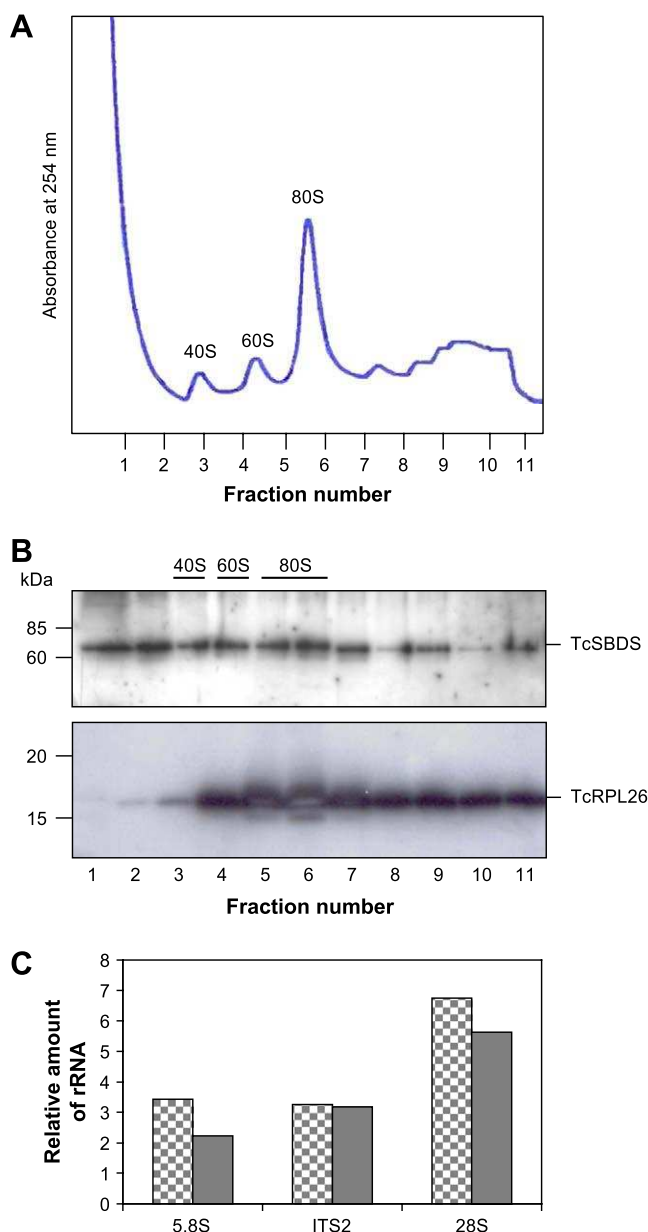


Fig. 2. Analysis of TcSBDS association to polysomes. *T. cruzi* polysomes were fractionated on sucrose density gradients and the fractions analyzed by immunoblotting to detect TcSBDS and TcRPL6. A, polysome profile as determined by measuring the absorbance at 254 nm. B, Immunoblot of the fractions indicated in A, showing TcSBDS (upper panel) and TcRPL26 (lower panel). C, Analysis of rRNAs that co-immunoprecipitate with TcSBDS. The graph shows the amount of 5.8S and 28S rRNAs and internal transcribed spacer 2 (ITS2) co-immunoprecipitated using an antiserum for TcSBDS relative to the amount of the same rRNAs and ITS2 co-immunoprecipitated using an antiserum for TcTCTP. Two independent co-immunoprecipitation experiments were performed and the results are represented by each bar color.

oligonucleotides: ONZ712 (5' GTG TCG CGA TGG ATG ACT TG 3') and ONZ713 (5' TGA TAC CAC TTA TCG CAC TTT GCT 3') for 5.8S rRNA quantitation; ONZ714 (5' ACA CAA CAA CAA CAC AAA ACT ACA AA 3') and ONZ715 (5' CCA CAC GCC ACT TTT TTT CTT T 3') for ITS2 quantitation and; ONZ716 (5' GCC CCC ACC CGT CTT G 3') and ONZ717 (5' CCC TTG CGC GTC TGT TTG 3') for the 28S rRNA quantitation. The reactions were performed in duplicates on a 7500 Real Time PCR System (Applied Biosystems) and the average value of the cycle threshold (Ct) for the samples co-immunoprecipitated with TcSBDS and TcTCTP was calculated. The relative amount of 5.8S and 28S rRNAs and ITS2 co-immunoprecipitated with TcSBDS was determined using the equation: $2^{-(\Delta Ct)}$, where $\Delta Ct = Ct_{TcSBDS} - Ct_{TcTCTP}$.

2.10. Production of rabbit antisera for TcSBDS and TcTCTP and immunoblot analysis

Rabbit antisera for TcSBDS and TcTCTP were generated by injecting rabbits with 100 µg of recombinant protein suspended in buffer containing 50 mM sodium acetate pH 5.2, 20 mM NaCl and 10 mM β-mercaptoethanol as described previously [18]. The antiserum for *T. cruzi* RPL26 was kindly provided by Samuel Goldenberg (IBMP/FIOCRUZ, Curitiba, Brazil). Samples (1 mL each) from the sucrose density gradient were precipitated with 1:1 volume of acetone, dried and suspended in SDS-PAGE sample buffer. The fractions were resolved by SDS-PAGE and transferred

to a PVDF membrane at 300 mA for 1 h in buffer containing 25 mM Tris-base, 200 mM glycine and 20% methanol. Subsequently, the membrane was blocked with TBST buffer (20 mM Tris-HCl pH 7.5; 150 mM NaCl; 0.05% Tween 20) containing 5% low fat milk. Membranes were incubated with a rabbit antiserum raised against TcSBDS, or with a mouse antiserum against TcRPL26 (kindly donated by Dr. Samuel Goldenberg) for 16 h, washed with TBST and incubated for 4 h with an anti rabbit IgG-HRP or anti-mouse IgG-HRP (GE Healthcare). The membranes were developed using the ECL Western Blotting Analysis System (GE Healthcare).

2.11. Amino acid sequence alignment

Multiple amino acid sequences were aligned using CLUSTALW [19] with default parameters. The NCBI accession numbers of the SBDS orthologs sequences are: NP 057122 (*Homo sapiens*), NP 013122 (*Saccharomyces cerevisiae*), NP 126299 (*Pyrococcus abyssi*), NP 069327 (*A. fulgidus*), XP 001686204 (*Leishmania major*), XP 844464 (*T. brucei*), NP 564488 (*Arabidopsis thaliana*). The amino acid sequence of the *T. cruzi* SBDS ortholog, TcSBDS, was obtained by cloning and sequencing the TcSBDS DNA as described above. The TcSBDS nucleotide sequence was deposited at the GenBank with the accession number EU715774.

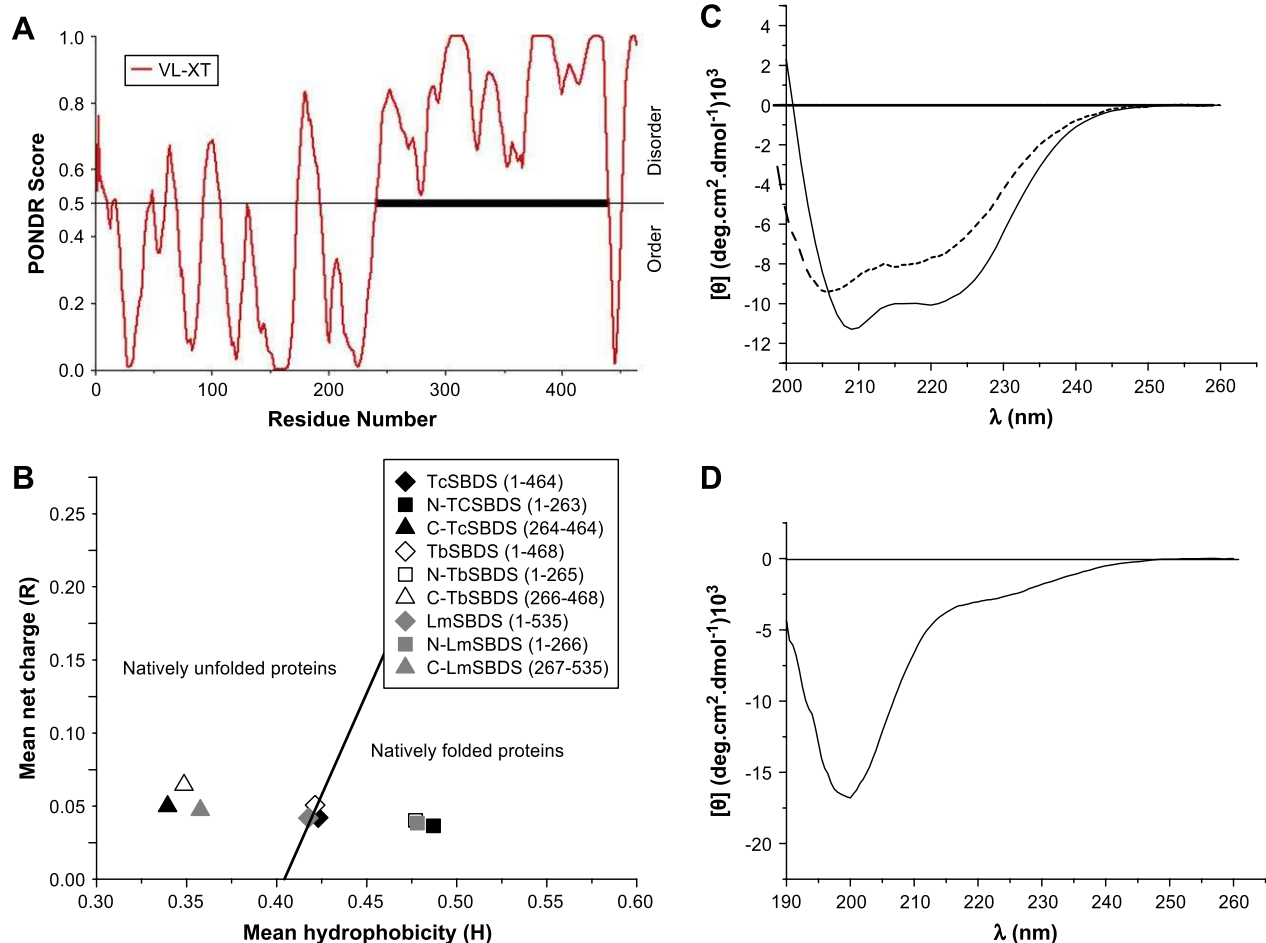


Fig. 3. Analysis of secondary structure elements. A, Graph showing the PONDR score for predicted disordered regions in TcSBDS. B, Graph showing the mean net charge (R) and mean hydrophobicity (H) ratio of SBDS from *T. cruzi*, *T. brucei* and *L. major*. The R/H ratios were plotted for full-length proteins and for N- and C-terminal regions, as indicated. The continuous black line ($R = 2.785H - 1.151$) represents the boundary between natively unfolded and folded proteins determined by Uversky et al. [15]. C, Circular dichroism analysis of TcSBDS full-length protein (dotted line) and N-terminal domain (solid line). D, Circular dichroism analysis of the C-terminal domain of TcSBDS.

3. Results

3.1. Cloning of the *T. cruzi* SBDS ortholog

The *L. major* (XP 001686204) and *T. brucei* (XP 844464) SBDS orthologs contain 535 and 468 amino acid residues, respectively. The sequence information annotated as *T. cruzi* SBDS orthologs (TcSBDS) available at the beginning of this work included two partial protein sequences deposited at the Genbank, comprising 114 (XP 817528) and 345 amino acids (XP 812524), respectively, and the locus Tc00.1047053506435.230 of the TIGR database (www.tigr.org/tdb/e2k1/tca1/tca1.shtml) that encodes a 252-residue protein. In order to define precisely the nucleotide sequence of TcSBDS, we designed primers initially based on the sequence of the locus Tc00.1047053506435.230. This sequence contains four guanosine nucleotides at positions 732–735 from the AUG translation start site. However, the fragment amplified in our laboratory using DNA template from the *T. cruzi* strain DM28C and the longer GenBank clone (XP 812524) contain only three guanosine nucleotides in this region, which indicates a different reading frame downstream from the codon 244. By combining the information from the three sequences mentioned above, we designed new primers to amplify a 1398 base pairs fragment by PCR using *T. cruzi* DM28C DNA as template. Sequencing analysis of the resulting PCR product revealed a 464 amino acid open reading frame. The DNA sequence of the TcSBDS clone used in this work was deposited at the Genbank with the accession number EU715774. The same TcSBDS coding sequence was amplified by RT-PCR by using total *T. cruzi* RNA as template, which confirmed that the TcSBDS is a 464-residue protein. An alignment of the TcSBDS 464-residue protein (Fig. 1) reveals a high similarity to its orthologs from other trypanosomatids in the extended C-terminal region and with other member of the SBDS family in the N-terminal region.

3.2. Analysis of TcSBDS association to polysomes and rRNA

Association of proteins to pre-ribosomal particles is an indication of their involvement with ribosome biogenesis. A TAP-Sdo1p fusion protein was found to co-fractionate with polysomes in sucrose density gradients with the bulk of TAP-Sdo1p found in the region of 60S ribosomes [9]. The human SBDS was found in the soluble fraction and in the fractions containing 60S ribosomes [5]. As shown in Fig. 2, TcSBDS was detected by immunoblotting in all fractions of the polysomal gradient but a large proportion of TcSBDS was found in the fractions containing 60S and 80S ribosomes and polysomes. TcRPL26 was used as a marker for 60S and 80S ribosome sedimentation (Fig. 2B). Thus, TcSBDS associates with ribosomes as do the other characterized orthologs. Association of TcSBDS to ribosomal RNA was also analyzed by co-immunoprecipitation assays. In these assays, 60S–80S ribosome fractions from sucrose density gradients were incubated with antibodies for TcSBDS immobilized on Protein A–Sepharose beads and the amount of rRNA or pre-rRNA co-immunoprecipitated was determined using quantitative PCR. This analysis revealed that TcSBDS co-immunoprecipitated higher amounts of the 5.8S and 28S rRNAs (Fig. 2C) relative to the co-immunoprecipitation with the control antiserum for TcTCTP. The ITS2 detected by quantitative PCR suggests that TcSBDS associates also to pre-ribosomal particles (Fig. 2C).

