

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

IVAN ROSA E SILVA

Caracterização bioquímica e biofísica de proteínas específicas  
envolvidas no SL *trans-splicing* de *Trypanosoma brucei*

São Carlos

2016



UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

IVAN ROSA E SILVA

Caracterização bioquímica e biofísica de proteínas específicas  
envolvidas no SL *trans-splicing* de *Trypanosoma brucei*

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada  
Opção: Física Biomolecular  
Orientador: Prof. Dr. Otavio Henrique Thiemann

Versão Corrigida

São Carlos

2016

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação do IFSC,  
com os dados fornecidos pelo(a) autor(a)

Silva, Ivan Rosa e  
Caracterização bioquímica e biofísica de proteínas  
específicas envolvidas no SL trans-splicing de  
Trypanosoma brucei / Ivan Rosa e Silva; orientador  
Otavio Henrique Thiemann - versão corrigida -- São  
Carlos, 2016.  
238 p.

Tese (Doutorado - Programa de Pós-Graduação em  
Física Biomolecular) -- Instituto de Física de São  
Carlos, Universidade de São Paulo, 2016.

1. SL trans-splicing. 2. Spliceossomo. 3.  
Proteínas Sm. 4. U5-200K. 5. Trypanosoma brucei. I.  
Thiemann, Otavio Henrique , orient. II. Título.

À minha família, pelo apoio ao longo do  
período de realização deste trabalho.



## AGRADECIMENTOS

Aos meus pais Benedito e Amarilda, e à minha irmã Iara, pelo apoio em todos os momentos, em nome de quem agradeço a todos da minha família.

Ao Prof. Dr. Otavio Henrique Thiemann, pela confiança, orientação e discussões construtivas durante a realização deste trabalho no Laboratório de Biologia Estrutural do Instituto de Física de São Carlos da Universidade de São Paulo (IFSC-USP), em São Carlos-SP.

Ao Dr. Kiyoshi Nagai, pela confiança, orientação e discussões construtivas na realização de parte deste trabalho durante estágio sanduíche no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido (LMB-MRC), em Cambridge, Inglaterra.

Ao Dr. Marco Túlio Alves da Silva, pela amizade, pelas discussões construtivas desde o começo do projeto, pelos ensinamentos e pela revisão de toda a tese.

Aos alunos de iniciação científica Camila e Guilherme, que acreditaram no projeto da U5-200K de *T. brucei*, pela dedicação, parceria e discussões construtivas.

A Natália, Geraldo, Vitor (B-2) e Fernanda, pelas discussões, incentivos e trocas de experiências. A Jéssica, Geraldo e Atílio, pela revisão de parte desta tese.

Aos técnicos dos laboratórios dos Grupos de Cristalografia e de Biofísica do IFSC-USP e do Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido (LMB-MRC), pelo apoio nos experimentos.

Aos professores do IFSC-USP, pelos ensinamentos e pelas discussões construtivas.

Ao Prof. Dr. Mario de Oliveira Neto da UNESP-Botucatu, pela discussão detalhada dos dados de SAXS das proteínas Sm durante e após o exame de qualificação de doutorado. Aos professores Dr. João Renato Carvalho Muniz e Dra Regina Maria Barretto Cicarelli, pelas contribuições neste mesmo exame.

A Chris O., pela ajuda com os experimentos de difração de raios-X no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido, em Cambridge.

Aos membros do grupo de pesquisa do Prof. Dr. Otavio H. Thiemann: Túlio, Teresa, Lívia Faim, Daiana, Fer, Mateus, Vitor (B-2) (antigos), Natália, Ana Laura, Alan, Thomás, Adriano, Jéssica, Renata, Camila (novos), em nome de quem agradeço a todos os amigos do laboratório.

Aos membros do grupo do spliceossomo, orientados pelo Dr. Kiyoshi Nagai: Lisa, Chris O., Sebastian, Kelly, Wojtek, Pei-Chun, Andy, Chris N., Max, Song, Lucas e Bennie.

Aos amigos da BIOMOL, pela amizade, pela união nos momentos difíceis, em nome de quem agradeço a todos aqueles que me acompanharam durante a minha formação.

Às funcionárias do serviço de Biblioteca e Informação do IFSC-USP, pela prestatividade em todos os momentos.

Ao Grupo de Biologia Molecular Estrutural e de Cristalografia de Proteínas do IFSC-USP, pela oportunidade de realização deste trabalho.

Ao Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido (LMB-MRC), pela oportunidade de realização de estágio de 9 meses e confiança no meu trabalho.

Ao Instituto de Física de São Carlos (IFSC) e à Universidade de São Paulo (USP), pela oportunidade de realização do projeto de pesquisa de doutorado.

Ao Laboratório Nacional de Luz Síncrotron (LNLS) – Campinas-SP, pela oportunidade de realização de experimentos de espalhamento de raios-X a baixo ângulo (SAXS) na linha SAXS2 e à sua equipe técnica.

Ao laboratório de luz síncrotron *Diamond Light Source* (DLS), Reino Unido, pela oportunidade de realização de experimentos de difração de raios-X em monocristais na linha I24 e à sua equipe técnica.

À CAPES e ao CNPq (processo 205627/2014-5), pela concessão de bolsas de doutorado e de doutorado sanduíche pelo Programa Ciência sem Fronteiras, respectivamente, e pelo apoio financeiro para a realização desta pesquisa.

Enfim, a todos que contribuíram de forma direta ou indireta para a realização deste trabalho.



“Quase todos os aspectos da vida são projetados no nível molecular, e sem o conhecimento das moléculas só podemos ter uma compreensão muito superficial da vida.”

Francis Crick



## RESUMO

SILVA, Ivan Rosa e. **Caracterização bioquímica e biofísica de proteínas específicas envolvidas no SL trans-splicing de *Trypanosoma brucei***. 2016. 238 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

O SL *trans-splicing* (do inglês, *spliced leader trans-splicing*) catalisado pelo spliceossomo em *Trypanosoma brucei* é responsável pelo processamento dos pré-mRNAs policistrônicos em mRNAs maduros. Esta maquinaria é associada a partir de pequenas partículas ribonucleoproteicas nucleares (snRNPs) U1, U2, U4/U6 e U5 constituídas de pequenos RNAs nucleares (snRNAs), um complexo canônico de sete proteínas Sm (SmB, SmD3, SmD1, SmD2, SmE, SmF e SmG) e fatores proteicos específicos. O núcleo de proteínas Sm de *T. brucei* apresenta variações com funções desconhecidas, como a substituição do heterodímero SmD3/SmB por Sm16,5K/Sm15K na snRNP U2, e de SmD3 por SSm4 na snRNP U4. Na primeira parte deste trabalho, investigou-se a interação destes diferentes complexos Sm recombinantes com os snRNAs U2, U4 e U5 obtidos por transcrição *in vitro*. Todos os complexos apresentaram alta afinidade pelo snRNA cognato. Observou-se, ainda, que apenas o núcleo Sm que contém Sm16,5K/Sm15K associado ao snRNA U2 interage com alta afinidade com U2A'/U2B''. Adicionalmente, foi obtida a estrutura cristalográfica de U2A'/U2B'' de *T. brucei*, que revela uma organização similar àquela já descrita para ortólogas de *Homo sapiens*. Entretanto, há um desvio de pelo menos 6 Å no ponto médio da alça carregada positivamente no domínio RRM de U2B'' para a acomodação do snRNA U2. Além disso, observou-se uma longa hélice- $\alpha$  adicional na extremidade C-terminal de U2A'. A análise dos três núcleos Sm de *T. brucei* a partir da combinação de modelagem molecular e espalhamento de raios-X a baixo ângulo revela estruturas de barril- $\beta$  altamente torcido com interior carregado positivamente para interação com snRNAs. A principal diferença entre as estruturas encontra-se nas extremidades C- e N-terminal dos variantes de proteínas Sm, possivelmente para interação com U2A' no braço 1 do spliceossomo, no caso de Sm15K/Sm16,5K, e com U5-220K e U5-200K no corpo do spliceossomo, no caso de SSm4. Na segunda parte deste trabalho, a expressão homóloga da proteína U5-200K de *T. brucei* completa e do produto truncado no seu cassete helicase/ATPase/Sec63 N-terminal levou à copurificação de um subcomplexo de snRNP U5 composto por U5-220K, U5-116K, U5-40K e U5-Cwc21, sendo que a proteína recombinante completa ainda copurificou as proteínas Sm. Experimentos de imunolocalização mostraram que a proteína U5-200K truncada não é direcionada ao núcleo, como é o caso da proteína completa. As células que expressam a proteína truncada apresentaram um defeito de crescimento significativo, e os processamentos de pré-mRNA por *cis*- e SL *trans-splicing* foram ligeiramente afetados, já que a proteína truncada não entra no núcleo, onde deveria exercer sua atividade. Os resultados apresentados indicam a formação de um subcomplexo de snRNP U5 ainda no citoplasma, sendo que as proteínas Sm devem ser um sinal para o seu transporte nuclear mediado por importina- $\beta$ . Em leveduras, a proteína Aar2 substituiu U5-200K no citoplasma, regulando assim a biogênese de snRNP U5, porém esta proteína não foi identificada em *T. brucei*. Os resultados apresentados neste trabalho contribuem como o primeiro estudo estrutural de proteínas spliceossomais de um parasita do homem e também com novas informações sobre a biogênese das partículas ribonucleoproteicas U2 e U5 de *T. brucei*.

Palavras-chave: SL *trans-splicing*. Spliceossomo. Proteínas Sm. U5-200K. *Trypanosoma brucei*.



## ABSTRACT

SILVA, Ivan Rosa e. **Biochemical and biophysical characterization of specific proteins involved in *Trypanosoma brucei* SL trans-splicing.** 2016. 238p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

The spliced-leader (SL) trans-splicing catalyzed by the spliceosome in *Trypanosoma brucei* is responsible for processing polycistronic pre-mRNAs into mature mRNAs. The spliceosome machinery is assembled by small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/U6 and U5 that are composed by small nuclear RNAs (snRNAs), a canonical complex of Sm proteins (SmB, SmD3, SmD1, SmD2, SmE, SmF, SmG) and specific factors. The Sm core peculiarly varies in *T. brucei*, where SmD3/SmB are replaced by Sm16.5K/Sm15K in U2 snRNP and SmD3 is substituted by SSm4 in U4 snRNP. In the first part of this thesis, we investigated the interaction of the different recombinant Sm cores with U2, U4 and U5 snRNAs obtained by *in vitro* transcription. All the protein complexes bind the cognate snRNA with high affinity. Only the Sm core that contains Sm16.5K/Sm15K associated with U2 snRNA interacts with the recombinant U2A'/U2B'' subcomplex. Additionally, the crystallographic structure of *T. brucei* U2A'/U2B'' was obtained, showing an overall organization similar to the one observed in the human counterpart. However, we observed a 6 Å deviation in the medium point of a positively charged turn in the RRM motif of U2B'' to accommodate U2 snRNA. Besides, a long  $\alpha$ -helix was observed in the C-terminal region of U2A'. Structural analysis of Sm core variations in *T. brucei* was proceeded using molecular modelling techniques associated with small angle X-ray scattering. The quaternary structure models show seven Sm proteins as  $\beta$ -barrels with positively charged interior for cognate snRNA interaction. The main difference among these Sm core structures resides in the C- and N-terminal regions of the variant proteins, probably enabling the interaction of Sm15K/Sm16,5K with U2A' in the spliceosome's arm 1, and the association of SSm4 with U5-220K and U5-200K in the spliceosome's body. In the second part of this thesis, homologous expression of full-length and N-terminally truncated U5-200K from *T. brucei* led to the copurification of a U5 snRNP subcomplex containing U5-220K, U5-116K, U5-40K and U5-Cwc21. The full-length U5-200K construct also copurified Sm proteins. Immunolocalization experiments showed that the truncated U5-200K protein is not directed to the nucleus as is the case for the full-length protein. Cells that expressed the truncated protein showed a significant growth defect and the pre-mRNA processing by cis- and SL trans-splicing was negatively affected since the truncated protein did not enter the nucleus where it should be active. The results suggest that a subcomplex of U5 snRNP begins to be assembled in the cytoplasm and the Sm proteins may be the signal for the nuclear transport mediated by  $\beta$ -importin. In yeast, Aar2 replaces U5-200K in the cytoplasm in another regulation step. However, Aar2 has not been identified in *T. brucei*. The results presented here contribute with the first structural study of spliceosomal proteins of a human parasite and give new insights into the biogenesis of U2 and U5 snRNPs in *T. brucei*.

Keywords: SL *trans-splicing*. Spliceosome. Sm proteins. U5-200K. *Trypanosoma brucei*.



## LISTA DE FIGURAS

- Figura 1.1 - Distribuição geográfica da doença do sono em 2014.** A – *T. brucei gambiense* e B – *T. brucei rhodesiense*..... 34
- Figura 1.2 - Distribuição geográfica da doença de Chagas entre 2006 e 2009** ..... 36
- Figura 1.3 - Distribuição geográfica dos novos casos de leishmaniose em 2013.** A – Leishmaniose visceral e B – Leishmaniose cutânea..... 37
- Figura 1.4 - Ciclo de vida do *T. brucei*** – 1- A forma tripomastigota metacíclica do *T. brucei* é injetada no hospedeiro humano pela mosca tsé-tsé. 2-4- *T. brucei* se transforma em tripomastigota sanguíneo, atinge outros fluidos corporais (linfa, líquido) e continua sua replicação por fissão binária. 5- A mosca tsé-tsé se infecta com as formas tripomastigotas sanguíneas. 6- No intestino médio da mosca, os parasitas se transformam em tripomastigotas procíclicos e se multiplicam por fissão binária. 7- Após deixar o intestino médio da mosca, os parasitas se transformam em epimastigotas e atingem sua glândula salivar (8), onde se reproduzem por fissão binária. *T. brucei*, então, transforma-se em tripomastigota metacíclico. O inseto infectado transmite o protozoário para novo hospedeiro humano (1), dando continuidade ao ciclo de vida do *T. brucei*..... 40
- Figura 1.5 – Arquitetura da célula do *T. brucei*.** A- Organelas e B- Posição do cinetoplasto e do flagelo nas diferentes formas celulares..... 41
- Figura 1.6 – Representação esquemática do genoma nuclear de *T. brucei*.** Estão representados os elementos genéticos de acordo com sua classe (rRNA, prociclina, CDS, SLRNA, snoRNA, tRNA, snRNA, SRP-RNA) e a RNA polimerase que os transcreve (RNA polimerase I, II ou III). Barras coloridas indicam a posição e o tamanho de cada elemento genético de cada cromossomo de *T. brucei*. Barras acima da linha preta que representa um cromossomo (C1-C11) correspondem à transcrição no sentido da esquerda para a direita e barras abaixo correspondem à transcrição no sentido contrário..... 43
- Figura 1.7 – Sinais de *splicing* no pré-mRNA.** A- Representação esquemática dos sinais de *splicing* consenso (em vermelho) no pré-mRNA de *T. brucei*, de mamíferos e de leveduras. B- Estrutura secundária proposta para RNA SL de *T. brucei*. Em roxo, identifica-se o éxon do RNA SL, que apresenta um capacete 5' de 7-metilguanosina hipermetilado (cap4). O sítio Sm do RNA SL está destacado, bem como a sequência GU do 5'SS..... 45
- Figura 1.8 – Mecanismos de *splicing* em eucariotos.** A1 e A2- Reações de SL *trans-splicing* e *cis-splicing*, respectivamente. B- Dinâmica de associação do spliceossomo durante o *cis-splicing*..... 49
- Figura 1.9 – Catálise das reações de transesterificação (A e B) a partir do posicionamento de íons  $Mg^{2+}$  pelo snRNA U6 para estabilizar os intermediários das reações**..... 50

<b>Figura 1.10 – snRNP U1.</b> Estruturas secundárias de snRNAs U1 de <i>T. brucei</i> e <i>H.sapiens</i> e modelo cristalográfico de snRNP U1 de <i>H. sapiens</i> <sup>55</sup> . Os sítios Sm estão evidenciados em vermelho.....	52
<b>Figura 1.11 – snRNP U2.</b> Estruturas secundárias de snRNAs U2 de <i>T. brucei</i> e <i>H. sapiens</i> e densidade eletrônica do núcleo de snRNP U2 de <i>S. pombe</i> a ~10Å de resolução obtida por criomicroscopia eletrônica <sup>54</sup> . Os sítios Sm estão evidenciados em vermelho.....	54
<b>Figura 1.12 – Tri-snRNP U4/U6.U5.</b> A1 e A2-Estrutura de criomicroscopia eletrônica da tri-snRNP U4/U6.U5 de <i>S. cerevisiae</i> <sup>56</sup> . B- Estruturas secundárias de snRNA U5 de <i>T. brucei</i> (B1) e <i>H. sapiens</i> (B2) e de di-snRNAs U4/U6 de <i>T. brucei</i> (C1) e <i>H. sapiens</i> (C2). Os sítios Sm estão evidenciados em vermelho.....	57
<b>Figura 1.13 – Dinâmica de associação do spliceossomo.</b> A- Composição das snRNPs e das proteínas spliceossomais de humanos. Estão grifadas as proteínas que apresentam homólogos em <i>T. brucei</i> e destacadas em vermelho as proteínas únicas deste organismo. B- Representação esquemática da composição do spliceossomo a cada etapa da catálise do <i>splicing</i> . O código de cores utilizado em A foi mantido. C- Rede de interações de moléculas de snRNAs a cada etapa.....	62
<b>Figura 1.14 – Representação esquemática da relação filogenética entre eucariotos,</b> evidenciando organismos onde o mecanismo de <i>splicing</i> está bem caracterizado (ponto vermelho) e os clados contendo organismos com spliceossomo reduzido (azul). Clados que contêm organismos onde o mecanismo de <i>trans-splicing</i> foi caracterizado são indicados em verde. Cm - <i>C. merolae</i> , Ec - <i>E. cuniculi</i> , Gl - <i>G. lamblia</i> , Tb - <i>T. brucei</i> .....	65
<b>Figura 2.1 – Conservação de aminoácidos em proteínas da família LSm.</b> Análise de conservação de aminoácidos a partir de alinhamento múltiplo de sequências de aminoácidos de proteínas LSm usando o modelo oculto de Markov (do inglês, <i>hidden Markov model</i> , HMM). O eixo vertical corresponde ao conteúdo relativo de informação medido em <i>bits</i> para cada posição na sequência de aminoácidos. Cilindros correspondem à estrutura secundária na forma de hélice- $\alpha$ (H) e triângulos à fita- $\beta$ (B). Alças (L) são destacadas em azul. Regiões de alta variabilidade de sequência de aminoácidos são evidenciadas em cinza.....	76
<b>Figura 2.2 – Estruturas terciária e quaternária de proteínas Sm/LSm.</b> A- Dímero de SmD1/SmD2 (código PDB 1B34) <sup>174</sup> , B- Hexâmero de Hfq de <i>E. coli</i> (código PDB 3GIB) <sup>20</sup> , C- Heptâmero de SmAP1 (código PDB 1LNx) <sup>180</sup> de <i>P. aerophilum</i> , D- Octâmero de LSm3 de <i>S. cerevisiae</i> (código PDB 3BW1) <sup>182</sup> e E- Trímero de LSm4 de <i>S. pombe</i> (código PDB 4EMH) <sup>178</sup> .....	78
<b>Figura 2.3 – Interação de proteínas Sm com snRNAs.</b> Comparação das estruturas dos núcleos Sm de snRNPs U4 (A) e U1 (B) de <i>H. sapiens</i> em interação com o snRNA cognato.....	79
<b>Figura 2.4 – Interação de U2A'/U2B'' com o snRNA U2 de H. sapiens.</b> Estrutura cristalográfica de U2A' (vermelho)/U2B'' (azul) <sup>68</sup> interagindo com a estrutura SL4 do snRNA U2 de <i>H sapiens</i> .....	82



**Figura 2.5 – Método de difusão de vapor e o diagrama de fases.** A- Representação esquemática de um experimento típico de difusão de vapor a partir de gota sentada (*sitting drop*). B- Diagrama de fases mostrando a curva de solubilidade de uma proteína e as regiões de nucleação espontânea, de crescimento cristalino e de formação de precipitado amorfo. Inicialmente, iguais concentrações de proteína e agentes precipitantes são misturadas e vapor de solvente se difunde até o reservatório do coquetel de precipitantes, de modo que a concentração de proteína e de precipitantes aumenta na gota ao passo que seu volume diminui. Uma vez que a região de nucleação espontânea é atingida, agregados cristalinos na forma de núcleos são formados e um cristal inicia o processo de crescimento. O número de moléculas de proteína em solução passa a diminuir devido ao crescimento do cristal e a solução supersaturada tende a se equilibrar com o(s) cristal(is), que atingem o tamanho máximo..... 95

**Figura 2.6 – Alinhamento múltiplo de sequências de aminoácidos de proteínas Sm de *H. sapiens* e *T. brucei*** feita a partir do servidor *Clustal Omega*<sup>194</sup>. Resíduos de aminoácidos evidenciados: hidrofóbicos em azul, Gly em amarelo, sítios de ligação a nucleotídeo: Asp ou Glu em rosa, aromáticos em roxo, Asn em verde, e Arg ou Lys em azul. Resíduos não conservados nos motivos Sm1 e Sm2 são destacados em vermelho. Quatro resíduos de aminoácidos que interagem diretamente com o sítio Sm no snRNA são destacados com o símbolo \*. Estrutura secundária: hélice- $\alpha$  em azul e fita- $\beta$  em rosa. A hélice- $\alpha 0$  é predita apenas para SSm4 e SmD2 de *T. brucei*..... 107

**Figura 2.7 – Comparação entre resíduos de aminoácidos conservados nos domínios Sm1 e Sm2 de proteínas Sm de *T. brucei* e a posição de interação no sítio Sm de snRNAs U1, U2, U4 e U5.** Análise dos sítios Sm no núcleo de snRNPs A- U1, B- U2 e C- U4, baseada nas estruturas cristalográficas de núcleos de snRNPs U1 e U4 de *H. sapiens*, e D- de conservação dos sítios Sm nos snRNAs U1, U2, U4 e U5 de *T. brucei* e *H. sapiens*. Em vermelho, são destacados os resíduos que variam nas proteínas Sm ortólogas de *T. brucei* e *H. sapiens*. O anel interno corresponde ao sítio Sm no snRNA. O próximo anel corresponde aos resíduos das proteínas Sm de *H. sapiens*, enquanto os seguintes correspondem às variantes em *T. brucei*..... 109

**Figura 2.8 – Alinhamento múltiplo de sequências de aminoácidos das proteínas U2A' e U2B'' de *H. sapiens* e *T. brucei*** feita a partir do servidor *Clustal Omega*<sup>194</sup> e comparação entre snRNAs U1 e U2. A- U2A', B- U2B''. Em azul estão evidenciados os resíduos de aminoácidos conservados característicos do motivo que determina a proteína. Em vermelho, são mostrados os resíduos não conservados nas posições determinantes dos motivos característicos. Estrutura secundária: hélice- $\alpha$  em azul e fita- $\beta$  em rosa. C- Comparação entre alças SL1 e SL2 de snRNAs U1 e U2, respectivamente. Em vermelho, são mostrados os nucleotídeos tipicamente conservados na primeira metade da alça de cada snRNA. Em azul, o primeiro nucleotídeo da segunda metade e, em verde, o par de nucleotídeos que fecha a alça..... 113

- Figura 2.9 – Perfis cromatográficos de purificação de proteínas Sm de *T. brucei*.** A- Confirmação de clonagem de SSm4/SmB. Cromatografia de Afinidade: A – SSm4/SmB, B- SmD1/SmD2, C- SmE/SmF/SmG, D- SmD3/SmB, E- Sm15K/Sm16,5K. F-B- Cromatografia de exclusão molecular dos sub-complexos Sm de *T. brucei*. C-G- SDS-PAGE 15%/Tricina dos subcomplexos SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB, Sm15k/Sm16,5K, SSm4/SmB de *T. brucei* após última etapa de purificação. MM – marcador de massa molecular. H- e I- Calibração da coluna SUPEROSE 12 10/300 (GE) para massa molecular e raio hidrodinâmico usando os seguintes padrões: citocromo c equino (MM = 12,4 kDa e  $R_H = 17,0 \text{ \AA}$ ), anidrase carbônica bovina (MM = 29,0 kDa e  $R_H = 23,6 \text{ \AA}$ ), BSA (MM = 66,0 kDa  $R_H = 33,9 \text{ \AA}$ ) e álcool desidrogenase de levedura (MM = 150,0 kDa e  $R_H = 46,0 \text{ \AA}$ ). O dextran azul é eluído em volume igual a 7,9 mL, o que permitiu calcular  $K_{av} = (V - V_{dextran}) / (24 \text{ mL} - V_{dextran})$ . Em todos os géis, a primeira coluna corresponde ao marcador de massa molecular, a segunda ao lisado, e da terceira em diante a lavagem e o eluído no gradiente de tampão B é indicado pela seta dupla vermelha..... 115
- Figura 2.10 – U2A'/U2B''.** A- Confirmação de clonagem em vetor pET28a por meio de análise de produto de restrição: o fragmento de 1122 pb corresponde à sequência que codifica para U2A'/U2B''. B- Cromatografia de afinidade a íons  $Ni^{2+}$  do subcomplexo U2A'/U2B'' recombinante. Proteínas foram eluídas entre 15 e 30% de tampão B, como observado em SDS-PAGE (C). D- Perfil de eluição de U2A'/U2B'' em resina de hidroxiapatita e SDS-PAGE (E) correspondente. As setas duplas vermelhas indicam os produtos eluídos em cada caso..... 119
- Figura 2.11 – Transcrição *in vitro* de snRNAs U2, U4 e U5.** A- Produtos de PCR dos moldes de snRNAs U2, U4 e U5 confirmando a clonagem em pUC18. B- Bandas únicas para snRNAs U2, U4 e U5 obtidos por transcrição *in vitro* observadas em eletroforese desnaturante (PAGE 15%/uréia)..... 120
- Figura 2.12 – Enovelamento de snRNAs U2 (A1, A2 e D1), U4 (B1, B2 e D2) e U5 (C1, C2 e D3) medido por espectroscopia de dicróismo circular.** São apresentados os espectros de dicróismo circular medidos entre 210 e 320 nm e gráfico de componente principal 1 *versus* temperatura. As bandas apresentadas como inserção correspondem à eletroforese em gel nativo (1% agarose), corado com brometo de etídeo ( $1 \mu\text{g mL}^{-1}$ ). As setas vermelhas indicam o aparecimento de mínimo e máximo locais de dicróismo circular com a diminuição de temperatura. As setas azuis indicam as bandas únicas em eletroforese nativa..... 120
- Figura 2.13 - Ensaios de interação de U2A'/U2B'' e núcleo Sm 2 ao snRNA U2.** A- Titulação de U2A'/U2B'' ao snRNA marcado com fluoresceína e B- ajuste da equação de Hill às medidas de fluorescência das bandas. C- Titulação de núcleo Sm 2 ao snRNA marcado com fluoresceína e D- ajuste da equação de Hill aos dados experimentais. .... 123
- Figura 2.14 - Formação do núcleo de snRNP U2.** A- Titulação de núcleo Sm 2 ao subcomplexo pré-formado contendo U2A'/U2B'' e snRNA U2. B- Titulação de U2A'/U2B'' ao núcleo Sm 2 pré-formado em interação com snRNA U2 ..... 124
- Figura 2.15 - Ensaios de interação de proteínas Sm com snRNA U2.** Titulação de proteínas dos núcleos A- Sm 1 e B- Sm 3 ao snRNA U2. Titulação de U2A'/U2B'' nos núcleos C- Sm 1 e D- Sm 2 pré-formados..... 125

<b>Figura 2.16 - Formação dos núcleos Sm 1 e 3.</b> Titulação de proteínas do núcleo Sm A-1 e B-3 em snRNAs U5 e U4, respectivamente. Ajuste dos dados experimentais ao modelo de Hill <sup>200</sup> para os núcleos D- Sm 1 e E- Sm3.....	126
<b>Figura 2.17- Estequiometria de interação entre os componentes do núcleo de snRNP U2.</b> A- Titulação de proteínas do núcleo Sm 2 ao snRNA U2. B- Titulação de U2A'/U2B'' ao snRNA U2. C- Titulação de U2A'/U2B'' ao núcleo Sm 2 pré-formado em interação com snRNA U2. As setas vermelhas indicam razão estequiométrica 1:1.....	127
<b>Figura 2.18 - Cromatografia de exclusão molecular dos núcleos Sm de <i>T. brucei</i> na ausência de snRNAs.</b> A- complexo 1 (Sm 1, canônico), B- complexo 2 (Sm 2) e C- complexo 3 (Sm 3). Para cada resultado de purificação, o perfil de eluição isocrática e o SDS-PAGE são mostrados. As setas duplas vermelhas indicam os produtos de eluição em cada caso.....	128
<b>Figura 2.19 - Ensaios de purificação dos núcleos de snRNPs U2 e U5 por cromatografia de troca iônica.</b> Perfis cromatográficos para purificação do núcleo de snRNP U2 em concentrações de sal A- alta (100 mM), B- intermediária (20 mM) e C- ausência de sal no tampão A. D- Perfil cromatográfico de purificação do núcleo snRNP U5. D- SPS-PAGE representativo de purificação do núcleo de snRNP U2 e U5, mostrando a dissociação dos complexos. As setas duplas vermelhas indicam os produtos eluídos.....	129
<b>Figura 2.20 - Cromatografia de exclusão molecular.</b> Padrão de eluição isocrática e SDS-PAGE para A1 e A2- núcleo de snRNP U2 e B1 e B2- núcleo Sm de snRNP U5. As setas duplas vermelhas indicam os produtos eluídos em cada caso.....	130
<b>Figura 2.21- Ensaios de otimização de cristalização por meio de semeadura.</b> São mostrados resultados representativos de diferentes proporções de amostra estoque de semente para a condição de cristalização G9 do <i>kit Morpheus II (Molecular Dimensions)</i> . Apresenta-se também resultado de NuPAGE 4-20% Bis-Tris ( <i>Thermoscientific</i> ), corado com nitrato de prata, para um dos cristais.....	134
<b>Figura 2.22- Padrões de difração de cristais obtidos em amostras do núcleo snRNP U2</b> - A- obtido <i>in house</i> , no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido, e B- medido na linha I24 da <i>Diamond Light Source</i> , Reino Unido.....	136
<b>Figura 2.23 - Análise de dados cristalográficos I.</b> A- Fator $B_{relativo}$ versus número da imagem. B- $R_{merge}$ versus número da imagem. C- Completude acumulada versus número da imagem...	137
<b>Figura 2.24 - Análise de dados cristalográficos II.</b> A- Análise do gráfico de Wilson. B- e C- Análise de $I/\sigma(I)$ médio e $CC_{1/2}$ , respectivamente, nas direções de $d_1$ (vermelho), $d_2$ (azul) e $d_3$ (verde).....	138

- Figura 2.25- Estrutura cristalográfica de U2A'/U2B'' de *T. brucei* I.** A e B- Conteúdo da cela unitária: dois heterodímeros de U2A'(azul)/U2B''(vermelho) na unidade assimétrica. Em B é apresentada a projeção da cela unitária no plano (*a,c*). Os círculos vermelhos correspondem à unidade assimétrica. Densidade eletrônica e modelo atômico C- logo após etapa de substituição molecular e D- logo após refinamento. A região C-terminal de U2A' está evidenciada. Nota-se, claramente, uma melhora na qualidade da densidade eletrônica na região após o refinamento, possibilitando a construção de uma hélice- $\alpha$  nesta região de U2A'..... 141
- Figura 2.26- Estrutura cristalográfica de U2A'/U2B'' de *T. brucei* II.** A- Gráfico de Ramachandran. B- Representação do conteúdo da cela unitária colorido de acordo com o valor do fator B por resíduo, do azul (menor valor) para o vermelho (maior valor). C- Representação das estruturas cristalográficas de U2A' (azul) e U2B'' (vermelho) de *T. brucei* na forma de *cartoon*. D- Topologia de U2A'. E- Topologia de U2B''..... 143
- Figura 2.27- Comparação da estrutura cristalográfica de U2A'/U2B'' de *T. brucei* e de *H. sapiens*.** A- Em vermelho e rosa estão representadas as estruturas de U2B'' de *T. brucei* e *H. sapiens*, respectivamente. Em azul mais forte e azul mais fraco estão representadas as estruturas de U2A' de *T. brucei* e *H. sapiens*, respectivamente. A molécula de SL4 do snRNA U2 de *H. sapiens* interage com U2B'' de esta espécie. B- Mapa de potencial eletrostático de superfície de U2A'/U2B'' de *T. brucei* obtido no programa APBS<sup>224</sup>. O potencial eletrostático é calculado em unidades  $k_bT e c^{-1}$  ( $\approx 25,9$  mV)..... 146
- Figura 2.28- Sm15K/Sm16,5K.** A- Modelo de estrutura 3D do dímero das proteínas Sm15K (vermelho) e Sm16,5K (azul). B e C- Gráficos de DOPE *score* por resíduo para os modelos de Sm15K e Sm16,5K, respectivamente, antes e após otimização. D- e E- Gráficos de Ramachandran para os modelos otimizados de Sm15K e Sm16,5K, respectivamente..... 148
- Figura 2.29 - Qualidade das estruturas cristalográficas do núcleo Sm de snRNPs U1 e U4 de *H. sapiens*.** A- e B- Propriedades das estruturas cristalográficas do núcleo Sm de snRNPs U1 e U4 de *H. sapiens*, respectivamente. C- e D- Gráficos de Ramachandran correspondentes. A qualidade dos modelos é medida pelas seguintes propriedades:  $R_{free}$ , que mede a similaridade entre os fatores de estrutura calculado e observado, a pontuação de sobreposição corresponde ao número de pares de átomos incomumente próximos, *outliers* do gráfico de Ramachandran e de cadeia lateral corresponde àqueles resíduos de aminoácidos com ângulos ( $\phi, \psi$ ) e  $\chi$  fora das regiões favoráveis, *outliers* de RSRZ ( $>2$ ) correspondem aos resíduos de aminoácidos modelados que não apresentam um bom ajuste com os dados observados transformados no espaço real, e, finalmente, a qualidade do modelo da cadeia principal do RNA baseada na estatística e ângulos de torção normalmente apresentados por estruturas de RNA publicadas..... 150
- Figura 2.30 - Avaliação da estereoquímica dos modelos dos complexos Sm de *T. brucei*,** a partir do Gráfico de Ramachandran. A1- Antes e A2- depois de otimização do complexo canônico (1), B- do complexo 2, e C1- antes e C2- depois da otimização do complexo 3..... 154

- Figura 2.31 – Visualização da superfície eletrostática dos modelos 3D dos cores Sm** A-canônico (complexo 1), B- contendo a variação Sm15k/Sm16,5K (complexo 2) e C- contendo a variação SSm4/SmB (complexo 3). O potencial eletrostático é calculado em unidades  $k_b T e_c^{-1}$  ( $\approx 25,9$  mV)..... 155
- Figura 2.32 – Proteínas Sm de *T. brucei*.** A- Visão transversal e lateral do motivo- $\beta$  característico das proteínas Sm de *T. brucei*, alinhadas no programa *PyMOL* (*Schrödinger, LLC*) e B- sobreposição dos modelos dos variantes dos núcleos Sm de *T. brucei*..... 156
- Figura 2.33– SAXS de complexos Sm de *T. brucei*.** A1, B1 e C1- Curvas de espalhamento, A2, B2 e C2- Gráficos de Guinier, e D, E e F- Gráficos de Krakty dos núcleos Sm 1, 2 e 3 de *T. brucei*, respectivamente..... 158
- Figura 2.34 – Curvas de distribuição de distâncias (p(r)) para os complexos A- Sm 1, B- Sm 2 e C- Sm 3.**..... 160
- Figura 2.35 – Envelopes moleculares dos complexos Sm de *T. brucei*.** A, B e C- Comparação entre curvas de espalhamentos experimentais e aquelas previstas para os modelos dos complexos 1,2 e 3, respectivamente, geradas pelo programa *CRY SOL*<sup>239</sup> e pelo programa *DAMMIN*<sup>236</sup>. Os envelopes moleculares médios gerados pelo programa *DAMMIN*<sup>236</sup> para os complexos 1, 2 e 3 são mostrados à direita. Visualização gerada pelo programa *PyMOL* (*Schrödinger, LLC*). São representados o envelope de SAXS (cinza) para cada complexo e as proteínas Sm (SmD3, Sm16,5K e SSm4 – azul, SmG – roxo, SmE – verde, SmF – azul-claro, SmD2 – amarelo, SmD1 – laranja e SmB ou Sm15K – vermelho) As setas vermelhas indicam a região do envelope molecular onde pode ser modelada a região C-terminal de SSm4..... 162
- Figura 2.36 – Sobreposição dos envelopes moleculares dos complexos Sm de *T. brucei* nos modelos de tri-snRNP U4/U6.U5 e do complexo spliceossomo que contém o íntron na forma de laço.** A- Parte da estrutura de criomicroscopia de tri-snRNP de *S. cerevisiae* está representada (Brr2 em vermelho, Prp8 em cinza, Snu114 em verde e proteínas Sm/LSm em azul, snRNAs U4, U6 e U5 em azul). O envelope molecular do complexo Sm 3 de *T. brucei* está representado em rosa e do complexo Sm canônico (de snRNP U5) é mostrado em ciano. B- Representação da estrutura de criomicroscopia do spliceossomo que contém o íntron na forma de laço, em cinza. O envelope molecular do complexo Sm 2 de *T. brucei* é apresentado em rosa, e o complexo Sm canônico (de snRNP U5), em ciano. Os snRNAs U2, U4, U6 e U5 são mostrados em azul. C- Região de snRNP U2 do spliceossomo. Apresenta-se, ainda, a sobreposição da estrutura do complexo Sm de levedura com aquela modelada para *T. brucei*, juntamente com o envelope molecular deste complexo. As estruturas cristalográficas de U2A'/U2B'' também foram sobrepostas, evidenciando-se as possíveis regiões de interação com snRNA U2 e com proteínas Sm..... 165
- Figura 2.37– Modelo de biogênese de núcleos Sm de *T. brucei*.** A associação das proteínas Sm aos snRNAs U4 e U5 depende da proteína SMN, que interage com SmB. Propõe-se que U2A'/U2B'' seja suficiente para garantir a especificidade de Sm15K/Sm16,5K pelo snRNA U2 em detrimento dos snRNAs U4 e U5..... 169

<b>Figura 3.1 – Arquitetura de domínios de U5-200K.</b> A- Representação esquemática de domínios conservados em U5-200K. B- Representação da estrutura tridimensional de U5-200K de <i>H. sapiens</i> (código PDB 4KIT) mostrando as regiões N-terminal e os cassetes helicase/ATPase/Sec63 N- e C-terminais, utilizando o mesmo código de cores de (A). As regiões N-terminal e cassete N-terminal são representados como <i>cartoon</i> e a região cassete C-terminal pe representada como superfície.....	174
<b>Figura 3.2 – Representação esquemática do vetor pC-PTP-NEO.</b> São mostradas as regiões de ligação da ORF de interesse (rosa), a extensão PTP (verde claro) e o gene de resistência a gentamicina (verde escuro). .....	179
<b>Figura 3.3 – Representação esquemática dos produtos de recombinação homóloga.</b> A- U5-200K-PTP completa. B- Região N-terminal + cassete helicase/ATPase/Sec63 N-terminal de U5-200K-PTP (incluindo região N-terminal).....	181
<b>Figura 3.4 – Representação esquemática das etapas de purificação em <i>tandem</i>.</b> A- Produto de expressão recombinante. B- Ligação à resina de IgG sepharose ( <i>GE</i> ). B- Ligação à resina com anticorpos anti-proteína C imobilizados ( <i>Roche</i> ) após digestão com protease TEV. D- Eluição com tampão contendo EGTA/EDTA.....	184
<b>Figura 3.5 – Representação esquemática das regiões de amplificação dos produtos de <i>cis</i> e <i>SL trans-splicing</i>.</b> São apresentados os pares de iniciadores (setas) para amplificação de região de pré-mRNA não processado (esquerda) e processado (direita), no caso de A- <i>cis</i> - e B- <i>SL trans-splicing</i> , respectivamente. As sequências de DNA e a respectiva T <sub>M</sub> são apresentados.....	185
<b>Figura 3.6 – Alinhamento múltiplo de sequências de U5-200K conservadas em <i>T. brucei</i>, <i>S. cerevisiae</i> e <i>H. sapiens</i>.</b> NT = região N-terminal e CT = região C-terminal.....	188
<b>Figura 3.7 – Clonagem e expressão homóloga de U5-200K de <i>T. brucei</i>.</b> A e B- Confirmação de inserção no DNA genômico de <i>T. brucei</i> 427 das construções de U5-200K completa/pC-PTP-NEO (1- controle negativo - cepa selvagem, 2- região amplificada de U5-200K completa) e U5-200K cassete N-terminal/pC-PTP-Neo (1- controle negativo – cepa selvagem, 2- região amplificada de U5-200K cassete N-terminal). C- Confirmação por <i>Western Blotting</i> da expressão homóloga das proteínas recombinantes com extensão PTP. 1- controle positivo, correspondente à proteína PSTK de <i>T. brucei</i> com extensão PTP (59 kDa). 2- U5-200K cassete N-terminal + PTP. 3-U5-200K completa + PTP. 4- Controle negativo – cepa selvagem de <i>T. brucei</i> 427 na forma procíclica.....	194
<b>Figura 3.8 – Curvas de crescimento de cultura de <i>T. brucei</i> na temperaturas de 26°C para cada população clonal.</b> As células foram contadas em câmara de Neubauer em microscópio óptico invertido. Segundo o teste T de Student aplicado aos resultados com hipótese nula igual à situação em que não há diferença entre as curvas em cada ponto, * indica diferença significativa com p<0,05, ** p<0,02, *** p<0,01 e **** p<0,0001.....	195

- Figura 3.9 – Purificação em *tandem* de U5-200K-PTP (A) e do seu cassete N-terminal-PTP (B).**  
MM = marcador de massa molecular, 1 = lisado celular (0,003%), 2- amostra que não interagiu com resina de IgG sepharose (0,003%), 3- eluato após tratamento com protease TEV (5%), 4- amostra que não interagiu com resina contendo anticorpo anti-proteína C imobilizado (5%), 5 e 6- eluato final (100%). As proteínas foram identificadas por espectrometria de massas no laboratório da Dra. Julia Chagas Pinheiro da Cunha, no Instituto Butantan..... 197
- Figura 3.10 – Imunolocalização de U200K-PTP em microscopia de fluorescência. *T. brucei* 427**  
na forma procíclica A- selvagem, B- expressando U5-200K-PTP completa e C- seu cassete N-terminal-PTP (B). A barra tem tamanho de 10  $\mu$ m..... 201
- Figura 3.11 – Processamento do pré-mRNA por *cis*- e *trans*-splicing por RT-PCR.** A- Controle 7SL, B- *Cis-splicing* – PAP, C- SL *trans-splicing*  $\alpha$ -tubulina. D- e E- Quantificação das bandas observadas em B. F- e G- Quantificação das bandas observadas em C. Cont. = células controle, 1 = células que expressam U5-200K-PTP completa e 2 = células que expressam cassete N-terminal de U5-200K-PTP..... 203
- Figura 3.12 – Modelos de maturação da snRNP U5 em levedura e *T. brucei*.** A- Em levedura, um complexo pré-snRNP U5, formado por snRNA U5, proteínas Sm, Snu114, Prp8 e Aar2 entra no núcleo, em transporte mediado por snuportina 1 e importina- $\beta$ . No núcleo, uma cinase desconhecida fosforila Aar2, que dá seu lugar para Brr2. O complexo maduro então interage com Dib1, Prp6 e Prp28. B- Já no caso de *T. brucei*, propõe-se que um complexo pré-formado de snRNP U5, contendo snRNA U5, proteínas Sm, U5-116K, U5-220K, U5-200K, U5-Cwc21 e U5-40K entre no núcleo em transporte mediado por, pelo menos, importina- $\beta$ . No núcleo, este complexo deve interagir com o subcomplexo formado por U5-15K e U5-102K ..... 205
- Figura A2.1 Avaliação da qualidade dos modelos obtidos por modelagem comparativa dos complexos Sm de *T. brucei*.** São mostrados, da esquerda para a direita, para cada cadeia polipeptídica nos três complexos, a distribuição dos valores de pontuação Z para estruturas cristalográficas e de ressonância magnética nuclear depositadas no PDB e a pontuação Z para o modelo em análise, seguida do gráfico de energia por resíduo de aminoácido medida em janelas de 10 (verde claro) e 40 resíduos (verde escuro), e da topologia da cadeia polipeptídica..... 231





## LISTA DE TABELAS

Tabela 2.1 – Características físico-químicas de proteínas Sm, U2A' e U2B'' de <i>T. brucei</i> .....	106
Tabela 2.2 – Condições de cristalização do kit <i>Morpheus II (Molecular Dimensions)</i> onde foram observados cristais para amostra de núcleo de snRNP U2.....	132
Tabela 2.3 – Estatísticas dos dados cristalográficos..	139
Tabela 2.4 – Análise de ausências sistemáticas na direção de <i>d2</i> .....	140
Tabela 2.5 – Estatística da última etapa de refinamento da estrutura cristalográfica de U2A'/U2B'' de <i>T. brucei</i> .....	142
Tabela 2.6 – Comparação entre os valores de <i>rmsd</i> resultantes da sobreposição entre as cadeias polipeptídicas observadas nos modelos cristalográficos de U2A'/U2B'' de <i>T. brucei</i> e de <i>H. sapiens</i> .....	145
Tabela 2.7 – Comparação entre as sequências de aminoácidos de proteínas Sm de <i>T. brucei</i> e de <i>H. sapiens</i> .....	147
Tabela 2.8 – Propriedades das interfaces de interação entre as proteínas Sm de <i>T. brucei</i> .	157
Tabela 2.9 – Parâmetros estruturais calculados por SAXS e a partir dos modelos de estrutura 3D dos complexos Sm de <i>T. brucei</i> .	161
Tabela 2.10 – Comparação entre massas moleculares calculadas por SAXS, cromatografia de exclusão molecular e a partir das sequências de aminoácidos dos complexos Sm de <i>T. brucei</i> .....	167
Tabela 3.1 – Proteínas copurificadas com U5-200K-PTP e identificadas por espectrometria de massas.	198
Tabela 3.2 – Proteínas copurificadas com o cassete N-terminal de U5-200K-PTP e identificadas por espectrometria de massas.....	199
Tabela A1.1 – Pequenas moléculas moduladoras do spliceossomo.	229



## LISTA DE ABREVIATURAS E SIGLAS

ASO	do inglês, <i>antisense oligonucleotide</i> ,
AVAI	anos de vida ajustados por incapacidade
BARP	do inglês, <i>brucei alanine-rich protein</i>
CATH	do inglês, <i>class, architecture, topology, homologous superfamily</i>
CBC	do inglês, <i>cap binding complex</i>
CD	do inglês, <i>circular dichroism</i>
CDS	do inglês, <i>coding DNA sequence</i>
DGC	do inglês, <i>directional gene clusters</i>
DTN	doenças tropicais negligenciadas
EJC	do inglês, <i>exon-junction complex</i>
EMSA	do inglês, <i>electromotility shift assay</i>
ESS/ISS	do inglês, <i>exonic/intronic splicing silencer</i>
ESSE/ISE	do inglês, <i>exonic/intronic splicing enhancer</i>
GPI	glicosil-fosfatidilinositol
IBC	do inglês, <i>intron-binding complex</i>
LL	do inglês, <i>likelihood</i>
LLG	do inglês, <i>log-likelihood gain</i>
LRR	leucina (L)-arginina (R)-arginina (R)
NTC	do inglês, <i>nineteen complex</i>
NTR	do inglês, <i>nineteen-related complex</i>
OMS	Organização Mundial da Saúde
PCA	do inglês, <i>principal component analysis</i>
PEG	polietilenoglicol
RES	do inglês, <i>retention and splicing complex</i>
RRM	do inglês, <i>RNA recognition motif</i>
SAXS	do inglês, <i>small angle X-ray scattering</i>
SDS-PAGE	do inglês, <i>sodium dodecyl-sulfate polyacrylamide gel electrophoresis</i>
Sm	<u>S</u> mith
SMN	do inglês, <i>survivor of motor neurons</i>
snoRNP	do inglês, <i>small nucleolar ribonucleoprotein</i>
snRNA	do inglês, <i>small nuclear ribonucleic acid</i>
snRNP	do inglês, <i>small nuclear ribonucleoprotein</i>
TAP	do inglês, <i>tandem affinity purification</i>
TPR	do inglês, <i>tetratricopeptide repeat</i>
UV	ultravioleta
VSG	do inglês, <i>variable surface glycoproteins</i>



## SUMÁRIO

<b>CAPÍTULO 1 – REVISÃO BIBLIOGRÁFICA</b> .....	31
<b>1.1 A família Trypanosomatidae e seu impacto médico e socioeconômico</b> .....	33
<b>1.2 <i>Trypanosoma brucei</i> como organismo modelo</b> .....	49
<b>1.2.1 Arquitetura da célula durante o ciclo de vida</b> .....	39
<b>1.2.2 Organização do genoma, replicação e transcrição</b> .....	42
<b>1.2.3 Processamento do pré-mRNA</b> .....	44
<b>1.3 Spliceossomo: arquitetura e função no <i>cis</i>- e no <i>SL trans-splicing</i></b> .....	46
<b>1.3.1 Catálise do <i>splicing</i></b> .....	47
<b>1.3.2 snRNPs: composição, função e biogênese</b> .....	51
<b>1.3.2.1 snRNP U1</b> .....	51
<b>1.3.2.2 snRNP U2</b> .....	52
<b>1.3.2.3 snRNP U5, di-snRNP U4/U6 e tri-snRNP U4/U6.U5</b> .....	55
<b>1.3.2.4 Biogênese de snRNPs</b> .....	59
<b>1.3.3 Dinâmica de associação das subunidades do spliceossomo para a catálise do <i>splicing</i></b> .....	61
<b>1.4 Regulação do spliceossomo</b> .....	65
<b>1.4.1 Relação entre a regulação do <i>spliceossomo</i> e a saúde humana</b> .....	68
<b>1.4.1.1 Doenças humanas diretamente relacionadas ao mal-funcionamento do spliceossomo</b> .....	68
<b>1.4.1.2 O <i>SL trans-splicing</i> de parasitas como alvo</b> .....	70
<b>CAPÍTULO 2 - ANÁLISE BIOQUÍMICA E ESTRUTURAL DOS NÚCLEOS Sm DAS snRNPs U2, U4 e U5 E DAS PROTEÍNAS U2A' E U2B'' DE <i>T. brucei</i></b> .....	73
<b>2.1 Introdução</b> .....	75
<b>2.1.1 A família das proteínas Sm</b> .....	75
<b>2.1.1.1 Interação de proteínas Sm/LSm com snRNAs</b> .....	78
<b>2.1.1.2 Caracterização de proteínas Sm e LSm de <i>T. brucei</i></b> .....	79
<b>2.1.2 U2A'/U2B''</b> .....	80
<b>2.1.2.1 U2A', U2B'' e sua interação com o snRNA U2</b> .....	80
<b>2.1.2.2 Caracterização de U2A'/U2B'' de <i>T. brucei</i></b> .....	83
<b>2.2 Justificativa e objetivos</b> .....	83
<b>2.3 Materiais e métodos</b> .....	84
<b>2.3.1 Análise das sequências de aminoácidos das proteínas Sm, U2A' e U2B'' de <i>T. brucei</i> por meio de ferramentas de Bioinformática</b> .....	84
<b>2.3.2 Clonagem, expressão e purificação de subcomplexos de proteínas Sm</b> .....	84

2.3.3 Clonagem, expressão e purificação de U2A'/U2B'' .....	87
2.3.4 Preparo dos snRNAs U2, U4 e U5 e sua caracterização por meio de espectroscopia de dicroísmo circular .....	89
2.3.4.1 Espectroscopia de dicroísmo circular .....	91
2.3.5 Reconstrução <i>in vitro</i> dos núcleos Sm de <i>T. brucei</i> .....	92
2.3.5.1 Ensaio de retardo de mobilidade eletroforética (EMSA) .....	92
2.3.5.2 Ensaio de purificação .....	93
2.3.6 Ensaio de cristalização, análise de dados de difração de raios-X e resolução de estrutura tridimensional.....	94
2.3.6.1 Cristalização.....	94
2.3.6.2 Coleta e análise de dados de difração de raios-X.....	96
2.3.6.3 Resolução do problema das fases, refinamento e análise do modelo de estrutura tridimensional.....	98
2.3.7 Modelagem da estrutura terciária e quaternária .....	101
2.3.8 Espalhamento de raios-X a baixo ângulo (SAXS).....	102
2.4 Resultados.....	105
2.4.1 Análise das sequências de aminoácidos das proteínas Sm, U2A' e U2B'' de <i>T. brucei</i> por meio de ferramentas de Bioinformática.....	105
2.4.1.1 Análise da conservação de aminoácidos das proteínas Sm e dos sítios de interação com snRNAs.....	107
2.4.1.2 Conservação de aminoácidos das proteínas U2A' e U2B'' e análise do sítio de interação com snRNA U2 .....	112
2.4.2 Clonagem, expressão e purificação das proteínas Sm .....	114
2.4.3 Clonagem, expressão e purificação de U2A'/U2B'' .....	118
2.4.4 Preparo dos snRNAs U2, U4 e U5 e sua caracterização por meio de espectroscopia de dicroísmo circular.....	119
2.4.5 Reconstrução <i>in vitro</i> dos núcleos Sm de <i>T. brucei</i> .....	122
2.4.5.1 Ensaio de retardo de mobilidade eletroforética (EMSA) .....	122
2.4.5.2 Ensaio de purificação .....	127
2.4.6 Ensaio de cristalização, análise de dados de difração de raios-X e resolução da estrutura tridimensional de U2A'/U2B'' .....	131
2.4.6.1 Ensaio de cristalização.....	131
2.4.6.2 Análise dos dados de difração de raios-X.....	135
2.4.6.3 Resolução do problema das fases, refinamento e análise do modelo de estrutura tridimensional.....	140
2.4.7 Modelagem molecular da estrutura terciária e quaternária das proteínas Sm de <i>T. brucei</i> .....	146

2.4.8 Caracterização dos núcleos Sm de snRNPs U2, U4 e U5 na ausência de snRNAs por meio de espalhamento de raios-X em baixo ângulo (SAXS).....	157
2.5 Discussão.....	167
<b>CAPÍTULO 3 - ANÁLISE BIOQUÍMICA DA PROTEÍNA U5-200K DE <i>T. brucei</i> E DE SEUS PARCEIROS DE INTERAÇÃO</b> .....	171
3.1 Introdução.....	173
3.1.1 U5-200K e a ativação do spliceossomo.....	173
3.1.2 U5-200K – estrutura e função helicase.....	173
3.1.3 U5-200K de <i>T. brucei</i> .....	177
3.2 Objetivos.....	177
3.3 Materiais e métodos.....	178
3.3.1 Análise da sequência de aminoácidos da proteína U5-200K de <i>T. brucei</i> .....	178
3.3.2 Clonagem e expressão homóloga de construções de U5-200K em <i>T. brucei</i> para experimentos de purificação em <i>tandem</i> .....	178
3.3.2.1 Amplificação das fases abertas de leitura e clonagem em vetor pC-PTP-NEO.....	178
3.3.2.2 Transfecção, seleção e verificação de expressão dos produtos recombinantes.....	179
3.3.2.3 Curvas de crescimento.....	182
3.3.3 Purificação dos produtos recombinantes e análise dos parceiros de interação por espectrometria de massas.....	182
3.3.4 Imunolocalização dos produtos recombinantes.....	185
3.3.5 Avaliação de produtos de <i>cis</i> - e SL <i>trans-splicing</i> por meio de RT-PCR.....	186
3.4 Resultados e discussão.....	187
3.4.1 Análise da estrutura primária de U5-200K.....	187
3.4.2 Clonagem e expressão homóloga de U5-200K de <i>T. brucei</i> .....	193
3.4.2.1 Amplificação de regiões de interesse de U5-200K, clonagem, transfecção em <i>T. brucei</i> e seleção de clones positivos.....	193
3.4.2.2 Curvas de crescimento.....	195
3.4.4 Purificação dos produtos recombinantes e análise dos parceiros de interação por espectrometria de massas.....	196
3.4.4 Imunolocalização dos produtos recombinantes.....	201
3.4.5 Avaliação de produtos de <i>cis</i> - e SL <i>trans-splicing</i> por meio de RT-PCR.....	202
3.5 Discussão.....	204

<b>CAPÍTULO 4 - CONSIDERAÇÕES FINAIS.....</b>	<b>207</b>
<b>4.1 Conclusão .....</b>	<b>209</b>
<b>4.2 Perspectivas.....</b>	<b>211</b>
<b>REFERÊNCIAS.....</b>	<b>213</b>
<b>APÊNDICE A1 .....</b>	<b>229</b>
<b>APÊNDICE A2 .....</b>	<b>231</b>



## CAPÍTULO 1

### REVISÃO BIBLIOGRÁFICA:

#### **ANÁLISE COMPARATIVA DO PROCESSAMENTO DO PRÉ-mRNA CATALISADO PELO SPLICEOSSOMO EM *Trypanosoma brucei*, EM SEU HOSPEDEIRO HUMANO E EM LEVEDURA**

Neste capítulo será introduzida a família dos tripanossomatídeos e seu impacto com relação à saúde humana, e o membro *Trypanosoma brucei* será apresentado como organismo modelo para o estudo comparativo de processos moleculares de eucariotos. O mecanismo de processamento do pré-mRNA catalisado pelo spliceossomo no parasita, conhecido como SL *trans-splicing*, será discutido em comparação com o *cis-splicing* caracterizado no hospedeiro humano e em leveduras. Adicionalmente, a relação estrutura-dinâmica-função do spliceossomo e sua regulação serão apresentados, destacando sua relação com a saúde humana.



## 1.1 A família Trypanosomatidae e seu impacto médico e socioeconômico

Nos últimos anos, avanços científicos e tecnológicos proporcionaram o aumento da expectativa de vida da população mundial. No entanto, o grupo das doenças tropicais negligenciadas (DTN) ainda apresenta sérias consequências no desenvolvimento socioeconômico de milhões de pessoas em todo o mundo.<sup>1</sup>

As 17 DTN\* priorizadas pela Organização Mundial da Saúde são doenças infecciosas causadas por bactérias, protozoários, helmintos e vírus, que se desenvolvem em regiões de pobreza, cujo quadro é agravado pelo calor e umidade do clima tropical. Em sua maioria, são patologias transmitidas por insetos vetores ou pela água, solo e mesmo sangue contaminados com formas infectantes. Estas doenças são caracterizadas por prosperar em áreas rurais remotas e periferias urbanas de diversos países da África subsaariana, Ásia e América Latina.<sup>1</sup>

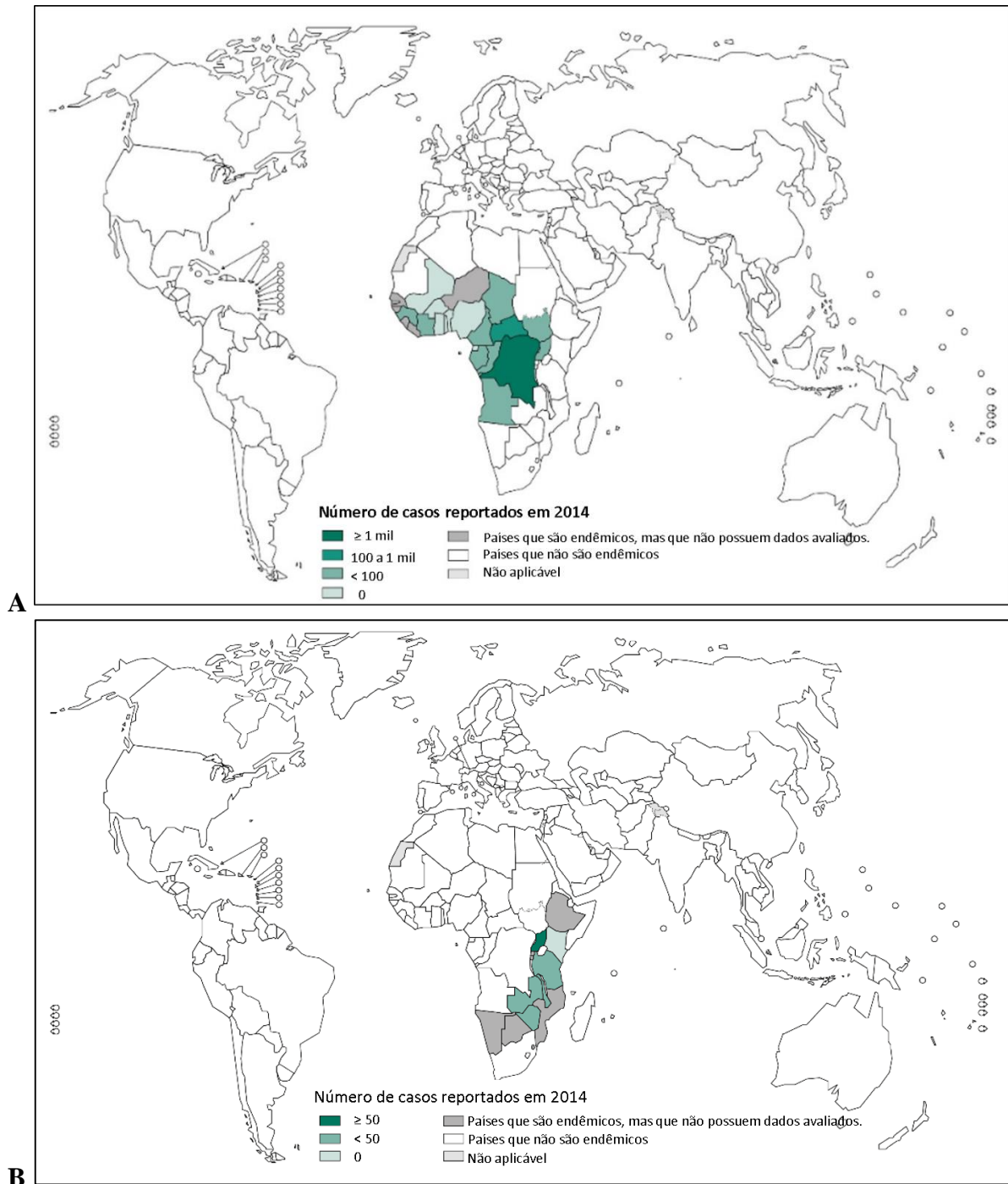
A família Trypanosomatidae é responsável por três DTNs (doença do sono, doença de Chagas e leishmanioses), e, por isso, é considerada de grande importância médica e econômica, uma vez que muitos de seus membros são parasitas e as doenças causadas por eles possuem caráter zoonótico.<sup>1</sup> Os protozoários flagelados pertencentes a esta família são classificados na ordem Trypanosomatida e na classe Kinetoplastida, dentro do filo Euglenozoa<sup>2</sup>, por apresentarem uma estrutura conhecida como cinetoplasto, o qual contém o material genético da mitocôndria.<sup>2-3</sup> Estes organismos são caracterizados pela presença de um único flagelo e seu nome reflete a movimentação em espiral (do grego, *trypano* = furador e *soma* = corpo).<sup>3</sup>

Em humanos, *Trypanosoma brucei*, *Trypanosoma cruzi* e *Leishmania sp.* causam, respectivamente, tripanossomíase africana humana ou doença do sono, tripanossomíase americana ou doença de Chagas, e leishmanioses.<sup>1</sup> Juntas, tais enfermidades acometem atualmente cerca de 20 milhões de pessoas, com mais de 2 milhões de novos casos anuais. Não existe vacina para essas doenças e o controle do inseto vetor é o método mais efetivo de prevenção. Além disso, associa-se o controle de bancos de sangue para se evitar infecção por transfusão sanguínea ou transplante de órgãos.<sup>1</sup>

A tripanossomíase africana humana afeta principalmente populações que vivem em áreas rurais remotas na África (Figura 1.1). Entretanto, vários casos foram reportados em regiões não-endêmicas, incluindo Europa e América, de turistas ou migrantes infectados nas regiões endêmicas.<sup>4</sup> Em 2009, o número de novos casos reportados diminuiu para menos de 10

\* As 17 DTN priorizadas pela OMS são: doença de Chagas, doença do sono e leishmanioses (causadas por protozoários), úlcera de Buruli, leprose, tracoma e boubá (causadas por bactérias), teníase e cisticercose, dracunculíase, equinococose, metagonimíase, filariose linfática, oncocercose, esquistossomose e helmintíases transmitidas pelo solo (causadas por helmintos), dengue e raiva (causadas por vírus).

mil pela primeira vez em 50 anos e estima-se que o número de casos atualmente seja de cerca de 30 mil.<sup>1,5-6</sup> As infecções causadas pelas subespécies *T. brucei gambiense* e *T. brucei rhodesiense* são transmitidas por espécies infectadas da mosca do gênero *Glossina*.<sup>1,5,7</sup> Com o tempo, o parasita que se encontra na corrente sanguínea acaba cruzando a barreira hematoencefálica e migra até o sistema nervoso central, causando sintomas como prostração, podendo levar ao coma e à morte.<sup>1</sup>

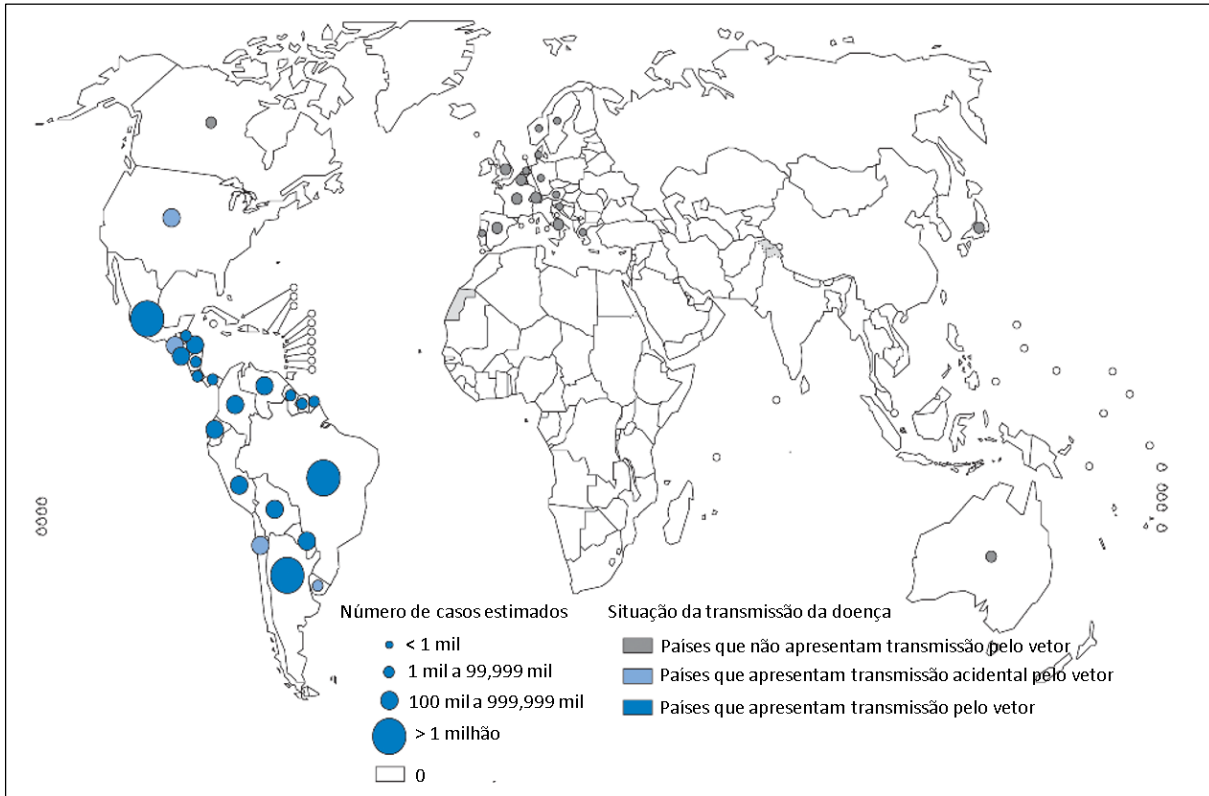


**Figura 1.1 - Distribuição geográfica da doença do sono em 2014. A – *T. brucei gambiense* e B – *T. brucei rhodesiense*.**

Fonte: Adaptada de WHO<sup>1</sup>.

Atualmente, o tratamento da doença do sono depende do estágio da doença. Suramina e pentamidina (descobertas em 1920 e 1940, nesta ordem) são usadas na fase inicial do tratamento da doença causada por *T. brucei rhodesiense* e *T. brucei gambiense*, respectivamente.<sup>1,8</sup> Ambas apresentam efeitos colaterais consideráveis, desde reações alérgicas a dano nas células pancreáticas, o que pode gerar hipoglicemia seguida de diabetes. Na segunda fase do tratamento, o melarsopol, descoberto em 1949, é administrado para os dois tipos de infecção. Este fármaco é derivado do arsênico e apresenta efeitos colaterais limitantes ao tratamento, como a encefalopatia, que pode ser fatal (de 3 a 10% dos casos). Também pode ser administrada a eflornitina (registrada em 1990), que é menos tóxica que o melarsopol, porém é efetiva somente contra *T. brucei gambiense*. Um tratamento combinado com nifurtimox e eflornitina foi introduzido em 2009.<sup>1,8</sup>

Já a doença de Chagas é transmitida principalmente por insetos da família Reduviidae, subfamília Triatominae, atingindo aproximadamente 10 milhões de pessoas em todo o mundo<sup>1,9</sup> (Figura 1.2). A presença da doença em países fora da América Latina se deve principalmente à migração de pessoas infectadas. Esta patologia, assim como a doença do sono, também pode ser transmitida por transfusão de sangue contaminado, de maneira congênita ou em acidentes de laboratório. Na fase crônica da doença, complicações cardíacas e intestinais podem aparecer, acometendo a qualidade de vida do infectado.<sup>1,9</sup>



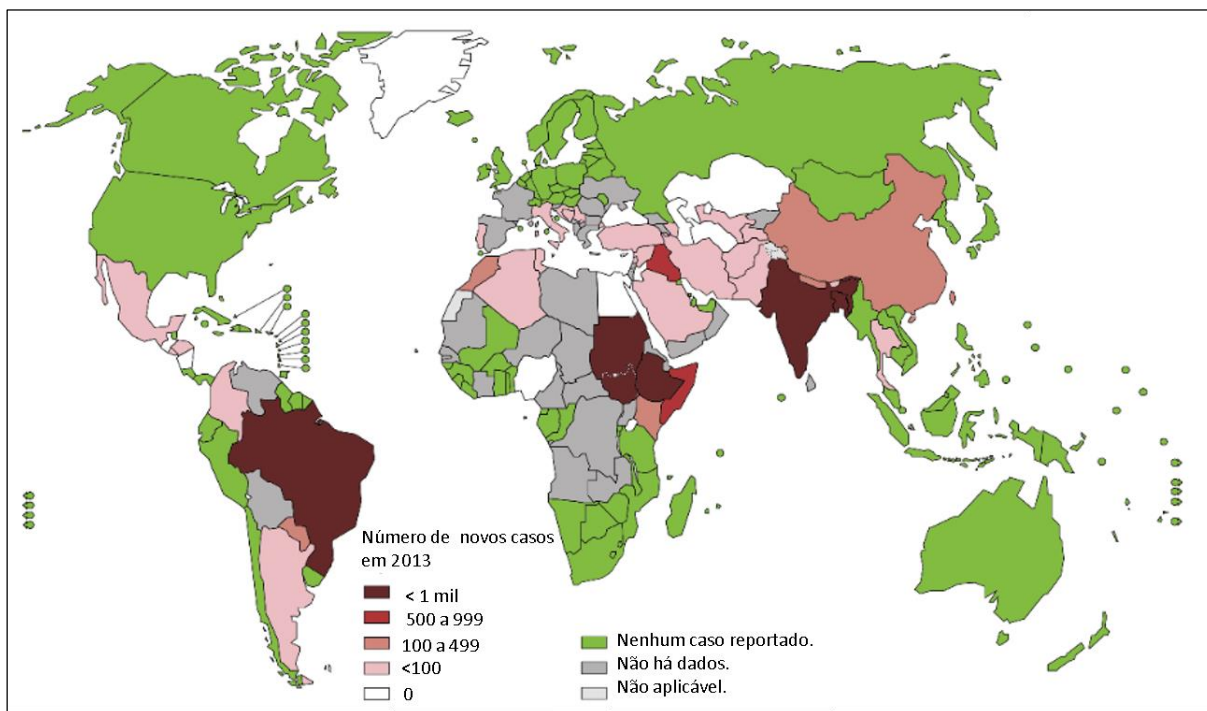
**Figura 1.2 - Distribuição geográfica da doença de Chagas entre 2006 e 2009.**

Fonte: Adaptada de WHO<sup>1</sup>.

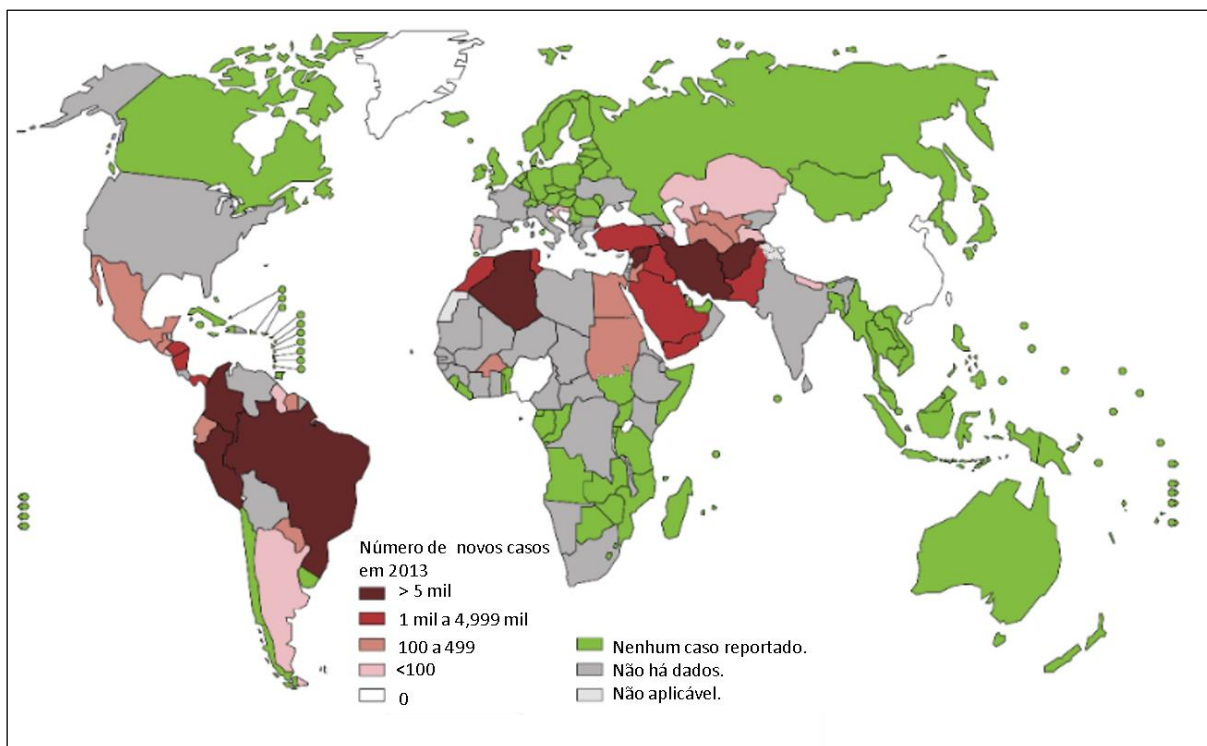
O fármaco benznidazol está registrado atualmente para o tratamento da doença de Chagas, sendo altamente efetivo na sua cura se administrado na fase aguda. Os efeitos colaterais comuns incluem alergia, perda de peso, neuropatia, entre outros.<sup>1, 10-11</sup>

No caso das leishmanioses, tanto as formas cutânea quanto visceral resultam de transmissão pela picada de fêmeas de insetos dípteros flebotomíneos do gênero *Lutzomyia*, no continente americano, e *Phlebotomus*, na Europa, África e Ásia.<sup>1</sup> São descritos aproximadamente 62 mil novos casos da doença por ano em países como Brasil, Colômbia, Paraguai, Venezuela, Panamá, Equador e Peru (Figura 1.3).<sup>1, 7, 12</sup> A forma mais severa de leishmaniose é a visceral, que leva a um quadro de febre alta, hepatoesplenomegalia, e anemia, podendo causar morte se não tratada.<sup>1</sup>

A



B



**Figura 1.3 - Distribuição geográfica dos novos casos de leishmaniose em 2013. A – Leishmaniose visceral e B – Leishmaniose cutânea.**  
 Fonte: Adaptada de WHO<sup>1</sup>.

O tratamento das leishmanioses depende da forma infectante e também do estágio da doença, devendo ser individualizado a cada caso. Em geral, opta-se por quimioterapia a partir da administração de antimoniais pentavalentes, como fármacos de primeira escolha, mas também podem ser indicados miltefosina, anfotericina B e paromomicina, entre outros, com efeitos colaterais severos que incluem cefaleia e leucopenia.<sup>1, 13</sup>

Além das tripanossomíases humanas, vários são os casos reportados da doença conhecida como nagana ou tripanossomíase africana animal, que é causada por algumas espécies de *Trypanosoma*, como *T. vivax* e *T. brucei*. Os protozoários são transmitidos pelas moscas tsé-tsé ao hospedeiro vertebrado, causando febre, fraqueza e letargia, podendo ser fatal. Estima-se que a presença de nagana leva à redução de 40 a 70% do rebanho, causando grande prejuízo econômico na indústria pecuária.<sup>14-15</sup>

Poucos são os estudos acerca dos custos efetivos de controle e tratamento das tripanossomíases e leishmanioses<sup>16-18</sup>. Trabalhos recentes mostraram que, em média, um indivíduo infectado com *Trypanosoma cruzi* resulta em 0,51 ano de vida ajustado por incapacidade (AVAI)\* e gastos acima de 400 dólares americanos anualmente.<sup>1, 17</sup> Globalmente, mais do que sete bilhões de dólares americanos são gastos anualmente na prevenção e no tratamento da doença, valor comparável ao gasto com doenças como câncer de cervical e aquela causada pelo rotavírus.<sup>17</sup> No caso da leishmaniose visceral, a estimativa é de mais de dois milhões de AVAIs<sup>18</sup> atualmente e o total gasto para prevenção e tratamento nos próximos anos é estimado em 100 milhões de dólares americanos por ano.<sup>1, 18</sup> Já para a tripanossomíase africana, estima-se entre 2 e 25 mil AVAIs anuais, sendo que o custo médio para tratamento recentemente vai de 25 a 500 dólares americanos por AVAI.<sup>1, 16, 19</sup>

O grande impacto desses parasitas na saúde humana e também na economia dos países em desenvolvimento tem direcionado investigações de seus processos moleculares. Dentre os vários mecanismos moleculares peculiares dos tripanossomatídeos, pode-se citar a predominância do processamento dos pré-mRNAs por *SL trans-splicing*, foco deste trabalho. Por ser um mecanismo ausente em vertebrados, em que ocorre somente o *cis-splicing*, tal processamento e as proteínas participantes podem ser consideradas importantes alvos parasito-específicos.<sup>20</sup>

\* O número de anos de vida ajustados por incapacidade (AVAI), do inglês *disability-adjusted life years* (DALY), é uma medida de carga de doença proposta por Murray e Lopes<sup>21</sup>, que corresponde à soma do número de anos perdidos por incapacidade (API) com o número potencial de anos de vida perdidos (APVP) em consequência de mortalidade precoce.



## 1.2 *Trypanosoma brucei* como organismo modelo

*T. brucei* é um organismo eucarioto unicelular que pode ser tratado como modelo para o estudo comparativo de mecanismos moleculares em eucariotos. Processos biológicos como edição de RNA, ancoramento de proteína em membrana celular por meio de glicosil-fosfatidilinositol (GPI), SL *trans-splicing* e variação antigênica foram primeiramente descobertos neste organismo, e depois observados e caracterizados em outros eucariotos.<sup>22</sup>

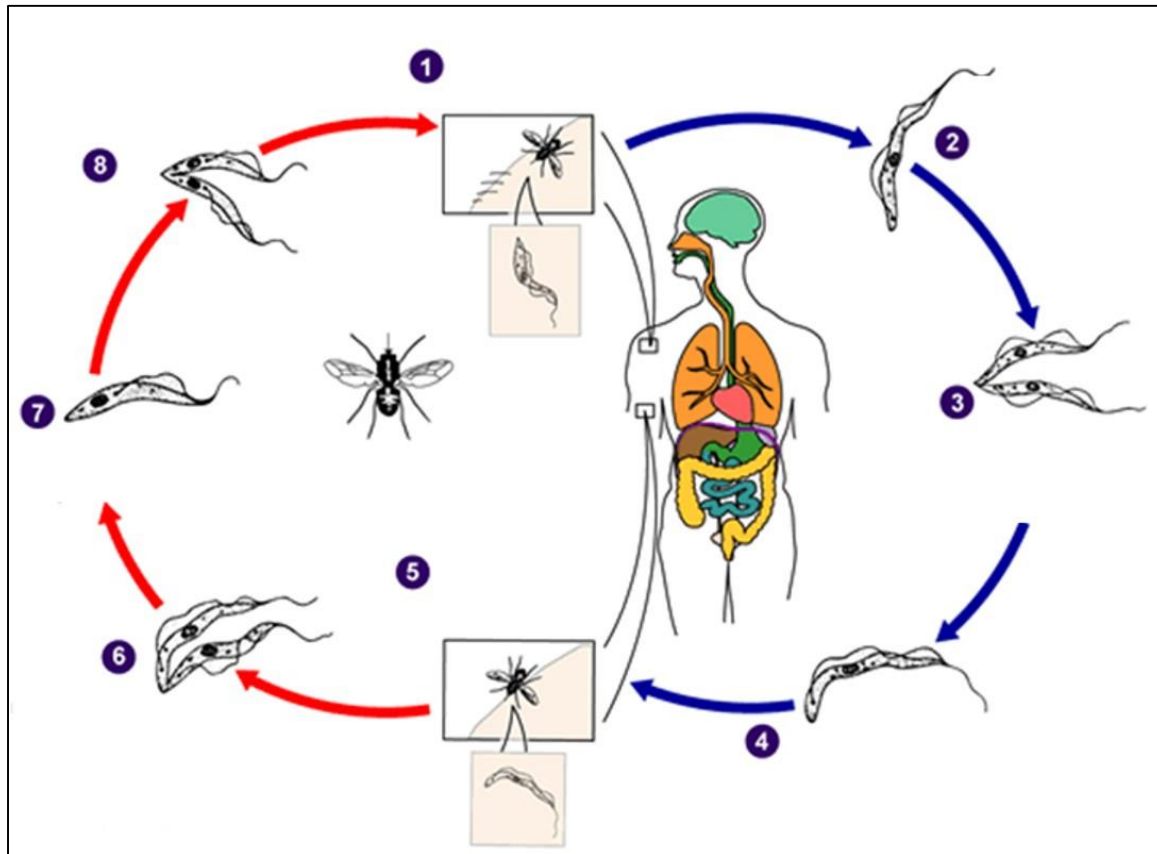
O sequenciamento completo de seu genoma<sup>23</sup> aliado ao desenvolvimento de técnicas como a deleção de um gene a partir de recombinação homóloga ou regulação negativa da expressão gênica por RNA de interferência permitem o estudo genético dos seus processos biológicos.<sup>22, 24</sup> Adicionalmente, os fatores necessários para crescimento e adaptação em diferentes fases do ciclo de vida podem ser investigados tanto em cultura como em modelos animais.<sup>22</sup> Além disso, *T. brucei* e outros membros da ordem Kinetoplastida contêm apenas uma mitocôndria<sup>25</sup> e organelas exclusivas que sofrem mudanças estruturais e funcionais durante seu ciclo de vida. Deste modo, este organismo também é de interesse para estudo de biogênese e função de organelas.<sup>22</sup>

### 1.2.1 Arquitetura da célula durante o ciclo de vida

A célula dos parasitas pertencentes à família Trypanosomatidae caracteriza-se por ser alongada, apresentando um citoesqueleto de microtúbulos altamente polarizados que define o formato da célula e se mantém intacto durante todo o seu ciclo de vida.<sup>22</sup>

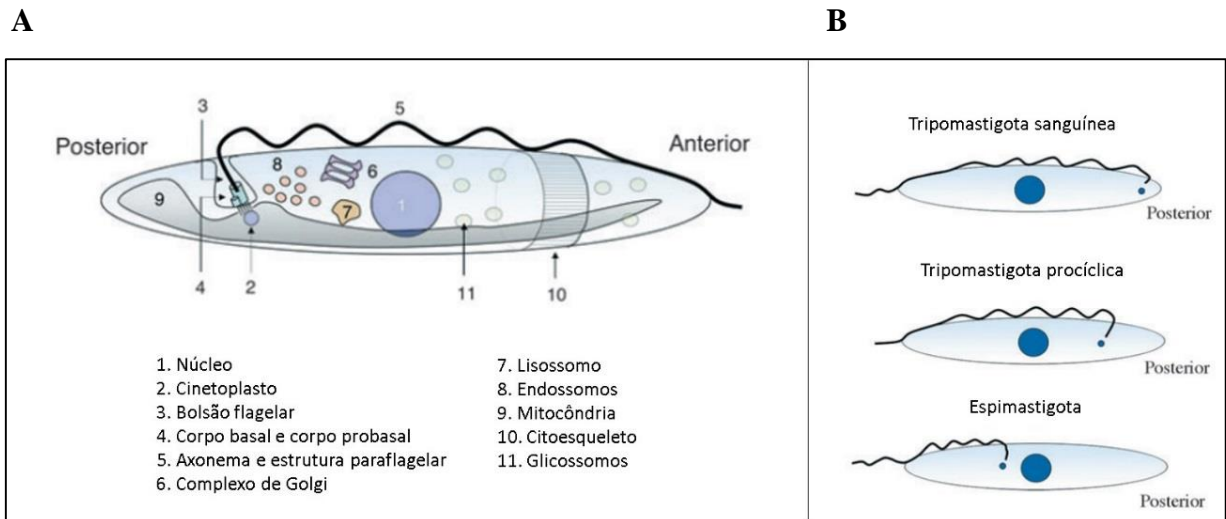
No caso de *T. brucei*, as células presentes na saliva da mosca tsé-tsé são conhecidas como tripomastigotas metacíclicas e são injetadas pelo inseto durante a alimentação de sangue do hospedeiro mamífero (Figura 1.4). Em humanos, os parasitas entram no sistema linfático e passam para a corrente sanguínea, onde se transformam em tripomastigotas sanguíneos (Figura 1.4). Na corrente sanguínea, estas formas são carregadas a outros órgãos e também atingem outros fluidos corporais, como o líquido, e continuam a se reproduzir por fissão binária.<sup>26-27</sup>

A mosca tsé-tsé se infecta com a forma tripomastigota sanguínea durante a alimentação de sangue contaminado e, no seu intestino médio, os parasitas se transformam em tripomastigotas procíclicos, onde também se multiplicam por fissão binária (Figura 1.4). Deixando esta região, os parasitas se transformam em epimastigotas e alçam a glândula salivar do inseto vetor, onde seguem a se multiplicar e se transformam em tripomastigotas metacíclicas<sup>26-27</sup> (Figura 1.4).



**Figura 1.4 - Ciclo de vida do *T. brucei*** – 1- A forma tripomastigota metacíclica do *T. brucei* é injetada no hospedeiro humano pela mosca tsé-tsé. 2-4- *T. brucei* se transforma em tripomastigota sanguíneo, atinge outros fluidos corporais (linfa, líquido) e continua sua replicação por fissão binária. 5- A mosca tsé-tsé se infecta com as formas tripomastigotas sanguíneas. 6- No intestino médio da mosca, os parasitas se transformam em tripomastigotas procíclicas e se multiplicam por fissão binária. 7- Após deixar o intestino médio da mosca, os parasitas se transformam em epimastigotas e atingem sua glândula salivar (8), onde se reproduzem por fissão binária. *T. brucei*, então, transforma-se em tripomastigota metacíclica. O inseto infectado transmite o protozoário para novo hospedeiro humano (1), dando continuidade ao ciclo de vida do *T. brucei*.  
Fonte: Adaptada de CDC<sup>26</sup>.