3.3. Analysis of TcSBDS secondary structure content

In order to provide insights into the functional role of the C-terminal extension found in the trypanosomatids' SBDS, we initially analyzed structural features of this segment. Theoretical predictions failed to identify either conserved domains or secondary

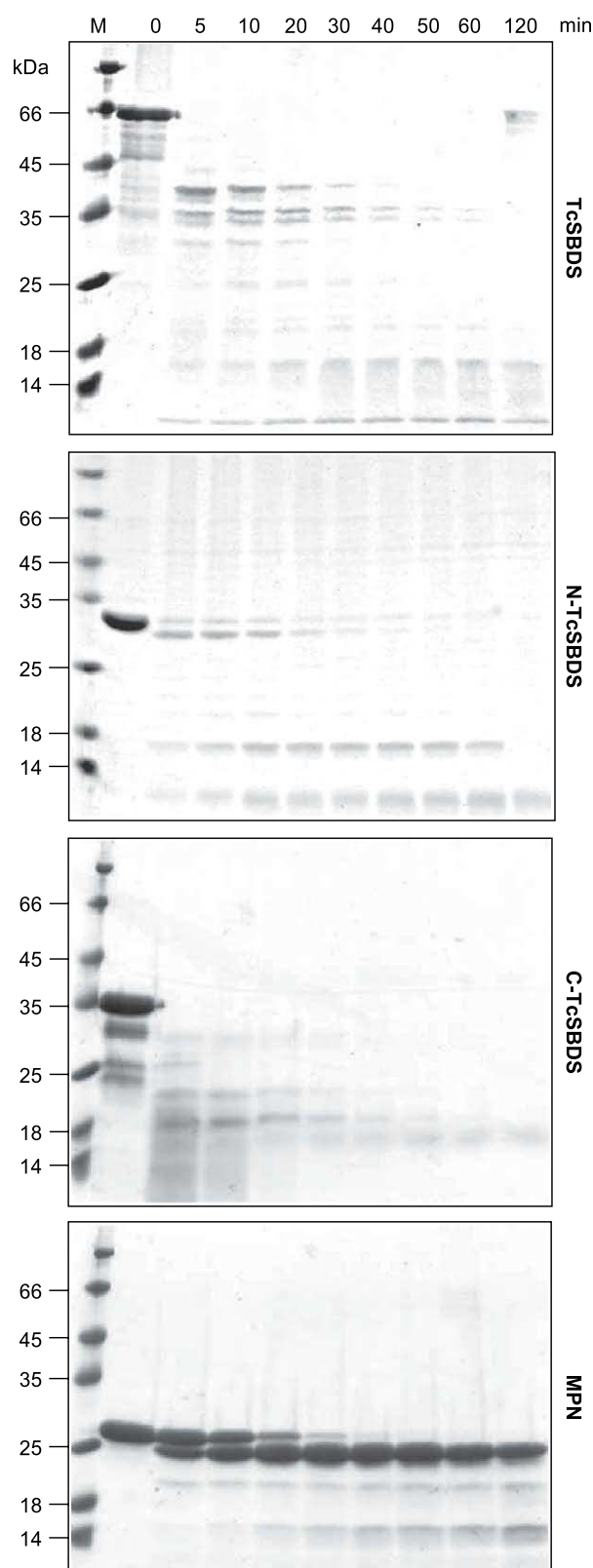


Fig. 4. Limited proteolysis analysis. Coomassie-stained SDS-PAGE gels showing limited proteolysis products. Reactions were prepared with a 1:100 (w:w) protein:trypsin ratio and aliquots were removed at various times, indicated above each lane. TcSBDS, full-length TcSBDS; N-TcSBDS, N-terminal domain of TcSBDS; C-TcSBDS, C-terminal domain of TcSBDS; MPN, MPN domain of the human Mov34 ortholog used as a control for trypsin proteolysis resistance (lower panel). Molecular weight markers are indicated on the left side of each panel.

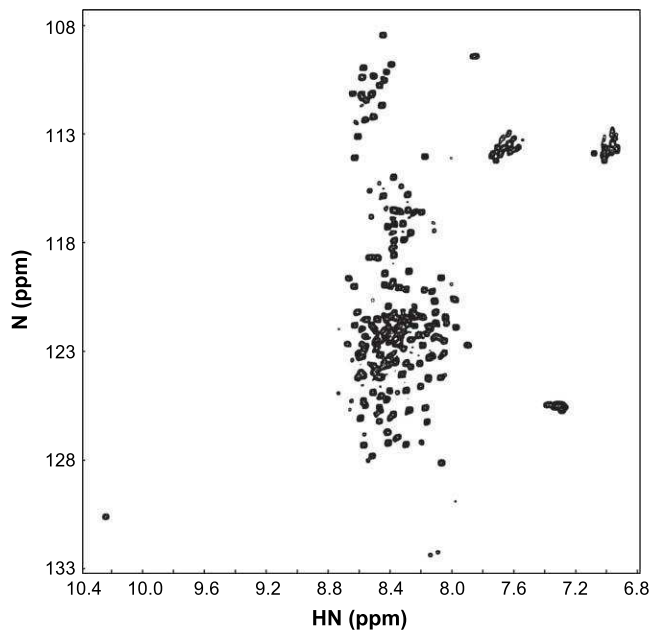


Fig. 5. Conformational analysis of the C-terminal domain of TcSBDS. The ^1H - ^{15}N -HSQC NMR spectrum shows the majority of the peaks clustered in between 8.0 ppm and 8.8 ppm in the ^1H dimension, indicating an unfolded protein.

structure elements in the extended C-terminal region of TcSBDS, indicating that it might be an intrinsically disordered segment [14] (Fig. 3A). Natively unfolded proteins have been described to possess a high ratio of charged residues and a low ratio of hydrophobic residues [15]. The mean net charge versus mean hydrophobicity ratio was plotted for TcSBDS and the SBDS orthologs from *L. major*

(LmSBDS) and *T. brucei* (TbSBDS). This analysis showed that the N-terminal region comprising the three-domain structure similar to SBDS orthologs from Archaea falls in the phase space of natively folded proteins whereas the C-terminal region falls in the phase space of natively unfolded proteins and the full-length protein shows an intermediate location in the plot (Fig. 3B).

In order to characterize the structure of TcSBDS, both the full-length protein and two truncated forms containing either the N-terminal (N-TcSBDS, residues 1–263) or the C-terminal region (C-TcSBDS, residues 264–464) were expressed in *E. coli* and purified. Circular dichroism analysis revealed a higher secondary structure content for the N-terminal domain as compared to the full-length protein (Fig. 3C). Consistent with this finding, the CD spectrum of the C-terminal region contains a negative peak near the 200 nm region, which is representative of unfolded proteins (Fig. 3D). High susceptibility to proteolysis is also an indication of unfolded protein regions. Therefore, the recombinant full-length TcSBDS, N-TcSBDS and the C-terminal region C-TcSBDS were submitted to limited proteolysis. As expected, C-TcSBDS was readily digested by trypsin (Fig. 4). Although TcSBDS and N-TcSBDS were more resistant than C-TcSBDS, they are also very sensitive to proteolysis if compared to the compact protease resistant MPN domain of human Mov34 [20] (Fig. 4). The conformation of the C-terminal domain was also analyzed qualitatively by NMR. As shown in Fig. 5, the ^{15}N -HSQC NMR spectrum revealed that the majority of the resonance peaks are clustered between 8.0 ppm and 8.8 ppm in the ^1H dimension, and this low dispersion is characteristic of unfolded structures [21].

3.4. Analysis of TcSBDS interaction with RNA

SBDS functions most probably as an adaptor protein interacting with RNA in the context of ribosome biogenesis and ribosome activation for translation. Based on this hypothesis, we decided to

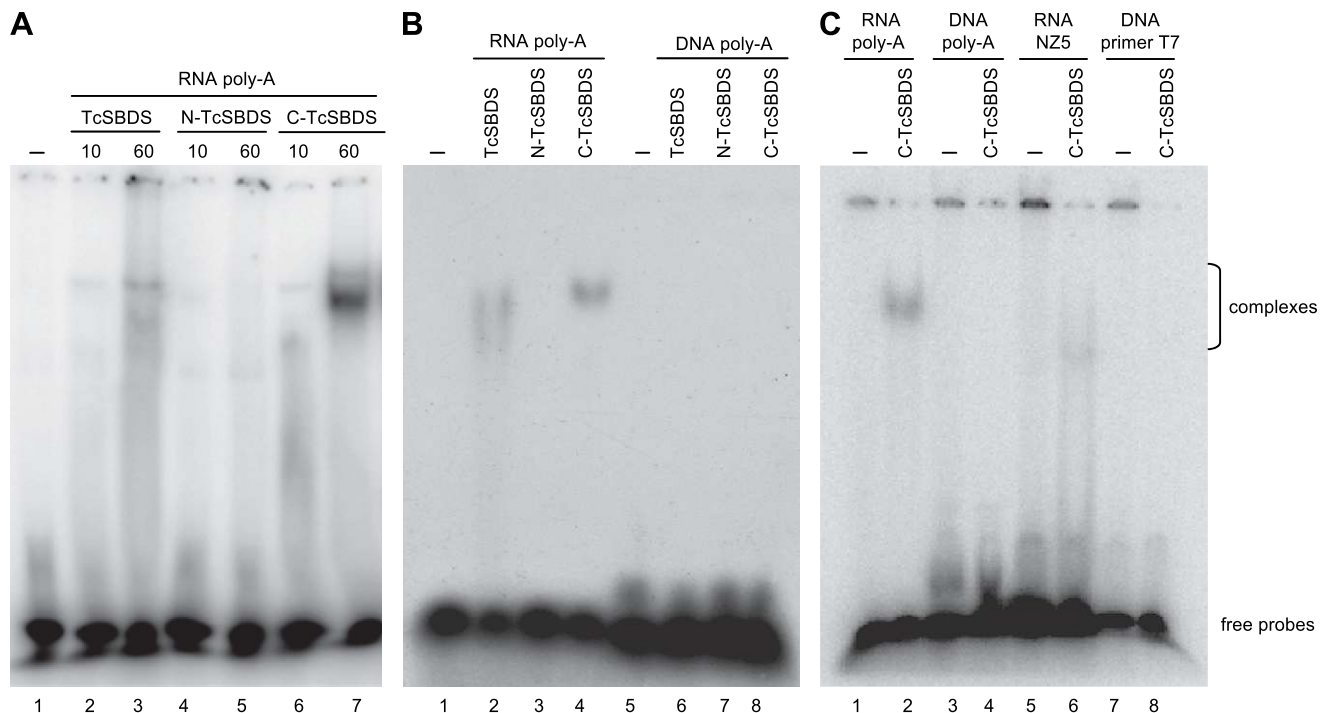


Fig. 6. Interaction of full-length and truncated TcSBDS forms with oligonucleotides. A, Autoradiograph of an electrophoretic mobility shift assay showing the interaction analysis of [^{32}P]-labeled poly- $\text{A}_{(25)}$ RNA with recombinant full-length TcSBDS, N-TcSBDS and C-TcSBDS. B, Autoradiograph of a shift assay showing the analysis of [^{32}P]-labeled RNA poly- $\text{A}_{(25)}$ and DNA poly- $\text{A}_{(20)}$ interaction with recombinant full-length TcSBDS, N-TcSBDS and C-TcSBDS. C, Autoradiograph of a shift assay showing the analysis of [^{32}P]-labeled RNA poly- $\text{A}_{(25)}$; DNA poly- $\text{A}_{(20)}$, a random 17-mer RNA sequence (RNA NZ5) and a 20-mer random DNA sequence interaction with recombinant C-TcSBDS. Protein–RNA complexes were resolved by electrophoresis under native conditions on polyacrylamide gels and visualized by autoradiography as described in the Materials and methods section.

test the interaction of TcSBDS with RNA in vitro by using electrophoretic mobility shift assays (EMSA). Interestingly, we found that the interaction of TcSBDS with RNA is mediated by the C-terminal natively unfolded region (Fig. 6A, lanes 2, 3, 6, 7). The specificity of the TcSBDS interaction with RNA was analyzed by performing parallel EMSA assays with RNA poly-A and DNA poly-A as ^{32}P -labeled probes. Band shifts were observed only for the EMSA performed with the RNA poly-A probe (Fig. 6B, lanes 2 and 4) both for the full-length TcSBDS and the C-terminal natively unfolded region, C-TcSBDS, indicating that TcSBDS shows higher affinity to RNA than to DNA. EMSA assays were also performed with C-TcSBDS using synthetic RNA (Fig. 6C, lanes 1, 2–5, 6) and DNA (Fig. 6C, lanes 3, 4–7, 8) sequences, which showed results similar to those obtained with the homopolymers RNA poly-A and DNA poly-A. These data indicate that TcSBDS shows a much higher affinity to RNA than to DNA.

Natively unfolded proteins or protein segments, which may fold and stabilize upon substrate binding, are commonly found in regulatory and signaling proteins and may allow recognition of different targets with high specificity. To determine whether the interaction with RNA induces conformational changes in C-TcSBDS, circular dichroism analyses were carried out in the presence of an RNA poly-A. The resulting spectrum of the mix C-TcSBDS and RNA differs slightly in the 200 and 220 nm regions from the theoretical sum of the spectra from C-TcSBDS and the RNA poly-A measured separately (Fig. 7A). Parallel control CD analyses of C-TcSBDS were performed with increasing concentrations of trifluoroethanol (TFE). A reduction in the negative band near the 200 nm region is observed in the presence of 5% and 10% of TFE and a gain of secondary structure is observed for TFE concentrations above 20% with the most negative peak migrating from the 200 nm to the 205 nm region and a negative band appearing near the 220 nm region (Fig. 7B). This control assay shows that depending on the milieu, C-TcSBDS can change conformation and eventually acquire secondary structure. A comparison of the CD spectra of the C-TcSBDS–RNA interaction with the spectrum of C-TcSBDS with 5% TFE indicates similar changes in the 200 nm region that might be due to conformation changes of C-TcSBDS following interaction with RNA.

4. Discussion

The available information on SBDS/Sdo1p function was obtained mainly from the yeast and mammalian systems while the structural information was obtained from an Archaeal ortholog. Here we provided evidence that the trypanosomatid ortholog may share similar functional and structural characteristics. In addition, novel features were identified in the *T. cruzi* protein that may be related to functional dissimilarities with the best characterized yeast or mammalian orthologs.

Yeast and human SBDS orthologs are 250 amino acid proteins and the length of the archaeal orthologs range from 226 to 242 residues. On the other hand, the SBDS orthologs from plants and from the phyla Chromalveolata and Excavata contain extended C-terminal regions that may reach up to 638 residues in *Plasmodium falciparum* (phylum Chromalveolata) and 535 residues in *L. major* (phylum Excavata) [11]. A group of plants and Chromalveolata organisms share C-terminal extension with a putative C2H2 zinc finger domain found in proteins that interact with nucleic acids [11]. However, sequence analyses did not identify conserved domains in the extended C-terminal of the Trypanosomatidae SBDS orthologs. The data shown in this work indicated that this C-terminal extension corresponds to a natively unfolded protein segment. The natively unfolded state of the recombinant C-TcSBDS protein was confirmed by a series of experiments, including CD to

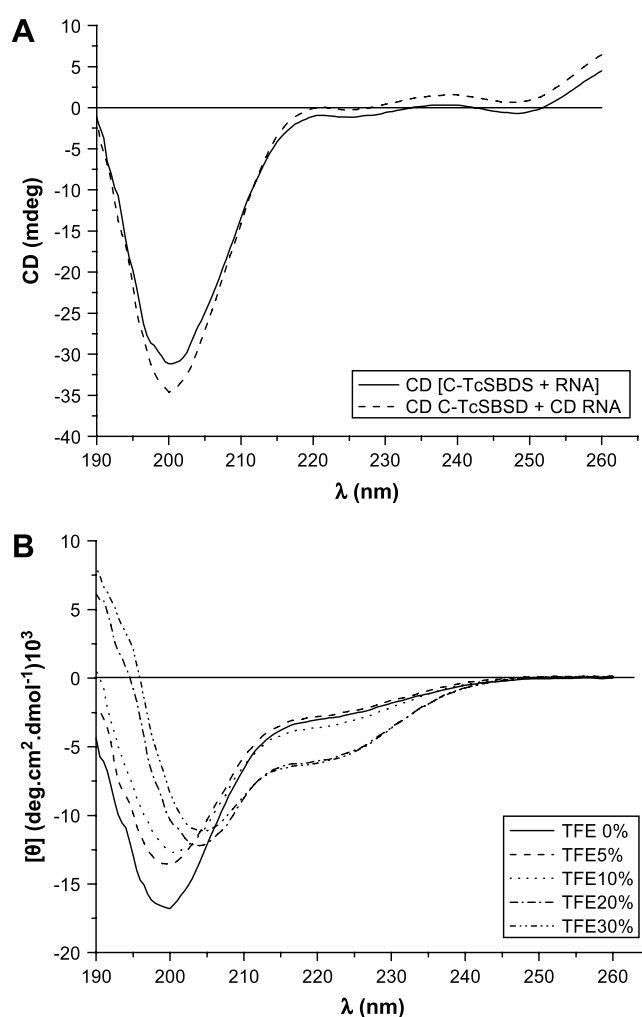


Fig. 7. Circular dichroism analysis of C-TcSBDS interaction with RNA and trifluoroethanol. A, Circular dichroism analysis of C-TcSBDS in the presence of RNA poly-A₍₂₅₎ RNA. The solid line represents the spectrum of C-TcSBDS with RNA and the dotted line represents the spectrum resulting from adding the individual spectra from C-TcSBDS and from the RNA poly-A₍₂₅₎. B, Circular dichroism analysis of C-TcSBDS in the presence of the indicated amounts of trifluoroethanol.

determine the secondary structure content, NMR analysis to determine the folding state and limited proteolysis to determine the accessibility of the recombinant C-TcSBDS to protease digestion which provides evidence on the compactness and folding state of the protein. Based on the amino acid sequence, the natively unfolded state of the C-terminal region can be extended to all trypanosomatids' SBDS proteins. A natively unfolded state for proteins or protein segments is quite common in proteins that are involved in multiple interactions, including transcriptional and translational regulators, and factors showing nucleic acid binding activity [22–25]. We clearly showed here that this C-terminal unfolded region of TcSBDS can bind RNA, and this binding induces conformational changes in the protein, as expected of a natively unfolded protein. The RNA interaction as determined by EMSA indicates a high affinity of C-TcSBDS for RNA.

Solution of the crystal structure of the *A. fulgidus* SBDS ortholog revealed a three-domain architecture in which the N-terminal domain was proposed to interact with RNA based on its similarity to the putative RNA-interacting protein encoded by the yeast open reading frame YHR087 [3,10]. TcSBDS retains residues of the domain 1 in the N-terminal region that were described by Boocock

et al. [11] both as invariant and as conserved. Although the domain 1 is predicted to mediate interaction with RNA, the N-terminal region of TcSBDS (N-TcSBDS) containing domains 1–2–3 did not show RNA interaction under the EMSA conditions used in this work. It is possible though that the N-terminal conserved region can interact with RNA only in the context of a ribosome, in which case we would not have been able to detect through these EMSA. It should be noted however, that direct RNA binding has not been previously shown for any of the other SBDS orthologs. Our results nevertheless indicate that the trypanosomatids' SBDS orthologs have evolved a C-terminal region with high capability to interact with RNA, which may suggest additional functions of this protein in these highly divergent eukaryotic organisms.

The polysome analysis described in this work showed that the *T. cruzi* SBDS protein cosediments with polysomes similarly to the yeast and human orthologs. Consistent with this result, we found that a TcSBDS antiserum co-immunoprecipitated both the 5.8S and the 28S rRNAs and the ITS2 of the pre-rRNA. Thus, the involvement of TcSBDS with ribosomal function/biosynthesis seems to be maintained in the trypanosomatid lineage. The rRNA primary transcript from trypanosomatids differs from yeast and humans in the arrangement of the large subunit rRNA, containing the 28S rRNA unit split into the 28S α and 28S β subunits and four additional small RNAs [26–28], and in the order of the processing events with the early cleavages taking place at the internal transcribed spacer 1 (ITS1) [29]. In addition, the rRNAs from *T. cruzi* are longer than the rRNAs from yeast, which results in larger RNA content in the final 80S ribosome structure [30]. These differences indicate that some steps of ribosome biogenesis in trypanosomatids may involve specific rRNA–protein interactions that are not found in organisms such as yeast, requiring compatible additional functions to be played by the biogenesis factors.

Acknowledgements

This work was supported by FAPESP grants CBME/CEPID-98/14138-2, SMOlBNet-00/10266-8 and 06/02083-7 and grant 02/12597-7 to BAC. JFO is a recipient of a FAPESP PhD fellowship. We thank Christian M. Probst, Daniela P. Pavoni and Alejandro Correa for technical support, Sergio Schenkman for providing the parasites and for critical reading of the manuscript and Samuel Goldenberg for the antiserum for TcRPL26.

References

- [1] Y. Dror, M.H. Freedman, Shwachman–Diamond syndrome, *Br. J. Haematol.* 118 (2002) 701–713.
- [2] O.P. Smith, Shwachman–Diamond syndrome, *Semin. Hematol.* 39 (2002) 95–102.
- [3] A. Savchenko, N. Krogan, J.R. Cort, E. Evdokimova, J.M. Lew, A.A. Yee, L. Sanchez-Pulido, M.A. Andrade, A. Bochkarev, J.D. Watson, M.A. Kennedy, J. Greenblatt, T. Hughes, C.H. Arrowsmith, J.M. Rommens, A.M. Edwards, The Shwachman–Bodian–Diamond syndrome protein family is involved in RNA metabolism, *J. Biol. Chem.* 280 (2005) 19213–19220.
- [4] S. Zhang, M. Shi, C.C. Hui, J.M. Rommens, Loss of the mouse ortholog of the Shwachman–Diamond syndrome gene (SBDS) results in early embryonic lethality, *Mol. Cell. Biol.* 26 (2006) 6656–6663.
- [5] K.A. Ganapathi, K.M. Austin, C.S. Lee, A. Dias, M.M. Malsch, R. Reed, A. Shimamura, The human Shwachman–Diamond syndrome protein, SBDS, associates with ribosomal RNA, *Blood* 110 (2007) 1458–1465.
- [6] C. Hesling, C.C. Oliveira, B.A. Castilho, N.I. Zanchin, 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 (2007) 4180–4195.
- [7] K.M. Austin, R.J. Leary, A. Shimamura, The Shwachman–Diamond SBDS protein localizes to the nucleolus, *Blood* 106 (2005) 1253–1258.
- [8] N.J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A.P. Tikuisis, T. Punna, J.M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M.D. Robinson, A. Paccanaro, J.E. Bray, A. Sheung, B. Beattie, D.P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M.M. Canete, J. Vlasblom, S. Wu, C. Orsi, S.R. Collins, S. Chandran, R. Haw, J.J. Rillstone, K. Gandi, N.J. Thompson, G. Musso, P. St Onge, S. Ghanny, M.H. Lam, G. Butland, A.M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J.S. Weissman, C.J. Ingles, T.R. Hughes, J. Parkinson, M. Gerstein, S.J. Wodak, A. Emili, J.F. Greenblatt, Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, *Nature* 440 (2006) 637–643.
- [9] T.F. Menne, B. Goyenechea, N. Sanchez-Puig, C.C. Wong, L.M. Tonkin, P.J. Ancliff, R.L. Brost, M. Costanzo, C. Boone, A.J. Warren, The Shwachman–Bodian–Diamond syndrome protein mediates translational activation of ribosomes in yeast, *Nat. Genet.* 39 (2007) 486–495.
- [10] S. Shammass, T.F. Menne, C. Hilcenko, S.R. Michell, B. Goyenechea, G.R. Boocock, P.R. Durie, J.M. Rommens, A.J. Warren, Structural and mutational analysis of the SBDS protein family. Insight into the leukemia-associated Shwachman–Diamond Syndrome, *J. Biol. Chem.* 280 (2005) 19221–19229.
- [11] G.R. Boocock, M.R. Marit, J.M. Rommens, Phylogeny, sequence conservation, and functional complementation of the SBDS protein family, *Genomics* (2006) 758–771.
- [12] F.R. Carneiro, T.C. Silva, A.C. Alves, T. Haline-Vaz, F.C. Gozzo, N.I. Zanchin, Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1, *Biochem. Biophys. Res. Commun.* 343 (2006) 260–268.
- [13] M.X. Li, D.C. Corson, B.D. Sykes, Structure determination by NMR: isotope labeling, *Methods Mol. Biol.* 173 (2002) 255–265.
- [14] P. Romero, Z. Obradovic, A.K. Dunker, Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genomics Inform.* 8 (1997) 110–124.
- [15] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41 (2000) 415–427.
- [16] M. Brecht, M. Parsons, Changes in polysome profiles accompany trypanosome development, *Mol. Biochem. Parasitol.* 97 (1998) 189–198.
- [17] K.H. Nielsen, B. Szamecz, L. Valasek, A. Jivotovskaya, B.S. Shin, A.G. Hinnebusch, Functions of eIF3 downstream of 48S assembly impact AUG recognition and GCN4 translational control, *EMBO J.* 23 (2004) 1166–1177.
- [18] E. Harlow, D. Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988, 748 pp.
- [19] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res.* 31 (2003) 3497–3500.
- [20] B.S.C. Alves, S. Oyama Jr., F.C. Gozzo, M. Sanches, B.G. Guimarães, N.I. Zanchin, Characterization of the human ortholog of Mov34 reveals eight N-terminal residues important for MPN domain stability, *Biochem. Biophys. Res. Commun.* 347 (2006) 608–615.
- [21] J. Cavanagh, W.J. Fairbrother, G. Palmer III, N.J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, 1996.
- [22] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradovic, A.K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol.* 323 (2002) 573–584.
- [23] L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, A.K. Dunker, The importance of intrinsic disorder for protein phosphorylation, *Nucleic Acids Res.* 32 (2004) 1037–1049.
- [24] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* 6 (2005) 197–208.
- [25] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, *Protein Sci.* 11 (2002) 739–756.
- [26] T.C. White, G. Rudenko, P. Borst, Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs, *Nucleic Acids Res.* 14 (1986) 9471–9489.
- [27] D.A. Campbell, K. Kubo, C.G. Clark, J.C. Boothroyd, Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA, *J. Mol. Biol.* 196 (1987) 113–124.
- [28] D.F. Spencer, J.C. Collings, M.N. Schnare, M.W. Gray, Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*, *EMBO J.* 6 (1987) 1063–1071.
- [29] T. Hartshorne, W. Toyofuku, Two 5'-ETS regions implicated in interactions with U3 snoRNA are required for small subunit rRNA maturation in *Trypanosoma brucei*, *Nucleic Acids Res.* 27 (1999) 3300–3309.
- [30] H. Gao, M.J. Ayub, M.J. Levin, J. Frank, The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components, *Proc. Natl. Acad. Sci. U.S.A.* 102 (29) (19 Jul 2005) 10206–10211.

3.2. ARTIGO II

Structure, dynamics and RNA interaction analysis of the human SBDS protein

Juliana F. de Oliveira, Mauricio L. Sforça, Tharin Blumenschein, Beatriz G. Guimarães, Carla C. Oliveira, Nilson I. T. Zanchin e Ana Carolina Zeri

STRUCTURE, DYNAMICS AND RNA INTERACTION ANALYSIS OF THE HUMAN SBDS PROTEIN

Juliana Ferreira de Oliveira[‡], Maurício L. Sforça[‡]; Tharin M. A. Blumenschein[§]; Beatriz G. Guimarães^{‡*}, Carla Columbano Oliveira[¶]; Nilson I. T. Zanchin[‡] and Ana-Carolina Zeri^{‡2}

[‡]Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS
Rua Giuseppe Maximo Scolfaro 10000
PO Box 6192, CEP13083-970, Campinas SP, Brazil

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

[§] School of Chemistry, University of East Anglia
Norwich - NR4 7TJ - United Kingdom

* Current adress: Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 GIF-SUR-YVETTE Cedex, France

²To whom correspondence should be addressed:
Centro de Biologia Molecular Estrutural
Laboratório Nacional de Luz Síncrotron
Campinas - SP – Brazil
Caixa Postal (PO Box) 6192 - CEP 13083-970
Tel. 55 19 3512 1119, Fax. 55 19 3512 1004
e-mail: aczeri@lnls.br

ABSTRACT

SDS (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 to mutations in the *SBDS* gene, which encodes a conserved protein showing orthologs in Archaea and eukaryotes. The SDS 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, the later being also 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 to understand the major unresolved issues regarding SBDS function, namely, the effect of the 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 the NMR experiments, for the flexible linker between the N-terminal and the central domains, 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 domain that concentrates most of the mutations described for the human SBDS.

Keywords: Shwachman-Bodian-Diamond Syndrome, SBDS protein, NMR protein structure, RNA interaction, protein dynamics.