Além das organelas características de eucariotos, *T. brucei* apresenta uma bolsa flagelar de onde emerge um flagelo que garante sua mobilidade<sup>28</sup>, além de uma única mitocôndria alongada que se estende pela célula e possui uma região conhecida como cinetoplasto, que contém DNA mitocondrial.<sup>25</sup> Na forma sanguínea do *T. brucei*, o cinetoplasto se posiciona na região posterior da célula, enquanto na sua forma tripomastigota procíclica esta estrutura se localiza entre o núcleo e a região posterior. Já na sua forma epimastigota, o cinetoplasto é encontrado na região anterior ao núcleo<sup>27</sup> (Figura 1.5).



**Figura 1.5 – Arquitetura da célula do *T. brucei*.** A- Organelas e B- Posição do cinetoplasto e do flagelo nas diferentes formas celulares.

Fonte: Adaptada de MATTHEWS et al<sup>27</sup>.

Em mamíferos, o parasita sobrevive livremente na corrente sanguínea, já que é capaz de escapar da resposta imune do hospedeiro por meio de variação antigênica de glicoproteínas da sua superfície (do inglês, *variable surface glycoproteins*, VSGs). Estas proteínas estão ligadas à membrana celular na sua região C-terminal por meio de âncoras de glicosil-fosfatidilinositol (GPI). Entretanto, após a ingestão pelo inseto, as formas sanguíneas substituem este recobrimento por outro menos denso composto de prociclinas, que também são ancoradas por GPI.<sup>29</sup>

Na forma tripomastigota sanguínea, a mitocôndria tem uma estrutura tubular simples devido à ausência de respiração mitocondrial nesta fase, já que a geração de energia depende apenas de glicólise em organelas especializadas conhecidas como glicossomos<sup>27, 30</sup>. Entretanto, as formas tripomastigotas procíclicas dependem da presença de uma mitocôndria altamente ativa.<sup>27</sup>

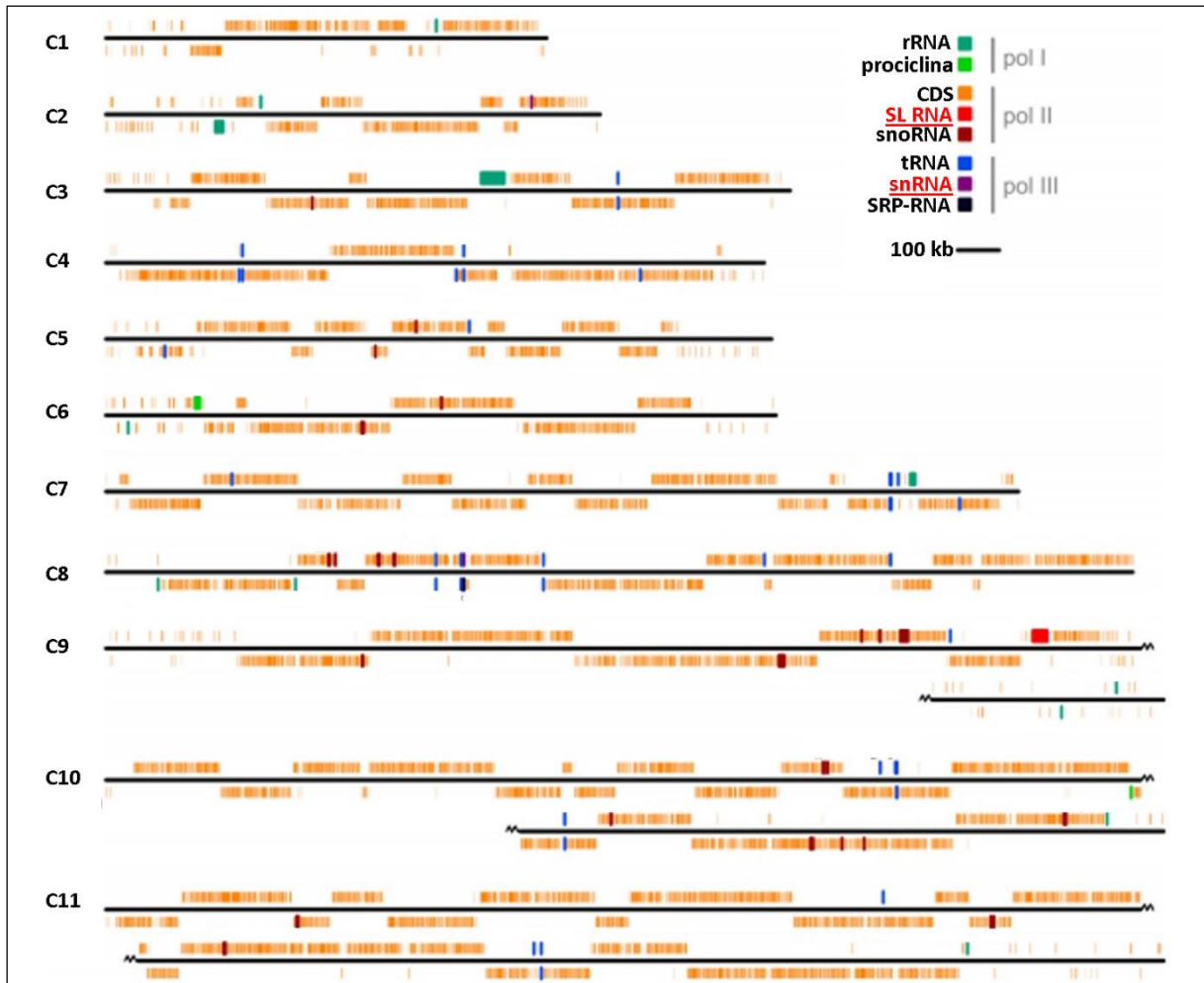
As formas epimastigotas geradas se ligam à parede da glândula salivar do inseto vetor por meio de sua membrana flagelar recoberta por proteínas ricas em arginina ancoradas por GPI, conhecidas como BARP (do inglês, *brucei alanine-rich protein*).<sup>31</sup> Após multiplicação, o parasita para de se dividir, readquire o recobrimento de VSG e é liberado ao lúmen da glândula salivar, onde está preparado para ser inoculado em um novo hospedeiro mamífero.<sup>27</sup>

### 1.2.2 Organização do genoma, replicação e transcrição

O conteúdo genético dos parasitas pertencentes à ordem Kinetoplastida encontra-se dividido entre o núcleo e o cinetoplasto. Em *T. brucei*, este corresponde a uma densa rede de DNA mitocondrial circular, na forma de maxicírculos (20 a 40 kb), que contém genes que codificam para proteínas mitocondriais, e minicírculos (0,5 a 1,0 kb), que codificam para pequenas moléculas de RNA que atuam como guias para a edição dos transcritos de maxicírculos por meio de incorporação ou deleção de uridinas.<sup>25, 27, 32</sup>

Entretanto, a maior parte de sua informação genética está compartimentalizada no núcleo, que possui aproximadamente 2,5  $\mu\text{m}$  de diâmetro. *T. brucei* é um exemplo de organismo diploide e possui de 26 a 35 Mb de DNA por genoma haploide, dependendo da cepa em análise. A região do genoma nuclear que corresponde aos cerca de oito mil genes que codificam para proteínas compreende 11 pares de cromossomos de tamanho entre 0,9 e 5,7 Mb. Adicionalmente, seu genoma nuclear contém por volta de 100 minicromossomos (30 a 150 kb) e cinco cromossomos de tamanho intermediário (200 a 900 pb), provavelmente aneuploides. Os minicromossomos abrigam um repertório de genes que codificam para VSG flanqueados por uma sequência repetitiva de 177 pares de base na região 5'.<sup>23, 27, 32-33</sup>

Vários aspectos da organização do genoma nuclear e regulação da expressão gênica de *T. brucei* não são usuais quando comparados a outros organismos eucariotos. O genoma nuclear é organizado em 388 agrupamentos direcionais de dois ou mais genes que codificam para proteínas (do inglês, *directional gene clusters*, DGCs), como pode ser observado na Figura 1.6. Cada gene individual é imediatamente seguido do próximo, podendo haver uma sequência intercistrônica de tamanho variável entre eles.<sup>23, 32</sup>



**Figura 1.6 – Representação esquemática do genoma nuclear de *T. brucei*.** Estão representados os elementos genéticos de acordo com sua classe (rRNA, prociclina, CDS, SLRNA, snoRNA, tRNA, snRNA, SRP-RNA) e a RNA polimerase que os transcreve (RNA polimerase I, II ou III). Barras coloridas indicam a posição e o tamanho de cada elemento genético de cada cromossomo de *T. brucei*. Barras acima da linha preta que representa um cromossomo (C1-C11) correspondem à transcrição no sentido da esquerda para a direita e barras abaixo correspondem à transcrição no sentido contrário. Fonte: Adaptada de DANIELS et al<sup>32</sup>.

A replicação do genoma nuclear de *T. brucei* é feita a partir de um pequeno número de origens de replicação, localizadas nas bordas de alguns DGCs.<sup>32, 34-36</sup> Algumas proteínas que fazem parte do complexo de iniciação de replicação foram identificadas e caracterizadas, porém ainda não foi apresentada a descrição completa do mecanismo de replicação deste organismo.<sup>35-</sup>

36

A transcrição dos genes desses agrupamentos é catalisada pela RNA polimerase II e é caracterizada como policistrônica, porém não se conhece precisamente a taxa de transcrição e as sequências de iniciação e terminação de transcrição da maioria dos DGCs.<sup>37</sup> A iniciação da transcrição está ligada às modificações em histonas em *T. brucei*.<sup>3</sup> Já a sua terminação ainda não é bem compreendida em *T. brucei*. Em outros tripanossomatídeos, a terminação da transcrição

parece estar associada à acumulação de um nucleotídeo modificado conhecido como J ( $\beta$ -D-glucopiranosiloximetiluracila).<sup>32-33, 39</sup>

A enzima RNA polimerase I transcreve não apenas as moléculas de RNA ribossomal (rRNA) como também genes que codificam para VSG e prociclina. Já a enzima RNA polimerase III transcreve um pequeno número de RNAs que são codificadores para proteínas que atuam no processo de tradução (tRNAs e rRNAs 5S), processamento de mRNA (snRNAs U1, U2, U4 e U5, que serão detalhados posteriormente) e outros processos celulares (como o RNA 7SL).<sup>32-33, 39</sup> Em *T. brucei*, os snRNAs U1, U2 e U4 recebem um capacete de 7-metilguanossina ( $m^7G$ ) hipermetilado na extremidade 5' como sinal de exportação ao citoplasma, com exceção do snRNA U5.<sup>20</sup>

Em contraste, os snRNAs U1, U2, U4 e U5 de mamíferos e levedura são transcritos pela RNA polimerase II, assim como cada pré-mRNA, e adquirem um capacete de 7-metilguanossina ( $m^7G$ ) na extremidade 5' como sinal de exportação ao citoplasma. Entretanto, tanto em *T. brucei* quanto nesses organismos, o snRNA U6 é transcrito pela RNA polimerase II, recebe um capacete de  $\gamma$ -monometilfosfato na extremidade 5' e esta molécula não deixa o núcleo celular.<sup>20,</sup>

40-41

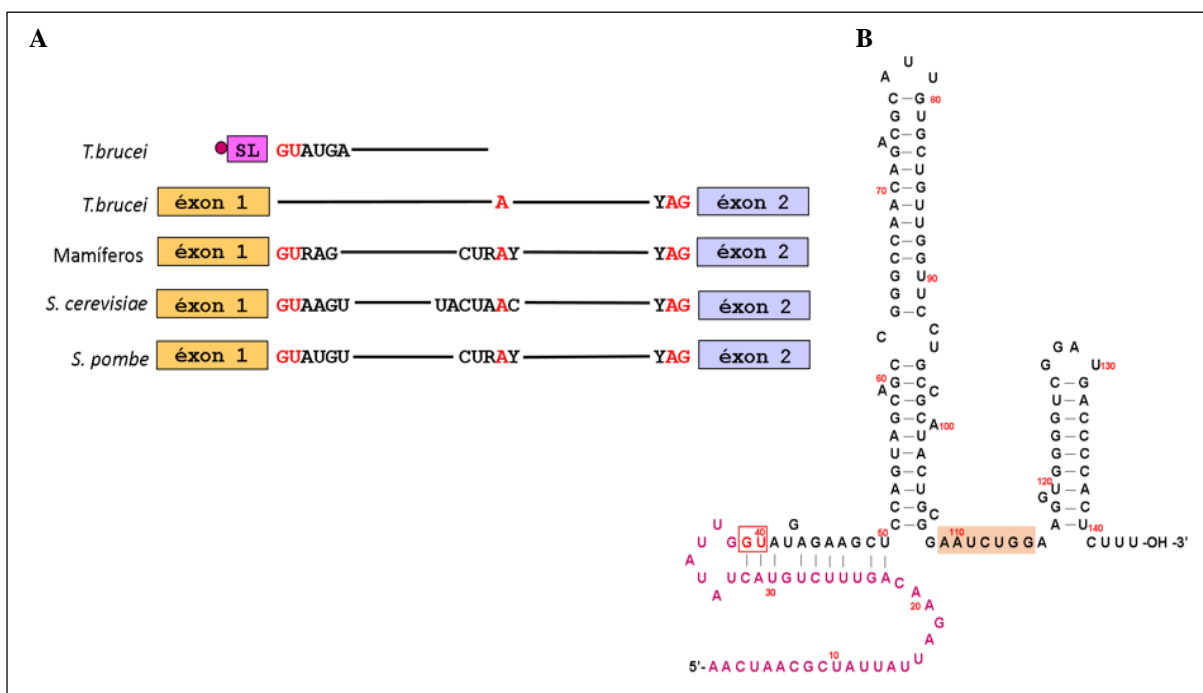
### 1.2.3 Processamento do pré-mRNA

O processamento dos transcritos policistrônicos a mRNAs monocistrônicos maduros envolve o SL *trans-splicing*, que consiste na ligação, na extremidade 5' de cada mRNA individual, de um pequeno éxon de RNA contendo capacete-5' hipermetilado de 7-metilguanossina (cap4), conhecido como, do inglês, *spliced leader* (SL). Esta molécula é transcrita pela RNA polimerase II em *T. brucei*<sup>39</sup> e o SL *trans-splicing* ocorre em poucos minutos após a sua síntese, concomitante à poliadenilação de sua extremidade 3'.<sup>42</sup> Duas únicas exceções são o gene PAP (Tb927.3.316) de *T. brucei*, codificante para a poli-A polimerase, e um gene putativo para uma RNA helicase (Tb927.8.1510), que apresentam um único íntron removido por *cis-splicing*.<sup>20, 43-45</sup>

Os pré-mRNAs contêm sinais consenso (Figura 1.7) para o processamento por SL *trans-splicing* ou por *cis-splicing* catalisados pelo spliceossomo. Entre eles, pode-se observar um sítio específico na extremidade 3' do éxon, conhecido como 3'SS (do inglês, 3' *splice site*), que apresenta o dinucleotídeo AG na junção entre um éxon e a próxima região intercistrônica, além de uma adenina dentro do éxon, marcando o ponto de ramificação (do inglês, *branch point*,

BP). Um trato de polipirimidinas (Py) é observado jusante a este (~20 nucleotídeos). Em leveduras, esta região não é observada.<sup>20, 43, 45</sup>

Diferente de eucariotos derivados, o sítio 5' de *splicing* (5'SS) está localizado no RNA SL, como mostrado na figura 1.7A. O éxon SL tem 39 nucleotídeos em *T. brucei* e o RNA SL tem 135 nucleotídeos, contendo, além do éxon SL, uma região não codificante de tamanho variável que forma duas estruturas secundárias em forma de grampo de cabelo, separadas por uma região de reconhecimento por proteínas Sm (Figura 1.7). Além disso, não são verificadas sequências de nucleotídeos consenso para o 3'SS, BP ou 5'SS entre tripanossomatídeos.<sup>20, 43, 45-46</sup>



**Figura 1.7 – Sinais de *splicing* no pré-mRNA.** A- Representação esquemática dos sinais de *splicing* consenso (em vermelho) no pré-mRNA de *T. brucei*, de mamíferos e de leveduras. B- Estrutura secundária proposta para SL RNA de *T. brucei*. Em roxo, identifica-se o éxon de SL RNA, que apresenta um capacete 5' de 7-metilguanosina hipermetilado (cap4). O sítio Sm do SL RNA está destacado, bem como a sequência GU do 5'SS.

Fonte: Adaptada de CROSS et al<sup>46</sup>.

O requerimento da adição de RNA SL ao pré-mRNA está relacionado à modulação dos níveis de mRNA maduro na célula. O RNA SL deve ser produzido em alta concentração para garantir o processamento de todos os transcritos. Esta molécula é transcrita na forma de pré-RNA SL, contendo um éxon e um íntron, e é codificada em um arranjo de repetições em *tandem* de 1,4 kb localizado no cromossomo 9 (Figura 1.6), que também é transcrito pela RNA polimerase II. Este é o único caso em que o promotor da RNA polimerase II está bem

caracterizado, mostrando-se bem conservado entre tripanossomatídeos.<sup>47-48</sup> Adicionalmente, outro mecanismo de regulação corresponde à estabilidade do transcrito maduro.<sup>32, 37</sup>

Concomitante ao SL *trans-splicing*, a adição de uma extensão poli-A na extremidade 3' dos transcritos maduros catalisada pelo complexo da poli-A polimerase (PAP) é bem conservada em tripanossomatídeos e é importante para a estabilização do mRNA, sua exportação para o citoplasma e tradução.<sup>37, 42</sup>

### 1.3 Spliceossomo: arquitetura e função no *cis*- e no SL *trans-splicing*

O spliceossomo é um complexo nuclear que interage com o RNA SL e o transcrito policistrônico nos sítios específicos de *splicing* para catalisar o processo de SL *trans-splicing* em tripanossomatídeos.<sup>20, 43-44</sup> No caso do *cis-splicing*, mais bem caracterizado em humanos e leveduras, o spliceossomo interage nos sítios localizados tanto na região de íntron quanto nas regiões de éxons (Figura 1.7).<sup>40, 49-52</sup> Em ambos os processos, este complexo é associado pela interação de pequenas partículas ribonucleoproteicas nucleares ricas em uridina (do inglês, *small nuclear ribonucleoparticules*, snRNPs - U1, U2, U5, U4/U6, U4/U6.U5 – descritas na seção 1.3.2), além de algumas subunidades multiproteicas (RES, NTC, IBC, EJC e NTR – descritas na seção 1.3.3). A nomenclatura das proteínas spliceossomais foi definida em estudos de *cis-splicing* de humanos e leveduras e é adotada pela comunidade científica também para tripanossomatídeos<sup>20, 40, 43</sup>.

Nos últimos 30 anos, estudos sobre o spliceossomo de humanos, leveduras e, mais recentemente de tripanossomatídeos, identificaram que as snRNPs são formadas por pequenos RNAs nucleares (do inglês, *small nuclear RNAs*, snRNAs - U1, U2, U4, U5 e U6) cujas sequências são ricas em uridina, e fatores proteicos específicos associados, como as proteínas Sm, que formam um anel hetero-heptamérico em torno dos snRNAs (seção 1.3.2) e recrutam proteínas específicas para a biogênese das snRNPs (seção 1.3.2.4). As pesquisas atuais em relação ao spliceossomo humano e de levedura se concentram em descrever a arquitetura do complexo em nível molecular e a relação com sua função enzimática, além da identificação dos pontos de regulação.<sup>20, 40</sup>

Os tripanossomatídeos possuem todos os cinco snRNAs (seção 1.3.2), porém o conhecimento acerca de fatores proteicos associados ao spliceossomo destes organismos ainda é limitado devido à divergência de suas sequências de aminoácidos com relação àqueles já bem caracterizados do spliceossomo de humanos e leveduras. Poucas proteínas foram diretamente identificadas pela anotação do genoma completo de *T. brucei*, *T. cruzi* e *L. major*, disponível



na base de dados TriTryp.<sup>53</sup> Os esforços das pesquisas acerca do *trans*-spliceossomo ainda se concentraram em identificar e caracterizar os complexos spliceossomais em termos de sua composição e função.<sup>20, 43, 45</sup>

A análise extensiva de sequências de aminoácidos das proteínas de *T. brucei* que conservam o domínio de reconhecimento de RNA (do inglês, *RNA recognition motif*, RRM) e o desenvolvimento de técnicas de purificação por afinidade em *tandem* (do inglês, *tandem affinity purification*, TAP) de proteínas spliceossomais combinadas com espectrometria de massas possibilitaram a identificação de muitas das proteínas que compõem o spliceossomo destes organismos. Entretanto, ainda não existem trabalhos publicados com a análise da relação estrutura-função das proteínas spliceossomais ou dos subcomplexos associados ao *trans*-spliceossomo de tripanossomatídeos.<sup>20, 43</sup>

Além da complexidade da composição do spliceossomo, este também é dinâmico para possibilitar a catálise de duas reações de transesterificação.<sup>49</sup> Em humanos, além dos snRNAs das partículas U1, U2, U4, U5 e U6, cerca de 170 proteínas estão relacionadas ao spliceossomo, porém nem todas são recrutadas para as duas reações de transesterificação<sup>40, 49, 54</sup>, como será descrito a seguir.

### 1.3.1 Catálise do *splicing*

O spliceossomo é associado pela adição sequencial de snRNPs ao pré-mRNA e ao RNA SL, no caso do SL *trans-splicing*<sup>20, 43</sup> (Figura 1.8), e somente ao pré-mRNA no caso do *cis-splicing*<sup>49-50</sup> (Figura 1.8), em um processo altamente complexo e dinâmico, com composição variada de snRNAs e fatores proteicos a cada etapa da catálise, como se observa na Figura 1.8. Exceto pela natureza intermolecular, o SL *trans-splicing* apresenta as mesmas etapas do *cis-splicing*, com duas reações de transesterificação.<sup>20, 40</sup>

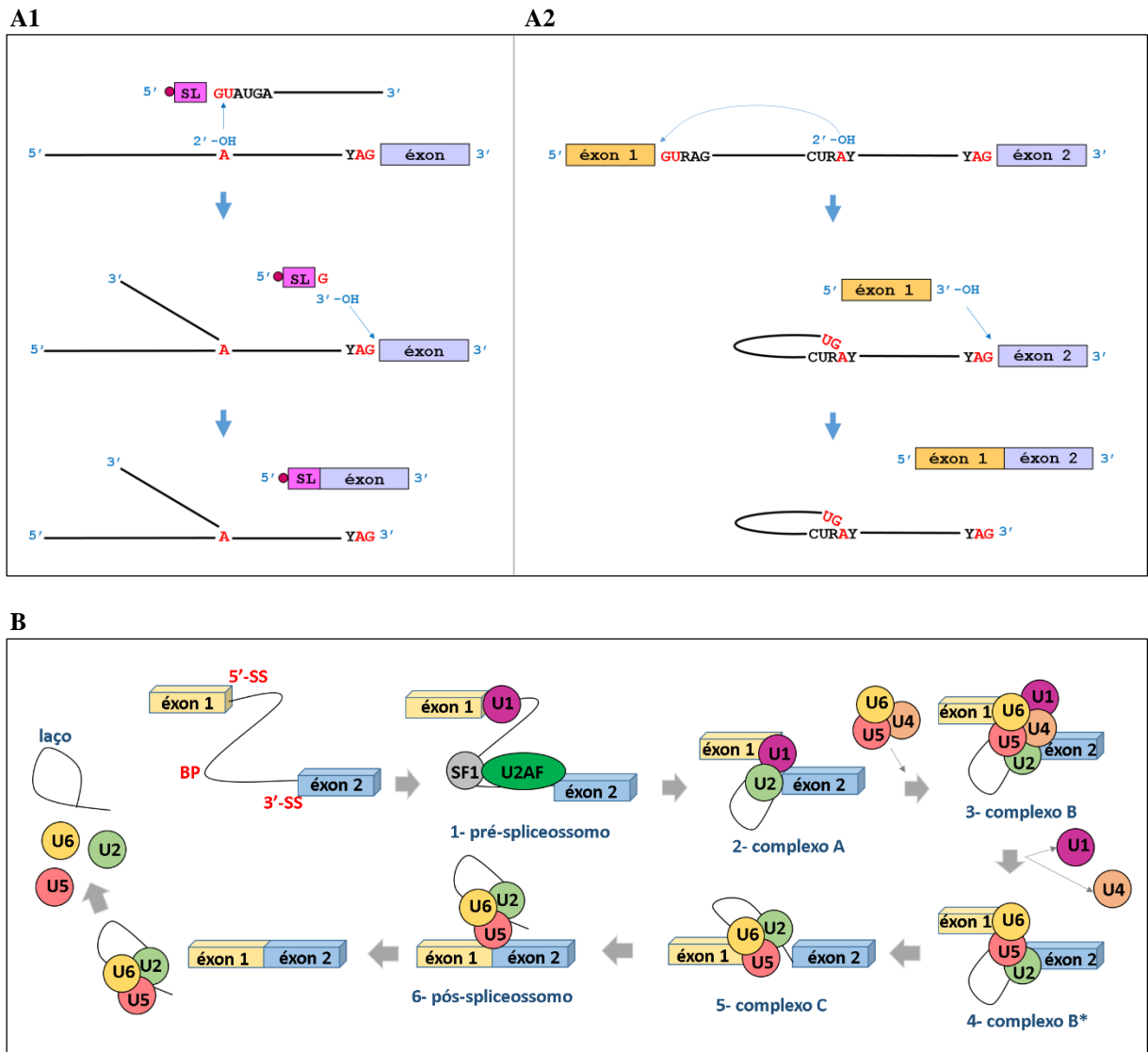
Inicialmente, propõe-se que o pareamento de bases do 5'SS presente no RNA SL com o 3'SS do pré-mRNA seja suficiente para definir o sítio de *splicing* 5' no SL *trans-splicing*<sup>20</sup>, o que difere do *cis-splicing*, onde observa-se o pareamento de bases entre o snRNA U1 e o sítio 5'SS no pré-mRNA.<sup>40, 55</sup> O fator proteico específico U2AF (Mud2 em levedura), na presença de SF1 (Msl5 em levedura), por sua vez, reconhece tanto o 3'SS como o BP, ambos no pré-mRNA. A partícula U2 é, então, recrutada e reconhece a sequência do BP. Nesse estágio é formado o complexo A<sup>20, 40</sup> (Figura 1.8).

Em seguida, um heterocomplexo formado por di-snRNP U4/U6 e snRNP U5, chamado tri-snRNP U4/U6.U5<sup>56</sup>, associa-se ao complexo A, formando o complexo B (Figura 1.8). Há,

então, um rearranjo estrutural, incluindo a entrada do complexo NTC (do inglês *nineteen complex*) e a saída da snRNP U4 (e U1, no caso do *cis-splicing*), o que leva à formação do complexo B<sup>at</sup> (spliceossomo ativado).<sup>20, 40, 49</sup> Para esta ativação, a proteína U5-200K (Brr2 em levedura) tem papel catalítico no desapareamento do *duplex* de snRNAs U4/U6, enquanto U5-100K (Prp28 em levedura) está associada à liberação de snRNP U1.<sup>57</sup> A investigação da função de U5-200K na snRNP U5 é um dos alvos desse trabalho.

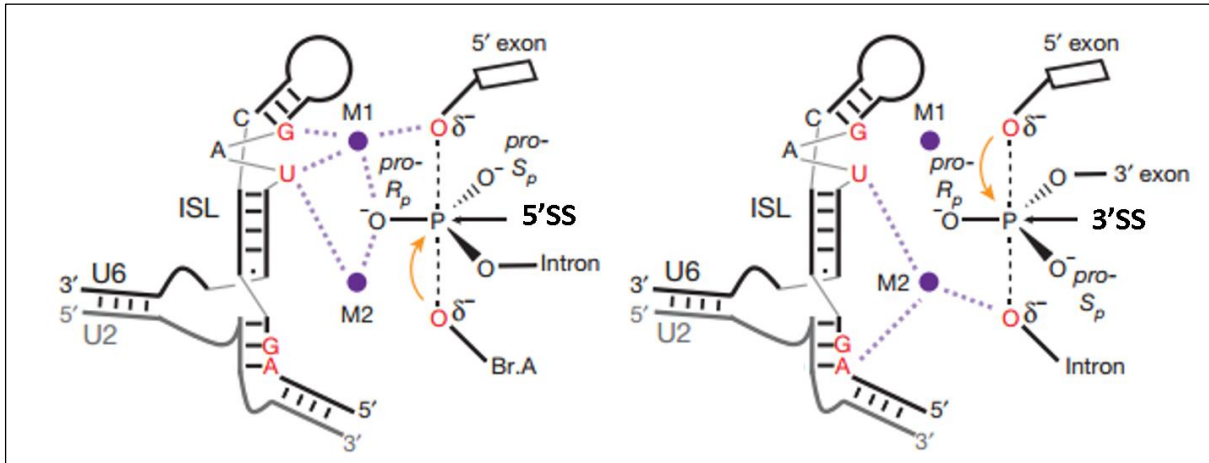
Posteriormente, há um rearranjo estrutural para formação do complexo B\*, que corresponde ao spliceossomo cataliticamente ativo. Ocorre então a primeira reação de transesterificação, em que o grupamento 2'-OH da adenosina do BP faz um ataque nucleofílico no átomo de fósforo da guanina da porção final do éxon do pré-mRNA no sítio 5'SS. Com a transferência da ligação fosfodiéster, há a formação do éxon SL 5' e a ligação do SL remanescente à adenosina do BP no pré-mRNA acceptor, resultando na formação de uma estrutura intermediária na forma de Y, e o spliceossomo é transformado no complexo C (Figura 1.8).<sup>20, 40, 49</sup> No caso do *cis-splicing*, o éxon 5' é ligado à adenosina do BP formando uma estrutura na forma de laço<sup>40</sup> (Figura 1.8).

Finalmente, ocorre a segunda reação de transesterificação, a partir do ataque nucleofílico do grupamento 3'-OH da sequência do éxon SL ao grupamento fosfato da extremidade 3' da junção intercistrôncia/éxon no complexo P. Os produtos são, portanto, o RNA maduro e uma estrutura na forma de Y.<sup>20</sup> Já no caso do *cis-splicing*, ocorre a reação equivalente para liberação do mRNA maduro e da estrutura na forma de laço<sup>40</sup> (Figura 1.8). Após esta etapa, o spliceossomo é dissociado e as snRNPs são recicladas para novas etapas de *splicing*.<sup>20, 40, 58</sup>



**Figura 1.8 – Mecanismos de *splicing* em eucariotos.** A1 e A2- Reações de SL *trans-splicing* e *cis-splicing*, respectivamente. B- Dinâmica de associação do spliceossomo durante o *cis-splicing*.  
Fonte: Adaptada de GUNZL et al<sup>20</sup>; WAHL<sup>49</sup>.

Durante as duas etapas da reação de *splicing*, o spliceossomo atua como uma ribozima, sendo que o snRNA U6 é responsável pelo posicionamento de dois íons  $Mg^{2+}$  (M1 e M2) que estabilizam os grupos abandonadores das reações de substituição nucleofílica do tipo SN2.<sup>52, 59</sup> Como se pode observar na Figura 1.9, o íon M2 ativa o nucleófilo e M1 estabiliza o grupo abandonador durante a primeira etapa. Já na segunda etapa, os papéis se invertem e M1 ativa o nucleófilo enquanto M2 estabiliza o grupo abandonador<sup>52, 59</sup> (Figura 1.9).



**Figura 1.9 –** Catálise das reações de transesterificação (A e B) a partir do posicionamento de íons  $Mg^{2+}$  pelo snRNA U6 para estabilizar os intermediários das reações.  
Fonte: FICA et al.<sup>59</sup>.

Muitos dos componentes proteicos do spliceossomo são indispensáveis para a catálise do *splicing*, seja atuando como helicase/ATPases ou como plataforma de interações para direcionar os snRNAs e os substratos para a catálise.<sup>52, 54</sup> Estes componentes e a respectiva função serão descritos nas próximas seções, em comparação com os fatores proteicos identificados para o *SL trans-splicing*.

Os complexos E, A, B, B<sup>at</sup>, B<sup>\*</sup> e C do spliceossomo não foram identificados em tripanossomatídeos até o momento.<sup>20</sup> Entretanto, muitas das proteínas relacionadas foram anotadas e parte delas caracterizadas. Adicionalmente, proteínas spliceossomais específicas de tripanossomatídeos foram identificadas, sendo que muitas dessas ainda não têm função definida.<sup>20, 60- 61</sup>

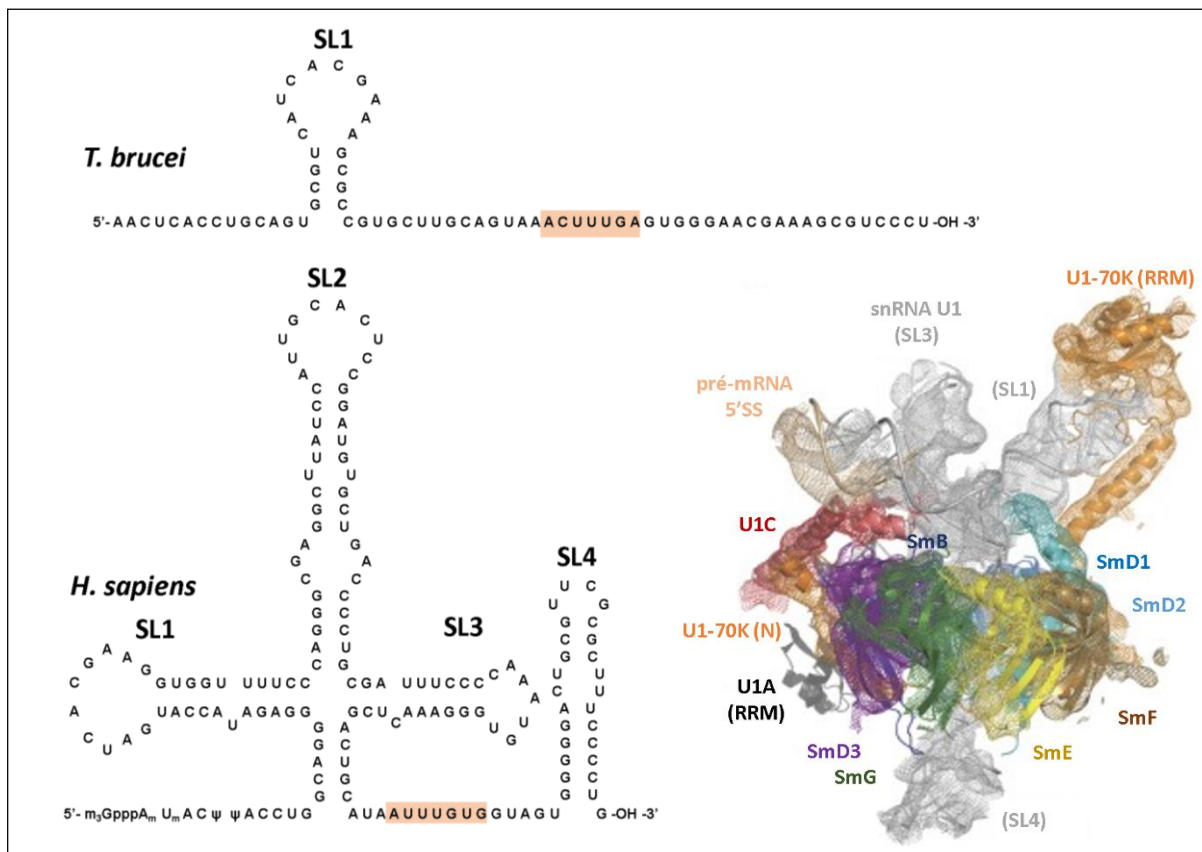
### 1.3.2 snRNPs: composição, função e biogênese

#### 1.3.2.1 snRNP U1

A partícula ribonucleoproteica U1 (snRNP U1) de humanos é caracterizada com um coeficiente de sedimentação igual a 12S e consiste de uma molécula de snRNA U1 com 165 nucleotídeos de extensão (Figura 1.10) e 10 proteínas associadas, entre elas as 7 proteínas Sm comuns a todas as snRNPs de humanos (SmB, SmD3, SmG, SmE, SmF, SmD2, SmD1), além de U1-70K, U1A e U1C.<sup>49, 55</sup> U1C estabiliza o pareamento de bases entre a extremidade 5' de snRNA U1 e a sequência do 5'SS<sup>55, 62</sup> (Figura 1.10).

Em leveduras, snRNP U1 é mais complexa e consiste de um snRNA U1 maior, com 568 nucleotídeos, além das 7 proteínas Sm e 10 proteínas adicionais (Prp39, Prp40, Snu71, Snu56, Snp1, Mud1, Luc7, Prp42, Nam8 e Yhc1). Snp1, Mud1 e Yhc1 são homólogas a U1-70K, U1A e U1C, respectivamente.<sup>40, 63</sup> Já no caso de tripanossomas, snRNP U1 varia em muitos aspectos. Primeiramente, U1-70K foi identificada a partir da sua identidade de sequência de aminoácidos com U1-70K humana<sup>64</sup> e, após purificação em *tandem* a partir da sua expressão recombinante em *T. brucei*, foram identificadas as proteínas U1C, cuja sequência de aminoácidos diverge da homóloga humana, e a proteína U1-24K, ausente em humanos e leveduras.<sup>65</sup> Neste experimento, as proteínas Sm canônicas também foram copurificadas. A proteína U1A, por sua vez, só foi identificada pela análise de sequências que conservam o domínio RRM em *T. brucei*.<sup>60</sup>

O snRNA U1 de *T. brucei* tem apenas 75 nucleotídeos e não possui as estruturas equivalentes a SL2 e SL3 (Figura 1.10), que tipicamente interagem com U1A.<sup>20, 65</sup> Estudos recentes de imunoprecipitação (do inglês, *individual-nucleotide resolution crosslinking-immunoprecipitation*, iCLIP) mostraram que não apenas o snRNA U1 mas também o RNA SL e o snRNA U6 interagem com U1C e U1-70K.<sup>66</sup> Adicionalmente, a ligação cruzada de U1C ao pré-mRNA do gene de PAP (que sofre *cis-splicing* em *T. brucei*) indica a atividade de snRNP U1 no reconhecimento do 5'SS, assim como em humanos e leveduras.<sup>66</sup> Deste modo, snRNP U1 deve representar uma ligação evolutiva entre as maquinarias de *cis-* e *SL trans-splicing*, sendo essencial para o reconhecimento do sítio 5' de *splicing* no SL RNA e nos *cis*-íntrons do pré-mRNA.



**Figura 1.10 – snRNP U1.** Estruturas secundárias de snRNAs U1 de *T. brucei* e *H. sapiens* e modelo cristalográfico de snRNP U1 de *H. sapiens*<sup>55</sup>. Os sítios Sm estão evidenciados em vermelho.  
Fonte: Adaptada de PALFI et al<sup>67</sup>; KONDO et al<sup>55</sup>.

### 1.3.2.2 snRNP U2

A snRNP U2 de humanos constitui-se de uma molécula de snRNA U2 e 7 proteínas Sm em torno do sítio Sm. Adicionalmente, este núcleo está associado às proteínas U2A' e U2B'<sup>68-69</sup> e aos complexos proteicos SF3a<sup>70</sup> e SF3b<sup>71</sup>, formando uma partícula de coeficiente de sedimentação igual a 17S. O complexo SF3a consiste das subunidades SF3a120, SF3a66 e SF3a60, e o complexo SF3b é composto de SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b e SF3b10, todas nomeadas de acordo com sua massa molecular aparente. Adicionalmente, CHERP, hPrp5, hPrp43, SPF31 e SPF45 estão associadas à partícula U2 em pequenas quantidades.<sup>49, 72-73</sup>

Esta partícula participa do reconhecimento da região BP do pré-mRNA no *cis-splicing* por meio de pareamento de bases com snRNA U2. Em seguida, é responsável por posicionar a adenosina do BP no centro catalítico do spliceossomo e, juntamente com o snRNA U6, o snRNA U2 forma uma rede de interações que posicionam os sítios reativos do pré-mRNA para que as reações de transesterificação aconteçam.<sup>20, 40, 49, 52, 54, 74</sup>

O snRNA U2 de humanos tem 187 pares de base e forma cinco estruturas de grampo-de-cabelo (SL1-4)<sup>72</sup> (Figura 1.11). A região 5' é altamente conservada e contém, além dessas estruturas, uma região de fita simples para interação com o snRNA U6, o íntron e proteínas específicas de snRNPs. Estruturas de ressonância magnética nuclear do SL1 do snRNA U2 de *H. sapiens* e *S. cerevisiae* mostram similaridade estrutural, apesar das modificações pós-transcricionais observadas para o snRNA U2 de humanos, em detrimento de *S. cerevisiae*, tornarem-no termodinamicamente mais estável.<sup>72</sup>

As regiões que flanqueiam o sítio de reconhecimento de BP (5'GUAGUA3', localizada entre SL1 e SL2) e SL1 são complementares ao snRNA U6, permitindo a formação de duas hélices intermoleculares.<sup>72</sup> Sugere-se que a estrutura SL1 tenha papel relevante no sítio ativo do spliceossomo. Adicionalmente, U2B'' se liga a SL4 e U2A' interage com U2B'', sendo que essa interação é essencial para a formação de spliceossomo estável em humanos e leveduras.<sup>68</sup> O complexo U2A'/U2B''/snRNA U2 será discutido em detalhes no capítulo 2 deste trabalho.

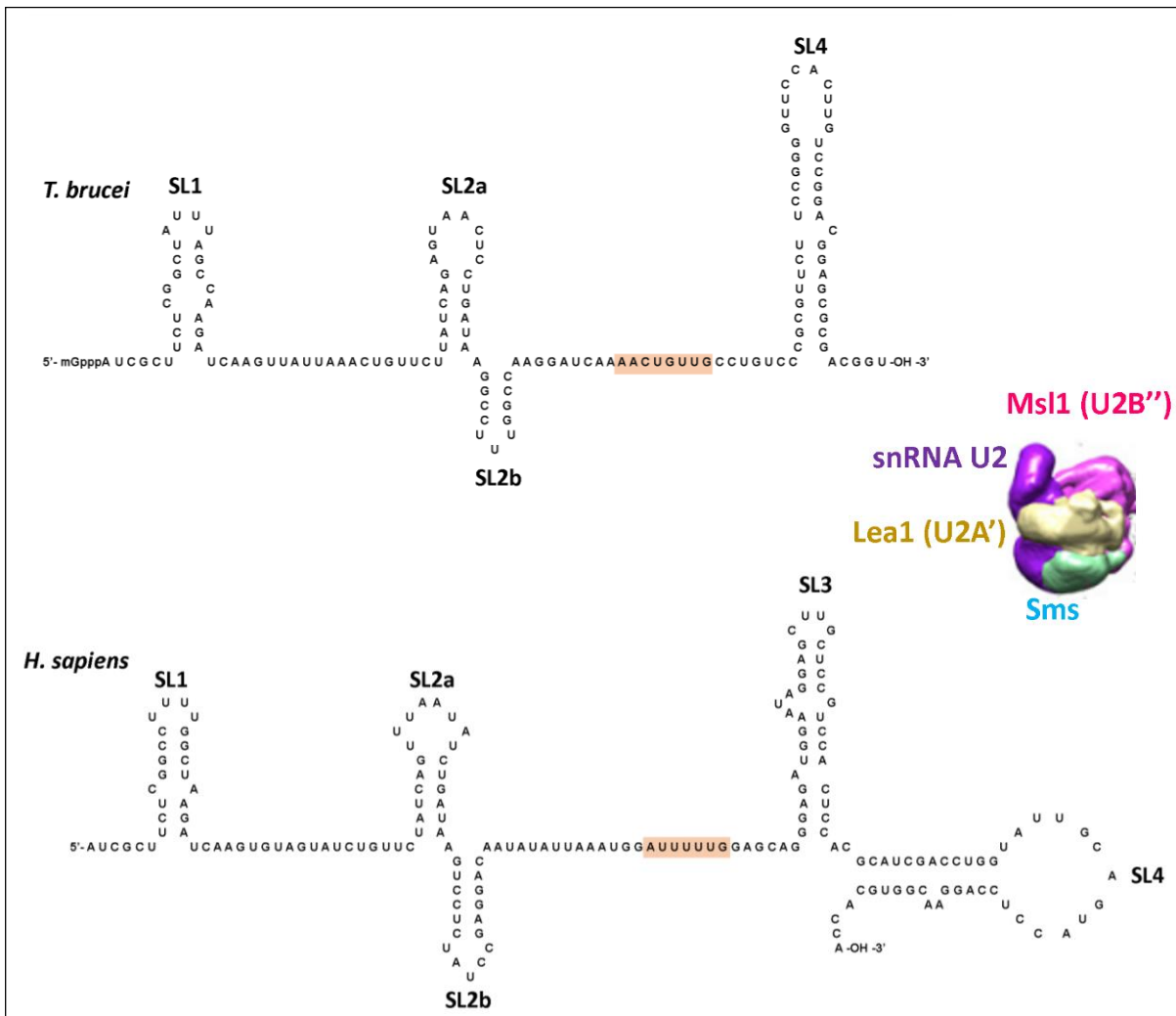
Recentemente, a estrutura do spliceossomo de *S. pombe*<sup>54</sup> a 3,6 Å de resolução máxima, obtida por criomicroscopia eletrônica, revelou a arquitetura de todas as nove proteínas de U2 snRNP, que compreende Lea1 (U2A' em humanos), Msl1 (U2B'' em humanos), anel Sm heptamérico, além do snRNA U2, no braço I do spliceossomo deste organismo.<sup>54</sup> Esta região está ligada ao corpo central por meio da interação entre as proteínas Cwf3/Syf1 e Cwf4/Syf3 e pelo snRNA U2 que, em complexo com o snRNA U6 e com o íntron na forma de laço, posiciona-se no sítio catalítico.<sup>54</sup> Entretanto, esta região na estrutura tem informação estrutural em baixa resolução (~10 Å)<sup>54</sup> (Figura 1.11).

O snRNA U2 de tripanossomatídeos (Figura 1.11) também é menor que o observado em humanos e leveduras, com apenas 150 nucleotídeos. Estão ausentes a estrutura SL3 e a sequência de reconhecimento de BP. A sequência de ligação a proteínas Sm não é usual e apresenta uma adenosina central.<sup>75-76</sup> Não foi observada uma sequência consenso para BP entre membros da família *Trypanosomatidae*, o que provavelmente reflete em um mecanismo diferente de reconhecimento do BP e do 3'SS durante o SL *trans-splicing*. Deve-se notar que a substituição da região de interação ao BP do snRNA U2 de levedura pela região equivalente de *T. brucei* resultou em spliceossomo não-funcional.<sup>20, 77</sup>

A primeira proteína específica de snRNP U2 a ser identificada em *T. brucei* foi a homóloga de U2A' (também conhecida como U2-40K), a partir da sua copurificação com o snRNA U2 em condições de alta estringência.<sup>67, 78</sup> A proteína U2B'', por sua vez, foi identificada por meio de análise de sequências de aminoácidos e observou-se que ela interage com a proteína U2A' como as homólogas de humanos e leveduras.<sup>79</sup> Em seguida, análises de

conservação do domínio RRM identificaram o fator proteico homólogo a SF3b49, membro da subunidade SF3b, cujo silenciamento gênico é letal para *T. brucei*.<sup>80</sup> A purificação desta proteína a partir do *T. brucei* seguida de espectrometria de massas levou à identificação de todas as proteínas da subunidade SF3b.<sup>20, 80</sup> Além disso, a única proteína de *T. brucei* anotada como homóloga a SF3a60 foi recentemente caracterizada, bem como seus parceiros de interação para formação da subunidade SF3a.<sup>81</sup>

A composição proteica da snRNP U2 de tripanossomatídeos também diverge. Foram identificadas as proteínas Sm15K e Sm16,5K em substituição a SmB e SmD3, respectivamente, no anel hetero-heptamérico que interage com o sítio Sm do U2 snRNA.<sup>82, 83</sup> Adicionalmente, a interação entre essas proteínas e U2A' foi postulada a partir de experimentos de reconstrução *in vitro* dos complexos proteínas-RNA.<sup>82</sup> A variação do anel Sm na snRNP U2 de *T. brucei* é uma característica exclusiva deste organismo e é um dos focos deste trabalho (capítulo 2).



**Figura 1.11 – snRNP U2.** Estruturas secundárias de snRNAs U2 de *T. brucei* e *H. sapiens* e densidade eletrônica do núcleo de snRNP U2 de *S. pombe* a ~10 Å de resolução obtida por criomicroscopia eletrônica.<sup>54</sup> Os sítios Sm estão evidenciados em vermelho.

Fonte: Adaptada de HARTSHORNE E AGABIAN<sup>76</sup>; SASHITAL et al<sup>72</sup>; YAN et al<sup>54</sup>.



### 1.3.2.3 snRNP U5, di-snRNP U4/U6 e tri-snRNP U4/U6.U5

A snRNP U5 é a maior ribonucleoproteína do spliceossomo humano, com coeficiente de sedimentação igual a 20S, sendo composta por uma molécula de snRNA U5 com 116 nucleotídeos de extensão (Figura 1.12-B2), além das sete proteínas Sm e oito proteínas específicas: U5-220K (hPrp8), U5-200K (hBrr2), U5-116K (hSnu114), U5-102K (hPrp6), U5-100K (hPrp28), U5-52K (hLin1), U5-40K e U5-15K (hDib1).<sup>56, 84</sup>

O snRNA U5 apresenta a estrutura na forma de grampo-de-cabelo SL1 altamente conservada (Figura 1.12-B1 e B2) que interage com sequências de éxons nos sítios 5'SS e 3'SS. Esta molécula está relacionada à estabilização do intermediário de *splicing* após a primeira reação de transesterificação e no alinhamento dos éxons para a segunda etapa catalítica.<sup>85,86</sup> As proteínas, por sua vez, estão associadas com a ativação e regulação do spliceossomo.<sup>56</sup>

U5-102K interage com U5-52K que, por sua vez, interage com U5-15K na partícula U5 (Figura 1.12-A1 e A2).<sup>87-88</sup> Já as proteínas U5-100K e U5-200K são helicases dependentes de ATP importantes no rearranjo da rede de interações de snRNAs para a ativação do spliceossomo.<sup>50-51, 56</sup> Vários estudos genéticos mostraram que Prp28 de levedura (homóloga a U5-100K de humanos) promove a liberação da partícula U1, processo coordenado com a liberação de snRNP U4 pela Brr2 (homóloga a U5-200K de humanos) e está envolvida na incorporação de tri-snRNP U4/U6.U5 ao spliceossomo.<sup>89</sup> U5-220K/Prp8 e U5-116K/Snu114, por sua vez, estão relacionadas à regulação da atividade de U5-200K/Brr2.<sup>50-51, 56</sup>

A partícula U6, por sua vez, não é usual e seu snRNA é o único transcrito pela RNA polimerase III em humanos. Esta partícula é encontrada em interação com snRNA U4, formando a di-snRNP U4/U6 (13S), a qual, por sua vez, é predominantemente encontrada em interação com snRNP U5, na forma de tri-snRNP U4/U6.U5 (25 S). O snRNA U6 (Figura 1.12-A1 e A2) possui 106 nucleotídeos e forma uma estrutura de grampo-de-cabelo muito parecida com o domínio catalítico V de íntrons do grupo II. Essa molécula se associa a sete proteínas LSM, que também formam um anel em torno do sítio Sm na sua extremidade 3'.<sup>90</sup> Já o snRNA U4 possui 142 nucleotídeos e interage com as sete proteínas Sm.<sup>91</sup> A estrutura do núcleo Sm da partícula U4 publicada recentemente<sup>91</sup> revela sítios de interação com snRNA U4 equivalentes aos observados no núcleo Sm da snRNP U1<sup>55</sup> (informação a ser detalhada no capítulo 2). Adicionalmente, a di-snRNP U4/U6 contém cinco proteínas específicas: 90K (hPrp3), 61K (hPrp31), 60K (hPrp4), 20K (CypH) e 15,5K (hSnu13)<sup>56, 92</sup>. As proteínas 90K, 60K e 20K formam um sub-complexo estável que é recrutado pela di-snRNP U4/U6 e está relacionado

com a estabilização da hélice II do duplex snRNA U4/U6.<sup>92</sup> Após a ativação do spliceossomo, snRNA U6 interage com o 5' SS e se alinha com snRNA U2.<sup>56, 92</sup>

Além das proteínas específicas de cada partícula, três outras são recrutadas para a formação da tri-snRNP U4/U6.U5: 110K (hSnu66), 65K (hSad1) e 27K.<sup>92</sup> As proteínas de tri-snRNP são bem conservadas e quase todas estão presentes em leveduras, com exceção de 27K, 40K e CypH.<sup>40, 50, 56</sup> U5-102K/Prp6 é necessária para a formação de tri-snRNP estável<sup>93</sup>, enquanto que U5-52K se dissocia após a formação de tri-snRNP U4/U6.U5.<sup>88</sup>

Recentemente, a estrutura de criomicroscopia eletrônica da tri-snRNP U4/U6.U5 de levedura, a 3,7 Å de resolução<sup>94</sup>, associada às estruturas de maior resolução de suas subunidades, revelaram a arquitetura deste complexo em forma triangular de aproximadamente 300 Å de dimensão máxima.<sup>56, 94-95</sup> A maior região de Prp8, que consiste de domínio semelhante a transcriptase reversa (do inglês, *reverse transcriptase*, RT), espaçador e do domínio semelhante a endonuclease do tipo-III, está localizado no centro do complexo e é identificado como uma plataforma para interações proteína-proteína e proteína-RNA que mantêm o complexo unido.<sup>51, 56, 96</sup> (Figura 1.12 A1-A2).

Na região N-terminal de Prp8 ocorre interação com Snu114, na qual foi identificada uma molécula de GTP ligada, e snRNA U5, que por sua vez interage com sete proteínas Sm. Na sua região C-terminal, Prp8 interage com Brr2 através de seu domínio Jab1/MPN, além do complexo Prp3-Prp4. A atividade GTPase ou de ligação de GTP de Snu114 foi sugerida como reguladora da função de Brr2 em levedura.<sup>56, 97-99</sup> (Figura 1.12 A1-A2).

Hélices de snRNAs U4/U6 e o grampo-de-cabelo da extremidade 5' de snRNA U4 foram identificados próximos ao cassete helicase N-terminal de Brr2 e da região N-terminal de Prp8. Snu13 e o domínio Nop de Prp31 se ligam à hélice da região 5' do snRNA U4 e o anel Sm de snRNP U4 interage na região entre os dois cassetes de Brr2. As proteínas LSm de snRNP U6, por sua vez, foram localizadas na região do braço flexível.<sup>56, 97</sup> (Figura 1.12 A1-A2).

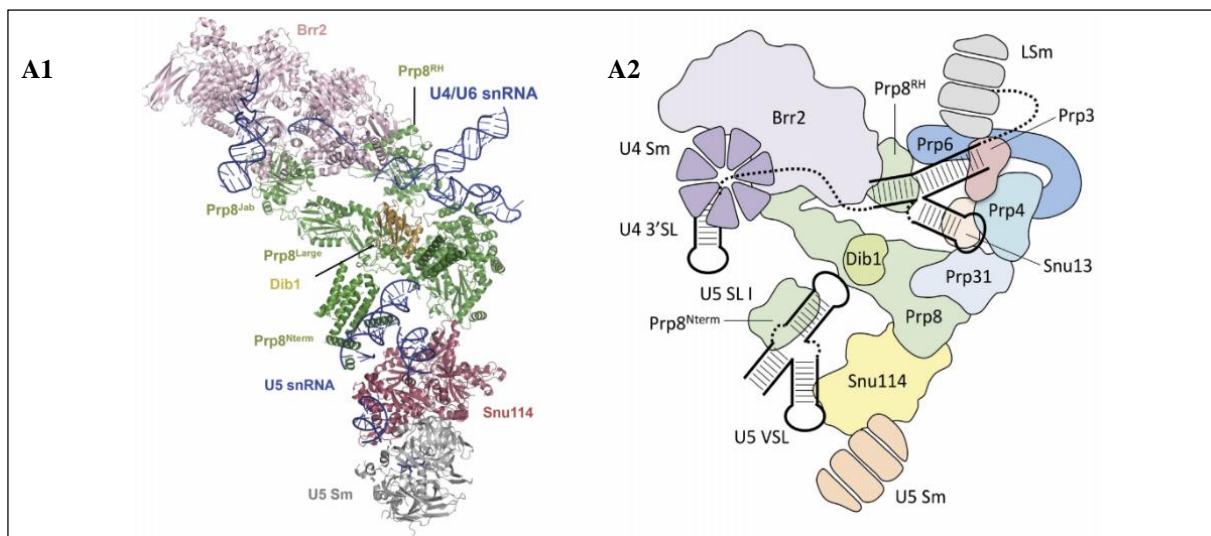
Dib1 faz contato com a alça I do snRNA U5 e esta interação é estabilizada pelas hélices do N-terminal de Prp8. Além disso, o sítio ativo do spliceossomo está localizado na Prp8, onde também se encontram um sítio de ligação a 5' SS do snRNA U6, as sequências do snRNA U6 que também pareiam com snRNA U2 e a alça I de snRNA U5 que pareia com o primeiro éxon e o segundo éxon. Esta superfície está parcialmente oculta pela Dib1, o que sugere seu papel potencial na regulação da incorporação dos componentes de RNA no sítio ativo do spliceossomo durante sua associação e ativação.<sup>56, 97</sup>

Em comparação com humanos e leveduras, *T. brucei* conserva snRNAs U4 e U6, que não apresentam diferença considerável em termos de estrutura secundária com relação aos

correspondentes naqueles organismos<sup>20, 97</sup> (Figura 1.12-C1 e C2). Entretanto, a proteína SSm4 substitui SmD3 no anel Sm de snRNP U4<sup>98</sup>, um dos focos deste trabalho. A razão para esta variação ainda não é compreendida. Já o snRNA U6 também se liga a um anel de proteínas LSm, homólogas às observadas em humanos e leveduras.<sup>20, 99</sup> Por outro lado, o snRNA U5 de *T. brucei* é o menor encontrado na natureza e não possui um capacete na extremidade 5'. Além disso, a estrutura na forma de grampo-de-cabelo na extremidade 3' é ausente e o anel Sm canônico se liga ao sítio Sm no snRNA U5.<sup>20, 100</sup>

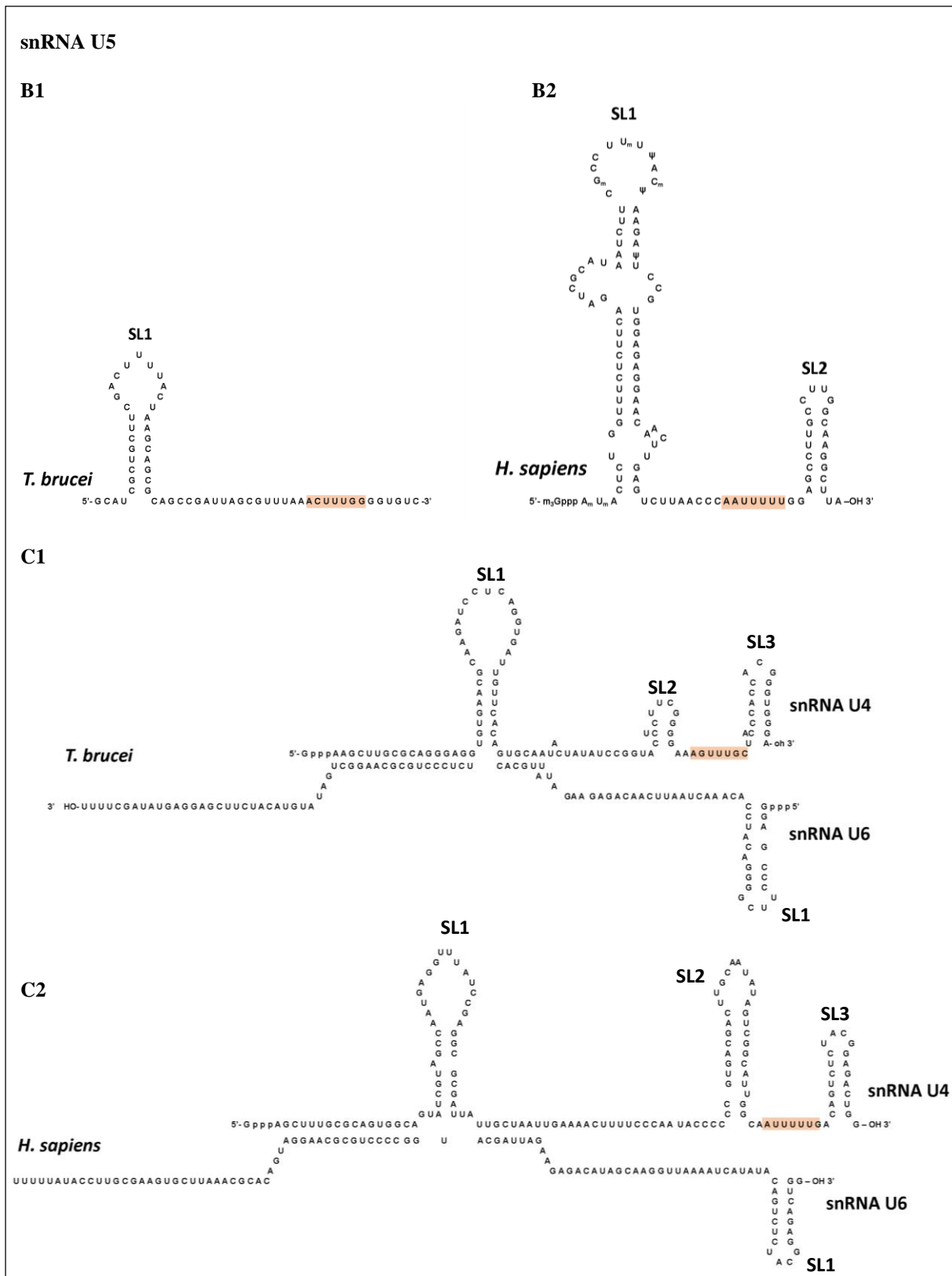
Snu13, membro da snRNP U4, e U5-200K, membro da snRNP U5, foram identificadas diretamente por mineração de dados a partir do sequenciamento do genoma de *T. brucei*<sup>20</sup>. As proteínas U5-220K<sup>101</sup>, U5-200K<sup>102</sup>, U5-116K, U5-40K<sup>60</sup> e U5-Cwc21<sup>60, 102</sup> foram identificadas a partir de purificação em *tandem* seguida de espectrometria de massas das proteínas SmD1 e SmB, em experimentos separados. A purificação de U5-Cwc21 levou à identificação, por espectrometria de massas, das mesmas proteínas. Esse resultado leva à sugestão de que U5-Cwc21 é membro estável da U5 snRNP em *T. brucei*, o que não é observado em humanos ou levedura.<sup>102</sup>

Adicionalmente, nota-se que U5-100K não foi identificada em nenhum dos experimentos. U5-102K e U5-15K, por sua vez, parecem formar um subcomplexo estável, como observado pela copurificação do heterodímero a partir da expressão em *T. brucei*.<sup>20, 60, 102</sup> Além disso, a autoclivagem da proteína U5-15K recombinante foi proposta recentemente, apresentando uma cinética lenta *in vitro*.<sup>103</sup> Entretanto, não se conhece a função desse mecanismo na função da proteína.<sup>103</sup> Além disso, observou-se que Prp31 de *T. brucei* é necessária para a formação de tri-snRNP U4/U6.U5.<sup>104</sup>



continua

continuação



**Figura 1.12 – Tri-snRNP U4/U6.U5.** A1 e A2- Estrutura de criomicroscopia eletrônica da tri-snRNP U4/U6.U5 de *S. cerevisiae*.<sup>56</sup> B- Estruturas secundárias de snRNA U5 de *T. brucei* (B1) e *H. sapiens* (B2) e C- de di-snRNAs U4/U6 de *T. brucei* (B1) e *H. sapiens* (B2). Os sítios Sm estão evidenciados em vermelho.

Fonte: NGUYEN et al<sup>56</sup>; MOTTRAM et al<sup>105</sup>; YU et al<sup>106</sup>.

### 1.3.2.4 Biogênese de snRNPs

Complexos macromoleculares são essenciais para as células e o ambiente celular onde são formados está repleto de moléculas, que acabam perturbando seu processo de formação dirigido por difusão, além de aumentar a probabilidade de interações não-produtivas. Deste modo, fatores transativadores de formação desses complexos são necessários para sequestrar subunidades de modo a impedi-las de formar interações não-específicas com outras moléculas do ambiente celular, além de catalisar a formação do complexo na disposição correta.<sup>107</sup>

Em humanos, os snRNAs são transcritos no núcleo celular pela RNA polimerase II (exceto o snRNA U6) e transportados ao citoplasma a partir da ligação aos complexos CBC (do inglês, *cap binding complex*), PHAX (do inglês, *phosphorylated adaptor for RNA export*) e Xpo1/CRM1 (exportina 1/região de manutenção do cromossomo 1). No citoplasma, PHAX é desfosforilado e continua associado à CBC/pré-snRNA, até a hipermetilação do capacete 5'.<sup>40, 107</sup>

A biogênese de uma snRNP humana inicia-se pela formação de um anel heteroheptamérico de proteínas Sm em torno do sítio Sm do snRNA (U1, U2, U4 e U5) no citoplasma. Apesar do reconhecimento do sítio Sm pelas 7 proteínas Sm ser suficiente para formação do núcleo estável da snRNP *in vitro*<sup>108</sup>, pelo menos 12 fatores formam os complexos PRMT5 e SMN (do inglês, *survivor of motor neuron*) que catalisam a biogênese de snRNPs *in vivo* em humanos e leveduras.<sup>107</sup>

Na fase inicial da biogênese do complexo de proteínas Sm/snRNA, os heterodímeros SmD1/SmD2 e SmD3/SmB interagem com a proteína pICln no citoplasma, e são recrutados pelo complexo PRMT5/WD45 para formar o supracomplexo 20S. Em seguida, SmD1/SmD2 e SmD3/SmB recebem dimetilações simétricas nas cadeias laterais de argininas dos domínios C-terminais de SmD1, SmB e SmD3. A proteína pICln atua como uma chaperona e previne as proteínas Sm de se associarem com moléculas de snRNA. O heterotrímero SmF/SmE/SmG é então adicionado e resulta na formação de um complexo intermediário na forma de anel com coeficiente de sedimentação igual a 6S, formado por SmD1/SmD2, SmF/SmE/SmG e pICln no complexo PRMT5/WD45.<sup>107</sup>

Na fase final, o complexo 6S pré-metilado e pICln-SmD3/SmB interagem com o complexo SMN, e pICln é eliminada. Então, Gemina2, presente no complexo SMN, coordena a ligação das 5 proteínas Sm (SmD1/SmD2/SmF/SmE/SmG)<sup>109</sup>, enquanto SmD3/SmB provavelmente se liga à proteína SMN, Gemina2 e/ou Gemina8.<sup>110</sup>

Em seguida, o snRNA se associa ao complexo e as proteínas Sm formam um anel em torno do sítio Sm. Finalmente, o complexo é transportado para o núcleo onde a snRNP madura se dissocia do complexo SMN. A associação é governada por energia térmica e o complexo SMN é identificado com uma máquina Browniana. A reimportação para o núcleo ocorre através da associação com importina- $\beta$  e snuportina-1.<sup>40, 107</sup> Após a reentrada no núcleo, as proteínas transportadoras se dissociam e os complexos de proteínas Sm/snRNA se localizam em corpos de Cajal. Os snRNAs U1, U2, U4 e U5 sofrem modificações (pseudouridilação e 2'-O-metilação) e as proteínas específicas de snRNP são então incorporadas para a formação de uma snRNP madura.<sup>40, 107</sup>

A partícula snRNP U6, por sua vez, contém o snRNA U6 que é transcrito pela RNA polimerase III, e não é exportado ao citoplasma. No núcleo, o snRNA U6 se liga ao complexo de 7 proteínas LSm (LSm2-8). Após a associação, este complexo segue para o nucléolo, onde ocorrem as mesmas modificações pós-transcricionais no snRNA e de lá segue para os corpos de Cajal. Assim, U6 interage com U4 por meio de pareamento de bases dos snRNAs.<sup>40, 107</sup>

Após a formação do núcleo composto por 7 proteínas Sm interagindo com o snRNA, outras proteínas específicas devem se associar para formar a partícula madura. Várias proteínas específicas de snRNP U1 e snRNP U2 (como U2A' e U2B'') foram caracterizadas como sendo transportadas para o núcleo de maneira independente do snRNA.<sup>111-115</sup> Entretanto, o conhecimento acerca da biogênese de snRNPs ainda está limitado.

Recentemente, um modelo para a biogênese de snRNP U4 foi proposto, no qual a partícula começa a ser constituída no citoplasma, com a formação do anel de proteínas Sm em torno do sítio Sm no snRNA U4. Este complexo se associa na presença de SMN e é transportado ao corpo de Cajal no núcleo, onde se liga às proteínas auxiliares NUFIP e ZNHIT3. A ligação de 15,5K a este complexo permite que Prp31 se associe de maneira estável à partícula para a subsequente formação de di- e tri-snRNPs.<sup>116</sup>

Por outro lado, durante a biogênese de snRNP U5 de levedura, o fator de *splicing* Aar2 interage, no citoplasma, com a região C-terminal de Prp8, nos domínios RT/Endonuclease. Prp8, por sua vez, forma um subcomplexo com Snu114, pelo menos. Este complexo proteico de proteínas específicas de snRNP U5 é transportado ao núcleo da célula, onde encontra o complexo proteínas Sm/snRNA U5. A fosforilação de Aar2 leva ao seu desligamento do complexo, dando lugar à proteína Brr2, helicase essencial para a ativação do spliceossomo. Assim, outras proteínas específicas se associam ao complexo, levando à formação da partícula U5 madura.<sup>117-118</sup>

Até o momento, nenhuma metilação foi identificada em proteínas Sm de *T. brucei* e apenas as proteínas SMN e Gemina2 foram caracterizadas.<sup>119-120</sup> A formação do anel Sm de snRNA U2 não é afetada quando a expressão de SMN é regulada negativamente, o que é consistente com o fato da reconstrução *in vitro* do núcleo Sm de snRNP U2 não requerer a presença de SMN. Os snRNAs U1, U4 e U5, por sua vez, têm sua concentração diminuída, o que sugere que SMN também seja importante para a estabilidade dessas moléculas. Adicionalmente, tanto o anel Sm canônico, que interage com snRNAs U1 e U5 e RNA SL, como o variante de snRNP U4 que apresenta SSm4 no lugar de SmD3, conservam a proteína SmB no anel, a qual interage diretamente com SMN.<sup>20, 82, 98, 120</sup>

A regulação negativa da expressão de SMN leva à perda de formação de di-snRNP U4/U6, o que mostra que interações para manutenção de di-snRNP U4/U6 requerem a formação do anel de proteínas Sm em torno do snRNA U4, primeiramente.<sup>120</sup> Já a transcrição dos snRNAs (exceto snRNA U6) e de RNA SL é feita pela RNA polimerase II. Foi observado que a transcrição de RNA SL ocorre em lugar definido no núcleo e a proteína SMN de *T. brucei* colocaliza nesta região, sugerindo que a formação de RNP SL ocorre exclusivamente no núcleo.<sup>44, 120</sup>

A deleção da proteína SMN leva ao acúmulo de RNA SL associado às proteínas Sm de forma não estável, o que está de acordo com o fato de que a reconstrução *in vitro* de proteínas Sm no RNA SL depende de SMN. A regulação negativa da expressão de SMN leva à acumulação de RNA SL com o cap4 parcialmente processado, similar ao resultado da regulação negativa da expressão da proteína SmD1, de modo que se supõe que o capeamento da extremidade 5' é dependente da associação correta do anel Sm no snRNA correspondente.<sup>120</sup>

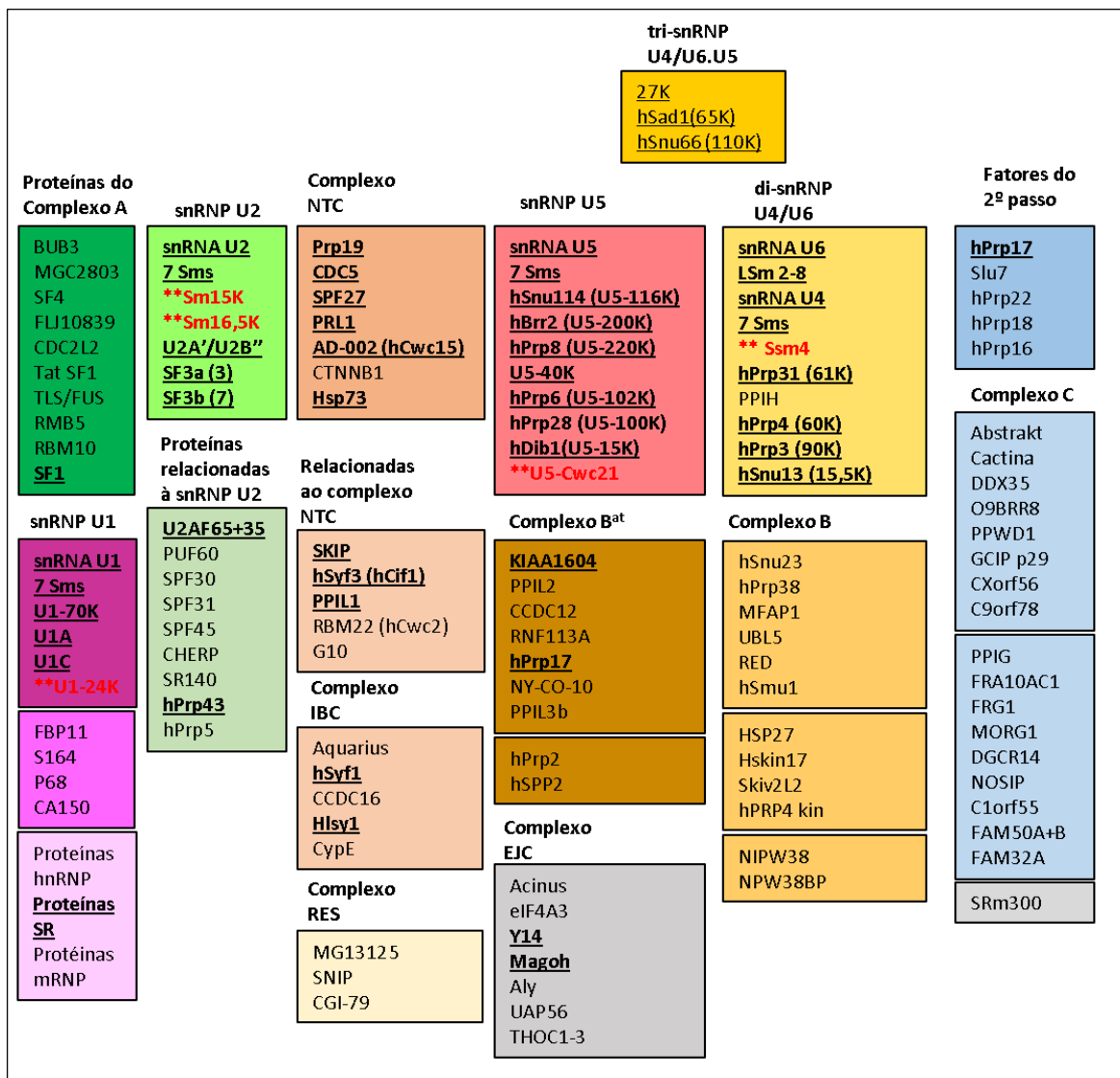
### 1.3.3 Dinâmica de associação das subunidades do spliceossomo para a catálise do *splicing*

Para que realize sua função de ribozima, o spliceossomo sofre uma série de mudanças na rede de interações RNA-RNA, proteína-RNA e proteína-proteína. Ocorre também uma drástica mudança na sua composição proteica.<sup>49, 121</sup> Além de snRNPs, o spliceossomo também recruta complexos proteicos adicionais para o remodelamento de sua estrutura. Entre eles, o NTC (do inglês, *nineteen complex*) estabiliza as snRNPs U5 e U6 após a dissociação de snRNP U4 na ativação do spliceossomo.<sup>122-123</sup> Já o IBC (do inglês, *intron-binding complex*) é necessário na etapa de ativação do complexo B. Adicionalmente, o EJC (do inglês, *exon-junction complex*) relaciona o processamento por *splicing* à estabilidade do mRNA, assim como

sua localização celular e o processo de tradução. Finalmente, o complexo proteico conhecido como NTR (do inglês *nineteen-related*) é necessário para a dissociação do spliceossomo após a liberação do mRNA maduro. Apenas o complexo NTC está bem caracterizado em *T. brucei*.<sup>122-124</sup>

A Figura 1.13A apresenta a composição das snRNPs com relação às proteínas spliceossomais de humanos e *T. brucei*. A dinâmica de associação do spliceossomo humano e de levedura em termos das proteínas que o compõem e da rede de interações de moléculas de RNA também é mostrada.

A

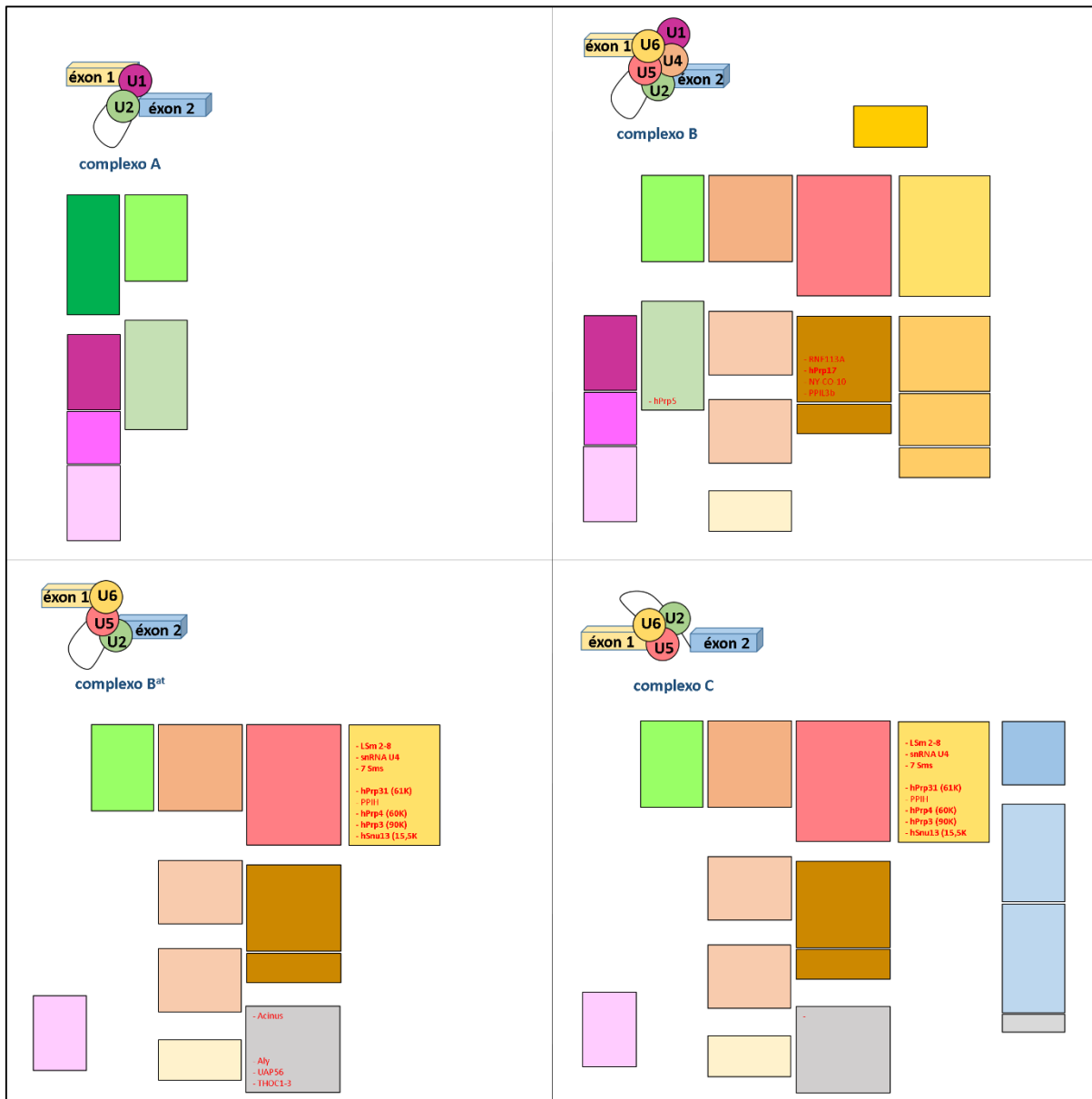


continua

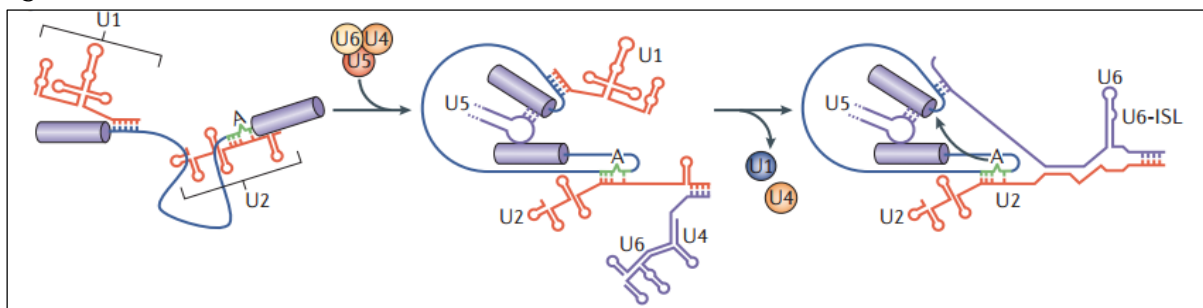


continuação

B



C



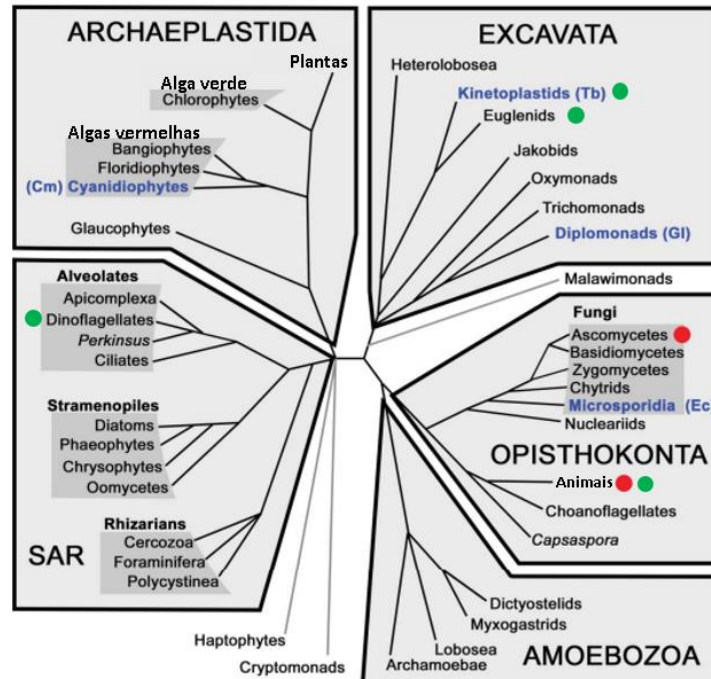
**Figura 1.13– Dinâmica de associação do spliceossomo.** A- Composição das snRNPs e das proteínas spliceossomais de humanos. Estão grifadas as proteínas que apresentam homólogas em *T. brucei* e destacadas em vermelho as proteínas únicas deste organismo. B- Composição do spliceossomo a cada etapa da catálise do *splicing*. O código de cores utilizado em A foi mantido. C- Rede de interações de moléculas de snRNAs a cada etapa.  
 Fonte: Adaptada de WAHL<sup>49</sup>.

Enzimas de modificação pós-traducional e peptidil-prolil cis/trans-isomerases (como Cwc27<sup>125</sup>) estão associadas a essa maquinaria, apesar de pouco estudadas até o momento.<sup>49, 121</sup> Por exemplo, ubiquitina e proteínas similares à ubiquitina copurificam com muitas proteínas spliceossomais. Além disso, fatores de *splicing* ubiquitinados também têm sido identificados, como a proteína membro do complexo NTC conhecida como Prp19, que possui um domínio *U-box* e modifica a proteína Prp3, componente de snRNP U4<sup>126</sup>. Já o domínio Jab1/MPN de Prp8, que é uma proteína componente essencial do centro catalítico do spliceossomo, apresenta atividade de ligação à ubiquitina.<sup>127</sup> Adicionalmente, em levedura, a deleção do gene *hub1* que codifica para uma proteína similar à ubiquitina leva a defeito de *splicing*.<sup>128</sup>

O elevado número de proteínas associadas à remodelagem do spliceossomo sugere a possibilidade de mais etapas intermediárias, além daquelas já conhecidas, o que levaria a um número maior de pontos de regulação. Além disso, há proteínas específicas que devem atuar apenas como suporte transiente para facilitar a organização de outros componentes do complexo.

Curiosamente, a caracterização bioquímica de proteínas spliceossomais de diferentes eucariotos como *Trypanosoma brucei* e *Drosophila melanogaster* também revelou proteínas spliceossomais específicas, sem homólogas evidentes nos organismos mais bem caracterizados e sem função conhecida até o momento. Adicionalmente, vários organismos foram caracterizados com número pequeno de íntrons no genoma e com um spliceossomo reduzido, como é o caso do próprio *T. brucei* (Excavata), além de *Giardia lamblia* (Excavata), *Cyanidioschyzon merolae* (Archaeplastida) e *Encephalitozoon cuniculi* (Opisthokonta) (Figura 1.14). *C. merolae* possui o menor spliceossomo funcional identificado até o momento, contendo apenas 43 proteínas, com nenhuma evidência para a existência de snRNP U1.<sup>129</sup>

Além disso, o mecanismo de SL *trans-splicing* catalisado pelo (*trans*-) spliceossomo foi identificado em organismos de diferentes filos. Além de tripanossomatídeos, membros de Euglenozoa (Excavata), organismos como rotíferos, cordados, nematodos, platelmintos, cnidários (Uniconta) e dinoflagelados (Chromalveolata) também apresentam RNA SL. Existe a hipótese de que um ancestral comum a organismos de pelo menos esses grupos de eucariotos apresentava o mecanismo de SL *trans-splicing*. Entretanto, a dificuldade em se traçar o caminho evolutivo deste tipo processamento do pré-mRNA está relacionada à quantidade limitada de informação disponível.<sup>130</sup>



**Figura 1.14 – Representação esquemática da relação filogenética entre eucariotos**, evidenciando organismos onde o mecanismo de *splicing* está bem caracterizado (ponto vermelho) e os clados contendo organismos com spliceossomo reduzido (azul). Clados que contêm organismos onde o mecanismo de SL *trans-splicing* foi caracterizado são indicados em verde. Cm - *C. merolae*, Ec - *E. cuniculi*, Gl - *G. lamblia*, Tb - *T. brucei*.

Fonte: Adaptada de HUDSON et al<sup>131</sup>.