INTRODUCTION

Currently, five human genetic syndromes are associated to loss-of-function mutations in genes encoding proteins involved in ribosome biogenesis, which include Diamond Blackfan Anemia (DBA), cartilage-hair hypoplasia (CHH), the Treacher Collins syndrome, Dyskeratosis Congenita (DC) and the Shwachman-Bodian-Diamond syndrome (SDS)¹. Mutations in the *SBDS* gene were reported for 90% of patients with the Shwachman-Bodian-Diamond syndrome (SDS; OMIM 260400). Most of 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 97% identity². SDS is an autosomal recessive disorder characterized by hematological dysfunction, pancreatic exocrine insufficiency and skeletal abnormalities, which usually manifests clinically in infancy or early childhood^{3; 4; 5}. SDS is also associated with a high propensity for malignant myeloid transformation into myelodysplasia and leukaemia⁶. Consistent with an essential function, targeted disruption of the *SBDS* gene leads to early embryonic lethality in mice⁷ 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 the high sequence conservation (Fig. 1) and wide distribution amongst species that includes Archaea and all eukaryotes. From a group of 159 genomes, 155 contain a single copy of the *SBDS* gene per haploid genome⁸. Regarding primary structure conservation, orthologs of SBDS can be separated into two major groups according to the extent of the C-terminal. One group includes Archaea, animals and fungi which possesses SBDS proteins of approximately 250 amino acid residues displaying a three-domain architecture⁸. The second group includes plants and protists, containing SBDS orthologs showing extended C-terminal regions ranging from 100–280 additional residues⁸. 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 region⁸ and the C-terminal extension of the *Trypanosoma cruzi* SBDS ortholog behaves as a natively unfolded protein segment that mediates interaction with RNA⁹.

Implication of SBDS in ribosome biogenesis has started with genome analyses that identified archaeal SBDS orthologs within highly conserved operons that contain RNA-processing genes¹⁰, and transcriptional profiling have described clustering of the yeast ortholog *SDO1* with rRNA processing factors^{11; 12}. Recent interaction studies using affinity capture followed by mass spectrometry analysis and yeast two-hybrid assays revealed association of both yeast and human SBDS with ribosome structural components and ribosome biogenesis factors^{13; 14; 15}. In yeast, Sdo1p is required for accurate synthesis and nuclear export of 60S ribosome subunits¹⁶. Sdo1p was also implicated in the release and recycling of the nucleolar shuttling factor Tif6p from pre-60S ribosomes¹⁶. This is a key step in translational activation of 60S ribosomes that involves also the elongation factor-like 1 (Efl1p) to facilitate the release of Tif6p from late cytoplasmic pre-60S subunits¹⁶. As for the human SBDS, further evidence of its involvement with ribosome biosynthesis includes its nucleolar localization¹⁷ and global decrease in rRNA synthesis reported for cells derived from SDS patients and for human skin fibroblasts knock down for SBDS¹⁸. 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 altered expression of several critical genes both at the transcriptional and translational levels¹⁹.

Structural information has also been used to infer SBDS molecular function. The structures of *Archaeoglobus fulgidus* SBDS ortholog (AfSBDS, PDB code **1T95**²⁰) and, more recently, the ortholog from *Methanothermobacter thermautotrophicus* (mthSBDS, PDB code **2WBM**²¹), were determined by X-ray crystallography, revealing a three-domain architecture. The two Archaeal proteins are very similar in sequence and structure, with 50% sequence identity and low

r.m.s.d deviations between equivalent domains. The crystals of mthSBDS²¹ have two molecules in the asymmetric unit and show an intrinsic interdomain flexibility. The SBDS N-terminal domain (termed *FYSH* for *Fungal*, *Yhr087wp*, *Shwachman* domain) displays a novel mixed α/β -fold and harbors the largest number of mutations related to the syndrome. In addition, the *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 gene *YHR087W*, which is implicated in RNA metabolism. 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}. Accordingly, the archaeal SBDS ortholog from *Pyrococcus abyssi* (PaSBDS) and yeast Sdo1p were shown to interact directly with RNA in electrophoresis mobility shift assays¹⁴. Sdo1p does not discriminate between the different RNA sequences from the yeast pre-rRNA and rRNAs tested in vitro¹⁴, a finding that does not help to answer the question whether SBDS orthologs bind specific cognate RNAs in vivo. As for the human SBDS, no band shift could be detected in electrophoresis mobility shift assays performed under the same conditions as for Sdo1p, indicating that it shows a low affinity for the RNAs tested. A study based on yeast complementation assays 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 manner⁸. These analyses have shown that the C-terminal is dispensable for complementation of a yeast deletion strain⁸. The fact that the C-terminal is dispensable raises doubts 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 the mutations in the SBDS protein affect its function leading to the SDS

syndrome. With this in mind, we sought to investigate the mechanism of human SBDS interaction with RNA and to correlate the 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 that showed to be highly similar to the structure of *A. fulgidus* 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 - 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 I.

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 $\beta 1$ - $\beta 2$ - $\alpha 1$ - $\alpha 2$ - $\beta 3$ - $\beta 4$ - $\alpha 3$ - $\alpha 4$, a strand helix hairpin - SHHP domain²³; a flexible linker, from residues 96 to 106 connects the N-terminal domain to the central domain, residues 107-167, which forms a three-helical bundle ($\alpha 5$ - $\alpha 6$ - $\alpha 7$); a short flexible linker from residues 168 to 172 connects the central to the C-terminal domain,

residues 173 to 250, comprising a four-stranded antiparallel β -sheet with two α -helices packing against the sheet's concave surface ($\beta 5$ - $\alpha 8$ - $\beta 6$ - $\beta 7$ - $\alpha 9$ - $\beta 8$), in a RRM (RNA Recognition Motif) 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 colored in yellow is shown in Fig. 2C. A few inter-domain NOEs could be assigned, mostly between the central and C-terminal domain, consistent with the higher flexibility seen for the N-terminal domain relative to the remaining of the protein. A few medium sized NOE signals were identified between residues at the N and central domains, as shown in detail in Figure 2D. The most striking are those between the side-chain of Lysine 151, in the central domain, and the alpha-carbon hydrogens of Serine 41 and Glycine 42 in the N-terminal domain. This indicates that the relative motions between the domains are not completely random but somewhat constrained, as shown in Figure 2B, where 7 low-energy structures are superposed via the central domain.

A subset of residues could not be assigned with the triple resonance strategy devised by Bax and co-workers²⁴, and some lie on an eleven-residue fragment (Ser96–Gln106) within the loop connecting domains 1 and 2 (Fig. 2C, highlighted in yellow). Missing peaks in the ¹⁵N-HSQC experiments are characteristic of regions in the polypeptide chain undergoing conformational exchange on an intermediate NMR time scale²⁵, and a search for alternative origins of extreme line broadening for this region excluded other explanations. Published NMR studies support the conformational exchange hypothesis²⁶ as well as the relaxation parameters measured for the protein, also in agreement with published results²⁷. The plot of transverse relaxation rate, R_2 , vs. residue number, as shown below, presents elevated values for residues next to the loop region, what suggests that those sites experience chemical exchange on microsecond to millisecond time scales²⁷. The corresponding residues in the yeast protein Yhr087w, also studied by NMR (²² PDB code **1NYN**) were not assigned, likely due to the intrinsic flexibility of the loop.

Comparison of the human SBDS to the 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 high evolutionary conservation between the 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. 3 A to C. The major difference between human SBDS (in green) and AfSBDS (blue) is the connecting loop between the N-terminal and central domains. While in the solution structure of human SBDS the geometry of this loop is largely undefined, 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 human SBDS solution structure and AfSBDS crystal structure resulted, respectively, in r.m.s.d. values of 1.6, 1.7 and 2.1 Å. A total of 70, 52 and 61 alpha carbons were superposed for each domain using the software SuperPose²⁸. The topologies of the domains are highly similar, with small differences in the length of some of the secondary structure elements. For the N-terminal domain of the lowest energy structure, beta strands $\beta 1$, $\beta 2$, and helix $\alpha 1$ are slightly shorter and strands $\beta 3$ and $\beta 4$ are longer for human SBDS with a small helical segment between them. A comparison between HsSBDS and mthSBDS yields similar results, with r.m.s.d. values for the independently superposed domains 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 can be seen Fig. 3 A to C, where the mthSBDS domains are depicted in yellow. The superposition was done with the software SuperPose²⁸ considering 72, 53 and 56 alpha carbons for each domain. Savchenko²² and co-workers described the NMR structure of the *S. cerevisiae* single domain protein Yhr087w which shows structural similarity to the N-terminal of AfSBDS. Alignment of 67 alpha carbons of Yhr087wp

and the N-terminal domain of human SBDS (Fig. 3D) results in an RMSD of 2.0Å. The major difference between these structures is a longer loop between strands $\beta 3$ and $\beta 4$ 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 done at 600 MHz. The most striking feature of the plots of relaxation rates is the difference in the dynamic properties between the N-terminal and the remaining domains. While the central and C-terminal domains behave like one entity, the relaxation times are significantly different from the N-terminal domain. The missing resonances for the residues in the loop connecting these domains reflect these inter-domain movements and the superposition of the structures in Fig. 2B show that even though the domains are rigid, there are significant inter-domain movements in solution.

Reduced spectral density mapping analysis²⁹ evaluates the spectral density function, J , at three different frequencies for each magnetic field: $J(0)$, increased by slow motions in the millisecond–microsecond timescale, and decreased by fast internal motions in the picosecond–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 thirds domains, while the N-terminal domain shows an increase in fast motions with respect to the other domains. The values for $J(0.87\omega_H)$, that indicate internal movements in the subnanosecond time scale 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 when compared to the central and C-terminal domains, and shows that it moves independently from the other two. Apparent rotational correlation times can be calculated for the different regions of the

molecule, according to the formula³⁰:

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

where ω_N is the Larmor frequency for ^{15}N in the magnetic field used. Applying the formula to individual residues, two distinct regions can be seen in SBDS: the N-terminal domain, with residues 15-95 presenting an apparent rotational correlation time 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 with an apparent rotational correlation time of 19.0 ± 3.9 ns at 600 MHz, and 18.5 ± 3.8 ns at 800 MHz. The first 10 and last 8 residues of SBDS show decreased apparent rotational correlation times, as it is normally the case with the N and C termini of proteins.

The larger standard deviation for the apparent rotational correlation times for the central and C-terminal domains is probably due to anisotropic 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 from the rest of the molecule, in a more isotropic way.

Characterization of the SBDS-RNA interaction

In order to selectively monitor SBDS during formation of the complex protein/RNA, uniformly ^{15}N -labeled SBDS was used and two-dimensional ^{15}N -HSQC spectra of the protein were recorded before and after 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. Fig. 6A shows a portion of the ^{15}N -HSQC spectra of mixtures of RNA:SBDS at four different points through the titration corresponding to molar ratios 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 the chemical shifts for the backbone residues that indicate 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 and bound states. 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 and bound forms. If the exchange rate is lower than the chemical shift difference, two peaks are observed, for the free and bound states. When the exchange rate is comparable with the chemical shift difference, the intermediate case, the peaks become broadened and may be unobservable²⁵. 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 concomitant intensity decrease (e.g. residue Ile72 in Fig. 6C) and a few cross peaks have changes in chemical shift (e.g. residue Arg26 in Fig. 6D). Most cross peaks are essentially unperturbed even at the highest RNA concentrations. This behavior is consistent with binding reactions in which interconversion between the RNA-bound and free forms of the protein occurs in the intermediate to fast NMR exchange regime, ranging from micro to milliseconds³¹. The experiments with a DNA oligonucleotide containing the corresponding sequence showed no significant perturbation of the resonances upon titration.

Amino acid residues of human SBDS involved in RNA interaction and structural mapping of the RNA binding site - The amino acids that experience changes in their chemical environment upon RNA titration are identified by the analysis of the 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 ¹⁵N-HSQC spectrum

recorded with a RNA:SBDS molar ratio of 2:1. These values are given in histogram form in Fig. 6D, which combines the changes in ¹H and ¹⁵N 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 ¹⁵N 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 dashed line in Figure 6D 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. 6C shows that all the peaks which 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. 6B, 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 RRM domains-RNA interactions studied by NMR³⁴. 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 RNABindR³⁵ indicates some of the residues in this region (Arg22, Lys25, Arg26) as

potential RNA binding sites, and does not predict any residues in the central or C-terminal domains. Taken together, these data provide strong evidence that SBDS interaction with RNA takes place via the N-terminal domain.

The electrostatic surface of HsSBDS, AfSBDS and mthSBDS models shown in Fig. 7, calculated and plotted with MolMol³⁶ reveal differences in the distribution of basic residues and the patch of positively charged residues is present in human SBDS in the region that shows 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.

DISCUSSION

Overall structure and domain dynamics - Evidence on the modular nature of the SBDS protein family was first suggested from the *A. fulgidus* SBDS crystal structure²⁰. 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 residues⁸. The C-terminal extension found in 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, evidence from experimental analyses performed with the *T. cruzi* ortholog⁹ and from theoretical predictions indicate that the C-terminal extension of the longer SBDS orthologs behave as an intrinsically disordered protein domain. The modular nature of SBDS was supported by evidence from genetic complementation assays of a $\Delta sdo1$ strain. These assays have shown that, except for the closely related *S. pombe* ortholog, full length SBDS proteins function in a species-specific manner⁸ although the N-terminal domain of Sdo1p can be exchanged with the equivalent domain of several eukaryotic species (*H. sapiens*, *C. elegans*, *D.*

melanogaster, *A. thaliana*, *L. major*). The results described by Boocock and co-workers⁸ have also led to the conclusion that the central domain confers species-specificity and that the C-terminal domain can be dispensable to complement this particular yeast strain⁸.

The major difference between the human SBDS and AfSBDS structures involves the flexibility of the connecting loops, especially the 11-residue loop connecting the N-terminal and central domains. 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 inter-domain flexibility suggested by the NMR data presented here is an important characteristic of the molecule that deserves further investigation. The interdomain flexibility found for the recently published Archaeal protein mthSBDS (PDB code **2WBM**) suggests that structural flexibility may be essential for 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; 14; 15}. Direct protein-protein interaction has been demonstrated for yeast Sdo1p and Nip7p¹⁴. 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, the sex-lethal and PABP proteins bind to single-stranded RNA using two RNA binding domains separated by a short 10-12 amino acid linker^{37; 38} and hnRNP A1 binds to single-stranded DNA or RNA, using two domains separated by a 17 amino acid linker³⁹. Nucleolin interacts with a stem-loop located at the 5' external transcribed sequence of the mammalian pre-ribosomal RNA using two domains of the RRM-type connected by a short four-residue linker⁴⁰. 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-loop⁴⁰. In the case of the polypyrimidine tract binding protein 1 (PTB1), which contains four RRM-type domains separated by three long linkers of 51, 91 and 23 residues respectively, although the structure of individual domains is highly similar in both the free and RNA-bound forms⁴¹, the RRM domains 3 and 4 connected by the long, flexible linker tumble independently in solution, having no fixed relative orientation⁴².

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, or it may play a function in protein interaction. Different RRM domains interact with RNAs with different affinities⁴¹ and more importantly, the RRM2 of human PTB1 has been implicated in protein-protein interaction⁴³. Similarly, the human PACT protein is formed by three RRM domains with RRM1 and RRM2 showing respectively high and intermediate binding affinities for RNA, 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 difference in electrostatic surfaces between archeal and human SBDS suggest that different domains may be responsible for RNA binding.

The finding that the C-terminal is dispensable for complementation of a yeast *Δsdo1* strain⁸ 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. Interestingly, our findings confirm with high accuracy a previous statement by Wessels and co-workers²³ 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.