## 1.4 Regulação do spliceossomo

Durante a associação do spliceossomo ao pré-mRNA, todos os sítios reativos são identificados por diferentes fatores várias vezes, o que aumenta a fidelidade do reconhecimento. Adicionalmente, RNA helicases são necessárias para a remodelagem da sua estrutura durante o ciclo de catálise do *splicing* e, até o momento, foram identificadas oito RNA helicases da família de helicases do tipo DExD/H-box envolvidas na ativação das mudanças estruturais do spliceossomo em humanos, leveduras e tripanossomatídeos. Estas enzimas são dependentes de hidrólise de ATP para modular a rede de interações entre moléculas de snRNA das snRNPs por meio da catálise da separação de ácidos nucleicos de fita dupla.<sup>132-133</sup>

(h)Prp5 e UAP56/Sub2 (humano/levedura) são recrutadas na fase de associação do spliceossomo no pré-mRNA para a formação do complexo A. (h)Prp5 está envolvida na catálise da mudança conformacional da partícula U2 e supõe-se que controle o reconhecimento de BP. UAP56/Sub2, por sua vez, parece estar envolvida na liberação das proteínas de ligação ao BP (SF1/Msl5 e U2AF/Mud2 em humano/levedura) para que snRNP U2 possa se ligar nessa região.<sup>132-133</sup>

Ambas as enzimas U5-100K/Prp28 (humano/levedura) e U5-200K/Brr2 (humano/levedura) fazem parte da tri-snRNP U4/U6.U5 e são necessárias para a ativação do spliceossomo, sendo responsáveis pela saída de snRNPs U1 e U4, respectivamente. Acredita-se que U5-100K desestabilize a interação de snRNA U1 com o 5'SS. Além disso, o recrutamento de tri-snRNP U4/U6.U5 ao spliceossomo é regulado pela sua fosforilação.<sup>132-133</sup> Adicionalmente, U5-100K está relacionada à revisão (do inglês *proofreading*) do reconhecimento correto do sítio 5'SS durante a associação do spliceossomo.<sup>132-133</sup> Já U5-200K atua no despareamento do di-snRNA U4/U6, e sua função será analisada em detalhes no capítulo 3.

A enzima Prp2 é recrutada para a catálise da dissociação de SF3a/b da partícula U2, levando à exposição do BP para a primeira etapa do *splicing*. Sua atividade requer alguns fatores proteicos como Spp2, Cwc22 e eIF4G-like, como caracterizado em levedura. Sítios para a ligação de Yju2 e Cwc25 são criados e a conformação do sítio catalítico do spliceossomo é estabilizada para a primeira reação de transesterificação. Prp16, por sua vez, é essencial na segunda etapa de catálise da reação de *splicing*, apesar de sua função ainda não estar determinada.<sup>132-133</sup>

Finalmente, Prp22 e Prp43 catalisam a dissociação do complexo do spliceossomo. A função primária de Prp22 é a liberação do mRNA maduro após a segunda reação de *splicing*. Esta enzima se liga diretamente na sequência de íntron à jusante do BP e transloca o mRNA após a ligação dos éxons. Já a enzima Prp43, que requer Ntr1 e Ntr2 como cofatores (presentes no complexo NTR), está envolvida na desassociação de complexos intermediários do spliceossomo que não foram associados corretamente.<sup>132-134</sup>

A cinética das RNA helicases e da proteína G Snu114 (U5-116K) possivelmente está acoplada com vias de descarte de intermediários mal-formados do spliceossomo. Acredita-se que Prp43 e seus cofatores do complexo NTR, por exemplo, estejam relacionados a mecanismos de revisão/fidelidade de associação do spliceossomo nos sítios de *splicing*.<sup>134</sup>

Outro nível de regulação do *splicing* está relacionado ao reconhecimento dos sítios de *splicing*. Vários organismos eucariotos contêm mais do que um íntron nos seus genes que codificam para proteínas. Deste modo, os pré-mRNAs maduros podem se originar de *splicing* alternativo, dada a flexibilidade na escolha dos sítios de *splicing* durante a catálise. Esse mecanismo, combinado com a presença de sítios alternativos de iniciação de transcrição e de clivagem ou poliadenilação da extremidade 3', possibilita a produção de várias proteínas únicas a partir do mesmo transcrito primário. O acoplamento cinético e físico do spliceossomo com

mecanismos de transdução de sinal, organização da cromatina e transcrição são fundamentais para a regulação do *splicing* alternativo.<sup>135</sup>

A escolha entre possíveis sítios de *splicing* alternativos é, em partes, determinada por fatores transativadores, que podem ser regulados de acordo com o estágio de desenvolvimento do organismo, o tecido, a fase de diferenciação ou por meio de estímulo externo.<sup>135</sup> Proteínas ricas em serina e arginina, conhecidas como proteínas SR, e ribonucleoproteínas nucleares heterogêneas (hnRNPs), respectivamente, promovem e interferem no recrutamento da maquinaria nos sítios de *splicing*. Ambas contêm regiões de interação com RNA por meio das quais reconhecem sequências potencializadoras ou silenciadoras de *splicing* (do inglês, *exonic/intronic splicing enhancer*, ESSE/ISE, ou *exonic/intronic splicing silencer*, ESS/ISS, respectivamente) que se localizam no pré-mRNA.<sup>136</sup> No caso de proteínas SR, essa interação é modulada por meio de sua fosforilação.<sup>137</sup>

No caso de tripanossomatídeos, onde a transcrição é policistrônica, a regulação genética é baseada em mecanismos pós-transcricionais. Recentemente, a análise de transcritos maduros nas diferentes fases do ciclo de vida de *T. brucei* levou à identificação de mais de 2,5 mil eventos de *splicing* alternativo, resultados da adição do éxon SL em diferentes posições no pré-mRNA.<sup>43, 138</sup> Entre as consequências deste mecanismo, estão incluídos: (1) inibição da tradução devido à perda do códon iniciador AUG, (2) inclusão ou exclusão de sequências de sinalização, podendo levar à diferente localização subcelular da proteína expressa, (3) inclusão ou exclusão de elementos de regulação, podendo alterar a estabilidade do mRNA produzido, e (4) uso de uma fase aberta de leitura alternativa.<sup>43,138</sup> Diferentes padrões de *splicing* são observados nas diferentes formas durante seu ciclo de vida e o mecanismo é importante para o desenvolvimento do parasita.<sup>138-139</sup> Notadamente, o mecanismo de evasão imune de tripanossomatídeos é dependente da expressão de VSGs, que correspondem de 7 a 11% dos transcritos que sofrem SL *trans-splicing* nas formas sanguíneas de *T. brucei*.<sup>138-139</sup>

Outro nível de regulação é encontrado em muitas plantas, fungos e animais, que contêm um segundo tipo de spliceossomo, conhecido como spliceossomo menor ou spliceossomo dependente de U12.<sup>140</sup> Em humanos, o spliceossomo menor é responsável pela excisão de menos que 0,5% dos íntrons. Estes são caracterizados pela presença de sequências AT-AC na extremidade 3' e 5'SS altamente conservado.<sup>140</sup> Apesar da snRNP U5 ser comum a ambos os spliceossomos, snRNPs U1, U2, U4 e U6 são substituídas por snRNPs U11, U12, U4atac e U6atac, respectivamente, cujos snRNAs são diferentes daqueles observados para as snRNPs canônicas. No caso de U11 e U12, fatores proteicos diferentes são recrutados (65K, 59K, 48K, 35K, 31K, 25K, 20K, Urp, além do complexo SF3b), enquanto que para U4atac e U6atac, as

mesmas proteínas foram observadas. O princípio de funcionamento do spliceossomo menor é basicamente o mesmo já caracterizado para o spliceossomo, exceto pelo fato de que snRNPs U11 e U12 formam uma di-snRNP estável, que se liga cooperativamente a 5'SS e BP.<sup>140</sup> Este tipo de controle não foi encontrado em tripanossomatídeos até o momento.

#### 1.4.1 Relação entre a regulação do *spliceossomo* e a saúde humana

##### 1.4.1.1 Doenças humanas diretamente relacionadas ao mal-funcionamento do spliceossomo

O spliceossomo é altamente susceptível a pequenas alterações na sua composição e na função das proteínas associadas. De fato, acredita-se que mais que 60% das mutações que levam a doenças humanas estejam relacionadas ao *splicing* ou ao *splicing* alternativo.<sup>141</sup>

Mutações nos genes que codificam para U5-220K, U5-200K, 90K e 61K têm sido relacionadas a uma doença conhecida como retinite pigmentosa em humanos, que se manifesta devido à degeneração contínua de células receptoras do tipo bastonete na retina. Pacientes que apresentam a doença têm sintomas como cegueira noturna e perda do campo de visão periférico. Atualmente, não há tratamento ou prevenção efetivos. A retinite pigmentosa tem origem genotípica bastante variada, sendo que mutações nesses genes que codificam para proteínas do spliceossomo levam a uma forma autossômica dominante, que ocorre em 15 a 20% dos casos da doença. Entretanto, ainda não se conhece a relação entre tais mutações e o mecanismo da doença.<sup>142</sup>

Recentemente, foram identificadas mutações no gene que codifica para a proteína U5-15K, membro da snRNP U5 de humanos, em indivíduos com a síndrome de Burn-McKeown (BMKS), uma condição rara caracterizada por atresia coanal bilateral, perda de audição, fissuras lábio palatinas e outros dimorfismos craniofaciais. Em levedura, a depleção do gene que codifica para esta proteína reduz o número de complexos de tri-snRNPs formados.<sup>143</sup>

Por outro lado, a deleção ou mutação do gene SMN1 que codifica para a proteína de sobrevivência do motoneurônio é a causa da atrofia muscular espinhal, uma doença neurodegenerativa com herança genética autossômica recessiva. A falta da proteína SMN leva à degeneração de motoneurônios- $\alpha$  localizados no corno anterior da medula espinhal, o que resulta em fraqueza e paralisia muscular proximal progressiva e simétrica. SMN é necessária para a associação dos núcleos das snRNPs no citoplasma antes da sua reimportação nuclear (seção 3.2.4). A diminuição da produção de snRNPs maduras tem sido ligada à doença.<sup>144-145</sup>

Acredita-se que a coleção de fatores de *splicing/splicing* alternativo que se associam ao spliceossomo no pré-mRNA possa ser única para um dado transcrito em um tipo de célula humana. Há evidências de que há o aumento do número de eventos de *splicing* anômalo em células cancerosas, onde fatores de *splicing* têm sua expressão diminuída comparando-se com células normais. Como exemplo, alterações em fatores de *splicing* que atuam na associação do spliceossomo nos sítios de reconhecimento no pré-mRNA, como SF3b155 associado à partícula U2 ou o fator de regulação SRSF1(ASF/SF2), têm sido apontadas como causas de doenças hematológicas.<sup>146-147</sup> Adicionalmente, mutações no snRNA U4atac levam à síndrome de Taybi-Linder (MOPD I).<sup>148</sup>

Apesar dos estudos acumulados durante as últimas décadas,<sup>40</sup> a base mecânica para as doenças humanas relacionadas aos fatores de *splicing* ainda não é conhecida. Não está claro como defeitos na maquinaria do spliceossomo, essencial para o organismo eucarioto, pode levar a efeitos locais que originam doenças tecido-específicas. Uma possibilidade é que a atividade de *splicing* reduzida seja particularmente danosa para tecidos altamente regenerativos que requerem proliferação celular abundante e uma alta taxa de transcrição e processamento de RNA. Atualmente, novos estudos começam a se voltar para a análise da dinâmica conformacional e composicional do spliceossomo e sua relação com a regulação da função de ribozima e do reconhecimento de sítios de *splicing/splicing* alternativo e como o seu malfuncionamento está relacionado com doenças humanas.<sup>149</sup>

Recentemente, metabólitos bacterianos que interagem com a subunidade SF3b de snRNP U2 foram identificados com atividade anti-tumoral, o que pode indicar que a patogenicidade de algumas bactérias possa estar envolvida com a inibição do spliceossomo.<sup>150-151</sup> Adicionalmente, um estudo recente mostrou que a superexpressão do oncogene MYC que codifica para um fator de transcrição confere estresse no *splicing* e que componentes do spliceossomo, como a proteína Bud31 e a subunidade SF3b, são possíveis alvos terapêuticos para câncer resultante da presença desse oncogene.<sup>152</sup>

O Apêndice A1 apresenta uma lista de pequenas moléculas relacionadas com algum efeito na função do spliceossomo. Inibidores de fosforilases de proteínas SR são os mais comuns, o que aponta para a importância de modificações pós-traducionais na regulação da maquinaria do spliceossomo. O desenvolvimento de oligonucleotídeos antisense (do inglês *AntiSense Oligonucleotides*, ASO) de até 25 nucleotídeos de extensão, complementares a um transcrito de RNA alvo, também são potenciais para o tratamento de doenças relacionadas ao *splicing*.<sup>153-154</sup>

### 1.4.1.2 O SL *trans-splicing* de parasitas como alvo

Apesar do controle da transcrição do RNA em parasitas ainda não ser bem compreendido, o genoma de muitos parasitas de humanos, animais e plantas já foi sequenciado e o número crescente de estudos acerca da maquinaria de SL *trans-splicing* tem levado a uma melhor descrição da regulação gênica nestes organismos eucariotos.<sup>139</sup> Além de tripanossomatídeos, esse mecanismo pós-transcricional também é encontrado em parasitas nematodos<sup>155</sup> e trematodos<sup>156</sup>, mas não foi observado nos seus hospedeiros vertebrados ou artrópodes<sup>157</sup> e, portanto, pode ser considerado como um processo parasita-específico.

No caso de tripanossomatídeos, os mecanismos ainda não bem entendidos de evasão imune podem indicar novos alvos para terapias contra as doenças causadas por eles. O controle da expressão de isoformas de antígenos pode resultar na diminuição da evasão imune ou mesmo da viabilidade dos parasitas. A maquinaria de SL *trans-splicing*, apresenta-se, portanto, como um alvo promissor para descoberta de novos fármacos potentes contra essas doenças.<sup>20, 139</sup>

A análise comparativa dos componentes do spliceossomo de *T. brucei* com aqueles do spliceossomo de seus hospedeiros, mais bem caracterizados até o momento, aponta para algumas diferenças na rede de interação proteína-proteína e proteína-RNA.<sup>20</sup> Entre elas, a variação dos núcleos Sm das snRNPs U2 e U4 se destaca como alvo potencial para a busca de um novo fármaco que module a biogênese de snRNPs em *T. brucei*. Por outro lado, a proteína U1-24K parece ser parasito-específica, porém sua função ainda não foi determinada.<sup>65</sup> Adicionalmente, U5-Cwc21 é essencial para a viabilidade de *T. brucei* e copurifica as proteínas específicas de U5 snRNP, apesar de não fazer parte desta partícula em humanos e leveduras.<sup>102</sup>

Até o momento, nenhum estudo se concentrou na busca por moléculas inibidoras que tenham o spliceossomo de parasitas como alvo. Por outro lado, vários esforços têm sido feitos para o estabelecimento de um método de diagnóstico eficiente para essas doenças, baseando-se no mecanismo específico de *splicing*.<sup>158</sup>

O diagnóstico acurado das doenças causadas por tripanossomatídeos é desafiador, já que a concentração de parasitas no sangue pode variar de milhares até poucas centenas de parasitas por mililitro de sangue, o que se encontra no limite de detecção de muitos métodos já descritos.<sup>159</sup> Recentemente, foi reportado pela primeira vez um método de diagnóstico de *T. brucei* por meio da detecção de SL RNA a partir de isolamento de RNA de amostra sanguínea, seguido de transcrição reversa e PCR em tempo real. A detecção de SL RNA se mostrou mais específica do que a detecção de DNA de parasita isolado de sangue contaminado. Testes de



diagnóstico baseados em detecção de ácidos nucleicos são geralmente mais sensíveis e apresentam alta especificidade.<sup>158</sup>

Apesar da prova de conceito de que o SL-RNA pode ser usado como biomarcador para a doença do sono, o procedimento pode ser transportado para detecção de doença de Chagas, leishmanioses e nagana. Nematodos também apresentam o mecanismo de SL *trans-splicing* e a detecção deste biomarcador também poderia ser aplicada como diagnóstico.<sup>158</sup>

Neste trabalho, serão apresentados estudos bioquímicos, biofísicos e estruturais da variação dos núcleos Sm das snRNPs no capítulo 2 e da enzima U5-200K, membro de snRNP U5 de *T. brucei*, no capítulo 3. Os resultados obtidos contribuem para uma melhor compreensão da evolução deste mecanismo entre organismos eucariotos. Além disso, estudos estruturais da variação dos núcleos Sm de *T. brucei* são as etapas iniciais na busca de novos fármacos para a doença do sono.



## CAPÍTULO 2

### **ANÁLISE BIOQUÍMICA E ESTRUTURAL DOS NÚCLEOS Sm DAS snRNPs U2, U4 e U5 E DAS PROTEÍNAS U2A' E U2B'' DE *T. brucei***

Neste capítulo, estuda-se a interação das proteínas Sm com snRNAs U2, U4 e U5 e um modelo de formação do núcleo de snRNP U2 é proposto, baseado na interação do núcleo Sm recombinante de *T. brucei* com o snRNA U2 seguida da ligação de U2A'/U2B''. Adicionalmente, apresenta-se a análise das sequências de aminoácidos das proteínas Sm e de U2A'/U2B'' de *T. brucei*, bem como a investigação de suas estruturas terciária e quaternária por meio de cristalografia, modelagem comparativa e espalhamento de raios-X a baixo ângulo. Os resultados aqui apresentados correspondem ao primeiro estudo estrutural de subcomplexos do trans-spliceossomo de um parasita humano.



## 2.1 Introdução

### 2.1.1 A família das proteínas Sm

As proteínas Sm foram descobertas na década de 1960 como pequenos antígenos nucleares alvos de anticorpos produzidos pela paciente Stephanie Smith, que apresentava um quadro de lúpus eritematoso sistêmico.<sup>160</sup> Esses antígenos ficaram conhecidos, em inglês, como *Smith antigens* (Sm Ag) e foram identificados em associação com complexos ribonucleoproteicos de extratos de células eucarióticas.<sup>161</sup> Nas décadas seguintes, diferentes grupos de pesquisa elucidaram as propriedades fisiológicas e bioquímicas de nove proteínas Sm (D1, D2, D3, E, F, G, N, B e sua variante de *splicing* B'), bem como sua estrutura tridimensional e especificidade por snRNAs (U1, U2, U4/U6 e U5), direcionando ao consenso de que as proteínas Sm eucarióticas atuam como uma estrutura base para a associação de uma snRNP.<sup>162</sup>

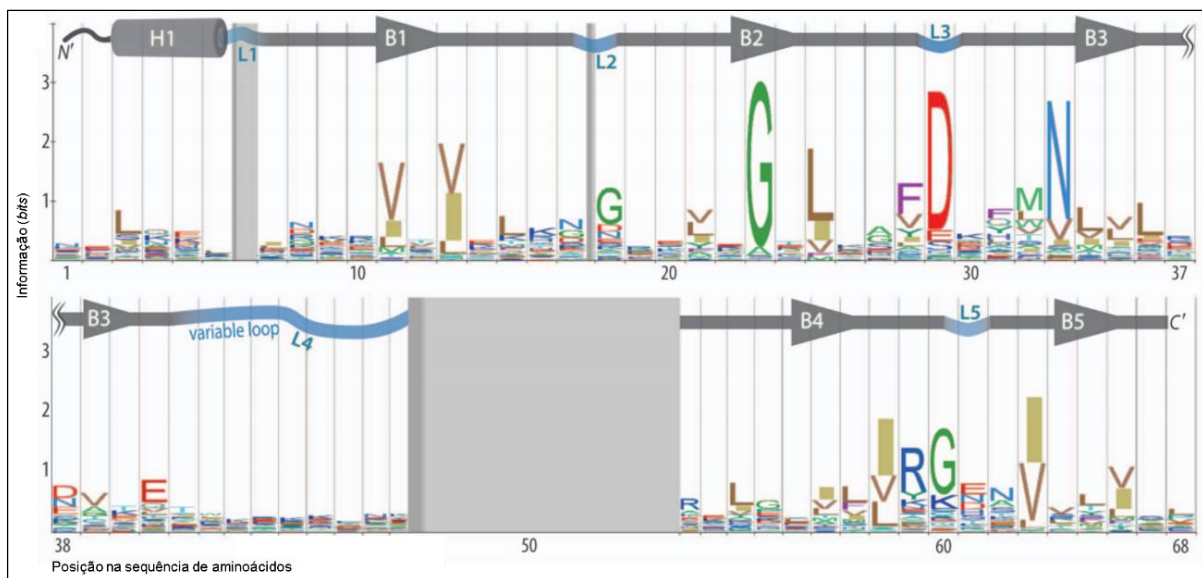
Em 1999, sete proteínas citosólicas homólogas às proteínas Sm, denotadas LSm1-7 (do inglês, *like-Sm*), foram identificadas em interação com o snRNA U6, e, mais tarde, também foram caracterizadas as proteínas nucleares LSm2-8.<sup>162-163</sup> Mais recentemente, as proteínas Gemina6 e Gemina7 se revelaram como um heterodímero com enovelamento Sm e função de chaperona no complexo SMN para assistir a formação do anel Sm.<sup>164-165</sup> Proteínas SmAPs homólogas às LSm eucarióticas também foram encontradas em algumas arqueias e permanecem até hoje sem função definida.<sup>166</sup>

Além da sua atuação na biogênese e função do spliceossomo, como descrito no capítulo 1, proteínas Sm/LSm eucarióticas estão relacionadas ao processamento de rRNA por snoRNPs (do inglês, *small nucleolar RNPs*)<sup>167</sup>, a *splicing* e maturação de tRNA baseados em RNase P<sup>168</sup>, ao processamento da extremidade 3' do mRNA de histonas pela snRNP U7<sup>169</sup>, ao decapeamento e decaimento de mRNA<sup>170</sup>, e à manutenção da extremidade do cromossomo pela telomerase<sup>171</sup>. Tal cenário aponta para uma função mais ampla das proteínas Sm/LSm na biogênese de ribonucleoproteínas envolvidas no metabolismo de RNA na célula eucariótica.

Considerando-se o domínio Bacteria, a proteína Sm conhecida como fator Hfq foi descrita no fim da década de 1960 como essencial para a replicação do bacteriófago Q $\beta$ . A caracterização bioquímica do fator Hfq revelou a formação de hexâmeros termoestáveis, em alta concentração intracelular. Estes hexâmeros se ligam preferencialmente a moléculas de RNA de fita simples ricas em nucleotídeos A/U, mas também são capazes de se ligar ao DNA no nucléolo. Em oposição à função passiva das proteínas Sm eucarióticas, o fator Hfq atua

como chaperona de RNA, facilitando o pareamento de bases entre RNAs de regulação (ncRNAs) e mRNAs alvos, o que resulta no controle da estabilidade de mRNAs e atividade de tradução.<sup>166, 172</sup>

As proteínas Sm e LSM constituem uma superfamília de proteínas ligadoras de RNA que contêm um motivo de sequência de aminoácidos composto de dois fragmentos (Sm1 e Sm2), separados por um espaçador variável (alça L4, Figura 2.1). Proteínas ortólogas são encontradas em espécies dos três domínios da vida (Bacteria, Archaea e Eukarya), o que implica em uma origem evolutiva antiga para esta superfamília, anterior à divergência entre organismos eucariotos e arqueias.<sup>166, 173</sup>



**Figura 2.1 – Conservação de aminoácidos em proteínas LSM.** Análise de conservação de aminoácidos a partir de alinhamento múltiplo de sequências de aminoácidos de proteínas LSM usando o modelo oculto de Markov (do inglês, *hidden Markov model*, HMM). O eixo vertical corresponde ao conteúdo relativo de informação medido em *bits* para cada posição na sequência de aminoácidos. Cilindros correspondem à estrutura secundária na forma de hélice- $\alpha$  (H) e triângulos à fita- $\beta$  (B). Alças (L) são destacadas em azul. Regiões de alta variabilidade de sequência de aminoácidos são evidenciadas em cinza.

Fonte: Adaptada de MURA et al<sup>166</sup>.

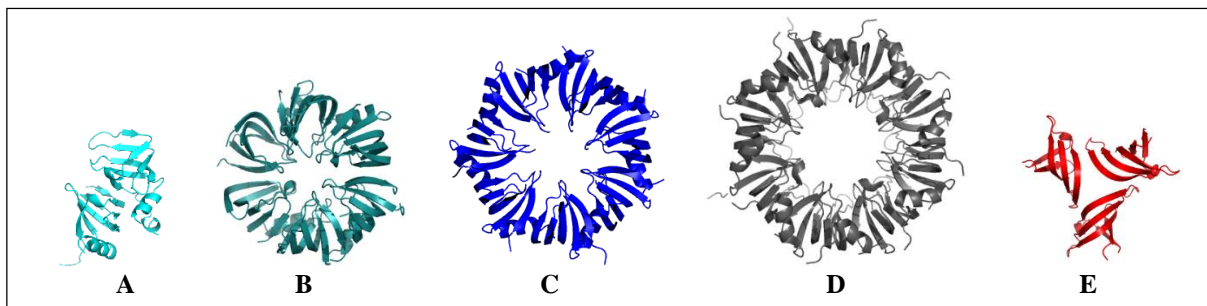
No motivo Sm1, resíduos de glicina na estrutura de fita- $\beta$ 2, aspartato na alça L3 e asparagina anterior à fita- $\beta$ 2 são mais frequentes nas suas respectivas posições, enquanto resíduos de aminoácidos hidrofóbicos são conservados na fita- $\beta$ 1. Já no caso do motivo Sm2, resíduos de aminoácidos hidrofóbicos flanqueiam resíduos de arginina e glicina altamente conservados em L5 (Figura 2.1). Todos estes resíduos foram reportados como sítios de interação específica com bases nitrogenadas de ácidos nucleicos em diferentes organismos.<sup>166</sup>

Em 1999, cristais de proteínas Sm humanas recombinantes foram obtidos pelo grupo do Dr. Kiyoshi Nagai, em Cambridge, Reino Unido, permitindo a determinação da estrutura

cristalográfica dos heterodímeros SmD1/SmD2 (código PDB 1B34) e SmD3/SmB (código PDB 1D3B).<sup>174</sup> Estas estruturas revelaram o enovelamento das proteínas Sm como sendo um barril do tipo SH3 composto de cinco fitas- $\beta$  antiparalelas altamente torcidas, com corte transversal em forma de U, podendo apresentar uma hélice- $\alpha$  na região N-terminal.<sup>166, 174-175</sup> Este enovelamento é classificado como sendo principalmente  $\beta$  de acordo com a classificação CATH (do inglês, *Class, Architecture, Topology, Homologous superfamily*). O motivo Sm1 consiste das fitas- $\beta$ 1,  $\beta$ 2 e  $\beta$ 3, e o motivo Sm2 de fitas- $\beta$ 4 e  $\beta$ 5. Apesar das sequências de aminoácidos de proteínas Sm e LSm de organismos dos três domínios da vida não apresentarem alta identidade, a estrutura tridimensional destas proteínas é altamente conservada, apresentando um desvio quadrático médio das coordenadas atômicas da cadeia principal entre 1 e 2 Å.<sup>166, 174-176</sup>

As proteínas Sm e LSm tendem a formar oligômeros cíclicos como estrutura quaternária. Experimentos de purificação de proteínas Sm recombinantes de levedura estabeleceram a estequiometria do anel Sm *in vivo* e confirmaram a ordem de ligação das proteínas Sm no anel eucariótico.<sup>177</sup> A primeira evidência da estrutura toroidal de oligômeros de proteínas Sm veio com experimentos de microscopia eletrônica de transmissão, no início da década de 2000.<sup>175-176</sup> Geometricamente, um toróide corresponde a uma superfície gerada pela revolução de um círculo no espaço tridimensional em torno de um eixo coplanar a este círculo, porém que não toca o círculo.

Desde a primeira estrutura tridimensional publicada para as proteínas Sm<sup>174</sup>, a combinação de estudos estruturais de baixa e alta resolução tem mostrado diferentes organizações oligoméricas para as proteínas Sm, como o trímero de LSm4 de *S. pombe* (código PDB 4EMH)<sup>178</sup>, os hexâmeros Sm(F/E/G)<sub>2</sub> de humano<sup>176</sup>, LSm(5/6/7)<sub>2</sub> de *S. pombe* (código PDB 4EMK)<sup>178</sup> e de Hfq de *E. coli* (código PDB 3GIB)<sup>179</sup>, o heptâmero de SmAP1 (código PDB 1LNX)<sup>180</sup> e de SmAP2 (código PDB 4XQ3)<sup>181</sup> de bactérias extremófilas, e o octâmero de LSm3 de *S. cerevisiae* (código PDB 3BW1)<sup>182</sup>, além dos heterodímeros humanos SmD1/SmD2 (código PDB 1B34)<sup>174</sup> e SmD3/SmB (código PDB 1D3B)<sup>174</sup> (Figura 2.2). Estudos recentes mostraram que proteínas SmAP de arqueias formam, espontaneamente, longas fibras quando puras em solução.<sup>172</sup>



**Figura 2.2 – Estruturas terciária e quaternária de proteínas Sm/LSm.** A- Dímero de SmD1/SmD2 (código PDB 1B34)<sup>174</sup>, B- Hexâmero de Hfq de *E. coli* (código PDB 3GIB)<sup>20</sup>, C- Heptâmero de SmA1 (código PDB 1LNK)<sup>180</sup> de *P. aerophilum*, D- Octâmero de LSm3 de *S. cerevisiae* (código PDB 3BW1)<sup>182</sup> e E- Trímero de LSm4 de *S. pombe* (código PDB 4EMH)<sup>178</sup>.

Fonte: Elaborada pelo autor usando o programa PyMOL (Schrödinger, LLC).

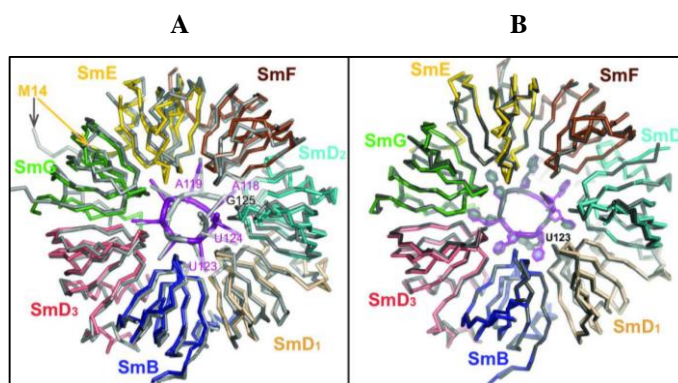
O enovelamento do tipo Sm também foi encontrado em proteínas não envolvidas no metabolismo de RNA ou DNA, com outras funções divergentes, como é o caso do canal MscS (código PDB 2OAU), que forma um homopentâmero. Este fato levanta a hipótese de que o enovelamento do tipo Sm possa ter surgido em eventos independentes na evolução da estrutura de proteínas.<sup>166</sup>

### 2.1.1.1 Interação de proteínas Sm/LSm com snRNAs

A hipótese de que os diferentes snRNAs interagem com as proteínas Sm através do orifício do anel Sm foi confirmada no final década de 2000, com a publicação das estruturas tridimensionais das snRNPs U1<sup>62</sup> e U4<sup>91</sup> de humano (códigos PDB 3PGW e 2Y9A, respectivamente) (Figura 2.3). O sítio Sm rico em uridina de cada molécula de snRNA (118-AAUUUUUGA-126, no caso de snRNA U4, e 125-AAUUUGUGG-133 para o snRNA U1, ambos de *H. sapiens*) interage sequencialmente com SmF, SmE, SmG, SmD3, SmB, SmD1 e SmD2 por meio de interações do tipo ligação de hidrogênio, pareamento  $\pi$ - $\pi$  e interação cátion- $\pi$  no bolsão formado por resíduos em posições equivalentes das alças L3 e L5 das proteínas Sm<sup>55,91</sup> (Figura 2.3).

Os últimos dois nucleotídeos se ligam menos especificamente entre SmD1 e SmD2, e SmE e SmF, respectivamente. Foi observado que a estrutura toroidal das proteínas Sm mantém a conformação tridimensional dos diferentes snRNAs.<sup>62,91</sup> O mesmo padrão de interações é observado no caso do complexo LSm2-8.<sup>90</sup>





**Figura 2.3 – Interação de proteínas Sm com snRNAs.** Comparação das estruturas dos núcleos Sm de snRNPs U4 (A) e U1 (B) de *H. sapiens* em interação com o snRNA cognato.  
Fonte: LI et al.<sup>91</sup>.

### 2.1.1.2 Caracterização de proteínas Sm e LSm de *T. brucei*

Experimentos iniciais de purificação por afinidade em tandem das partículas U2, U4/U6 e U5 de *T. brucei* e sequenciamento das proteínas identificadas por espectrometria de massas por Palfi *et al.*, nas décadas de 1990 e 2000, permitiram a identificação das proteínas SmD1, SmD2, SmE, SmF e SmG de *T. brucei* e levaram à conclusão de que snRNAs de *T. brucei* ligam um conjunto comum de proteínas Sm.<sup>20, 67, 183</sup> As proteínas SmB e SmD3, por sua vez, foram identificadas somente a partir de mineração de dados do genoma de *T. brucei*, comparando-se com sequências de proteínas de *splicing* de outros organismos.<sup>20, 67, 183-184</sup> Curiosamente, as proteínas Sm de *T. brucei* não apresentam reação cruzada com anticorpos anti-Sm humanos<sup>20, 67, 183-184</sup> e possuem uma identidade de sequência de aminoácidos bastante baixa, apesar de conservarem os motivos Sm1 e Sm2.<sup>20, 67, 183-184</sup>

Experimentos funcionais subsequentes mostraram que a proteína SmG de *T. brucei* complementa a função de levedura cujo gene para esta proteína foi nocauteado<sup>184</sup>. Adicionalmente, o silenciamento dos genes codificantes para SmD1 e SmE por meio de experimento de RNA de interferência em *T. brucei* levaram à diminuição dos níveis celulares de snRNAs U1, U2, U4 e U6.<sup>20, 67, 183-184</sup>

Proteínas LSm2-8 de *T. brucei* foram identificadas em 2004, mas somente a função de LSm8 e LSm3 foi caracterizada, sendo as outras reconhecidas apenas pela análise das suas sequências de aminoácidos, levando à errônea identificação inicial das proteínas LSm2 e LSm5 como proteínas LSm. Um segundo estudo distinguiu as proteínas LSm2 e LSm5 verdadeiras e foi observado que o silenciamento de uma única proteína LSm leva à perda do snRNA U6.<sup>20, 99</sup>

Em 2006, Wang e colaboradores<sup>82</sup> identificaram, por meio de purificação por afinidade e coimunoprecipitação de componentes de snRNP U2, as proteínas Sm16,5K (antes conhecida como LSm2) e Sm15K (antes conhecida como LSm5) em substituição às proteínas SmD3 e SmB, respectivamente, no núcleo Sm da partícula snRNP U2.<sup>82</sup> Um resíduo de guanosina do sítio de ligação a Sm de U2 snRNA é determinante para formação do núcleo Sm variante de snRNP U2.<sup>82</sup>

Em 2007, outra variação foi observada, sendo que Tkacz e colaboradores<sup>185</sup> identificaram a proteína SSm4 em substituição à proteína SmD3 na partícula snRNP U4 em *T. brucei*. Neste trabalho, os pesquisadores mostraram que o silenciamento de SSm4 mediado por RNA de interferência leva à redução específica dos níveis celulares de snRNP U4 e a sua caracterização bioquímica revelou a formação do anel Sm correto para a partícula snRNP U4 apenas com a presença da proteína SSm4.<sup>185</sup>

Acredita-se que essa variação no núcleo Sm de snRNPs seja importante para a associação das partículas ribonucleoproteicas e que possa facilitar a função das snRNPs no processo de *splicing* por meio de interações específicas com os snRNAs. Teoricamente, tal variação também pode facilitar o recrutamento de diferentes enzimas modificadoras de capeamento no núcleo das snRNPs, porém, não há correlação entre o tipo de capacete no snRNA e o tipo de núcleo Sm.<sup>20, 82, 186</sup>

## 2.1.2 U2A'/U2B''

### 2.1.2.1 U2A', U2B'' e sua interação com o snRNA U2

As proteínas U2A' e U2B'', juntamente com as proteínas Sm, são as únicas que permanecem como parte da snRNP U2 durante todo o ciclo de *splicing*. Acredita-se que U2A' provê interações proteína-proteína que estabilizam a snRNP U2 e o pré-spliceossomo, e que U2B'' é a base na qual U2A' se liga à snRNP U2.<sup>69</sup>

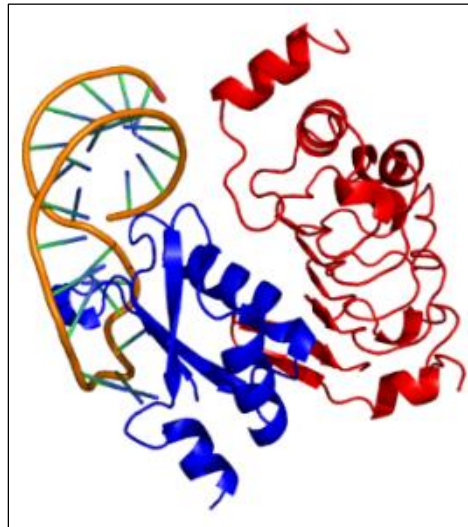
U2A' de humanos é uma proteína modular que contém um domínio N-terminal de 180 resíduos de aminoácidos caracterizado pela presença de sequências repetidas de resíduos de leucina (do inglês, *leucine-rich repeat*, LRR<sup>187</sup>) e um domínio C-terminal que é predito como desordenado, porém com função desconhecida até o momento. O motivo LRR foi identificado em várias de sequências de proteínas em todas as formas de vida e é marcado por uma sequência consenso LxxLxLxxNxL, sendo x qualquer resíduo de aminoácido.<sup>68,187</sup>

Este domínio possui enovelamento do tipo  $\alpha/\beta$  e apresenta uma arquitetura curvada, adequada para interação proteína-proteína (Figura 2.4). Apesar das suas funções diferentes, proteínas que contêm o motivo LRR, em geral, interagem com outra proteína. Cada motivo LRR contribui com uma fita- $\beta$  para a folha- $\beta$  que define a região côncava do domínio LRR. As fitas- $\beta$  são entrelaçadas com uma variedade de elementos de estrutura secundária, como hélice- $\alpha$ , hélice de poliprolina II, hélice  $3_{10}$ , volta- $\beta$  e pequena fita- $\beta$ .<sup>187</sup>

No caso da proteína U2A' de humanos, observou-se que a região côncava de seu domínio LRR se liga à proteína U2B''. Este domínio apresenta uma estrutura na forma de grampo de cabelo  $\beta$  (do inglês,  *$\beta$ -hairpin*) na extremidade N-terminal e uma alça na extremidade C-terminal que atuam como braços para interação com U2B'', que, por sua vez, conserva dois motivos de reconhecimento de RNA (do inglês, *RNA recognition motif*, RRM), separados por uma alça (Figura 2.4).<sup>68</sup>

A função de ligação de RNA do domínio RRM, presente em U2B'', é conservada em proteínas com diferentes funções em todas as formas de vida. Cada domínio RRM consiste de dois motivos: ribonucleoproteína 1 (RNP1), que possui como consenso a sequência [R/K]-G-[F/Y]-[G/A]-[F/Y]-[I/L/V]-x-[F/Y], e o segundo motivo ribonucleoproteína (RNP2), que se encontra a aproximadamente 30 resíduos de RNP1 na extremidade N-terminal, cuja sequência consenso é [I/L/V]-[F/Y]-[I/L/V]-x-N-L, onde x representa qualquer resíduo de aminoácido.<sup>68, 188</sup> A primeira estrutura tridimensional do domínio RRM foi determinada para a proteína spliceossomal U1A<sup>189</sup> no início da década de 1990 e foi confirmada com a determinação da estrutura de U2B'' pelo mesmo grupo alguns anos depois.<sup>68, 188</sup>

O enovelamento do tipo RRM é composto por duas hélices- $\alpha$  empacotadas contra quatro fitas- $\beta$  antiparalelas, resultando em uma topologia descrita como sendo  $\beta\alpha\beta\beta\alpha\beta$ , de modo a formar um sanduíche  $\alpha/\beta$ . Os motivos RRM consenso se localizam nas duas fitas- $\beta$  centrais, com RRM1 em  $\beta 1$  e RRM2 em  $\beta 3$ .<sup>68, 188</sup> No cocrystal contendo a estrutura SL4 do snRNA U2, o motivo RRM1 de U2B'' e o domínio LRR de U2A' de *H. sapiens* (Figura 2.4), o motivo RRM1 interage entre o RNA na superfície da folha- $\beta$  e o motivo LRR na estrutura  $\alpha 1$  na face oposta do motivo RRM1<sup>68</sup> (Figura 2.4).



**Figura 2.4 – Interação de U2A'/U2B'' com o snRNA U2 de *H. sapiens*.** Estrutura cristalográfica de U2A' (vermelho)/U2B'' (azul) <sup>68</sup> interagindo com a estrutura SL4 do U2 snRNA de *H sapiens*.  
 Fonte: Elaborada pelo autor usando o programa *PyMOL* (Schrödinger, LLC).

Ambas as proteínas spliceossomais U1A e U2B'' de humanos conservam o domínio RRM, porém apresentam afinidade e especificidade diferentes por SL2 de snRNA U1 e SL4 de snRNA U2. U1A interage fortemente e com cooperatividade positiva com SL2 de snRNA U1 e U2B'' com SL4 de snRNA U2 (ambas apresentando constante de dissociação na ordem de nanomolar). A constante de dissociação é cinco vezes menor para U1A/snRNA U1 do que para U2A'/snRNA U2 na presença de U2B''.

Experimentos com oócitos de *Xenopus sp.* em que o snRNA U2 foi inativado mostraram que o *splicing* do pré-mRNA pode ser recuperado quando snRNA U2 exógeno é expresso. Quando SL4 é deletada no snRNA U2 exógeno, resultando na consequente perda de interação com U2A'/U2B'', observa-se defeito no *splicing*. Neste caso, o complexo A do spliceossomo (que contém snRNP U1 e U2AF) se acumula nas células e os complexos B e C do spliceossomo (que contém snRNP U2) não são detectados. Resultados similares foram observados para leveduras quando SL4 de snRNA U2 e/ou as proteínas homólogas de U2A' e U2B'' foram deletadas.<sup>190-191</sup>

### 2.1.2.2 Caracterização de U2A'/U2B'' de *T. brucei*

A primeira proteína específica de snRNP U2 a ser identificada em *T. brucei* foi U2-40K, ortóloga de U2A', a partir da sua copurificação com o snRNA U2 em condições de alta estringência.<sup>67, 78</sup> A proteína U2B'', por sua vez, foi identificada por meio de análise de sequências de aminoácidos e observou-se que ela interage com a proteína U2-40K, como as homólogas de humanos e leveduras.<sup>79</sup>

Acredita-se que as proteínas específicas de snRNP U2 de *T. brucei*, Sm16,5K e Sm15K, interajam com a proteína U2A', que por sua vez interage com a proteína U2B''. Este complexo deve interagir especificamente com a extremidade 3' do snRNA U2. Em humanos, U2A' está separada do core Sm por uma estrutura secundária em forma de grampo de cabelo no snRNA U2 (SL3).<sup>82</sup>

## 2.2 Justificativa e objetivos

Apesar da informação funcional acumulada desde o final da década de 1990 a respeito de proteínas Sm e LSm de eucariotos, o seu papel na biogênese de snRNPs e na função do spliceossomo ainda não é bem compreendido. Da mesma forma, a função da variação das proteínas Sm em *T. brucei* nas snRNPs U2 e U4, bem como a necessidade da interação de U2A'/U2B'' para a especificidade a snRNA U2, ainda não estão elucidadas. Deste modo, os objetivos específicos deste trabalho incluem:

- a. Reconstruir *in vitro* os núcleos das snRNPs U2, U4 e U5 de *T. brucei* a partir de subcomplexos de proteínas Sm (SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB, Sm15K/Sm16,5K e Ssm4/SmB) e de U2A'/U2B''.
- b. Investigar a interação entre os complexos Sm de *T. brucei* e os snRNAs U2, U4 e U5.
- c. Investigar a interação do subcomplexo U2A'/U2B'' com o snRNA U2 de *T. brucei*.
- d. Analisar a estrutura terciária e quaternária dos diferentes núcleos de proteínas Sm de *T. brucei* *in silico* e *in vitro*, em comparação com os ortólogos de *H. sapiens*.

## 2.3 Materiais e métodos

Buscando cumprir os objetivos propostos, uma série de experimentos foi realizada *in silico* e *in vitro*, como descrito a seguir:

### 2.3.1 Análise das sequências de aminoácidos das proteínas Sm, U2A' e U2B'' de *T. brucei* por meio de ferramentas de Bioinformática

A predição *in silico* de propriedades de proteínas é bastante útil na sua investigação funcional e também como guia para o trabalho experimental. Deste modo, ferramentas de Bioinformática foram empregadas para a análise das proteínas Sm, U2A' e U2B'' de *T. brucei*.

As sequências de aminoácidos destas proteínas foram obtidas na base de dados *TriTrypDB*<sup>53</sup>. Os parâmetros físico-químicos de cada proteína foram analisados a partir de sua sequência de aminoácidos por meio da ferramenta *PROTPARAM* do servidor *ExpPASy*<sup>192</sup> e a análise dos domínios conservados foi feita por meio do servidor *SMART*<sup>193</sup>. Um alinhamento múltiplo com sequências homólogas foi gerado por meio do programa *Clustal Omega*<sup>194</sup> e a predição de estrutura secundária foi feita por meio do servidor *PSIPRED*<sup>195</sup>. Deste modo, foi possível comparar as sequências de aminoácidos de proteínas de *T. brucei* com aquelas de ortólogas.

### 2.3.2 Clonagem, expressão e purificação de subcomplexos de proteínas Sm

Construções contendo sequências de DNA codificantes para os subcomplexos SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB e Sm15K/Sm16,5K em vetor de expressão pQE30 (*Qiagen*) foram gentilmente cedidas pelo professor Dr. Christian Kambach, do Instituto Paul Scherrer (Suíça). Estes plasmídeos foram construídos para que a primeira proteína de cada subcomplexo seja expressa com uma extensão de seis histidinas na sua extremidade N-terminal de modo que a(s) sua(s) parceira(s) de interação seja(m) copurificada(s) por afinidade.

Por outro lado, as fases abertas de leitura de SSm4/SmB foram obtidas a partir do DNA genômico do *T. brucei* extraído a partir de 1 mL de cultura de *T. brucei* 427 em sua forma procíclica, em fase *logarítmica* de crescimento ( $\sim 5 \times 10^6$  células mL<sup>-1</sup>). As células foram centrifugadas a 3000xg à temperatura ambiente por um minuto e lavadas com tampão A1 (conhecido na literatura, em inglês, como *Tryps Wash Buffer*: 20 mM Tris/HCl pH 7,5, 100 mM NaCl, 3 mM MgCl<sub>2</sub>). O sedimento foi ressuscitado em tampão A2 (conhecido na literatura, em inglês, como tampão *TELT*: 50 mM Tris/HCl pH 8,0, 62,5 mM EDTA, 2,5 mM

LiCl, 4% Triton X-100). DNA foi extraído usando 250  $\mu\text{L}$  de fenol:clorofórmio:álcool isoamil (25:24:1) e centrifugado a 16.000xg por 5 minutos à temperatura ambiente e esse processo foi repetido por duas vezes. Três volumes de etanol 100% gelado e 0,1 volume de acetato de sódio 3M pH 7,0 foram adicionados e a solução foi centrifugada por 15 minutos a 16.000xg a 4°C. O sobrenadante foi descartado e 500  $\mu\text{L}$  de etanol 70% gelado foram adicionados. A solução foi novamente centrifugada por cinco minutos a 16.000xg a 4°C e o produto foi desidratado usando rotação a vácuo (em *speedvac*) durante cinco minutos. Finalmente, o sedimento foi ressuspenso em 50  $\mu\text{L}$  de água deionizada estéril.

A amplificação das fases abertas de leitura (do inglês, *open reading frame*, ORF) dos genes codificantes para as proteínas SSm4 (Tb927.7.6380) e SmB (Tb927.2.4540) a partir de DNA genômico de *T. brucei* foi feita por meio de reação em cadeia da DNA polimerase recombinante de alta fidelidade *Phusion* (*Thermoscientific*), nas condições descritas pelo fabricante, a partir de oligonucleotídeos iniciadores *sense* 5' G/GATCCGATGTCCGGTAGA GGCGAGTAGACC 3' (*BamHI*) e *anti-sense* 5' GC/GGCCGCTCACTCTTTCT TTGCTTTGGTCC 3' (*NotI*) para SSm4 e *sense* 5' CA/TATGGGCCACCAAATA TGCTTCACAACATC 3' (*NdeI*) e *anti-sense* 5' C/TCGAGTCAATCGCGTTTCC GCTTGGC 3' (*XhoI*) para SmB. O resultado foi analisado em gel de agarose preparativo 1% contendo 0,5  $\mu\text{g mL}^{-1}$  de *SyberSafe* (*Thermoscientific*), em tampão TAE 1X (8 mM Tris-acetato pH 8,0, 0,2 mM EDTA).

Os fragmentos amplificados foram purificados a partir do gel de agarose preparativo utilizando-se *Wizard® SV Gel Clean-Up System* (*PROMEGA*), de acordo com as condições do fabricante. Os fragmentos purificados foram clonados em vetor de clonagem pJET1.2 (*Thermoscientific*) e o produto foi transformado em células competentes *E. coli* DH5 $\alpha$  preparadas por meio do método do cloreto de cálcio<sup>196</sup>.

Colônias cresceram em placas LB-ágar (10 g L<sup>-1</sup> triptona, 5 g L<sup>-1</sup> extrato de levedura, 10 g L<sup>-1</sup> NaCl, 1,5% ágar), foram inoculadas em 5 mL de meio de cultura LB contendo ampicilina (20  $\mu\text{g mL}^{-1}$ ) e mantidas a 37°C por 16 horas sob agitação. Em seguida, procedeu-se com extração do DNA plasmidial por meio do kit *GeneJET™ Plasmid Miniprep* (*Fermentas Life Sciences*), de acordo com as condições do fabricante, e a verificação de ligação foi feita por meio de reação em cadeia da DNA polimerase, como descrito anteriormente. Os plasmídeos foram então digeridos utilizando-se as endonucleases de restrição correspondentes.

O vetor de expressão pETDuet-1 (*Novagen*), planejado para a coexpressão de duas ORFs alvo, contém dois sítios de clonagem (MCS1 e MCS2), cada um precedido de um promotor T7/operador lac e um sítio de ligação a ribossomo (RBS), além de gene de resistência

para ampicilina. A ORF para a proteína SSm4 foi inserida em MCS1 para expressão de proteína recombinante em fusão N-terminal com uma extensão de seis histidinas, enquanto a ORF para SmB foi inserida em MCS2. Após reação de ligação com T4 DNA ligase (*Thermoscientific*), o volume total da mistura de ligação foi utilizado na transformação de células competentes *E. coli* DH5 $\alpha$ .

Novamente, procedeu-se com extração do DNA plasmidial por meio do kit *GeneJET™ Plasmid Miniprep* (*Fermentas Life Sciences*), de acordo com as condições do fabricante. A confirmação da ligação foi feita por meio de PCR usando os iniciadores específicos para o vetor de clonagem e os resultados foram verificados a partir de gel de agarose 1%.

Células da cepa de *E. coli* BL21 rosetta1 (DE3) foram transformadas com a construção SSm4/SmB/pETDuet1 utilizando-se o mesmo procedimento descrito anteriormente e foram inoculadas, inicialmente, em 5 mL de meio de cultura LB contendo ampicilina (20  $\mu\text{g mL}^{-1}$ ) e cloroanfenicol (34  $\mu\text{g mL}^{-1}$ ) e incubadas por 16 horas sob agitação de 250 rpm a 37°C. Após crescimento das células, inocularam-se 10 mL dessa amostra em 1 L de meio de cultura LB líquido, mantendo sob agitação a 37°C até se atingir um valor de densidade óptica (DO) igual a 0,6, medido no comprimento de onda de 600 nm em espectrofotômetro. Testes de expressão foram realizados a partir de indução com diferentes concentrações de IPTG (0,1, 0,5 e 1,0 mM) a 20°C e 37°C. Em seguida, as células foram centrifugadas a 12.000xg por 15 minutos a 4°C e o sedimento foi armazenado a -20°C por até um mês.

Por outro lado, os subcomplexos SmD3/SmB, SmE/SmF/SmG, SmD1/SmD2 e Sm16,5K/Sm15K foram coexpressos usando vetor de expressão pEQ30 (*Qiagen*) em cepa *E. coli* SG13009 em meio de cultura LB e indução com 0,1, 0,5 e 1,0 mM IPTG durante a 20°C e 37°C. De maneira similar, as células foram centrifugadas a 12.000xg por 15 minutos a 4°C e o sedimento foi armazenado a -20°C por até um mês.

O sedimento correspondente à coexpressão de cada subcomplexo Sm foi solubilizado em tampão B (50 mM TRIS-HCl pH 8,0, 300 mM NaCl, 2% glicerol, 5 mM DTT e 10 mM Imidazol, na proporção de 50 mL de tampão B para cada 1 L de cultura) acrescido de um comprimido de coquetel de inibidores de protease (*complete™* livre de EDTA, *Roche Diagnosis Ltda*) e submetido à ultrassonicação por seis ciclos no equipamento 550 *Sonic Dismembrator* (*Fisher Scientific*). Cada ciclo corresponde a uma ultrassonicação de 30 segundos com intervalo de um minuto de repouso. As células lisadas foram centrifugadas por cinco minutos a 20.000xg a 4°C, de modo a se separar as frações solúvel (sobrenadante) e insolúvel (sedimento).



A copurificação das proteínas Sm foi realizada em coluna *HiTrap IMAC HP* (Qiagen) acoplada a sistema Akta Purifier10 (*GE*) a partir de um gradiente de concentração de imidazol até 500 mM em tampão B. Exceto para SSm4/SmB, a extensão de seis histidinas foi clivada com protease TEV (250 µg mL<sup>-1</sup>) por 16 horas a 4°C e o produto foi aplicado novamente à coluna *HiTrap IMAC HP* (Qiagen). Amostras contendo subcomplexos que já não apresentavam a extensão de histidinas não interagiram com a resina e foram coletadas e concentradas. Adicionalmente, cromatografia de exclusão molecular em coluna SUPEROSE 12 10/300 (*GE*) ou SUPERDEX200 10/300 (*GE*) foi empregada em tampão B, na ausência de imidazol. Proteínas com massa molecular conhecida foram utilizadas para calibração de uma relação entre massa molecular e volume de eluição permitindo, assim, uma estimativa da massa molecular dos subcomplexos purificados.

Todas as etapas de purificação foram analisadas em gel desnaturante (SDS-PAGE 20% ou NuPAGE 4-20% Bis-Tris (*Thermoscientific*)), corado com *comassie brilliant blue* 0,010% e descorado com ácido acético 10%. As amostras purificadas foram submetidas à concentração por ultrafiltração em membrana *Millipore* com corte de 10 kDa, sob rotação de 2.000xg a 4°C. A concentração proteica das amostras resultantes foi medida utilizando-se o espectrofotômetro *Nanodrop<sup>TM</sup> 1000* (*Thermoscientific*) à partir do coeficiente de extinção molar teórico para cada subcomplexo e comparada com aquela obtida pelo método de Bradford.<sup>197</sup>

Amostras puras dos subcomplexos Sm foram armazenadas a -80°C e enviadas em nitrogênio líquido para o Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas (LMB-MRC) do Reino Unido, onde o aluno Ivan Rosa e Silva realizou experimentos de interação com RNA e de cristalização durante estagio de nove meses.

### 2.3.3 Clonagem, expressão e purificação de U2A'/U2B''

Todos os experimentos com U2A'/U2B'' de *T. brucei* foram realizados sob supervisão do Dr. Kiyoshi Nagai no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido, em Cambridge.

As sequências das fases abertas de leitura de U2A' (Tb927.10.14360) e U2B'' (Tb927.3.3480) de *T. brucei* foram otimizadas para expressão heteróloga em *E. coli* usando a ferramenta de otimização de códons da empresa *Integrated DNA Technologies* (IDT) (<https://www.idtdna.com/CodonOpt>) e sintetizada de maneira a conter a seguinte construção: sítio de *NdeI* + sítio de reconhecimento por TEV + U2A' + sítio de ligação a ribossomo (do inglês, *ribosome binding site*, RBS) + U2B'' + sítio de *XhoI* (onde RBS =

AATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG e sítio de reconhecimento por TEV = GAGAATCTTTATTTTCAG) ligada em vetor de clonagem pMK\_RQ, que contém gene de resistência a canamicina.

Produtos de restrição com *NdeI* e *XhoI* foram clonados em vetor de expressão pET28a(+) e transformados em células eletrocompetentes *E. coli* BL21 DE3 RIL usando protocolo padrão<sup>196</sup>. A ligação foi confirmada por análise de restrição com as mesmas enzimas e o produto foi sequenciado. Nestas condições, este vetor leva à expressão de U2A' contendo uma extensão N-terminal de 6 resíduos histidinas e de U2B'' não fusionada à mesma extensão sob o controle de promotor T7/operador lac.

Colônias cresceram em placas contendo meio sólido TYE-ágar (15 g L<sup>-1</sup> ágar, 8 g L<sup>-1</sup> NaCl, 10 g L<sup>-1</sup> bactotripton, 5 g L<sup>-1</sup> extrato de levedura) acrescidas de 50 µg mL<sup>-1</sup> de canamicina a 37°C por 16 horas. Uma pré-cultura foi preparada a partir do inóculo de uma colônia em meio de cultura LB contendo 50 µg mL<sup>-1</sup> de canamicina. O produto foi diluído (1:100) em 1 L de meio de cultura LB e a cultura foi mantida a 37°C/250 rpm até se atingir DO<sub>600</sub> ~0,6, quando a expressão proteica foi induzida pela adição de 1 mM IPTG.

Após 16 horas a 20°C, as células foram centrifugadas a 4.500 rpm (rotor Beckman SL8.1) por 30 minutos e o sedimento foi ressuspensionado em tampão contendo uma pastilha de coquetel de inibidores de protease (*Roche Diagnosis Ltda*), na proporção de 50 mL de tampão de lise para cada 1 L de cultura. As células ressuspensionadas foram ultrassonicadas em pelo menos seis ciclos de 30 segundos, com um minuto de parada e o lisado foi centrifugado por 40 minutos a 20.000xg a 4°C. O sobrenadante clarificado foi aplicado a uma coluna de afinidade HisTrap HP column (*GE*) nas mesmas condições de tampão em um gradiente linear de concentração de imidazol até 500 mM usando o sistema Akta Prime (*GE*) em câmara fria.

As frações correspondentes ao pico principal foram dialisadas contra tampão C (10 mM fosfato de potássio pH 7,4, 100 mM KCl, 10 mM DTT) e a amostra resultante foi aplicada a uma coluna de hidroxiapatita (25 g de hidroxiapatita sólida dissolvida em 50 mL de tampão C – Biorad), lavada com tampão C e eluída em gradiente linear de tampão D (tampão C + 6% sulfato de amônio) em 40 mL. As frações eluídas foram concentradas por ultrafiltração (Amicon 10 kDa - Sigma Aldrich Ltda) e dialisadas contra tampão E (50 mM Tris/HCl pH 8, 100 mM KCl, 1 mM DTT, 2% glicerol) e armazenadas a -80°C. A concentração da amostra foi avaliada em espectrofotômetro *Nanodrop™ 1000 (Thermoscientific)* a partir do coeficiente de extinção molar teórico para o subcomplexo e conferida utilizando-se o método de Bradford<sup>197</sup>. A pureza das amostras foi analisada por eletroforese em gel desnaturante NuPAGE 4-12% Bis-Tris (*Thermoscientific*) (180V/55 minutos) corado com 0,25% comassie *brillant blue*.

### 2.3.4 Preparo dos snRNAs U2, U4 e U5 e sua caracterização por meio de espectroscopia de dicroísmo circular

Moldes de DNA para a transcrição *in vitro* de snRNAs U2, U4 e U5 foram baseados em predições de estrutura 2D publicadas previamente, descritas no capítulo 1 (seção 1.3). Oligonucleotídeos foram sintetizados de modo a se construir a região 3' de snRNA U2 (SLII + sítio Sm + SLIV), U4 snRNA (SLII + sítio Sm + SLIII) e snRNA U5 completo na seguinte disposição: sítio de *HindIII* + promotor de T7 RNA polimerase + sequência codificante para RNA + sítio de *BsmAI* + sítio de *BamHI*, onde a sequência correspondente ao promotor de T7 RNA polimerase é dada por AATTAATACGACTCACTATA. O sítio *BsmAI* foi substituído pelo sítio de *BsrDI* no caso do molde de snRNA U5.

Os oligonucleotídeos para cada molde de DNA foram hibridizados por meio de tratamento térmico. Tubos contendo oligonucleotídeos em concentração igual a 10  $\mu\text{M}$ , diluídos em água deionizada estéril e livre de nucleases, foram incubados por cinco minutos a 98°C e transferidos para um béquer contendo água fervente. Este foi deixado à temperatura ambiente até que o equilíbrio térmico fosse atingido. O resultado foi analisado em gel de agarose 1,5% (1X tampão TBE – 0,22 mM Tris pH 8,3, 180 mM borato de sódio, 5 mM EDTA).

O número de moléculas de DNA molde para transcrição de cada snRNA de interesse foi amplificado por PCR utilizando-se oligonucleotídeos específicos. O produto foi purificado a partir de extração com fenol/clorofórmio e precipitação com etanol/isopropanol, da mesma maneira como descrito na seção 2.3.2. As amostras resultantes foram digeridas com enzimas de restrição específicas para clonagem em vetor pUC18 (*Thermoscientific*) digerido nos sítios de *BamHI* e *HindIII*. Os produtos de ligação foram transformados em células eletrocompetentes *E. coli* TG1 e transformantes foram selecionados em placas TYE-ágar contendo 100  $\mu\text{g mL}^{-1}$  ampicilina, 16  $\mu\text{L}$  X-Gal 50  $\text{mg mL}^{-1}$  e 4  $\mu\text{L}$  IPTG 1M. PCR de colônia e análise de restrição com enzima PacI (*Thermoscientific*), que cliva em um sítio específico na sequência do promotor de T7 RNA polimerase, foram empregados para confirmar colônias positivas.

Para se obter um grande número de plasmídeos que pudessem servir como molde para a transcrição *in vitro*, experimentos de purificação plasmidial em larga escala foram realizados com a ajuda da Dra. Kelly Nguyen a partir de 1 L de culturas de *E. coli* TG1 contendo cada construção plasmidial. Resumidamente, as células foram centrifugadas (8.000 rpm, 30 minutos, rotor SLA3000, centrífuga SORVALL), ressuspensas em 30 mL de tampão conhecido como TGE (25 mM Tris/HCl pH 8, 50 mM glicose, 10 mM EDTA) e 5  $\text{mg mL}^{-1}$  lisozima, 80 mL de 1% SDS e 0,2 M NaOH foram adicionados para a lise. O conteúdo foi centrifugado a 8.000 rpm

por 10 minutos (rotor SLA3000, centrífuga SORVALL) e o sobrenadante foi filtrado em papel de filtro. Foi então adicionada RNase A (0,5 mL – 10 mg mL<sup>-1</sup>) e o produto foi precipitado com etanol 100%. A amostra foi ressuspensa em 360 µL de brometo de etídeo (5 mg mL<sup>-1</sup>) acrescida de 17,2 g de cloreto de cézio.

Bandas de DNA foram separadas por ultracentrifugação a 55.000 rpm a 20°C por 16 horas (rotor Becmann VTi 65.2/tubos selados). As bandas vermelhas correspondentes ao DNA foram coletadas usando seringa e agulha. A amostra resultante foi misturada com butan-2-ol puro e saturado com água até que a fase aquosa se tornasse límpida. Finalmente, as amostras contendo os plasmídeos foram dialisadas contra tampão TE (10 mM Tris/HCl pH8, 0,1 mM EDTA) e armazenadas a -20°C.

Os plasmídeos contendo região promotora de T7 RNA polimerase e as sequências codificantes para snRNAs U2, U4 e U5 foram linearizados com a enzima de restrição que cliva na extremidade 3' da sequência codificadora para o RNA de interesse. Reações de digestão em larga escala foram feitas em 5 mL de mistura de reação contendo 1 mg de DNA e 100 unidades de enzima de restrição, incubada a 37°C por 16 horas. As amostras de DNA foram purificadas a partir de extração com fenol/clorofórmio e precipitação com etanol/isopropanol, como descrito anteriormente.

Experimentos de transcrição *in vitro* foram realizados em 5 mL de mistura de reação contendo 4 mM ATP, UTP, GTP e CTP, 40 mM Tris/HCl pH 8, 20 mM MgCl<sub>2</sub>, 2 mM espermidina, 100 mg mL<sup>-1</sup> de plasmídeo linearizado ou 20 µg µL<sup>-1</sup> de produto de PCR e 15 µg µL<sup>-1</sup> de T7 RNA polimerase. Cada reação foi incubada a 37°C de 4-6 horas. O produto foi extraído com fenol/clorofórmio e precipitado com isopropanol/etanol, como descrito previamente. Em seguida, a amostra resultante foi dissolvida em tampão de amostra (95% (v/v) formamida, 5 mM EDTA, 0,1% (w/v) xileno cianol, 0,1% (w/v) azul de bromofenol) e aplicada em PAGE 15% em 1X TBE buffer contendo 8 M uréia a 35 W. As bandas foram coradas com 0,1% azul de toluidina.

Amostras de RNA transcrito foram dialisadas contra tampão 10 mM Tris/HCl pH 8 e o produto foi novamente precipitado com isopropanol/etanol, ressuspensa em água e armazenado a -20°C. Uma fração das amostras de RNA foi marcada com fluoresceína tiosemicarbazida na extremidade 3'. Brevemente, o sedimento correspondente a 100 µg de RNA foi ressuspensa em 100 µL de solução de oxidação fresca (0,1 M periodato de sódio, 0,1 M acetato de sódio pH 5) e incubado por 90 minutos à temperatura ambiente. A oxidação foi suprimida pela adição de 11 µL de KCl (2M) em gelo por 10 minutos. Sedimento de KIO<sub>4</sub> foi removido e fluoresceína semicarbazida foi adicionada na concentração final de 50 mM, e a

amostra foi incubada à temperatura ambiente por quatro horas. A amostra resultante foi purificada por meio de extração com fenol/clorofórmio e precipitação com isopropanol/etanol. Em seguida, o material marcado foi purificado das moléculas não marcadas a partir de SDS-PAGE 15% contendo uréia, como descrito anteriormente. A banda amarela foi excisada do gel e mantida por 16 horas em 600 µL de 0,3 M acetato de sódio a 4°C sob agitação. A amostra resultante foi novamente precipitada com etanol e o pellet foi ressuspensionado em água e armazenado a -20°C.

O envelhecimento das moléculas de RNA foi alcançado por meio de tratamento térmico em tampão de ligação (20 mM Tris/HCl pH 7,5, 100 mM NaCl ou KCl, 1,5 mM MgCl<sub>2</sub>, 1 mM DTT, 2% glicerol) ou tampão de hibridização (10 mM Tris/HCl pH 8 e 10 mM MgCl<sub>2</sub>). Os tubos contendo amostras de RNA foram incubados a 98°C por três minutos, por cinco minutos à temperatura ambiente e, em seguida, mantidos no gelo por 30 minutos.

#### 2.3.4.1 Espectroscopia de dicroísmo circular

A maioria das macromoléculas biológicas são opticamente ativas e a atividade óptica está intimamente relacionada à sua conformação. Portanto, a técnica de espectroscopia de dicroísmo circular (sigla, do inglês, para *circular dichroism*, CD) foi empregada para a confirmação do envelhecimento de snRNAs U2, U4 e U5. Os espectros de CD foram obtidos utilizando-se um espectropolarímetro JASCO J-815 (JASCO), equipado com um sistema de controle de temperatura do tipo PELTIER PTC 423S/15, no Laboratório de Biofísica Molecular do IFSC-USP. Os espectros foram coletados na região do UV distante, no intervalo de comprimento de onda de 320 a 200 nm, com resolução de 0,1 nm e taxa de 100 nm min<sup>-1</sup>, com resposta de dois segundos, utilizando-se uma cubeta de quartzo de caminho óptico de 0,5 cm para amostras em concentração de 0,10 a 0,30 mg mL<sup>-1</sup>.

O programa *Spectra Manager* (JASCO) foi usado para registro dos dados e o programa *Origin PRO 8.0* (*Origin Lab Corporation*) foi empregado para análise dos resultados. Os espectros foram medidos em termos de elipticidade ( $\theta$ ), que é proporcional à diferença de absorção de luz circularmente polarizada à esquerda e à direita, e foram transformados para valores absolutos de dicroísmo circular ( $\Delta\epsilon$ ) molar, em cm<sup>2</sup>mmol<sup>-1</sup>, por meio da seguinte equação:

$$\Delta\epsilon = \frac{\theta}{32.980 \times C \times L \times N} \quad (1)$$

onde  $C$  é a concentração da amostra medida em  $\text{mol.L}^{-1}$ ,  $L$  é o caminho óptico dado em centímetros e  $N$  é o número de nucleotídeos da molécula de RNA.

Os espectros de CD obtidos nas diferentes temperaturas foram submetidos a uma análise de componentes principais (do inglês, *principal components analysis*, PCA) no programa *CDTools*.<sup>198,199</sup> A técnica se baseia em uma análise estatística multivariada de modo a se reduzir a dimensionalidade do conjunto de dados para um número mínimo de variáveis que descrevem os dados originais.<sup>199</sup> Uma descrição detalhada desta técnica pode ser encontrada em Jolliffe, 2002<sup>199</sup>.

### 2.3.5 Reconstrução *in vitro* dos núcleos Sm de *T. brucei*

De posse de amostras puras de subcomplexos de proteínas Sm, U2A'/U2B'' e snRNAs U2, U4 e U5 dos núcleos de snRNPs de *T. brucei*, ensaios de reconstrução *in vitro* foram preparados em tampão de ligação (20 mM Tris/HCl pH 7,5, 100 mM NaCl ou KCl, 1,5 mM  $\text{MgCl}_2$ , 1 mM DTT, 2% glicerol). Para experimentos envolvendo baixas concentrações de RNA, amostras de snRNAs marcados com fluoresceína na concentração de 2 nM foram incubadas com concentrações crescentes de proteínas Sm e/ou U2A'/U2B'' por 30 minutos a 30°C e, em seguida, mantidas no gelo por pelo menos duas horas. No caso de experimentos em alta concentração de RNA, concentrações de até 2  $\mu\text{M}$  de snRNA não-marcado foram utilizadas.

#### 2.3.5.1 Ensaio de retardo de mobilidade eletroforética (EMSA)

O ensaio de retardo de mobilidade eletroforética (do inglês, *electromotility shift assay*, EMSA) baseia-se na alteração de mobilidade eletroforética de uma molécula marcada com um fluoróforo, neste caso uma molécula de snRNA, após interação com proteína. Eletroforese em gel de agarose 0,5% nativo preparado com tampão TB 0,5X (27 g de Tris base e 13,25 g de ácido bórico em 1 L de água milli-Q) foi aplicada para se investigar a formação de complexos entre snRNAs e proteínas Sm e também com U2A'/U2B'' utilizando-se EMSA. Os géis foram submetidos à diferença de potencial elétrico de 45 V 90 minutos em câmara fria. O equipamento *Typhoon TRIO imager (GE)* foi utilizado para se quantificar a intensidade de fluorescência das moléculas de RNA marcadas com fluoresceína. Neste caso, os ensaios foram realizados em concentração de RNA igual a 2 nM, o que permitiu a determinação da constante de dissociação aparente dos complexos a partir do ajuste dos dados experimentais ao modelo de Hill<sup>200</sup>, descrito pela equação:

$$I = I_{inicial} + (I_{final} - I_{inicial}) \frac{[L]^n}{K_{d,ap} + [L]^n} \quad (2)$$

onde  $I$  é a intensidade de fluorescência medida para cada banda no gel,  $[L]$  é a concentração de ligantes livres,  $K_{d,ap}$  é a constante de dissociação aparente e  $n$  é o coeficiente de Hill, que descreve a cooperatividade do sistema. Tal ajuste foi realizado no programa *Origin PRO 8.0* (*Origin Lab Corporation*). Experimentos em alta concentração de RNA e proteínas não marcados foram utilizados para a determinação da estequiometria de ligação.

### 2.3.5.2 Ensaio de purificação

Primeiramente, os núcleos de proteínas Sm de snRNPs U2, U4 e U5 de *T. brucei* foram reconstituídos *in vitro* na ausência de RNA em tampão de ligação. Cerca de 500  $\mu$ L de cada amostra em concentração de 2  $\mu$ M foram preparados na razão estequiométrica 1:1:1 de cada subcomplexo específico para cada núcleo Sm e mantidos a 30°C por 30 minutos e, em seguida, em gelo por pelo menos duas horas. Cromatografia de exclusão molecular em SUPEROSE 12 10/300 GL (*GE*) foi empregada para a avaliação da formação dos complexos proteicos na ausência de RNA. Proteínas com massa molecular e raio hidrodinâmico conhecidos foram utilizadas como padrão para uma estimativa dessas propriedades dos complexos Sm purificados. Os produtos de purificação foram avaliados por SDS-PAGE seguido de coloração com prata (*SilverQuest*®, *Invitrogen*).

Sabendo-se que cromatografia de troca aniônica em coluna Mono Q 5/50 GL (*GE*) havia sido utilizada com sucesso para a purificação dos núcleos de snRNPs U1 e U4 de *H. sapiens* obtidos por reconstrução *in vitro* no laboratório do Dr. Kiyoshi Nagai, a purificação dos núcleos das snRNPs U2, U4 e U5 contendo proteínas Sm e snRNAs de *T. brucei* foi testada utilizando-se a mesma técnica e as mesmas colunas cromatográficas. Cromatografia de exclusão molecular em SUPERDEX 200 5/150 GL (*GE*) e SUPERDEX 200 10/300 GL (*GE*) também foi empregada.

Amostras contendo complexos reconstruídos *in vitro* em volume de 500  $\mu$ L consistiram de 2  $\mu$ M de amostra de snRNA e 2  $\mu$ M de amostras de subcomplexos de proteínas na razão molar 1:1 em tampão de ligação, como descrito previamente. A coluna Mono Q 5/50 GL (*GE*) foi equilibrada com tampão E1 (20 mM Tris/HCl pH 7,5, 0-100 mM NaCl ou KCl, 1,5 mM MgCl<sub>2</sub>, 1 mM DTT, 2% glicerol). As amostras foram eluídas em diferentes gradientes lineares de tampão E2 (20 mM Tris/HCl pH 7,5, 1 M NaCl ou KCl, 1,5 mM MgCl<sub>2</sub>, 1 mM DTT, 2%

glicerol). Adicionalmente, cromatografia de exclusão molecular em SUPERDEX 200 5/150 GL (GE) e SUPERDEX 200 10/300 GL (GE) foi aplicada em tampão E1 contendo 50 mM NaCl ou KCl. Em todos os casos, a absorvância de luz UV a 260 nm (RNA) e a 280 nm (proteína) foi monitorada. As amostras foram então analisadas por meio de eletroforese em NuPAGE 4-12% Bis-Tris (180V/55 minutos) seguida de coloração com prata (*SilverQuest®*, *Invitrogen*).

Tentativas de concentração das amostras de complexos proteínas-RNA foram realizadas por meio de ultrafiltração (filtro Amicon, corte de 10 kDa - *Sigma Aldrich Ltda*) e ultracentrifugação (90 000 rpm - rotor TLA100, *Beckman*).

### **2.3.6 Ensaios de cristalização, análise de dados de difração de raios-X e resolução de estrutura tridimensional**

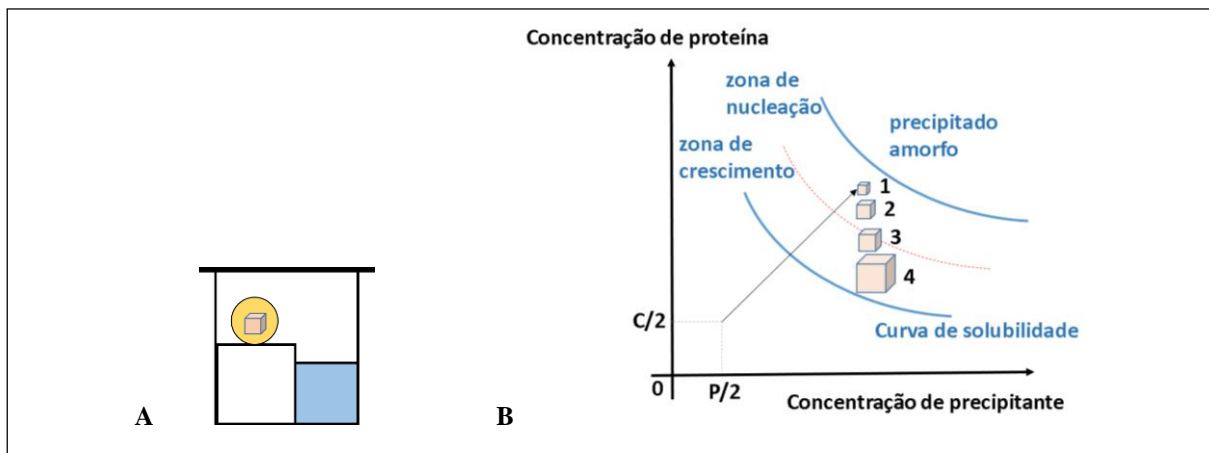
#### **2.3.6.1 Cristalização**

O processo de cristalização de uma proteína é crucial para a determinação de sua estrutura atômica pelo método cristalográfico e se concentra na redução da sua solubilidade de maneira controlada na tentativa de que as macromoléculas se organizem de maneira periódica durante a separação de fase para formação de um cristal único que tenha poder de difração de raios-X. A solubilidade da proteína pode ser reduzida por meio da adição de agentes precipitantes (sais, solventes orgânicos e polietilenoglicóis (PEG)) e/ou pela remoção de solvente da solução de proteína. Temperatura e pH também afetam a solubilidade proteica. Assim que o limite de solubilidade da macromolécula é excedido, a solução se torna supersaturada e, em condições adequadas, as moléculas em excesso se agregam formando núcleos e a solução proteica se separa em proteína agregada (possivelmente na forma de cristais) mais solução saturada, na zona metaestável do diagrama de fases (Figura 2.5).<sup>201-203</sup>

O tempo para se atingir o equilíbrio tem grande influência no resultado final, que pode ser desde um precipitado amorfo ou microcristais até um cristal único e grande. O procedimento mais comum utilizado para cristalização de proteínas é o método de difusão de vapor, que se baseia em um processo de equilíbrio entre duas soluções, uma gota e um reservatório, através da fase de vapor, em ambiente fechado. A gota possui uma solução menos concentrada de precipitantes e, portanto, maior pressão de vapor do que o reservatório, e tende a equilibrar a sua pressão de vapor com a pressão de vapor da solução. Ocorre, então, a difusão de compostos voláteis da gota para o reservatório, tornando a gota concentrada até um nível ótimo para cristalização da proteína. Desde que o sistema entre em equilíbrio, esta condição é mantida até



que a cristalização seja completada.<sup>201-203</sup> Entre os métodos que utilizam este princípio está o método da gota sentada (*sitting drop* – Figura 2.5), que foi empregado neste trabalho.



**Figura 2.5 – Método de difusão de vapor e o diagrama de fases.** A- Representação esquemática de um experimento típico de difusão de vapor a partir de gota sentada (*sitting drop*). Gota mostrada em laranja e reservatório em azul. B- Diagrama de fases mostrando a curva de solubilidade de uma proteína e as regiões de nucleação espontânea, de crescimento cristalino e de formação de precipitado amorfo. Inicialmente, iguais concentrações de proteína (C) e agentes precipitantes (P) são misturadas e vapor de solvente se difunde até o reservatório do coquetel de precipitantes, de modo que a concentração de proteína e de precipitantes aumenta na gota ao passo que seu volume diminui. Uma vez que a região de nucleação espontânea é atingida (1), agregados cristalinos na forma de núcleos são formados e um cristal inicia o processo de crescimento (2-4). O número de moléculas de proteína em solução passa a diminuir devido ao crescimento do cristal e a solução supersaturada tende a se equilibrar com o(s) cristal(is), que atingem o tamanho máximo. Fonte: Adaptada de RUPP<sup>203</sup>.

A dificuldade no controle do processo de cristalização advém da complexidade das interações fracas entre as macromoléculas, as quais são flexíveis e apresentam superfície irregular.<sup>201-203</sup> Portanto, a homogeneidade conformacional, a flexibilidade e a distribuição de carga na sua superfície são fatores determinantes da cristalizabilidade de uma macromolécula. Quando se trata da cristalização de um complexo de macromoléculas recombinantes, objetivo deste trabalho, as etapas de reconstituição *in vitro* e purificação (seção 2.3.5) são essenciais para se garantir homogeneidade do complexo a ser cristalizado tanto em termos de composição como na sua conformação.

Dada a complexidade do processo de cristalização, faz-se necessária a variação sistemática de condições físico-químicas no espaço multiparamétrico que abrange diferentes concentrações de diferentes agentes precipitantes, tampões e aditivos, além de concentração de proteína e temperatura. Deste modo, amostras de núcleo snRNP U2, núcleo snRNP U2 + U2A'/U2B'' e núcleo snRNP U5 obtidas por reconstrução *in vitro* foram submetidas a ensaios de cristalização pelo método de difusão de vapor em gota sentada a 20°C em placas de 96 poços utilizando-se condições de cristalização comerciais (*Crystal screen 1/2/Lite*, *Grid screen ammonium sulfate/PEG ion/Sodium chloride/PEG 6000*, *MemFac*, *Quick screen*, *PEG/ion*,

*Natrix, Index, SaltRX 1/2 – Hampton Research, Wizard screen 1/2/cryo 1/cryo 2 – Rigaku, JBS1-10, Pi minimal screen, Pi-PEG screen – Jenna BioScience, Clear strategy screen 1/2, MemStart, MemSys, Morpheus I/II – Molecular dimensions, JCSG+ – Qiagen*) preparadas em robô de cristalização TTP Mosquito (TTP Labtech). O rastreamento de condições ótimas de cristalização baseadas nos resultados iniciais foi realizado em placas de 24 poços pelo mesmo método à partir da adição de volumes iguais (1 a 2  $\mu\text{L}$ ) de solução de proteína e de solução do reservatório.

Métodos de cristalização envolvendo o controle da nucleação por meio de semeadura (do inglês, *seeding*)<sup>204-205</sup> também foram empregados na tentativa de melhorar a qualidade dos cristais obtidos. Inicialmente, procedeu-se com o preparo de um estoque de sementes, onde 10  $\mu\text{L}$  de solução do reservatório foram misturados com a solução do reservatório da condição que continha muitos cristais pequenos, que foram transferidos para um microtubo contendo esferas de vidro, onde foram agitados por 3 minutos. O procedimento foi repetido até se obter 50  $\mu\text{L}$  de solução. Em seguida, foi preparada uma diluição seriada de fragmentos de cristais quebrados (estoque de sementes) na solução do reservatório nas seguintes proporções: 1:5, 1:25, 1:125 e 1:625.<sup>204-205</sup> Estes estoques foram resfriados em nitrogênio líquido e mantidos a  $-80^\circ\text{C}$ .

No experimento de *microseeding*<sup>204-205</sup>, os microcristais (sementes) foram introduzidos em solução de proteína em concentração menor do que a necessária para a nucleação, o que foi baseado nos experimentos iniciais de cristalização. Deste modo, cada gota de cristalização continha 0,6  $\mu\text{L}$  de amostra de proteínas, 0,4  $\mu\text{L}$  de solução de reservatório e 0,2  $\mu\text{L}$  de estoque de sementes nas diferentes diluições. Adicionalmente, foi empregada uma variação deste método conhecida, em inglês, como *streak seeding*<sup>204-205</sup>, que consiste do uso de um bigode de gato para transporte de microcristais de um estoque de sementes para uma gota de cristalização pré-equilibrada. Em ambos os casos, as concentrações de precipitantes e macromolécula deveriam posicioná-la na zona de crescimento (metaestável) no diagrama de fases.

### 2.3.6.2 Coleta e análise de dados de difração de raios-X

A difração de raios-X nos monocristais obtidos foi testada *in house* no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas (LMB-MRC) em Cambridge, Reino Unido, usando um detector do tipo placa de imagem MAR345 a 200 mm do cristal ( $\Delta\phi = 0,5^\circ$ /tempo de exposição por imagem = 30 minutos), com fonte de raios-X do tipo ânodo rotatório e comprimento de onda igual a 1,54 Å à temperatura de 80 K. Conjuntos completos de dados foram coletados com auxílio do Dr. Chris Oubridge na linha luz I24 da *Diamond Light Source*

(DLS), que possui um detector Pilatus 6 M. Conjunto de dados nativos ( $\Delta\varphi = 0,15^\circ$ / distância cristal-detector igual a 500 mm/ temperatura = 80 K) foram coletados com pequenas doses de raios-X (tempo de exposição igual a 0,02 segundos com transmitância igual a 41,3%) em comprimento de onda de 0,9795 Å.

Cada ponto no padrão de difração é também conhecido como reflexão por corresponder à intensidade ( $I_{hkl}$ ) de uma onda eletromagnética com amplitude  $|\vec{F}_{hkl}|$  e fase  $\varphi_{hkl}$  refletida em um plano de Bragg do cristal descrito pelos índices de Miller ( $h, k, l$ ).<sup>203, 206</sup> A intensidade das reflexões parciais e totais medidas foram relacionadas aos diferentes índices de Miller no processo de indexação e integradas usando o programa *iMosflm*<sup>207</sup>, o qual também realiza o refinamento dos parâmetros que caracterizam a cela unitária.

Sabendo-se que mudanças na intensidade do feixe de raios-X durante a coleta de dados, dano causado pela radiação, mosaicidade do cristal, entre outros, influenciam nos valores de intensidades medidas para cada reflexão total, e, considerando-se que cada átomo em uma macromolécula real terá um possível deslocamento individual, foi feito o escalonamento das reflexões observadas em diferentes imagens de difração de raios-X para uma escala comum usando o programa *AIMLESS*<sup>208</sup>. O processo de escalonamento resulta na determinação de um fator de escala linear global ( $k$ ), bem como na correção do fator B global, que contabiliza o deslocamento individual dos átomos.<sup>203, 208</sup>

Durante a coleta de dados, múltiplos dados redundantes para reflexões relacionadas por simetria foram medidos e precisaram ser fusionados em um conjunto de dados únicos que representa a unidade assimétrica do cristal no espaço recíproco, o que foi feito também utilizando o programa *AIMLESS*<sup>208</sup>. As intensidades equivalentes, portanto, foram usadas para a determinação de um conjunto único de amplitudes de fatores de estrutura ( $|\vec{F}_{hkl}|$ ), que correspondem à raiz quadrada da intensidade total escalonada para cada ( $h, k, l$ ). A multiplicidade do conjunto de dados corresponde ao número de reflexões totais medidas dividido pelo número de reflexões totais únicas, já se considerando a simetria do cristal. Além disso, a qualidade dos dados de difração coletados também foi avaliada pela razão sinal-ruído calculada como a média  $\langle I/\sigma(I) \rangle$  para todas as reflexões em uma dada camada de resolução e pelos fatores conhecidos como  $R_{\text{merge}}$  (do inglês, *linear merging R-value*) e  $R_{\text{rim}}$  (do inglês, *redundancy-independent merging R-value*). Todas estas propriedades estatísticas dos dados experimentais também foram calculadas utilizando-se o programa *AIMLESS*<sup>208</sup>.

Os fatores R correspondem ao módulo da diferença entre a intensidade medida e a intensidade média calculada para cada ( $h, k, l$ ) somada para todas as reflexões medidas, dividida

pela soma de todas as intensidades medidas e levando-se em consideração a redundância dos dados<sup>203, 208</sup>. Assim, os fatores R dependem do fator de escala. Além disso, o fator  $R_{\text{merge}}$  acaba sendo linearmente dependente da redundância dos dados. Por isso, o fator  $R_{\text{rim}}$  é definido com um termo multiplicativo que o torna independente desta. O coeficiente de correlação linear (CC), por sua vez, é independente do fator de escala e é definido como<sup>203, 208</sup>

$$CC = \sigma_{xy} / \sigma_x \sigma_y \quad (3)$$

onde  $\sigma_x$  e  $\sigma_y$  representam a variância das intensidades medidas separadas em dois conjuntos de dados ( $x$  e  $y$ ).  $CC_{1/2}$  é definida como a correlação linear entre duas metades do conjunto de dados experimentais separadas aleatoriamente.