The pattern of chemical shift changes of N-terminal domain residues identified by RNA titration is consistent with the 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 of 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 RRM domains-RNA interactions studied by NMR³⁴.

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, TcSBDS⁹, 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 domain⁹, 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-terminal of trypanosomatidae SBDS⁹ mediate additional contacts with RNA, although based only on the set of information available to date, the possibility that SBDS might play additional functions on RNA metabolism in these organisms cannot be ruled out yet.

The yeast ortholog of SBDS, Sdo1p, does not require a specific RNA sequence for RNA binding in electrophoresis mobility shift assays¹⁴. These results are corroborated by recently published data on the archeal mthSBDS that does not show sequence specificity for binding²¹. The actual RNA target sequence of human SBDS is not known and it does not produce band shifts in electrophoresis mobility shift assays, performed under the same conditions as for Sdo1p. However, the data presented in this work show that it interacts with RNA but not with a corresponding single-stranded DNA. The line broadening caused by the intermediate exchange regime was not sufficient for a precise calculation of the affinity constant for RNA, and also hampered further dynamics and orientation studies with the protein-RNA complex, aimed at mapping changes in the domains movements and relative orientations.

Mutations in human SBDS and implication on RNA interaction - Most of the disease-related mutations in human SBDS⁴⁶ 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, what 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 could destabilize the structure, such as the L71P mutation, which causes loss of protein expression and is located next to a residue perturbed by RNA interaction, Ile72. 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 - The SBDS cDNA was amplified by PCR from a human fetal brain cDNA library using 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-TEV⁴⁷ 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 [¹³C]glucose and 1 g/L [¹⁵N]ammonium chloride. Expression was induced at an OD₆₀₀ of 0.8-1.0 for 4 h by adding IPTG to a 0.5 mM final concentration. Triply-labeled (¹⁵N, ¹³C, ²H) SBDS was prepared as described by Li and collaborators⁴⁸ with 12 h of induction at 30 °C. The cells were harvested by centrifugation at 5000g for 10 min and suspended in buffer A (50 mM Tris–HCl pH 7.2, 500 mM NaCl, 10 mM β-mercaptoethanol, 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% weight/volume) for 1 h at 4 °C, with agitation. The supernatant fraction was isolated by centrifugation (23,000g) 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-to-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% w/w TEV protease at 25 °C for 24 h in buffer C (50 mM Tris-HCl pH 8.0, 0.05 mM EDTA; 1 mM 1,4-dithiothreitol and 50 mM NaCl). Two linker residues are 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, 1 mM 1,4-dithiothreitol) 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 buffer containing 20 mM sodium phosphate pH 7.2, 50 mM NaCl and 1 mM 1,4-dithiothreitol (DTT). SDS-PAGE, Circular Dichroism and Dynamic Light Scattering reassured the protein purity and identity, and the protein identity was confirmed by mass

spectrometry, with 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 NOESY (80 ms mixing time, chosen after a build up curve with other mixing time values); ^{15}N -edited 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 mixing time); HBHA(CO)NH^{24, 49; 50; 51}. Triply-labeled (^{15}N , ^{13}C , ^2H) SBDS was used to record the experiments HNCACB⁵⁰ and HN(CO)CACB⁵² at 800 MHz. Dynamics experiments were conducted at both 600 and 800 MHz. NMR data were processed using the NMRPipe software⁵³ and analyzed using NMRView⁵⁴.

Structure Calculation - The structure of SBDS protein was calculated in a semi-automated iterative manner with the program CYANA version 2.1⁵⁵, 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% $^2\text{H}_2\text{O}$; hydrogen bond donors were identified by the presence of an amide peak in the ^{15}N -HSQC recorded after 2h30min. Hydrogen bonding constraints were then added to the structure calculation protocol. The structures obtained were further refined by restrained minimization and molecular dynamic studies using CNS⁵⁶. 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-NMR⁵⁷ and structural statistics are presented in Table I.

Relaxation Measurements - For the 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 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 17, 34, 68, 102, 136, 170 and 204 ms at 800 MHz. In all the experiments at 600 MHz, a relaxation delay of 3 s was used, while 8 s were used at 800 MHz. For the heteronuclear NOE measurements 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 the 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 8s. In both cases, peak volumes were fitted to a single exponential decay function using the program NMRView.

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 20-mer corresponding single-stranded oligodeoxyribonucleotide (DNA = 5'-TAA TAC GAC TCA CTA TAG GG-3') was used as a control. For practical convenience we used the sequence of the promoter of the T7 bacteriophage RNA polymerase for the oligonucleotides. Synthetic oligonucleotides were purchased from IDT-Integrated DNA Technologies, Inc. The titrations with RNA or DNA oligonucleotides were followed by recording 2D ^{15}N -HSQC spectra of [^{15}N]SBDS protein. Titrations were performed by direct addition of small aliquots (1 to 16.3 μL) of 20-mer oligonucleotide (15 mM) to the NMR tube containing 350 μL [^{15}N]SBDS (0.35 mM) to obtain titration points of oligonucleotide:SBDS at ratios 0:1, 0.25:1, 0.5:1, 1:1, 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.

ACCESSION NUMBERS: The coordinates have been deposited at the Protein Data Bank under code **2KDO**, and experimental data was deposited at the Biological Magnetic Resonance Data Bank, BMRB under accession number **16119**.

ACKNOWLEDGMENTS

This work was supported by FAPESP grants CBME/CEPID 98/14138-2 to NITZ and ACZ, and 06/02083-7 to NZ. JFO is a recipient of a FAPESP PhD fellowship. We thank The Wolfson Foundation for its support to the UEA-Wolfson Molecular Structure Centre at the University of East Anglia.

REFERENCES

1. Liu, J. M. & Ellis, S. R. (2006). Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood* 107, 4583-8.
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. *Nature Genetics* 33, 97-101.
3. Bodian, M., Sheldon, W. & Lightwood, R. (1964). Congenital hypoplasia of the Exocrine Pancreas. *Acta Paediatr.* 53, 282-93.
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, 645-63.
5. 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. & Durie, P. R. (1999). Shwachman syndrome: phenotypic manifestations of sibling sets and isolated cases in a large patient cohort are similar. *J Pediatr* 135, 81-8.
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, 279-84.
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. *Molecular Cell Biology* 26, 6656-63.
8. Boocock, G. R., Marit, M. R. & Rommens, J. M. (2006). Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics* 87, 758-71.
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, 475-83.
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, 240-52.

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. *Nature Genetics* 31, 255-65.
12. Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., W, W. Z., Mitsakakis, N., Ryan, O. W., Datta, N., Jojic, V., Pal, C., Canadien, V., Richards, D., Beattie, B., Wu, L. F., Altschuler, S. J., Roweis, S., Frey, B. J., Emili, A., Greenblatt, J. F. & Hughes, T. R. (2003). A panoramic view of yeast noncoding RNA processing. *Cell* 113, 919-33.
13. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandhi, K., Thompson, N. J., Musso, G., Onge, P. S., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A. & Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637-43.
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, 287-98.
15. Ball, H. L., Zhang, B., Riches, J., Gandhi, R., Li, J., Rommens, J. M. & Myers, J. S. (2009). SBDS is a multi-functional protein implicated in cellular stress responses. *Human Molecular Genetics* [Epub ahead of print].
16. Menne, T. F., Goyenechea, B., Sanchez-Puig, N., Wong, C. C., Tonkin, L. M., Ancliff, P. J., Brost, R. L., Costanzo, M., Boone, C. & Warren, A. J. (2007). The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nature Genetics* 39, 486-95.
17. Austin, K. M., Leary, R. J. & Shimamura, A. (2005). The Shwachman-Diamond SBDS protein localizes to the nucleolus. *Blood* 106, 1253-8.
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, 1458-65.
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, 4180-95.
20. Shamas, C., Menne, T. F., Hilcenko, C., Michell, S. R., Goyenechea, B., Boocock, G. R., Durie, P. R., Rommens, J. M. & Warren, A. J. (2005). Structural and mutational analysis of the SBDS protein family. Insight into the leukemia-associated Shwachman-Diamond Syndrome. *J Biol Chem* 280, 19221-9.
21. Ng, C. L., Waterman, D. G., Koonin, E. V., Walters, A. D., Chong, J. P., Isupov, M. N., Lebedev, A. A., Bunka, D. H., Stockey, P. G., Ortiz-Lombardia, M. &

- Antson, A. A. (2009). Conformational flexibility and molecular interactions of an archaeal homologue of the Shwachman-Bodian-Diamond syndrome protein. *BMC Structural Biology* 9, 32.
22. Savchenko, A., Krogan, N., Cort, J. R., Evdokimova, E., Lew, J. M., Yee, A. A., Sánchez-Pulido, L., Andrade, M. A., Bochkarev, A., Watson, J. D., Kennedy, M. A., Greenblatt, J., Hughes, T., Arrowsmith, C. H., Rommens, J. M. & Edwards, A. M. (2005). The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. *J Biol Chem* 280, 19213-20.
 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. *Journal of Cell Science* 119, 370-9.
 24. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, 93-158.
 25. Cavanagh, J., Fairbrother, W. J., III, A. G. P., Skelton, N. J. & Rance, M. (2007). *Protein NMR spectroscopy: principles and practice*. 2nd edit, Academic Press.
 26. Yao, S., Liu, M. S., Masters, S. L., Zhang, J. G., Babon, J. J., Nicola, N. A., Nicholson, S. E. & Norton, R. S. (2006). Dynamics of the SPRY domain-containing SOCS box protein 2: flexibility of key functional loops. *Protein Science* 15, 2761-72.
 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, 30448-59.
 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, W590-4.
 29. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral density function mapping using ¹⁵N relaxation data exclusively. *J Biomol NMR* 6, 153-62.
 30. Bracken, C., Carr, P. A., Cavanagh, J. & 3rd., A. G. P. (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, 2133-46.
 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, 17053-9.
 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, 456-62.

34. Bae, E., Reiter, N. J., Bingman, C. A., Kwan, S. S., Lee, D., Jr, G. N. P., Butcher, S. E. & Brow, D. A. (2007). Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J Mol Biol* 367, 1447-58.
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, W578-84.
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., Muto, Y. & Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature* 398, 579-85.
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, 835-45.
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, 1102-15.
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, 6870-81.
41. Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. & Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309, 2054-7.
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, 3132-41.
43. Rideau, A., Gooding, C., Simpson, P. J., Monie, T. P., Lorenz, M., Huttelmaier, S., Singer, R. H., Matthews, S., Curry, S. & Smith, C. W. (2006). A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nature Struct Mol Biol* 13, 839-48.
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, 175-86.
46. Costa, E. & Santos, R. (2008). Hematologically important mutations: Shwachman-Diamond syndrome. *Blood Cells Mol Dis* 40, 183-4.
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, 260-8.
48. Li, M. X., Corson, D. C. & Sykes, B. D. (2002). Structure determination by NMR. Isotope labeling. *Methods Mol Biol* 173, 255-65.
49. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly ¹³C/¹⁵N-enriched 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. *Journal of Magnetic Resonance B* 101, 201-205.
51. Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). ^1H - ^1H correlation via isotropic mixing of ^{13}C magnetization: A new three-dimensional approach for assigning ^1H 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, 6291-3.
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., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst D* 54, 905-921.
57. Laskowski, R. A., Rullmann, 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.

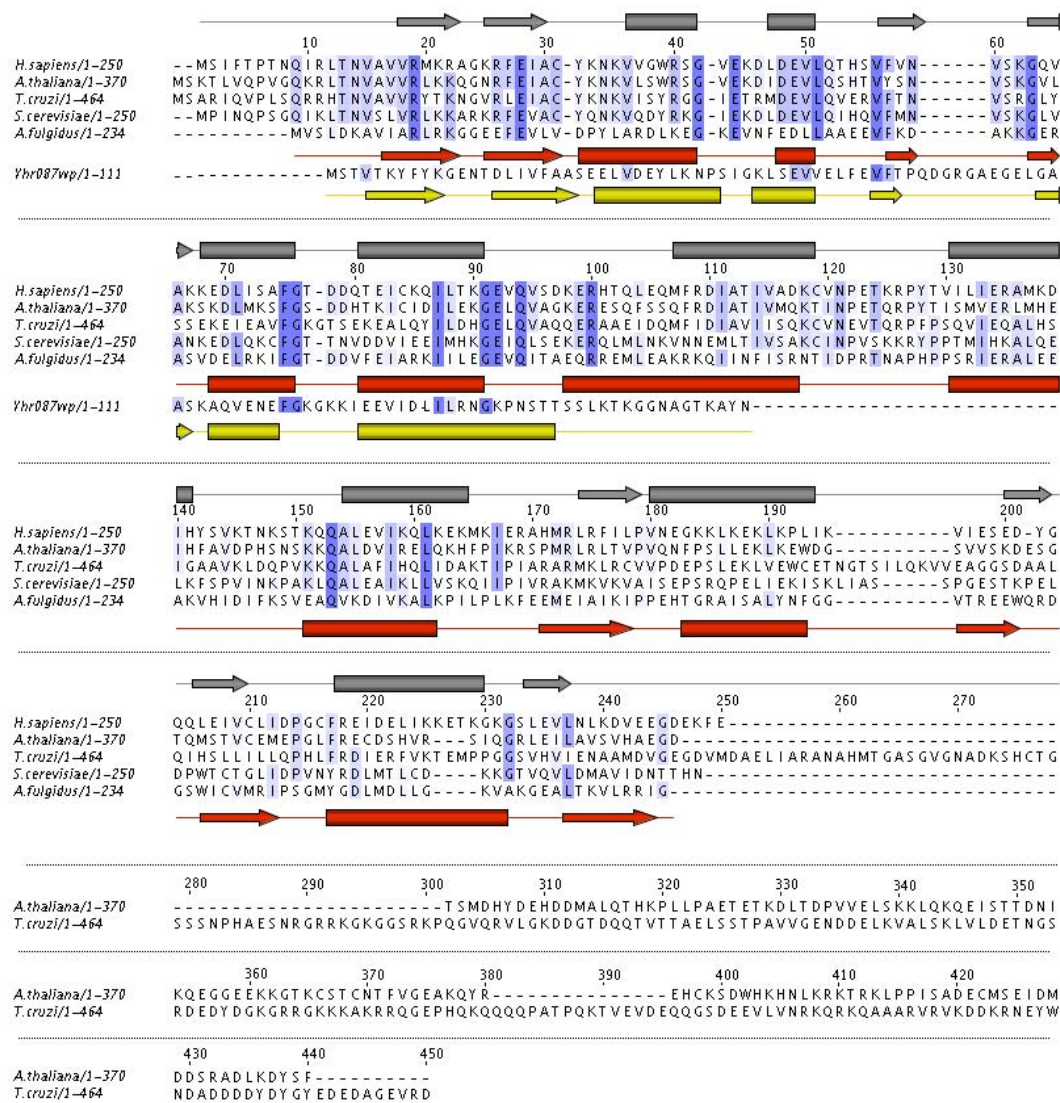


Fig. 1. Sequence alignment of SBDS orthologs and Yhr087wp protein. The sequences of SBDS orthologs shown are from human (*H. sapiens*), plant (*A. thaliana*), trypanosomatid (*T. cruzi*), yeast (*S. cerevisiae*) and *Archaea* (*A. fulgidus*). Yhr087wp is from *S. cerevisiae*. Conserved residues are highlighted in blue. The alignment was generated by ClustalW. Secondary structure elements for human SBDS, AfSBDS and Yhr087wp are colored in gray, red and yellow, respectively. The numbering corresponds to the amino acid sequence for human SBDS.