A análise inicial da simetria cristalina foi feita pelo programa *POINTLESS*<sup>209</sup>. Adicionalmente, foi realizada a análise do conteúdo de solvente do cristal por meio do cálculo do coeficiente de Matthews ( $V_M$ )<sup>210</sup>, que corresponde à razão entre o volume da unidade assimétrica e a massa molecular das macromoléculas presentes dentro desta. O conteúdo de solvente ( $\tau$ ) corresponde, portanto, a

$$\tau = 1 - 1,23/V_M \quad (4)$$

### 2.3.6.3 Resolução do problema das fases, refinamento e análise do modelo de estrutura tridimensional

O primeiro objetivo do experimento de cristalografia é a determinação da densidade eletrônica ( $\rho(x, y, z)$ ) da(s) macromolécula(s) que espalharam raios-X. Esta é calculada a partir da transformada de Fourier inversa dos fatores de estrutura para cada ( $h, k, l$ ):<sup>203, 211</sup>

$$\rho(x, y, z) = \frac{1}{V} \sum_{\vec{h}=-\infty}^{+\infty} \vec{F}(\vec{h}) \exp(-2\pi i \vec{h} \cdot \vec{r}) \quad (5)$$

onde  $\vec{h} = h \hat{a}^* + k \hat{b}^* + l \hat{c}^*$  (espaço recíproco),  $\vec{r} = x \hat{a} + y \hat{b} + z \hat{c}$  (espaço real) e  $\hat{a}$ ,  $\hat{b}$  e  $\hat{c}$  caracterizam a cela unitária do cristal. Cada fator de estrutura é descrito como:

$$\vec{F}(\vec{h}) = |\vec{F}(\vec{h})| \exp(i\varphi_{hkl}) \quad (6)$$

onde  $|\vec{F}(\vec{h})|$  é a amplitude e  $\varphi_{hkl}$  é a fase da onda eletromagnética. Entretanto, apenas a amplitude da onda pode ser determinada à partir da intensidade detectada para cada  $(h, k, l)$  no experimento de difração.<sup>203, 211</sup>

Assim, o programa *PHASER*<sup>212-213</sup> foi empregado para a determinação inicial das fases pelo método de substituição molecular. Um modelo de estrutura atômica conhecida do heterodímero U2A'/U2B'' de *H. sapiens*<sup>68</sup> (código PDB 1A9N) foi posicionado na unidade assimétrica a partir de rotações e translações até que os dados de difração experimentais e aqueles calculados para o modelo tivessem a melhor correspondência. As fases deste modelo nesta nova orientação foram calculadas e usadas para a obtenção de um primeiro mapa de densidade eletrônica a ser refinado. Para evitar a influência das cadeias laterais dos resíduos de aminoácidos do modelo posicionado na unidade assimétrica e das regiões de alça que têm sequências de aminoácidos com comprimentos diferentes entre o modelo, optou-se por utilizar o programa *CHAINSAW*<sup>214</sup>, que realizou o truncamento do modelo no primeiro átomo da cadeia lateral de cada resíduo que é diferente daquele observado para a(s) proteína(s) de interesse.

A substituição molecular automatizada implementada no programa *PHASER*<sup>212-213</sup> combina a correção dos dados anisotrópicos com otimização de funções de translação e rotação baseadas em verossimilhança (do inglês, *likelihood*, *LL*) e análise de empacotamento para resolver a estrutura cristalográfica em estudo. Uma função de verossimilhança (*LL*) descreve a probabilidade de se observar um valor experimental ( $|\vec{F}_{obs}(h, k, l)|$ ) dado um modelo teórico ( $\vec{F}_{calc}(h, k, l)$ ), sendo que cada átomo no cristal contribui com um desvio relacionado à variação de sua posição (que afeta a fase) e na sua ocupação (que afeta a amplitude). Assim, a função *LL* é tratada como uma distribuição gaussiana.<sup>212-213</sup>

Como as fases utilizadas para a reconstrução do mapa de densidade eletrônica provêm do modelo posicionado na unidade assimétrica, esta densidade eletrônica inicial está enviesada, incompleta e incorreta em várias regiões. Para se obter um mapa mais realístico, coeficientes de Fourier que levam em conta erros e incompletezas do modelo devem ser usados para o cálculo de  $\rho(x, y, z)$ . A reconstrução automática do modelo atômico no mapa de densidade eletrônica foi, portanto, realizada por meio do programa *ARP/wARP*<sup>215</sup>, com o objetivo de diminuir o viés do modelo inicial.

Finalmente, o modelo de estrutura atômica foi refinado usando os programas *Refmac5*<sup>216</sup> e *Phenix-refine*<sup>217</sup>, também baseados em máxima verossimilhança<sup>218</sup>. Os parâmetros do modelo que incluem a posição de cada átomo e fator B (exceto ocupação), foram utilizados na otimização do modelo até se alcançar a maior concordância entre as amplitudes

das reflexões observadas e aquelas calculadas para o modelo. O resultado foi avaliado por meio do fator R (ou  $R_{\text{work}}$ ):

$$R = \frac{\sum_{\vec{h}=-\infty}^{+\infty} [F_{\text{obs}}(\vec{h}) - F_{\text{calc}}(\vec{h})]}{\sum_{\vec{h}=-\infty}^{+\infty} F_{\text{obs}}(\vec{h})} \quad (7)$$

O fator  $R_{\text{free}}$  foi calculado para validação cruzada da mesma maneira que  $R_{\text{work}}$  para um conjunto de reflexões selecionadas aleatoriamente que corresponde a uma pequena porcentagem do total e que não foi usada durante o refinamento.<sup>203, 218</sup>

O programa *Coot*<sup>219</sup> foi utilizado para a inspeção do ajuste do modelo à densidade eletrônica calculada contornada a  $1\sigma$ , onde

$$\sigma = \left\{ \frac{1}{g} \sum_{\vec{r}} [(\rho(\vec{r}))^2 - \langle \rho \rangle^2] \right\}^{0,5} \quad (8)$$

Os programas *Coot*<sup>219</sup>, *PyMOL* (*Schrödinger, LLC*) e *Chimera 1.9*<sup>220</sup> foram usados na visualização do modelo cristalográfico refinado. A avaliação da geometria do modelo de estrutura cristalográfica<sup>221</sup> obtido foi feita a partir da ferramenta *MolProbity*<sup>222</sup> e a estrutura quaternária foi avaliada no servidor *PISA*<sup>223</sup>. Adicionalmente, a superfície eletrostática foi gerada pelas ferramentas *APBS* e *PDB2PQR*<sup>224</sup>, usando campo de força molecular *Amber* e pH 7,5.

O desvio quadrático médio (do inglês, *root mean square deviation, rmsd*) esperado para as distâncias entre átomos equivalentes nas estruturas tridimensional de *T. brucei* e de *H. sapiens* foi medido no programa *PyMOL* (*Schrödinger, LLC*) e foi comparado com aquele obtido pela seguinte relação teórica:<sup>225</sup>

$$\Delta = 0,4 \exp[1,87(1 - i)] \quad (9)$$

onde  $i$  corresponde à fração de resíduos conservados para o alinhamento entre duas sequências comparadas. Neste trabalho,  $i$  corresponde à fração de resíduos conservados em cada complexo avaliado.

### 2.3.7 Modelagem da estrutura terciária e quaternária

Dada a evidência de que a estrutura tridimensional de proteínas Sm é mais conservada evolutivamente que suas sequências de aminoácidos, técnicas de modelagem molecular por homologia e por alinhamento (do inglês, *threading*) foram empregadas para a análise estrutural dos núcleos Sm de *T. brucei*. Em ambos os casos, informações acerca da sua estrutura tridimensional obtida por cristalografia, ressonância magnética nuclear ou até mesmo criomicroscopia eletrônica são levadas em consideração.<sup>226-227</sup>

No caso da modelagem molecular por homologia, foi utilizado o programa *Modeller*<sup>228</sup>, que se baseia na satisfação, de maneira ótima, de restrições espaciais geradas a partir de uma estrutura molde conhecida relacionadas com a proteína-alvo por meio de um alinhamento de sequências de aminoácidos. Estas restrições são expressas na forma de uma função de densidade de probabilidade para a localização de cada átomo na proteína.<sup>228</sup>

Por outro lado, o servidor *I-Tasser*<sup>227</sup> foi empregado para execução da técnica de modelagem por alinhamento. Fragmentos de sequências de aminoácidos que se alinham com a sequência alvo são procurados em um banco de dados de estruturas de proteínas. Esta etapa é direcionada pela similaridade local entre a estrutura secundária predita para a proteína em estudo e a apresentada pelo fragmento molde daquela região. A qualidade do alinhamento é medida baseando-se na significância estatística do melhor alinhamento. A organização dos fragmentos em uma estrutura tridimensional é feita por meio de simulação de Monte Carlo por *replica-exchange* até que se encontre um modelo otimizado.<sup>227</sup> Caso existam regiões da sequência de interesse que não alinhem com nenhuma sequência com estrutura conhecida, o servidor *I-Tasser*<sup>227</sup> utiliza modelagem *Ab initio*. Esta última se utiliza de um campo de força molecular que define a interação entre os átomos que compõem a sequência primária em análise e de condições iniciais do sistema.<sup>227</sup>

A maioria das proteínas de *T. brucei*, quando comparadas com ortólogas de *H. sapiens*, apresentam valores de identidade de sequências de aminoácidos próximos a 30% e encontram-se no limiar da faixa de identidade ideal para o uso da técnica de modelagem comparativa. A escolha entre modelagem por homologia ou por alinhamento, portanto, depende da qualidade do alinhamento entre as sequências de aminoácidos. O desvio quadrático médio (*rmsd*) esperado para as distâncias entre átomos equivalentes nas estruturas 3D após sobreposição entre os modelos dos complexos Sm de *T. brucei* e seus homólogos de *H. sapiens* foi medido utilizando-se o programa *PyMOL* (*Schrödinger, LLC*) e seu valor foi comparado com aquele obtido pela equação (9).

A avaliação da qualidade dos modelos de estrutura terciária e quaternária das proteínas em estudo foi feita a partir da análise de sua geometria por meio dos servidores *PDBsum*<sup>229</sup> e *ProSa*<sup>230</sup>, e sua visualização foi possível utilizando os programas *PyMOL* (*Schrödinger, LLC*) e *Chimera* 1.9<sup>220</sup>. Finalmente, propriedades hidrodinâmicas dos modelos gerados foram obtidas a partir do programa *HYDROPRO*<sup>231</sup> para comparação com aquelas medidas experimentalmente (seções 2.3.5.2 e 2.3.8).

### 2.3.8 Espalhamento de raios-X a baixo ângulo (SAXS)

A avaliação experimental da estrutura em baixa resolução dos complexos Sm de *T. brucei* na ausência de RNA foi possível por meio de espalhamento de raios-X a baixo ângulo (do inglês, *small angle X-ray scattering*, SAXS). Amostras puras dos núcleos Sm de *T. brucei* purificadas na ausência de RNA em concentrações de 1,5, 3,0, 6,0 e 12,0 mg mL<sup>-1</sup> em tampão 50 mM TRIS-HCl pH 8,0, 300 mM NaCl, na presença de 5 mM DTT, foram submetidas a um feixe de raios-X ( $\lambda = 1,48 \text{ \AA}/288 \text{ K}$ ) e a intensidade ( $I$ ) da radiação espalhada foi medida em detector MAR-165 na linha SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP), em função do vetor de espalhamento, cujo módulo é dado por

$$q = \frac{4\pi}{\lambda} \text{sen}(\theta) \quad (10)$$

onde  $\theta$  é igual à metade do ângulo de espalhamento.

A análise sistemática dos dados de espalhamento de raios-X foi feita, inicialmente, através do programa *FIT2D*<sup>232</sup>, normalizando-se os dados pela intensidade do feixe e multiplicando-se pela absorção da amostra. O espalhamento da solução tampão foi subtraído da curva de espalhamento da solução proteica, resultando na medida  $I_{exp}(q)$ , que foi transformada em escala absoluta a partir da seguinte relação:<sup>233</sup>

$$I_{abs}(q) = \frac{I_{exp}(q)}{I_{exp,\acute{a}gua}(q)} \times \frac{d\Sigma}{d\Omega} \quad (11)$$

onde  $I_{exp}(q)$  é o espalhamento medido experimentalmente,  $I_{exp,\acute{a}gua}(q)$  corresponde ao espalhamento experimental de água deionizada a 288 K e  $\frac{d\Sigma}{d\Omega}$  é a seção de choque diferencial para o espalhamento de água igual a  $1,632 \times 10^2 \text{ cm}^{-1}$  em escala absoluta.<sup>233</sup>



A partir dos dados de espalhamento de raios-X foi possível uma caracterização comparativa dos complexos Sm por meio do cálculo do raio de giro, que corresponde à raiz quadrada da distância quadrática média entre pontos da densidade eletrônica de cada complexo e seu centro de massa. O raio de giro de cada complexo Sm foi obtido utilizando-se a aproximação de Guinier<sup>234</sup>, válida tipicamente para  $qR_g < 1,3$ :

$$I(q) = I(0)\exp\left(-\frac{R_g^2 q^2}{2}\right) \quad (12)$$

onde  $I(0)$  é a intensidade no ângulo zero. A partir do espalhamento experimental também foi possível calcular o volume de Porod ( $V$ ), dado por:

$$V = \left[2\pi^2 \frac{I(0)}{Q}\right] \quad (13)$$

onde  $Q$  corresponde ao invariante

$$Q = \int_0^\infty [I_{exp}(q) - A]q^2 dq \quad (14)$$

onde  $A$  é uma constante apropriada subtraída de cada ponto experimental para se obter um decaimento proporcional a  $q^{-4}$  em altos ângulos, de acordo com a lei de Porod para partículas homogêneas em solução. Também foi possível calcular a curva de Krakty<sup>234</sup> para cada conjunto de dados experimentais.

Adicionalmente, a distribuição de distâncias  $p(r)$  foi calculada pelo método que resolve a transformada de Fourier inversa, implementado no programa *GNOM*<sup>235</sup>.

$$p(r) = \frac{r}{2\pi^2} \int_0^\infty q I(q) \sin(rq) dq \quad (15)$$

Modelos de baixa resolução para o envelope molecular de cada complexo Sm foram gerados pelo método *ab initio* por meio do programa *DAMMIN*<sup>236</sup>, assumindo simetria P1. Arrefecimento simulado (do inglês, *simulated annealing*) é usado na procura por um modelo compacto que concorde bem com os dados experimentais, através de minimização de uma função do parâmetro discrepância:<sup>236</sup>

$$\chi = \left\{ \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{N(i)} \left[ \left( I_{\text{exp}}^{(i)}(q_j) - I_{(q_j)}^{(i)} \right) / \sigma(q_j) \right]^2 \right\}^{0,5} \quad (16)$$

onde  $M$  corresponde ao número de curvas ( $i = 1, 2, \dots, M$ ),  $N^{(i)}$  é o número de pontos na  $i$ -ésima curva e  $\sigma(q_j)$  denota erros experimentais. Foram realizadas 10 reconstruções *ab initio* independentes com o programa *DAMMIN*<sup>236</sup>, que foram promediadas no programa *DAMAVR*<sup>237</sup>. O programa *SUPCOMB*<sup>238</sup> foi utilizado para sobreposição de envelope de SAXS com modelos tridimensionais obtidos a partir de modelagem por homologia.

O programa *CRY SOL*<sup>239</sup> foi utilizado para a comparação do espalhamento de raios-X a baixo ângulo teórico dos modelos da estrutura tridimensional dos complexos Sm de *T. brucei* com aquele obtido experimentalmente. O refinamento de corpo rígido desses modelos foi realizado no programa *SASREF*<sup>240</sup>, que também utiliza algoritmos baseados em arrefecimento simulado. Finalmente, foram utilizados três métodos diferentes para o cálculo da massa molecular a partir dos dados de espalhamento de raios-X a baixo ângulo:

**A** – A partir da expressão proposta por Feigin e Svergun<sup>241</sup> e Orthaber<sup>233</sup>:

$$MM = [N_A I_{abs}(0)/c] / \Delta \rho_M^2 \quad (17)$$

onde  $I_{abs}(0)/c$  corresponde ao espalhamento absoluto no ângulo zero dividido pela concentração,  $N_A$  é o número de Avogadro e

$$\Delta \rho_M = [\rho_{M,prot} - (\rho_{solv} \bar{v})] r_0 \quad (18)$$

onde  $\rho_{M,prot} = 3,22 \times 10^{23} \text{ eg}^{-1}$  é o número de elétrons por massa seca de proteína,  $\rho_{solv} = 3,34 \times 10^{23} \text{ ecm}^{-3}$  é o número de elétrons por volume de solvente aquoso,  $\bar{v}$  é o volume parcial específico de proteína, assumido como sendo igual a  $0,74 \text{ cm}^3 \text{ g}^{-1}$  para proteínas globulares, e  $r_0 = 2,8179 \times 10^{-13} \text{ m}$  é o comprimento de espalhamento de um elétron.

**B** – Por meio do programa *SAXSMoW*<sup>242</sup>, que usa dados de espalhamento experimental em escala relativa. Este programa calcula a massa molecular de uma macromolécula com volume  $V$  a partir da seguinte equação:

$$MM = V \cdot \rho_{M,prot} = \left[ 2\pi^2 \frac{I_{exp}(0)}{Q} \right] \rho_{M,prot} \quad (19)$$

Portanto, tal método não requer a medida de concentração de proteína ou a medida da diferença entre densidade eletrônica de proteína e de tampão, como no método A.

C – A partir do volume dos átomos DAM<sup>236</sup> (empírico):

$$MM (kDa) = V_{DAM}/2,0 \quad (20)$$

## 2.4 Resultados

### 2.4.1 Análise das sequências de aminoácidos das proteínas Sm, U2A' e U2B'' de *T. brucei* por meio de ferramentas de Bioinformática

A Tabela 2.1 apresenta um resumo das características físico-químicas das proteínas Sm, U2A' e U2B'' de *T. brucei* e de seus complexos. A maior parte dos oligômeros de proteínas Sm e U2A'/U2B'' de *T. brucei* apresenta ponto isoelétrico teórico acima de 9,00, o que indica uma carga total positiva em pH fisiológico. Esse resultado é esperado para proteínas que interagem com molécula de RNA, cuja carga total é negativa.

O coeficiente de extinção molar e a massa molecular teóricos apresentados na Tabela 2.1 foram utilizados para o cálculo da concentração de amostras de subcomplexos, como descrito nas seções 2.3.2 e 2.3.3, e na análise das etapas de purificação dos mesmos. Deve-se notar que somente as proteínas SmE, SmF e Sm16,5K possuem ao menos um resíduo de triptofano em sua estrutura primária. Isso resulta em coeficiente molar teórico com uma imprecisão maior que 10% para a maioria dos subcomplexos. Deste modo, o método de Bradford<sup>197</sup> também foi utilizado para a checagem da concentração das amostras purificadas e o resultado sempre foi comparável.

Tabela 2.1 – Características físico-químicas de proteínas Sm, U2A' e U2B'' de *T. brucei*.

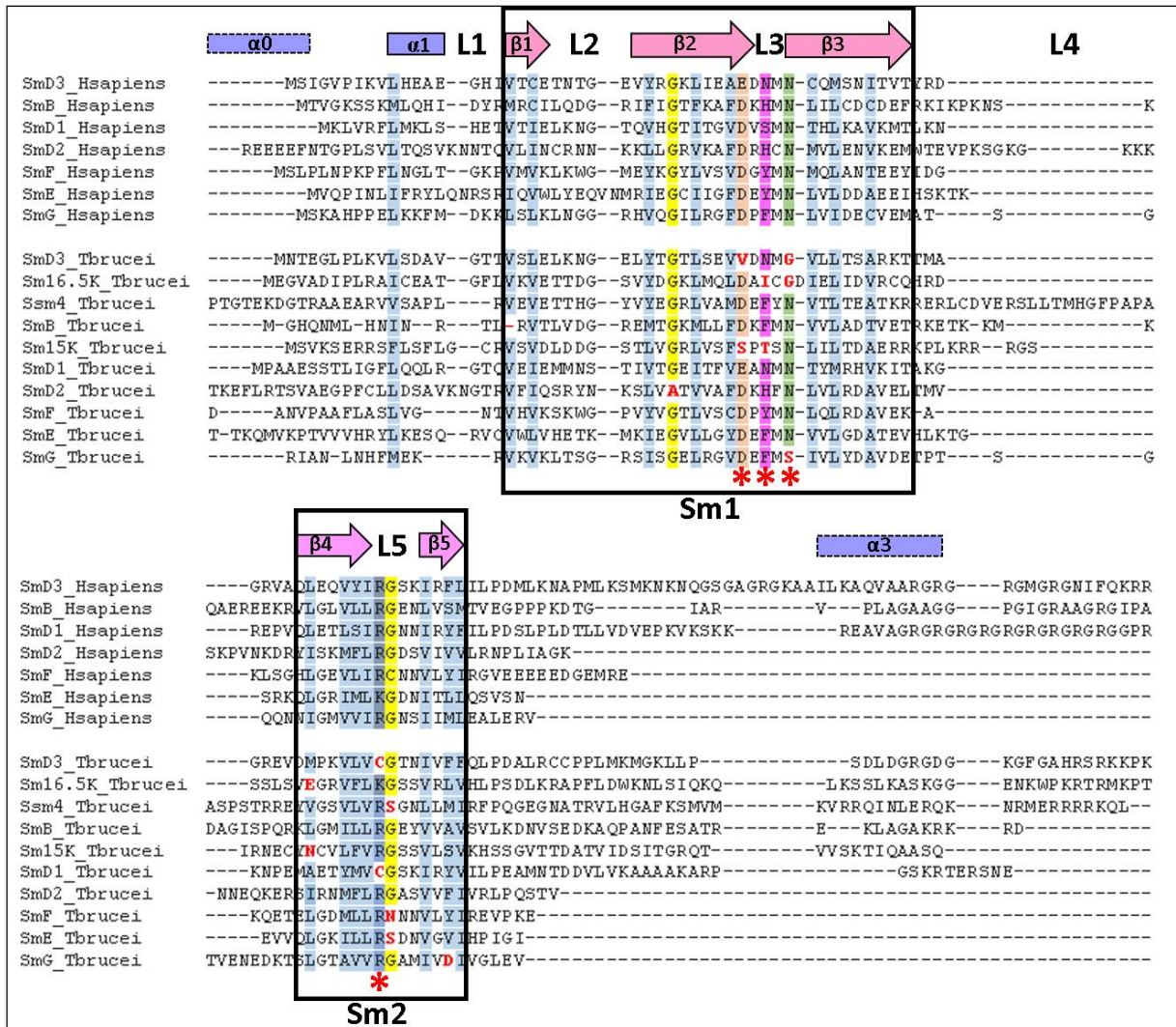
Proteínas	Código na base de dados <i>TriTrypDB</i> <sup>53</sup>	Massa molecular teórica (Da)	pI teórico	Coefficiente de extinção molar teórico a 280 nm (M <sup>-1</sup> cm <sup>-1</sup> )
SmD3	Tb927.4.890	12363,5	9,30	1490*
Sm16,5K	Tb927.10.4950	14743,3	9,68	12490
SSm4	Tb927.7.6380	23213,6	10,19	5960*
SmB	Tb927.2.4540	12331,4	9,92	1490*
Sm15K	Tb927.6.4340	12791,6	10,27	1490*
SmD1	Tb927.7.3120	11713,5	7,86	4470*
SmD2	Tb927.2.5850	12518,5	10,00	1490*
SmF	Tb927.9.10250	8361,6	5,68	9970
SmE	Tb927.6.2700	9655,4	7,93	8480
SmG	Tb927.11.14310	8922,3	5,76	1490*
SmD3/SmB	-	24676,9	9,71	2980*
Sm16,5K/ Sm15K	-	27516,9	9,92	13980
SSm4/SmB	-	35526,9	10,31	7450*
SmD1/SmD2	-	24214,0	9,51	5960*
SmE/SmF/SmG	-	26903,3	6,61	19940
Complexo 1 (SmD3/SmB+ SmE/SmF/SmG+ SmD1/SmD2)	-	75758,3	9,34	28880
Complexo 2 (Sm16,5K/Sm15K+ SmE/SmF/SmG+ SmD1/SmD2)	-	78598,3	9,46	39880
Complexo 3 (SSm4/SmB+ SmE/SmF/SmG+ SmD1/SmD2)	-	86608,4	9,59	33350
U2A' (U2-40K)	Tb927.10.14360	36492,4	6,29	13980
U2B''	Tb927.3.3480	13626,5	10,06	10430
U2A'/U2B''	-	50100,9	9,13	24410

\* Sequências de aminoácidos que não apresentam o resíduo de triptofano.

Fonte: Elaborada pelo autor a partir de resultados da ferramenta *ProtParam* do servidor *ExPASy*<sup>192</sup>

### 2.4.1.1 Análise da conservação de aminoácidos das proteínas Sm e dos sítios de interação com snRNAs

A Figura 2.6 mostra a análise de conservação das seqüências de aminoácidos de proteínas Sm de *T. brucei* e de *H. sapiens* acompanhada da predição de estrutura secundária.



**Figura 2.6** – Alinhamento múltiplo de seqüências de aminoácidos de proteínas Sm de *H. sapiens* e *T. brucei* feita a partir do servidor *Clustal Omega*<sup>194</sup>. Resíduos de aminoácidos evidenciados: hidrofóbicos em azul, Gly em amarelo, sítios de ligação a nucleotídeo: Asp ou Glu em rosa, aromáticos em roxo, Asn em verde, e Arg ou Lys em azul. Resíduos não conservados nos motivos Sm1 e Sm2 são destacados em vermelho. Os quatro resíduos de aminoácidos que interagem diretamente com o sítio Sm no snRNA são destacados com o símbolo \*. Estrutura secundária: hélice- $\alpha$  em azul e fita- $\beta$  em rosa. A hélice- $\alpha 0$  é predita apenas para SSm4 e SmD2 de *T. brucei*.

Fonte: Elaborada pelo autor a partir do resultado do servidor *Clustal Omega*<sup>194</sup>.

Os motivos Sm1 (hélice- $\alpha 1$  e fitas- $\beta 1$ ,  $\beta 2$  e  $\beta 3$ ) e Sm2 (fitas- $\beta 4$  e  $\beta 5$ ) são degenerados e contêm apenas alguns dos resíduos de aminoácidos conservados em todas as proteínas da família Sm. Os resíduos de metionina e asparagina na fita- $\beta 3$  do motivo Sm1 e os resíduos arginina e glicina na fita- $\beta 4$  do motivo Sm2 são os mais bem conservados na família Sm, como

já destacado na Figura 2.1 (seção 2.1.1). Entretanto, dentre todas as proteínas Sm de *T. brucei*, apenas a proteína SmB apresenta todos os quatro resíduos conservados.

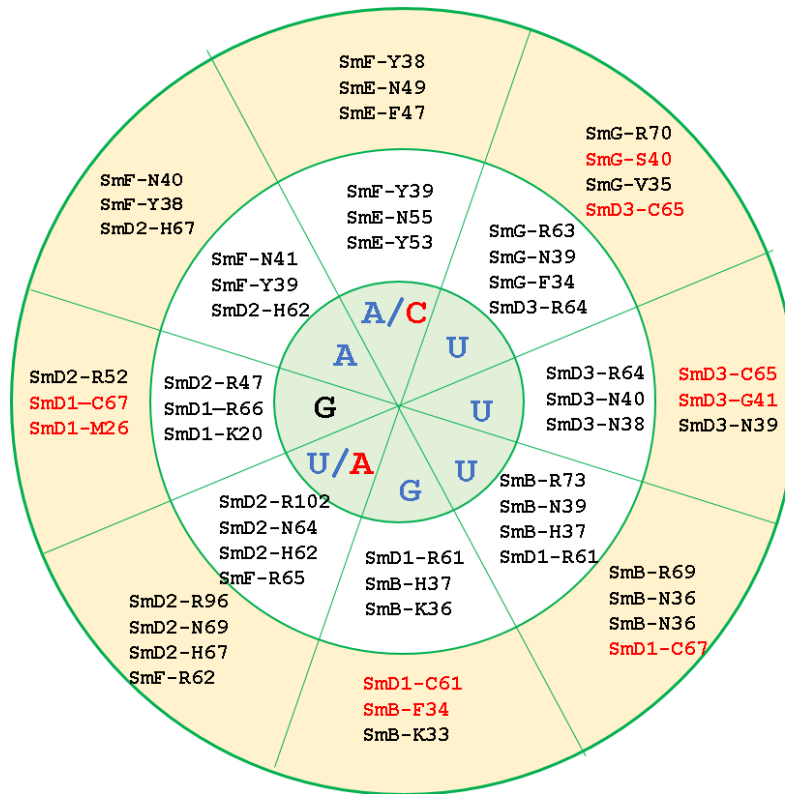
Uma região de alça variável em tamanho entre as fitas- $\beta 3$  e  $\beta 4$ , denominada alça L4, separa os domínios Sm1 e Sm2. A proteína SSm4 possui a maior separação entre estas regiões. Esta proteína também possui extremidades N- e C-terminal bastante extensas em comparação com as outras. SmD2 de *T. brucei*, por sua vez, também possui uma extremidade N-terminal mais extensa, com a predição de uma hélice- $\alpha$  adicional nessa região. Por outro lado, hélices- $\alpha$  adicionais são preditas na extremidade C-terminal das proteínas SmB, SmD1, Sm15K e Sm16,5K deste organismo (Figura 2.6). Estas regiões adicionais podem estar envolvidas em interações específicas para se manter a estrutura dos núcleos de snRNP em *T. brucei* e possivelmente com outras proteínas spliceossomais. Estudos adicionais são necessários para elucidar sua função.

Como descrito no capítulo 1 (seção 1.3.2.4), os motivos ricos em arginina/glicina (RG) na extremidade C-terminal das proteínas SmD1, SmD3 e SmB de *H. sapiens* são dimetilados pelo complexo PRMT5/WD45.<sup>110, 243, 244</sup> As metilações levam a um aumento da afinidade dos núcleos Sm pelo complexo SMN, o que se mostra como um mecanismo de regulação importante na biogênese de snRNPs.<sup>110, 243, 244</sup> Entretanto, esses motivos não são conservados nas ortólogas de *T. brucei*, sendo que apenas SmD3 apresenta padrão de sequência de aminoácidos com as mesmas características físico-químicas de motivos RG na extremidade C-terminal. Porém, nenhuma metilação foi caracterizada para SmD3 de *T. brucei* até o momento.

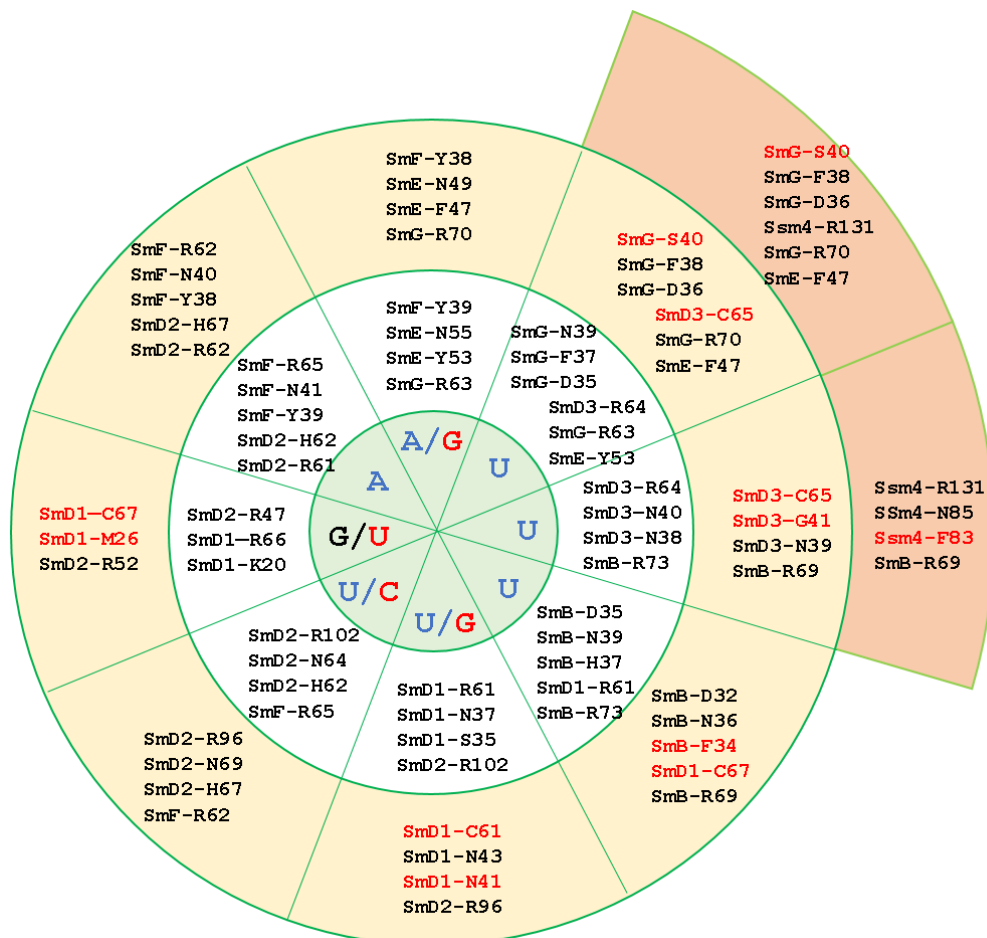
É notável que as proteínas Sm de *T. brucei* apresentam variação nas posições consenso dos motivos Sm1 e Sm2. Contudo, as posições hidrofóbicas são praticamente inalteradas (destacadas em azul, na Figura 2.6), o que está em acordo com a observação prévia de que uma mutação nesses resíduos leva à morte de células de levedura.<sup>91</sup>

A Figura 2.7, a seguir, mostra uma comparação entre os sítios Sm dos snRNAs U1, U2, U4 e U5 de *T. brucei* e *H. sapiens*, bem como a correspondência entre resíduos de aminoácidos de proteínas Sm de *T. brucei* preditos como responsáveis pela interação com os sítios Sm de snRNAs U1, U2 e U4 com base nas estruturas cristalográficas de núcleos de snRNPs U1<sup>55</sup> e U4<sup>91</sup> de *H. sapiens* e no alinhamento de sequências de aminoácidos apresentado na Figura 2.7.

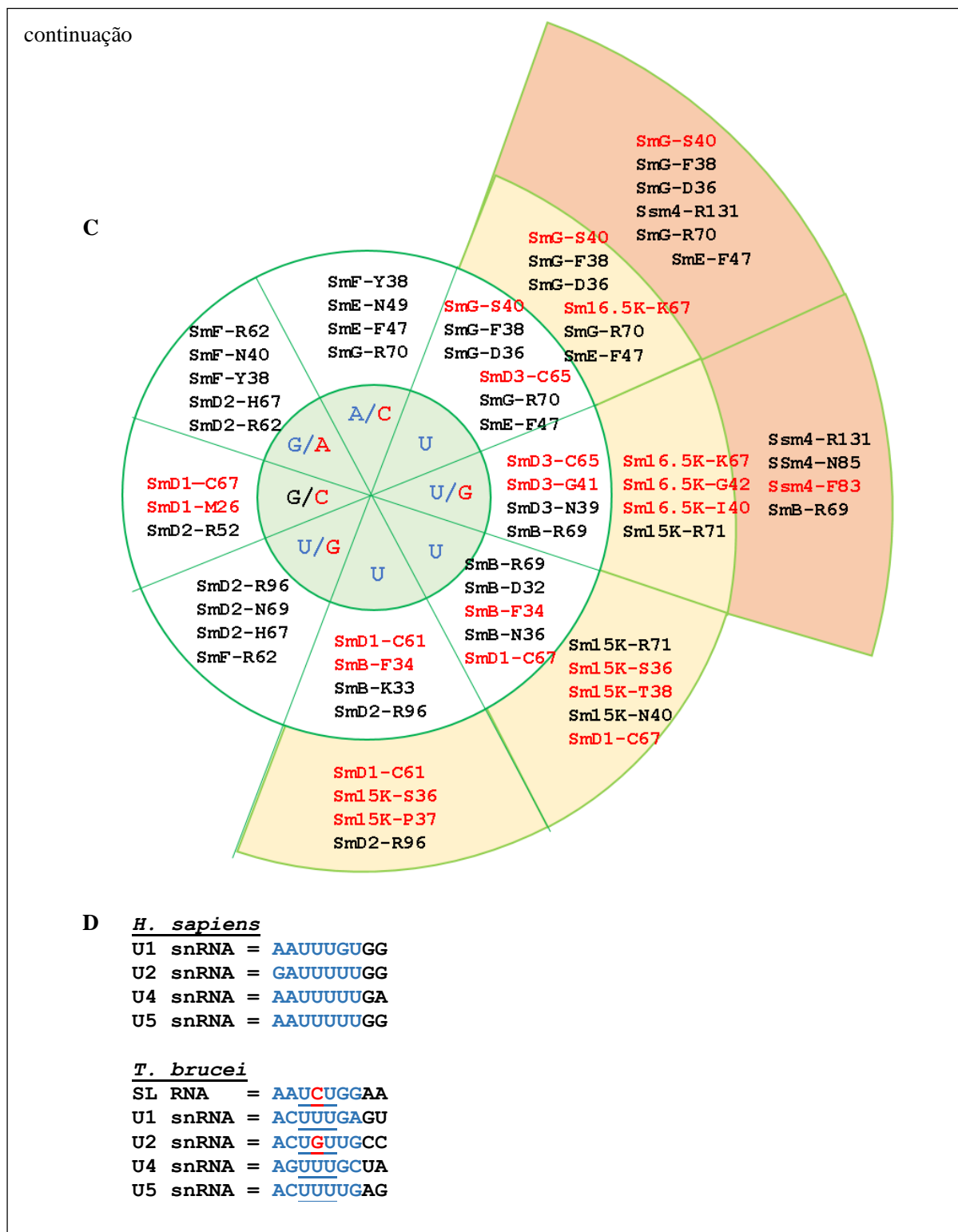
A



B



continua



**Figura 2.7 – Comparação entre resíduos de aminoácidos conservados nos domínios Sm1 e Sm2 de proteínas Sm de *T. brucei* e a posição de interação no sítio Sm de snRNAs U1, U2, U4 e U5. Análise dos sítios Sm no núcleo de snRNPs A- U1, B- U4 e C- U2, baseada nas estruturas cristalográficas de núcleos de snRNPs U1 e U4 de *H. sapiens*, e D- de conservação dos sítios Sm nos snRNAs U1, U2, U4 e U5 de *T. brucei* e *H. sapiens*. Em vermelho, são destacados os resíduos que variam nas proteínas Sm ortólogas de *T. brucei* e *H. sapiens*. O anel interno corresponde ao sítio Sm no snRNA. O próximo anel corresponde aos resíduos das proteínas Sm de *H. sapiens*, enquanto os seguintes correspondem às variantes em *T. brucei*.**

Fonte: Elaborada pelo autor.



Em *T. brucei*, o núcleo Sm de snRNP U1 é formado pelas sete proteínas Sm canônicas (SmF/SmE/SmG/SmD3/SmB/SmD1/SmD2). Entretanto, o sítio Sm do snRNA U1 deste organismo difere nas posições 2 e 7, onde uma adenina é substituída por uma citosina na posição 2 e uma guanina é substituída por uma uracila na posição 7 (Figura 2.7-D). Os resíduos de aminoácidos de SmE que interagem diretamente com a posição 2 (A126) do snRNA U1 de *H. sapiens* por meio de ligações de hidrogênio e interação hidrofóbica são conservados em *T. brucei*, enquanto que os resíduos correspondentes nas proteínas SmB e SmD1 que interagem na posição 7 (G130 no snRNA U1 de *H. sapiens*) variam, porém, mantendo a capacidade de realizar ligação de hidrogênio e interação do tipo  $\pi$ - $\pi$  (Figura 2.7-A).

Observa-se, também, que resíduos de aminoácidos de SmD1, SmD3 e SmG em posição para interação com snRNA U1 não são conservados (coloridos em vermelho na Figura 2.7-A), sendo que apenas a variação SmD3-N40 (*H. sapiens*) para SmD3-G41 (*T. brucei*) corresponde à perda de capacidade de realização de ligação de hidrogênio equivalente. Esta interação é importante para a estabilização da posição 4 (U128 no snRNA U1 de *H. sapiens*), porém outros dois resíduos de SmD3 de *T. brucei* (C65 e N39) podem atuar como doadores/receptores de ligação de hidrogênio nesta posição (Figura 2.7-A).

Já no caso do núcleo de snRNP U4 de *T. brucei* (Figura 2.7-B), há a substituição de SmD3 por SSm4. Os snRNAs U4 de *T. brucei* e *H. sapiens* variam nas posições 2, 6, 7 e 8 (A para G em 2, U para G em 6, U para C em 7 e G para U em 8, em *H. sapiens* e *T. brucei*, respectivamente – Figura 2.7-B). A comparação dos sítios Sm revela que todos os resíduos de aminoácidos que interagem nas posições 2, 7 e 8 do snRNA são conservados, enquanto que na posição 6 os resíduos de aminoácidos conservam suas características físico-químicas. A única variação evidente corresponde ao resíduo SSm4-F83 que substitui SmD3-N38 de *H. sapiens*, perdendo a capacidade de realizar ligação de hidrogênio na posição, porém ganhando uma possível interação hidrofóbica. A proteína SmD3 de *T. brucei*, por outro lado, não conserva os resíduos responsáveis pela interação com snRNA U4 nestas posições (Figura 2.7-B).

Apesar de não se conhecer a estrutura tridimensional do sítio Sm da snRNP U2, é possível comparar as sequências de snRNAs e seus sítios de interação cuja estrutura é conhecida para se avaliar a variação do núcleo Sm de snRNP U2 de *T. brucei* em nível de estrutura primária. Uma comparação com o sítio Sm no snRNA U2 de *H. sapiens* revela variação nas posições 1, 2, 4, 8 e 9 (G para A em 1, A para C em 2, U para G em 4, G para C em 8 e 9, para *H. sapiens* e *T. brucei*, respectivamente) (Figura 2.7-D).

No caso do núcleo da snRNP U2 de *T. brucei*, SmD3 é substituída por Sm16,5K e SmB é substituída Sm15K. Diferentemente de snRNP U4 de *T. brucei*, a variação de proteínas Sm

em snRNP U2 ocorre em posição correspondente à modificação de nucleotídeo no sítio Sm do snRNA U2, mais especificamente na posição 4. A variação de uma uracila para uma guanina nesta posição leva à possibilidade da realização de uma ligação de hidrogênio extra com o snRNA. Curiosamente, a maioria dos resíduos de aminoácidos de Sm15K e Sm16,5K não são conservados em relação às posições equivalentes em SmB e SmD3 de *H. sapiens*, como pode-se notar nas variações: SmD3-C65 para Sm16,5K-K67, SmD3-N39 para Sm16,5K-I40, SmB-D32 para Sm15K-S36, SmB-F34 para Sm15K-T38 e smB-K33 para Sm15K-P37 (coloridos em vermelho na Figura 2.7-C).

Informações em nível atômico do sítio Sm de snRNP U2 de *T. brucei* podem revelar a especificidade da interação com snRNA U2 em detrimento do complexo canônico, indicando possíveis pontos para planejamento racional de fármaco contra este tripanossomatídeo. Adicionalmente, a interação do núcleo Sm da snRNP U2 com U2A'/U2B'' também se mostra relevante, podendo se tratar de um ponto de discrepância entre os complexos de snRNP U2 de *H. sapiens* e *T. brucei*.

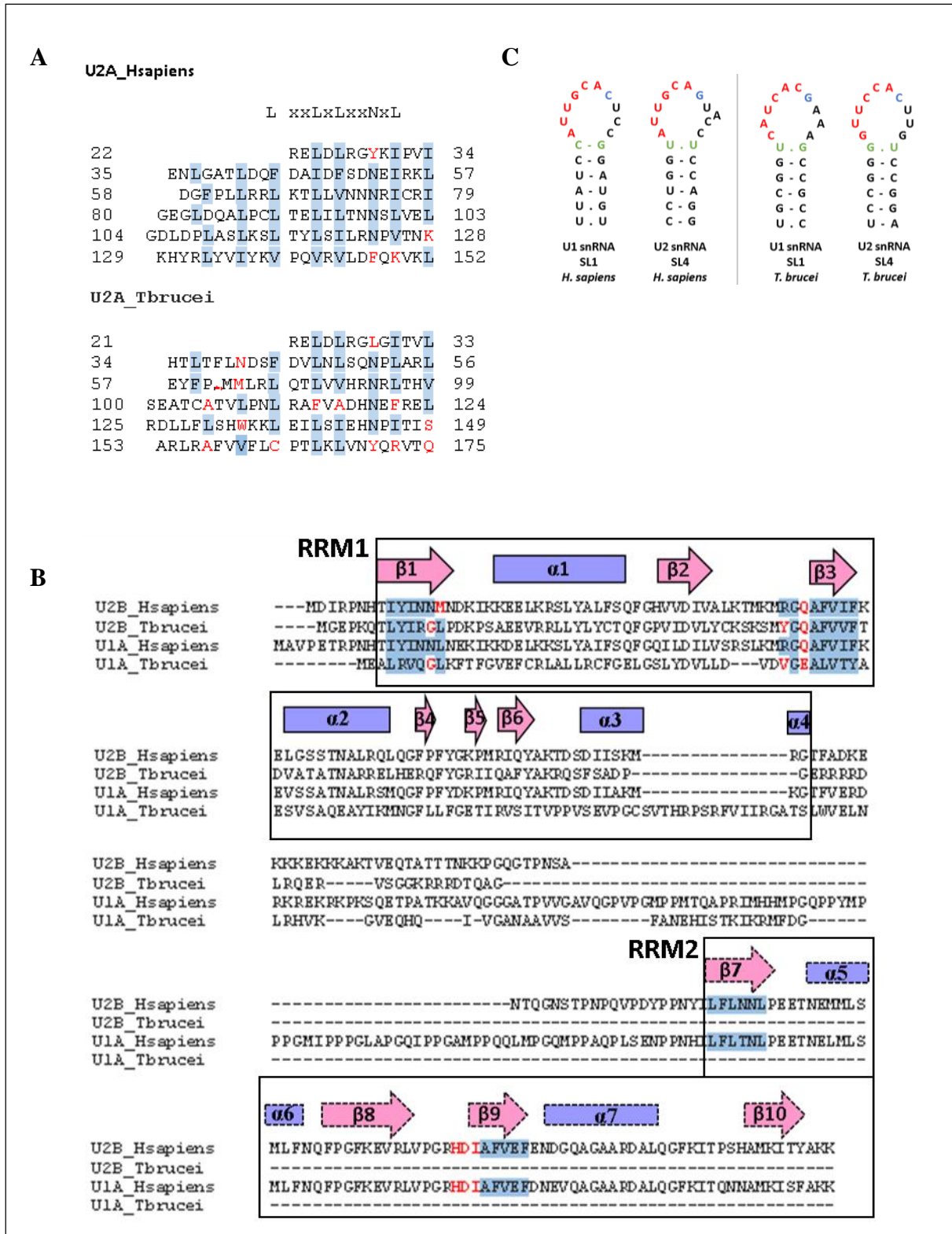
#### 2.4.1.2 Conservação de aminoácidos das proteínas U2A' e U2B'' e análise do sítio de interação com snRNA U2

Os motivos LRR de U2A' de *T. brucei* são bastante degenerados, como se pode observar na Figura 2.8-A, em comparação com os mesmos da ortóloga de *H. sapiens*. Já no caso de U2B'', apenas o primeiro domínio RRM é conservado em *T. brucei* (Figura 2.8-B). Além disso, não há informação funcional a respeito do segundo domínio RRM de *H. sapiens*.

Por outro lado, o alinhamento da sequência de aminoácidos de U2B'' de *T. brucei* e de *H. sapiens* com a proteína U1A dos mesmos organismos mostra que ambas conservam domínios RRM (Figura 2.8-B). Como no caso de U2B'' de *T. brucei*, a proteína U1A deste organismo também apresenta apenas o primeiro domínio RRM. Ambas interagem especificamente com uma alça no snRNA correspondente (SL4 no snRNA U2 e SL2 no snRNA U1), mostrados na Figura 2.8-C, e a ligação é sequência-específica.

A Figura 2.8-C mostra que os seis primeiros nucleotídeos da alça SL4 do snRNA U2 de *H. sapiens* são idênticos aos correspondentes no snRNA U1 de *H. sapiens*. Entretanto, a análise das alças SL1 e SL4 dos snRNAs U1 e U2 de *T. brucei*, respectivamente, mostra que ambas as metades são diferentes entre si. Uma comparação da estrutura cristalográfica do complexo U1A/snRNA U1-SL1 humano<sup>55</sup> (código PDB 4PJO) com informações em nível atômico da estrutura de U2B''/snRNA U2-SL4<sup>190</sup> (código PDB 1A9N) do mesmo organismo revela que a

especificidade na interação ocorre na segunda metade da alça, região onde interagem os resíduos do final da fita-β2 que variam entre U2B'' e U1A de *H. sapiens*.<sup>55,190</sup> Além disso, os resíduos que estão próximos ao par de bases que fecha a alça (M14 e E16) em U2B'' são substituídos por de L e E na proteína U1A de *H. sapiens* e *T. brucei*.



**Figura 2.8** – Alinhamento múltiplo de seqüências de aminoácidos das proteínas U2A' e U2B'' de *H. sapiens* e *T. brucei*. A- U2A', B- U2B''. Em azul estão evidenciados os resíduos de aminoácidos conservados característicos do motivo que determina a proteína. Em vermelho, são mostrados os resíduos não conservados nas posições determinantes dos motivos característicos. Estrutura secundária: hélice- $\alpha$  em azul e fita- $\beta$  em rosa. C- Comparação entre alças SL1 e SL2 de snRNAs U1 e U2, respectivamente. Em vermelho, são mostrados os nucleotídeos tipicamente conservados na primeira metade da alça de cada snRNA. Em azul, o primeiro nucleotídeo da segunda metade e, em verde, o par de nucleotídeos que fecha a alça.  
 Fonte: Elaborada pelo autor a partir de resultados do servidor *Clustal Omega*<sup>194</sup>.

#### 2.4.2 Clonagem, expressão e purificação das proteínas Sm

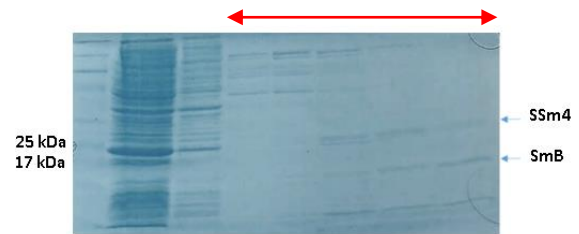
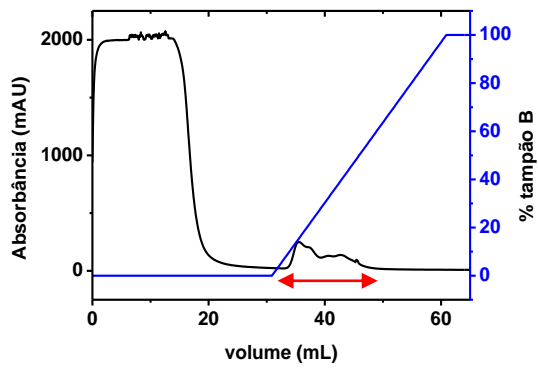
Estudos bioquímicos e estruturais de proteínas Sm necessitam de amostras puras de proteínas recombinantes. Portanto, as fases abertas de leitura dos genes que codificam para as proteínas SmB e SSm4 foram amplificadas, clonadas em vetor pJET1.2/*blunt* (*Thermoscientific*), e confirmadas por PCR e sequenciamento no laboratório de Biofísica do Instituto de Física de São Carlos – Universidade de São Paulo, e subclonadas em vetor de expressão pETDuet1 (*Life Technology*). A construção gênica foi transformada em *E. coli* BL21 rosetta1 competente e, após testes de expressão variando-se temperatura entre 20°C e 37°C e de concentração de indutor IPTG (0,1, 0,5 e 1,0 mM), observou-se maior rendimento solúvel para 12 horas de expressão a 20°C com 0,5 mM IPTG em meio de cultura LB (resultados não mostrados).

O teste de expressão descrito anteriormente também foi empregado para a expressão dos subcomplexos SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB, Sm15K/Sm16,5K, cujas construções gênicas foram transformadas em *E. coli* SG13009 competente, e o maior rendimento de expressão solúvel também ocorre para 12 horas de expressão a 20°C com 0,5 mM IPTG. Em seguida à coexpressão dos subcomplexos Sm, o conteúdo solúvel de cada lisado celular foi copurificado por cromatografia de afinidade e o resultado é mostrado na Figura 2.9-A-E. Apenas uma das proteínas de cada subcomplexo é produzida com extensão de 6 histidinas e a(s) proteína(s) parceira(s) de interação é(são) copurificada(s).

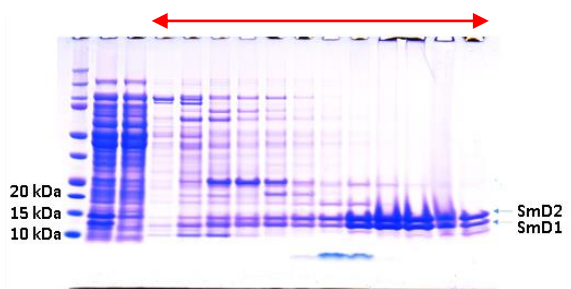
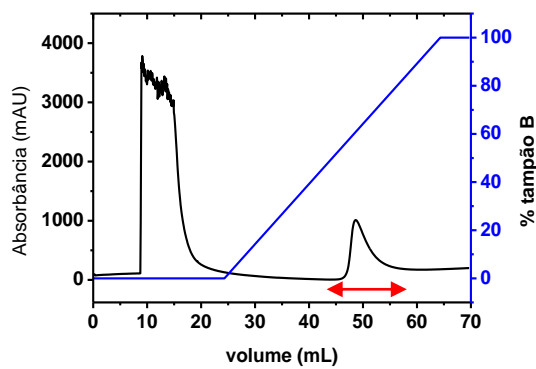
Após clivagem da cauda de 6 histidinas com TEV protease, exceto para o subcomplexo SSm4/SmB que não possui sítio de clivagem, os subcomplexos foram purificados por cromatografia de exclusão molecular (Figura 2.9F). Os gráficos de calibração da coluna SUPEROSE 12 10/300 (*GE*) são mostrados na Figura 2.9-H-I. As amostras puras foram analisadas em SDS-PAGE 20% ou NuPAGE 4-20% Bis-Tris (*Thermoscientific*), corados com *comassie brilliant blue* 0,010% e descorados com ácido acético 10% (Figura 2.9-G).

**A**

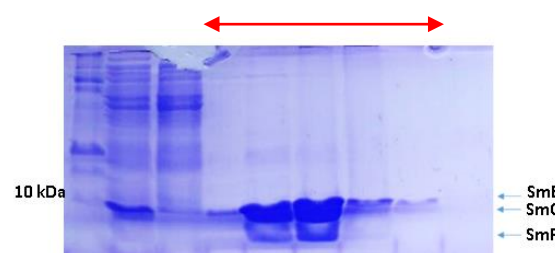
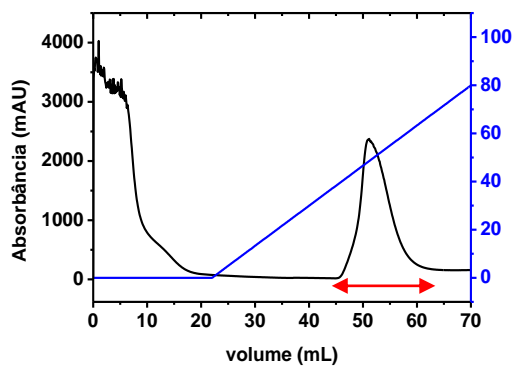
SSm4/SmB

**B**

SmD1/SmD2

**C**

SmE/SmF/SmG

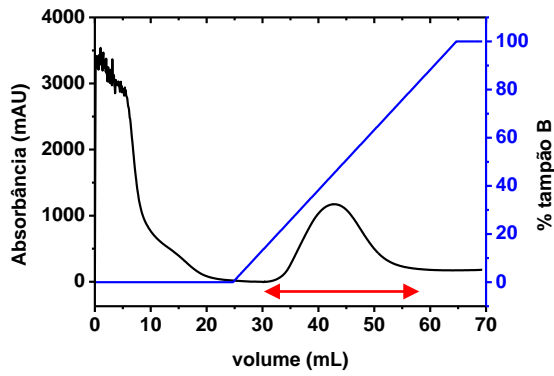


continua

continuação

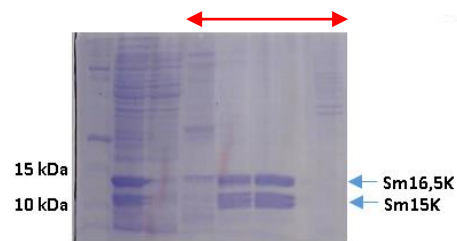
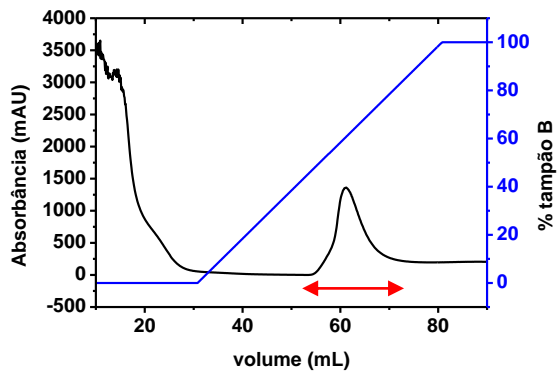
**D**

SmD3/SmB

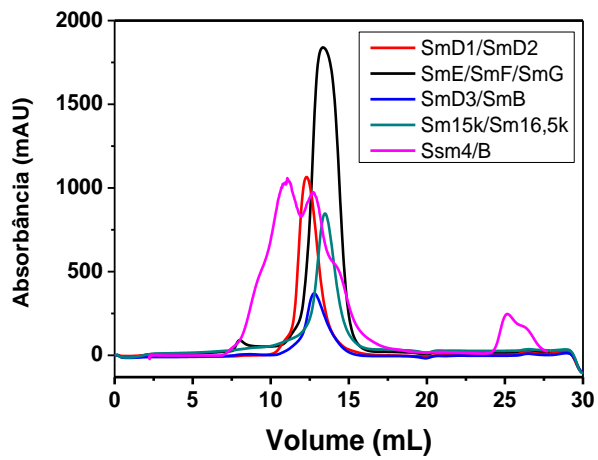


**E**

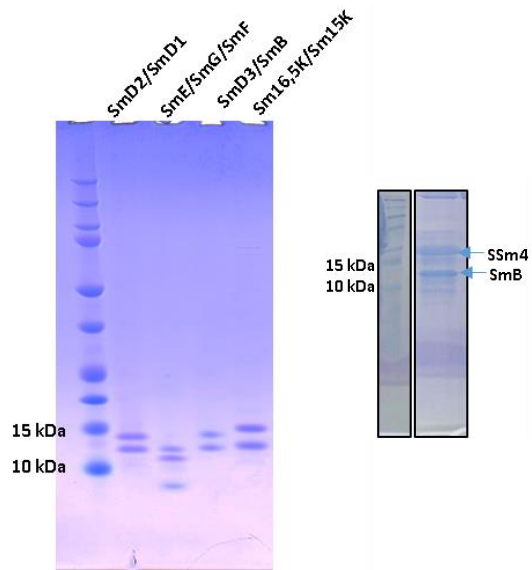
Sm15K/Sm16,5K



**F**

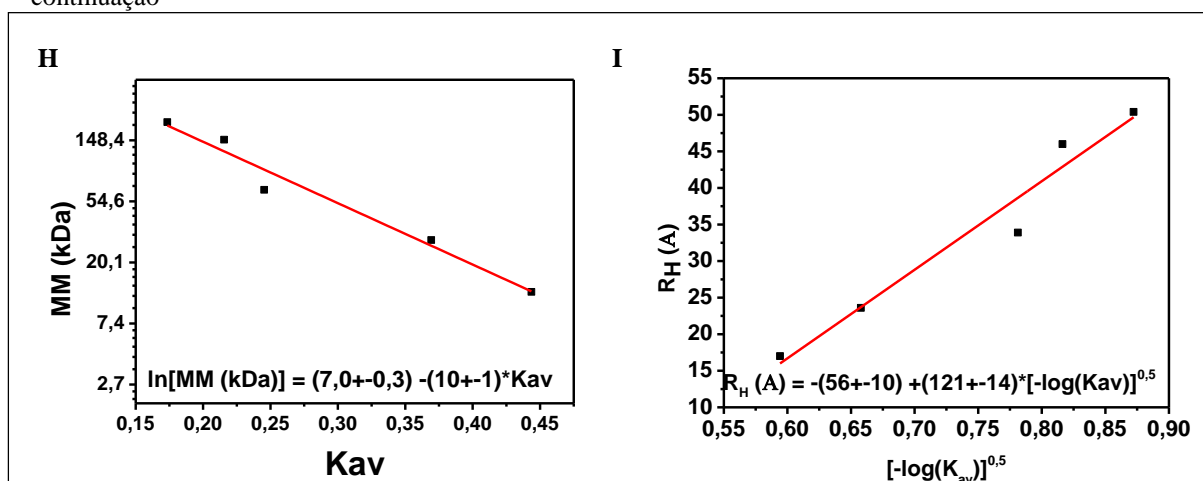


**G**



continua

continuação



**Figura 2.9 – Perfis cromatográficos de purificação de proteínas Sm de *T. brucei*.** Cromatografia de Afinidade: A – SSm4/SmB, B- SmD1/SmD2, C- SmE/SmF/SmG, D- SmD3/SmB, E- Sm15K/Sm16,5K. F-B- Cromatografia de exclusão molecular dos sub-complexos Sm de *T. brucei*. C-G- SDS-PAGE 15%/Tricina dos subcomplexos SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB, Sm15K/Sm16,5K, SSm4/SmB de *T. brucei* após última etapa de purificação. H- e I- Calibração da coluna SUPEROSE 12 10/300 (GE) para massa molecular e raio hidrodinâmico usando os seguintes padrões: citocromo c equino (MM = 12,4 kDa e  $R_H = 17,0 \text{ \AA}$ ), anidrase carbônica bovina (MM = 29,0 kDa e  $R_H = 23,6 \text{ \AA}$ ), BSA (MM = 66,0 kDa  $R_H = 33,9 \text{ \AA}$ ) e álcool desidrogenase de levedura (MM = 150,0 kDa e  $R_H = 46,0 \text{ \AA}$ ). O dextran azul é eluído em volume igual a 7,9 mL, o que permitiu calcular  $K_{av} = (V - V_{dextran}) / (24 \text{ mL} - V_{dextran})$ . Em todos os géis, a primeira coluna corresponde ao marcador de massa molecular, a segunda ao lisado, e da terceira em diante a lavagem e o eluído no gradiente de tampão B é indicado pela seta dupla vermelha. Fonte: Elaborada pelo autor.

Nota-se, na Figura 2.9-A-E, que os diferentes subcomplexos apresentam diferente afinidade pela coluna que contém íons  $\text{Ni}^+$  imobilizados, sendo que o subcomplexo SSm4/SmB é o que apresenta menor afinidade, já que é eluído em menor concentração de tampão B. Um pico único relativamente alargado é observado para as purificações de SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB e Sm15K/Sm16,5K, enquanto que, para a purificação de SSm4/SmB, observam-se dois picos, sendo que somente o segundo pico corresponde ao heterodímero de interesse.

A massa molecular experimental e o raio hidrodinâmico calculados a partir de curva de calibração da coluna SUPEROSE 12 10/300 (GE) são iguais a  $(75 \pm 30) \text{ kDa}$  e  $(34 \pm 14) \text{ \AA}$  para SmD1/SmD2,  $(36 \pm 12) \text{ kDa}$  e  $(26 \pm 11) \text{ \AA}$  para SmE/SmF/SmG,  $(54 \pm 20) \text{ kDa}$  e  $(31 \pm 14) \text{ \AA}$  para SmD3/SmB,  $(33 \pm 3) \text{ kDa}$  e  $(26 \pm 11) \text{ \AA}$  para Sm15K/Sm16,5K e  $(56 \pm 6) \text{ kDa}$  e  $(31 \pm 14) \text{ \AA}$  para SSm4/SmB, correspondendo, respectivamente, às seguintes combinações de oligômeros:  $(\text{SmD1-SmD2})_3$ ,  $(\text{SmE-SmF-SmG})_1$ ,  $(\text{SmD3-SmB})_2$ ,  $(\text{Sm15k-Sm16,5k})_1$ ,  $(\text{SSm4-SmB})_2$ . Curiosamente, o subcomplexo humano SmE/SmF/SmG foi previamente caracterizado como hexamérico,<sup>176</sup> diferentemente do heterotrímero observado para o subcomplexo recombinante de *T. brucei*. Na unidade assimétrica cristalina de SmD1/SmD2<sup>15</sup> e

SmD3/SmB<sup>15</sup> humanas foram identificados heterodímeros, também em contraste ao que se foi observado para as proteínas Sm de *T. brucei*. Deve-se notar que o erro associado a cada medida apresentada provém da propagação do desvio ao ajuste linear dos dados.

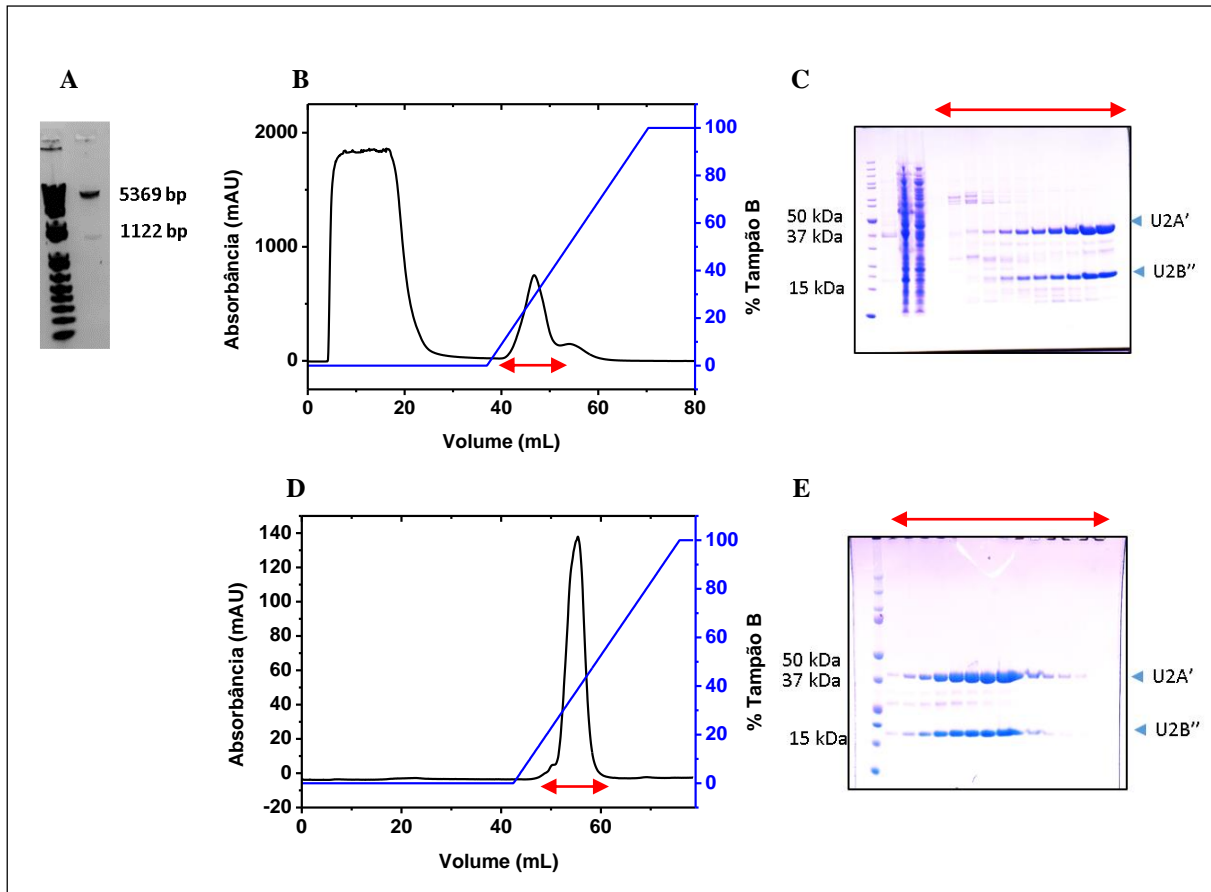
Acredita-se que a formação desses oligômeros de ordem superior de proteínas Sm recombinantes não esteja necessariamente relacionada à função dessas proteínas e/ou à associação dos complexos Sm de *T. brucei in vivo*. Como a interação entre proteínas Sm ocorre por meio da formação de uma folha- $\beta$ , é possível que a oligomerização observada seja devido à alta concentração dessas interfaces expostas, o que não deve acontecer no ambiente celular.

Todas as amostras puras de proteínas Sm apresentaram razão de absorvância de luz ultravioleta a 260 nm e a 280 nm próxima a 0,6, o que indica ausência de contaminação com nucleotídeos/ácidos nucleicos endógenos. A quantidade total, em massa, para 1L de cultura de cada subcomplexo após purificação, foi de 30,0 mg para SmD1/SmD2, 6,4 mg para SmE/SmF/SmG, 120,0 mg para SmD3/SmB, 4,2 mg para Sm16,5K/Sm15K e 4,0 mg para SSm4/SmB de conteúdo total de proteína pura.

### 2.4.3 Clonagem, expressão e purificação de U2A'/U2B''

As fases abertas de leitura de U2A'/U2B'' foram clonadas com sucesso e as sequências de DNA de U2A'/U2B'' foram confirmadas por sequenciamento. As proteínas recombinantes foram expressas e copurificadas, primeiramente a partir de cromatografia de afinidade (Figura 2.10). A amostra pura apresentou razão de absorvância a 260 nm e a 280 nm igual a 1,23, o que indica que estava contaminada com ácido nucleico endógeno, provavelmente por interação inespecífica com o domínio RRM de U2B''. Foi necessária, portanto, outra etapa de purificação, em coluna de hidroxiapatita, para eliminação da contaminação (Figura 2.10). A mesma razão medida para a amostra resultante é igual a 0,55, o que aponta para uma amostra final pura.





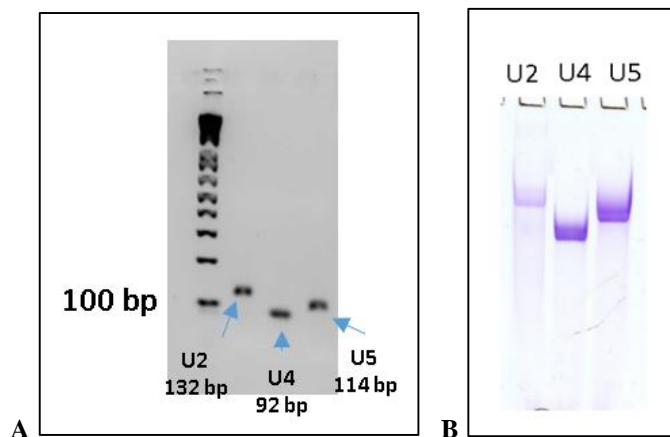
**Figura 2.10 – U2A'/U2B''.** A- Confirmação de clonagem em vetor pET28a por meio de análise de produto de restrição: o fragmento de 1122 pb corresponde à sequência que codifica para U2A'/U2B''. B- Cromatografia de afinidade a íons Ni<sup>+</sup> do subcomplexo U2A'/U2B'' recombinante. Proteínas foram eluídas entre 15 e 30% de tampão B, como observado em SDS-PAGE (C). D- Perfil de eluição de U2A'/U2B'' em resina de hidroxiapatita e SDS-PAGE (E) correspondente. As setas duplas indicam os produtos eluídos em cada caso.

Fonte: Elaborada pelo autor.

O rendimento de purificação de U2A'/U2B'' recombinante a partir de 1 L cultura de *E. coli* foi igual a 10 mg no total.

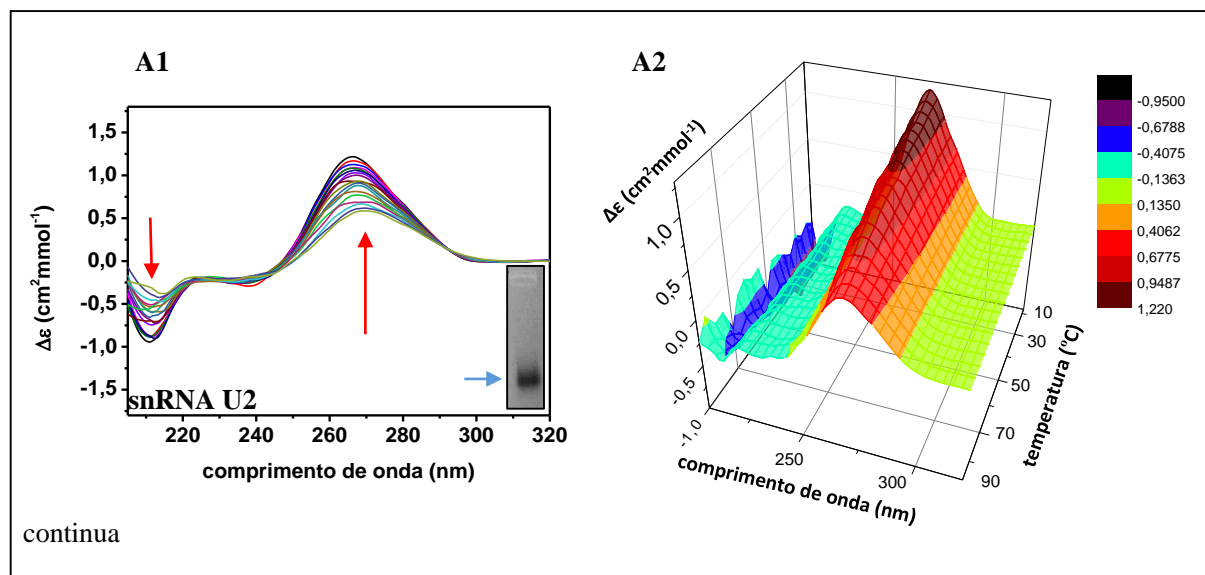
#### 2.4.4 Preparo dos snRNAs U2, U4 e U5 e sua caracterização por meio de espectroscopia de dicroísmo circular

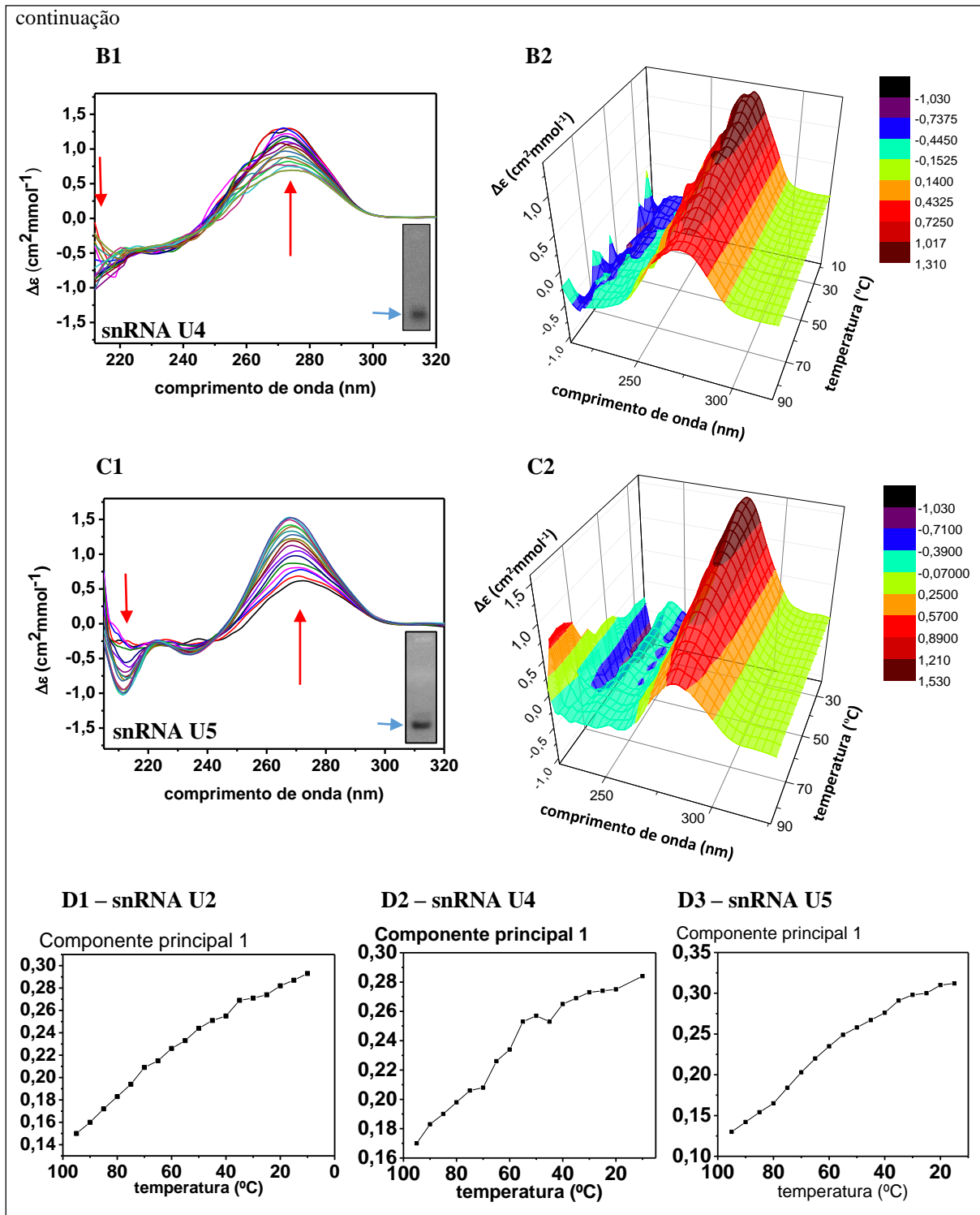
Moldes de DNA para a transcrição *in vitro* dos snRNAs U2, U4 e U5 foram obtidos por PCR e clonados em vetor pUC18 (*Thermoscientific*) entre os sítios de *Bam*HI e *Hind*III, como se observa na Figura 2.11-A. A transcrição *in vitro* pela T7 RNA polimerase foi confirmada por eletroforese em condição desnaturante (Figura 2.11-B):



**Figura 2.11** – Transcrição *in vitro* de snRNAs U2, U4 e U5. A- Produtos de PCR dos moldes de snRNAs U2, U4 e U5 confirmando a clonagem em pUC18. B- Bandas únicas para snRNAs U2, U4 e U5 obtidos por transcrição *in vitro* observadas em eletroforese desnaturante (PAGE 15%/uréia).  
Fonte: Elaborada pelo autor.

Eletroforese em condição não desnaturante de 20  $\mu\text{L}$  de amostra de snRNA U2, U4 e U5 a 2  $\mu\text{M}$  (máxima concentração utilizada nos experimentos de reconstrução *in vitro* dos núcleos de snRNPs) mostra banda única predominante em cada caso (Figura 2.12), mesmo em alta concentração. Desta forma, descarta-se a formação de oligômeros nas condições analisadas neste trabalho. Assim, espectroscopia de discoísmo circular foi empregada para caracterizar as o enovelamento térmico das moléculas de RNAs, como mostrado na Figura 2.12:





**Figura 2.12 – Enovelamento de snRNAs U2 (A1, A2 e D1), U4 (B1, B2 e D2) e U5 (C1, C2 e D3) medido por espectroscopia de dicroísmo circular.** São apresentados os espectros de dicroísmo circular medidos entre 210 e 320 nm e gráfico de componente principal 1 *versus* temperatura. As bandas apresentadas como inserção correspondem à eletroforese em gel nativo (1% agarose), corado com brometo de etídeo ( $1\ \mu\text{g mL}^{-1}$ ). As setas vermelhas indicam o aparecimento de mínimo e máximo locais de dicroísmo circular com a diminuição de temperatura. As setas azuis indicam as bandas únicas em eletroforese nativa.

Fonte: Elaborada pelo autor usando o programa *Origin PRO 8.0* (Origin Lab Corporation).

Em temperaturas moderadas, observa-se um máximo intenso para comprimento de onda de aproximadamente 265 nm e um mínimo por volta de 210 nm, o que é indicativo da formação de hélices do tipo A para os três snRNAs de *T. brucei* analisados. O primeiro pico é devido ao pareamento de bases nitrogenadas por interações do tipo  $\pi$ - $\pi$  entre os anéis aromáticos e o segundo é derivado de transições na cadeia principal.<sup>245</sup> À medida que ocorre a diminuição de temperatura, ambos os picos aumentam em intensidade afastando-se do eixo adicrótico. Observa-se, ainda, um pequeno desvio na direção do vermelho para o pico positivo quando a temperatura é alta.

A análise da componente principal 1 de cada um dos conjuntos de espectros em função da temperatura não indica uma transição clara entre dois estados (não enovelado – enovelado). Entretanto, o aumento de intensidade dos picos sugere o enovelamento do RNA na forma de hélice do tipo A, como esperado.

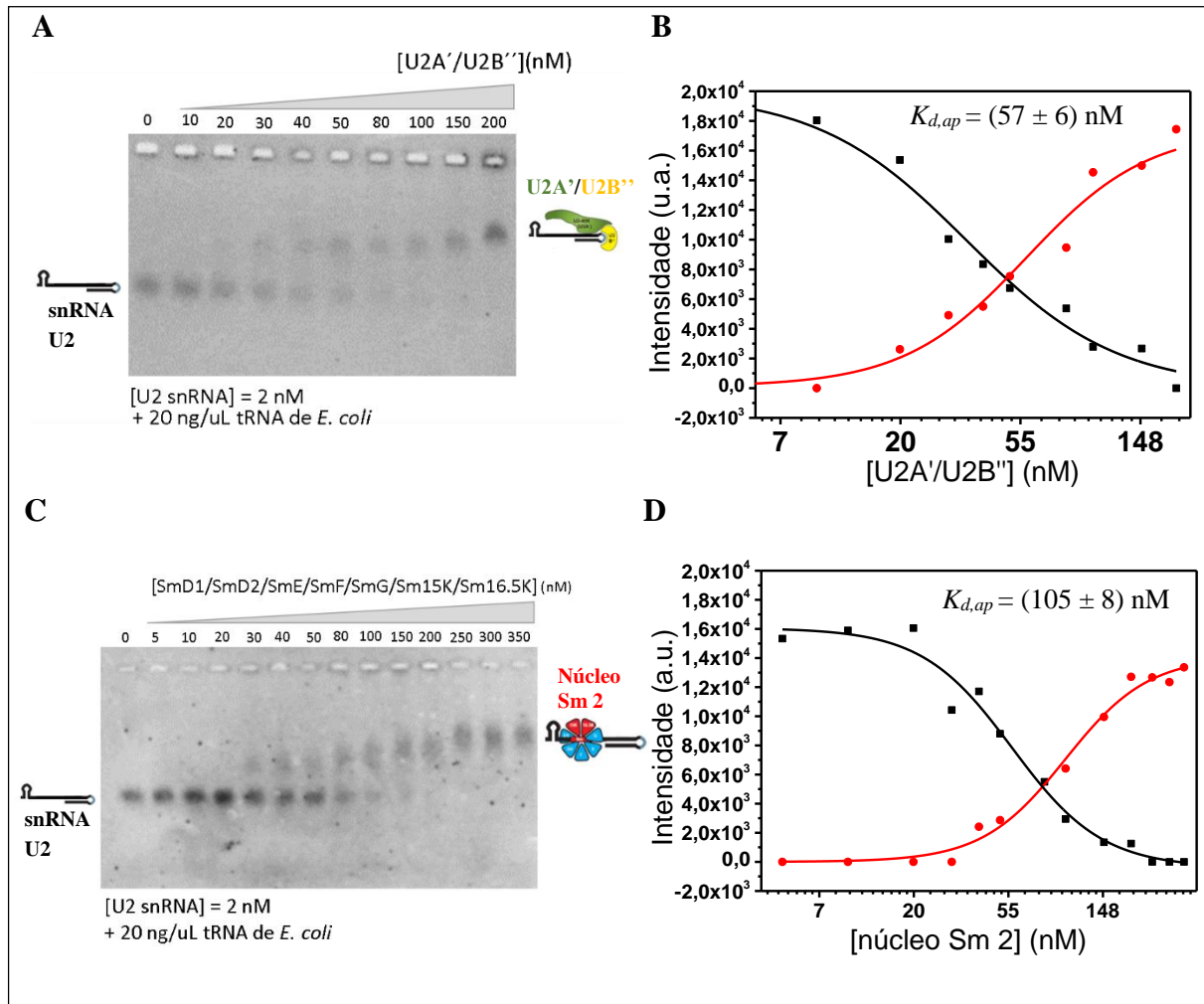
#### 2.4.5 Reconstrução *in vitro* dos núcleos Sm de *T. brucei*

A reconstrução *in vitro* comparativa dos núcleos Sm de *T. brucei* foi possível a partir das amostras puras de snRNAs enovelados, subcomplexos Sm e U2A'/U2B'', como descrito nas seções a seguir.

##### 2.4.5.1 Ensaios de retardo de mobilidade eletroforética (EMSA)

Inicialmente, a interação entre U2A'/U2B'' e a região 3' do snRNA U2 de *T. brucei* (SLII + sítio Sm + SLIV) marcado com fluoresceína foi investigada *in vitro*. A Figura 2.13-A mostra um retardo na mobilidade eletroforética devido à formação do complexo. A intensidade de fluorescência de cada banda foi quantificada e a equação de Hill foi ajustada aos dados experimentais, revelando uma constante de dissociação aparente ( $K_{d,ap}$ ) igual a  $57 \pm 6$  nM e um coeficiente de Hill ( $n$ ) igual a  $1,6 \pm 0,4$  (Figura 2.13-B). Este resultado é comparável àqueles previamente publicados para a interação de U2A'/U2B'' com snRNA U2 de *H. sapiens* (<100 nM).<sup>68-69, 190</sup>

A ligação da mesma construção de snRNA U2 com as proteínas SmD1/SmD2, SmE/SmF/SmG e Sm15K/Sm16,5K, na razão estequiométrica 1:1:1, formando o núcleo Sm 2, foi investigada pela mesma técnica (Figura 2.13-C), e resultou em  $K_{d,ap} = 105 \pm 8$  nM e  $n = 1,9 \pm 0,2$  (Figura 2.13-D). A interação de proteínas Sm de *H. sapiens* com di-sRNA U4/U6, por exemplo, apresenta constante de dissociação da mesma ordem de grandeza.<sup>246</sup>

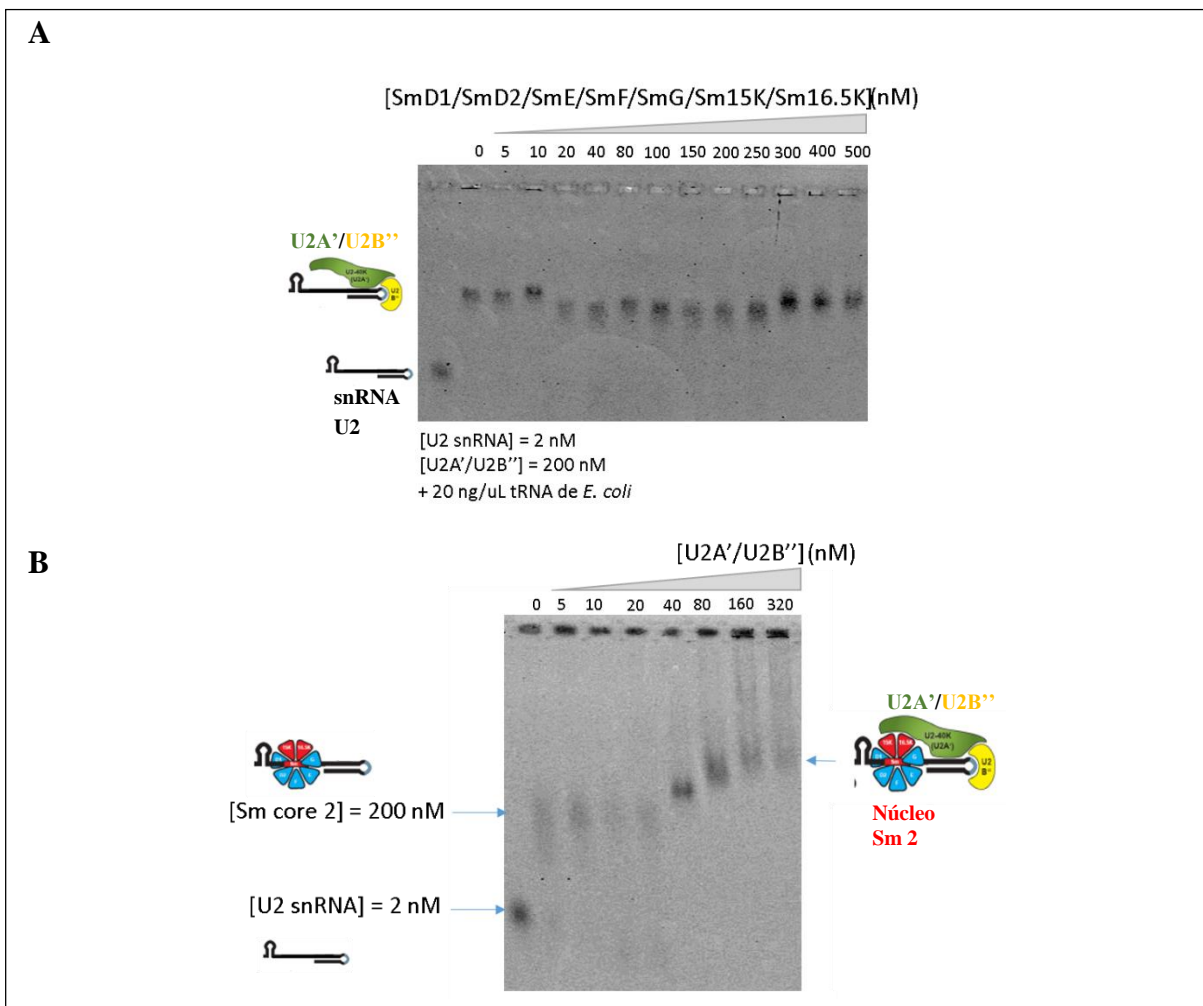


**Figura 2.13 - Ensaios de interação de U2A'/U2B'' e núcleo Sm 2 ao snRNA U2.** A- Titulação de U2A'/U2B'' ao snRNA marcado com fluoresceína e B- ajuste da equação de Hill às medidas de fluorescência das bandas. C) Titulação de núcleo Sm 2 ao snRNA marcado com fluoresceína e D- ajuste da equação de Hill aos dados experimentais.  
Fonte: Elaborada pelo autor.

Após a confirmação da interação *in vitro* do subcomplexo U2A'/U2B'' e do núcleo Sm 2, separadamente, com a região 3' do snRNA U2 de *T. brucei*, decidiu-se analisar se a interação prévia de um destes subcomplexos afeta a afinidade do outro pela ligação com o snRNA U2. Deste modo, as proteínas Sm que formam o núcleo Sm 2 (em razão estequiométrica 1:1:1) foram inicialmente tituladas em amostra de subcomplexo U2A'/U2B''/snRNA U2 (200 nM de U2A'/U2B'' e 2 nM de snRNA U2 marcado com fluoresceína) e não foi observado retardo de mobilidade eletroforética, o que significa ausência de interação nessas condições (Figura 2.14-A).

Por outro lado, um grande retardo de mobilidade eletroforética (do inglês, *supershift*) foi observado quando o subcomplexo U2A'/U2B'' foi titulado em amostra de núcleo Sm 2/snRNA U2 marcado com fluoresceína (Figura 2.14-B). Este resultado sugere a associação

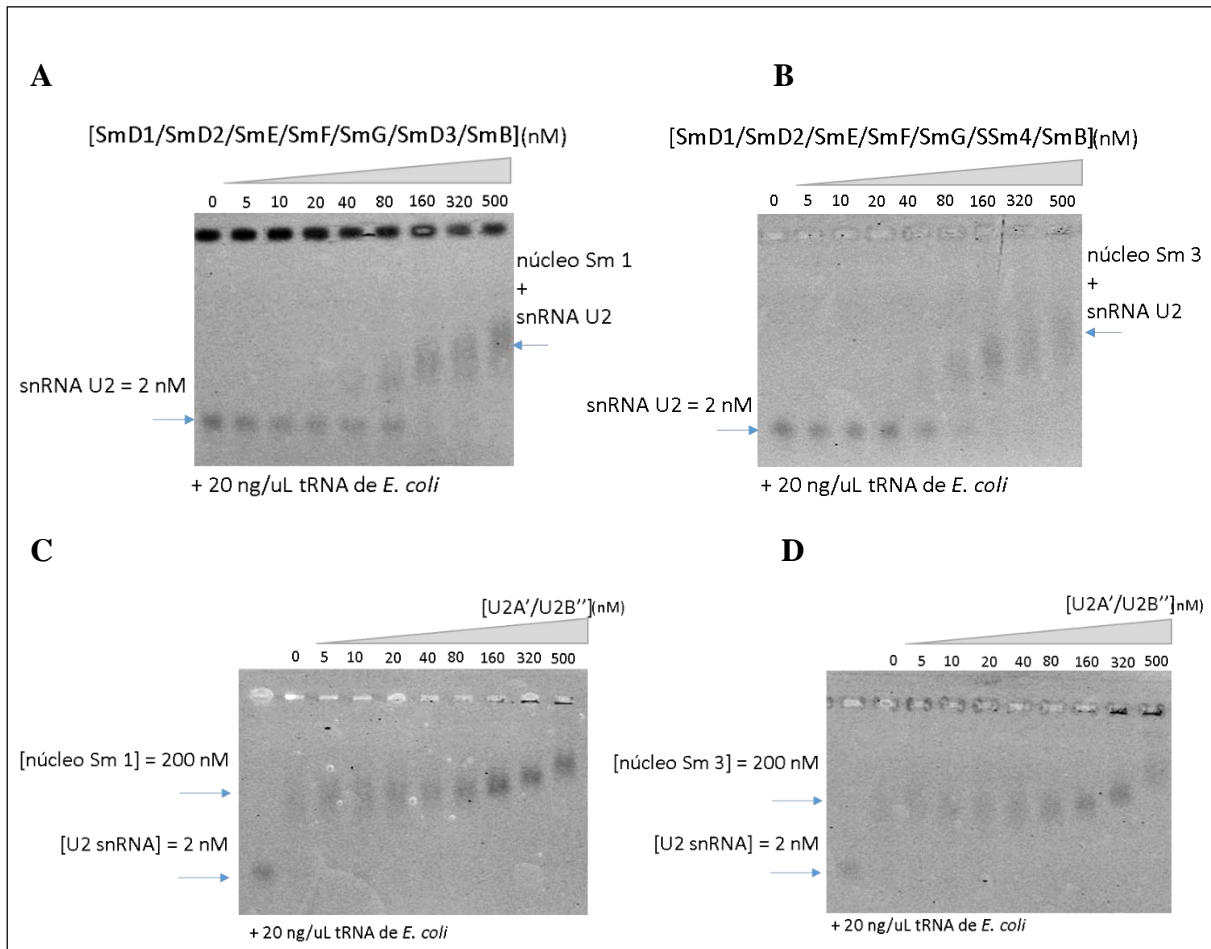
ordenada do núcleo Sm de snRNP U2 de *T. brucei*, onde primeiro as proteínas Sm interagem com o snRNA U2 e, em seguida, U2A'/U2B'' interage com o núcleo Sm pré-formado.



**Figura 2.14 - Formação do núcleo de snRNP U2.** A- Titulação de núcleo Sm 2 ao subcomplexo pré-formado contendo U2A'/U2B'' e snRNA U2. B- Titulação de U2A'/U2B'' ao núcleo Sm 2 pré-formado em interação com snRNA U2.  
Fonte: Elaborada pelo autor.

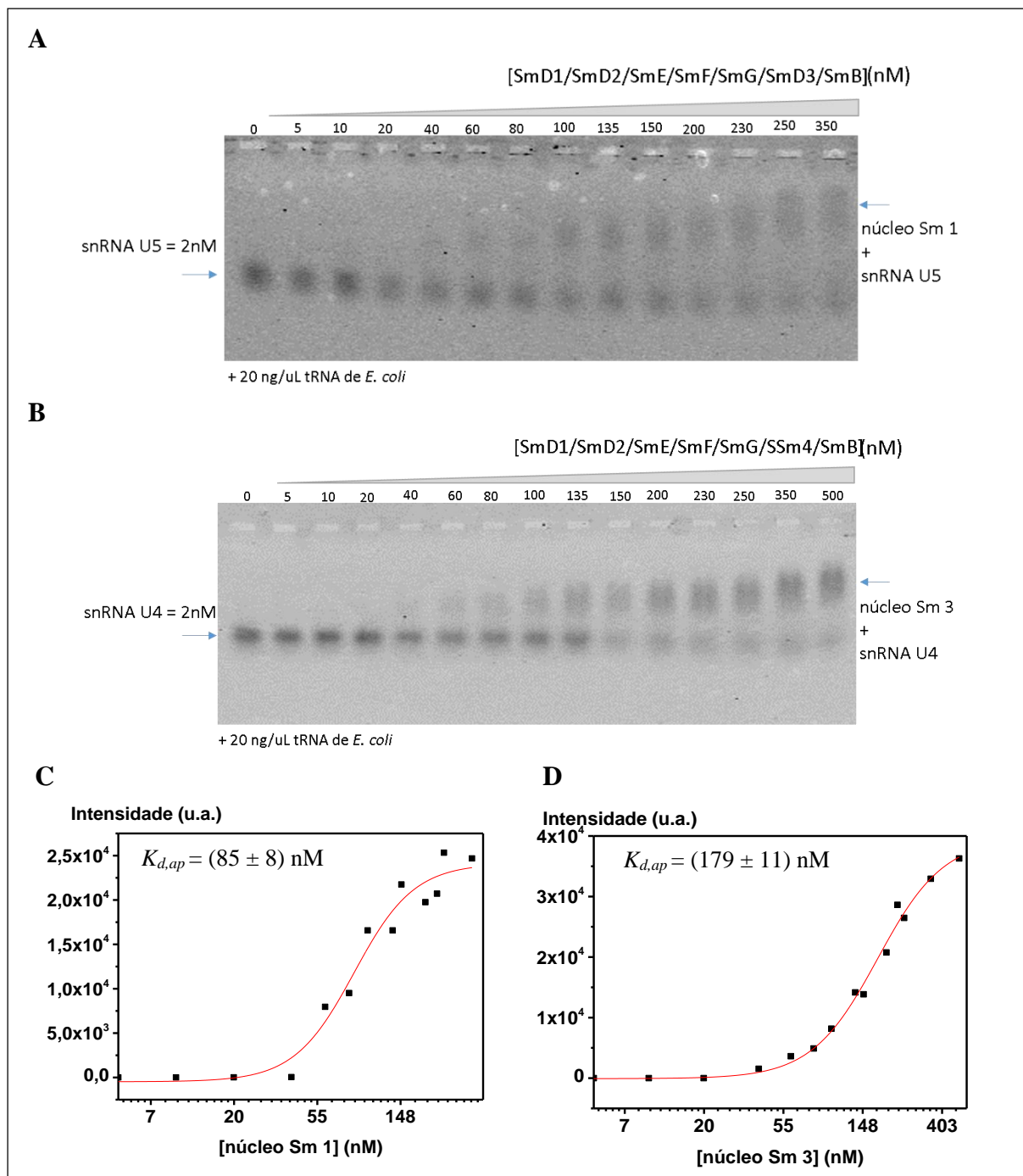
A especificidade da interação de proteínas Sm recombinantes com o snRNA U2 foi analisada a partir da titulação dos outros dois núcleos de proteínas Sm diferentes (núcleo Sm 1 - SmD1/SmD2, SmE/SmF/SmG, SmD3/SmB, e núcleo Sm 3 - SmD1/SmD2, SmE/SmF/SmG, SSm4/SmB) à mesma construção da molécula de snRNA U2 marcada com fluoresceína. Observa-se retardo na mobilidade eletroforética em ambos os casos, o que indica que não há especificidade de interação do snRNA U2 com proteínas Sm recombinantes *in vitro* (Figura 2.15-A-B). Deve-se notar, entretanto, que as bandas correspondentes aos complexos dos núcleos Sm 1 e 3 interagindo com snRNA U2 são difusas, o que indica que a população de complexos formados não é homogênea. Por outro lado, U2A'/U2B'' apresenta maior afinidade

para interação com o núcleo Sm 2 pré-formado do que para com os outros possíveis (Figura 2.15-C-D).



**Figura 2.15 - Ensaio de interação de proteínas Sm com snRNA U2.** Titulação de proteínas dos núcleos A- Sm 1 e B- Sm 3 ao snRNA U2. Titulação de U2A'/U2B'' nos núcleos Sm C-1 e D-2 pré-formados. Fonte: Elaborada pelo autor.

Já no caso da interação dos núcleos Sm 1 e 3 com os respectivos snRNAs U5 e U4 cognatos (Figura 2.16-A-B), a análise de EMSA resultou em  $K_{d,ap} = 85 \pm 8$  nM e  $n = 2,8 \pm 0,7$  para o complexo do núcleo Sm 1 interagindo com snRNA U5 (Figura 2.16-C) e  $K_{d,ap} = 179 \pm 11$  nM e  $n = 2,4 \pm 0,3$  para o complexo do núcleo Sm 3 interagindo com snRNA U4 (Figura 2.16-D). Em ambos os casos, o retardo na mobilidade dos complexos nos géis não chega a ser completo. Isso pode ser devido a um possível envelhecimento parcial da população de moléculas de snRNA em solução. Deste modo, para o cálculo das constantes de interação, considerou-se que todas as moléculas marcadas cujo envelhecimento está correto interagiram com todas as proteínas Sm na mais alta concentração. Devido a essa dificuldade experimental, optou-se por não avaliar a interação dessas moléculas de snRNA com as proteínas Sm de todos os diferentes núcleos Sm.



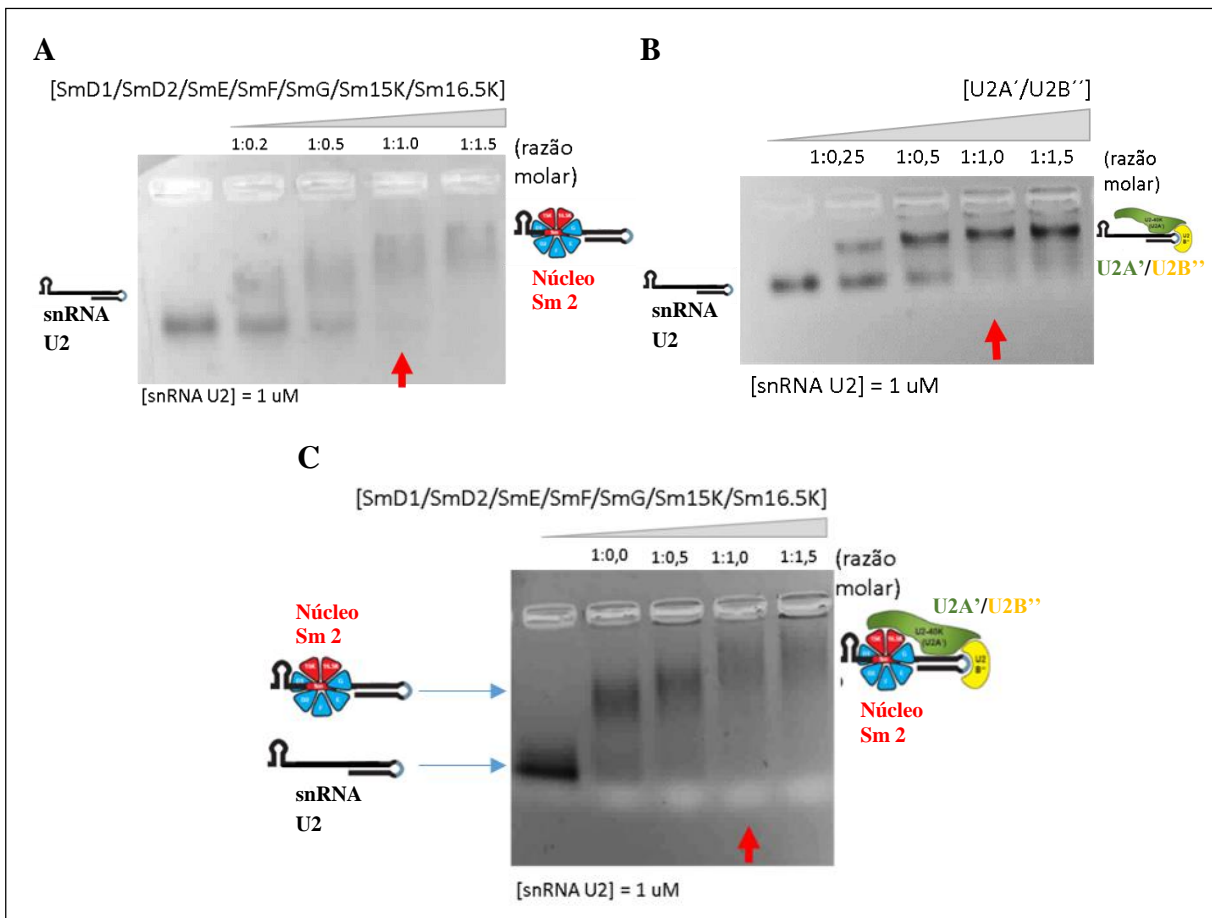
**Figura 2.16 - Formação dos núcleos Sm 1 e 3.** Titulação de proteínas do núcleo A- Sm 1 e B- Sm 3 em snRNAs U5 e U4, respectivamente. Ajuste dos dados experimentais ao modelo de Hill<sup>200</sup> para os núcleos D- Sm 1 e E- Sm 3.

Fonte: Elaborada pelo autor.

Finalmente, a estequiometria de ligação do snRNA U2 para os diferentes subcomplexos de proteínas foi avaliada em altas concentrações (Figura 2.17). Note que o núcleo Sm 2 se liga ao snRNA U2 em razão estequiométrica 1:1 (Figura 2.17-A) e o mesmo é observado para a ligação de U2A'/U2B'' (Figura 2.17-B). O super-deslocamento observado para a interação de



U2A'/U2B'' com o núcleo Sm 2 pré-formado com snRNA U2 indica estequiometria 1:1:1 para a formação dessa estrutura (Figura 2.17-C).

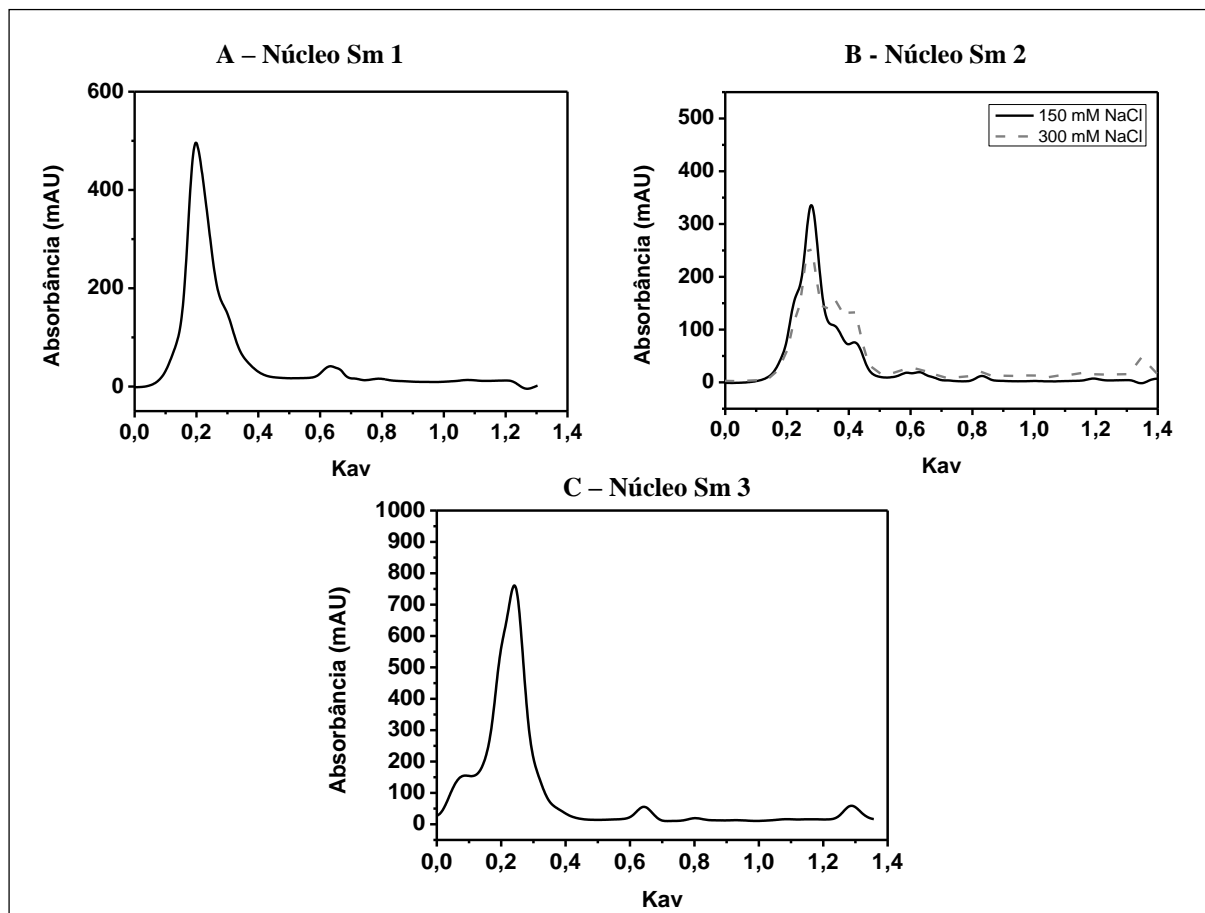


**Figura 2.17-** Estequiometria de interação entre os componentes do núcleo de snRNP U2. A- Titulação de proteínas do núcleo Sm 2 ao snRNA U2. B- Titulação de U2A'/U2B'' ao snRNA U2. C- Titulação de U2A'/U2B'' ao núcleo Sm 2 pré-formado em interação com snRNA U2. As setas vermelhas indicam razão estequiométrica 1:1.

Fonte: Elaborada pelo autor.

#### 2.4.5.2 Ensaio de purificação

Os resultados anteriores mostraram que os núcleos Sm recombinantes se formam na presença dos snRNAs cognatos. A associação *in vitro* dos núcleos de proteínas Sm na ausência de snRNAs foi avaliada por cromatografia de exclusão molecular à partir de amostras reconstituídas *in vitro* na ausência de RNA em tampão de ligação, que contém 100 mM NaCl, e eluídas no mesmo tampão, porém contendo 300 mM NaCl (Figura 2.18-A-C). Apenas no caso do complexo 2 foi observada uma tendência de dissociação do complexo (Figura 2.18-B). Portanto, para este complexo Sm foi necessária a eluição com tampão contendo 150 mM NaCl, para a qual observa-se um pico predominante correspondente ao núcleo Sm 2 (Figura 2.18-B).

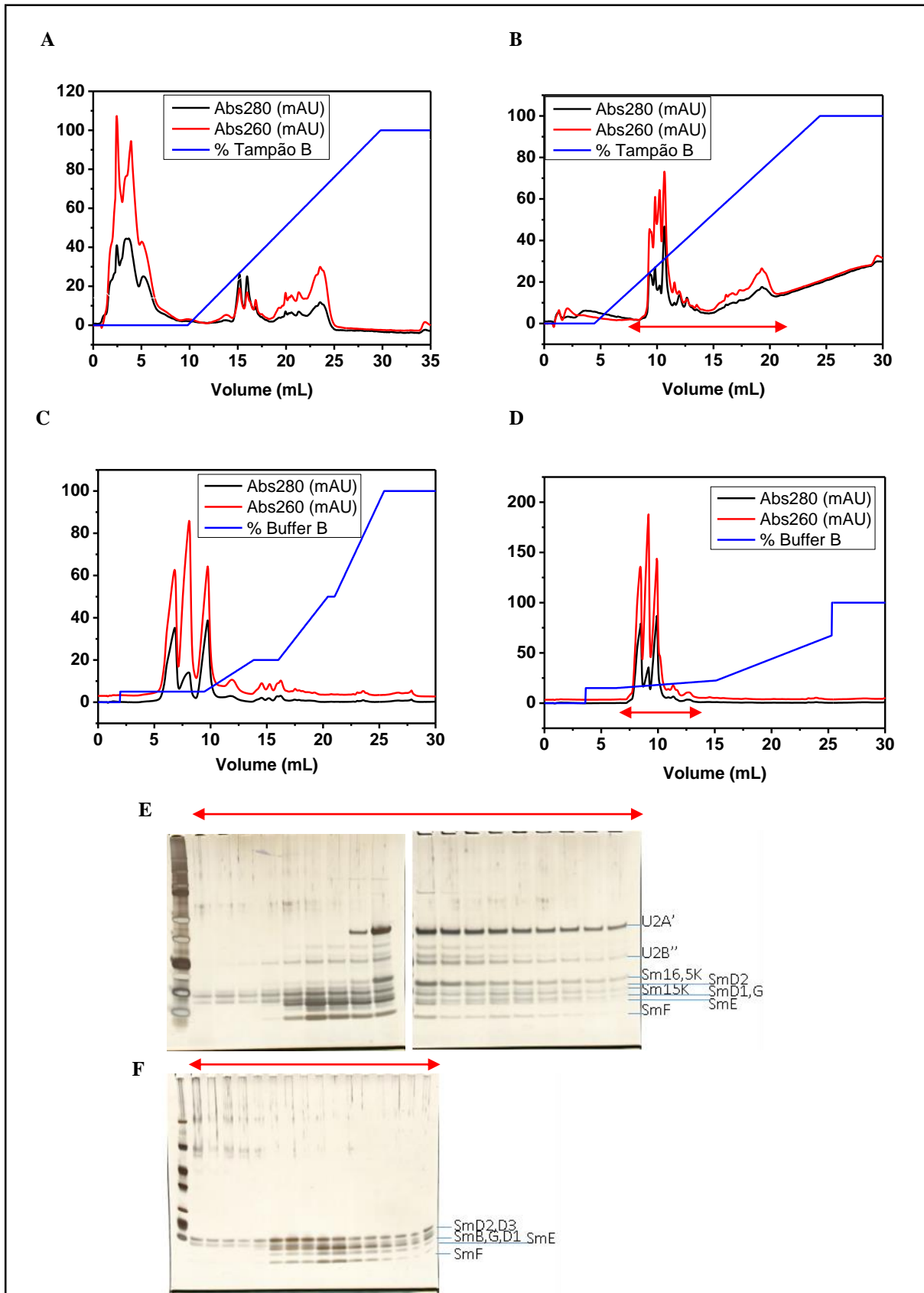


**Figura 2.18** - Cromatografia de exclusão molecular dos núcleos Sm de *T. brucei* na ausência de snRNAs. A- complexo 1 (Sm 1, canônico), B- complexo 2 (Sm 2) e C- complexo 3 (Sm 3). Para cada resultado de purificação, o perfil de eluição isocrática e o SDS-PAGE são mostrados.  
 Fonte: Elaborada pelo autor usando o programa *Origin PRO 8.0* (Origin Lab Corporation).

Nota-se um único pico predominante para o núcleo Sm 1 (canônico) na ausência de snRNA (Figura 2.18-A) com massa molecular e raio hidrodinâmico (calculados a partir das curvas de calibração mostradas na Figura 2.9-H-I, seção 2.4.2) iguais a  $(99 \pm 40)$  kDa e  $(39 \pm 8)$  Å, respectivamente. Já para o núcleo Sm 2, que contém Sm15K/Sm16,5K, também na ausência de snRNA (Figura 2.18-B), o pico majoritário corresponde a uma massa molecular de  $(73 \pm 30)$  kDa e a um raio hidrodinâmico igual a  $(34 \pm 7)$  Å. Finalmente, para o núcleo Sm 3, que contém SSm4/SmB, na ausência de snRNA (Figura 2.18-C), observa-se um pico com máximo correspondendo a uma massa molecular de  $(102 \pm 40)$  kDa e raio hidrodinâmico igual a  $(39 \pm 12)$  Å. Estes resultados serão comparados com aqueles obtidos no experimentos de SAXS, descritos na seção 2.4.8.

Em seguida, dado o sucesso na reconstrução *in vitro* dos núcleos de proteínas Sm em interação com snRNAs avaliado por meio de ensaios de retardo na mobilidade eletroforética, tentativas de purificação dos mesmos foram realizadas. O primeiro teste aplicado foi por meio de cromatografia de troca iônica do núcleo da snRNP U2 de *T. brucei* (snRNA U2 /núcleo Sm

2/U2A'/U2B'') em diferentes condições de tampão, como se observa na Figura 2.19-A-C para experimentos representativos. Adicionalmente, foram realizados os mesmos testes para purificação do núcleo Sm da snRNP U5 (núcleo Sm 1) (Figura 2.18-D):



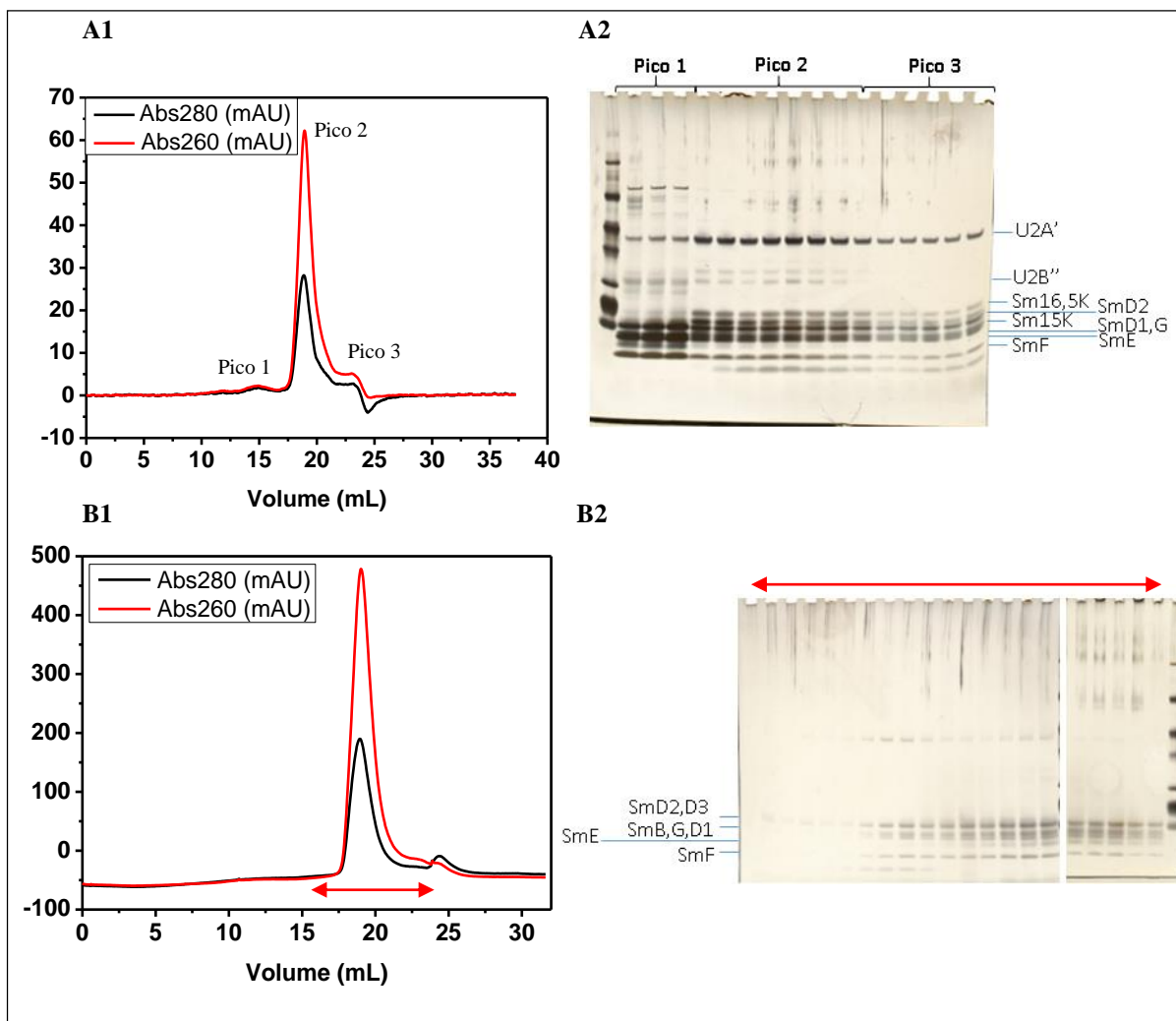
**Figura 2.19 - Ensaios de purificação dos núcleos de snRNPs U2 e U5 por cromatografia de troca iônica.**

Perfis cromatográficos para purificação do núcleo de snRNP U2 em concentrações de sal A- alta (100 mM), B- intermediária (20 mM) e C- ausência de sal no tampão A. D- Perfil cromatográfico de purificação do núcleo snRNP U5. D- SPS-PAGE representativo de purificação do núcleo de snRNP U2 e U5, mostrando a dissociação dos complexos. As setas duplas vermelhas indicam os produtos eluídos em cada caso.

Fonte: Elaborada pelo autor.

Observa-se que, mesmo para condições em que o tampão A não possui sal, o complexo correspondente ao núcleo de snRNP U2 tende a se dissociar durante a eluição. O mesmo acontece para o núcleo Sm de snRNP U5 (núcleo Sm 1). Visto que não foi possível purificar com sucesso os núcleos Sm 1 e 2, não foram realizados experimentos de purificação por troca iônica do núcleo Sm 3, cuja constante de dissociação do complexo de proteínas Sm com o snRNA U4 é maior.

Optou-se pela purificação em condições físico-químicas mais brandas em cromatografia de exclusão molecular com eluição isocrática dos núcleos Sm 1 e 2 em interação com snRNA cognato. Neste caso, picos coincidentes de absorvância de luz UV em comprimento de onda igual a 280 nm e 260 nm foram observados para ambos os complexos (Figura 2.20-A1-A2) e SDS-PAGEs confirmaram as proteínas Sm (Figura 2.20-B1-B2).



**Figura 2.20 - Cromatografia de exclusão molecular.** Padrão de eluição isocrática e SDS-PAGE para A1 e A2- núcleo de snRNP U2 e B1 e B2- núcleo Sm de snRNP U5. As setas duplas vermelhas indicam os produtos eluídos em cada caso. As setas duplas vermelhas indicam os produtos de eluição.  
Fonte: Elaborada pelo autor.

Curiosamente, as proteínas são eluídas tardiamente em ambos os casos, o que indica que a coluna não é inerte às moléculas nas condições dos experimentos. Portanto, não foi possível calcular a massa molecular dos complexos contendo RNA a partir destes dados experimentais.




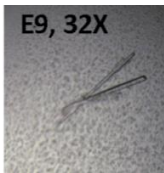
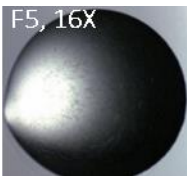
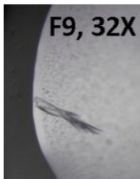

Não houve sucesso nas tentativas de concentração das amostras resultantes das purificações. Quando se aplicou ultrafiltração, notou-se que a medida de absorvância de luz UV a 280 nm e a 260 nm para o eluato e para a amostra retida no filtro corresponde à situação em que as moléculas de snRNA passam direto pelo filtro e as moléculas de proteínas são retidas. Por outro lado, quando se aplicou ultracentrifugação, não houve formação de sedimento, mesmo em diferentes concentrações de glicerol (de 0% a 10%).

## 2.4.6 Ensaios de cristalização, análise de dados de difração de raios-X e resolução da estrutura tridimensional de U2A'/U2B''

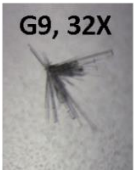
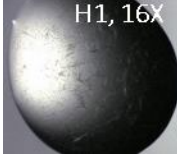



### 2.4.6.1 Ensaios de cristalização

Apesar do insucesso na purificação e concentração dos complexos Sm de *T. brucei* em interação com os snRNAs cognatos, optou-se por dar continuidade ao projeto com experimentos de cristalização de amostras obtidas por reconstrução *in vitro*. Inicialmente, foram preparadas reações de reconstituição *in vitro* dos núcleos Sm de snRNPs U2 (Sm 2) e U5 (Sm1), como descrito previamente, nas concentrações de 20  $\mu\text{M}$  de RNA e 17  $\mu\text{M}$  de solução de proteínas na razão 1:1:1 dos subcomplexos Sm. No caso das amostras do núcleo snRNP U2 contendo U2A'/U2B'', cristais na forma de pequenas agulhas se formaram a 20°C em algumas condições de cristalização do *kit* Morpheus II (*Molecular Dimensions*) pelo método de difusão de vapor em gota sentada (do inglês, *sitting drop*), como mostra a Tabela 2.2. Este *kit* consiste de uma mistura de precipitantes, tampões e aditivos na razão 1/2 [mistura de precipitantes] + 1/10 [mistura de aditivos] + 1/10 [sistema tampão] + 3/10 dH<sub>2</sub>O.

Tabela 2.2 - Condições de cristalização do kit *Morpheus II (Molecular Dimensions)* onde foram observados cristais para amostra de núcleo de snRNP U2.

Condição	Componentes	Gota contendo 0,2 $\mu\text{L}$ de solução de proteína e 0,2 $\mu\text{L}$ de solução de reservatório	Gota contendo 0,5 $\mu\text{L}$ de solução de proteína e 0,5 $\mu\text{L}$ de solução de reservatório
A1	90 mM LiNaK 0,1 M sistema tampão 4 (pH 6,5) 50 % v/v mistura de precipitantes 5		-
D1	2 mM oxometalatos 0,1 M sistema tampão 4 (pH 6,5) 50 % v/v mistura de precipitantes 5	-	
E1	2 mM lantanídeos 0,1 M sistema tampão 4 (pH 6,5) 50 % v/v mistura de precipitantes 5	-	
E9	2 mM lantanídeos 0,1 M sistema tampão 6 (pH 8,5) 50 % v/v mistura de precipitantes 5	-	
F5	100 mM monossacarídeos II 0,1 M sistema tampão 5 (pH 7,5) 50 % v/v mistura de precipitantes 5		-
F9	100 mM monossacarídeos II 0,1 M sistema tampão 6 (pH 8,5) 50 % v/v mistura de precipitantes 5	-	
G1	100 mM aminoácidos II 0,1 M sistema tampão 4 (pH 6,5) 50 % v/v mistura de precipitantes 5	-	

continua

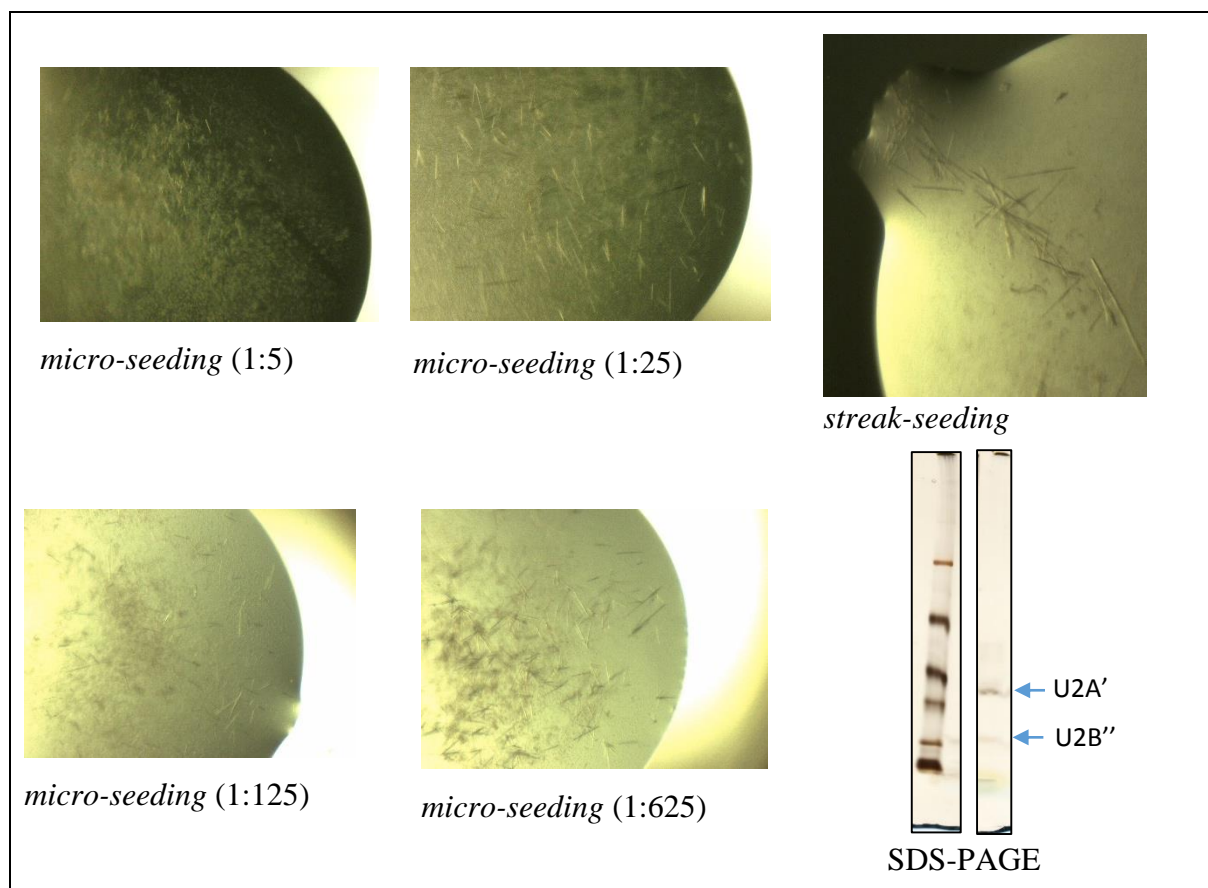
continuação			
G9	100 mM aminoácidos II 0,1 M sistema tampão 6 (pH 8,5) 50 % v/v mistura de precipitantes 5	-	
H1	40 mM poliaminas 0,1 M sistema tampão 5 (pH 6,5) 50 % v/v mistura de precipitantes 5		
H5	40 mM poliaminas 0,1 M sistema tampão 4 (pH 7,5) 50 % v/v mistura de precipitantes 5		-
H9	40 mM poliaminas 0,1 M sistema tampão 6 (pH 8,5) 50 % v/v mistura de precipitantes 5	-	

\* A composição de cada mistura de precipitante, sistema tampão e aditivos é descrita abaixo:

- 0,9 M LiNaK = 0,3 M sulfato de lítio, 0,3 M sulfato de sódio, 0,3 M sulfato de potássio.
  - 0,02 M oxometalatos = 0,005 M cromato de sódio tetrahidratado, 0,005 M molibdato de sódio desidratado, 0,005 M tungstato de sódio desidratado, 0,005 M ortovanadato de sódio
  - 100 mM monossacarídeos II = 0,2 M xilitol, 0,2 M mio-inositol, 0,2 M D-(-)-frutose, 0,2 M ramnose monohidratado, 0,2 M D-sorbitol
  - 100 mM aminoácidos II = 0,2 M hidrocloreto de DL-arginina, 0,2 M DL-treonina, 0,2 M monocloreto de DL-histidina monohidratado, 0,2 M hidrocloreto de DL-5-hidroxilisina, 0,2 M trans-4-hidroxi-L-prolina
  - 40 mM poliaminas = 0,1 M tetracloroeto de espermina, 0,1 M trihidrocloreto de espermidina, 0,1 M dihidrocloreto de 1,4-diaminobutano, 0,1 M monocloreto de DL-ornitina
  - 1 M sistema tampão 4 = 1 M MOPSO, 1 M Bis-Tris
  - 1 M sistema tampão 5 = 1 M BES, 1 M trietilamina (TEA)
  - 1 M sistema tampão 6 = 1 M Gly-Gly, 1 M AMPD
  - 50 % v/v mistura de precipitantes 5 = 30% w/v PEG 3000, 40% v/v 1, 2, 4-butanotriol, 2% w/v NDSB 256
- Fonte: Elaborada pelo autor.

Nota-se que, para as gotas com menor volume, em geral, houve a formação de muitos cristais pequenos, enquanto para as gotas de maior volume, cristais maiores foram observados em menor número. Tal resultado vai ao encontro de observações experimentais descritas na literatura de que, para um dado volume de reservatório, gotas menores tendem a equilibrar mais rápido do que gotas de volume maior.<sup>203</sup> Além disso, estima-se que, para uma gota de até 2  $\mu\text{L}$  e um reservatório de até 500  $\mu\text{L}$ , o equilíbrio seja atingido entre 3 a 6 dias. No caso dos cristais obtidos, o tempo para o aparecimento dos primeiros cristais foi de aproximadamente 15 dias. Isso indica que possa ter ocorrido um evento de dissociação e/ou proteólise de uma ou mais proteínas do núcleo de snRNP U2 antes ou durante o processo de cristalização.

Tentativas de reprodução das condições de cristalização mostradas acima a partir dos componentes individuais utilizando-se o mesmo método em volume maior (em placas de 24 poços) não foram bem sucedidas. Com o objetivo de se obter monocristais maiores, técnicas de *microseeding* e *streak-seeding* foram empregadas em gotas contendo 1  $\mu\text{L}$  de solução de proteína e 1  $\mu\text{L}$  de solução do reservatório adquirida diretamente da empresa *Molecular Dimensions* (Figura 2.21).



**Figura 2.21 - Ensaio de otimização de cristalização por meio de sementeira.** São mostrados resultados representativos de diferentes proporções de amostra estoque de semente para a condição de cristalização G9 do *kit Morpheus II (Molecular Dimensions)*. Apresenta-se também resultado de NuPAGE 4-20% Bis-Tris (*Thermoscientific*), corado com nitrato de prata, para um dos cristais. Fonte: Elaborada pelo autor.

Observa-se na Figura 2.21 a formação de cristais maiores para uma maior diluição de sementes, onde um menor número de núcleos leva à formação de menor quantidade de cristais. No caso do *streak-seeding*, cristais maiores, ainda na forma de agulhas, foram obtidos na linha onde o bigode de gato tocou a gota de cristalização. Entretanto, não foram observados macrocristais únicos nas gotas.

O maior cristal obtido na forma de barra atingiu  $\sim 250 \mu\text{m}$  em uma dimensão e  $\sim 20 \mu\text{m}$  na outra dimensão após  $\sim 60$  dias em gota contendo 500 nL de solução de proteínas e 500 nL de solução de reservatório. *Microseeding* e *streak-seeding* em placas de 24 poços (método de



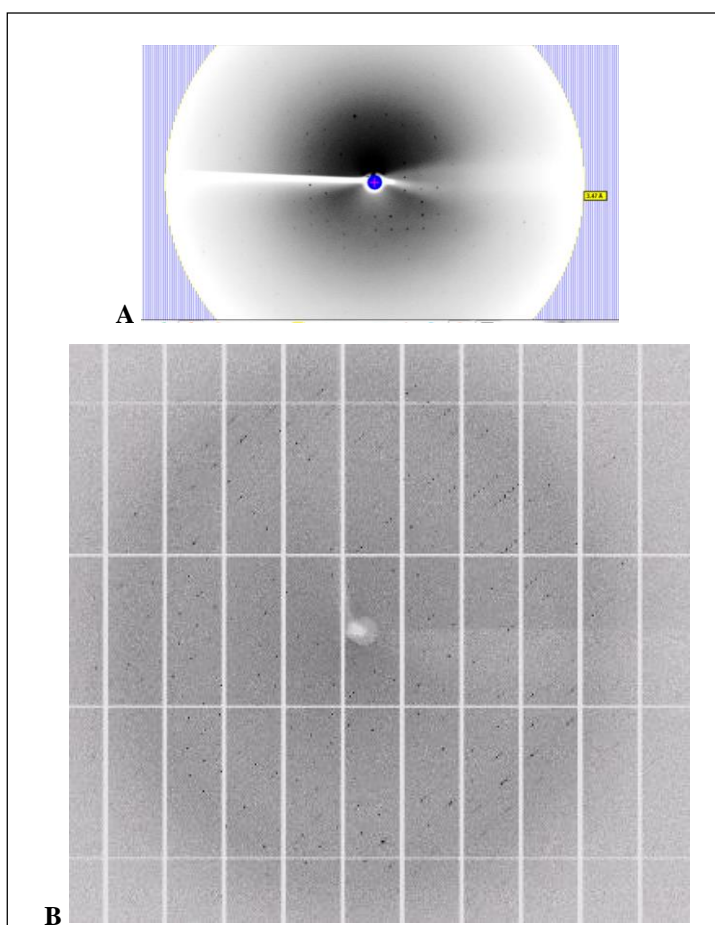
difusão de vapor em gota sentada) também levaram a cristais com pelo menos 100  $\mu\text{m}$  na dimensão maior. Um dos cristais foi lavado por três vezes com a solução do reservatório e, então, dissolvido em 10  $\mu\text{L}$  de água milli-Q e, a esta solução, foram adicionados 10  $\mu\text{L}$  de tampão de amostra (5X concentrado). A amostra resultante foi fervida por 5 minutos e aplicada em NuPAGE 4-20% Bis-Tris (*Thermoscientific*), corado com nitrato de prata.

O gel mostra claramente bandas com massas moleculares correspondentes às proteínas U2A' e U2B'' (Figura 2.21). Entretanto, a região de proteínas Sm não está bem resolvida no gel. Além disso, a presença de RNA no cristal não foi avaliada. Sabendo-se que os cristais obtidos continham pelo menos U2A'/U2B'' e que as condições de cristalização continham agentes crioprotetores, optou-se por resfriar em nitrogênio líquido 27 cristais com pelo menos 100  $\mu\text{m}$  na dimensão maior retirados diretamente das gotas de cristalização. Estes cristais foram armazenados por até um mês para então se realizar o procedimento de coleta de dados de difração de raios-X.

#### 2.4.6.2 Análise dos dados de difração de raios-X

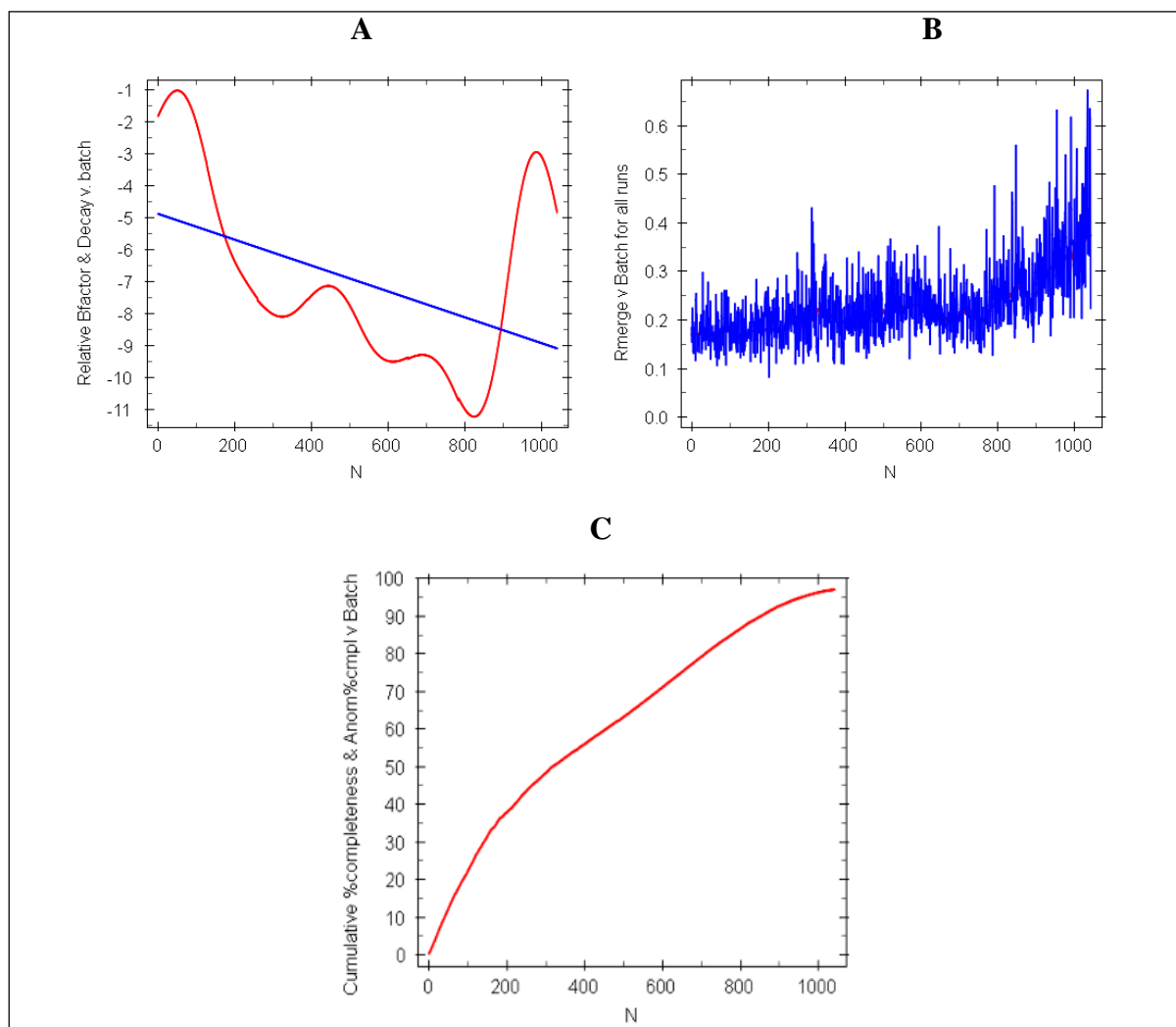
A coleta de dados de difração de raios-X *in house* no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas, Cambridge, Reino Unido, em um cristal obtido na condição H1 do kit Morpheus II (*Molecular Dimensions*) (Tabela 2.2) usando um detector do tipo placa de imagem MAR345 ( $\lambda = 1,54 \text{ \AA}$ , distância cristal-detector = 200 mm) mostrou um padrão de difração com reflexões até 3,5  $\text{\AA}$  de resolução. A indexação das imagens de difração pelo programa *iMosflm*<sup>207</sup> resultou em uma simetria monoclinica P2, porém não foi descartada a possibilidade de simetria P2<sub>1</sub>. Os parâmetros da cela unitária foram determinadas como sendo  $a = 59,7 \text{ \AA}$ ,  $b = 42,9 \text{ \AA}$  e  $c = 144,7 \text{ \AA}$  e ângulos de  $\alpha = \gamma = 90^\circ$  e  $\beta = 102,7^\circ$  (Figura 2.22A).

Em seguida, os 27 cristais resfriados em nitrogênio líquido foram enviados para a fonte de luz síncrotron *Diamond Light Source* (DLS) em Oxfordshire, Reino Unido. Dados de difração de raios-X foram coletados na linha de luz I24, que possui feixe de raios-X com microfoco de 10  $\mu\text{m}$  por 10  $\mu\text{m}$ , ideal para difração em cristais pequenos. Foi observado que uma rápida interrupção do feixe de nitrogênio a 80 K incidente no cristal (rápido *annealing*) levou a uma melhora no padrão de difração, tanto em relação à resolução quanto à separação dos pontos de difração (Figura 2.22B).



**Figura 2.22 - Padrões de difração de cristais obtidos em amostras do núcleo snRNP U2** A- obtido *in house*, no Laboratório de Biologia Molecular do Conselho de Pesquisas Médicas do Reino Unido, e B- medido na linha I24 da *Diamond Light Source*, Reino Unido.  
Fonte: Elaborada pelo autor.

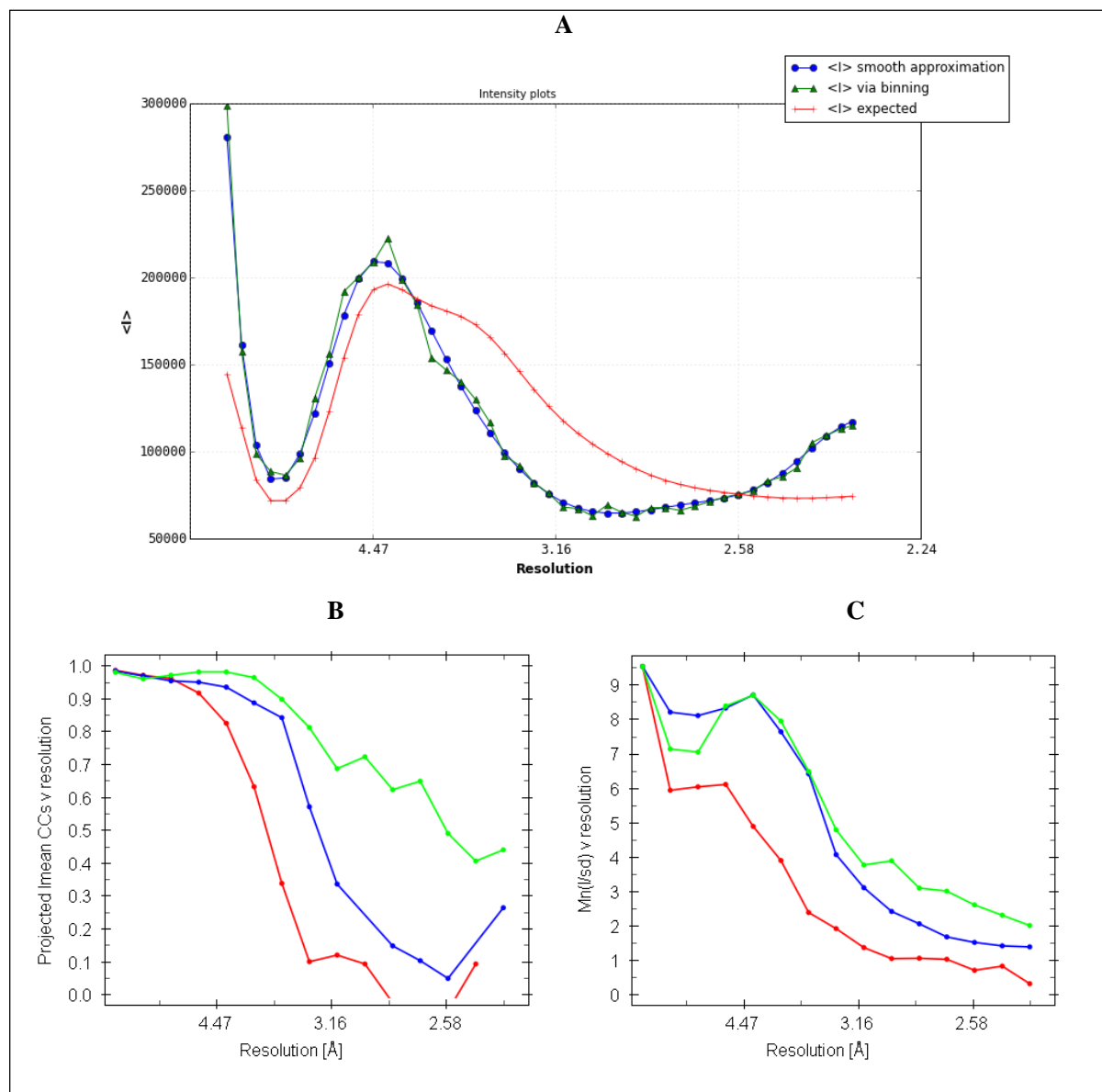
Os dados de difração de raios-X (Figura 2.22-B) foram processados utilizando-se os programas *iMosflm*<sup>207</sup> e *AIMLESS*<sup>208</sup>. Os parâmetros da cela unitária refinados para o cristal com maior poder de difração foram  $a = 62,08 \text{ \AA}$ ,  $b = 43,06 \text{ \AA}$  e  $c = 146,08 \text{ \AA}$  e  $\alpha = \gamma = 90^\circ$  e  $\beta = 99,89^\circ$ . Uma análise do fator-B relativo e do  $R_{\text{merge}}$  por imagem indica dano ocasionado por radiação a partir da imagem de número 1050. Deste modo, o conjunto de dados foi truncado de tal forma a resultar em um conjunto de dados o mais completo possível, de acordo com a análise da porcentagem de completude acumulada por imagem (Figura 2.23-A-C). Adicionalmente, não foi identificado sinal anômalo aparente no conjunto de dados.



**Figura 2.23 - Análise de dados cristalográficos I.** A- Fator  $B_{relativo}$  versus número da imagem. B-  $R_{merge}$  versus número da imagem. C- Completeza acumulada versus número da imagem.  
Fonte: Elaborada pelo autor a partir dos resultados do programa AIMLESS<sup>209</sup>.

Considerando-se as primeiras 1050 imagens de difração de raios-X, a maioria das reflexões parciais observadas (com mosaicidade = 1,42) foram integradas em 56.853 reflexões completas, das quais 21.885 eram únicas. O maior pico diferente da origem no mapa de Patterson (não mostrado) tem apenas 7% da altura do pico da origem, o que aponta para a ausência de possível pseudotranslação. Adicionalmente, o escalonamento de Wilson estima um fator B isotrópico global igual a  $26,24 \text{ \AA}^{-2}$  e um fator de escala  $\log(k) = -3,51$ . O gráfico de Wilson, do qual esses dois parâmetros foram derivados, apresenta uma derivada primeira positiva a partir de  $2,6 \text{ \AA}$ , o que indica possível anisotropia (Figura 2.24-A).

A análise de  $I/\sigma(I)$  médio e do coeficiente de correlação ( $CC$ ) por resolução revela um comportamento diferente para cada uma das três direções independentes de difração ( $d_1 = 0,97 h + 0,23 l$  (vermelho),  $d_2 = k$  (azul) e  $d_3 = -0,11 h + 0,99 l$  (verde)) (Figura 2.24-B-C):



**Figura 2.24 - Análise de dados cristalográficos II.** A- Análise do gráfico de Wilson. B- e C- Análise de  $I/\sigma(I)$  médio e  $CC_{1/2}$ , respectivamente, nas direções de  $d_1$  (vermelho),  $d_2$  (azul) e  $d_3$  (verde).  
Fonte: Elaborada pelo autor a partir dos resultados do programa AIMLESS<sup>209</sup>.

Os dados mostram que a maior resolução medida encontra-se na direção de  $d_3$ , enquanto a difração na direção de  $d_1$  vai até uma menor resolução (Figura 2.24 e Tabela 2.3). Adicionalmente, a análise da estatística das intensidades das reflexões medidas não indica presença de geminação (do inglês, *twinning*) ( $\langle I^2 \rangle / \langle I \rangle^2 = 1,950$  (ausência = 2,000/presença = 1,500),  $\langle F^2 \rangle / \langle F \rangle^2 = 0,816$  (ausência = 0,785/presença = 0,885),  $\langle |E^2| - 1 \rangle = 0,696$  (ausência = 0,736/presença = 0,541) para reflexões acêntricas, e  $\langle I^2 \rangle / \langle I \rangle^2 = 2,973$  (ausência = 3,000/presença = 2,000),  $\langle F^2 \rangle / \langle F \rangle^2 = 0,673$  (ausência = 0,637/presença = 0,785),  $\langle |E^2| - 1 \rangle = 0,936$  (ausência = 0,968/ presença = 0,736) para reflexões cêntricas. Já o teste L para reflexões acêntricas resulta em  $\langle L \rangle = 0,444$  (ausência = 0,500/presença = 0,375)

e  $\langle L^2 \rangle = 0,270$  (ausência = 0,333/presença = 0,200). A variação observada pode ser devida à anisotropia dos dados de difração.

A Tabela 2.3 mostra as estatísticas dos dados nativos de difração de raios-X, incluindo todos os dados até a maior resolução. Entretanto, melhores estatísticas são obtidas truncando-se os dados na resolução de 2,66 Å (Tabela 2.3):

**Tabela 2.3 - Estatísticas dos dados cristalográficos.**

Propriedades globais e na última camada de resolução	Conjunto de dados original	Conjunto de dados truncados a 2,66Å de resolução máxima
Limites de resolução (Å)	71,95-2,35 (2,43-2,35)	71,95-2,66 (2,79-2,66)
Fator B global (Wilson, Å <sup>2</sup> )	26,24	37,24
$R_{merge}$	0,228 (0,874)	0,165 (0,614)
$R_{meas}$	0,280 (1,015)	0,233 (0,863)
$R_{pim}$	0,197 (0,713)	0,164 (0,607)
Número de observações	80.155 (5.880)	56.853 (7.617)
Número de observações únicas	31.165 (2.620)	21.885 (2.905)
$\langle I/\sigma(I) \rangle$	3,0 (1,0)	3,9 (1,4)
$CC_{1/2}$	0,970 (0,089)	0,972 (0,308)
Completeza (%)	96,9 (84,1)	98,2 (99,3)
Multiplicidade	2,6 (2,2)	2,6 (2,6)
Limites de resolução reais (Å)		
$d_1 = 0,97h + 0,23l$	3,22	3,22
$d_2 = k$	2,66	2,66
$d_3 = -0,11h + 0,99l$	2,35	2,66

Fonte: Elaborada pelo autor a partir dos resultados do programa *AIMLESS*<sup>209</sup>.

O programa *POINTLESS*<sup>209</sup> prediz que o cristal apresenta grupo especial  $P2_1$  com probabilidade de 0,823, dadas as ausências sistemáticas ao longo do eixo  $(0,k,0)$  quando  $k$  é ímpar. Entretanto, uma análise cuidadosa da lista de reflexões medidas (Tabela 2.4) mostra que as reflexões ímpares deste eixo apresentam  $I/\sigma(I)$  maior que zero, apesar de menores que as observadas para  $k$  par.

Tabela 2.4 – Análise de ausências sistemáticas na direção de  $d_2$ .

$k$	$I_{exp}$	$I_{exp}/\sigma(I_{exp})$
1	2	0,85
2	570	167,68
3	27	14,93
4	1043	167,95
6	77	18,93
7	8	3,36
8	359	39,90
9	17	4,01
10	33	7,00
11	13	1,87
12	44	7,67
13	26	3,24
14	52	8,09
16	9	1,59
17	2	0,28
18	9	1,66

Fonte: Elaborada pelo autor a partir dos resultados do programa *AIMLESS*<sup>209</sup>.

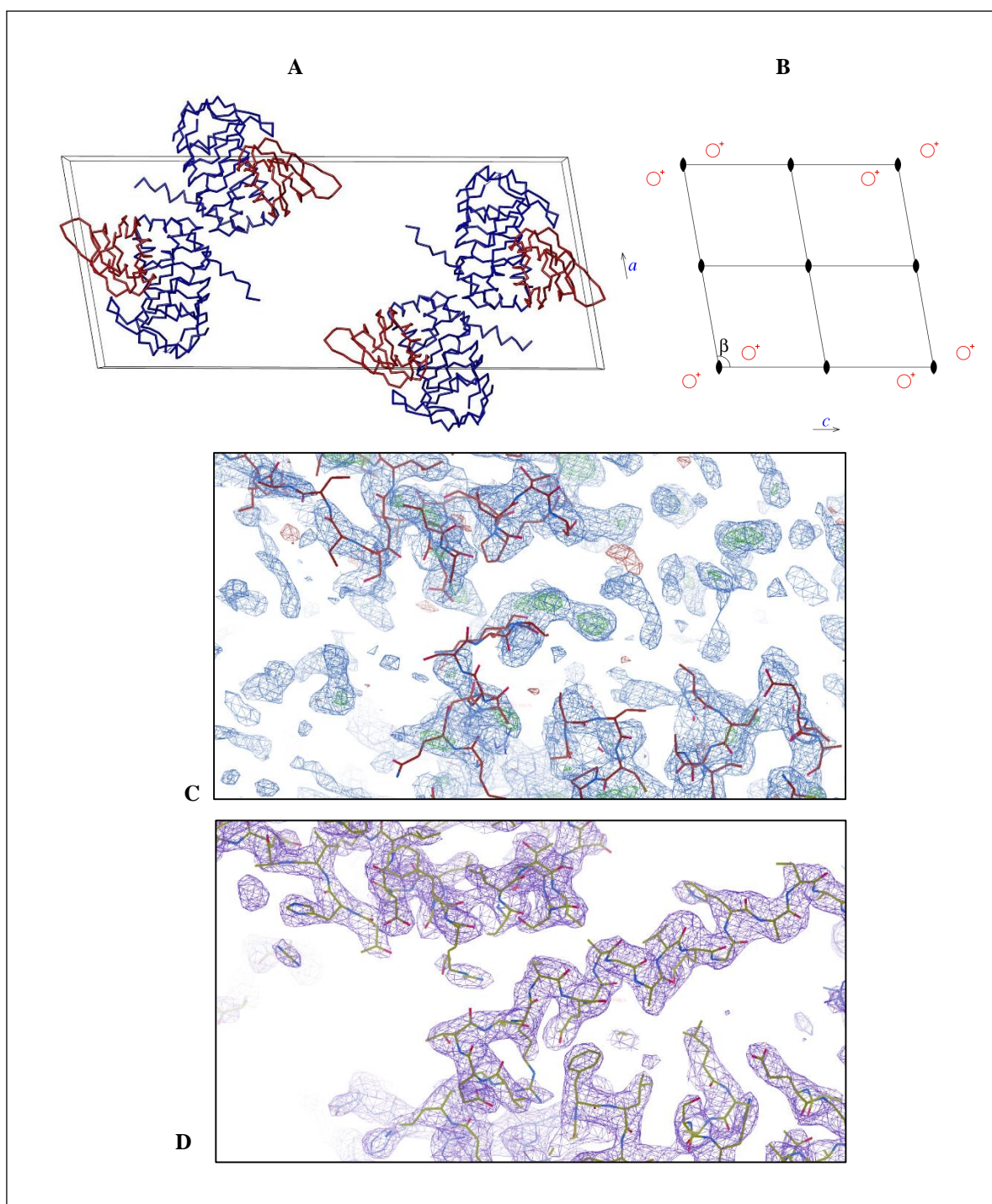
A análise automática do coeficiente de Mathews<sup>27</sup> resultou em  $2,32 \text{ \AA}^3 \text{ kDa}^{-1}$  e uma porcentagem de solvente igual a 47% para um conteúdo de proteína com massa molecular igual a 83 kDa na unidade assimétrica. Este resultado indica que o complexo formado por núcleo Sm 2/snRNA U2/U2A'/U2B'' não foi cristalizado por completo. Associado a esse resultado, deve-se notar que o SDS-PAGE mostra claramente a presença das proteínas U2A'/U2B''.

### 2.4.6.3 Resolução do problema das fases, refinamento e análise do modelo de estrutura tridimensional

A estrutura cristalográfica de U2A'/U2B'' de *H. sapiens* (código PDB 1A9N<sup>26</sup>) foi utilizada no programa *PHASER*<sup>212</sup> como modelo de busca para substituição molecular a partir dos dados de difração de raios-X inicialmente processados em simetria monoclinica P2<sub>1</sub>. O programa testou o faseamento nos grupos espaciais P2 e P2<sub>1</sub> e a melhor solução foi encontrada para P2 ( $top \text{ LLG} = 534,976$  e  $top \text{ TFZ} = 23,6$ ) com duas cópias de cada proteína na unidade assimétrica. Entretanto, uma solução também foi encontrada em simetria P2<sub>1</sub> ( $top \text{ LLG} = 202,410$  e  $top \text{ TFZ} = 12,4$ ). Portanto, seguiu-se com o refinamento de ambas as soluções em paralelo.

Os valores iniciais de  $R_{work}$  e  $R_{free}$  após refinamento de corpo rígido no programa *phenix.refine*<sup>217</sup> para o modelo resolvido em simetria P2 foram 0,44 e 0,45, respectivamente.

Por outro lado, o modelo resolvido em simetria  $P2_1$  apresentou  $R_{\text{work}}$  igual a 0,47 e  $R_{\text{free}}$  igual a 0,48. Os modelos foram então enviados ao programa *ARP/wARP*<sup>247</sup> para reconstrução de cadeias laterais e otimização da densidade eletrônica. Após inspeção visual no programa *Coot*<sup>209</sup>, cadeias laterais ausentes no modelo foram adicionadas. As figuras 2.25-A-B mostram representações das duas unidades assimétricas dentro da cela unitária do cristal com simetria monoclinica P2 e as Figuras 2.25-C-D mostram a evolução do modelo ajustado na densidade eletrônica antes e após refinamento.



**Figura 2.25 - Estrutura cristalográfica de U2A'/U2B'' de *T. brucei* I.** A e B- Conteúdo da cela unitária: dois heterodímeros de U2A'(azul)/U2B''(vermelho) na unidade assimétrica. Em B é apresentada a projeção da cela unitária no plano (*a,c*). Os círculos vermelhos correspondem à unidade assimétrica. Densidade eletrônica e modelo atômico C- logo após etapa de substituição molecular e D- logo após refinamento. A região C-terminal de U2A' está evidenciada. Nota-se, claramente, uma melhora na qualidade da densidade eletrônica na região após o refinamento, possibilitando a construção de uma hélice- $\alpha$  nesta região de U2A'.

Fonte: Elaborada pelo autor nos programas *Coot*<sup>219</sup> e *Chimera*1.9<sup>220</sup>.

Uma análise visual, no programa *Coot*<sup>219</sup>, das densidades eletrônicas calculadas não revelou densidade adicional que poderia corresponder a proteínas Sm ou a uma molécula de RNA. Por outro lado, observou-se densidade eletrônica adicional na extremidade C-terminal das duas cópias de U2A' (Figura 2.25-D). Foi possível adicionar manualmente 35 resíduos a esta região da cadeia polipeptídica de cada cópia de U2A' usando o programa *Coot*<sup>209</sup> para refinamento no espaço real. Estes resíduos formam uma estrutura de hélice- $\alpha$  nesta extremidade (Figura 2.25-D).

A estatística dos dados após última etapa de refinamento realizada antes da escrita deste documento encontra-se na Tabela 2.5:

**Tabela 2.5 – Estatística de refinamento da estrutura cristalográfica de U2A'/U2B'' de *T. brucei***

Propriedades	
Número de moléculas na unidade assimétrica	4
Número átomos (resíduos) refinados	4.008 (83)
Faixa de resolução (Å)	71,95 – 2,66
Número de reflexões usadas no refinamento	20.504
Número de reflexões para validação cruzada	1.381
$R_{\text{free}}$	0,2787 (0,3896)
R	0,2380 (0,3569)
Fator B global médio (Å <sup>2</sup> )	53,50
<i>rmsd</i> - fator B (átomos ligados na cadeia principal) (Å <sup>2</sup> )	1,410
<i>rmsd</i> - fator B (átomos ligados nas cadeias laterais) (Å <sup>2</sup> )	1,587
<i>rmsd</i> de comprimento de ligações (Å)	0,014
<i>rmsd</i> de ângulos de ligação (°)	1,506
Correlação entre os coeficientes $F_{\text{obs}}$ e $F_{\text{calc}}$	0,9075
Correlação entre os coeficientes $F_{\text{obs}}$ e $F_{\text{calc}}$ (livre)	0,8520
Gráfico de Ramachandran*	
Regiões favoráveis (%)	91,5
Regiões generosamente permitidas (%)	7,5
Regiões não-permitidas (%)	1,0

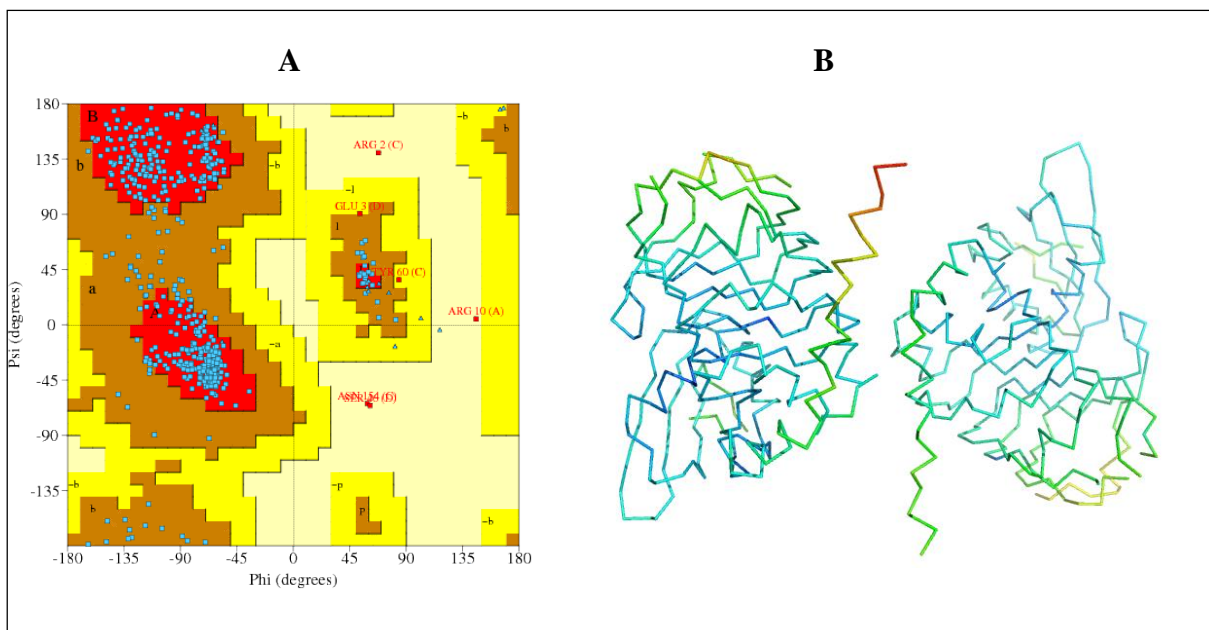
\*De acordo com as regiões definidas no programa *MolProbity*<sup>222</sup>.

Fonte: Elaborada pelo autor a partir dos resultados do programa *MolProbity*<sup>222</sup>.



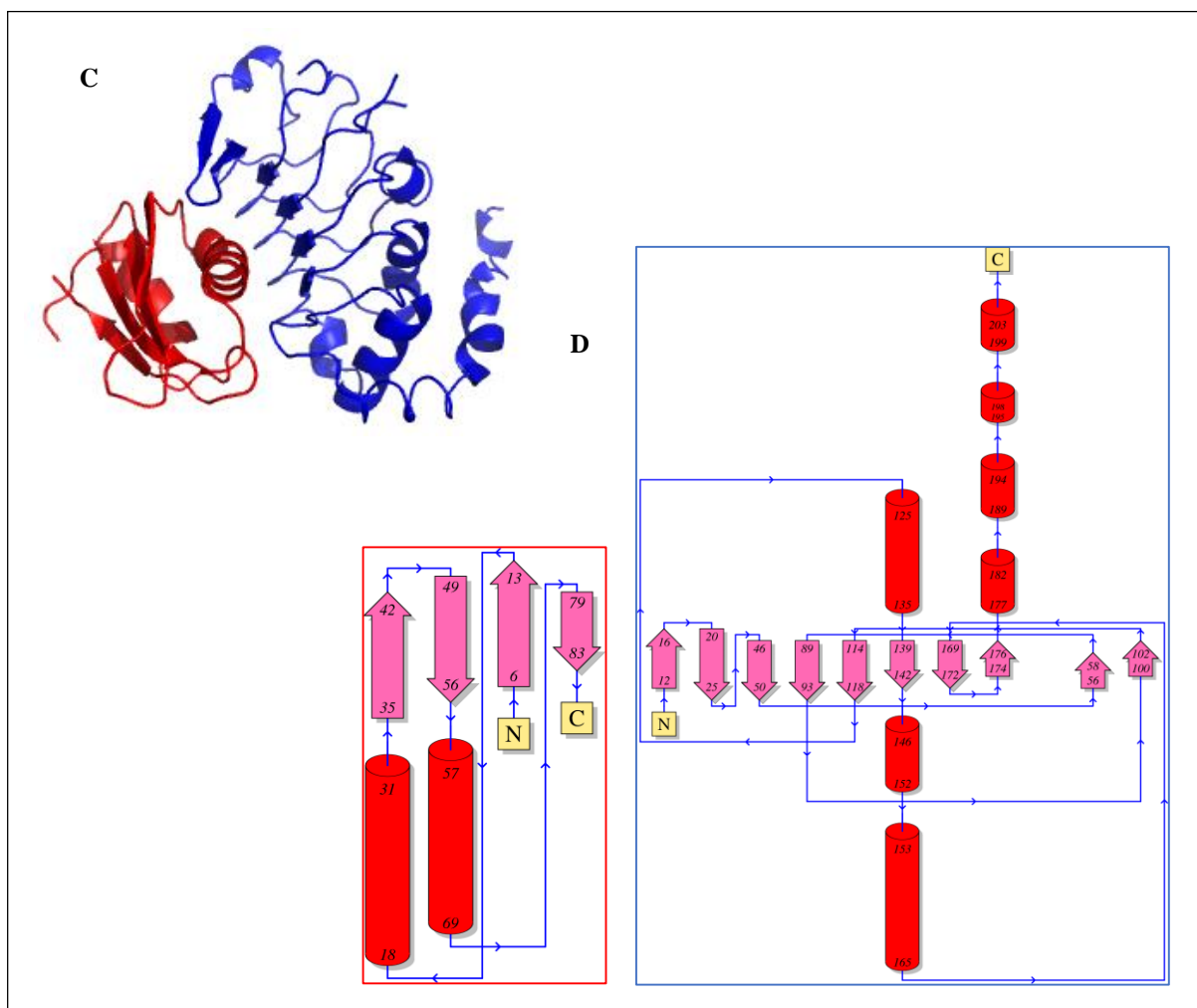
A maior parte dos resíduos de aminoácidos da estrutura cristalográfica encontra-se em regiões permitidas do gráfico de Ramachandran (91,5%), sendo que os resíduos localizados em regiões desfavoráveis (1,0% do total) estão em regiões de alças longas na estrutura (Tabela 2.5 e Figura 2.26-A). Regiões de alça e a extremidade C-terminal apresentam maior valor para o fator-B (Figura 2.26-B), o que corresponde a regiões de maior flexibilidade estrutural.

A comparação da estrutura tridimensional das cadeias A e C, que correspondem a U2A', revela *rmsd* igual a 0,3 Å e entre as cadeias B e D, que correspondem a U2B'', *rmsd* igual a 0,2 Å, o que indica estruturas similares. A análise da estrutura cristalográfica de U2A' das duas cadeias mostra que resíduos de 20 a 150 formam uma estrutura de solenoide, sendo também observada uma folha- $\beta$  antiparalela na região interna do domínio LRR. Entretanto, as alças apontadas para a superfície formam estruturas desordenadas ou pequenas hélices- $\alpha$ , como se observam na Figura 2.26-B. As regiões N- e C-terminais de U2A' se enovelam como estruturas em hélice- $\alpha$  mais longas. Por outro lado, a estrutura em hélice- $\alpha$  de U2B'' interage com a superfície côncava da folha- $\beta$  de U2A' (Figura 2.26-B).



continua

continuação



**Figura 2.26- Estrutura cristalográfica de U2A'/U2B'' de *T. brucei* II.** A- Gráfico de Ramachandran. B- Representação do conteúdo da célula unitária colorido de acordo com o valor do fator B por resíduo, do azul (menor valor) para o vermelho (maior valor). C- Representação das estruturas cristalográficas de U2A' (azul) e U2B'' (vermelho) de *T. brucei* na forma de *cartoon*. D- Topologia de U2A'. E- Topologia de U2B''.

Fonte: Elaborada pelo autor nos programas *Coot*<sup>219</sup> e *PyMOL*(Schrödinger, LLC).

As interações entre U2A' e U2B'' de cada heterodímero presente na unidade assimétrica do cristal são equivalentes, sendo que 22 resíduos da cadeia A-U2A' (área de superfície igual a 963 Å<sup>2</sup>) interagem com 18 resíduos da cadeia B-U2B'' (área de superfície igual a 974 Å<sup>2</sup>) por meio de 13 ligações de hidrogênio e duas pontes salinas, enquanto 21 resíduos da cadeia A-U2A' (área de superfície igual a 996 Å<sup>2</sup>) interagem com 17 resíduos da cadeia B-U2B'' (área de superfície igual a 988 Å<sup>2</sup>) por meio de 11 ligações de hidrogênio e duas pontes salinas. A energia de interação calculada pelo programa *PISA*<sup>223</sup> é igual a -4,4 kcal mol<sup>-1</sup> para a interface entre as cadeias A e B e igual a -4,0 kcal mol<sup>-1</sup> para a outra interface, sendo que o programa classifica essas interfaces com alta probabilidade de serem reais e não apenas contatos cristalinos.

A comparação da estrutura de U2A'/U2B'' de *T. brucei* na ausência de snRNA U2 aqui apresentada com aquela publicada pelo grupo do Dr. Kiyoshi Nagai na década de 1990 para U2A'/U2B'' de *H. sapiens* em interação com snRNA U2 mostra que o enovelamento das proteínas é conservado, com *rmsd* de posições atômicas da cadeia principal igual a até 1,0 Å (Tabela 2.6). Algumas diferenças são observadas em posições relevantes de cada proteína, como se pode observar na Figura 2.27.

**Tabela 2.6 – Comparação entre os valores de *rmsd* resultantes da sobreposição entre as cadeias polipeptídicas observadas nos modelos cristalográficos de U2A'/U2B'' de *T. brucei* e de *H. sapiens*.**

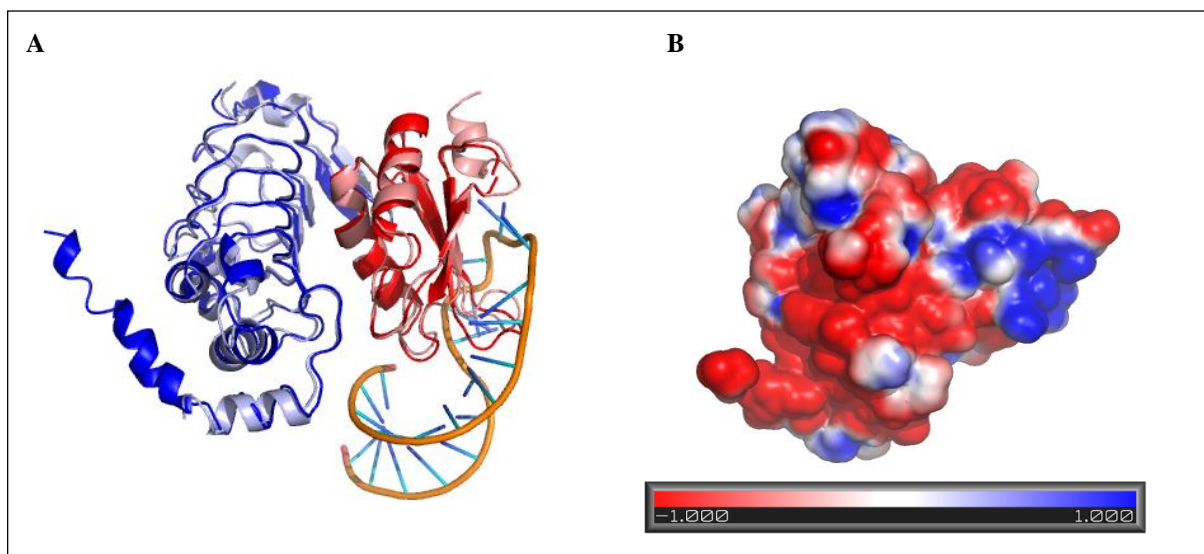
		U2A' de <i>T. brucei</i> ( <i>rmsd</i> , Å)		U2B'' de <i>T. brucei</i> ( <i>rmsd</i> , Å)	
		cadeia A	cadeia C	cadeia B	cadeia D
U2A' de	cadeia A	1,0 (1,4)	1,0 (1,4)	-	-
<i>H. sapiens</i>	cadeia C	1,0 (1,4)	1,0 (1,4)	-	-
U2B'' de	cadeia B	-	-	0,7 (1,1)	0,6 (1,1)
<i>H. sapiens</i>	cadeia D	-	-	0,8 (1,1)	0,6 (1,1)

\* Entre parênteses encontra-se o valor teórico obtido a partir da equação (9) para o desvio quadrático médio das distâncias entre átomos nas estruturas das cadeias polipeptídicas de U2A' e U2B'' nos modelos cristalográficos das proteínas de *T. brucei* e *H. sapiens*.