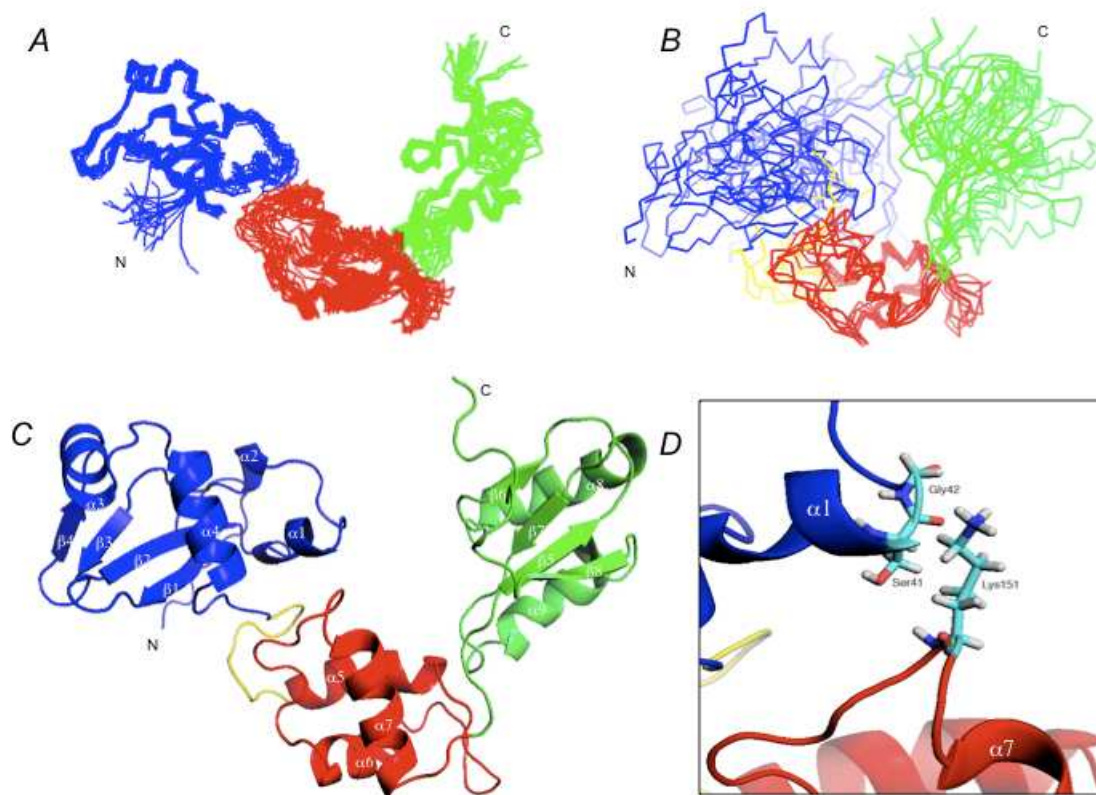


Fig. 2. Structure of human SBDS protein. *A*, 20 lowest energy structures with folded domains independently superposed, and the flexible loop left out for clarity of presentation. *B*, Superposition by the central domain of 7 representative low energy structures illustrating the broad range of inter-domain motions. A few NOE signals confirm the lower spread of movements between the central and C-terminal domains, 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 the X-ray structures), with the unassigned loop connecting the N-terminal and central domains colored in yellow. N-terminal domain is shown in blue, central domain in red and C-terminal domain in green colors. *D*, Zoom of region between N and central domains showing residues that give rise to medium intensity NOE signals.

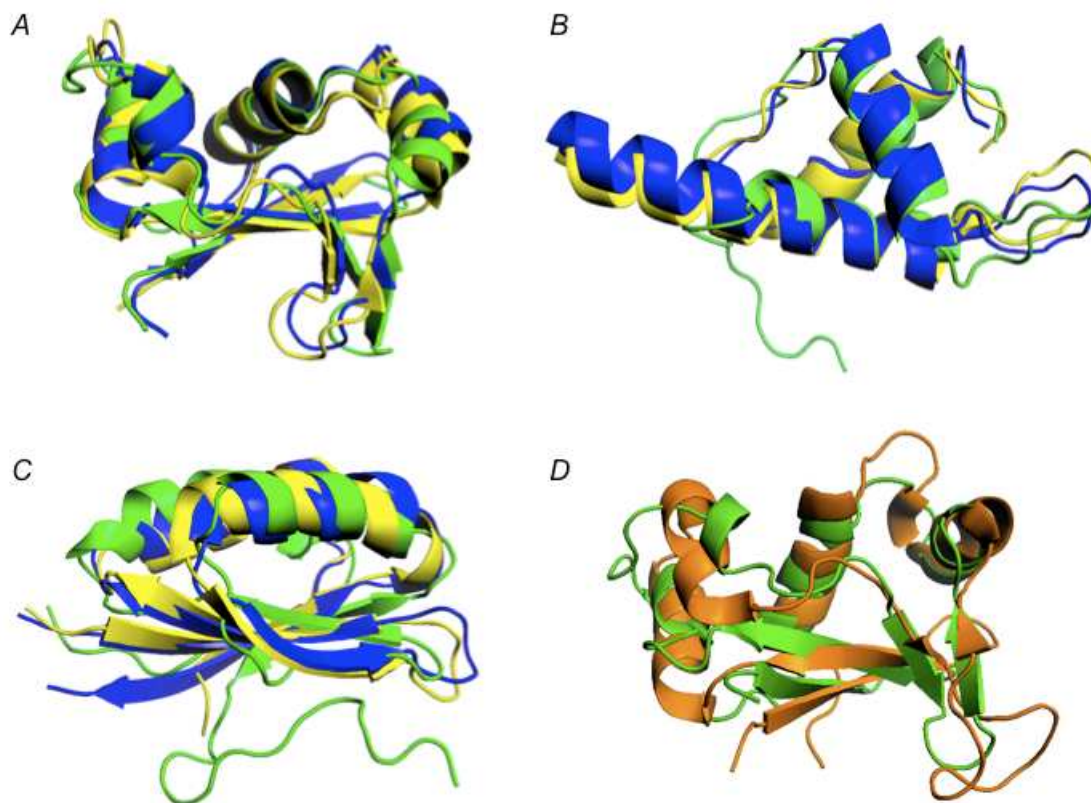


Fig. 3. Structural comparison of the domains of human SBDS with orthologs from Archaea and with yeast Yhr087w. *A*, *B*, *C*, Structure superposition of N-terminal (*A*), central (*B*) and C-terminal (*C*) domains of human SBDS (PDB 2KDO) (green) with the respective domains of *A. fulgidus* SBDS (PDB 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.

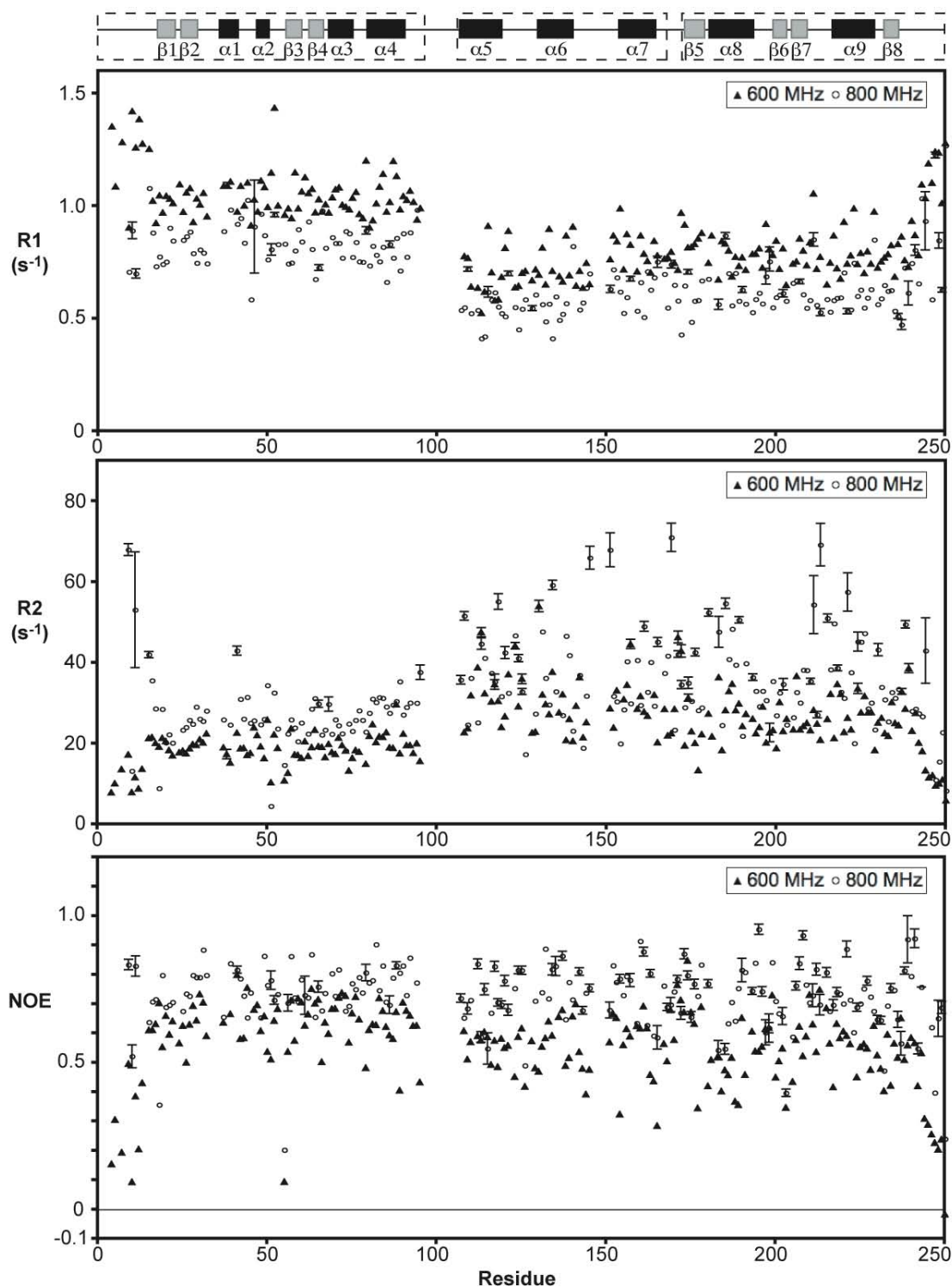


Fig. 4. Relaxation parameters *versus* residue number. Top to bottom: R_1 , R_2 and heteronuclear ^1H - ^{15}N NOE derived from data collected at 600 and 800 MHz, plotted against residue number. Residues for which no result is shown correspond either to prolines or to missing/overlapped cross-peaks that could not be analyzed quantitatively. The domain limits are shown by dashed boxes around the sequence representation. Error bars were smaller than point marker and are not visible for some points.

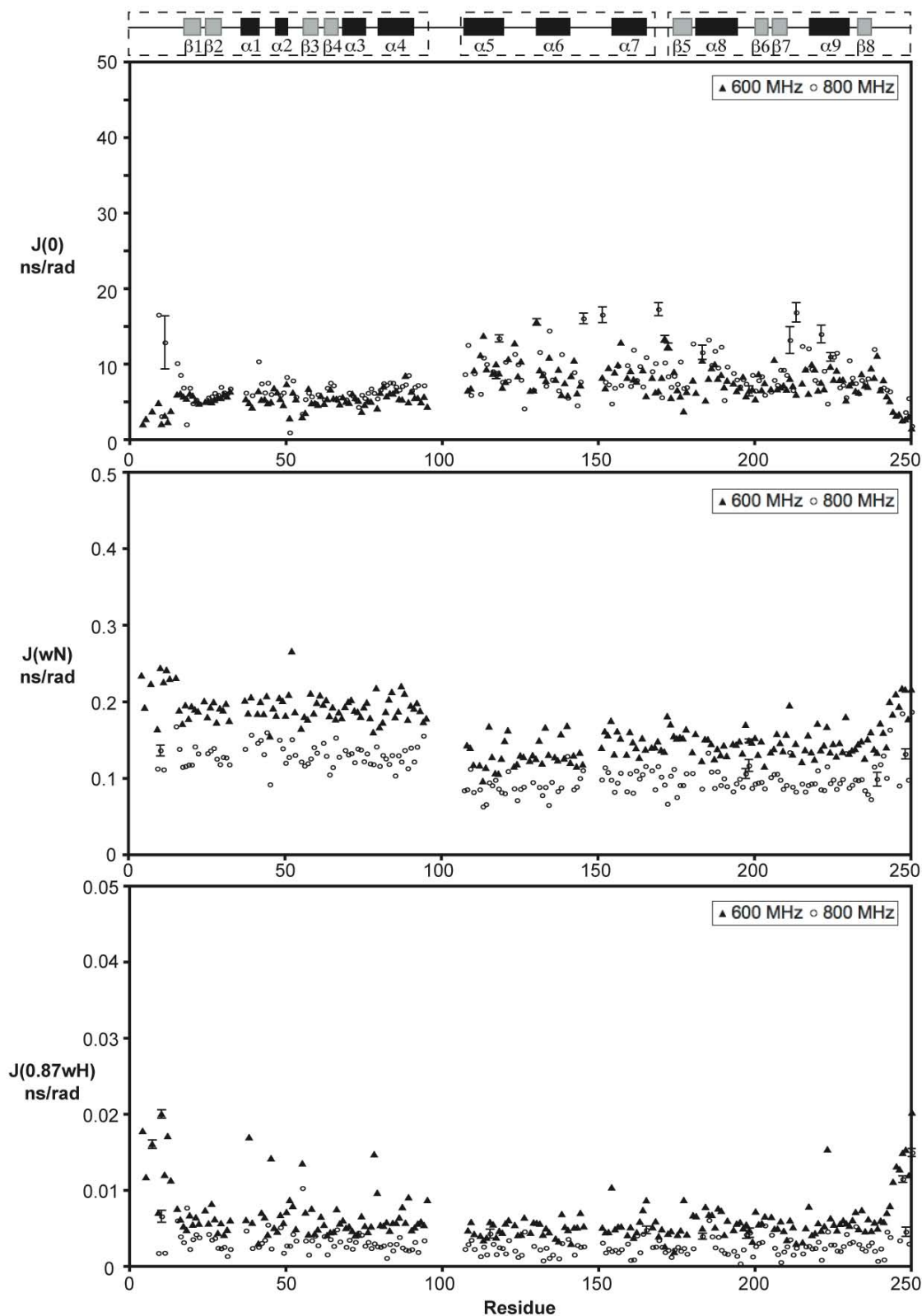


Fig. 5. Plots of reduced spectral density mapping data for human SBDS at 600 and 800 MHz. The domain limits are shown by dashed boxes around the sequence representation. Error bars were smaller than point marker and are not visible for some points.