Fonte: Elaborada pelo autor.

As interações entre SL4 de snRNA U2 e a proteína U2B'' observadas na estrutura do heterodímero de *H. sapiens*<sup>15</sup> podem ser divididas nas seguintes regiões do snRNA: primeira metade da alça, segunda metade da alça, par de bases nitrogenadas na base da alça e na região de hélice (seção 2.4.1.2). A comparação das estruturas cristalográficas de U2A'/U2B'' de *T. brucei* com as ortólogas de *H. sapiens* mostra que os resíduos F52, Y9 e K45 são conservados na primeira em posições propícias para interações do tipo  $\pi$ - $\pi$  com as bases nitrogenadas U9, C10, C11 e A12, enquanto os resíduos T44 e L42 são conservados para interação com a outra metade da alça no snRNA U2 de *T. brucei*. Finalmente, os resíduos K16, D15 e L13 são conservados para interação com a base da alça. Entretanto, o resíduo A18 substitui K22 de U2B'' de *H. sapiens* na região de interação com a hélice de SL4.

Adicionalmente, observa-se uma abertura da quarta alça, entre as fitas- $\beta$ 2 e  $\beta$ 3, sendo que o resíduo S46, localizado no ponto médio desta alça, está a 6,3 Å de distância da posição equivalente na estrutura de *H. sapiens*. Esta movimentação está diretamente relacionada à interação com a molécula de RNA. Além disso, esta região tem potencial eletrostático positivo (azul, Figura 2.27-B) para interação com a molécula de RNA, a qual tem carga parcial negativa na região de interação.



**Figura 2.27 – Comparação da estrutura cristalográfica de U2A'/U2B'' de *T. brucei* e de *H. sapiens*.** A- Em vermelho e rosa estão representadas as estruturas de U2B'' de *T. brucei* e *H. sapiens*, respectivamente. Em azul escuro e azul claro estão representadas as estruturas de U2A' de *T. brucei* e *H. sapiens*, respectivamente. A molécula de SL4 do snRNA U2 de *H. sapiens* interage com U2B'' de esta espécie. B- Mapa de potencial eletrostático de superfície de U2A'/U2B'' de *T. brucei* obtido no programa APBS<sup>224</sup>. O potencial eletrostático é calculado em unidades de  $k_b T e_c^{-1}$  ( $\approx 25,9$  mV).  
Fonte: Elaborada pelo autor.

#### 2.4.7 Modelagem molecular da estrutura terciária e quaternária das proteínas Sm de *T. brucei*

Apesar de não se ter obtido sucesso na cristalização dos núcleos Sm de snRNPs de *T. brucei*, a modelagem molecular comparativa, aliada ao experimento de espalhamento de raios-X a baixo ângulo (apresentado na seção 2.4.8), possibilitou a obtenção de informação estrutural dos diferentes complexos. Para tal, foi necessária a busca por proteínas Sm cuja estrutura tridimensional é conhecida. Como descrito na seção 2.1.1, várias proteínas Sm têm estrutura resolvida, entretanto, as proteínas Sm de *H. sapiens* apresentam a maior relação identidade de sequência de aminoácidos/porcentagem de cobertura de alinhamento, as quais são mostradas na Tabela 2.7.

Tabela 2.7 – Comparação entre as seqüências de aminoácidos de proteínas Sm de *T. brucei* e de *H. sapiens*.

Proteínas Sm de <i>T. brucei</i>	Proteínas Sm de <i>H. sapiens</i>	Porcentagem de identidade (%)	Porcentagem de cobertura (%) na proteína de <i>T. brucei</i>
SmD3	SmD3	36	88
Sm16,5K	SmD3	30	79
SSm4	SmD3	33	14
SmB	SmB	43	66
Sm15K	SmB	23	59
SmD1	SmD1	43	66
SmD2	SmD2	41	73
SmF	SmF	50	83
SmE	SmE	44	93
SmG	SmG	38	96

\*Estão destacadas em vermelho as proteínas com baixa relação identidade/cobertura de alinhamento de sequências de aminoácidos entre ortólogos de *T. brucei* e *H. sapiens*.

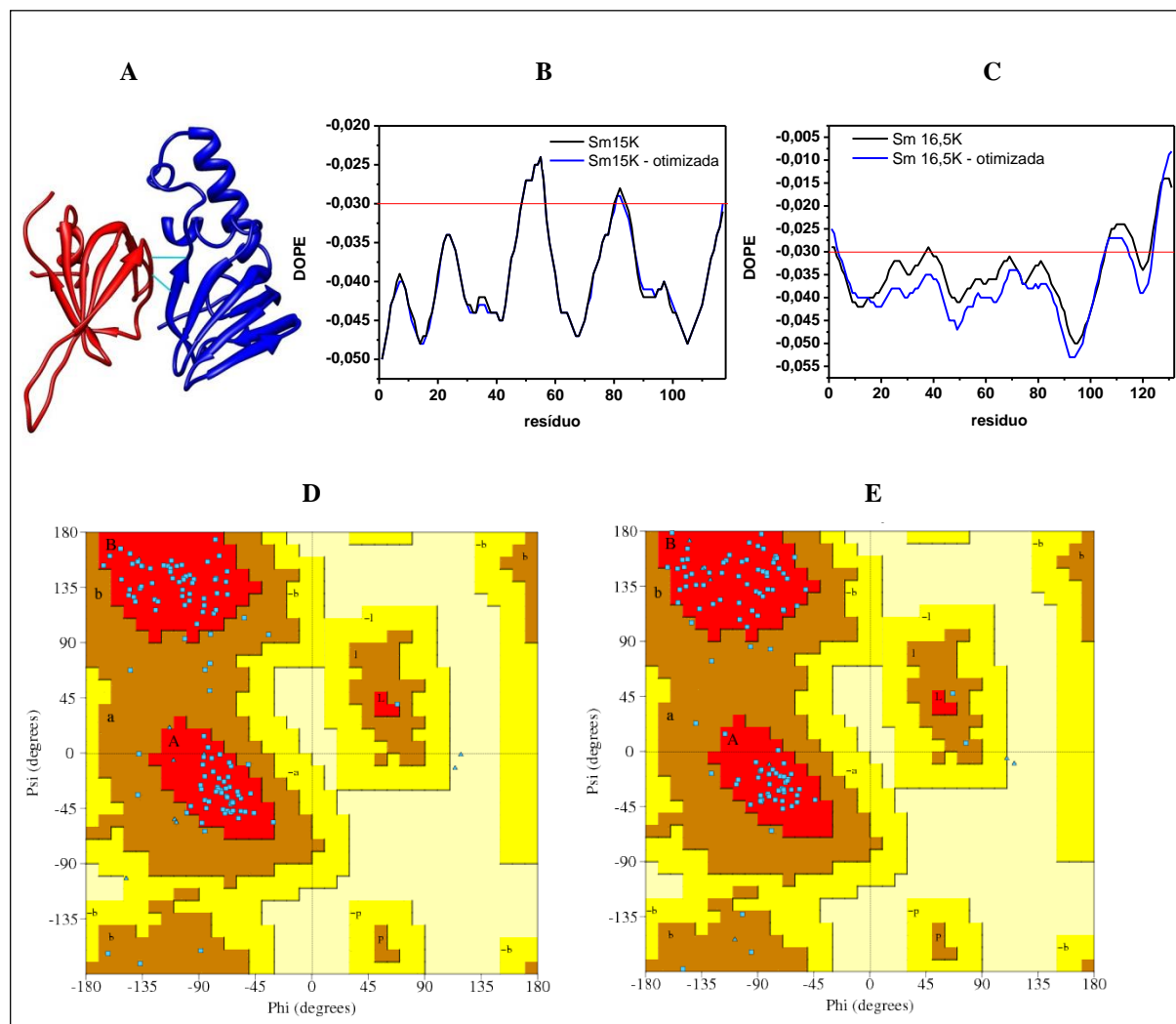
Fonte: Elaborada pelo autor.

A maioria destes valores encontra-se no limiar para uma possível modelagem das suas estruturas 3D por homologia ou por alinhamento, não sendo viável uma modelagem por homologia para as proteínas SSm4 e Sm15K com base nos modelos humanos. Os dímeros de Sm15K/Sm16,5K e SmB/SSm4 são estruturas exclusivas de *T. brucei* e apresentam baixa identidade de seqüência de aminoácidos e cobertura de alinhamento com homólogos cuja estrutura já foi resolvida experimentalmente.

Portanto, as técnicas de modelagem comparativa e de alinhamento foram empregadas para sua modelagem. No caso da técnica de modelagem por alinhamento, utilizando-se o servidor *I-Tasser*, a predição de estrutura terciária foi feita de forma automatizada para cada uma das seqüências de Sm15K, Sm16,5K e SSm4. O melhor modelo para cada proteína foi escolhido baseado no parâmetro de confiança *C-score* igual a -1,45, -3,13 e -4,67 para Sm15K, Sm16,5K e SSm4, respectivamente. Este resultado mede a convergência dos parâmetros durante a simulação computacional e varia de -5 a 2, sendo maior quanto mais confiável o modelo. Nota-se que os três modelos gerados apresentam um valor negativo para este parâmetro, sendo que o modelo gerado para a estrutura de SSm4 não é confiável por estar no limite inferior de *C-score*. Portanto, o modelo gerado para a proteína SSm4 foi desconsiderado e seguiu-se com a avaliação da estereoquímica dos modelos de Sm15K e Sm16,5K (Figura 2.28-D-E).

Em ambos os casos, alguns resíduos da região C-terminal apresentaram baixa qualidade, quando analisado o parâmetro energia proteica otimizada discreta (sigla do inglês, DOPE) por

resíduo (Figura 2.28-B-C), um potencial estatístico que mede a energia de cada modelo gerado, o que pode estar relacionado à baixa identidade destas regiões com sequências de aminoácidos de estrutura conhecida. Uma otimização das regiões de baixa qualidade foi feita utilizando-se o programa *Modeller 9.10* seguida de análise estereoquímica e do parâmetro DOPE por resíduo.



**Figura 2.28 – Sm15K/Sm16,5K.** A- Modelo de estrutura 3D do dímero das proteínas Sm15K (vermelho) e Sm16,5K (azul). B e C- Gráficos de DOPE *score* por resíduo para os modelos de Sm15K e Sm16,5K, respectivamente, antes e após otimização. D- e E- Gráficos de Ramachandran para os modelos otimizados de Sm15K e Sm16,5K, respectivamente.

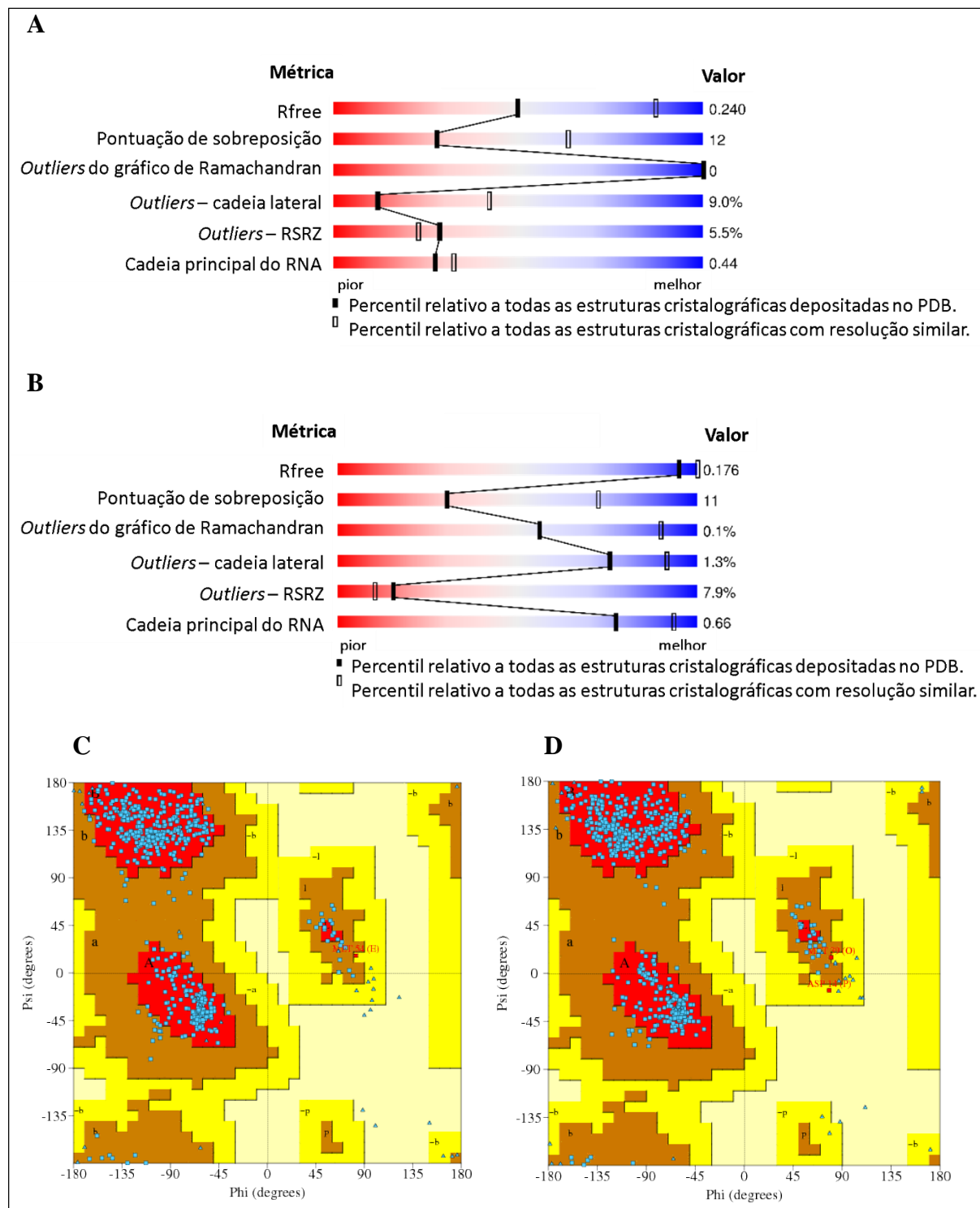
Fonte: Elaborada pelo autor.

O modelo otimizado de Sm15K apresenta 87,6% dos resíduos em regiões mais favoráveis do gráfico de Ramachandran e 12,4% nas regiões adicionalmente favoráveis, enquanto o modelo otimizado de Sm16,5K apresenta 87,0% na primeira e 13,0% dos resíduos na segunda. Estes modelos foram alinhados com o modelo do dímero SmD3/SmB humano (código PDB 1D3B<sup>174</sup>) para construção do modelo do dímero de Sm15K/Sm16,5K mostrado na Figura 2.28-A. Por meio do servidor *PISA*<sup>223</sup>, foi possível calcular uma área da interface igual a 937,3 Å<sup>2</sup> e energia livre de interação ( $\Delta G$ ) entre os monômeros igual a - 7,4 kcal mol<sup>-1</sup>.

O dímero é mantido por interações de hidrogênio para formação de uma folha- $\beta$  antiparalela entre as fitas  $\beta 5$  de Sm15K e  $\beta 4$  de Sm16,5K, além de contatos hidrofóbicos no interior da interface e contatos polares nas extremidades entre cadeias laterais dos resíduos de aminoácidos da interface.

Para a modelagem por homologia das proteínas Sm do núcleo canônico de *T. brucei* (SmB/SmD3/SmD1/SmD2/SmE/SmF/SmG – núcleo Sm 1), utilizou-se o programa *Modeller* 9.10<sup>248</sup>, baseando-se nos modelos cristalográficos dos núcleos Sm de snRNPs U1<sup>55</sup> (U1 – código PDB 2PJO) e U4<sup>91</sup> de *H. sapiens* (U4 – código PDB 2WZJ), obtidos a partir de dados de difração de raios-X com resolução máxima igual a 3,3 Å e 3,6 Å pelo grupo do Dr. Kiyoshi Nagai. Estas estruturas puderam ser utilizadas como moldes para a construção de 100 modelos estruturais do núcleo Sm canônico de *T. brucei*, a partir do alinhamento de sequências de aminoácidos já discutido na seção 2.4.1.1 deste capítulo. A Figura 2.29-A-B mostra um resumo das propriedades geométricas dos moldes do núcleo Sm canônico de *H. sapiens* com melhor geometria na unidade assimétrica do cristal e de validação destas estruturas cristalográficas.

O modelo cristalográfico das proteínas Sm de *H. sapiens* que formam o núcleo de snRNP U1 apresenta 89,4 % dos resíduos em regiões mais favoráveis, 10,4% nas regiões adicionalmente favoráveis e 0,2% em regiões generosamente favoráveis do gráfico de Ramachandran, enquanto o modelo cristalográfico das mesmas proteínas no núcleo de snRNP U4 da mesma espécie apresenta 90,2% na primeira, 9,4% na segunda e 0,4% dos resíduos na terceira região do gráfico de Ramachandran (Figura 2.29-C-D). Além disso, propriedades que descrevem a qualidade da estrutura cristalográfica do núcleo snRNP U1 humano encontram-se em regiões de menor percentil relativo a estruturas cristalográficas de resolução similar, enquanto que as mesmas propriedades para a estrutura do núcleo Sm de snRNP U4 humana encontram-se nas regiões de maior percentil, o que indica sua melhor qualidade (Figura 2.29-A-B).



**Figura 2.29 – Qualidade das estruturas cristalográficas do núcleo Sm de snRNPs U1 e U4 de *H. sapiens*.** A- e B- Propriedades das estruturas cristalográficas do núcleo Sm de snRNPs U1 e U4 de *H. sapiens*, respectivamente. C- e D- Gráficos de Ramachandran correspondentes. A qualidade dos modelos é medida pelas seguintes propriedades: Rfree, que mede a similaridade entre os fatores de estrutura calculado e observado, a pontuação de sobreposição corresponde ao número de pares de átomos incomumente próximos, outliers do gráfico de Ramachandran e de cadeia lateral corresponde àqueles resíduos de aminoácidos com ângulos ( $\phi, \psi$ ) e  $\chi$  fora das regiões favoráveis, outliers de RSRZ ( $>2$ ) correspondem aos resíduos de aminoácidos modelados que não apresentam um bom ajuste com os dados observados transformados no espaço real, e, finalmente, a qualidade do modelo da cadeia principal do RNA baseada na estatística e ângulos de torção normalmente apresentados por estruturas de RNA publicadas.

Fonte: Elaborada pelo autor.



Os modelos do núcleo Sm 1 (canônico) obtidos por modelagem comparativa foram avaliados segundo o valor de DOPE e o modelo com menor valor de energia foi utilizado para otimização de alças através dos métodos *Variable Target Function* (VTFM) e dinâmica molecular, implementados no programa. Também foi necessária a imposição de formação de estrutura secundária em algumas regiões, de acordo com uma análise do alinhamento tridimensional do molde com o modelo final e a comparação com a predição de estrutura secundária.

Antes da otimização, 86,0 % dos resíduos de aminoácidos do modelo estavam em regiões mais favoráveis do gráfico de Ramachandran, 12,1% nas regiões adicionalmente favoráveis, 0,9% em regiões generosamente favoráveis e 0,9% em regiões desfavoráveis. Após otimização, o resultado foi de 89,0% na primeira, 9,6% na segunda, 0,6% na terceira e 0,9% na quarta regiões (Figura 2.30-A1-A2). Tanto os resíduos em regiões generosamente favoráveis quanto os resíduos em regiões não favoráveis se encontram em grandes alças, principalmente na alça L4. Portanto, o modelo para o núcleo Sm canônico de *T. brucei* está de acordo com a estereoquímica de seus moldes de *H. sapiens*.

O servidor *ProSa*<sup>230</sup> indica uma pontuação, conhecida com *Z-score*, que mede a qualidade do modelo por meio da medida da energia total da estrutura com respeito a uma função distribuição de energia de conformações aleatórias, fornecendo uma faixa de valores característicos para proteínas nativas, comparadas às proteínas com estrutura conhecida. O Apêndice A2 mostra esta comparação para todas as cadeias do modelo do núcleo Sm 1. Em média, o *Z-score* foi de - 3, valor este que se encontra dentro do limite esperado para proteínas cuja estrutura é conhecida e que tem mesmo tamanho de sequência de aminoácidos. Além disso, o cálculo da energia média para cada resíduo de aminoácido da estrutura (Apêndice A2), feito considerando-se janelas de 10 e 40 resíduos em torno de cada posição na estrutura primária, revelou algumas regiões de alta energia, que correspondem basicamente às extremidades das cadeias polipeptídicas.

Para a modelagem do complexo que contém o dímero Sm15K/Sm16,5K em substituição a SmB/SmD3 (núcleo Sm 2), inicialmente, optou-se por alinhar o modelo obtido por alinhamento para este dímero com o modelo do *core* Sm canônico. Porém, foram observadas regiões de sobreposição de átomos dos monômeros das extremidades com o dímero Sm15K/Sm16,5K, principalmente em alças grandes, apesar das interfaces de interação apresentarem boa qualidade de modelagem. Isso foi devido à modelagem separada do dímero Sm15K/Sm16,5K e do pentâmero SmD1/SmD2/SmE/SmG/SmF, que não leva em consideração a flexibilidade das alças.

Deste modo, o alinhamento de Sm15K e Sm16,5K com proteínas SmD3/SmB do núcleo Sm humano, baseado naquele gerado pelo servidor *I-Tasser*, foi utilizado como arquivo de entrada no programa *Modeller 9.10*, apesar da baixa identidade de sequência de aminoácidos. Assim, foi possível modelar o complexo heptamérico usando os mesmos passos de modelagem comparativa feitos para o núcleo Sm canônico, com resultado bastante satisfatório. A avaliação da estereoquímica do modelo final pelo gráfico de Ramachandran é mostrada na Figura 2.30-B, que mostra 83,9 % dos resíduos de aminoácidos do modelo estavam em regiões mais favoráveis, 13,6% nas regiões adicionalmente favoráveis, 1,9% em regiões generosamente favoráveis e 0,6% em regiões desfavoráveis do gráfico de Ramachandran (Figura 2.30-B). Os resíduos presentes em regiões não favoráveis se encontram em grandes alças, principalmente na alça L4, como no modelo do complexo canônico. Tentativas de otimização não levaram a melhora significativa na qualidade do modelo.

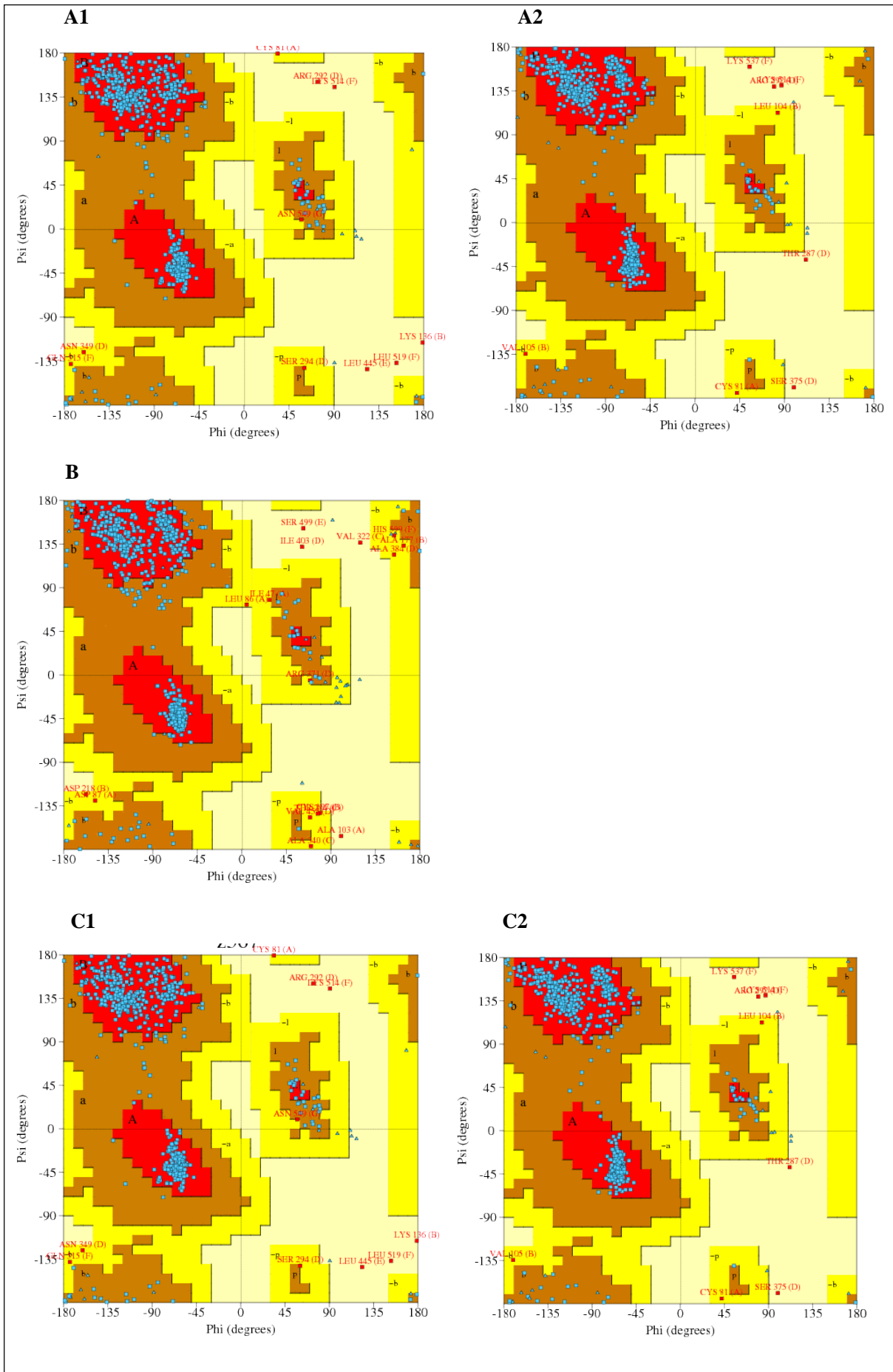
O resultado de *Z-score* para as cadeias polipeptídicas do modelo do núcleo Sm 2 foi, em média, em torno de -3, como no caso do núcleo Sm canônico. Entretanto, as cadeias correspondentes à variação Sm16,5K não apresenta boa qualidade, com *Z-score* igual -2,00 e a maioria dos resíduos de aminoácidos com alta energia média, o que já era esperado dada a baixa qualidade do alinhamento com a estruturas primária de SmD3 de *H. sapiens*.

Com base na estratégia bem sucedida de modelagem da variação do complexo canônico para Sm15K/Sm16,5K substituindo SmB/SmD3 (núcleo Sm 3), optou-se por modelar a variação SSm4 em substituição a SmD3 da mesma maneira. Porém, o alinhamento gerado pelo programa *I-Tasser* contra a proteína SmD3 não cobre toda a sequência de SSm4. Por outro lado, o alinhamento com a proteína Sm16,5K leva a maior cobertura. Portanto, para a modelagem deste complexo com variação SSm4 em lugar de SmD3, foi usada uma abordagem multi-moldes com as estruturas de códigos PDB 2PJO<sup>55</sup> e 2WZJ<sup>91</sup> e do modelo da variação Sm15K/Sm16,5K no lugar de SmB/SmD3, obtido anteriormente. Um modelo otimizado para esta nova variação foi obtido com êxito, para o qual é mostrado o gráfico de Ramachandran antes e depois de otimização (Figura 2.30-C1-C2).

Antes da otimização, 86,0% dos resíduos de aminoácidos do modelo estavam em regiões mais favoráveis do gráfico de Ramachandran, 12,1% nas regiões adicionalmente favoráveis, 0,9% em regiões generosamente favoráveis e 0,9% em regiões desfavoráveis. Após otimização, o resultado foi de 89,0% na primeira, 9,6% na segunda, 0,6% na terceira e 0,9% na quarta (Figura 2.30-C1-C2) e, como observado para os outros complexos modelados, neste caso s resíduos cuja geometria da cadeia principal não é favorável encontram-se principalmente na alça L4.

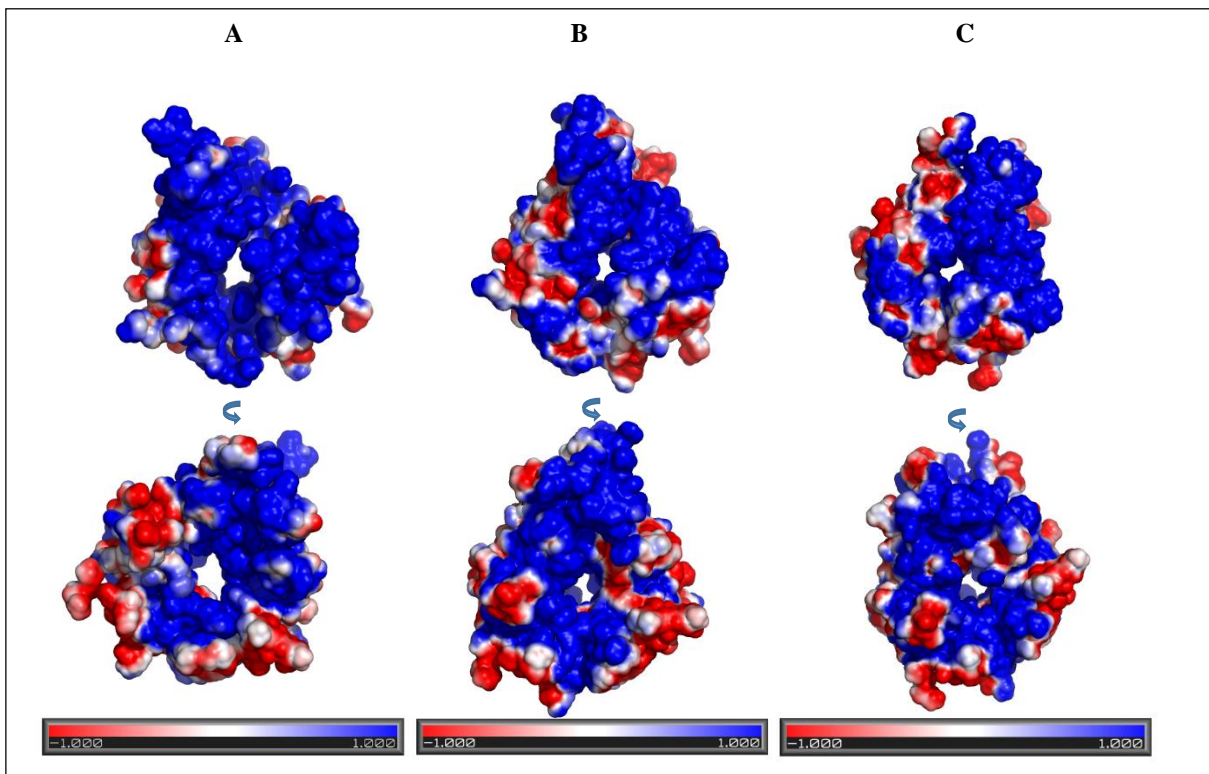
O resultado de *Z-score* para as cadeias polipeptídicas de SSm4 e SmB do modelo do núcleo Sm 3 indica qualidade de modelagem inferior, apesar de estar dentro do valor esperado para proteínas de mesmo tamanho cuja estrutura é conhecida. A análise de energia média por resíduo de SSm4 mostra que a maioria dos resíduos de aminoácidos exibe alta energia média, o que já era esperado dada a baixa qualidade do alinhamento com a estruturas primária de SmD3 de *H. sapiens*.

A sobreposição da estrutura tridimensional do modelo do núcleo Sm 1 de *T. brucei* com ambos os modelos cristalográficos dos núcleos Sm de snRNP U1 e U4 resulta em *rmsd* para a cadeia principal de cada complexo igual a 1,1 e 1,3 Å, respectivamente. Já para o núcleo Sm 2, os valores são de 1,2 e 2,1 Å, respectivamente, enquanto que, para o núcleo Sm 3, esses resultados são iguais a 1,2 e 1,8 Å, respectivamente. Os valores teóricos calculados pela equação (9) para *rmsd* das distâncias entre os átomos das cadeias principais do núcleo Sm canônico e das variações 2 e 3 de *T. brucei* com relação ao núcleo Sm canônico de *H. sapiens* são 1,4, 1,3 e 1,4, respectivamente. Os resultados mostram alta conservação do enovelamento do tipo Sm.



**Figura 2.30 – Avaliação da estereoquímica dos modelos dos complexos Sm de *T. brucei***, a partir do Gráfico de Ramachandran. A1- Antes e B2- depois de otimização do complexo canônico (1), B- do complexo 2, e C1- antes e C2- depois da otimização do complexo 3.  
Fonte: Elaborada pelo autor.

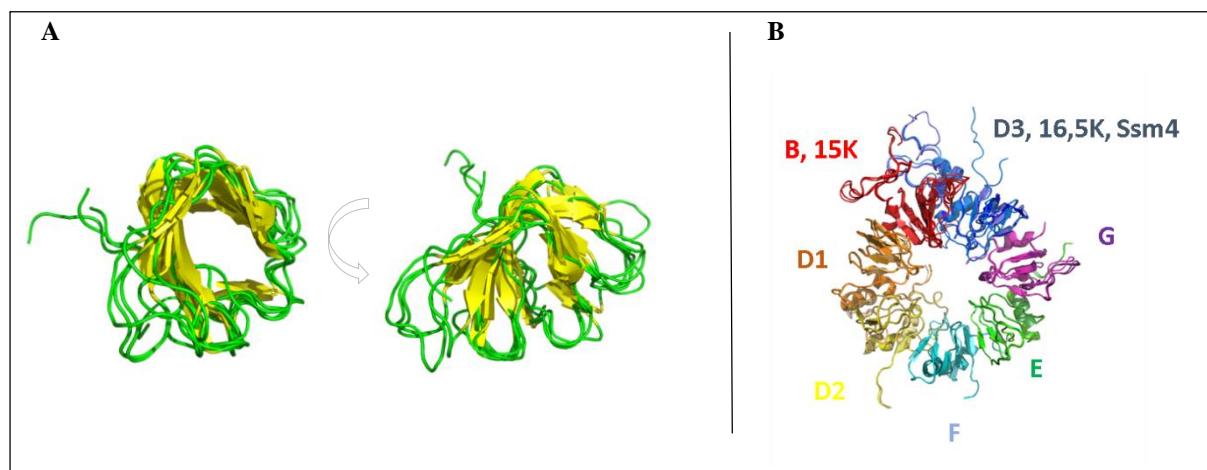
A Figura 2.31 mostra o potencial eletrostático de superfície calculado para os anéis Sm de *T. brucei* a pH 7,5 e pode-se identificar uma região altamente positiva para interação com os diferentes snRNAs no interior dos anéis.



**Figura 2.31 – Visualização da superfície eletrostática dos modelos 3D dos cores Sm** A-canônico (núcleo Sm 1), B- contendo a variação Sm15k/Sm16,5k (núcleo Sm 2) e C- contendo a variação SSm4/SmB (núcleo Sm 3). O potencial eletrostático é calculado em unidades de  $k_B T e_c^{-1}$  ( $\approx -25,9$  mV).  
Fonte: Elaborada pelo autor usando o programa APBS<sup>224</sup>.

Como comparação, a alça L4 é a primeira região de interação com RNAs ricos em adenina na proteína Hfq de *Bacteria*, homóloga de proteínas Sm.<sup>179</sup> A variação na sequência de aminoácidos desta alça modula o potencial eletrostático nesta face do toróide.<sup>179</sup> Nota-se que esta superfície é predominantemente positiva para a interação com as moléculas de snRNA, que são predominantemente negativas devido aos grupamentos fosfato expostos.

Adicionalmente, a Figura 2.32-A, a seguir, mostra a seção transversal do motivo sanduiche- $\beta$  de todas as proteínas Sm de *T. brucei* alinhadas utilizando-se o programa PyMOL (*Schrödinger, LLC*). Adicionalmente, apresenta-se a sobreposição da estrutura quaternária dos três núcleos Sm de *T. brucei* na Figura 2.32-B. Note a variação nas regiões de alça (L4) no interior dos anéis e nas posições das proteínas SSm4, Sm15K e Sm16,5K.



**Figura 2.32 – Proteínas Sm de *T. brucei*.** A- Visão transversal e lateral do motivo- $\beta$  característico das proteínas Sm de *T. brucei*, alinhadas no programa *PyMOL* (*Schrödinger, LLC*) e B- sobreposição dos modelos dos variantes dos núcleos Sm de *T. brucei*.

Fonte: Elaborada pelo autor usando o programa *PyMOL* (*Schrödinger, LLC*).

Os valores de área de interface entre os monômeros que formam o anel heteroheptamérico e a respectiva energia livre de interação, calculados pelo servidor *PISA*<sup>223</sup>, variam entre 750 e 1100 Å<sup>2</sup> e -7,0 e -12,0 kcal/mol, respectivamente (Tabela 2.8). Como observado anteriormente para o dímero Sm15K/Sm16,5K modelado por alinhamento, as interfaces são formadas principalmente por interações de hidrogênio fracas entre cadeias principais e hidrofóbicas entre cadeias laterais da fita- $\beta$ 5 do monômero *i* com a fita- $\beta$ 4 do monômero *i*+1, formando uma folha anti-paralela. Algumas pontes salinas auxiliam na estabilização das interfaces, principalmente nas extremidades.

A estrutura semelhante a uma barril- $\beta$  torcido formada por 7 monômeros é estável e está de acordo com o que se foi observado para o motivo  $\beta$ -*propeller*, para o qual é possível mostrar geometricamente que um motivo de 7 membros é mais estável que outras possíveis combinações.<sup>91</sup> Os valores de área de superfície e energia livre de interação obtidos estão de acordo com aqueles verificados para a manutenção da estrutura quaternária do núcleo de snRNP U4 humano<sup>91</sup> (código PDB 2Y9A).

**Tabela 2.8 – Propriedades das interfaces de interação entre as proteínas Sm de *T. brucei*.**

	Núcleo Sm 1		Núcleo Sm 2		Núcleo Sm 3	
	Área	$\Delta G$	Área	$\Delta G$	Área	$\Delta G$
	( $\text{\AA}^2$ )	( $\text{kcal mol}^{-1}$ )	( $\text{\AA}^2$ )	( $\text{kcal mol}^{-1}$ )	( $\text{\AA}^2$ )	( $\text{kcal mol}^{-1}$ )
SmD3 – SmB	880,2	-15,9	-	-	-	-
Sm16,5K – Sm15K	-	-	1.229,7	-12,4	-	-
SSm4 – SmB	-	-	-	-	880,2	-15,9
Sm15K – SmD1	-	-	915,8	-9,9	-	-
SmB – SmD1	916,2	-6,8	-	-	916,2	-6,8
SmD1 – SmD2	1.088,4	-12,2	1.182,7	-8,2	1.088,4	-12,2
SmD2 – SmF	888,8	-18,2	717,8	-9,6	888,8	-18,2
SmF – SmE	835,6	-16,6	926,6	-11,4	835,6	-16,6
SmE – SmG	1.070,5	-12,8	833,6	-9,0	1.070,5	-12,8
SmG – SmD3	642,6	-7,9	-	-	-	-
SmG – Sm16,5K	-	-	829,8	-10,3	-	-
SmG – SSm4	-	-	-	-	642,6	-7,9

Fonte: Elaborada pelo autor a partir dos resultados do servidor PISA<sup>223</sup>.

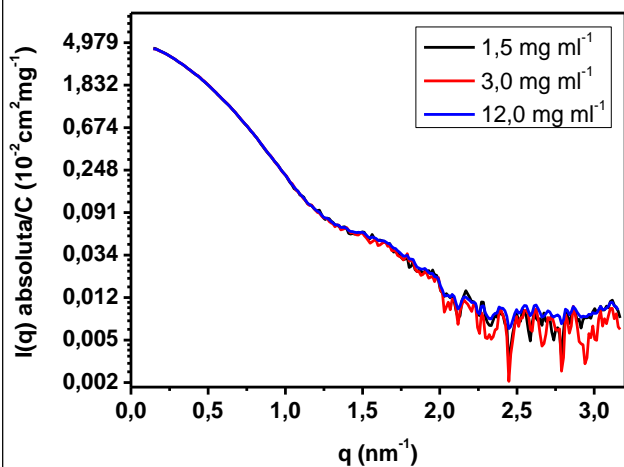
As propriedades hidrodinâmicas dos modelos dos complexos Sm serão mostradas e analisadas juntamente com os valores experimentais obtidos por espalhamento de raios-X a baixo ângulo, a seguir.

#### 2.4.8 Caracterização dos núcleos Sm de snRNPs U2, U4 e U5 na ausência de snRNAs por meio de espalhamento de raios-X em baixo ângulo (SAXS)

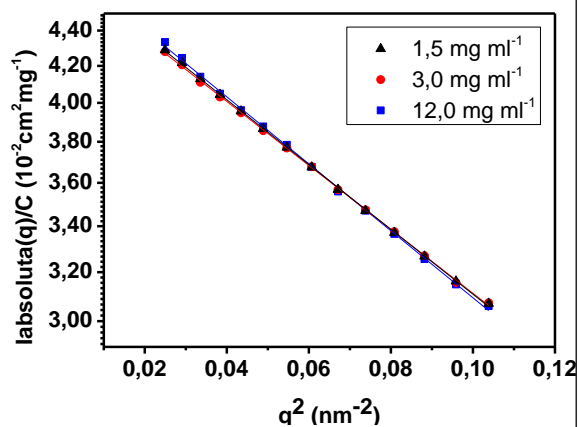
Primeiramente, foi medido o espalhamento de raios-X a baixo ângulo de água destilada, que foi descontado do espalhamento do porta-amostras no vácuo, resultando em um valor constante igual a  $8,26 \times 10^{-5} \text{ cm}^{-1}$ . Este resultado foi usado para se calcular o espalhamento de raios-X nas amostras dos complexos Sm de *T. brucei* ajustado para a escala absoluta, mostrado na Figura 2.33 (A1, B1 e C1) para todas as concentrações medidas. Os gráficos de Guinier e de Kratky correspondentes também são apresentados (Figura 2.33 – Guinier: A2, B2 e C2, Kratky: D, E e F).

Núcleo Sm 1

A1

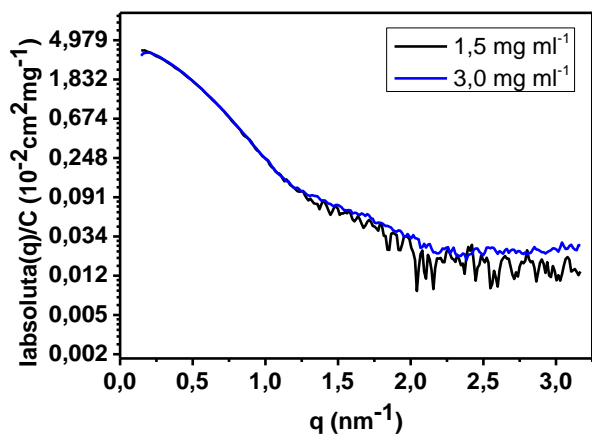


A2

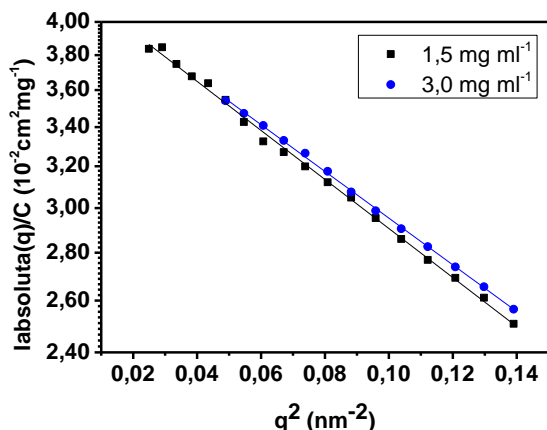


Núcleo Sm 2

B1

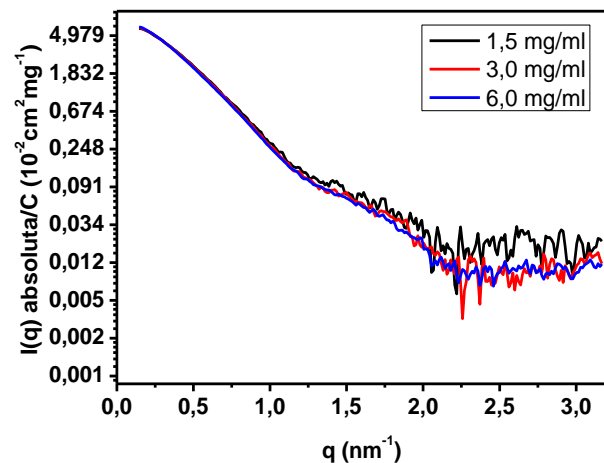


B2

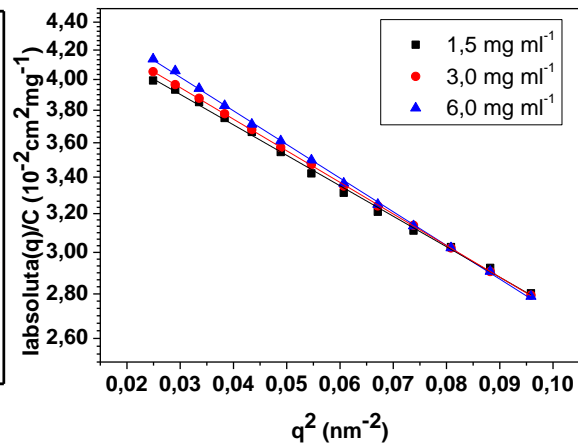


Núcleo Sm 3

C1



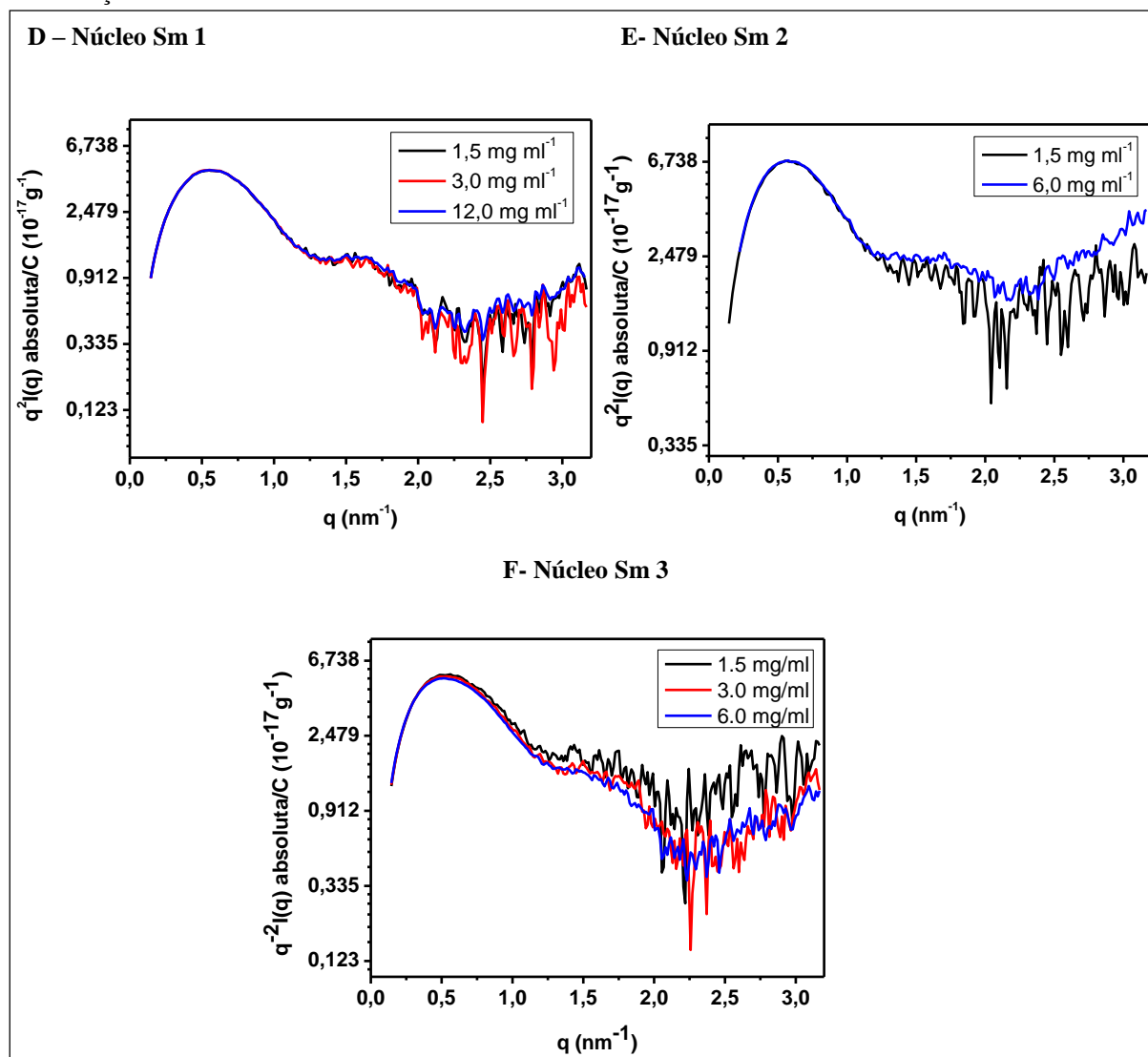
C2



continua



continuação



**Figura 2.33** – SAXS de complexos Sm de *T. brucei*. A1, B1 e C1- Curvas de espalhamento, A2, B2 e C2- Gráficos de Guinier, e D, E e F- Gráficos de Kratky dos núcleos Sm 1, 2 e 3 de *T. brucei*, respectivamente.

Fonte: Elaborada pelo autor usando o programa *Origin PRO 8.0* (*Origin Lab Corporation*).

A partir da Figura 2.33 é possível observar que todas as curvas de espalhamento apresentam o mesmo comportamento com um máximo local na região  $1,25 \text{ nm}^{-1} < q < 2,25 \text{ nm}^{-1}$  entre dois mínimos locais indicativos de esfericidade. Observa-se, também, que os dados de concentração mais baixa apresentam menor relação sinal/ruído em ângulos de espalhamento maiores. Os gráficos de Kratky (Figura 2.33-D-F) apresentam um máximo local por volta de  $0,5 \text{ nm}^{-1}$  e outro entre  $1,5$  e  $2,0 \text{ nm}^{-1}$ , o que aponta para estruturas globulares com uma pequena extensão. Além disso, a inclinação positiva a partir de  $q = 2,5 \text{ nm}^{-1}$  em todas as concentrações medidas para os três diferentes complexos indica a presença de

pequenas regiões flexíveis, provavelmente devido à presença de grandes alças apontando para a superfície.

Os gráficos de Guinier para amostras mais concentradas apresentam uma maior inclinação, o que indica possível formação de (hetero-)oligômeros maiores em solução. Portanto, o cálculo de  $R_g$  foi realizado apenas para os resultados de concentração mais baixa para cada núcleo Sm de *T. brucei*. Entretanto, como os dados a mais alta concentração têm maior relação sinal/ruído, estes foram usados para o cálculo das curvas de distribuição de distâncias. Como se observa na Figura 2.34, as curvas  $p(r)$  apresentam um máximo na região  $4 \text{ nm} < r < 5 \text{ nm}$ . Além disso, verifica-se uma assimetria na curva, o que indica uma assimetria na estrutura quaternária dos complexos.

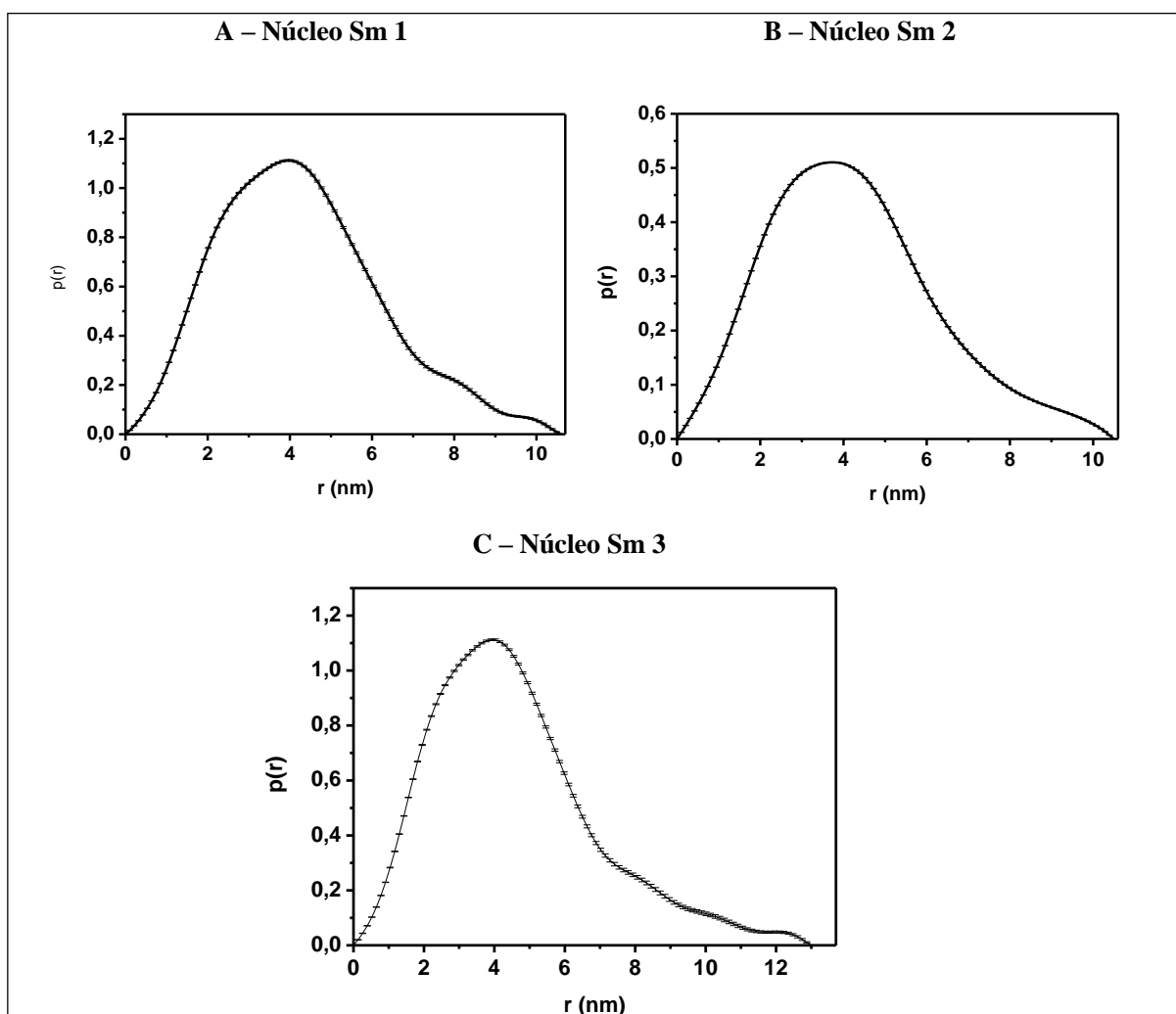


Figura 2.34 – Curvas de distribuição de distâncias ( $p(r)$ ) para os complexos Sm A-1, B-2 e C-3.

Fonte: Elaborada pelo autor usando o programa *Origin PRO 8.0* (*Origin Lab Corporation*).

A Tabela 2.9 mostra os parâmetros estruturais obtidos para os três complexos Sm de *T. brucei* a partir dos dados de SAXS, em comparação com aqueles estimados para os modelos obtidos por homologia descritos na seção 2.4.7.

**Tabela 2.9 – Parâmetros estruturais calculados por SAXS, cromatografia de exclusão molecular analítica e a partir dos modelos de estrutura 3D dos complexos Sm de *T. brucei*.**

	Núcleo Sm 1	Núcleo Sm 2	Núcleo Sm 3
<b>Parâmetros estruturais</b>			
I(0) ( $10^{-2} \text{ cm}^2 \text{ mg}^{-1}$ )/C, Guinier	4,74 ± 0,01	4,25 ± 0,04	4,75 ± 0,03
I(0) ( $10^{-2} \text{ cm}^2 \text{ mg}^{-1}$ )/C, GNOM	4,52 ± 0,04	4,21 ± 0,03	5,20 ± 0,04
R <sub>g</sub> (nm), Guinier	3,63 ± 0,01	3,37 ± 0,01	3,90 ± 0,02
R <sub>g</sub> (nm), GNOM	3,31 ± 0,01	3,31 ± 0,01	3,95 ± 0,02
D <sub>máx</sub> (nm)	10,6	10,5	12,9
Volume de Porod (nm <sup>3</sup> )	158,50	131,0	189,3
<b>Parâmetros estruturais obtidos a partir dos modelos</b>			
R <sub>g</sub> (nm), R <sub>H</sub> (nm)*	2,86 (3,72)	29,4 (38,2)	3,21 (4,17)
D <sub>máx</sub> (nm)	9,51	9,60	12,00
Volume de Porod (nm <sup>3</sup> )	140,4	140,0	154,5
Volume DAM (nm <sup>3</sup> )	162,5	151,9	185,5
Volume seco (nm <sup>3</sup> )**	91,7	95,1	104,8
<b>Parâmetro estrutural obtido a partir de cromatografia de exclusão molecular</b>			
R <sub>H</sub> (nm)	3,9 ± 0,8	3,4 ± 0,7	3,9 ± 1,0

\*Raio hidrodinâmico estimado a partir da seguinte relação:  $R_H = 1,3 R_g$

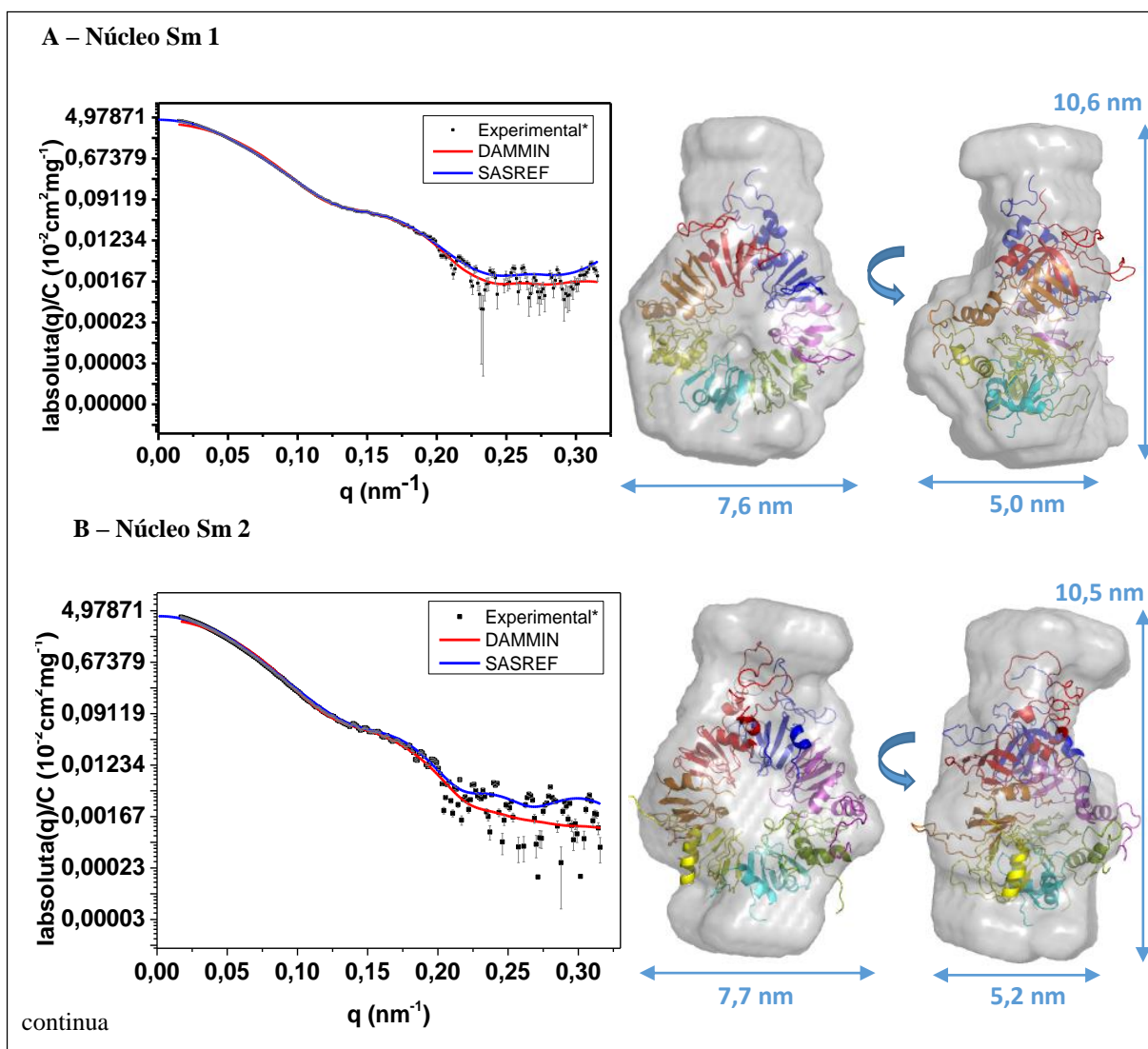
\*\* Calculado a partir das sequências de aminoácidos (massa molecular predita) de cada complexo.

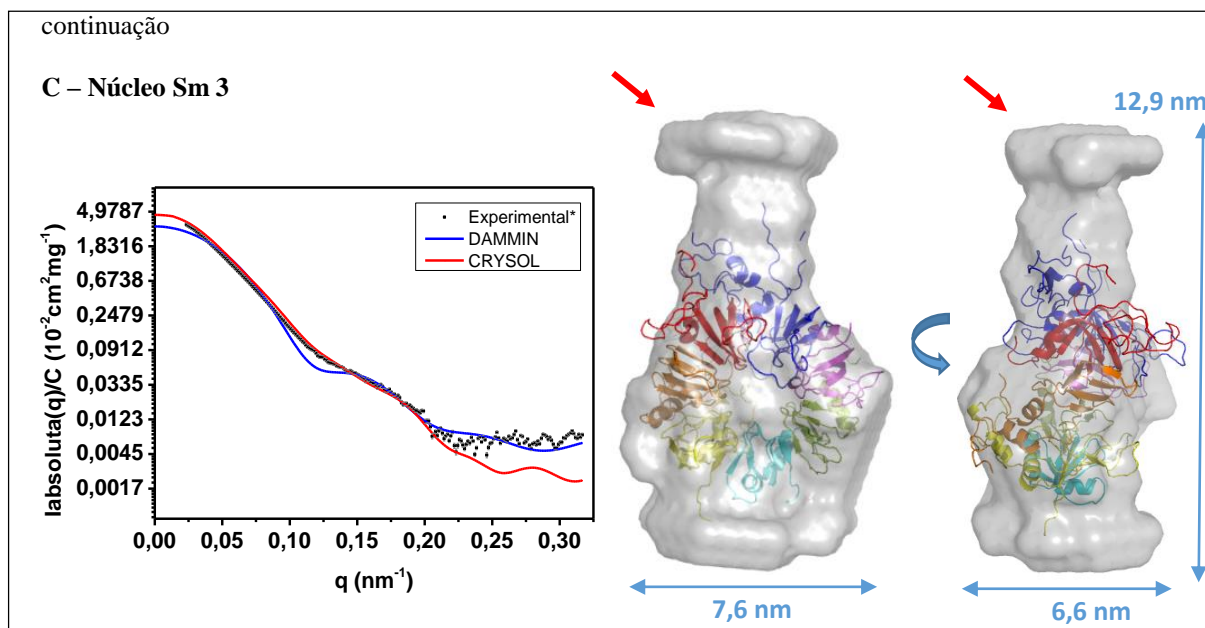
Fonte: Elaborada pelo autor.

Observa-se, na Tabela 2.9, que a variação entre os valores de raio de giro obtidos para cada complexo Sm pela aproximação de Guinier e pelo programa *GNOM*<sup>235</sup> é de menos que 5%. Observa-se, também, que o núcleo 3, que possui a variação SSm4/SmB, tem um raio de giro maior que os outros complexos, o que está de acordo com a presença da proteína SSm4 no lugar da proteína SmD3, já que a proteína SSm4 é consideravelmente maior que SmD3 e Sm16,5K.

Por meio do programa *CRY SOL*<sup>239</sup>, foi possível comparar as curvas de espalhamento preditas para os modelos de estrutura quaternária dos complexos Sm de *T. brucei* com as curvas de espalhamento experimentais. O valor da discrepância  $\chi$  entre a curva experimental e a curva teórica foi de 6,7 para o núcleo Sm 1, 6,6 para o núcleo Sm 2 e 7,6 para o núcleo Sm 3. Deste modo, optou-se por seguir com refinamento de corpo rígido no programa *SASREF*<sup>240</sup> e o resultado foi  $\chi$  igual a 4,9 para o núcleo Sm 1 e 4,8 para o núcleo Sm 2, sendo que não foi possível refinar o modelo do núcleo Sm 3 (Figura 2.35). Quanto mais próximo de 1, mais semelhante é a curva teórica com a experimental.

Foram realizadas 10 reconstruções *ab initio* independentes assumindo simetria P1 no programa *DAMMIN*<sup>236</sup>, que foram promediadas no programa *DAMAVR*<sup>237</sup>. Os modelos gerados para os envelopes moleculares dos três complexos Sm foram alinhados com seus respectivos modelos de estrutura quaternária, usando o programa *SUPCOMB*<sup>238</sup> (Figura 2.35), onde a discrepância espacial normalizada (do inglês, *normalized spatial discrepancy, NSD*) é minimizada.





**Figura 2.35– Envelopes moleculares dos complexos Sm de *T. brucei*.** A, B e C- Comparação entre curvas de espalhamentos experimentais e aquelas previstas para os modelos dos complexos 1,2 e 3, respectivamente, geradas pelo programa *CRY SOL*<sup>239</sup> e pelo programa *DAMMIN*<sup>236</sup>. Os envelopes moleculares médios gerados pelo programa *DAMMIN*<sup>236</sup> para os complexos 1, 2 e 3 são mostrados à direita. Visualização gerada pelo programa *PyMOL*(*Schrödinger, LLC*). São representados o envelope de SAXS (cinza) para cada complexo e as proteínas Sm (SmD3, Sm16,5K e SSm4 – azul, SmG – roxo, SmE – verde, SmF – azul-claro, SmD2 – amarelo, SmD1 – laranja e SmB ou Sm15K – vermelho). As setas vermelhas indicam a região do envelope molecular onde pode ser modelada a região C-terminal de SSm4.

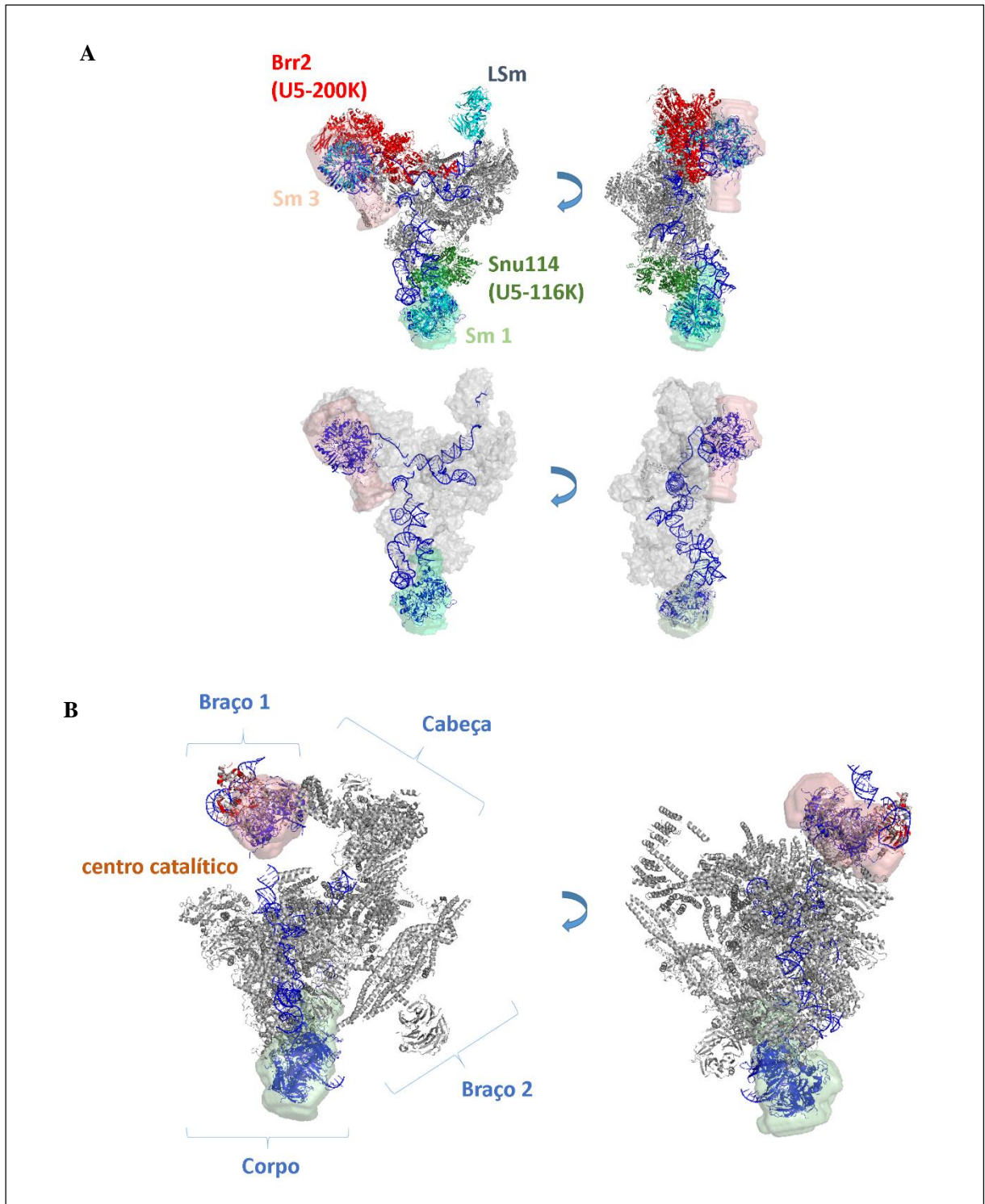
Fonte: Elaborada pelo autor usando os programas *Origin PRO 8.0* (*Origin Lab Corporation*) (A-C) e *PyMOL* (*Schrödinger, LLC*) (D-F).

Observa-se, na Figura 2.35, uma boa sobreposição entre os modelos obtidos por modelagem comparativa de todos os complexos Sm de *T. brucei* e os respectivos envelopes de SAXS, bem como um bom alinhamento entre as curvas de espalhamento experimental e a curvas de espalhamento teóricas destes modelos. No caso do núcleo Sm 3, a maior discrepância entre os dados pode ser devida à modelagem pouco precisa da região C-terminal da proteína SSm4, prevista como hélice- $\alpha$ , porém com baixa identidade de alinhamento com a proteína SmD3 de *T. brucei*. Esta região pode ser verificada na Figura 2.35-C, onde há espaço para o enovelamento de estruturas em forma de hélice- $\alpha$ .

Os valores mínimos de *NSD* para o alinhamento dos modelos dos núcleos Sm 1, 2 e 3 nos respectivos envelopes de SAXS foram iguais a 1,3, 1,4 e 1,8, sendo que quanto mais próximo de zero, mais similares os modelos. A discrepância entre as estruturas de modelagem comparativa e os envelopes resultantes é devida às regiões cujo alinhamento não existe ou é pobre, que foram excluídas dos modelos obtidos por modelagem por homologia e resultaram em buracos nos envelopes (Figura 2.35).

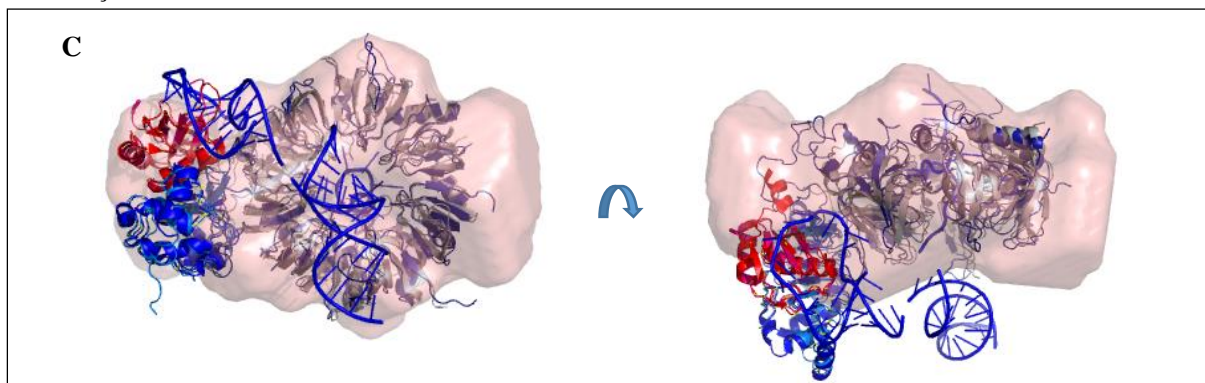
A comparação dos envelopes de SAXS dos três núcleos Sm variantes de *T. brucei* com as estruturas da tri-snRNP U4/U6.U5<sup>94</sup> e do complexo do spliceossomo que contém o íntron na forma de laço (pós-spliceossomo)<sup>54</sup> de leveduras, recentemente resolvidas por criomicroscopia eletrônica, indica as regiões onde as variações nos anéis Sm de *T. brucei* devem se posicionar. Para tal, assume-se a conservação evolutiva da estrutura dos complexos do spliceossomo. A sobreposição do envelope molecular do anel Sm 3 com a região que interage com o snRNA U4 na extremidade esquerda da tri-snRNP U4/U6.U5 (que corresponde ao corpo do spliceossomo) revela que existe uma região do envelope que não se alinha com proteínas Sm (Figura 2.36-A, superfície colorida em rosa). Acredita-se que esta região corresponde às alças extras e estruturas de hélice- $\alpha$  não modeladas das extremidades N- e C-terminal de SSm4. Esta região encontra-se em contato direto com a extremidade C-terminal de U5-200K e, possivelmente, na região de interação entre U5-220K e U5-116K (Figura 2.36-A).

Por outro lado, as proteínas Sm do núcleo 1 (canônico) interagem diretamente com U5-116K, na base do corpo de tri-snRNP U4/U6.U5 e do spliceossomo. A região extra do envelope molecular obtido por SAXS apresenta sobreposição com o snRNA U5 em ambas as estruturas (Figura 2.36-A, superfície colorida em verde claro), o que indica que as alças flexíveis de SmD3 e SmB (não modeladas) devem atuar na estabilização da estrutura desta molécula de RNA. Já o envelope de SAXS do núcleo 2, variante em *T. brucei*, foi localizado na região do braço 1 do spliceossomo (Figura 2.36-B, superfície colorida em vermelho). A região extra no envelope molecular referido, que corresponde às alças de Sm16,5K e Sm15K não modeladas, apresenta sobreposição com a região N-terminal da proteína U2A'. Este resultado indica que esta região de U2A' pode interagir diretamente com Sm16,5K e Sm15K (Figura 2.36-C), para a estabilização da estrutura de snRNP U2.



continua

continuação



**Figura 2.36** – Sobreposição dos envelopes moleculares dos complexos Sm de *T. brucei* nos modelos de tri-snRNP U4/U6.U5 e do complexo spliceossomo que contém o íntron na forma de laço. A- Parte da estrutura de criomicroscopia de tri-snRNP de *S. cerevisiae* está representada (Brr2 em vermelho, Prp8 em cinza, Snu114 em verde e proteínas Sm/LSm em azul, snRNAs U4, U6 e U5 em azul). O envelope molecular do núcleo Sm 3 de *T. brucei* está representado em rosa e do núcleo Sm canônico (de snRNP U5) é mostrado em ciano. B- Representação da estrutura de criomicroscopia do spliceossomo que contém o íntron na forma de laço, em cinza. O envelope molecular do núcleo Sm 2 de *T. brucei* é apresentado em rosa, e o núcleo Sm canônico (de snRNP U5), em ciano. Os snRNAs U2, U4, U6 e U5 são mostrados em azul. C- Região de snRNP U2 do spliceossomo. Apresenta-se, ainda, a sobreposição da estrutura do núcleo Sm de levedura com aquela modelada para *T. brucei*, juntamente com o envelope molecular deste complexo. As estruturas cristalográficas de U2A'/U2B'' também foram sobrepostas, evidenciando-se as possíveis regiões de interação com snRNA U2 e com proteínas Sm.

Fonte: Elaborada pelo autor a partir de estruturas publicadas por Nguyen et al<sup>94</sup>; Yan et al<sup>54</sup>.

Finalmente, o cálculo da massa molecular de cada complexo a partir dos dados de SAXS pelo primeiro método, que utiliza a curva de espalhamento em escala absoluta, resultou em valores subestimados, enquanto que o segundo método, pelo programa *SAXSMoW*<sup>242</sup>, resultou em valores superestimados. Nota-se que os resultados obtidos para massa do núcleo Sm 2 se aproximam mais do valor esperado, assim como se observou para a massa calculada a partir dos dados de cromatografia de exclusão molecular.



**Tabela 2.10** – Comparação entre massas moleculares calculadas por SAXS, cromatografia de exclusão molecular e a partir das sequências de aminoácidos dos complexos Sm de *T. brucei*.

	Núcleo Sm 1	Núcleo Sm 2	Núcleo Sm 3
$MM_{\text{experimental,A}}$ (kDa)	66,0 (-13%)	59,1 (-25%)	66,1 (-24%)
$MM_{\text{experimental,B}}$ (kDa)	99,1 (30%)	79,4 (1%)	97,8 (13%)
$MM_{\text{experimental,C}}$ (kDa)	81,2 (7%)	76,0 (-4%)	92,8 (-9%)
$MM_{\text{Cromatografia de Exclusão Molecular}}$ (kDa)	$99 \pm 40$ (30%)	$73 \pm 30$ (-7%)	$102 \pm 40$ (19%)
$MM_{\text{predita}}$ (kDa)	75,8	78,6	86,6

\* Entre parênteses encontram-se os desvios relativos ao valor teórico calculados com  $\Delta = (MM_{\text{experimental}} - MM_{\text{predita}}) / MM_{\text{predita}}$ .  
Fonte: Elaborada pelo autor.

## 2.5 Discussão

As partículas ribonucleoproteicas são as unidades básicas para a associação do spliceossomo no pré-mRNA. Em tripanossomatídeos, a composição do spliceossomo varia em termos de proteínas das snRNPs e de fatores associados, que podem estar relacionados especificamente à catálise do SL *trans-splicing*. As proteínas Sm formam um anel em torno de cada snRNA e atuam como sinalização para transporte nuclear de subcomplexos pré-formados de snRNPs.<sup>20</sup> Como a maioria das proteínas de *T. brucei*, as proteínas Sm apresentam baixa identidade de sequência de aminoácidos com as ortólogas bem caracterizadas em humanos e leveduras. As sequências reconhecidas por metilases na extremidade C-terminal, por exemplo, não são conservadas. Entretanto, a distribuição de cargas de modo a resultar em uma superfície de potencial eletrostático positivo no interior do anel em pH fisiológico está conservada, propiciando a interação de uma molécula de RNA, carregada negativamente.

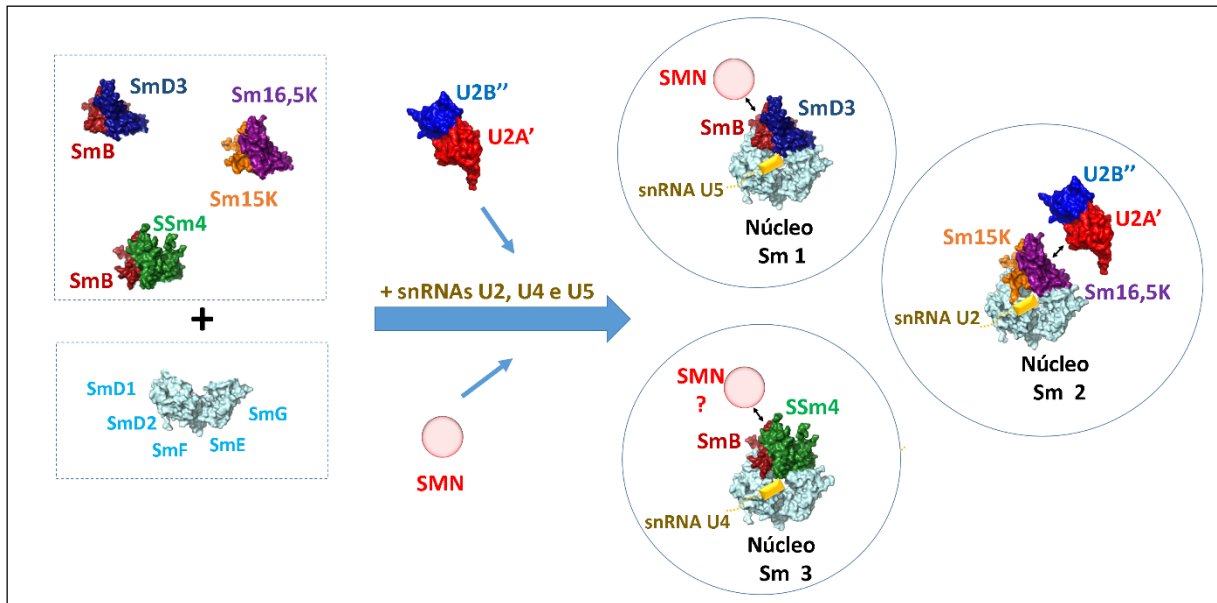
Dentre as proteínas específicas do spliceossomo de *T. brucei*, as variações mais curiosas estão nos núcleos de proteínas Sm das snRNPs U2 e U4. As proteínas SmD3 e SmB são substituídas por Sm16,5K e Sm15K na partícula U2, enquanto a proteína SSm4 está no lugar de SmD3 na partícula U4<sup>98, 185</sup>. A comparação dos sítios de interação de proteínas Sm com snRNAs publicados recentemente para snRNPs U1<sup>62</sup> e U4<sup>91</sup> de *H. sapiens* mostra que os resíduos de SmD3 humana em posição para interação com snRNA U4 são conservados em SSm4 e não em SmD3 de *T. brucei*. O mesmo não acontece para as proteínas Sm16,5K e

Sm15K. Entretanto, detalhes em nível atômico ainda são necessários para a descrição das interações que mantêm a estrutura de tais complexos moleculares.

Neste capítulo, os núcleos das snRNPs U2, U4 e U5 de *T. brucei* foram analisados. Primeiramente, foi necessário coexpressar e copurificar por cromatografias de afinidade e de exclusão molecular os subcomplexos variantes SSm4/SmB e Sm15K/Sm16,5K, além dos canônicos SmD3/SmB, SmD1/SmD2, SmE/SmF/SmG de *T. brucei*. O subcomplexo U2A'/U2B'' recombinante de *T. brucei* também foi copurificado. Além disso, as regiões snRNAs U2, U4 e U5 deste organismo necessárias para interação com estas proteínas foram obtidas por meio de transcrição *in vitro*. Espectros de dicroísmo circular confirmaram formação de hélice de RNA tipo A em cada caso. Modificações pós-transcricionais no snRNA U2 de eucariotos derivados, incluindo metilações em resíduos específicos, também são importantes para seu enovelamento correto e sua função no spliceossomo.<sup>106</sup> Entretanto, nenhuma modificação foi descrita para o snRNA U2 de *T. brucei* até o momento. Adicionalmente, bandas únicas em eletroforese nativa eliminam a possibilidade de formação *in vitro* de oligômeros de RNA na faixa de concentrações utilizadas.

A interação das proteínas dos três variantes dos núcleos Sm de *T. brucei* e os snRNAs U2, U4 e U5 cognatos resultou em constantes de dissociação da ordem de dezenas de nanomolares, indicando energia de interação na mesma ordem de grandeza relatada para os complexos Sm de *H. sapiens*. Além disso, observou-se que qualquer um dos três complexos Sm interage com o snRNA U2 e a seleção do variante correto do núcleo Sm para interação com snRNA U2 parece ocorrer a partir da interação com U2A'/U2B''. Experimentos de EMSA mostraram que U2A'/U2B'' interage com o núcleo Sm 2 (1 SmD1/SmD2 : 1 SmE/SmF/SmG : 1 Sm15K/Sm16,5K) já associado ao snRNA U2. Por outro lado, quando o núcleo Sm 2 é titulado em amostra do subcomplexo U2A'/U2B''/snRNA U2 pré-formado, a interação não é observada. Esses resultados levam à proposição de um modelo de associação do núcleo da snRNP U2 de *T. brucei*, em que as proteínas Sm do núcleo 2 formam um subcomplexo com snRNA para então U2A'/U2B'' interagem especificamente com este subcomplexo pré-formado (Figura 2.37).

Como descrito no capítulo 1, na seção 1.3.2.4, a formação de snRNP U2 de *T. brucei* não requer a proteína SMN, que atua como cofator para a associação dos outros núcleos Sm<sup>120</sup>,<sup>186</sup> e é um sinal para o transporte nuclear, como já mostrado em leveduras.<sup>145</sup> Neste caso, a variação Sm15K/Sm16,5K em *T. brucei* poderia atuar como sinal para tal, apesar de nenhum experimento *in vivo* até o momento ter apontado para este cenário.



**Figura 2.37 – Modelo de biogênese de núcleos Sm de *T. brucei*.** A associação das proteínas Sm aos snRNAs U4 e U5 depende da proteína SMN, que interage com SmB. Propõe-se que U2A'/U2B'' seja suficiente para garantir a especificidade de Sm15K/Sm16,5K pelo snRNA U2 em detrimento dos snRNAs U4 e U5.

Fonte: Elaborada pelo autor a partir dos resultados do capítulo 2 e Jae et al<sup>186</sup>.

A proteína U2B'' de *T. brucei* conserva o primeiro domínio RRM para interação com a alça de SL4 do snRNA U2, enquanto a proteína U2A' conserva os seis domínios LLR, que se enovelam formando uma folha- $\beta$  para interação direta com U2B''. A região N-terminal de U2A' possivelmente interage com as proteínas Sm16,5K/Sm15K do anel Sm pré-formado, porém ainda não se conhece as interfaces de interação. Também não está claro se a interação de U2A'/U2B'' com o subcomplexo composto pelas proteínas Sm do núcleo 2 interagindo com o snRNA U2 ocorreria ainda no citoplasma, sendo que a proteína U2A' de *T. brucei* foi identificada majoritariamente no núcleo.<sup>78, 249</sup>

Experimentos de cristalização de amostras de reconstrução *in vitro* do núcleo Sm 2 na presença de snRNA foram planejados com o objetivo de se investigar esta possível interação. Entretanto, tais ensaios resultaram em cristais contendo apenas U2A'/U2B'' de *T. brucei*. A estrutura deste subcomplexo, na ausência de snRNA, é similar àquela publicada para U2A'/U2B''/snRNA U2 de *H. sapiens*<sup>68</sup>, com até 1,0 Å de *rmsd* entre as posições atômicas nas cadeias principais correspondentes. Algumas diferenças foram observadas, principalmente na região de interação com o snRNA U2, onde a alça localizada entre as fitas- $\beta$ 2 e  $\beta$ 3 de U2B'' deve se movimentar para acomodar a alça da molécula de RNA.

A estrutura tridimensional das proteínas Sm na forma de sanduíche- $\beta$  altamente torcido e a estrutura quaternária na forma de anel dos complexos Sm de *T. brucei* foram modeladas com base na sua homologia com ortólogas cujas estruturas são conhecidas experimentalmente.