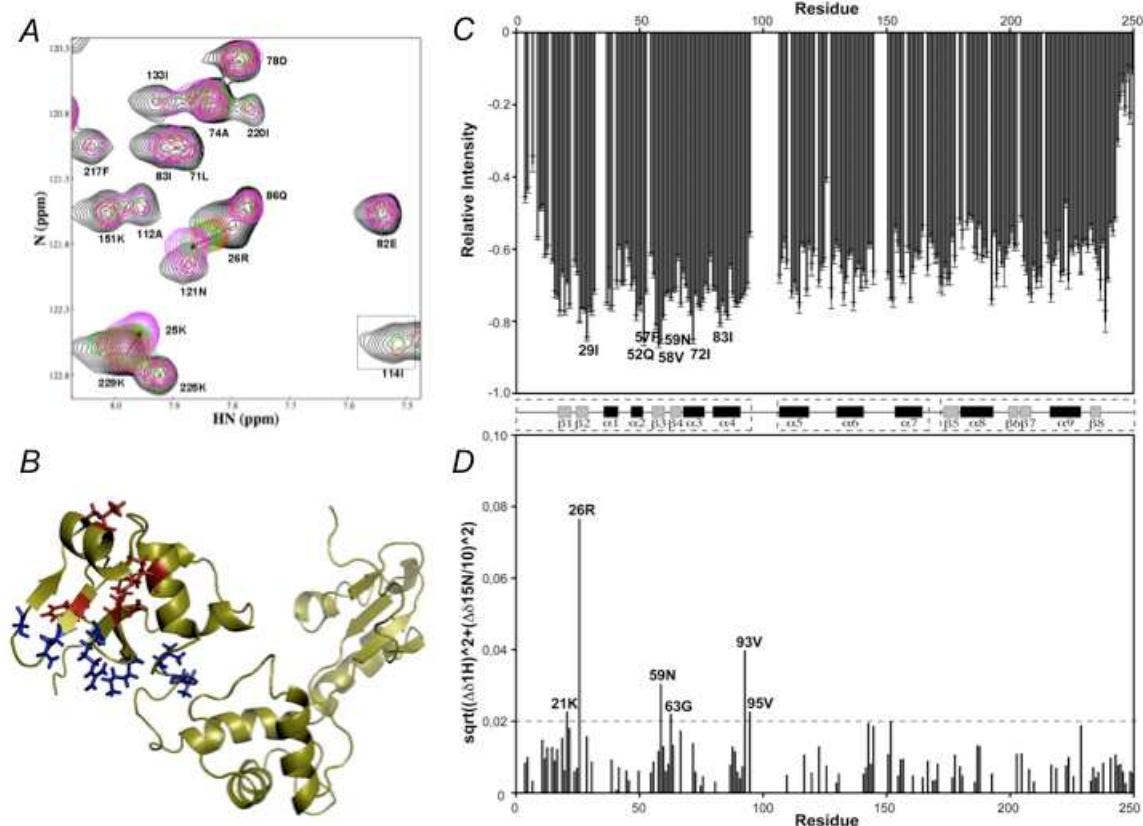


Fig. 6. Changes in the NMR spectra of SBDS upon RNA binding. *A*, Section of the ^1H - ^{15}N -HSQC spectra of free SBDS (black) superimposed on a series of spectra of the protein bound to 20-mer RNA. The peaks in red, green and pink represent spectra with RNA:SBDS ratios of 0.25:1; 0.5:1 and 2:1 respectively. *B*, 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 colored red; residues with RNA-induced changes in cross peak chemical shifts are colored blue. 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. *C*, Plot of the fractional changes of peak intensities for individual backbone amide resonances of SBDS upon addition of 20-mer RNA, with stoichiometric ratio RNA:SBDS 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. The 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). *D*, The graph shows the scaled chemical shift differences between the HSQC spectra of the protein and the 2:1 RNA:protein mixture, with most affected residues labeled. Scaling was done according to the equation $\Delta\delta(^{15}\text{N}+^1\text{H}) = ((\Delta\delta(^{15}\text{N}/10))^2 + (\Delta\delta(^1\text{H}))^2)^{1/2}$ and the dashed line indicates the 0.02 ppm threshold.

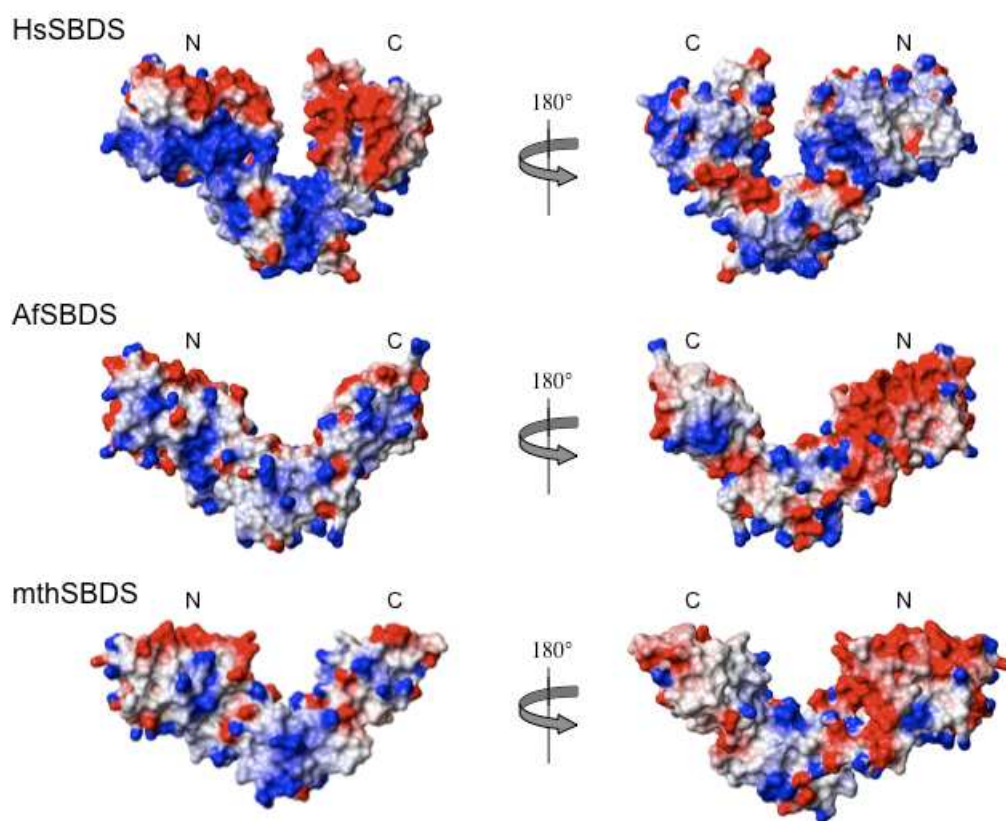


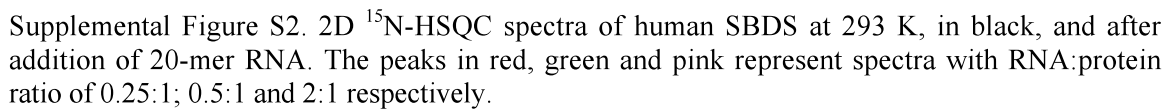
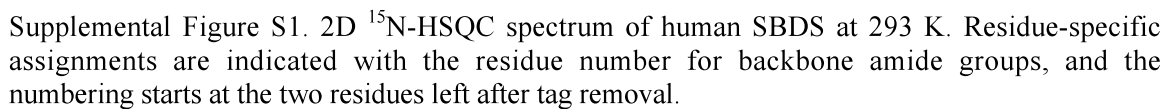
Fig. 7. Electrostatic surfaces of human SBDS, AfSBDS and mthSBDS colored from blue (positive) to red (negative). Figures generated using MOLMOL.

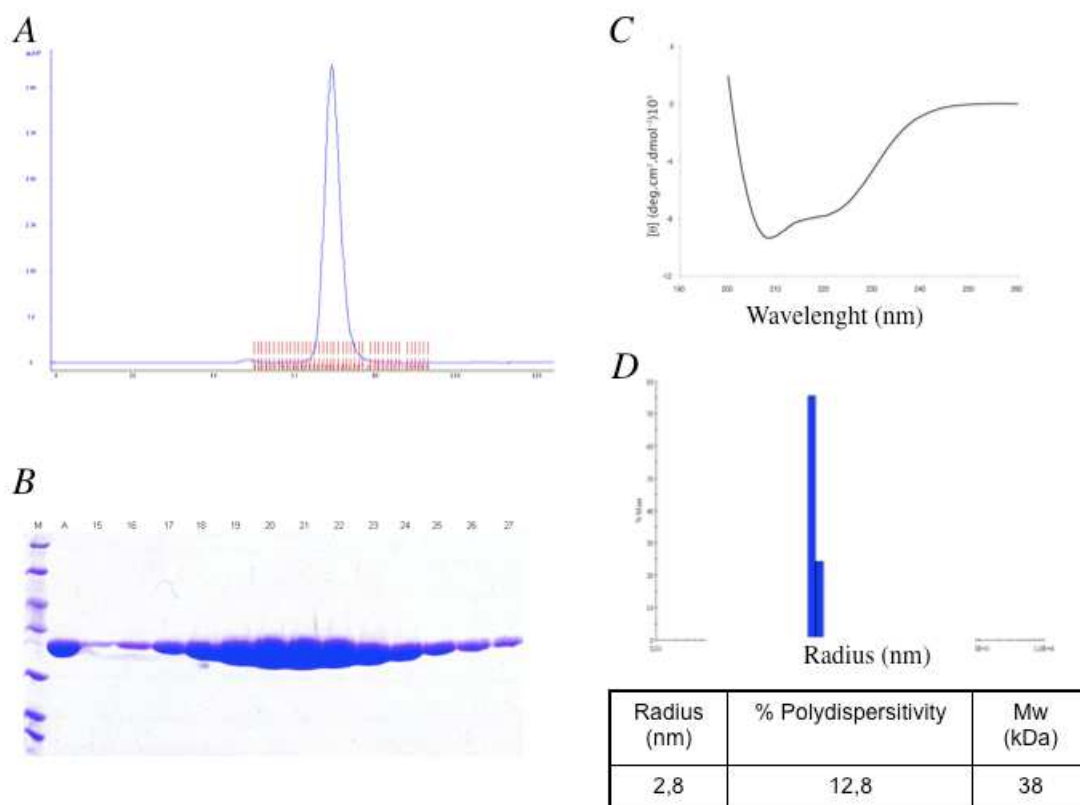
TABLE 1 – Experimental restraints and structural statistics for human SBDS.

Distance restraints		
All		2673
Intraresidue		617
Sequential		921
Medium range ($i+1 < i+4$)		556
Long range ($i > i+5$)		522
Hydrogen bonds		57
Residual target function, $\text{\AA}^2$		8.49 ± 1.12
Residual constraints violations		
Distance violations $> 0.2 \text{ \AA}$		0
Angle violation $> 5^\circ$		0
Van der Waals violation $> 0.5 \text{ \AA}$		0
Average Pairwise		
RMSD N-terminal domain (residues 9 to 95)		
Backbone atoms ($\text{\AA}$)		1.20 ± 0.21
All heavy atoms ($\text{\AA}$)		2.19 ± 0.33
RMSD central domain (residues 107 to 145, 151 to 167) ^a		
Backbone atoms ($\text{\AA}$)		1.91 ± 1.06
All heavy atoms ($\text{\AA}$)		3.09 ± 1.21
RMSD C-terminal domain (residues 173 to 236)		
Backbone atoms ($\text{\AA}$)		0.86 ± 0.08
All heavy atoms ($\text{\AA}$)		1.65 ± 0.12
RMSD for secondary structure residues ^b		
Backbone atoms ($\text{\AA}$)		0.45 ± 0.27
All heavy atoms ($\text{\AA}$)		1.43 ± 0.52
Ramachandran plot		
Most favored region (%)		75.4
Additionally allowed regions (%)		23.6
Generously allowed regions (%)		1.0
Disallowed regions (%)		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, 233-236.





Supplemental Figure S3 . HsSBDS biophysical characterization. *A*, size-exclusion chromatography profile, with collected fractions marked in red. *B*, SDS-PAGE electrophoresis (12,5% gel) of fractions from the size exclusion chromatography. *C*, Circular Dichroism plot of the purified protein. *D*, Dynamic Light Scattering plot of HsSBDS protein sample, confirming its monomeric form and molecular weight.

CAPÍTULO 4 - CONCLUSÕES E PERSPECTIVAS

A determinação da sequência correta da proteína SBDS de *Trypanosoma cruzi* permitiu a identificação de uma região adicional em seu C-terminal, presente também em ortólogas de plantas e demais tripanosomatídeos. Estudos estruturais e funcionais realizados com o C-terminal estendido da TcSBDS possibilitaram caracterizá-lo como um fragmento intrinsecamente desenovelado, capaz de realizar interações com RNA. Ensaio funcionais de interação com RNA ribossomal indicaram o envolvimento da TcSBDS na biossíntese de ribossomos, sugerindo que a função da SBDS é mantida em tripanossomatídeos. O fato de alguns aspectos do metabolismo de rRNA em *T. cruzi* serem diferentes dos demais eucariotos pode explicar a presença de uma região adicional na proteína SBDS de *T. cruzi*, possivelmente envolvida em interações adicionais e específicas com rRNA.

Ensaio de cristalização foram realizados utilizando-se uma construção truncada da proteína TcSBDS (na ausência da região C-terminal desenovelada), porém esta proteína se mostrou bastante suscetível a agregação e precipitação. Análises por proteólise limitada indicaram a ausência de um domínio compacto e estável na ortóloga de *T. cruzi* para a realização de ensaios de cristalização, impossibilitando o estudo estrutural dessa proteína por cristalografia de raios X.