Apesar da baixa identidade de sequência de aminoácidos, a estrutura tridimensional é bem conservada. Foi possível observar que as alças L4 de tamanho variável das estruturas das proteínas Sm se encontram apontadas para o interior do anel. Além disso, as alças L4 contêm aminoácidos carregados positivamente e modulam o potencial eletrostático no interior dos anéis Sm para interação com os snRNAs em pH fisiológico. A diferença principal entre os três núcleos Sm variantes em *T. brucei* encontra-se nas extremidades N- e C-terminal e nas alças L4 das proteínas SmD3, SmB, Sm16,5K, Sm15K e SSm4. Estas regiões de alças devem estar relacionadas a interações com outras proteínas para formação de partículas ribonucleoproteicas estáveis no trans-spliceossomo.

Adicionalmente, a reconstrução *in vitro* dos núcleos Sm na ausência de snRNA também foi bem-sucedida, como observado em cromatografia de exclusão molecular. Deste modo, foi possível analisar a estrutura em baixa resolução dos núcleos Sm por meio de experimentos de espalhamento de raios-X a baixo ângulo (SAXS). Os parâmetros hidrodinâmicos e de massa molecular obtidos são comparáveis com aqueles calculados para os modelos de estrutura quaternária e também com os obtidos a partir de cromatografia de exclusão molecular analítica. O envelope molecular do núcleo Sm 3 apresenta uma porção mais estendida que a dos outros complexos, que corresponde à região onde as extremidades N- e C-terminal da proteína SSm4 se encaixam.

A comparação dos envelopes de SAXS para os variantes dos núcleos Sm com as estruturas do tri-snRNP U4/U6.U5 e do complexo spliceossomal pós-catalítico que contém o íntron na forma de laço publicados recentemente mostra o anel canônico interagindo com a proteína U5-116K na base do tri-snRNP U4/U5.U6 e do corpo do spliceossomo. SmD3 e SmB estão em posições voltadas para interação com a hélice do snRNA U5. Por outro lado, o anel variante que contém Sm15K e Sm16,5K encontra-se no braço 1 do spliceossomo, na entrada do sítio ativo. As proteínas variantes apontam para U2A', indicando possível interface de interação na região N-terminal desta proteína. Adicionalmente, o núcleo Sm que contém SSm4 se localiza entre os dois cassetes helicase/ATPase de U5-200K, sendo que a superfície do envelope de SAXS correspondente às extremidades de SSm4 aponta para a região entre o C-terminal de U5-200K e U5-220K.

Novos estudos estruturais serão necessários para se compreender os detalhes das interações que mantêm os núcleos de snRNPs de *T. brucei* em comparação com aquelas observadas no spliceossomo do hospedeiro humano, na busca por pequenas moléculas moduladoras seletivas e específicas para as variações dos núcleos Sm de tripanossomatídeos.

## CAPÍTULO 3

### **ANÁLISE BIOQUÍMICA DA PROTEÍNA U5-200K DE *T. brucei* E DE SEUS PARCEIROS DE INTERAÇÃO: IMPLICAÇÃO NA BIOGÊNESE DE snRNP U5**

Neste capítulo serão apresentados os resultados de expressão homóloga de U5-200K de *T. brucei* completa e truncada no seu cassete helicase/ATPase N-terminal ligadas a uma extensão C-terminal de epítomos de proteínas A e C para purificação em *tandem*. As proteínas copurificadas indicam a formação de um subcomplexo de snRNP U5 ainda no citoplasma entre U5-220K, U5-200K, U5-116K, U5-40K e U5-Cwc21. Este é um ponto de diferença fundamental entre a biogênese de snRNP U5 em *T. brucei* e em leveduras, já que este organismo não conserva a proteína Aar2, responsável por regular a entrada de U5-200K na partícula U5. As células de *T. brucei* que expressam um dos alelos do gene desta proteína na sua forma truncada apresentam um pequeno defeito de crescimento. Também se observa efeito negativo no processamento de pré-mRNA por SL *trans-splicing*.



### 3.1 Introdução

#### 3.1.1 U5-200K e a ativação do spliceossomo

O centro catalítico do spliceossomo é composto pelo di-snRNA U2/U6, que interage com os substratos de maneira a posicioná-los para as duas reações de transesterificação, e a estrutura ISL no snRNA U6, que se encontra adjacente e coordena dois íons  $Mg^{2+}$  (M1 e M2) que estabilizam os grupos abandonadores das reações de substituição nucleofílica do tipo SN2. Entretanto, nenhuma das duas estruturas está pré-formada no complexo A do spliceossomo, como descrito no capítulo 1 (seção 1.3).<sup>51, 54, 59, 94-95, 250</sup>

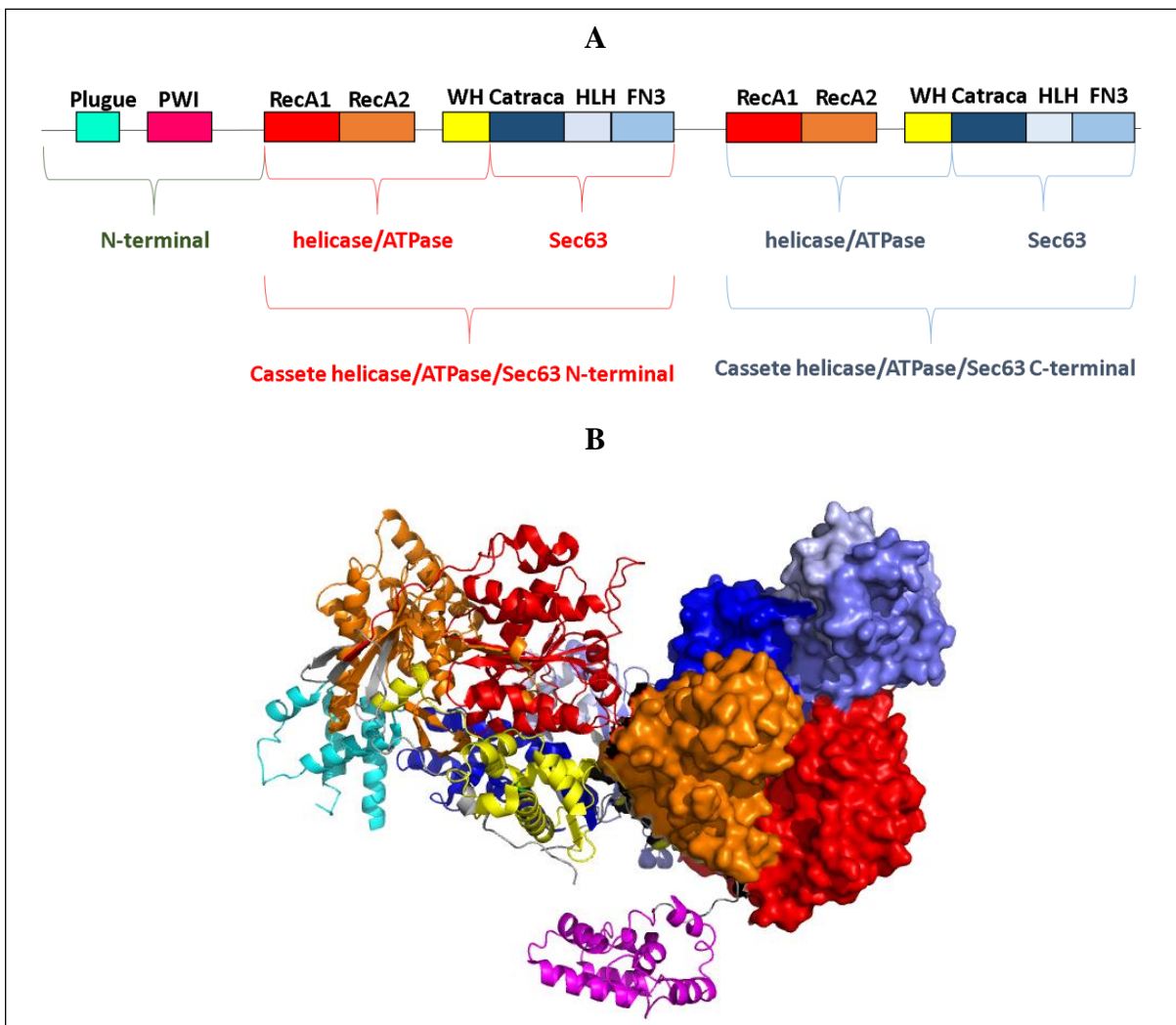
O snRNA U6 encontra-se, no complexo B, extensivamente pareado com o snRNA U4 nas regiões SL1 e SL2 dentro da tri-snRNP U4/U6.U5. Portanto, uma etapa fundamental na ativação do spliceossomo é o desapareamento do di-snRNA U4/U6, liberando snRNA U4 e permitindo o pareamento dos snRNAs U6 e U2 no complexo B\*. A enzima U5-200K (conhecida como Brr2 em levedura) tem função helicase e utiliza energia da hidrólise de ATP para induzir a dissociação do duplex de snRNAs U4/U6.<sup>57, 118, 251-252</sup>

#### 3.1.2 U5-200K – estrutura e função helicase

A proteína U5-200K é classificada no subgrupo de helicases similares a Ski2 da superfamília SF2, sendo também conhecida como uma RNA helicase que contém o motivo DExD/H.<sup>57,118</sup> As estruturas tridimensionais de U5-200K de *H.sapiens* e de Brr2 de *S. cerevisiae* são caracterizadas por conservar as seguintes regiões (de acordo com a nomenclatura descrita na literatura, Figura 3.1):

- a. Região N-terminal, que consiste de duas estruturas: um domínio conhecido como plugue, composto de dois grampos de cabelo  $\alpha$ , e o domínio PWI, conectado aos elementos vizinhos por meio de alças flexíveis. Ambos os domínios plugue e PWI estão separados por um gancho intercassete (IC) com conformação estendida e uma dobra de 90° no centro. Um segundo gancho, conhecido como gancho N-terminal (NC), separa o domínio PWI da região N-terminal do cassete helicase/ATPase 1.<sup>57, 118, 253-254</sup>

- b. Dois cassetes helicase/ATPase/Sec63 em *tandem* - cada cassete é composto por seis domínios: dois domínios RecA seguidos dos domínios WH (do inglês, *winged helix*), e os domínios catraca, HLH (do inglês, *helix-loop-helix*) e FN3 (do inglês, *fibronectin3-like*). Os três primeiros domínios são referidos como cassete helicase/ATPase e os três últimos são referidos como domínio Sec63. Os domínios WH, catraca e HLH são similares aos conservados na DNA helicase Hel308.<sup>255</sup> Além disso, os domínios RecA1, RecA2, WH e catraca estão arranados em forma circular formando, respectivamente, a base e o topo de um túnel para acomodar uma fita de ácido nucleico. Um grampo de cabelo  $\beta$  que se estende de RecA1 até o domínio catraca tem função de separar duas fitas de ácido nucleico. O domínio HLH, por sua vez, está posicionado entre catraca e FN3.<sup>57, 118</sup>



**Figura 3.1– Arquitetura de domínios de U5-200K.** A- Representação esquemática de domínios conservados em U5-200K. B- Representação da estrutura tridimensional de U5-200K de *H. sapiens* (código PDB 4KIT) mostrando as regiões N-terminal e os cassetes helicase/ATPase/Sec63 N- e C-terminais, utilizando o mesmo código de cores de A-. As regiões N-terminal e cassete N-terminal são representados como *cartoon* e a região cassete C-terminal pe representada como superfície. Fonte: Adaptada de SANTOS et al<sup>57</sup>; NGUYEN et al<sup>118</sup>.



As regiões a. e b. (Figura 3.1) são essenciais para *splicing* eficiente e viabilidade celular, sendo que a região N-terminal é necessária para associação estável de U5-200K à tri-snRNP U4/U6.U5. Santos et al<sup>57</sup> mostraram, recentemente, que o cassete helicase/ATPase C-terminal de Brr2 apresenta atividade de ligação a ATP, porém não é capaz de hidrolisar o nucleotídeo, já que não conserva todos os resíduos de aminoácidos necessários para tal função. Entretanto, a atividade ATPase no cassete N-terminal é aumentada na presença do cassete C-terminal. Deste modo, apoiados por experimentos *in vitro* de interação com snRNAs e atividade ATPase/helicase de mutantes de Brr2, os autores propõem, para este segundo cassete, um papel de regulação da atividade do primeiro cassete helicase/ATPase.<sup>57, 118</sup>

As mutações em U5-200K de *H. sapiens* que estão relacionadas à retinite pigmentosa descrita no capítulo 1 (1.4.1.1) ocorrem em dois conjuntos de resíduos de aminoácidos, sendo um deles (R681C, P681H, V683L e Y689C) localizado entre os dois domínios RecA do cassete N-terminal, e o outro (S1087L e R1090L) situado na região de catraca, também no domínio Sec63 N-terminal. Tais mutações estão relacionadas à diminuição na atividade catalítica de U5-200K e, conseqüentemente, do spliceossomo, como causa da doença.<sup>57, 118</sup>

A proteína U5-200K é descrita como membro da snRNP U5 juntamente com U5-220K (hPrp8), U5-116K (hSnu114), U5-102K (hPrp6), U5-100K (hPrp28), U5-52K (hLin1), U5-40K e U5-15K (hDib1) em *H. sapiens* e se mantém nesta partícula quando ela se associa à di-snRNP U4/U6 para formação de tri-snRNP U4/U6.U5 e dos complexos do spliceossomo, como descrito no capítulo 1 (seção 1.3.3). As estruturas de criomicroscopia eletrônica de tri-snRNPs U4/U6.U5 de levedura<sup>94</sup> e *H. sapiens*<sup>250</sup> mostram a proteína U5-200K em posições completamente diferentes. A proteína Brr2 de *S. cerevisiae* encontra-se ligada ao domínio Jab1 de Prp8 (U5-220K), que interage com o domínio EM desta mesma proteína, e ao núcleo de proteínas Sm de snRNP U4 em uma região localizada a cerca de 20 nm de distância da posição em que U5-200K é encontrada na tri-snRNP U4/U6.U5 de *H. sapiens*. Nesta última estrutura, o cassete C-terminal (b.) de U5-200K interage com Snu114 e Sad1, enquanto sua região N-terminal (a.) interage diretamente com o domínio RNase H de U5-220K e com hPrp6.<sup>57, 94, 118,</sup>

250

O di-sRNA U4/U6 tem as extremidades 3' livres como fita simples e ambas poderiam ser substrato para a helicase. Entretanto, tais extremidades estão oclusas em interação com outro snRNA formando uma estrutura de grampo de cabelo e/ou interagindo com proteínas Sm. O mecanismo de reconhecimento da extremidade correta por meio da proteína U5-200K ainda não é conhecido. A estrutura da proteína Brr2 de *S. cerevisiae* presente na tri-snRNP U4/U6.U5

resolvida por criomicroscopia eletrônica<sup>94</sup> mostra que a proteína está ligada à extremidade de fita simples do snRNA U4. Nesta estrutura, a separação entre os domínios RecA do cassete N-terminal é mais alargada e o separador na forma de grampo de cabelo  $\beta$  encontra-se próximo à região de fita dupla entre os snRNAs U4 e U6.<sup>94</sup>

Antes da ativação do spliceossomo, U5-200K já encontra seu substrato na tri-snRNP U4/U6.U5. A proteína Prp8 (U5-220K) compete pela ligação ao snRNA U4 por meio do seu domínio RNase-H, enquanto a extensão C-terminal do seu domínio Jab1 ocupa a entrada do túnel formado pelos domínios RecA1-A2 e WH na proteína U5-200K. Além disso, a região N-terminal desta proteína pode apresentar atividade de autoinibição, competindo pela ligação ao snRNA no domínio plugue e restringindo a flexibilidade conformacional da região helicase devido à ligação do gancho intercassete e do domínio PWI. A proteína Sad1, necessária para a formação de tri-snRNP U4/U6.U5, deve estar relacionada à estabilização deste estado de autoinibição. Adicionalmente, a provável atividade GTPase da proteína U5-116K (hSnu114) tem sido relacionada à ativação de U5-200K, porém o mecanismo molecular ainda não está claro.<sup>57, 94, 118, 250</sup>

Outro ponto de regulação de U5-200K está na própria biogênese da partícula U5, já que o fator de *splicing* Aar2 a substitui no complexo pré-snRNP U5 ainda no citoplasma, como descrito no capítulo 1 (seção 1.3.2.4). O sinal para que Aar2 ceda seu lugar à proteína Brr2 na partícula U5 já no núcleo da célula é a fosforilação da primeira proteína.<sup>57, 118</sup>

### 3.1.3 U5-200K de *T. brucei*

Poucos dados estão disponíveis em relação à U5-200K de tripanossomatídeos (Tb927.5.2290). Dois trabalhos identificaram simultaneamente a proteína em *T. brucei* utilizando ferramentas de bioinformática.<sup>20, 60-61, 102</sup> Da Silva e colaboradores<sup>102</sup> confirmaram a localização de U5-200K de *T. brucei* no núcleo celular, onde o spliceossomo realiza sua função, e todos os snRNAs foram coimunoprecipitados, provavelmente por interação entre complexos U5-snRNP e outras snRNPs.<sup>20, 102</sup>

A arquitetura de domínios da proteína U5-200K de *T. brucei*, analisada em detalhes na seção 3.4.1, apresenta variação com relação às ortólogas de *H. sapiens* e *S. cerevisiae*. Além disso, a proteína Aar2 não foi identificada em *T. brucei* e o mecanismo de biogênese de snRNP U5 não é conhecido.<sup>20, 60-61, 102</sup>

## 3.2 Objetivos

Dado o acúmulo de informação a respeito da proteína U5-200K de *H. sapiens* e *S. cerevisiae* nos últimos dois anos, em detrimento da ortóloga de *T. brucei*, os objetivos específicos do trabalho descrito neste capítulo são:

- a. Analisar os parceiros de interação de U5-200K de *T. brucei* completa e de seu cassete helicase/ATPase N-terminal, em comparação com ortólogas de *H. sapiens* e *S. cerevisiae*.
- b. Determinar a localização celular dos produtos recombinantes.
- c. Analisar o efeito do truncamento da região C-terminal da proteína no mecanismo de *splicing*.

### 3.3 Materiais e métodos

#### 3.3.1 Análise da sequência de aminoácidos da proteína U5-200K de *T. brucei*

A análise de domínios conservados na sequência de aminoácidos da proteína U5-200K de *T. brucei* foi feita por meio do servidor *SMART*<sup>193</sup> e o alinhamento múltiplo com sequências homólogas de *H. sapiens* e *S. cerevisiae* foi gerado por meio do programa *Clustal Omega*<sup>194</sup>. Adicionalmente, a ferramenta *PROTPARAM* do servidor *Expasy*<sup>192</sup> foi usada para estimativa de propriedades físico-químicas. Por outro lado, a predição de estrutura secundária foi feita por meio do servidor *PSIPRED*<sup>195</sup> e a análise da estrutura terciária de U5-200K de *H. sapiens* e de Brr2 de *S. cerevisiae* foi feita no programa *PyMOL* (*Schrödinger, LLC*) e no servidor *PDBSum*<sup>229</sup>.

#### 3.3.2 Clonagem e expressão homóloga de construções de U5-200K em *T. brucei* para experimentos de purificação em *tandem*

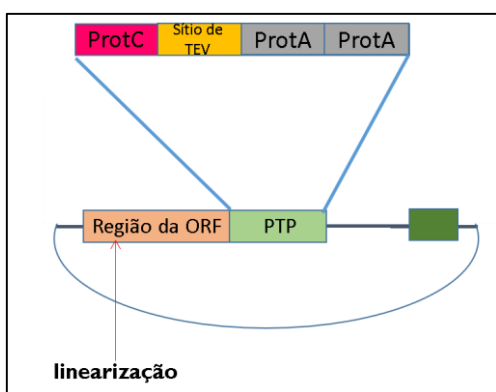
A técnica de purificação por afinidade em *tandem* (sigla do inglês, *tandem affinity purification* - TAP) é útil para purificação de complexos proteicos em condições nativas. Neste trabalho, uma extensão C-terminal contendo dois epítomos de proteína A de *Staphylococcus aureus* e um epítomo de proteína C da vitamina K humana, separados por uma região de reconhecimento da protease do vírus de tabaco (TEV), foi adicionada a diferentes construções da proteína U5-200K de *T. brucei* a partir do vetor de recombinação homóloga pC-PTP-NEO<sup>67</sup>.

##### 3.3.2.1 Amplificação das fases abertas de leitura e clonagem em vetor pC-PTP-NEO

Oligonucleotídeos foram planejados para a amplificação de regiões da fase aberta de leitura da U5-200K de *T. brucei* de modo que o produto recombinante de expressão homóloga fosse uma construção da U5-200K completa contendo a extensão PTP (ProtA–TEV–ProtC) na porção C-terminal e um produto recombinante correspondendo ao cassete helicase/ATPase N-terminal também contendo a extensão PTP na sua região C-terminal. A primeira construção corresponde ao intervalo da 5920<sup>a</sup> à 9714<sup>a</sup> base nitrogenada (795 pb) e os oligonucleotídeos desenhados são 5' GGGCCCCGTG GCGTGAGAGCGAAATTTTGG 3' (*sense*) e 5' CGGCCGTCCCGC TGCAAAGGTTACGTC 3' (*antisense*), enquanto que, para a segunda construção, que compreende a região de 3150<sup>a</sup> à 3849<sup>a</sup> base nitrogenada (700 pb), os oligonucleotídeos são 5' GGGCCCAGTACAGTCACCAGCGCGTC 3' (*sense*) e 5' CGGCCCG

TATTGGTGCAACCAGGTCGTG 3' (*antisense*). Estão destacados os sítios de restrição para as endonucleases *ApaI* e *EagI*. A amplificação destas regiões foi feita a partir do DNA genômico de *T. brucei* (capítulo 2, seção 2.3.2) por meio de reação em cadeia da DNA polimerase *Phusion* (*Thermoscientific*), de acordo com as especificações do fabricante, e verificada em gel de agarose 1% corado com 0,5  $\mu\text{g mL}^{-1}$  de *SyberSafe* em tampão TAE 1X. Os produtos foram ligados, separadamente, em vetor de clonagem pJET (*Thermoscientific*) de acordo com as condições descritas pelo fabricante e confirmados por sequenciamento no laboratório de biofísica do Instituto de Física de São Carlos – Universidade de São Paulo.

A digestão com endonucleases de restrição tanto destas construções em vetor de clonagem (*ApaI* e *EagI* - *BioLabs*) como do vetor pC-PTP-NEO<sup>67</sup> (*ApaI* e *NotI* - *BioLabs*) seguiu as recomendações do fabricante das enzimas e os produtos de digestão foram purificados de gel de agarose 1% utilizando-se o *kit Wizard® SV Gel Clean-Up System* (*PROMEGA*), seguido de reação de ligação com T4 DNA ligase (*Thermoscientific*), de acordo com as condições do fabricante. O volume total da mistura de ligação foi utilizado na transformação de células competentes *E. coli* DH5 $\alpha$  como descrito no (capítulo 2, seção 2.3.2). Os plasmídeos recombinantes foram extraídos utilizando-se o *kit GeneJET™ Plasmid Miniprep* (*Fermentas Life Sciences*).



**Figura 3.2 – Representação esquemática do vetor pC-PTP-NEO.** São mostradas as regiões de ligação da ORF de interesse (rosa), a extensão PTP (verde claro) e o gene de resistência a gentamicina (verde escuro).  
Fonte: Elaborada pelo autor a partir de PALFI et al<sup>67</sup>.

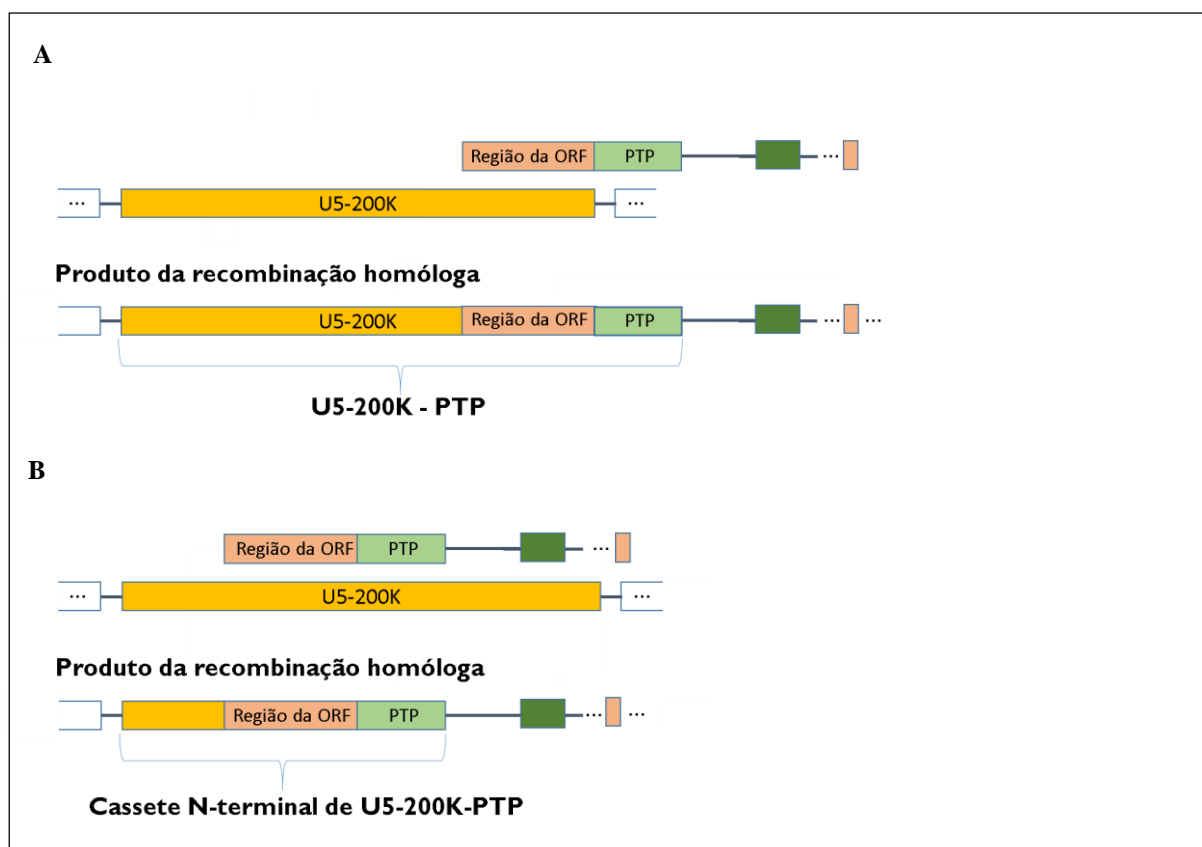
### 3.3.2.2 Transfecção, seleção e verificação de expressão dos produtos recombinantes

Para transfecção em células de *T. brucei* 427 na forma procíclica, o plasmídeo correspondente à construção recombinante U5-200K-PTP completa foi linearizado com a endonuclease de restrição *XcmI* (*Biolabs*), enquanto o plasmídeo correspondente à construção cassete N-terminal de U5-200K-PTP foi linearizado com a enzima *BSpEI* (*Biolabs*), de acordo

com as recomendações do fabricante. O produto linearizado foi purificado de gel de agarose 1% como descrito anteriormente.

Um total de 10 µg de plasmídeo de cada construção foi utilizado para transfecção em células *T. brucei* 427 na forma prócíclica em fase log de crescimento ( $5 \times 10^6$  parasitas mL<sup>-1</sup>), de acordo com o protocolo descrito por Palfi et al<sup>67</sup>. Brevemente, um total de  $7 \times 10^7$  células foi centrifugado a 2.500xg por cinco minutos a 4°C, lavado e ressuspensionado em 1,5 mL de meio de eletroporação (três partes de *cytomix* para uma parte de solução de sacarose tamponada – *cytomix* = 120 mM KCl, 0,15 mM CaCl<sub>2</sub>, 10 mM K<sub>2</sub>HPO<sub>4</sub>, 25 mM Hepes, 2 mM EDTA e 5 mM MgCl<sub>2</sub> em pH 7,6 e sacarose tamponada = 277 mM sacarose, 16 mM KPO<sub>4</sub>, 1 mM MgCl<sub>2</sub> em pH 7,4). O total de DNA plasmidial linearizado (10 µg) foi adicionado a 450 µL de células ressuspensionadas em meio de eletroporação para cada construção de U5-200K. O resultado foi submetido à eletroporação por dois pulsos sucessivos (1.500 V e 25 µF) separados por um intervalo de 10 segundos no eletroporador GeneRuler II (*BioRad*). O produto foi transferido para 10 mL de meio de cultura SDM79<sup>257</sup> (suplementado com 10% (v/v) soro fetal bovino, 5 mg L<sup>-1</sup> hemina e 100 mg L<sup>-1</sup> de mistura penicilina/estreptomicina) contendo 20 µg mL<sup>-1</sup> de gentamicina. Um volume de 250 µL de produto foi pipetado em placa de 96 poços para a seleção de parasitas recombinantes, os quais foram transferidos para 10 mL de meio SDM-79<sup>257</sup> suplementado contendo, nesta etapa, 40 µg mL<sup>-1</sup> de geneticina (G418).

Após transfecção, ocorre a recombinação homóloga, como apresentado esquematicamente na Figura 3.3. O resultado esperado é de que um dos alelos do gene que codifica para U5-200K no genoma de *T. brucei* receba a extensão PTP na sua extremidade C-terminal, de modo que resulte em dois clones, sendo que um deles expressa homologamente a proteína U5-200K-PTP completa e outro a proteína que consiste da região do cassete N-terminal da mesma proteína ligado à extensão PTP em sua extremidade C-terminal.



**Figura 3.3 – Representação esquemática dos produtos de recombinação homóloga.** A- U5-200K-PTP completa. B- Região N-terminal + cassete helicase/ATPase/Sec63 N-terminal de U5-200K-PTP (incluindo região N-terminal).  
Fonte: Elaborada pelo autor.

O conteúdo de DNA genômico das populações clonais de *T. brucei* foi extraído pelo mesmo método empregado na seção 2.3.2 para avaliação da inserção correta dos fragmentos lineares das construções em pC-PTP-NEO. Para tal, foram utilizados dois oligonucleotídeos iniciadores específicos, sendo que um deles parecia com a sequência da proteína C (protC-TAP3' - (5'-TCGCTAGCAGTAGTTGGAATATCATATTTCCCA TCAATAAGACGAGG-3') e o outro com uma região de U5-200K fora da sequência contida no fragmento linearizado (conhecido como iniciador 5'-out - 5' GCTGTCGAACCTCAACTGCTG 3' para a construção completa de U5-200K-PTP, e 5' TCCTTCGTGCTCTTTATGAG 3' para a construção contendo apenas o cassete N-terminal).

A confirmação da expressão homóloga de U5-200K-PTP e do seu cassete N-terminal-PTP recombinantes foi feita por meio de *Western Blotting*. O sedimento correspondente à centrifugação de 1 mL de cultura de *T. brucei* na fase logarítmica de crescimento (5 minutos a 2.500xg/temperatura ambiente) foi ressuscitado em 40 µL de tampão de amostra (5X) e fervido por cinco minutos, sendo que apenas 10 µL do produto foi aplicado em SDS-PAGE

10% (180V/1 h). As bandas de proteína foram transferidas para membrana de nitrocelulose (100 V/1h – *Mini Trans-Blot® Cell - Biorad*), em tampão de transferência gelado (25 mM Tris, 192 mM glicina, 20% metanol gelado e água milli-Q para completar 1 L de tampão). Em seguida, a membrana foi bloqueada com 5% (m/v) leite dissolvido em PBST 1x (PBS 1x, 0,1% Tween20) sob agitação por uma hora à temperatura ambiente. Foi adicionado anticorpo primário anti-proteína A (*Sigma*) em diluição 1:4000 em PBST 1x/leite 2% por duas horas sob agitação à temperatura ambiente, seguido de três lavagens com o mesmo tampão. Finalmente, a membrana foi incubada com anticorpo secundário anti-IgG de coelho (*KPL*) em diluição 1:5000 por uma hora sob agitação à temperatura ambiente. Após mais três lavagens com mesmo tampão, seguiu-se com revelação em solução de luminol (6 mL de solução contendo 0,01M Tris 1M pH 8,5, 90 mM de ácido p-cumático e 250 mM de luminol adicionados de 20 µL de H<sub>2</sub>O<sub>2</sub>) e observação no equipamento C-Digit (*Li-COR*).

### 3.3.2.3 Curvas de crescimento

O crescimento das culturas de células recombinantes a 26°C foi acompanhado em comparação à cultura de *T. brucei* cepa 427 selvagem na forma procíclica. A contagem de parasitas foi realizada periodicamente em câmara de Neubauer<sup>258</sup> e a análise estatística comparativa do crescimento das culturas foi feita por meio do teste T de Student<sup>259</sup>. Neste teste paramétrico, comparam-se as médias dos números de células contadas para cada condição diferente e em cada ponto da curva de crescimento, dados os desvios padrões de cada medida, com base na hipótese de que as médias são diferentes para cada condição. A hipótese nula é a condição de igualdade entre as médias. São considerados significativos aqueles resultados com 95% ( $p < 0,05$ ) de probabilidade de refutarem a hipótese nula.

### 3.3.3 Purificação dos produtos recombinantes e análise dos parceiros de interação por espectrometria de massas

Um total de cinco litros de cultura de *T. brucei* recombinante, cultivados em meio de SDM79<sup>257</sup> suplementado até atingir a sua fase logarítmica de crescimento ( $\sim 10^6$  células), foi centrifugado por 10 minutos a 3.300xg a 4°C. O sobrenadante foi descartado e o sedimento foi lavado por duas vezes em 40 mL de tampão conhecido como *trypsin wash* gelado (seção 2.3.2 do capítulo 2). Em seguida, o sedimento foi lavado novamente com 30 mL de uma solução também conhecida como tampão de transcrição (150 mM sacarose, 20 mM L-glutamato de potássio, 20

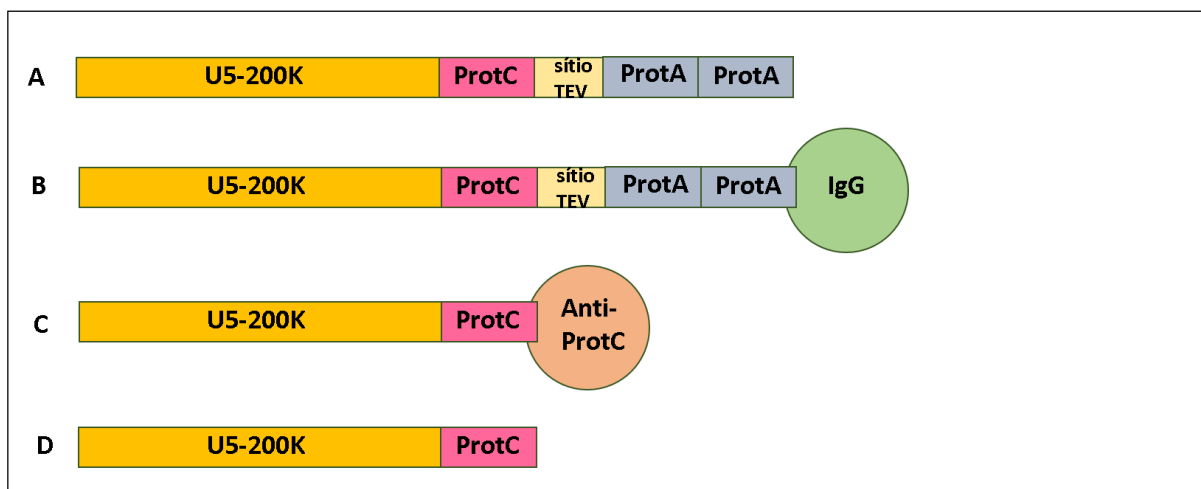


mM Hepes/KOH pH 7,7e 3 mM MgCl<sub>2</sub>) e ressuspendido na proporção 1 volume de sedimento : 1 volume de tampão de transcrição gelado. O resultado foi incubado por 30 minutos no gelo. Após esta etapa, um comprimido de coquetel de inibidores de proteases (*mini complete*<sup>TM</sup>, livre de EDTA, *Roche Diagnosis Ltda*) e 1:1000 (v:v) de DTT (1M) foram adicionados de acordo com o volume de células ressuspendidas. As células foram lisadas mecanicamente em prensa francesa (*Thermoscientific*) em seis ciclos de 2.000 psi (um minuto cada). O lisado celular (conhecido como Y4) correspondente a cada construção recombinante foi aliquoteado em frações de 1 mL e congelado em nitrogênio líquido e mantido a -80°C por até seis meses.

Para purificação, 10 amostras correspondentes a cada construção de U5-200K foram descongeladas e cada uma foi misturada com 100 uL de tampão de extração gelado (20 mM Tris/HCl pH 7,7, 150 mM KCl, 3 mM MgCl<sub>2</sub>, 0,5 mM DTT, 1% Tween20). O produto foi incubado no gelo por 20 minutos, sendo invertido por 3 vezes a cada cinco minutos. Os tubos foram então centrifugados a 21.000xg por 15 minutos a 4°C e o sobrenadante foi cuidadosamente transferido para novo tubo previamente gelado, onde foi adicionado meio comprimido de coquetel de inibidores de proteases (*mini complete*<sup>TM</sup>, livre de EDTA, *Roche Diagnosis Ltda*). O produto foi aplicado em coluna 260 uL de resina *IgG sepharose Fast Flow (GE)* previamente equilibrada com 50 mL de tampão PA-150 (20 mM Tris/HCl pH 7,7, 150 mM KCl, 3 mM MgCl<sub>2</sub>, 0,5 mM DTT, 0,1% Tween20) e mantido sob rotação por duas horas em câmara fria. A resina foi lavada com mais 15 mL de tampão PA-150. Em seguida, foram adicionados 2 mL de tampão de proteólise (20 mM Tris/HCl pH 7,7, 150 mM KCl, 3 mM MgCl<sub>2</sub>, 1 mM DTT, 0,1% Tween20) e 300 U de de TEV protease (*Invitrogen*), mantidos sob rotação por 16 horas a 4°C.

O eluato obtido após proteólise com TEV para cada construção de U5-200K-PTP foi coletado. Foram adicionados 4 mL de tampão conhecido como PC-150 (150 mM KCl, 20 mM Tris/HCl pH 7,7, 3 mM MgCl<sub>2</sub>, 0,5 mM DTT, 0,1% Tween20 e 1 mM CaCl<sub>2</sub>) e coletados. Foram então adicionados 7,5 uL de CaCl<sub>2</sub> (1 M) e 500 µL de solução de coquetel de inibidores de proteases (meio comprimido em tampão PA-150) e o conteúdo foi adicionado à 260 uL de resina contendo anticorpos monoclonais anti-proteína C imobilizados (*Roche*), previamente equilibrado com 50 mL de tampão PC-150. O produto foi mantido sob rotação por duas horas em câmara fria. Em seguida, a fração não ligada correspondente a cada purificação foi coletada e a resina foi lavada com mais 30 mL de tampão PC-150, mantendo sob rotação por 10 minutos a cada 10 mL de lavagem. Finalmente, foram adicionados 600 uL de tampão de eluição (5 mM Tris/HCl, 10 mM EGTA, 5 mM EDTA, 0,01 mg mL<sup>-1</sup> leupeptina) e o produto foi deixado sob rotação por mais 5 minutos à temperatura ambiente. Após a coleta do eluato correspondente, a

etapa de eluição foi repetida por mais duas vezes. A Figura 3.4 apresenta um resumo das etapas de purificação em *tandem* utilizadas.



**Figura 3.4 – Representação esquemática das etapas de purificação em *tandem*.** A- Produto de expressão recombinante. B- Ligação à resina de IgG sepharose (GE). B- Ligação à resina com anticorpos anti-proteína C (Roche) imobilizados após digestão com protease TEV. D- Eluição com tampão contendo EGTA/EDTA.

Fonte: Elaborada pelo autor.

As amostras coletadas em cada etapa foram submetidas a SDS-PAGE 10% corado com Sypro®Ruby (Thermoscientific). No caso do eluato da resina contendo anticorpo anti-proteína C, adicionou-se 15  $\mu$ L de resina *StrataClean* (Agilent), que foi mantida por 20 minutos sob rotação em contato com as amostras de interesse, e centrifugada por cinco minutos a 3.000xg. O sedimento foi ressuspendido em 10  $\mu$ L de tampão de amostra (5X), fervido por cinco minutos e todo o conteúdo foi aplicado no mesmo gel. A visualização do resultado foi feita em TYPHOON TRIO (GE). Em seguida, as bandas foram recortadas do gel, congeladas a  $-20^{\circ}\text{C}$  e enviadas para análise por espectrometria de massas no laboratório da Dra. Júlia Chagas Pinheiro da Cunha no Centro de Toxinas, Resposta-Imune e Sinalização Celular do Instituto Butantan, em São Paulo.

As bandas foram descoradas por três horas à temperatura ambiente com solução de metanol 50% e ácido acético 5% em água milli-Q, para se remover o SDS. Em seguida, os fragmentos de gel foram desidratados com 100% acetonitrila por cinco minutos, a qual foi evaporada à vácuo. O resultado foi reduzido pela adição de 10 mM DTT, incubando-se por 30 minutos à temperatura ambiente, e as amostras foram alquiladas pela adição de 30  $\mu$ L de solução iodoacetamida (50 mM de iodoacetamida em solução de 100 mM bicarbonato de amônio) por 30 minutos à temperatura ambiente. Finalmente, os fragmentos foram lavados com 100 mM bicarbonato de amônio por 10 minutos e desidratados novamente com 100%

acetonitrila e reidratados com 100 mM bicarbonato de amônio. Este procedimento foi repetido por mais uma vez e os fragmentos desidratados resultantes foram proteolisados com 50 ng  $\mu\text{L}^{-1}$  tripsina (*Sigma*) em solução de 50 mM bicarbonato de amônio durante 16 horas a 4°C.

Às amostras proteolisadas, adicionou-se solução 5% de ácido fórmico de modo a se recobrir todo o gel, e incubou-se por 10 minutos à temperatura ambiente. Após coletar o sobrenadante, foi adicionada mais solução 5% ácido fórmico em 50% acetonitrila, incubando-se novamente à temperatura ambiente por 10 minutos. As amostras resultantes foram evaporadas à vácuo e ressuspendidas em 0,5% ácido fórmico e dessalinizadas em resinas do tipo *zip-tip* (*Millipore*).

O experimento de espectrometria de massas das amostras resultantes foi realizado em LTQ-Orbitrap Velos (*ThermoScientific*), a 1,9 kV e 200 °C, em modo de aquisição dependente dos dados, com cobertura de 200 a 2000 m/z e resolução de 3.000 Da. Os dados foram analisados nos programas *Andromeda Search engine*<sup>260</sup> e as sequências de proteínas foram identificadas a partir de busca em banco de dados de sequências de *T. brucei* (*TriTrypDB*<sup>53</sup>).

### 3.3.4 Imunolocalização dos produtos recombinantes

Um total de  $4 \times 10^6$  parasitas de culturas correspondentes a cada construção de U5-200K de *T. brucei* foram centrifugados 600xg por 10 minutos à temperatura ambiente, o sedimento foi lavado por duas vezes em 1 mL de PBS 1X (pH 7,4, *Sigma*) gelado, e finalmente ressuspendido em 100  $\mu\text{L}$  do mesmo tampão. Gotas de 40  $\mu\text{L}$  foram adicionadas a poços de uma lâmina previamente tratada com 0,1% poli-L-lisina e mantido em repouso à temperatura ambiente por 5 minutos para adesão. Os parasitas foram fixados com solução de paraformaldeído 2% em PBS 1X gelado, incubando-se por 20 minutos à temperatura ambiente e lavados por três vezes com PBS 1X gelado.

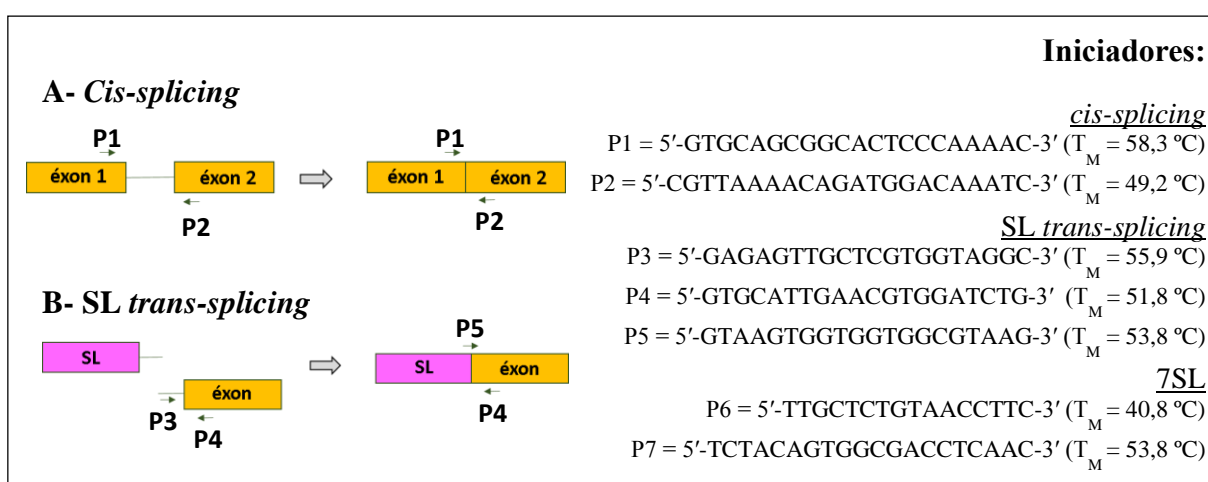
As células foram permeabilizadas com solução Triton X-100 0,3% em PBS 1X gelado, incubando-se por três minutos à temperatura ambiente, lavadas por três vezes em PBS 1X e bloqueadas com solução 3% BSA em PBS 1X por uma hora à temperatura ambiente. Em seguida, as lâminas foram incubadas por uma hora à temperatura ambiente com anti-proteína A de coelho (*Sigma*) em diluição 1:40.000 diluído em solução de bloqueio. Após três lavagens com PBS 1X gelado, as células foram então incubadas por uma hora à temperatura ambiente com solução de bloqueio contendo o anticorpo secundário Alexa 594 (*Invitrogen*) em diluição 1:400 e DAPI 10  $\mu\text{g ml}^{-1}$ . Após três lavagens com PBS 1X gelado, adicionou-se *vectashield*

*mounting media* (VectorLabs) para se evitar fotodegradação dos fluoróforos e as lâminas foram vedadas com esmalte. Finalmente, as lâminas foram visualizadas em microscópio de fluorescência *ECLIPSE E600* (Nikon) com contraste de fase e com filtro para comprimentos de onda na região de ultravioleta e as imagens foram analisadas no programa *NIS-Elements BR3.0* (Nikon).

### 3.3.5 Avaliação de produtos de *cis*- e *SL trans-splicing* por meio de RT-PCR

O total de RNA de culturas de *T. brucei* 427 na forma procíclica recombinantes (10 mL) foi extraído por meio do *kit NucleoSpin RNA purification* (Macherey-Nagel), que inclui tratamento com DNase I para se evitar contaminação dos preparos com DNA endógeno. Em seguida, moléculas de cDNA foram produzidas usando *SuperScript III first-strand cDNA synthesis* (ThermoFisher), de acordo com as especificações do fabricante, a partir de oligonucleotídeos randômicos e (dT)<sub>n</sub>. Os oligonucleotídeos randômicos são sequências de seis nucleotídeos que possuem todas as combinações de bases nitrogenadas possíveis, capaz de se alinhar à maioria das moléculas de cDNA. Oligonucleotídeos (dT)<sub>n</sub>, por sua vez, reconhecem a sequência de (A)<sub>n</sub> na extremidade 3' de transcritos maduros.

Finalmente, oligonucleotídeos iniciadores específicos foram desenhados para amplificação de produtos de *splicing* dos genes PAP (poli-A polimerase) e  $\alpha$ -tubulina, como representado na Figura 3.5.



**Figura 3.5 – Representação esquemática das regiões de amplificação dos produtos de *cis* e *SL trans-splicing*.**

São mostrados os pares de iniciadores (setas) para amplificação de região de pré-mRNA não processado (esquerda) e processado (direita), no caso de A- *cis*- e B- *SL trans-splicing*, respectivamente. As sequências de DNA e as respectivas T<sub>M</sub> são apresentados.

Fonte: Elaborada pelo autor.

### 3.4 Resultados

Exceto pela análise da sequência de aminoácidos da proteína U5-200K de *T. brucei*, os resultados aqui apresentados foram obtidos com a ajuda de Camila Maria Boralli, durante seu treinamento como aluna de iniciação científica, e do pós-doutorando Dr. Marco Túlio Alves da Silva no Laboratório de Biologia Estrutural do grupo de Cristalografia do Instituto de Física de São Carlos da Universidade de São Paulo.

#### 3.4.1 Análise da estrutura primária de U5-200K

De acordo com o programa *ProtParam* do servidor *Expasy*<sup>192</sup>, a proteína U5-200K de *T. brucei* possui 249,3 kDa e pI teórico igual a 5,85. O alinhamento da sequência de aminoácidos de U5-200K de *T. brucei* com as ortólogas de humano e de levedura resulta em 33% e 31% de identidade, respectivamente. A análise do alinhamento múltiplo de sequências, predição de domínios e de predição de estrutura secundária indica que a sequência de aminoácidos de U5-200K de *T. brucei* também pode ser dividida nas regiões N-terminal (contendo os domínios plugue e PWI) e nos cassetes helicase/ATPase/Sec63 N- e C-terminal, que incluem, cada um, dois domínios RecA (compostos pela alça Q e pelos motivos I, Ia, Ib, II e III, IV, IVa, IVb, V e VI) e um domínio WH além de dois domínios Sec63, cada um composto pelos domínios catraca, HLH e domínio similar a FN3 (Figura 3.6).

Na região N-terminal de U5-200K de *T. brucei*, o domínio plugue conserva um padrão de resíduos hidrofóbicos em regiões das duas primeiras hélices- $\alpha$  e resíduos carregados na terceira. O gancho IC, por sua vez, não é conservado e nota-se que, em humano, essa região da proteína também não tem sequência conservada. O domínio PWI, por sua vez, não é canônico e não conserva a sequência característica de resíduos de fenilalanina, triptofano e isoleucina na extremidade N-terminal. Na proteína U5-200K de *T. brucei*, ainda, a última estrutura na forma de hélice- $\alpha$  não é conservada. Em outras proteínas, este domínio serve como sítio de interação não-específica com o RNA. Finalmente, o gancho NC se mostra bem conservado para a proteína em análise (Figura 3.6).

Na superfamília SF1 de helicases, os motivos do tipo Walker (I e II) na região de RecA1 são responsáveis pela interação com  $Mg^{2+}$ -ATP/ADP. No motivo I, o grupamento  $\alpha$ -amino da lisina conservada no cassete helicase/ATPase N-terminal interage com os grupos fosfatos  $Mg^{2+}$ -ATP/ADP e a hidroxila da treonina interage com  $Mg^{2+}$ , enquanto que a carboxila do aspartato adjacente no motivo II coordena o íon de  $Mg^{2+}$  e o glutamato atua como base catalítica para a

hidrólise de ATP.<sup>261-262</sup> Note que o motivo I presente em RecA1 dos cassetes N- e C- terminais de todas as proteínas analisadas conservam ambos os resíduos de lisina e treonina característicos. Já no motivo II, os resíduos de asparagina e histidina característicos são conservados nos dois cassetes, enquanto o resíduo de glutamato adjacente ao primeiro aspartato é mutado para asparagina ou histidina (Figura 3.6).

Por outro lado a ligação da região 3' do snRNA deve ocorrer nos motivos Ia e Ib (RecA1), e IV e IVa (RecA2), sendo estabilizada na proximidade da região básica do domínio catraca de Sec63, o qual também provê o direcionamento da fita de RNA<sup>261, 262</sup>. Os resíduos característicos de todos estes domínios são conservados apenas no cassete N-terminal de cada proteína, sendo que os motivos IVa e V não apresentam resíduos conservados no cassete C-terminal de U5-200K de *T. brucei* (Figura 3.6).

O mesmo ocorre para os domínios Sec63. Nas três sequências de proteína comparadas, o domínio Sec63 do cassete C-terminal não conserva todos os resíduos característicos de cada motivo, sendo que para a região correspondente ao domínio Sec63 C-terminal de *T. brucei*, poucos são os resíduos conservados, o que dificulta a análise de domínios automatizada no servidor *SMART*<sup>193</sup> e este acaba não o encontrando. Algumas helicases de ácidos nucleicos também conservam um segundo cassete helicase/ATPase originado de duplicação gênica. Entretanto, como na U5-200K de *T. brucei*, observa-se acúmulo de mutações inativadoras (Figura 3.6).

Dentre os dois grupos de resíduos cujas mutações estão relacionadas à doença retinite pigmentosa em humanos (seção 3.1.2), apenas o resíduo equivalente a Y689 na U5-200K de *H. sapiens* está mutado para fenilalanina na ortóloga de *T. brucei*, o que provavelmente não altera as possíveis interações deste resíduo e não deve implicar em perda de função da proteína (Figura 3.6). Na figura 3.6, a seguir, encontra-se o alinhamento dos cassetes C- e N-terminal de U5-200K de *T. brucei* juntamente com Brr2 de *S. cerevisiae*.

```

CT_U5-200K_Tbrucei      -----
CT_Brr2_Scerevisiae    -----
CT_U5-200K_Hsapiens    -----
NT_U5-200K_Tbrucei      -----
NT_Brr2_Scerevisiae    MTEHETKDKAKKIREIYRYDEMSENKVLKVDKRFMNTSQNPQRDAEISQPKSMSGRI SAKD
U5-200K_Hsapiens       -----MADV TARS LQY EYKANSNLV LQADRSLIDRTRR---DEPTGEVLSLVGKLEGR

```

```

CT_U5-200K_Tbrucei      -----
CT_Brr2_Scerevisiae    -----
CT_U5-200K_Hsapiens    -----
NT_U5-200K_Tbrucei      -----MMAAITRYSPHSERT
NT_Brr2_Scerevisiae    MGQGLCNNINKLKENVAVE-K-TGKSASLKKIQQHNTILN--SSSDPRLHYYPKDP SN
U5-200K_Hsapiens       MGDKAQRTKPKQ-MQEERRAKRRRDEDRHDINKMKG-YTLLSEGIDEMVGIIYKPKET

```

Domínio plugue

continua

continuação

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 RVHYERFLRDIRRWLPPGLDDSIGEVAEETLSIVCRAGSGGFS----GSGAATQIQNV  
 VETYEQILQ---WVTEVLGNDIP-----HDLIIGTADILIRQLKENEENEDGNIERK  
 RETYEVLLS---FIQAALGD-QP-----RDILCGAADEVLAVLKNEKLRD---KERR

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 KQLETLFDSPLSAETLNEVLQYGRLLIDDFVSTEDPEGAHTNNVVAEDGLSDLLLMOEGKG  
 EKIQHHELGINIDSLKFNELVKLMKNTDYETHPDNSNKQA-----VAILADDEKS  
 KEIDLLLQ-QTDDTRYHVLVNLGKKITDYGGDKIIONMDD-NIDETYGVNVQFESDEEEG

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 QRKPSDMSSSDSDGDTQKADYGKGRDLMKFAADVDSINDGFDDQNDDDANCTNAIRVTF  
 DEE-----EVTEMSNSANVLGGEINDNEDDDEBYDYNDEVVNS  
 DE-----DVYGEVREEASDDDMEGDEAVVR---



CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 E---EVACN-----PNYIRDSLRRLFPQTIEECDLQAE  
 KKKKRALPNIENDIIRLSDSKTSNIESVPIYSIDFFMQRKLRLSELGYKDTSVIQDLSE  
 -----CTLSANLVASGFLMSKKKDLHPRDIDAFWLQRQLSRF--YDDAIVSQKKAD

Gancho IC

Dominio PWI

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 RVLQYCGKRSVDQLTLETQLTAFLLGGYDDEAVTDMWIATVAASRWLDIVYGMSPASKQTQKE  
 KILNDIETLEHNPVALEQRLVDLLK-FE---NISLAEFILKNRSTIFWGIKRLAKST-ENE  
 EVLEILKT-ASDDRECENQLVLLG-FN---TFDFIKVLRQHRMMILYCTLLASAQSEAE

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 KSLVMDAVKEHARTDRVVERL-----YQS---ITGKELDN---KPH-SGERGN  
 IPNLIKRMVAK-----GLNDLVEQYKFRETHSKRELDGDDQPQSSEAKRT  
 KERIMGKMEADPELSKFLYQLHETEKEDLIREERSRREKVRQSRMDTDLE----TMDLDQ



CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 U5-200K\_Hsapiens

-----  
 -----  
 -----  
 DADGSKPLRRVDLQAYAFKDERTPHQYTHAVVPQGTKRAVFEHDEVLPPPTASAT--EN  
 KFSNBAIPPVIDLEKIKFDESSKLMVTVKVSLPEGSFKRVKPYDEIHIPAPSKPVID--  
 GGEALAPRQVLDLEDLVFTQGSHEFMANKRCQLPDGSRFRQRKGYEEVHVPAKPKPFGESE

Gancho NC

CT\_U5-200K\_Tbrucei  
 CT\_Brr2\_Scerevisiae  
 CT\_U5-200K\_Hsapiens  
 NT\_U5-200K\_Tbrucei  
 NT\_Brr2\_Scerevisiae  
 NT\_U5-200K\_Hsapiens

-----  
 -----  
 -----  
 LPCTPLAVFPEWARPAFPGV-TQ  
 YELKEITSLPDMCQEAFFPSSETT  
 EQLLPVEKLPKYAQAGFEGF-KT

continua

continuação

CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> FFQIQHDLVAPIMENRGESFFVAFPPCSGKTAVAEI FNKIQSQVFESE-LYNSNDSVVFVGSGRCTGRTAMAEI FNPIQTQVFNT-VYNSDDNVFVGAPTCSGKTICAEF LNPMQSKIFEC-AFPSDENMLVCAPTGAKTNVAMM LNPIQSKVFAHA-AFEGDSNMLICAPTCSGKTNLALL LNRIQSKLYRA-ALETDENLLCAPTGAKTNVALM                 </pre> <p style="text-align: center;"><b>Alça-Q                      Motivo I</b></p>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> FILKFLLECAHSRSANCVSPVCGGNEEENNDGGESILRTEQKLYLTATEACAMRYNE ALLNHWR-----QNKCRAVYINPSCGKIDFLLSD AILRMLLQ-----SSGRCVYITPHEALAEQVYMD AILRAISNMMSR-----TCLVNLRELKVVYVAPMKALVQEVVRT TVLKALSHYHPK-----TRKLMLSAFKI VYLAIPKALVQEVRE CMLREIKGHINMD-----CTINVDDFKI IYLAIPMSLVQEMVGS                 </pre> <p style="text-align: center;"><b>Motivo Ia</b></p>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> WRYKFGBELNQKVAKL EPHYGEELAIKAEKVRGATIIISCGSSFAPLL RHG-AMDFLSAIT WNKRFSHLACGKII--MKLCMDPSLNLKLLAKSHVLLATPVQFELL SRMRQRKNIQSLE WYKFKQDRLNKKV--L-LTGETS TD LKL LKGNIIISTPERMDILSRMRQRKNIQSLE FSARLEPL--GLAV--IELTGDSCANQAQVCGAQLIVTT PERMDVVTRKSVELCVASL LK FQRRLAFL--GIRV--AELTGDSSRLSRKQIDETQVLVSTPERMDITRNSNMLAIVELVR FGKRLATY--GITV--AELTGDHQLCKEISATQIIVCTPERMDITRKGERTYTVQLVR                 </pre> <p style="text-align: center;"><b>Motivo Ib (parte 1)      Motivo Ib (parte 2)</b></p>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> HVIIVDHVHLL-RAPEGRMMEECVARLQSKPYIVN-RCQC PARLLALSYPLISCTEVSRRM LMIYDDAHEISQCVYCAVYETLISPMI---FIATQ-LEKRI RFVCLSNCLANARDFGEWA LFVVDEVHLL-GEENCPVLEVICSRMR---YISSQ-IERPIRIVALSSSLSNARQVAHWL LLIIDEVHLL-HNE RGPVLEAIVARTL---LQQQHRCEAGIRLVGLSATLPPNFHDVASFL LLIIDEIHLH-DDRGPVLESIVARTF---WASKY-CQYYPRII GLSATLPPNYEDVCRFL LLIILDEIHLH-DDRGPVLEALVAPAI---RNLEM-TQEDVRLI GLSATLPPNYEDVATFL                 </pre> <p style="text-align: center;"><b>Motivo II                      Motivo III</b></p>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> KV-PTARQYNYGNSYRQLRVRLEA--VROFSARSRYAAAATISVLKLLQNDRYAVSPCVIF GM-TKSNIIYNFSPSERIEPLEINIQSFRQVVEHISFNFSMLQMAFRAASAAAAGNPNSSVF GC-SATSTFNHFNVRPVPLELHIQCFMISHTQTRLLSMAKPVYHAIT-IRHSPKPVIVF QVDRQRGLFVFDSSYRPIPLQQTFCVAVKVRGCTNQAAVMNLLVYDKVLAAATEGAQSLVF RV-PKRGCLFYFDSSFRPCPLSQQFCGIKRMSLKLKAMNDACYEKVLESINEGNQIIVF RVDPAKGLFYFDNSFRPVPLEQTYVGTITRKAIRKQIMNEIVYKIMRHAG-RNQVLFV                 </pre> <p style="text-align: center;"><b>Motivo IV</b></p>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> VPTAQDAEELARRIVLRCRDFVPTD-----ACEDVEDRQLALLLASC LPSRRDCMEVASAFMKF--SKAIEWDMLNAREE--QI---VPYIEKLTGCHLRAPLKHG VPSRQQLTALDILITCAADIQQRFLHCTEK--DL---IPYLEKLSDS TLKRTLLNG VHSRKD TDYALTYIVRQVDDKRTHYFVRPGSDS EQVLRREAVSDPNSLRPSIQQMLPFG VHSRRETSRTATWLRNKFAEENITHKLTRNDACSRQILKTE---AANVLDPSLRKLEISG VHSRRETKTARAIRDMCLKFDTLGLFLREGSASTEVLRTE---AEQCKML ELKDL L P YC                 </pre>
CT_U5-200K_Thrucei CT_Brr2_Scerevisiae CT_U5-200K_Hsapiens NT_U5-200K_Thrucei NT_Brr2_Scerevisiae NT_U5-200K_Hsapiens	<pre> VAYMHRGSSLLDELNI TEKVERPARHPKTDPLPLYLVC SFRAAWRLPAALFATAFVCAA VGI LYRGMASNDERIVKRLY-----EYGAVSVLLISK-DCSAFACTDEVIIL- VGYLHEGLSPMERLRVQLF-----SSGAIQVVVASRSLCUMNVAHLVIIM- FAVHHAGMSREERSLVESLF-----AGCHVRVLCVTS TLAUGVNLP AHQVLIK- IGTHHAGLTPSDRSLSKDLF-----ADGLLQVLVCTATLAUGVNLP AHQVLIK- FAIHHAGMT RVDRTLVKDLF-----ADKHIQVLVSTATLAUGVNLP AHQVLIK-                 </pre> <p style="text-align: center;"><b>Motivo IVa                      Motivo IVb      Motivo IVc</b></p>

continua



continuação

```

CT_U5-200K_Tbrucei ERIGNVCEEDKSEVCGDFATDCSVSELQMTSPA-----LNRAVVYCRAAARRWVWSR
CT_Brr2_Scerevisiae ---G---TNLYDGAHKH-YMPTYTINE LLEMVGLASGN-DS-MAGKVLILTSHNMKAYYRK
CT_U5-200K_Hsapiens ---D---TQYYNCKIHA-YVDYPIYDVLQMVGHANRPLQD-DRCRCVIMCQGSKKDFPKK
NT_U5-200K_Tbrucei ---G---TRVFNAAKGE-T ELLNALDVLQMFGRAGRVCFGSTLGRAAVITS AEDLQYYLS
NT_Brr2_Scerevisiae ---G---TDVYSPEKGS-WEQLSPQDVLQMLGRAGRPRYD-TFEGEIIITDQSNVQYYLS
NT_U5-200K_Hsapiens ---G---TQVYSPEKGR-WTELGALDILQMLGRAGRPOYD-TRGEGILITSHGELQYYLS
    
```

**Grupo-de-cabelo β Motivo VI**

```

CT_U5-200K_Tbrucei LLNDPLPIESHLYPEDFRDINTAVAQGRADMVDLRLILQSHYFLHHLRNTNLHFYCVF
CT_Brr2_Scerevisiae FLIEPLPTESYLQY--I IHDTLNNEIAMSIIQSKQDCVDWFYTSYFYRRIHVNPSSYYGVR
CT_U5-200K_Hsapiens FLYEPLPVESHLDH--CMHDHFMAEIVTKTIENKQDAVDYLWTFFLYRMTQNPNYNLQ
NT_U5-200K_Tbrucei VLNNQLPIESQMMR--RUVDMLMAEVVLCGHI TNLDEGVRWLQRTYLYIRMRRAPEIYGAR
NT_Brr2_Scerevisiae VLNQQLPIESQFVS--KLVDNLMAEVVAGNICKRNDAVNWLAATYLYVRLAS PMLYKVP
NT_U5-200K_Hsapiens LLNQQLPIESQMWV--KLPDMLMAEIVLGNVQNAKD AVNWLCYAYLYIRMLRSPTLYGIS
    
```

**Domínio WH**

```

CT_U5-200K_Tbrucei SKR-----DIPAYASEFARSVIASLQRTCCVTSSSNSAS-----STGCDGDELTAQPTPR
CT_Brr2_Scerevisiae D--TSP-HGISVFLSNLWETC LMDLVSSFI EIDDT EAVTAEVNGGDDEATEIIS TLSM
CT_U5-200K_Hsapiens C--ISH-RHLSDHLSELVEQT LSDLEQSKCISIEDEM-----DVAPLML
NT_U5-200K_Tbrucei A--SSNDPLLRLHGNIVHTAADDLRPSQMVEYD SNT-----HRIATSL
NT_Brr2_Scerevisiae D--ISSDGLKKRFRESLVHSA LCI LKEQELVLYDAEN-----DVI EATDL
NT_U5-200K_Hsapiens HDDLKGDP LLDQRRLD LVHTAALMLDKMNLVKYDKKT-----GNFQVTEL
    
```

```

CT_U5-200K_Tbrucei GVAAMRHGI
CT_Brr2_Scerevisiae GLIASHYGV
CT_U5-200K_Hsapiens GMIAAYYI
NT_U5-200K_Tbrucei GRIASHYYL
NT_Brr2_Scerevisiae GNIASSFYI
NT_U5-200K_Hsapiens GRIASHYYI
    
```

```

CT_U5-200K_Tbrucei SIESLEAIDRAVSANSRCGDSVTRMWRVIASCCVEL--TPQHVGDAARIVD
CT_Brr2_Scerevisiae SFFTIQSFVSSLSNTS T----LKNMLYVLS T-AVEFESI PLRCDRALLV-
CT_U5-200K_Hsapiens NYTTIELFMSMLNARKT----VRGLIEIISN-AAEYENIPI RHEDNL LR-
NT_U5-200K_Tbrucei TTTSMATYLYLCNTMH----DVDLFRF SM-SKEFSHI IVRPEEQSQLQ-
NT_Brr2_Scerevisiae NHASMDVYNRELD EHT T----QIDLFRF SM-SEEFKYVSVRYE EKRE LK-
NT_U5-200K_Hsapiens TNDTVQTYNQLKPTLS----EIELFRVFSL-SEEFKNI TVREE EKLELQ-
    
```

**Domínio catraca**

```

CT_U5-200K_Tbrucei IAE LSALHVVARAF PHTYD VQYLDMD FSKGWTRVHL LVL AHCARMFVL PVS CDDGCV EAE
CT_Brr2_Scerevisiae ---KLSKRLPLRFPKH---TFS-----
CT_U5-200K_Hsapiens ---QLAQKVPKHLNMP---KFN-----
NT_U5-200K_Tbrucei ---YLL ENAPIAVRES---RYT-----
NT_Brr2_Scerevisiae ---QLLEKAPIPIRED---IDD-----
NT_U5-200K_Hsapiens ---KLL ERVPIPVRES---IEE-----
    
```

```

CT_U5-200K_Tbrucei RGEIDALHPFASRSSLAVEPQLLLQIPHSV----ARPLEDLNVL LPVVDLVRCVSEIF
CT_Brr2_Scerevisiae -----GSVSFRKVFLLLQAYFSRL ELPV-DFQNDLKD IL EKVVPLINVVVDIL
CT_U5-200K_Hsapiens -----DP-HVKTNLLQAHLSRMQLSA-ELQSDTEELSKAIRLIQACVDVL
NT_U5-200K_Tbrucei -----P--LAKINVLQCYISNMNLQCL PLMSEM SYVKD SAQRILRALYEIS
NT_Brr2_Scerevisiae -----P--LAKVNVLLQSYFSQLRFE CFALNSDMVF IHQNA GRLLPAMFEIC
NT_U5-200K_Hsapiens -----P--SAKINVLQAFISQLKLECFALMADMVYVTQSACRLMRAIFEIFV
    
```

continua

continuação

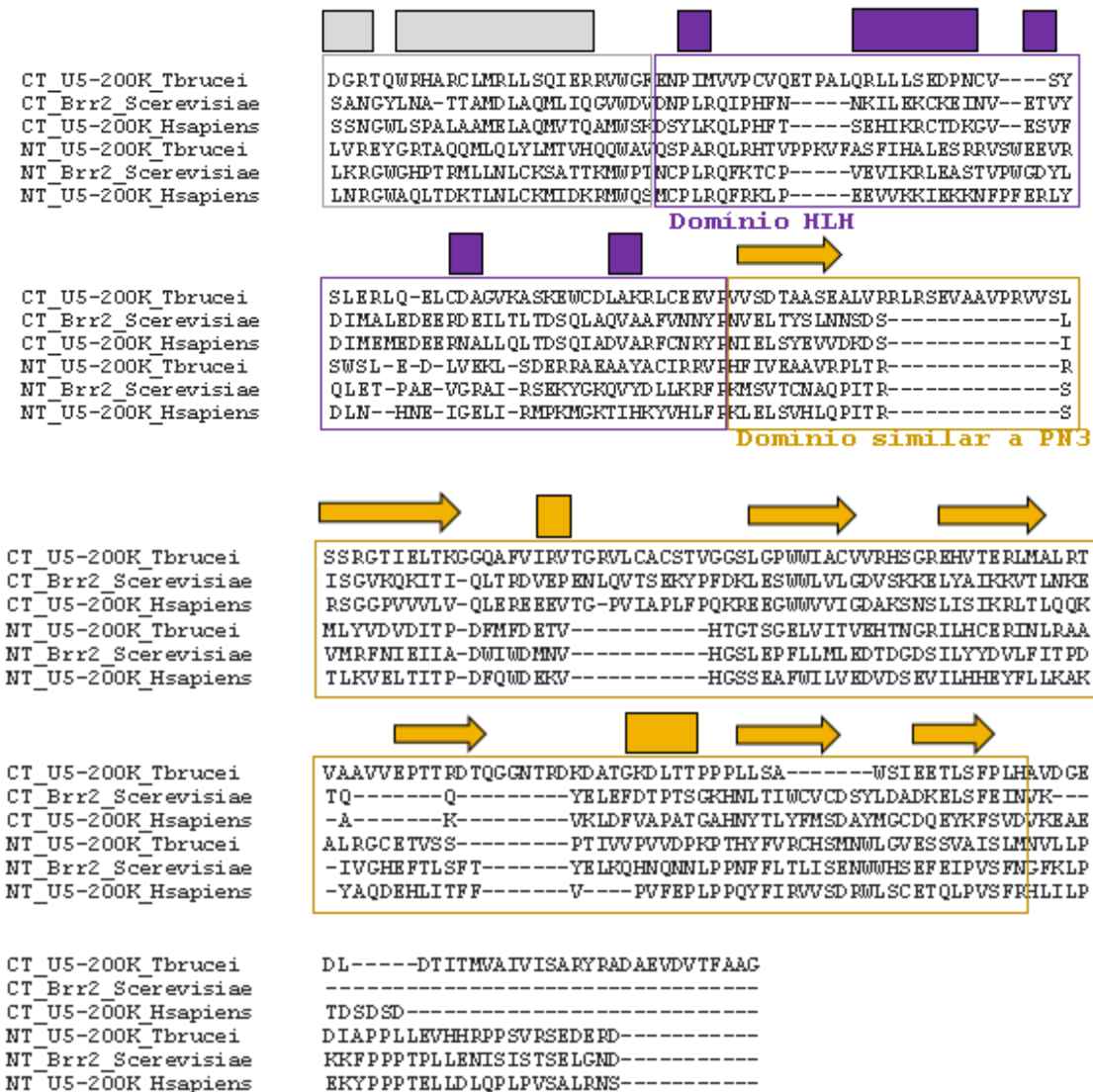


Figura 3.6 – Alinhamento múltiplo de seqüências de U5-200K conservadas em *T. brucei*, *S. cerevisiae* e *H. sapiens*. NT = região N-terminal e CT = região C-terminal.

Fonte: Elaborada pelo autor.

O domínio Sec63 do cassete N-terminal de Brr2 de levedura interage diretamente com Prp8 e Snu66, além do snRNA U4, enquanto a região entre os dois cassetes helicase/ATPase interage com o núcleo de proteínas Sm (SmB, SmD1 e SmD2).<sup>57, 94, 118</sup> A interação Prp8 foi mostrada como essencial para a ativação do *cis-splicing* e a ausência do cassete C-terminal contendo o domínio helicase e o domínio Sec63 na proteína Brr2 de *S. cerevisiae* levou à morte celular.<sup>57, 94, 118</sup> No caso da proteína de *T. brucei*, ainda não se sabe qual a função desta região C-terminal não conservada.

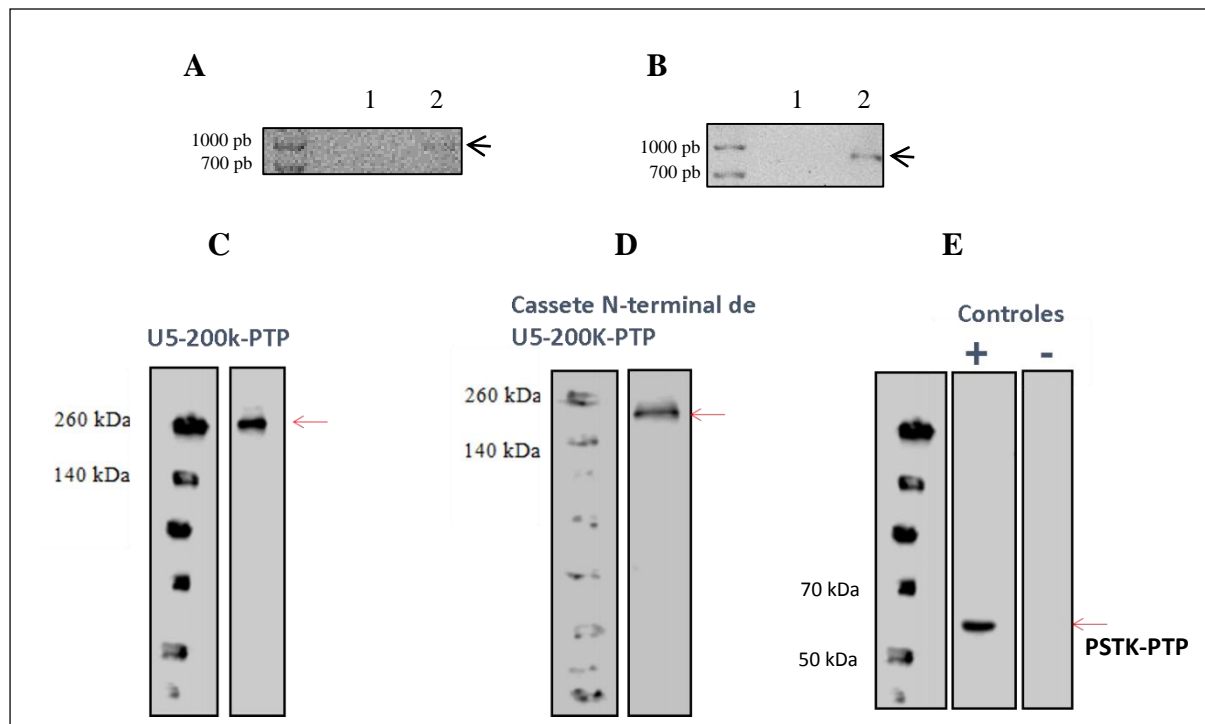
### 3.4.2 Clonagem e expressão homóloga de U5-200K de *T. brucei*

Dada a verificação de que o cassete helicase/ATPase/Sec63 C-terminal de U5-200K não é conservado em *T. brucei*, foi planejada a expressão homóloga do cassete helicase/ATPase/Sec63 N-terminal (mais sua extremidade N-terminal que contém os domínios plugue e PWI) ligada à extensão PTP na extremidade C-terminal para comparação com a expressão homóloga da proteína completa.

#### 3.4.2.1 Amplificação de regiões de interesse de U5-200K, clonagem, transfecção em *T. brucei* e seleção de clones positivos

As regiões de interesse da ORF de U5-200K de *T. brucei* foram amplificadas a partir do DNA genômico e ligadas em vetor de recombinação homóloga pC-PTP-NEO<sup>67</sup>. Após confirmação por sequenciamento, o plasmídeo linearizado foi transfectado em células de *T. brucei* 427 na forma procíclica. O vetor linear pC-PTP-NEO<sup>67</sup> contendo as sequências de U5-200K fusionadas à extensão PTP foi inserido no genoma do *T. brucei* e clones positivos foram obtidos para as construções de U5-200K completa e aquela contendo somente o cassete N-terminal. A Figura 3.7 mostra a confirmação por PCR dos clones positivos, com a banda entre 700 pb e 1000 pb correspondendo à amplificação de regiões de 627 pb e 893 pb, (mais a região de extensão PTP) para a construção completa e para o cassete N-terminal, respectivamente.

Os clones positivos também foram analisados para expressão homóloga das proteínas recombinantes por *Western Blotting* utilizando-se o anticorpo anti-ProtA, que reconhece a extensão PTP, e o resultado é mostrado na Figura 3.7. A extensão PTP adiciona 18,9 kDa à massa molecular da proteína expressa e, portanto, espera-se, para a proteína completa com extensão PTP uma massa molecular igual a 268,2 kDa, enquanto que, para a construção contendo o cassete N-terminal de U5-200K de *T. brucei* com extensão PTP, a massa molecular esperada é de 155,9 kDa.

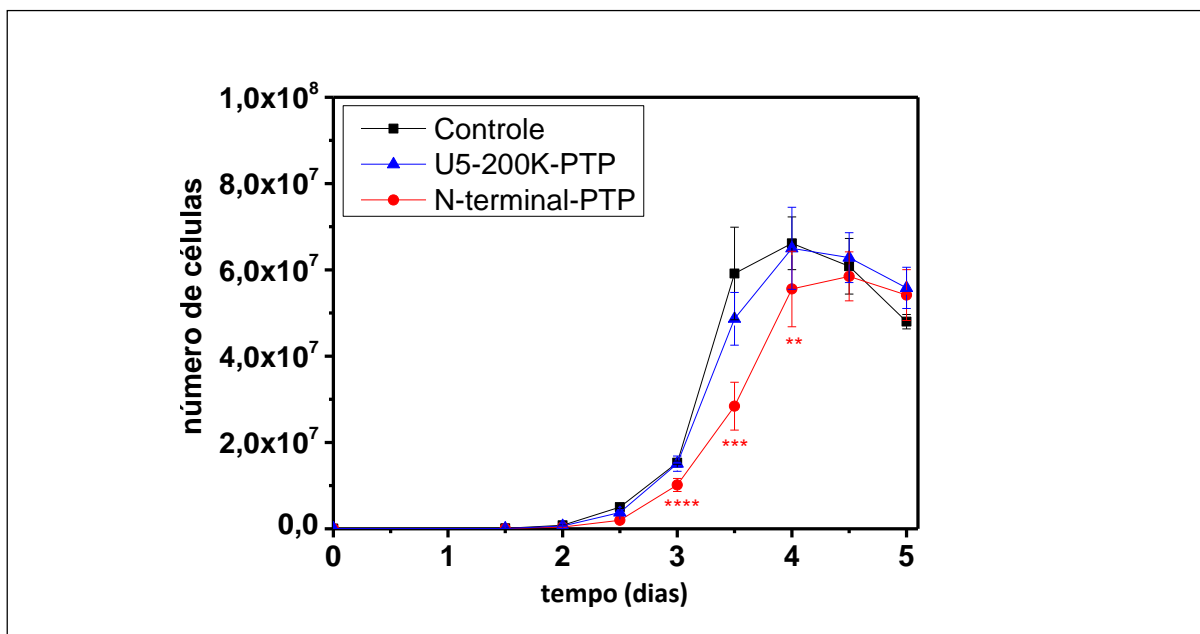


**Figura 3.7 – Clonagem e expressão homóloga de U5-200K de *T. brucei*.** A e B- Confirmação de inserção no DNA genômico de *T. brucei* 427 das construções de U5-200K completa/pC-PTP-NEO (1- controle negativo-cepa selvagem, 2- região amplificada de U5-200K completa) e U5-200K cassette N-terminal/pC-PTP-Neo (1- controle negativo-cepa selvagem, 2- região amplificada de U5-200K cassette N-terminal). Confirmação por *Western Blotting* da expressão homóloga das proteínas recombinantes com extensão PTP: C- U5-200K cassette N-terminal + PTP, D- U5-200K completa + PTP e E- controles positivo (correspondente à proteína PSTK de *T. brucei* com extensão PTP - 59 kDa) e negativo – cepa selvagem de *T. brucei* 427 na forma procíclica.  
 Fonte: Elaborada pelo autor.

Observa-se, na Figura 3.7, bandas no tamanho esperado para ambas as construções após reação de amplificação por PCR e *Western Blotting*, confirmando a inserção correta no DNA genômico e expressão das proteínas recombinantes com extensão PTP. Este resultado corresponde ao primeiro trabalho reportado na literatura em que o vetor pC-PTP-NEO<sup>67</sup> é utilizado com sucesso para o truncamento de um dos alelos de um gene em estudo. Deste modo, foi possível comparar o efeito do truncamento de U5-200K a partir de sua expressão homóloga em *T. brucei* sob a regulação do promotor nativo. Assim, interações proteína-proteína puderam ser analisadas comparativamente entre diferentes construções de U5-200K, como mostrado nas próximas seções.

### 3.4.2.2 Curvas de crescimento

As curvas de crescimento de cada população clonal a 26°C em meio de cultura SDM79<sup>257</sup> estão representadas no gráfico a seguir (Figura 3.8):



**Figura 3.8** – Curvas de crescimento de cultura de *T. brucei* na temperatura de 26°C para cada população clonal. As células foram contadas em câmara de Neubauer em microscópio óptico invertido. Segundo o teste T de Student aplicado aos resultados com hipótese nula igual à situação em que não diferença entre as curvas em cada ponto, \* indica diferença significativa com  $p < 0,05$ , \*\*  $p < 0,02$ , \*\*\*  $p < 0,01$  e \*\*\*\*  $p < 0,0001$ .  
Fonte: Elaborada pelo autor.

Todas as curvas apresentam o comportamento de crescimento canônico, com uma fase lag por dois dias, seguida de um período de crescimento exponencial (fase *log*) de dois dias, uma fase estacionária durante um dia, seguida de morte celular (Figura 3.8). Nota-se defeito de crescimento estatisticamente significativo durante a fase *log* das células que expressam o cassete N-terminal de U5-200K-PTP (Figura 3.8). Os resultados implicam na importância da função do cassete C-terminal de U5-200K de *T. brucei*.

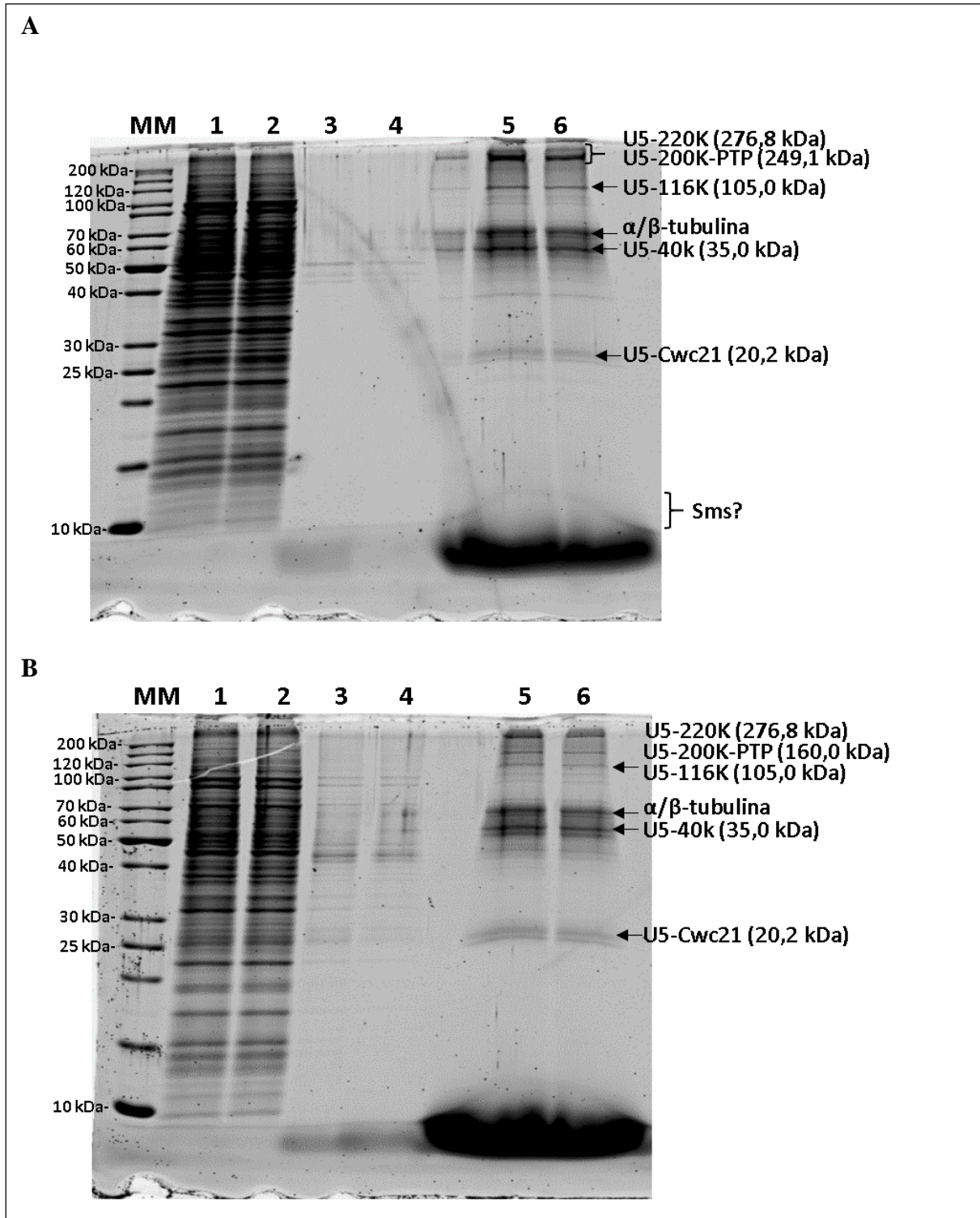
Entretanto, a ausência deste domínio não leva à morte celular, o que pode ser explicado pelo fato de que o alelo selvagem continua sendo transcrito e a proteína U5-200K selvagem deve competir com o cassete N-terminal-PTP pela ligação às proteínas que compõem a snRNP U5 e o spliceossomo. Nota-se ainda, que o crescimento celular não é alterado quando a célula expressa a proteína recombinante U5-200K-PTP completa, o que indica que a extensão PTP não deve ser tóxica para célula, como já observado para várias outras proteínas spliceossomais de *T. brucei*.<sup>60,66,81,102,120,124</sup>

Para uma análise mais aprofundada do efeito do truncamento da proteína U5-200K na atividade de *splicing* em *T. brucei*, foram realizados experimentos de purificação dos produtos recombinantes na forma de complexos macromoleculares, sua imunolocalização e também a análise dos produtos de *cis-* e *SL trans-splicing* nas populações clonais, apresentados a seguir.