A dificuldade encontrada em cristalizar as proteínas SBDS de levedura e humana, YLR022c e HsSBDS, e melhorar a ordem interna dos cristais da ortóloga de Archaea (PaUPF0023), se devem, provavelmente, à flexibilidade existente entre os domínios da proteína, sugerida pela estrutura cristalográfica das homólogas de *A. fulgidus* (Shammas *et al.*, 2005) e *M. thermotrophicus* (Long *et al.*, 2009) e confirmada pela estrutura em solução da SBDS humana resolvida por RMN como parte deste trabalho. A técnica de RMN permitiu não somente o estudo estrutural da HsSBDS, mas também a análise da sua dinâmica em solução e a investigação da sua interação com RNA, permitindo o mapeamento da região de ligação.

A estrutura da proteína SBDS humana, igualmente às homólogas de AfSBDS e mthSBDS, consiste de três diferentes domínios bem estruturados. Foi verificada a existência de mobilidade conformacional entre os domínios N-terminal e central da proteína e certa flexibilidade entre os domínios central e C-terminal. O movimento relativo entre os domínios também foi observado nos experimentos de dinâmica

realizados por RMN. Essa flexibilidade estrutural deve ser necessária para a interação simultânea da proteína com diferentes parceiros.

Na superfície eletrostática da proteína HsSBDS foram encontradas duas prováveis regiões de interação: no N-terminal existe uma região positivamente carregada, um possível alvo para a ligação de moléculas de RNA; e no C-terminal verifica-se uma região com carga negativa, podendo mediar interações com proteínas. Análises de interação utilizando-se RMN confirmaram a capacidade da proteína SBDS humana de interagir diretamente com RNA e possibilitaram o mapeamento da região de interação, que se sobrepõe à região positivamente carregada no N-terminal da proteína. No N-terminal também se encontram as principais mutações relacionadas à síndrome SDS, sugerindo uma possível relação entre a doença e a função da proteína de interação com RNA.

Esse trabalho abre oportunidade para a realização de estudos estruturais com as demais ortólogas da SBDS (de Archaea e levedura) através da técnica de RMN, uma vez que os protocolos de expressão e purificação dessas proteínas encontram-se padronizados e a estabilidade dessas proteínas foi verificada. Diferenças na estrutura da proteína de *P. abyssi* podem ser responsáveis por sua alta estabilidade térmica.

Análises de interação de proteínas da família SBDS com RNA e outros parceiros moleculares podem ser feitas por RMN. Embora o enovelamento das ortólogas seja bastante semelhante, a superfície eletrostática das proteínas de Archaea e humana são consideravelmente diferentes, inclusive na região de interação com RNA. Estudos devem ser feitos para analisar se a proteína de Archaea é capaz de interagir com RNA e se a interação ocorre na mesma região da proteína humana.

Estudos estruturais e de interação com RNA também devem ser feitos com mutantes da proteína SBDS humana, com o intuito de verificar se ocorrem alterações estruturais nas proteínas mutadas e, ainda, se as mutações relacionadas à síndrome humana SDS podem influenciar na interação da proteína com RNA.

REFERÊNCIAS BIBLIOGRÁFICAS

Abergel, C. (2004) Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. *Acta Cryst. D*, 60:1413-1416.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. **Molecular Biology of the Cell**. 4a ed., Garland Science, New York, 2002.

Austin KM, Leary RJ, Shimamura A. (2005) The Shwachman-Diamond SBDS protein localizes to the nucleolus. *Blood*, 106:1253-1258.

Ball HL, Zhang B, Riches J, Gandhi R, Li J, Rommens JM, Myers JS. (2009) SBDS is a multi-functional protein implicated in cellular stress responses. *Human Molecular Genetics*, doi:10.1093/hmg/ddp316.

Bergfors, T. (2003) Seeds to crystals. *Journal of Structural Biology*, 142:66-76.

Biertümpfel C, Basquin J, Suck D, Sauter C. (2002) Crystallization of biologicals molecules using agarose gel. *Acta Cryst. D*, 58:1657-1659.

Bodenhausen G, Ruben DJ. (1980) Natural abundance N-15 by enhanced heteronuclear spectroscopy. *Chem Phys Lett*, 69:185–189.

Bodian M, Sheldon W, Lightwod R. (1964) Congenital hypoplasia of the exocrine pancreas. *Acta Paediatr*, 53:282-293.

Bondos SE & Bicknell A. (2003) Detection and prevention of protein aggregation before, during, and after purification. *Analytical Biochemistry*, 316:223-231.

Boocock GR, Morrison J, Popovic M, Richards N, Ellis L, Durie P, Rommens J. (2003) Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet*, 33:97-10139.

Boocock GRB, Marit MR, Rommens JM. (2006) Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics*, 87:758-771.

Brooun A, Foster SA, Chrencik JE, Chien YET, Kolatkar AR, Streiff M, Ramage P, Widmer H, Weckbecker G, Kuhn P. (2007) Remedial strategies in structural

proteomics: Expression, purification, and crystallization of the Vav1/Rac1 complex. *Protein Expression and Purification*, 53:51-62.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gross P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst. D*, 54:905-921.

Carneiro FR, Silva TC, Alves AC, Haline-Vaz T, Gozzo FC, Zanchin NI. (2006) Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem. Biophys. Res. Commun.*, 343:260-268.

Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. (2007) His-tag impact on structure. *Acta Cryst. D*, 63:295-301.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6:227-293.

Dennis PP, Omer A, Lowe T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, 40:509-519.

Erdos M, Alapi K, Balogh I, Oroszlan G, Rakoczi E, Sumegi J, Marodi L. (2006) Severe Shwachman-Diamond syndrome phenotype caused by compound heterozygous missense mutations in the *SBDS* gene. *Exp Hematol*, 34:1517-1521.

Evans, J. **Biomolecular NMR Spectroscopy**. 2^a ed, New York, Oxford University Press Inc., 1996.

Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. (1995) Spectral density function mapping using ¹⁵N relaxation data exclusively. *Journal of biomolecular NMR*, 6:153-162.

Fatica A & Tollervey D. (2002) Making ribosomes. *Curr. Opin. Cell. Biol*, 14:313-318.

Fromont-Racine M, Senger B, Saveanu C, Fasiolo F. (2003) Ribosome assembly in

eukaryotes. *Gene*, 313:17-42.

Fu Z, Du Bois GC, Song SP, Harrison RW, Weber IT. (1999) Improving the diffraction quality of MTCP-1 crystals by post-crystallization soaking. *Acta Cryst. D*, 55:5-7.

Ganapathi KA, Austin KM, Lee CS, Dias A, Malsch MM, Reed R, Shimamura A. (2007) The human Shwachman-Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood*, 110: 1458-1465.

Gesteland RF, Cech TR, Atkins JF. **The RNA World**. 2a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999.

Güntert P, Mumenthaler C, Wüthrich K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, 273:283-298.

Güntert P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.*, 278:353-378.

Harp JM, Timm DE, Bunick GJ. (1998) Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Cryst. D*, 54:622-628.

Hesling C, Oliveira CC, Castilho BA, Zanchin NIT. (2007) The Shwachman-Bodian-Diamond syndrome associated protein interacts with HsNip7 and its down-regulation affects gene expression at the transcriptional and translational levels. *Experimental Cell Research*, 313:4180-4195.

Johnson B, Blevins RA. (1994) NMRView: A computer program for the visualization and analysis os NMR data. *J. Biomol. NMR*, 4:603-614.

Kiss T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, 20:3617-3622.

Lecompte O, Ripp R, Thierry J, Moras D, Poch O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.*, 30:5382-5390.

Leong Ng C, Waterman DG, Koonin EV, Walters AD, Chong JPJ, Isupov MN,

- Lebedev AA, Bunka DHJ, Stockley PG, Ortiz-Lombardía M, Antson AA. (2009) Conformational flexibility and molecular interactions of an archaeal homologue of the Shwachman-Bodian-Diamond syndrome protein. *BMC Structural Biology*, doi:10.1186/1472-6807-9-32.
- Li MX, Corson DC, Sykes BD. (2002) Structure Determination by NMR: Isotope Labeling. *Methods in Molecular Biology*, Human Press Inc., Totowa, NJ, Unites States, 173:255-265.
- Liu JM, Elli SR. (2006) Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood*, 107:4583-4588.
- Luz JS, Georg RC, Gomes CH, Machado-Santelli GM, Oliveira CC. (2009) *Sdo1p*, the yeast orthologue of Shwachman-Bodian-Diamond syndrome protein, binds RNA and interacts with nuclear rRNA-processing factors. *Yeast*, 26:287-298.
- Maden BE. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, 39:241-303.
- Maki JA, Schnobrich DJ, Culver GM. (2002) The DnaK chaperone system facilitates 30S ribosomal subunit assembly. *Mol. Cell*, 10:129-138.
- Menne TF, Goyenechea B, Sanchez-Puig N, Wong CC, Tonkin LM, Ancliff PJ, Brost RL, Costanzo M, Boone C, Warren AJ. (2007) The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat Genet*, 39:486-495.
- Nakashima E, Mabuchi A, Makita Y, Masuno M, Ohashi H, Nishimura G, Ikegawa S. (2004) Novel SBDS mutations caused by gene conversion in Japanese patients with Shwachman-Diamond syndrome. *Hum Genet*, 114:345-348.
- Newman J. (2006) A review of techniques for maximizing diffraction from a protein crystal in stilla. *Acta Cryst. D*, 62:27-31.
- Nierhaus KH. (1991) The assembly of prokaryotic ribosomes. *Biochimie*, 73:739-755.
- Pervushin K, Riek R, Wider G, Wüthrich K. (1997) Attenuated T₂ relaxation by mutual

cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.*, 94:12366-71.

Renzone G, Vitale RM, Scaloni A, Rossi M, Amodeo P, Guagliardi A. (2007) Structural characterization of the functional regions in the archaeal protein Sso7d. *Proteins*, 67:189-197.

Roberts, G. **NMR of Macromolecules – A Practical Approach**. 2^a ed, New York, Oxford University Press Inc., 1995.

Rodger A & Norden B. Circular dichroism & linear dichroism. Oxford University Press, 1997.

Rozenski J, Crain PF, McCloskey JA. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, 27:196-197.

Saridakis E & Chayen NE. (2000) Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. *Protein Science*, 9:755-757.

Sattler M, Fesik SW. (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure*, 4:1245-1249.

Sattler M, Schleucher J, Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in NMR Spectroscopy*, 34:93-158.

Sattler M, Simon B. (2002) NMR of large proteins. *The ELSO Gazette*, 11.

Savchenko A, Krogan N, Cort JR, Evdokimova E, Lew JM, Yee AA, Sanchez-Pulido L, Andrade MA, Bochkarev A, Watson JD, Kennedy MA, Greenblatt J, Hughes T, Arrowsmith CH, Rommens JM, Edwards AM. (2005) The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism. *J Biol Chem*, 280:19213-19220.

Shammas C, Menne TF, Hilcenko C, Michell SR, Goyenechea B, Boocock GRB, Durie PR, Rommens JM, Warren AJ. (2005) Structural and mutational analysis of the

SBDS protein family. Insight into the leukemia-associated Shwachman-Diamond Syndrome. *J Biol Chem*, 280:19221-19229.

Shwachman H. (1964) The syndrome of pancreatic insufficiency and bone marrow dysfunction. *J Pediatr*, 65:645-663.

Slupsky CM, Boyko RF, Booth VK, Sykes BD. (2003) Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments. *J. Biomol. NMR*, 27:313-21.

Tollervey D & Kiss T. (1997) Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell. Biol*, 9:337-342.

Tschochner H & Hurt E. (2003) Pre-ribosomes on the road from the nucleolus to the cytoplasm. *TRENDS in Cell Biology*, 13:255-263.

Venema J & Tollervey D. (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annual Review of Genetics*, 33:261-311.

Wishart DS, Sykes BD, Richards FM. (1992) The Chemical Shift Index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31:1647-1651.

Wishart DS & Sykes BD. (1994) The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J. Biomol. NMR*, 2:171-80.

Woloszynek JR, Rothbaum RJ, Rawls AS, Minx PJ, Wilson RK, Mason PJ, Bessler M, Link DC. (2004) Mutations of the *SBDS* gene are present in most patients with Shwachman-Diamond syndrome. *Blood*, 104:3588-3590.

Wool IG. (1979) The structure and function of eukaryotic ribosomes. *Ann. Rev. Biochem.*, 48:719-754.

Wüthrich K. **NMR of Proteins and Nucleic Acids**. John Wiley & Sons, New York, 1986.

Yan SZ, Beeler JA, Chen Y, Shelton RK, Tang WJ. (2001) The Regulation of Type 7

Adenylyl Cyclase by Its C1b Region and *Escherichia coli* Peptidylprolyl Isomerase, SlyD. *The Journal of Biological Chemistry*, 276:8500-8506.

Zeri AC, Mesleh MF, Nevzorov AA, Opella SJ. (2003) Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. *PNAS*, 100:6458-6463.

Zhang S, Shi M, Hui C, Rommens JM. (2006) Loss of the mouse ortholog of the shwachman-diamond syndrome gene (*Sbds*) results in early embryonic lethality. *Mol Cell Biol*, 26:6656-6663.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação/tese de Mestrado/Doutorado intitulada "**Estudos Estruturais e Funcionais de Proteínas da Família SBDS com Ênfase nas Ortólogas de *Trypanosoma cruzi* e Humana**":

() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

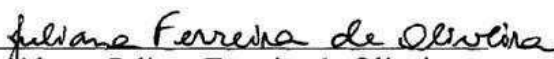
() está inserido no **Projeto CIBio/IB/UNICAMP** (Protocolo nº _____), intitulado _____;

() tem autorização da **Comissão de Ética em Experimentação Animal/IB/UNICAMP** (Protocolo nº _____);

() tem autorização do **Comitê de Ética para Pesquisa com Seres Humanos/FCM/UNICAMP** (Protocolo nº _____);

(X) tem autorização de comissão de bioética ou biossegurança externa à UNICAMP.

Especificar: Comissão Interna de Biossegurança da ABTLuS – Associação Brasileira de Tecnologia de Luz Síncrotron (Processo nº BGG03.02).

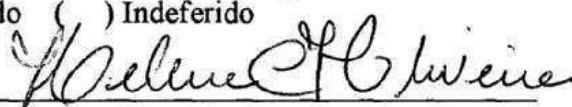

Aluno: Juliana Ferreira de Oliveira


Orientador: Ana Carolina de Mattos Zeri

Para uso da Comissão ou Comitê pertinente:

(X) Deferido () Indeferido

Nome:

Função: 
Prof. Dra. HELENA COUTINHO F. DE OLIVEIRA
Presidente
Comissão Interna de Biossegurança
CIBio/IB - UNICAMP