### 3.4.3 Purificação dos produtos recombinantes e análise dos parceiros de interação por espectrometria de massas

Como observado no capítulo 1 (seção 1.3.2), a técnica de purificação em *tandem* de proteínas spliceossomais a partir de expressão homóloga em *T. brucei*, em combinação com mineração de dados do genoma deste organismo, permitiu a identificação de grande parte das proteínas que compõem as partículas ribonucleoproteicas deste organismo. Entretanto, muitas das proteínas específicas descritas em cada complexo do spliceossomo humano ainda não tiveram homólogas identificadas em *T. brucei*. Sabendo-se que a proteína U5-200K é parte da snRNP U5 em leveduras e humanos, posicionando-se no centro catalítico do spliceossomo,<sup>121, 256</sup> foi realizada a purificação da ortóloga de *T. brucei* com extensão PTP na extremidade C-terminal e de seu produto truncado no domínio helicase/ATPase N-terminal para caracterização do complexo spliceossomal copurificado.

A purificação dos produtos recombinantes foi analisada em SDS-PAGE 10% (Figura 3.9). Observa-se que proteínas contaminantes foram eluídas nas lavagens e após tratamento com a protease TEV, e aquelas que interagem especificamente com U5-200K foram copurificadas ao final do experimento (Figura 3.9). Além disso, as proteínas isoladas para cada uma das construções recombinantes foram concentradas e enviadas para análise por espectrometria de massas. O resultado é mostrado nas Tabelas 3.1 e 3.2 para as proteínas com pelo menos um peptídeo único detectado. Foram excluídas das tabelas as proteínas comumente detectadas como contaminantes em experimentos de purificação em *tandem*, que incluem chaperonas, elementos de transposição,  $\alpha$ - e  $\beta$ -tubulinas, proteínas ribossomais, entre outras.<sup>120,124,266</sup>



**Figura 3.9 – Purificação em *tandem* de U5-200K-PTP (A) e do seu cassete N-terminal-PTP (B).** MM = marcador de massa molecular, 1 = lisado celular (0,003%), 2- amostra que não interagiu com resina de IgG sepharose (0,003%), 3- eluato após tratamento com protease TEV (5%), 4- amostra que não interagiu com resina contendo anticorpo anti-proteína C imobilizado (5%), 5 e 6- eluato final (100%). As proteínas foram identificadas por espectrometria de massas no laboratório da Dra. Julia Chagas Pinheiro da Cunha, no Instituto Butantan.  
Fonte: Elaborada pelo autor.

**Tabela 3.1 – Proteínas copurificadas com U5-200K-PTP e identificadas por espectrometria de massas.**

Código na <i>TriTrypDB</i> <sup>53</sup>		Pontuação	Cobertura (%)	# Peptídeos únicos	# Peptídeos	# PSMs	# Resíduos	MM (kDa)	pI teórico
Tb927.5.2 290	U5-200K, putativa (Brr2)	2052,94	29,06	62	62	76	2237	249,1	6,24
Tb427tmp. 211.2420	Produto não especificado (U5-220K/ Prp8)	1897,89	27,92	70	70	81	2403	276,8	8,27
Tb427tmp. 01.7080	Produto não especificado (U5-116K - Snu114)	619,86	20,84	20	20	30	974	105,4	5,27
Tb427tmp. 01.2940	Produto não especificado (U5-40K)	160,00	27,30	8	8	9	326	35,0	6,74
Tbg972.9. 1770	U5-Cwc21 hipotética, conservada	152,52	28,42	5	5	5	183	20,4	9,77
Tb427.07. 3120	Produto não especificado (SmD1)	75,94	15,09	1	1	1	106	11,7	8,15
Tb427.02. 4540	Produto não especificado (SmB)	74,76	31,19	4	4	4	109	12,3	9,92
Tb427.06. 2700	Produto não especificado (SmE)	57,45	24,42	2	2	2	86	9,6	8,19
Tb427tmp. 01.5915	Produto não especificado (SmG)	-	10,98	1	1	1	82	8,9	6,05
Tb427.10. 2900	Produto não especificado ( $\beta$ -importina)	48,93	1,04	1	1	1	866	95,0	4,78
Tb09.211. 3610	Enzima ativadora de ubiquitina E1, putativa	89,50	2,06	3	2	2	1214	134,6	6,70

\* # PSM corresponde ao número de espectros de peptídeos assinalados.

Fonte: Elaborada pelo autor.



**Tabela 3.2 - Proteínas copurificadas com o cassete N-terminal de U5-200K-PTP e identificadas por espectrometria de massas.**

Código na <i>TriTrypDB</i> <sup>53</sup>		Pontuação	Cobertura (%)	# Peptídeos únicos	# Peptídeos	# PSMs	# Resíduos	MM (kDa)	pI teórico
Tb427tmp.211.2420	Produto não especificado (U5-220K/Prp8)	257,42	5,16	12	12	13	2403	276,8	8,27
Tb427.05.2290	U5-200K, putativa (Brr2)	794,06	11,67	27	27	31	2237	249,3	6,29
Tb427tmp.01.7080	Produto não especificado (U5-116K/Snu114)	334,78	16,43	15	15	16	974	105,4	5,27
Tbg972.9.1770	U5-Cwc21, hipotética, conservada	190,93	32,79	6	6	6	183	20,4	9,77
Tb427tmp.01.2940	Produto não especificado (U5-40K)	62,66	6,75	2	2	2	326	35,0	6,74
Tb427.10.540	Sub2, putativa	41,38	2,29	1	1	1	436	49,2	6,24
Tb11.01.1680	Poliubiquitina, putativa	48,24	11,82	2	1	1	685	76,6	7,66

\* # PSM corresponde ao número de espectros de peptídeos assinalados.

Fonte: Elaborada pelo autor.

Observa-se, em ambos os géis (Figura 3.9), que o padrão de proteínas purificadas é praticamente o mesmo nos dois casos, exceto pela presença de algumas bandas menos intensas, que podem ser devidas a interações menos específicas de proteínas abundantes na célula. Ao final do experimento, o resultado é a copurificação de um subcomplexo da snRNP U5 que consiste de U5-220K, U5-200K-P, U5-116K, U5-40K e U5-Cwc21 para ambas as construções completa e truncada de U5-200K-PTP. Além disso, proteínas Sm e uma  $\beta$ -importina foram copurificadas com a construção completa com pontuação significativa de acordo com o experimento de espectrometria de massas. Em comparação, a purificação em *tandem* da proteína Brr2 de *S. cerevisiae* (homóloga de U5-200K) levou à identificação recente das proteínas componentes da tri-snRNP U4/U6.U5 deste organismo.<sup>94</sup>

Curiosamente, os membros de snRNP U5 conhecidos como U5-15K (Dib1) e U5-102K (Prp6) não foram copurificados com este subcomplexo de snRNP U5 de *T. brucei*. As proteínas U5-100K (Prp28) e U5-52K, conservadas em snRNP U5 de mamíferos, também não foram copurificadas com U5-200K-PTP. A mineração de dados do genoma de *T. brucei* também não

aponta a conservação de tais proteínas. Por outro lado, a proteína U5-40K, ausente em levedura e presente em *H. sapiens*, foi detectada neste trabalho com massa molecular entre 50 e 60 kDa. Como a sua massa molecular teórica é de ~35 kDa, sugere-se uma possível modificação pós-traducional. Deve-se notar que peptídeos que codificam para regiões da sequência de aminoácidos de uma enzima de ativadora de ubiquitina foram detectados com pontuação significativa para o produto de copurificação de U5-200K-PTP completa e para poliubiquitina no caso do produto de copurificação de U5-200K-PTP truncada. Experimentos futuros são necessários para se investigar a natureza de tal modificação pós-traducional de U5-40K.

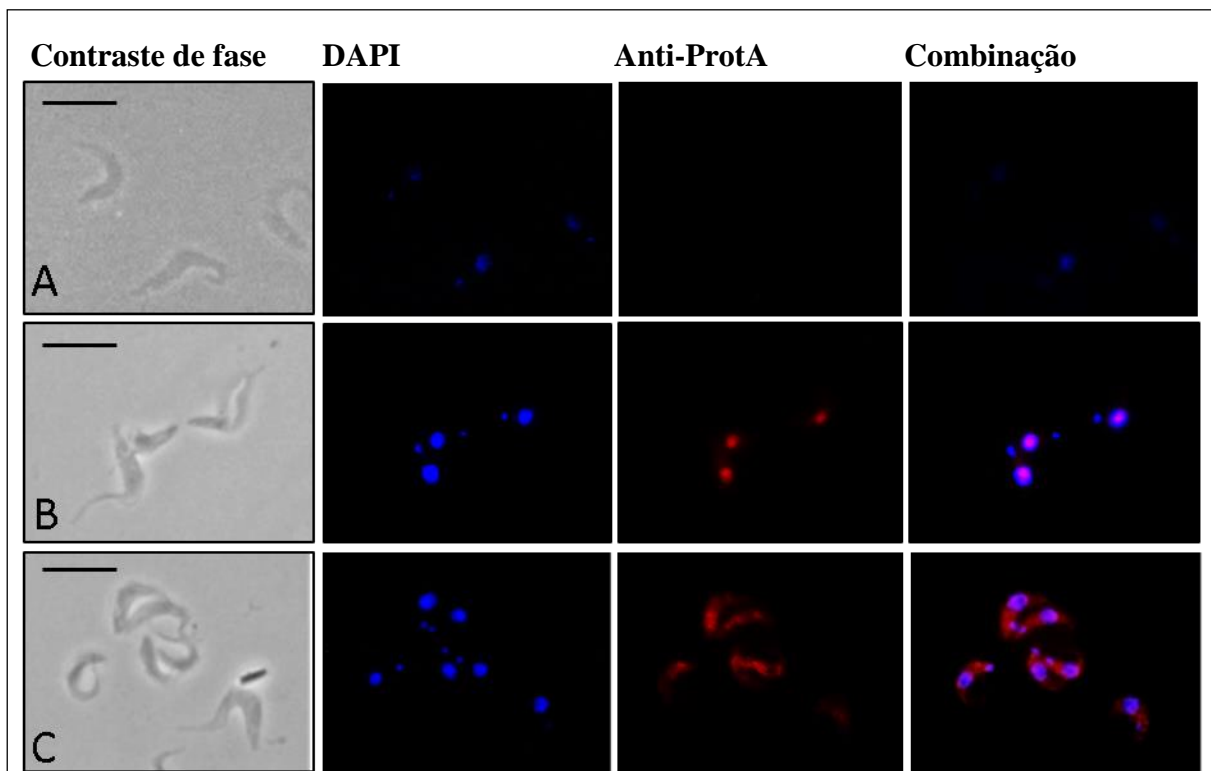
Outra proteína copurificada com U5-200K-PTP de *T. brucei* foi U5-Cwc21. Em levedura, Prp8, Snu114 e snRNA U5 interagem geneticamente com Cwc21 (ortóloga de SRm300/SRRM2 em humano), que não é essencial para a viabilidade do organismo.<sup>263-264</sup> Acredita-se que Cwc21 de levedura esteja envolvida no posicionamento do sítio 3'SS na transição para o segundo passo catalítico de *splicing*, sendo recrutada principalmente em condições de estresse, como em alta temperatura.<sup>263-264</sup> Por outro lado, um experimento de RNA de interferência em *T. brucei* realizado por da Ambrósio et al<sup>60</sup> mostrou que esta proteína é essencial para a viabilidade deste organismo e sua purificação em *tandem* levou ao isolamento do mesmo complexo de proteínas copurificado neste trabalho, o que reforça o fato de U5-Cwc21 ser membro da snRNP U5 em *T. brucei*.

Adicionalmente, a detecção de pequena quantidade de importina- $\beta$  copurificada com U5-200K-PTP completa sugere sua localização nuclear, já que esta proteína é requerida, juntamente com snuportina I para transporte nuclear de subcomplexos de snRNPs. Além disso, entre os contaminantes predominantes em ambas as purificações de U5-200K de *T. brucei* encontram-se as proteínas  $\alpha$ - e  $\beta$ -tubulina, que são abundantes na célula. Entretanto, ainda não foi publicado um estudo detalhado investigando a possível interação entre estas proteínas e o spliceossomo. Contaminantes minoritários comumente verificados em experimentos de purificação em *tandem* também foram detectados no experimento de espectrometria de massas.<sup>120,124</sup> No caso da purificação de U5-200K-PTP completa, foram verificadas as chaperonas HSP90 e retroelementos RHS (*retrotransposon hot spot*) das famílias 1, -4 e -5,<sup>265</sup> que estão principalmente concentrados no núcleo celular. Por outro lado, a purificação da construção N-terminal de U5-200K-PTP levou ao isolamento de pequenas quantidades de chaperona HSP70, de proteínas ribossomais e de proteínas de metabolismo celular, que estão localizadas no citoplasma.<sup>124, 266</sup> Os resultados indicam uma localização diferenciada para as duas construções de U5-200K de *T. brucei*, que foi melhor investigada por experimento de imunolocalização (seção 3.4.4)

Além disso, um peptídeo único para a RNA helicase Sub2 (UAP56 em *H. sapiens*) foi detectado na purificação de U5-200K truncada em seu N-terminal, localizada no citoplasma. hSub2 tem sido relacionada como integrante do complexo EJC no transito núcleo-citoplasma do mRNA maduro.<sup>267-268</sup> É a primeira vez que outra RNA helicase, além de U5-200K e Prp43, é detectada em experimentos de copurificação de complexos spliceossomais de *T. brucei*.

### 3.4.4 Imunolocalização dos produtos recombinantes

As proteínas recombinantes contendo extensão PTP no C-terminal foram analisadas quanto à sua localização subcelular por meio de microscopia de fluorescência. A Figura 3.10 mostra que a proteína U5-200K-PTP colocaliza com o marcador de núcleo celular e cinetoplasto DAPI, enquanto que o cassete N-terminal-PTP desta proteína permanece no citoplasma da célula (Figura 3.10).



**Figura 3.10 – Imunolocalização de U200K-PTP em microscopia de fluorescência.** *T. brucei* 427 na forma procíclica A- selvagem, B- expressando U5-200K-PTP completa e C- seu cassete N-terminal-PTP (B). A barra tem tamanho de 10  $\mu\text{m}$ .

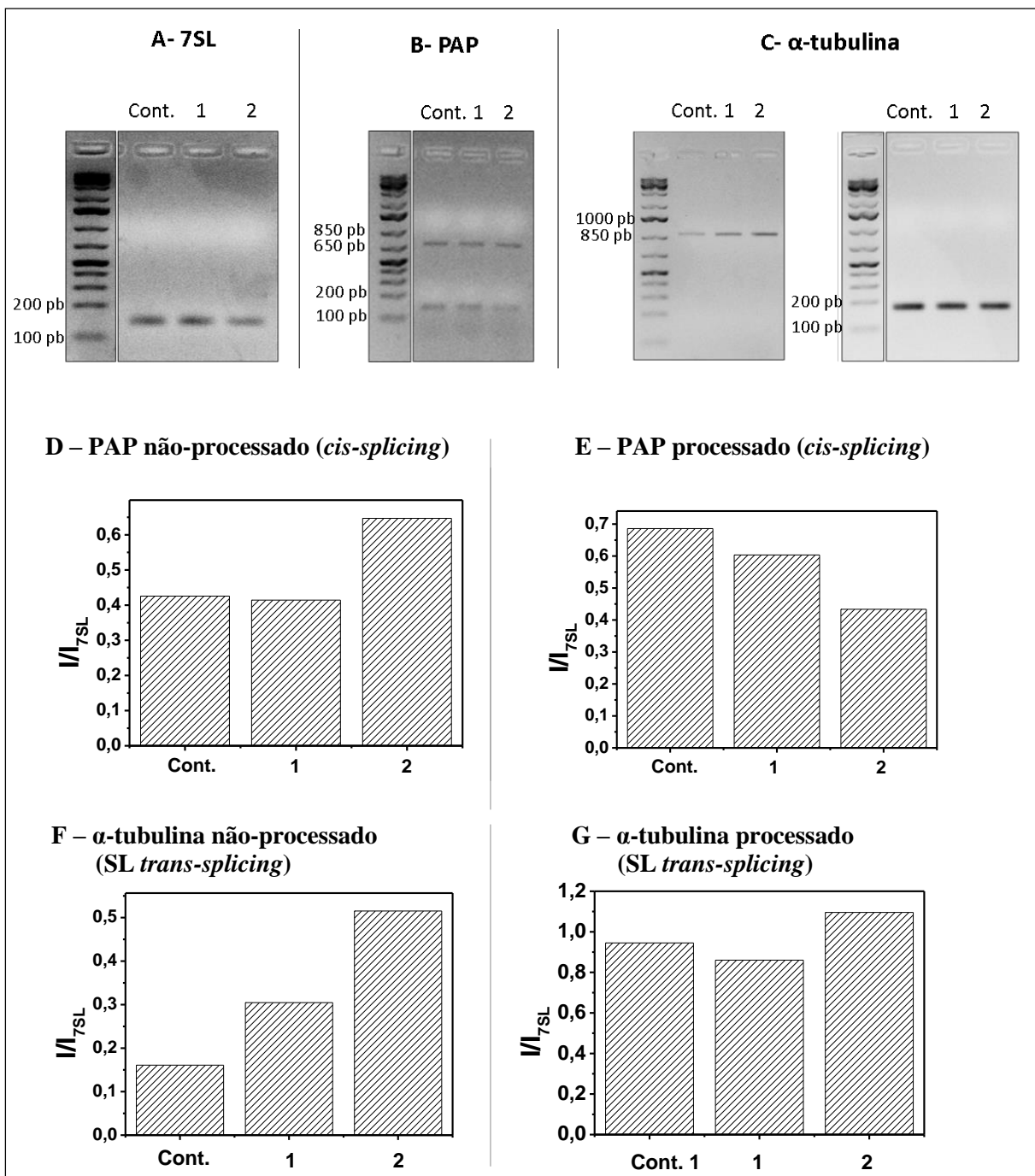
Fonte: Elaborada pelo autor.

A combinação destes resultados com aqueles mostrados na seção anterior sugere a pré-formação de um subcomplexo de snRNP U5 ainda no citoplasma contendo U5-220K, U5-220K, U5-116K, U5-40K e U5-Cwc21. O cassete helicase/ATPase/Sec63 C-terminal deve interagir com proteína(s) envolvida(s) em transporte nuclear e, na sua ausência, não há sinalização para direcionamento do subcomplexo ao núcleo. Curiosamente, as proteínas Sm não são copurificadas com o cassete N-terminal de U5-200K. Além disso, além das proteínas Sm, também foi detectado peptídeo único para importina- $\beta$  em análise de espectrometria de massas do produto copurificado de U5-200K-PTP completa.

Como mencionado no capítulo 1 (seção 1.3.2.4), acredita-se que a interação de SmB com SMN em *T. brucei* seja essencial para a biogênese de snRNP U5.<sup>61, 120</sup> Na ausência de SmB, o subcomplexo de snRNP U5 formado com o cassete N-terminal de U5-200K-PTP, como observado neste trabalho, o subcomplexo não deveria entrar no núcleo, como verificado. Sabendo-se que a construção cassete N-terminal-PTP de U5-200K de *T. brucei* não entra no núcleo, onde deveria exercer sua função no spliceossomo, esta população clonal disponibiliza de apenas um alelo funcional do gene que codifica para esta proteína, cujo papel é central no spliceossomo. Deste modo, seguiu-se com experimentos para análise dos produtos de *cis*- e *SL trans-splicing* em *T. brucei* (seção 4.4.5).

### 3.4.5 Avaliação de produtos de *cis*- e *SL trans-splicing* por meio de RT-PCR

A avaliação semi-quantitativa dos produtos de *cis*- e *SL trans-splicing* dos genes de PAP e  $\alpha$ -tubulina foi feita por RT-PCR, como mostrado na Figura 3.11.



**Figura 3.11 – Processamento do pré-mRNA por *cis*- e *trans-splicing* por RT-PCR.** A- Controle 7SL, B- *Cis-splicing* – PAP, C- SL *trans-splicing*  $\alpha$ -tubulina. D- e E- Quantificação das bandas observadas em B. F- e G- Quantificação das bandas observadas em C. Cont. = células controle, 1 = células que expressam U5-200K-PTP completa e 2 = células que expressam cassete N-terminal de U5-200K-PTP.

Fonte: Elaborada pelo autor.

Observa-se que, no caso do *cis-splicing* do gene que codifica para a proteína PAP, ambos os RNAs não-processados e processados apresentam quantidade (comparativa ao RNA controle 7SL, que não sofre *splicing*) relativamente iguais nas células controle e naquelas que expressam U5-200K-PTP, enquanto a quantidade de RNA processado diminui para as células

que expressam o cassete N-terminal-PTP. Já no caso do SL *trans-splicing* do pré-mRNA do gene que codifica para  $\alpha$ -tubulina, o número de moléculas de RNAs não-processados é menor do que o de mRNA maduro para as células controle (Figura 3.11).

Entretanto, no caso das células que expressam U5-200K-PTP, o número de moléculas de pré-mRNA não processado é maior (Figura 3.11), o que pode indicar que a extensão PTP afeta a produtividade do spliceossomo, provavelmente por diminuir a atividade helicase/ATPase de U5-200K. As células que expressam o seu cassete N-terminal-PTP, por sua vez, apresentam um número de moléculas não processadas consideravelmente maior, como esperado, já que as células expressam somente um alelo funcional.

### 3.5 Discussão

O rearranjo do centro catalítico do spliceossomo é dependente, entre outros fatores, da atividade helicase da enzima U5-200K, integrante da snRNP U5, da tri-snRNP U4/U6.U5 e dos complexos do spliceossomo. Esta enzima deve estar sob alta regulação, já que encontra seu substrato já na tri-snRNP U4/U6.U5, antes mesmo dos complexos do spliceossomo se associarem ao pré-mRNA.<sup>132,250-251,256</sup>

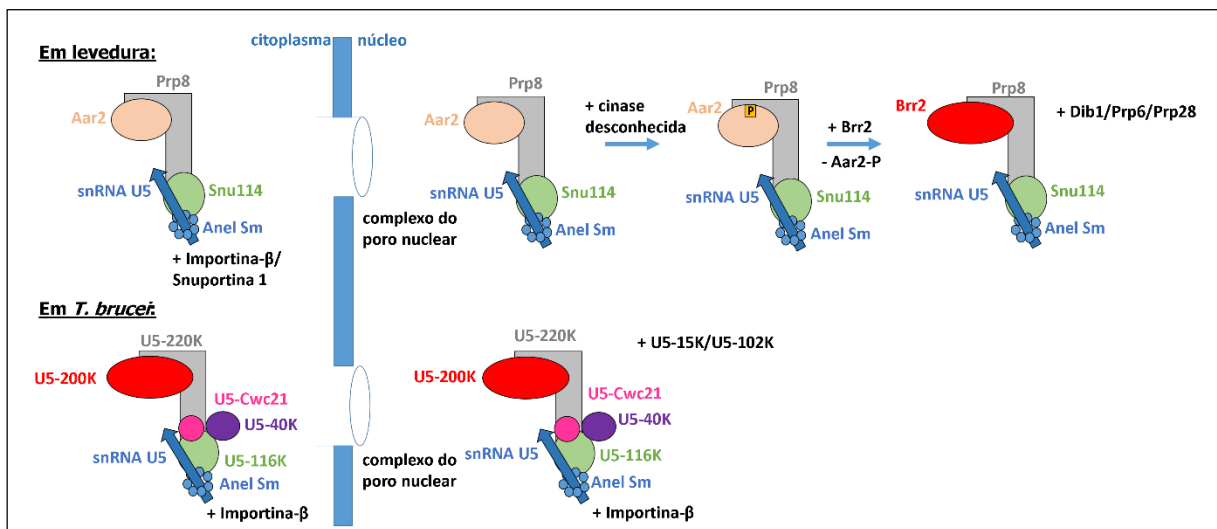
A proteína U5-200K de *T. brucei*, membro da família de helicases Ski2 (superfamília SF2), conserva os domínios plugue e PWI no N-terminal e o cassete helicase/ATPase/Sec63 N-terminal. Entretanto, várias mutações foram acumuladas no cassete C-terminal, o que faz com que este não seja identificado automaticamente por algoritmos de predição de domínios baseados em alinhamento múltiplo de sequências de proteínas da mesma família.

A expressão de U5-200K-PTP codificada por um dos alelos do gene de U5-200K em *T. brucei* não altera significativamente o crescimento da cultura das células na forma procíclica a 26°C. Além disso, experimento de purificação em *tandem* de U5-200K-PTP leva à copurificação de um subcomplexo de snRNP U5 composto de U5-220K, U5-200K, U5-116K, U5-40K e U5-Cwc21, além de proteínas Sm, localizado no núcleo celular, como era de se esperar, já que é o ambiente onde o spliceossomo deve realizar sua função.

Por outro lado, a expressão de um dos alelos do gene que codifica para U5-200K em *T. brucei* na sua forma truncada no cassete N-terminal resulta em um defeito de crescimento significativo a 26°C. A purificação em *tandem* desta proteína recombinante leva à copurificação do mesmo subcomplexo de snRNP U5, com exceção das proteínas Sm. Este resultado sugere a pré-formação de snRNP U5 ainda no citoplasma e o sinal para o transporte nuclear provavelmente está nas proteínas Sm. As células que a expressam têm apenas o alelo selvagem

sendo expresso como proteína funcional e as proteínas de snRNP U5 devem competir pela interação com as proteínas selvagem e recombinante truncada e, por isso, menos complexos spliceossomais funcionais devem estar disponíveis no núcleo celular. Isso pode explicar o defeito de crescimento das células que expressam a proteína truncada. Além disso, a análise de produtos de *cis*- e *trans*-splicing nestas células mostra que ambos os processamentos são afetados.

Os resultados apresentados neste capítulo apontam para um cenário de biogênese de snRNP U5 em *T. brucei* diferente daquele já caracterizado para levedura.<sup>117</sup> Pela primeira vez, observou-se que um subcomplexo de snRNP U5 se forma no citoplasma na presença de U5-200K. Em levedura, a proteína Aar2 substitui U5-200K no citoplasma e é fosforilada no núcleo por uma fosforilase ainda desconhecida como sinal para entrada de U5-200K em seu lugar. Humanos também conservam a proteína Aar2<sup>117</sup>. Entretanto, este mecanismo de regulação parece estar ausente em *T. brucei*, que não conserva Aar2 (Figura 3.12). Experimentos adicionais são necessários para se avaliar a possibilidade da fosforilação de U5-200K de *T. brucei* para a regulação de sua função, em analogia ao que se observa em levedura.<sup>117</sup>



**Figura 3.12 – Modelos de maturação da snRNP U5 em levedura e *T. brucei*.** A- Em levedura, um complexo pré-snRNP U5, formado por snRNA U5, proteínas Sm, Snu114, Prp8 e Aar2 entra no núcleo, em transporte mediado por snuportina 1 e importina-β. No núcleo, uma cinase desconhecida fosforila Aar2, que dá seu lugar para Brr2. O complexo maduro então interage com Dib1, Prp6 e Prp28. B- Já no caso de *T. brucei*, propõe-se que um complexo pré-formado de snRNP U5, contendo snRNA U5, proteínas Sm, U5-116K, U5-220K, U5-200K, U5-Cwc21 e U5-40K entre no núcleo em transporte mediado por, pelo menos, importina-β. No núcleo, este complexo deve interagir com o subcomplexo formado por U5-15K e U5-102K.  
Fonte: Elaborada pelo autor.

As proteínas U5-15K e U5-102K, não copurificadas com U5-200K-PTP, haviam sido obtidas na forma de heterocomplexo U5-15K/U5-102K por da Silva et al<sup>102</sup> a partir de

expressão homóloga de U5-15K-PTP em *T. brucei*. As estruturas de criomicroscopia de tri-snRNPs U4/U6.U5 de levedura e de humano mostraram recentemente que estas proteínas interagem entre si.<sup>56</sup> Além disso, a ortóloga de U5-15K está ligada à proteína (h)Prp8 no sítio ativo do spliceossomo e deve exercer função de regulação, sendo que após a ativação do spliceossomo, esta proteína é liberada juntamente com U5-102K<sup>102</sup>. A aluna de doutorado Msc. Ana Laura Lima, em sua dissertação de mestrado, identificou a autoproteólise da extremidade C-terminal de U5-15K de *T. brucei*.<sup>103</sup> É possível que esta proteólise esteja relacionada à regulação de sua função no spliceossomo, provavelmente na etapa de ativação do complexo B.

Adicionalmente, este trabalho apresenta a proteína U5-Cwc21 como parte integral da partícula U5 de *T. brucei*. Em levedura, esta proteína está relacionada ao posicionamento do 3'SS na segunda etapa de *splicing*, principalmente em pré-mRNAs onde o BP não é conservado, como é o caso em *T. brucei*.



## **CAPÍTULO 4**

### **CONSIDERAÇÕES FINAIS**

Neste capítulo são apresentadas a conclusão e as perspectivas deste trabalho.



## 4.1 Conclusão

A revisão bibliográfica apresentada no capítulo 1 introduz a família Trypanosomatidae como responsável por três das dezessete doenças negligenciadas classificadas pela Organização Mundial da Saúde (OMS), que acometem centenas de milhões de pessoas em todo o mundo atualmente. O organismo *Trypanosoma brucei*, membro desta família e causador da doença do sono, é considerado um organismo modelo para o estudo comparativo de processos moleculares de eucariotos, já que é de fácil cultivo em laboratório, seu conteúdo genético já se encontra sequenciado e ferramentas para estudos funcionais estão disponíveis. A comparação do mecanismo de processamento do pré-mRNA catalisado pelo spliceossomo no parasita, conhecido como SL *trans-splicing*, com o *cis-splicing* caracterizado no hospedeiro humano e em leveduras, revelou alguns pontos de diferença fundamental em termos de conteúdo proteico e de snRNAs desta maquinaria. Deste modo, tal mecanismo pode ser considerado um alvo em potencial para a busca de fármacos contra doenças causadas por tripanossomatídeos.

Os snRNAs do *trans-spliceossomo*, em geral, são menores e não conservam as principais regiões de reconhecimento de sinais de *splicing*. Para a liberação de um íntron na forma de Y, ao invés de um íntron na forma de laço como no *cis-splicing*, o *trans-spliceossomo* de *T. brucei* se associa a partir de snRNPs U1, U2 e U4/U6.U5, que contêm proteínas específicas, como é o caso de U1-24K na snRNP U1, das variantes de proteínas Sm (Sm15K/Sm16,5K na snRNP U2 e SSm4 na snRNP U4), e de U5-Cwc21 na snRNP U5, cuja função ainda é desconhecida.

O estudo dos núcleos de snRNPs U2, U4 e U5 de *T. brucei* apresentado no capítulo 2 é a primeira análise estrutural de que se tem notícia de proteínas spliceossomais de um parasita humano. A estrutura cristalográfica de U2A'/U2B'' de *T. brucei* mostra que a organização tridimensional do heterodímero é conservada, sendo que deve haver uma movimentação da alça do domínio RRM de U2B'' para a interação com a estrutura SL4 do snRNA U2. A proteína U2A', que interage com U2B'' na região de folha- $\beta$ , revela uma estrutura enovelada como uma longa hélice- $\alpha$  na sua extremidade C-terminal, distante da região de interação com RNA.

A combinação de modelagem molecular comparativa com experimentos de SAXS permite a identificação de possíveis regiões de interação entre os complexos de proteínas Sm com as proteínas componentes das partículas U2 e U4/U6.U5, a partir da sobreposição com estruturas de alta resolução de tri-snRNP U4/U6.U5 e do spliceossomo pós-catalítico de levedura. No caso do anel Sm canônico, as extremidades de SmB e SmD3 devem interagir com o snRNA U5, enquanto as regiões C- e N-terminal das variantes Sm15K e Sm16,5K na snRNP U2 devem interagir diretamente com U2A', possivelmente na sua região N-terminal. Por outro

lado, a variante SSm4 deve interagir com a região C-terminal de U5-200K e a região N-terminal de U5-220K.

As proteínas Sm de *T. brucei* interagem fortemente com o snRNA cognato. Curiosamente, snRNAs não-cognatos também se associam a estas proteínas. Pelo menos no caso de snRNP U2, a análise comparativa de interação de proteínas Sm do núcleo que contém Sm15K/Sm16,5K, bem como do subcomplexo U2A'/U2B'', com snRNAs cognatos e não-cognatos aponta para um cenário de biogênese desta partícula iniciando-se com a interação das proteínas Sm ao sítio rico em uridinas no snRNA U2 de *T. brucei*, localizado entre as hélices SL2b e SL3. Em seguida, U2A'/U2B'' se associam a este pré-complexo, formando o núcleo de snRNP U2. A associação ordenada de proteínas específicas ao snRNA U2 pode estar relacionada à especificidade do anel Sm que contém Sm15K/Sm16,5K à partícula U2.

A biogênese da partícula U5 de *T. brucei*, por sua vez, foi investigada no capítulo 3. A expressão homóloga comparativa de U5-200K completa e truncada no seu cassete helicase/ATPase N-terminal em *T. brucei* mostra localização subcelular diferenciada para as proteínas recombinantes. A proteína completa está localizada no núcleo, onde forma um complexo com U5-220K, U5-116K, U5-40K, U5-Cwc21 e proteínas Sm antes de ser incorporada ao spliceossomo. Já a proteína truncada não é transportada ao núcleo, o que leva a um defeito de crescimento de células em cultura. Este resultado pode ser explicado pelo efeito negativo no processamento do pré-mRNA por *cis* e *trans-splicing*, já que apenas a proteína U5-200K selvagem deve fazer parte do spliceossomo no núcleo. A proteína truncada compete pela ligação às proteínas U5-220K, U5-116K, U5-40K e U5-Cwc21, no citoplasma, o que deve resultar em um menor número de spliceossomos ativos no núcleo.

Deste modo, os experimentos de espectrometria de massas dos produtos copurificados e de imunolocalização de U5-200K-PTP apresentados no capítulo 3 indicam que um subcomplexo de snRNP U5 contendo U5-220K, U5-200K, U5-116K, U5-40K e U5-Cwc21 é transportado ao núcleo em mecanismo mediado por importina- $\beta$ . No núcleo, este subcomplexo deve se associar ao heterodímero U5-15K/U5-102K. Este é um ponto de diferença fundamental entre *T. brucei* e eucariotos superiores, que conservam a proteína Aar2, responsável por regular a associação de U5-200K à snRNP U5, somente no núcleo.

Os resultados apresentados nesta tese contribuem significativamente para a caracterização comparativa das variações dos núcleos Sm das snRNPs U2, U4 e U5 de *T. brucei*, em particular, para a biogênese das partículas U2 e U5. Novos estudos que objetivem o entendimento da função do *trans-spliceossomo* de tripanossomatídeos e que busquem pequenas moléculas moduladoras desta função poderão se apoiar nos dados aqui apresentados em

complemento à literatura científica disponível sobre o tema, que foi acumulada principalmente nos últimos vinte anos.

## 4.2 Perspectivas

Como perspectivas do trabalho, temos:

- Obter estruturas de mais alta resolução de complexos Sm de *T. brucei* a partir de experimentos de cristalografia. Para isso, pode-se utilizar de tratamento com proteases dos complexos Sm obtidos a partir de reconstrução *in vitro*, eliminando-se assim regiões flexíveis que inibir a formação de cristais.
- Realizar experimentos de cristalização do complexo U2A'/U2B'' de *T. brucei* em interação com SL4 do snRNA U2, para comparação com a estrutura cristalográfica obtida neste trabalho na ausência de snRNA.
- Investigar os parceiros de interação de Sm16,5K por meio de experimento de purificação em *tandem* a partir de expressão homóloga de proteína recombinante contendo extensão PTP. O complexo de proteínas copurificadas poderá ser submetido à análise estrutural por meio de criomicroscopia eletrônica.
- Analisar a estrutura do subcomplexo da snRNP U5 copurificado com U5-200K-PTP por criomicroscopia eletrônica.



## REFERÊNCIAS

- 1 WORLD HEALTH ORGANIZATION - WHO. **Neglected tropical diseases**. Disponível em: <[http://www.who.int/neglected\\_diseases/en/](http://www.who.int/neglected_diseases/en/)>. Acesso em: maio 2016.
- 2 ADL, S. M. et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. **Journal of Eukaryotic Microbiology**, v. 52, n. 5, p. 399-451, 2005.
- 3 SIMPSON, A. G. B., STEVENS, J. R., LUKES, J. The evolution and diversity of kinetoplastid flagellates. **Trends in Parasitology**, v. 22, n. 4, p. 168-174, 2006.
- 4 SUDARSHI, D., BROWN, M. Human African trypanosomiasis in non-endemic countries. **Clinical Medicine**, v. 15, n. 1, p. 70-73, 2015.
- 5 SIMARRO, P. P. et al. The atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. **International Journal of Health Geographics**, v. 9, p. 57, 2010. doi: 10.1186/1476-072X-9-57.
- 6 FRANCO, J. R. et al. Epidemiology of human African trypanosomiasis. **Clinical Epidemiology**, v. 6, p. 257-75, 2014 2014.
- 7 SIMARRO, P. P. et al. The human African trypanosomiasis control and surveillance programme of the world health organization 2000-2009: the way forward. **Plos Neglected Tropical Diseases**, v. 5, n. 2, e1007, 2011. doi: 10.1371/journal.pntd.0001007.
- 8 FRANCO, J. R. et al. The journey towards elimination of gambiense human African trypanosomiasis: not far, nor easy. **Parasitology**, v. 141, n. 6, p. 748-760, 2014.
- 9 MESSENGER, L. A., MILES, M. A., BERN, C. Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. **Expert Review of Anti-Infective Therapy**, v. 13, n. 8, p. 995-1029, 2015.
- 10 ALONSO-PADILLA, J., RODRIGUEZ, A. high throughput screening for anti-trypanosoma cruzi drug discovery. **Plos Neglected Tropical Diseases**, v. 8, n. 12, e3259, 2014. doi: 10.1371/journal.pntd.0003259. eCollection 2014.
- 11 URBINA, J. A. Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. **Acta Tropica**, v. 115, n. 1-2, p. 55-68, 2010.
- 12 KEVRIC, I., CAPPEL, M. A., KEELING, J. H. New world and old world leishmania infections: a practical review. **Dermatologic clinics**, v. 33, n. 3, p. 579-93, 2015.
- 13 BEZERRA DE MENEZES, J. P. et al. Advances in development of new treatment for leishmaniasis. **Biomed Research International**, v. 2015, ID 815023, 2015. doi: 10.1155/2015/815023.
- 14 STEVERDING, D. The history of African trypanosomiasis. **Parasites & Vectors**, v. 1, n. 3, 2008. doi: 10.1186/1756-3305-1-3.
- 15 MAJEKODUNMI, A. O. et al. A longitudinal survey of African animal trypanosomiasis in domestic cattle on the Jos Plateau, Nigeria: prevalence, distribution and risk factors. **Parasites & Vectors**, v. 6, n. 1, p. 239, 2013.
- 16 MOLYNEUX, D. H. Neglected tropical diseases: now more than just 'other diseases' - the post-2015 agenda. **International Health**, v. 6, n. 3, p. 172-180, 2014.
- 17 LEE, B. Y. et al. Global economic burden of Chagas disease: a computational simulation model. **Lancet Infectious Diseases**, v. 13, n. 4, p. 342-348, 2013.

- 18 MARINHO, D. S. et al. Health economic evaluations of visceral leishmaniasis treatments: a systematic review. **Plos Neglected Tropical Diseases**, v. 9, n. 2, e0003527, 2015. doi: 10.1371/journal.pntd.0003527. eCollection 2015.
- 19 FEVRE, E. M. et al. The burden of human African trypanosomiasis. **Plos Neglected Tropical Diseases**, v. 2, n. 12, e333, 2008. doi: 10.1371/journal.pntd.0000333.
- 20 GUNZL, A. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? **Eukaryotic Cell**, v. 9, n. 8, p. 1159-1170, 2010.
- 21 MURRAY, C. J. L., LOPEZ, A. D. Alternative projections of mortality and disability by cause 1990-2020: global burden of disease study. **Lancet**, v. 349, n. 9064, p. 1498-1504, 1997.
- 22 SERRICCHIO, M., BUETIKOFER, P. Trypanosoma brucei: a model micro-organism to study eukaryotic phospholipid biosynthesis. **Febs Journal**, v. 278, n. 7, p. 1035-1046, 2011.
- 23 BERRIMAN, M. et al. The genome of the African trypanosome *Trypanosoma brucei*. **Science**, v. 309, n. 5733, p. 416-422, 2005.
- 24 SUBRAMANIAM, C. et al. Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. **Eukaryotic Cell**, v. 5, n. 9, p. 1539-1549, 2006.
- 25 SCHNEIDER, A. Unique aspects of mitochondrial biogenesis in trypanosomatids. **International Journal for Parasitology**, v. 31, n. 13, p. 1403-1415, 2001.
- 26 CENTERS FOR DISEASE CONTROL AND PREVENTION – CDCP 2015. Disponível em: <<http://www.cdc.gov/parasites/sleepingsickness.html>>. Acesso em: maio 2016.
- 27 MATTHEWS, K. R. The developmental cell biology of *Trypanosoma brucei*. **Journal of Cell Science**, v. 118, n. 9, p. 2078-2078, 2005.
- 28 LANGOUSIS, G., HILL, K. L. Motility and more: the flagellum of *Trypanosoma brucei*. **Nature Reviews Microbiology**, v. 12, n. 7, p. 505-518, 2014.
- 29 RODITI, I., LINIGER, M. Dressed for success: the surface coats of insect-borne protozoan parasites. **Trends in Microbiology**, v. 10, n. 3, p. 128-134, 2002.
- 30 PARSONS, M. Glycosomes: parasites and the divergence of peroxisomal purpose. **Molecular Microbiology**, v. 53, n. 3, p. 717-724, 2004.
- 31 URWYLER, S. et al. A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. **Molecular Microbiology**, v. 63, n. 1, p. 218-228, 2007.
- 32 DANIELS, J.-P., GULL, K., WICKSTEAD, B. Cell biology of the trypanosome genome. **Microbiology and Molecular Biology Reviews**, v. 74, n. 4, p. 552-569, 2010.
- 33 MAREE, J. P., PATTERTON, H.-G. The epigenome of *Trypanosoma brucei*: a regulatory interface to an unconventional transcriptional machine. **Biochimica et Biophysica Acta: gene regulatory mechanisms**, v. 1839, n. 9, p. 743-750, 2014.
- 34 DE MELO GODOY, P. D. et al. Trypanosome prereplication machinery contains a single functional orcl/cdc6 protein, which is typical of archaea. **Eukaryotic Cell**, v. 8, n. 10, p. 1592-1603, 2009.
- 35 BENMERZOUGA, I. et al. Trypanosoma brucei Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. **Molecular Microbiology**, v. 87, n. 1, p. 196-210, 2013.
- 36 CALDERANO, S. G. et al. Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics. **Nucleic Acids Research**, v. 43, n. 5, p. 2655-2665, 2015.



- 37 CLAYTON, C. E. Networks of gene expression regulation in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 195, n. 2, p. 96-106, 2014.
- 38 SIEGEL, T. N. et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. **Genes & Development**, v. 23, n. 9, p. 1063-1076, 2009.
- 39 DAS, A., BANDAY, M., BELLOFATTO, V. RNA polymerase transcription machinery in trypanosomes. **Eukaryotic Cell**, v. 7, n. 3, p. 429-434, 2008.
- 40 MATERA, A. G., WANG, Z. A day in the life of the spliceosome. **Nature Reviews Molecular Cell Biology**, v. 15, n. 2, p. 108-121, 2014.
- 41 SINGH, R., REDDY, R. Gamma-monomethyl phosphate - a cap structure in spliceosomal U6 small nuclear-Rna. **Proceedings of the National Academy of Sciences of the United States of America**, v. 86, n. 21, p. 8280-8283, 1989.
- 42 ULLU, E., MATTHEWS, K. R., TSCHUDI, C. Temporal-order of RNA-processing reactions in trypanosomes - rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. **Molecular and Cellular Biology**, v. 13, n. 1, p. 720-725, 1993.
- 43 PREUSSER, C., JAE, N., BINDEREIF, A. mRNA splicing in trypanosomes. **International Journal of Medical Microbiology**, v. 302, n. 4-5, p. 221-224, 2012.
- 44 MICHAELI, S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. **Future Microbiology**, v. 6, n. 4, p. 459-474, 2011.
- 45 LASDA, E. L., BLUMENTHAL, T. Trans-splicing. **Wiley Interdisciplinary Reviews-RNA**, v. 2, n. 3, p.417-434, 2011.
- 46 CROSS, M. et al. Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma-brucei* - structural organization and protein-components of the spliced leader rnp. **Molecular and Cellular Biology**, v. 11, n. 11, p. 5516-5526, 1991.
- 47 LEE, J. H. et al. Spliced leader RNA gene transcription in *Trypanosoma brucei* requires transcription factor TFIIH. **Eukaryotic Cell**, v. 6, n. 4, p. 641-649, 2007.
- 48 MICHAELI, S. Spliced leader RNA silencing (SLS) - a programmed cell death pathway in *Trypanosoma brucei* that is induced upon ER stress. **Parasites & Vectors**, n. 5, p. 107, 2012. doi: 10.1186/1756-3305-5-107.
- 49 WAHL, M. C., LUEHRMANN, R. SnapShot: spliceosome dynamics I. **Cell**, v. 161, n. 6, p. 1474-U267, 2015.
- 50 CHEN, W., MOORE, M. J. The spliceosome: disorder and dynamics defined. **Current Opinion in Structural Biology**, v. 24, p. 141-149, 2014. doi: 10.1016/j.sbi.2014.01.009.
- 51 GALEJ, W. P. et al. Structural studies of the spliceosome: zooming into the heart of the machine. **Current Opinion in Structural Biology**, v. 25, p. 57-66, 2014. doi: 10.1016/j.sbi.2013.12.002.
- 52 HANG, J. et al. Structural basis of pre-mRNA splicing. **Science**, v. 349, n. 6253, p. 1191-1198, 2015.
- 53 ASLETT, M. et al. *TriTrypDB*: a functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, p. D457-D462, 2010. doi: 10.1093/nar/gkp851.
- 54 YAN, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. **Science**, v. 349, n. 6253, p. 1182-1191, 2015.
- 55 KONDO, Y. et al. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. **eLife**, v. 4, e04986, 2015. doi: 10.7554/eLife.04986.

- 56 NGUYEN, T. H. D. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. **Nature**, v. 523, n. 7558, p. 47-52, 2015.
- 57 SANTOS, K. F. et al. Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 43, p. 17418-17423, 2012.
- 58 FOURMANN, J.-B. et al. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. **Genes & Development**, v. 27, n. 4, p. 413-428, 2013.
- 59 FICA, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. **Nature**, v. 503, n. 7475, p. 229-234, 2013.
- 60 AMBROSIO, D. L. et al. Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors. **Eukaryotic Cell**, v. 8, n. 7, p. 990-1000, 2009.
- 61 TKACZ, I. D. et al. Analysis of spliceosomal proteins in trypanosomatids reveals novel functions in mrna processing. **Journal of Biological Chemistry**, v. 285, n. 36, p. 27982-27999, 2010.
- 62 WEBER, G. et al. Functional organization of the Sm core in the crystal structure of human U1 snRNP. **Embo Journal**, v. 29, n. 24, p. 4172-4184, 2010.
- 63 SCHWER, B., CHANG, J., SHUMAN, S. Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5\**Mud2* branchpoint-binding complex and other spliceosome assembly factors. **Nucleic Acids Research**, v. 41, n. 15, p. 7485-7500, 2013.
- 64 PALFI, Z., LANE, W. S., BINDEREIF, A. Biochemical and functional characterization of the cis-spliceosomal U1 small nuclear RNP from *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 121, n. 2, p. 233-243, 2002.
- 65 PALFI, Z. et al. U1 small nuclear RNP from *Trypanosoma brucei*: a minimal U1 snRNA with unusual protein components. **Nucleic Acids Research**, v. 33, n. 8, p. 2493-2503, 2005.
- 66 PREUSSER, C. et al. Genome-wide RNA-binding analysis of the trypanosome U1 snRNP proteins U1C and U1-70K reveals cis/trans-spliceosomal network. **Nucleic Acids Research**, v. 42, n. 10, p. 6603-6615, 2014.
- 67 PALFI, Z. et al. Affinity purification of *Trypanosoma-brucei* small nuclear ribonucleoproteins reveals common and specific protein-components. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 20, p. 9097-9101, 1991.
- 68 PRICE, S. R., EVENS, P. R., NAGAI, K. Crystal structure of the spliceosomal U2B '-U2A ' protein complex bound to a fragment of U2 small nuclear RNA. **Nature**, v. 394, n. 6694, p. 645-650, 1998.
- 69 WILLIAMS, S. G., HALL, K. B. Binding affinity and cooperativity control U2B''snRNA/U2A' RNP formation. **Biochemistry**, v. 53, n. 23, p. 3727-3737, 2014.
- 70 LIN, P.-C., XU, R.-M. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. **Embo Journal**, v. 31, n. 6, p. 1579-1590, 2012.
- 71 WILL, C. L. et al. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. **Embo Journal**, v. 21, n. 18, p. 4978-4988, 2002.
- 72 SASHITAL, D. G. et al. Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. **RNA: a publication of the RNA society**, v. 13, n. 3, p. 328-338, 2007.
- 73 SCHWER, B., SHUMAN, S. Structure-function analysis and genetic interactions of the Yhc1, Smd3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of Smd3 with U2 snRNP subunit Lea1. **RNA: a publication of the RNA society**, v. 21, n. 6, p. 1173-1186, 2015.

- 74 ZHAO, C. et al. Conformational heterogeneity of the protein-free human spliceosomal U2-U6 snRNA complex. **RNA: a publication of the RNA society**, v. 19, n. 4, p. 561-573, 2013.
- 75 GUNZL, A. et al. Assembly of the U2 small nuclear ribonucleoprotein from *Trypanosoma-brucei* - a mutational analysis. **Journal of Biological Chemistry**, v. 268, n. 18, p. 13336-13343, 1993.
- 76 HARTSHORNE, T., AGABIAN, N. A new U2 Rna secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. **Genes & Development**, v. 4, n. 12A, p. 2121-2131, 1990.
- 77 MIRAGLIA, L. et al. Limited functional equivalence of phylogenetic variation in small nuclear-RNA - yeast U2 RNA with altered branchpoint complementarity inhibits splicing and produces a dominant lethal phenotype. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 16, p. 7061-7065, 1991.
- 78 CROSS, M. et al. The trans-spliceosomal U2 snRNP protein 40K of *Trypanosoma-brucei* - cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. **Embo Journal**, v. 12, n. 3, p. 1239-1248, 1993.
- 79 PREUSSER, C., PALFI, Z., BINDEREIF, A. Special Sm core complex functions in assembly of the U2 small nuclear ribonucleoprotein of *Trypanosoma brucei*. **Eukaryotic Cell**, v. 8, n. 8, p. 1228-1234, 2009.
- 80 MANFUL, T., CRISTODERO, M., CLAYTON, C. DRBD1 is the *Trypanosoma brucei* homologue of the spliceosome-associated protein 49. **Molecular and Biochemical Parasitology**, v. 166, n. 2, p. 186-189, 2009.
- 81 NYAMBEGA, B. et al. Proteins associated with SF3a60 in *T. brucei*. **Plos One**, v. 9, n. 3, e91956, 2014. doi: 10.1371/journal.pone.0091956.
- 82 WANG, P. et al. Sm core variation in spliceosomal small nuclear ribonucleoproteins from *Trypanosoma brucei*. **Embo Journal**, v. 25, n. 19, p.4513-4523, 2006.
- 83 TKACZ, I. D. et al. Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in *Trypanosoma brucei*. **RNA: a publication of the RNA society**, v. 13, n. 1, p. 30-43, 2007.
- 84 LIU, S. B. et al. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. **RNA: a publication of the RNA society**, v. 12, n. 7, p. 1418-1430, 2006.
- 85 NEWMAN, A. J. The role of U5 snRNP in pre-mRNA splicing. **Embo Journal**, v. 16, n. 19, p. 5797-5800, 1997.
- 86 TURNER, I. A. et al. Roles of the U5 snRNP in spliceosome dynamics and catalysis. **Biochemical Society Transactions**, v. 32, n. 6, p. 928-931, 2004.
- 87 LAGGERBAUER, B. et al. The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. **RNA: a publication of the RNA society**, v. 11, n. 5, p. 598-608, 2005.
- 88 NIELSEN, T. K. et al. Structural basis for the bifunctionality of the U5 snRNP 52K protein (CD2BP2). **Journal of Molecular Biology**, v. 369, n. 4, p. 902-908, 2007.
- 89 PRICE, A. M. et al. An unanticipated early function of DEAD-box ATPase Prp28 during commitment to splicing is modulated by U5 snRNP protein Prp8. **RNA: a publication of the RNA society**, v. 20, n. 1, p. 46-60, 2014.
- 90 ZHOU, L. et al. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. **Nature**, v. 506, n. 7486, p. 116-120, 2014.
- 91 LEUNG, A. K. W., NAGAI, K., LI, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. **Nature**, v. 473, n. 7348, p. 536-U285, 2011.

- 92 NOTTROTT, S., URLAUB, H., LUHRMANN, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. **Embo Journal**, v. 21, n. 20, p. 5527-5538, 2002.
- 93 LUETZELBERGER, M. et al. The N-terminus of Prp1 (Prp6/U5-102K) is essential for spliceosome activation in vivo. **Nucleic Acids Research**, v. 38, n. 5, p. 1610-1622, 2010.
- 94 NGUYEN, T. H. D. et al. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 angstrom resolution. **Nature**, v. 530, n. 7590, p. 298-302, 2016.
- 95 WAN, R. et al. The 3.8 angstrom structure of the U4/U6.U5 tri-snRNP: insights into spliceosome assembly and catalysis. **Science**, v. 351, n. 6272, p. 466-475, 2016.
- 96 GALEJ, W. P. et al. Crystal structure of Prp8 reveals active site cavity of the spliceosome. **Nature**, v. 493, n. 7434, p. 638-643, 2013.
- 97 AMBROSIO, D. L., SILVA, M. T. A., CICARELLI, R. M. B. Cloning and molecular characterization of *Trypanosoma cruzi* U2, U4, U5, and U6 small nuclear RNAs. **Memorias do Instituto Oswaldo Cruz**, v. 102, n. 1, p. 97-105, 2007.
- 98 JAE, N. et al. Essential role of a trypanosome U4-Specific Sm core protein in small nuclear ribonucleoprotein assembly and splicing. **Eukaryotic Cell**, v. 9, n. 3, 2010.
- 99 TKACZ, I. D. et al. Identification of the heptameric Lsm complex that binds U6 snRNA in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 160, n. 1, p. 22-31, 2008.
- 100 XU, Y. X., BENSHLOMO, H., MICHAELI, S. The U5 RNA of trypanosomes deviates from the canonical U5 RNA: The *Leptomonas collosoma* U5 RNA and its coding gene. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 16, p. 8473-8478, 1997.
- 101 LUCKE, S. et al. Trans mRNA splicing in trypanosomes: cloning and analysis of a PRP8-homologous gene from *Trypanosoma brucei* provides evidence for a U5-analogous RNP. **Embo Journal**, v. 16, n. 14, p. 4433-4440, 1997.
- 102 DA SILVA, M. T. A. et al. New insights into trypanosomatid U5 small nuclear ribonucleoproteins. **Memorias do Instituto Oswaldo Cruz**, v. 106, n. 2, p. 130-138, 2011.
- 103 LIMA, A. L. D. **Caracterização estrutural da proteína spliceossomal de *Trypanosoma brucei* U5-15K**. 2015. 99p. (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.
- 104 LIANG, X. H. et al. Analysis of spliceosomal complexes in *Trypanosoma brucei* and silencing of two splicing factors Prp31 and Prp43. **Molecular and Biochemical Parasitology**, v. 145, n. 1, p. 29-39, 2006.
- 105 MOTTRAM, J. et al. Isolation and sequence of 4 small nuclear U-RNA genes of *Trypanosoma-brucei* subsp *brucei* - identification of the U2-RNA, U4-RNA and U6-RNA analogs. **Molecular and Cellular Biology**, v. 9, n. 3, p. 1212-1223, 1989.
- 106 YU, Y. T., SHU, M. D., STEITZ, J. A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. **Embo Journal**, v. 17, n. 19, p. 5783-5795, 1998.
- 107 NEUENKIRCHEN, N. et al. Reconstitution of the human U snRNP assembly machinery reveals stepwise Sm protein organization. **EMBO Journal**, v. 34, n. 14, p.1925-1941, 2015.
- 108 RAKER, V. A. et al. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. **Molecular and Cellular Biology**, v. 19, n. 10, p. 6554-6565, 1999.
- 109 ZHANG, R. et al. Structure of a key intermediate of the SMN complex reveals gemin2's crucial function in snRNP assembly. **Cell**, v. 146, n. 3, p. 384-395, 2011.

- 110 CHARI, A. et al. An assembly chaperone collaborates with the SMN complex to generate spliceosomal snRNPs. **Cell**, v. 135, n. 3, p. 497-509, 2008.
- 111 FEENEY, R. J., ZIEVE, G. W. Nuclear-exchange of the U1 and U2 snRNP-specific proteins. **Journal of Cell Biology**, v. 110, n. 4, p. 871-881, 1990.
- 112 KAMBACH, C., MATTAJ, I. W. Intracellular-distribution of the U1a protein depends on active-transport and nuclear-binding to U1 snRNA. **Journal of Cell Biology**, v. 118, n. 1, p. 11-21, 1992.
- 113 KAMBACH, C., MATTAJ, I. W. Nuclear transport of the U2 snRNP-specific U2B" protein is mediated by both direct and indirect signaling mechanisms. **Journal of Cell Science**, v. 107, n. 7, p. 1807-1816, 1994.
- 114 ROMAC, J. M. J., GRAFF, D. H., KEENE, J. D. The U1 small nuclear ribonucleoprotein (snRNP) 70k protein is transported independently of U1 snRNP particles via a nuclear-localization signal in the RNA-binding domain. **Molecular and Cellular Biology**, v. 14, n. 7, p. 4662-4670, 1994.
- 115 HETZER, M., MATTAJ, I. W. An ATP-dependent, Ran-independent mechanism for nuclear import of the U1A and U2B " spliceosome proteins. **Journal of Cell Biology**, v. 148, n. 2, p. 293-303, 2000.
- 116 BIZARRO, J. et al. NUFIP and the HSP90/R2TP chaperone bind the SMN complex and facilitate assembly of U4-specific proteins. **Nucleic Acids Research**, v. 43, n. 18, p. 8973-89, 2015.
- 117 WEBER, G. et al. Structural basis for dual roles of Aar2p in U5 snRNP assembly. **Genes & Development**, v. 27, n. 5, p. 525-540, 2013.
- 118 NGUYEN, T. H. D. et al. Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. **Structure**, v. 21, n. 6, p. 910-919, 2013.
- 119 KROISS, M. et al. Evolution of an RNP assembly system: a minimal SMN-complex facilitates formation of U snRNPs in *Drosophila melanogaster*. **Faseb Journal**, v. 105, n. 29, p. 10045-10050, 2008.
- 120 PALFI, Z. et al. SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. **Genes & Development**, v. 23, n. 14, p. 1650-1664, 2009.
- 121 WAHL, M. C., WILL, C. L., LUEHRMANN, R. The spliceosome: design principles of a dynamic RNP machine. **Cell**, v. 136, n. 4, p. 701-718, 2009.
- 122 CHANARAT, S., STRAESSER, K. Splicing and beyond: the many faces of the Prp19 complex. **Biochimica et Biophysica Acta: molecular cell research**, v. 1833, n. 10, p. 2126-2134, 2013.
- 123 CHAN, S. P., CHENG, S. C. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. **Journal of Biological Chemistry**, v. 280, n. 35, p. 31190-31199, 2005.
- 124 AMBROSIO, D. L., BADJATIA, N., GUENZL, A. The spliceosomal PRP19 complex of trypanosomes. **Molecular Microbiology**, v. 95, n. 5, p. 885-901, 2015.
- 125 ULRICH, A., WAHL, M. C. Structure and evolution of the spliceosomal peptidyl-prolyl cis-trans isomerase Cwc27. **Acta Crystallographica Section D: biological crystallography**, v. 70, p. 3110-3123, 2014.
- 126 SONG, E. J. et al. The Prp19 complex and the Usp4(Sart3) deubiquitinating enzyme control reversible ubiquitination at the spliceosome. **Genes & Development**, v. 24, n. 13, p. 1434-1447, 2010.
- 127 BELLARE, P. et al. Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. **RNA: a publication of the RNA society**, v. 12, n. 2, p. 292-302, 2006.
- 128 BELLARE, P. et al. A role for ubiquitin in the spliceosome assembly pathway. **Nature Structural & Molecular Biology**, v. 15, n. 5, p. 444-451, 2008.

- 129 STARK, M. R. et al. Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 11, p. E1191-E1200, 2015.
- 130 BITAR, M. et al. The spliced leader trans-splicing mechanism in different organisms: molecular details and possible biological roles. **Frontiers in Genetics**, v. 4, n. 199, p. 199, 2013. doi: 10.3389/fgene.2013.00199.
- 131 HUDSON, A. J. et al. Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms. **RNA Biology**, v. 12, n. 11, p. 1-8, 2015.
- 132 LIU, Y.-C., CHENG, S.-C. Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. **Journal of Biomedical Science**, v. 22, n. 54, 2015. doi: 10.1186/s12929-015-0161-z.
- 133 OZGUR, S. et al. The conformational plasticity of eukaryotic RNA-dependent ATPases. **FEBS Journal**, v. 282, n. 5, p. 850-863, 2015.
- 134 SEMLOW, D. R., STALEY, J. P. Staying on message: ensuring fidelity in pre-mRNA splicing. **Trends in Biochemical Sciences**, v. 37, n. 7, p. 263-273, 2012.
- 135 KORNBLIHTT, A. R. et al. Alternative splicing: a pivotal step between eukaryotic transcription and translation. **Nature Reviews Molecular Cell Biology**, v. 14, n. 3, p. 153-165, 2013.
- 136 ZHU, J., MAYEDA, A., KRAINER, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. **Molecular Cell**, v. 8, n. 6, p. 1351-1361, 2001.
- 137 FU, X.-D., ARES JR, M. Context-dependent control of alternative splicing by RNA-binding proteins. **Nature Reviews Genetics**, v. 15, n. 10, p. 689-701, 2014.
- 138 NILSSON, D. et al. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. **Plos Pathogens**, v. 6, n. 8, e1001037, 2010. doi: 10.1371/journal.ppat.1001037.
- 139 HULL, R., DLAMINI, Z. The role played by alternative splicing in antigenic variability in human endoparasites. **Parasites & Vectors**, v. 7, p. 53-53, 2014. doi: 10.1186/1756-3305-7-53.
- 140 TURUNEN, J. J. et al. The significant other: splicing by the minor spliceosome. **Wiley Interdisciplinary Reviews-RNA**, v. 4, n. 1, p. 61-76, 2013.
- 141 WANG, G.-S., COOPER, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. **Nature Reviews Genetics**, v. 8, n. 10, p. 749-761, 2007.
- 142 DAIGER, S. P., BOWNE, S. J., SULLIVAN, L. S. Perspective on genes and mutations causing retinitis pigmentosa. **Archives of Ophthalmology**, v. 125, n. 2, p. 151-158, 2007.
- 143 WIECZOREK, D. et al. Compound heterozygosity of low-frequency promoter deletions and rare loss-of-function mutations in TXNL4A causes Burn-McKeown syndrome. **American Journal of Human Genetics**, v. 95, n. 6, p. 698-707, 2014.
- 144 BAIONI, M. T. C., AMBIEL, C. R. Spinal muscular atrophy: diagnosis, treatment and future prospects. **Jornal de Pediatria**, v. 86, n. 4, p. 261-270, 2010.
- 145 LI, D. K. et al. SMN control of RNP assembly: From post-transcriptional gene regulation to motor neuron disease. **Seminars in Cell & Developmental Biology**, v. 32, p. 22-29, 2014. doi: 10.1016/j.semcdb.2014.04.026.
- 146 FACKENTHAL, J. D., GODLEY, L. A. Aberrant RNA splicing and its functional consequences in cancer cells. **Disease Models & Mechanisms**, v. 1, n. 1, p. 37-42, 2008.
- 147 SILIPO, M., GAUTREY, H., TYSON-CAPPER, A. Deregulation of splicing factors and breast cancer development. **Journal of Molecular Cell Biology**, v. 7, n. 5, p. 388-401, 2015.

- 148 JAFARIFAR, F. et al. Biochemical defects in minor spliceosome function in the developmental disorder MOPD I. **RNA: a publication of the RNA society**, v. 20, n. 7, p. 1078-1089, 2014.
- 149 WAHL, M. C., LUHRMANN, R. SnapShot: spliceosome dynamics III. **Cell**, v. 162, n. 3, p. 690-690, 2015.
- 150 BONNAL, S., VIGEVANI, L., VALCARCEL, J. The spliceosome as a target of novel antitumour drugs. **Nature Reviews Drug Discovery**, v. 11, n. 11, p. 847-859, 2012.
- 151 LAGISETTI, C. et al. Optimization of antitumor modulators of pre-mRNA splicing. **Journal of Medicinal Chemistry**, v. 56, n. 24, p. 10033-10044, 2013.
- 152 HSU, T. Y. T. et al. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. **Nature**, v. 525, n. 7569, p. 384+, 2015.
- 153 KOLE, R., LEPPERT, B. J. Targeting mRNA splicing as a potential treatment for duchenne muscular dystrophy. **Discovery Medicine**, v. 14, n. 74, p. 59-69, 2012.
- 154 HAVENS, M. A., DUELLI, D. M., HASTINGS, M. L. Targeting RNA splicing for disease therapy. **Wiley Interdisciplinary Reviews-RNA**, v. 4, n. 3, p. 247-266, 2013.
- 155 KRAUSE, M., HIRSH, D. A trans-spliced leader sequence on actin messenger-RNA in *C-elegans*. **Cell**, v. 49, n. 6, p. 753-761, 1987.
- 156 RAJKOVIC, A. et al. Asplliced leader is present on a subset of messenger-RNAs from the human parasite *Schistosoma-mansoni*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 87, n. 22, p. 8879-8883, 1990.
- 157 DOURIS, V., TELFORD, M. J., AVEROF, M. Evidence for multiple independent origins of trans-splicing in metazoa. **Molecular Biology and Evolution**, v. 27, n. 3, p. 684-693, 2010.
- 158 GONZALEZ-ANDRADE, P. et al. Diagnosis of trypanosomatid infections targeting the spliced leader RNA. **Journal of Molecular Diagnostics**, v. 16, n. 4, p. 400-404, 2014.
- 159 CHAPPUIS, F. et al. Options for field diagnosis of human African trypanosomiasis. **Clinical Microbiology Reviews**, v. 18, n. 1, p. 133-146, 2005.
- 160 TAN, E. M., KUNKEL, H. G. Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. **Journal of Immunology**, v. 176, n. 3, p. 1297-1304, 2006.
- 161 LERNER, M. R. et al. 2 novel classes of small ribonucleoproteins detected by antibodies associated with lupus-erythematosus. **Science**, v. 211, n. 4480, p. 400-402, 1981.
- 162 ACHSEL, T. et al. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. **EMBO Journal**, v. 18, n. 20, p. 5789-5802, 1999.
- 163 TANG, W. et al. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. **Nature**, v. 484, n. 7393, p. 260-U152, 2012.
- 164 PELZ, J.-P. et al. Crystallizing the 6S and 8S spliceosomal assembly intermediates: a complex project. **Acta Crystallographica Section D: biological crystallography**, v. 71, n. 10, p. 2040-2053, 2015.
- 165 SENG, C. O. et al. The SMN structure reveals its crucial role in snRNP assembly. **Human Molecular Genetics**, v. 24, n. 8, p. 2138-2146, 2015.
- 166 MURA, C. et al. Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. **RNA Biology**, v. 10, n. 4, p. 636-651, 2013.
- 167 TOMASEVIC, N., PECULIS, B. A. Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. **Molecular and Cellular Biology**, v. 22, n. 12, p. 4101-4112, 2002.

- 168 KUFEL, J. et al. Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. **Molecular and Cellular Biology**, v. 22, n. 14, p. 5248-5256, 2002.
- 169 SCHUMPERLI, D., PILLAI, R. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. **Cellular and Molecular Life Sciences**, v. 61, n. 19-20, p. 2560-2570, 2004.
- 170 THARUN, S. et al. Yeast Sm-like proteins function in mRNA decapping and decay. **Nature**, v. 404, n. 6777, p. 515-518, 2000.
- 171 SETO, A. G. et al. *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. **Nature**, v. 401, n. 6749, p. 177-180, 1999.
- 172 MURINA, V. N. et al. Supramolecular organization of Hfq-like proteins. **Biochemistry**, v. 80, n. 4, p. 441-448, 2015.
- 173 VERETNIK, S. et al. Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. **Plos Computational Biology**, v. 5, n. 3, e1000315, 2009. doi: 10.1371/journal.pcbi.1000315.
- 174 KAMBACH, C. et al. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. **Cell**, v. 96, n. 3, p. 375-387, 1999.
- 175 STARK, H., LUEHRMANN, R. Cryo-electron microscopy of spliceosomal components. **Annual Review of Biophysics and Biomolecular Structure**, v. 35, p. 435-457, 2006.
- 176 PLESSEL, G., LUHRMANN, R., KASTNER, B. Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (EFG) protein heteromer and the intact snRNP core. **Journal of Molecular Biology**, v. 265, n. 2, p. 87-94, 1997.
- 177 WALKE, S. et al. Stoichiometry of the Sm proteins in yeast spliceosomal snRNPs supports the heptamer ring model of the core domain. **Journal of Molecular Biology**, v. 308, n. 1, p. 49-58, 2001.
- 178 WU, D. et al. Crystal structures of Lsm3, Lsm4 and Lsm5/6/7 from *Schizosaccharomyces pombe*. **Plos One**, v. 7, n. 5, e36768, 2012. doi: 10.1371/journal.pone.0036768
- 179 LINK, T. M., VALENTIN-HANSEN, P., BRENNAN, R. G. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 46, p. 19286-19291, 2009.
- 180 MURA, C. et al. The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). **Protein Science**, v. 12, n. 4, p. 832-847, 2003.
- 181 MAERTENS, B. et al. The heptameric SmAP1 and SmAP2 Proteins of the *Crenarchaeon sulfobolus solfataricus* bind to common and distinct RNA Targets. **Life-Basel**, v. 5, n. 2, p. 1264-1281, 2015.
- 182 NAIDOO, N. et al. Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: Implications for Lsm ring organisation and recruitment. **Journal of Molecular Biology**, v. 377, n. 5, p. 1357-1371, 2008.
- 183 PALFI, Z. et al. The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 16, p. 8967-8972, 2000.
- 184 MANDELBOIM, M. et al. Silencing of Sm proteins in *Trypanosoma brucei* by RNA interference captured a novel cytoplasmic intermediate in spliced leader RNA biogenesis. **Journal of Biological Chemistry**, v. 278, n. 51, p. 51469-51478, 2003.
- 185 TKACZ, I. D. et al. Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in *Trypanosoma brucei*. **RNA: a publication of the RNA society**, v. 13, n. 1, p. 30-43, 2007.
- 186 JAE, N. et al. snRNA-specific role of SMN in trypanosome snRNP biogenesis in vivo. **Rna Biology**, v. 8, n. 1, p. 90-100, 2011.



- 187 BELLA, J. et al. The leucine-rich repeat structure. **Cellular and Molecular Life Sciences**, v. 65, n. 15, p. 2307-2333, 2008.
- 188 KIELKOPF, C. L., LUCKE, S., GREEN, A. R. U2AF homology motifs: protein recognition in the RRM world. **Genes & Development**, v. 18, n. 13, p. 1513-1526, 2004.
- 189 OUBRIDGE, C. et al. Crystal-structure at 1.92 angstrom resolution of the rna-binding domain of the U1a spliceosomal protein complexed with an RNA hairpin. **Nature**, v. 372, n. 6505, p. 432-438, 1994.
- 190 CASPARY, F., SERAPHIN, B. The yeast U2A 'U2B " complex is required for pre-spliceosome formation. **Embo Journal**, v. 17, n. 21, p. 6348-6358, 1998.
- 191 TANG, J., ABOVICH, N., ROSBASH, M. Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B" protein. **Molecular and Cellular Biology**, v. 16, n. 6, p. 2787-2795, 1996.
- 192 GASTEIGER, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3784-3788, 2003.
- 193 SCHULTZ, J. et al. SMART: a web-based tool for the study of genetically mobile domains. **Nucleic Acids Research**, v. 28, n. 1, p. 231-234, 2000.
- 194 SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 1, p. 539, 2011. doi: 10.1038/msb.2011.75.
- 195 BUCHAN, D. W. A. et al. Scalable web services for the PSIPRED protein analysis workbench. **Nucleic Acids Research**, v. 41, n. W1, p. W349-W357, 2013.
- 196 HANAHAN, D., JESSEE, J., BLOOM, F. R. Plasmid transformation of *Escherichia-coli* and other bacteria. **Methods in Enzymology**, v. 204, p. 63-113, 1991. doi:10.1016/0076-6879(91)04006-A.
- 197 BRADFORD, M. M. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. **Analytical Biochemistry**, v. 72, n. 1-2, p. 248-254, 1976.
- 198 MIKULECKY, P. J., FEIG, A. L. Heat capacity changes in RNA folding: application of perturbation theory to hammerhead ribozyme cold denaturation. **Nucleic Acids Research**, v. 32, n. 13, p. 3967-3976, 2004.
- 199 JOLLIFFE, I. T. **Principal component analysis**. 2<sup>nd</sup> ed. New York: Springer, 2002.
- 200 WEISS, J. N. The Hill equation revisited: uses and misuses. **Faseb Journal**, v. 11, n. 11, p. 835-841, 1997.
- 201 OCHI, T. et al. Perspectives on protein crystallisation. **Progress in Biophysics & Molecular Biology**, v. 101, n. 1-3, p. 56-63, 2009.
- 202 JEN, A., MERKLE, H. P. Diamonds in the rough: protein crystals from a formulation perspective. **Pharmaceutical Research**, v. 18, n. 11, p. 1483-1488, 2001.
- 203 RUPP, B. **Biomolecular crystallography**. Abingdon, New York: Garland Science, Taylor & Francis Group, 2010. 809 p. ISBN-10 0815340818.
- 204 D'ARCY, A. et al. Microseed matrix screening for optimization in protein crystallization: what have we learned? **Acta Crystallographica Section F: structural biology communications**, v. 70, p. 1117-1126, 2014.
- 205 BERGFORS, T. Seeds to crystals. **Journal of Structural Biology**, v. 142, n. 1, p. 66-76, 2003.
- 206 DAUTER, Z. Data-collection strategies. **Acta Crystallographica Section D: biological crystallography**, v. 55, n. 10, p. 1703-1717, 1999.

- 207 BATTYE, T. G. G. et al. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. **Acta Crystallographica Section D: biological crystallography**, v. 67, n. 4, p. 271-281, 2011.
- 208 EVANS, P. R., MURSHUDOV, G. N. How good are my data and what is the resolution? **Acta Crystallographica Section D: biological crystallography**, v. 69, n. 7, p. 1204-1214, 2013.
- 209 EVANS, P. Scaling and assessment of data quality. **Acta Crystallographica Section D: biological crystallography**, v. 62, n. 1, p. 72-82, 2006.
- 210 MATTHEWS, B. W. Solvent content of protein crystals. **Journal of Molecular Biology**, v. 33, n. 2, p. 491-497, 1968.
- 211 TAYLOR, G. The phase problem. **Acta Crystallographica Section D: biological crystallography**, v. 59, p. 1881-1890, 2003.
- 212 MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658-674, 2007.
- 213 READ, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. **Acta Crystallographica Section D: biological crystallography**, v. 57, n. 10, p. 1373-1382, 2001.
- 214 STEIN, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. **Journal of Applied Crystallography**, v. 41, p. 641-643, 2008. doi:10.1107/S0021889808006985.
- 215 LANGER, G. G. et al. Visual automated macromolecular model building. **Acta Crystallographica Section D: biological crystallography**, v. 69, n. 4, p. 635-641, 2013.
- 216 MURSHUDOV, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. **Acta Crystallographica Section D: biological crystallography**, v. 67, n. 4, p. 355-367, 2011.
- 217 AFONINE, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. **Acta Crystallographica Section D: biological crystallography**, v. 68, n. 4, p. 352-367, 2012.
- 218 MURSHUDOV, G. N., VAGIN, A. A., DODSON, E. J. Refinement of macromolecular structures by the maximum-likelihood method. **Acta Crystallographica Section D: biological crystallography**, v. 53, n. 3, p. 240-255, 1997.
- 219 EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 4, p. 486-501, 2010.
- 220 PETTERSEN, E. F. et al. UCSF chimera - a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605-1612, 2004.
- 221 URZHUMTSEVA, L. et al. Crystallographic model quality at a glance. **Acta Crystallographica Section D: biological crystallography**, v. 65, n. 3, p. 297-300, 2009.
- 222 CHEN, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 1, p. 12-21, 2010.
- 223 KRISSINEL, E., HENRICK, K. Inference of macromolecular assemblies from crystalline state. **Journal of Molecular Biology**, v. 372, n. 3, p. 774-797, 2007.
- 224 BAKER, N. A. et al. Electrostatics of nanosystems: application to microtubules and the ribosome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 18, p. 10037-10041, 2001.
- 225 CHOTHIA, C., LESK, A. M. The relation between the divergence of sequence and structure in proteins. **Embo Journal**, v. 5, n. 4, p. 823-826, 1986.

- 226 KACZANOWSKI, S., ZIELENKIEWICZ, P. Why similar protein sequences encode similar three-dimensional structures? **Theoretical Chemistry Accounts**, v. 125, n. 3-6, p. 643-650, 2010.
- 227 BAZZOLI, A., TETTAMANZI, A. G. B., ZHANG, Y. Computational protein design and large-scale assessment by I-TASSER structure assembly simulations. **Journal of Molecular Biology**, v. 407, n. 5, p. 764-776, 2011.
- 228 MARTI-RENO, M. A. et al, A. Comparative protein structure modeling of genes and genomes. **Annual Reviews on Biophysics and Biomolecular Structure**, v. 29, p. 291-325, 2000. doi: 10.1146/annurev.biophys.29.1.291.
- 229 LASKOWSKI, R. A. PDBsum: summaries and analyses of PDB structures. **Nucleic Acids Research**, v. 29, n. 1, p. 221-222, 2001.
- 230 WIEDERSTEIN, M., SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, n. 2, p. W407-W410, 2007.
- 231 ORTEGA, A., AMOROS, D., GARCIA DE LA TORRE, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. **Biophysical Journal**, v. 101, n. 4, p. 892-898, 2011.
- 232 HAMMERSLEY, A. P. Fit2D: an introduction and overview. **European Synchrotron Radiation Facility Internal Report ESRF97HA02T**, v. 68, 1997.
- 233 ORTHABER, D., BERGMANN, A., GLATTER, O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. **Journal of Applied Crystallography**, v. 33, p. 218-225, 2000.
- 234 GUINIER, A., FOURNET, G. **Small-angle scattering of X-rays**. New York: John Wiley & Sons, 1955.
- 235 SVERGUN, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal of Applied Crystallography**, v. 25, p. 495-503, 1992.
- 236 SVERGUN, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophysical Journal**, v. 76, n. 6, p. 2879-2886, 1999.
- 237 VOLKOV, V. V., SVERGUN, D. I. Uniqueness of ab initio shape determination in small-angle scattering. **Journal of Applied Crystallography**, v. 36, n. 3, p. 860-864, 2003.
- 238 KOZIN, M. B., SVERGUN, D. I. Automated matching of high- and low-resolution structural models. **Journal of Applied Crystallography**, v. 34, n. 1, p. 33-41, 2001.
- 239 SVERGUN, D., BARBERATO, C., KOCH, M. H. J. CRY SOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. **Journal of Applied Crystallography**, v. 28, n. 6, p. 768-773, 1995.
- 240 PETOUKHOV, M. V., SVERGUN, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. **Biophysical Journal**, v. 89, n. 2, p. 1237-1250, 2005.
- 241 SVERGUN, D. I. Small-angle X-ray and neutron scattering as a tool for structural systems biology. **Biological Chemistry**, v. 391, n. 7, p. 737-743, 2010.
- 242 FISCHER, H. et al. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. **Journal of Applied Crystallography**, v. 43, n. 1, p. 101-109, 2010.
- 243 BRAHMS, H. et al. The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. **Journal of Biological Chemistry**, v. 275, n. 22, p. 17122-17129, 2000.

- 244 BRAHMS, H. et al. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. **RNA**, v. 7, n. 11, p. 1531-1542, 2001.
- 245 WOODY, R. W. Circular-dichroism. **Biochemical Spectroscopy**, v. 246, p. 34-71, 1995.
- 246 HARDIN, J. W. et al. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. **Nucleic Acids Research**, v. 43, n. 22, p. 10963-10974, 2015.
- 247 LANGER, G. et al. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. **Nature Protocols**, v. 3, n. 7, p. 1171-1179, 2008.
- 248 SANCHEZ, R., SALI, A. Comparative protein structure modeling : introduction and practical examples with modeller. **Protein Structure Prediction: methods and protocols**, v. 143, p. 97-129, 2000.
- 249 PALFI, Z., BINDEREIF, A. Immunological characterization and intracellular-localization of transspliceosomal small nuclear ribonucleoproteins in *Trypanosoma-brucei*. **Journal of Biological Chemistry**, v. 267, n. 28, p. 20159-20163, 1992.
- 250 AGAFONOV, D. E. et al. Molecular architecture of the human U4/U6.U5 tri-snRNP. **Science**, v. 351, n. 6280, p. 1416-1420, 2016.
- 251 ZHANG, L. et al. Brr2 plays a role in spliceosomal activation in addition to U4/U6 unwinding. **Nucleic Acids Research**, v. 43, n. 6, p. 3286-3297, 2015.
- 252 HUANG, Y.-H. et al. Sad1 counteracts Brr2-mediated dissociation of U4/U6. U5 in tri-snRNP homeostasis. **Molecular and Cellular Biology**, v. 34, n. 2, p. 210-220, 2014.
- 253 ABSMEIER, E. et al. The large N-terminal region of the Brr2 RNA helicase guides productive spliceosome activation. **Genes & Development**, v. 29, n. 24, p. 2576-2587, 2015.
- 254 ABSMEIER, E. et al. A noncanonical PWI domain in the N-terminal helicase-associated region of the spliceosomal Brr2 protein. **Acta Crystallographica Section D: structural biology**, v. 71, n. 4, p. 762-771, 2015.
- 255 PENA, V. et al. Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. **Molecular Cell**, v. 35, n. 4, p. 454-466, 2009.
- 256 KOMETA, I., MAGNUS, M., BUJNICKI, J. M. Structural bioinformatics of the human spliceosomal proteome. **Nucleic Acid Research**, v. 40, n. 15, p. 7046-7065, 2012.
- 257 BRUN, R., SCHONENBERGER, M. Cultivation and invitro cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. **Acta Tropica**, v. 36, n. 3, p. 289-292, 1979.
- 258 STROBER, W. Appendix 3A monitoring cell growth. **Current Protocols in Immunology**. v. 21, 2001. doi: 10.1002/0471142735.ima03as21.
- 259 BLAIR, R. C., HIGGINS, J. J., SMITLEY, W. D. S. On the relative power of the u and t tests. **British Journal of Mathematical & Statistical Psychology**, v. 33, p. 114-120, 1980. doi: 10.1111/j.2044-8317.1980.tb00783.x.
- 260 COX, J. et al. Andromeda: a peptide search engine integrated into the *MaxQuant* environment. **Journal of Proteome Research**, v. 10, n. 4, p. 1794-1805, 2011.
- 261 BUTTNER, K., NEHRING, S., HOPFNER, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. **Nature Structural & Molecular Biology**, v. 14, n. 7, p. 647-652, 2007.
- 262 LINDER, P., KOWALCZYKOWSKI, S. Helicases and NTP-Driven nucleic acid machines: structure, function and roles in human disease. **Nucleic Acids Research**, v. 34, n. 15, p. 4081-4081, 2006.

- 263 GRAINGER, R. J. et al. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. **RNA: a publication of the RNA society**, v. 15, n. 12, p. 2161-2173, 2009.
- 264 GAUTAM, A. et al. Cwc21p promotes the second step conformation of the spliceosome and modulates 3' splice site selection. **Nucleic Acids Research**, v. 43, n. 6, p. 3309-3317, 2015.
- 265 BRINGAUD, F. et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. **Eukaryotic Cell**, v. 1, n. 2, p. 315-315, 2002.
- 266 XU, X. L. et al. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. **Protein Expression and Purification**, v. 72, n. 2, p. 149-156, 2010.
- 267 STRASSER, K., HURT, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. **Nature**, v. 413, n. 6856, p. 648-652, 2001.
- 268 SERPELONI, M. et al. An essential nuclear protein in trypanosomes is a component of mRNA transcription/export pathway. **Plos One**, v. 6, n. 6, e20730, 2011. doi: 10.1371/journal.pone.0020730.
- 269 XARGAY-TORRENT, S. et al. The splicing modulator sudemycin induces a specific antitumor response and cooperates with ibrutinib in chronic lymphocytic leukemia. **Oncotarget**, v. 6, n. 26, p. 22734-22749, 2015
- 270 BONNAL, S., VIGEVANI, L., VALCÁRCEL, J. The spliceosome as a target of novel antitumour drugs. **Nature Reviews Drug Discovery**, v. 11, n. 11, p.847-859, 2012
- 271 EFFENBERGER, K. A. et al. The natural product N-palmitoyl-l-leucine selectively inhibits late assembly of human spliceosomes. **Journal of Biological Chemistry**, v. 290, n. 46, p. 27524-27531, 2015.
- 272 LAGISETTI, C. et al. Optimization of antitumor modulators of pre-mRNA splicing. **Journal of Medicinal Chemistry**, v. 56, n. 24, p. 10033–10044, 2013.
- 273 PAWELLEK, A. et al. Identification of small molecule inhibitors of pre-mRNA splicing. **Journal of Biological Chemistry**, v. 289, n. 50, p. 34683-34698, 2014.
- 274 NARYSHKIN, A. et al. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. **Science**, v. 345, n. 6197, p. 688-693, 2014.



## APÊNDICE A1

Tabela A1.1 – Pequenas moléculas modulares do spliceossomo.

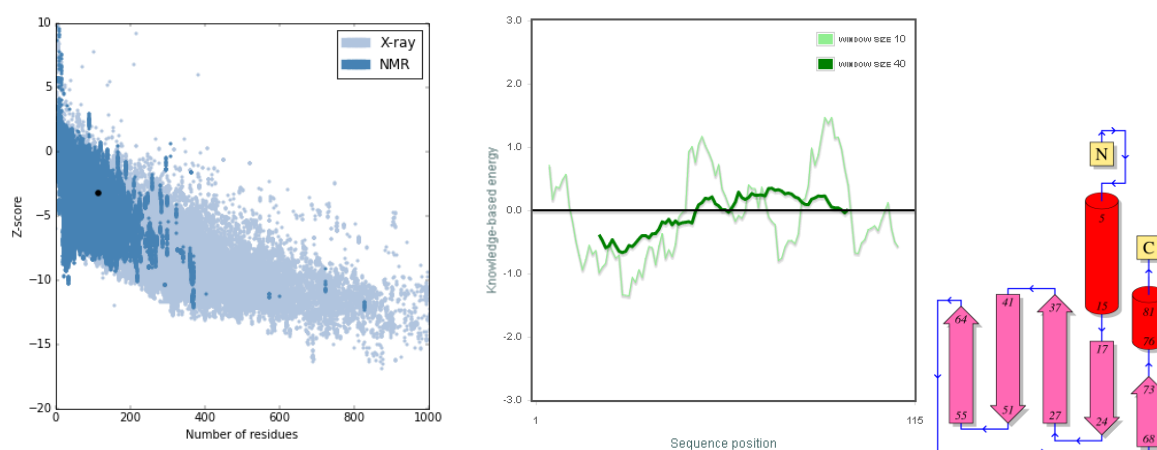
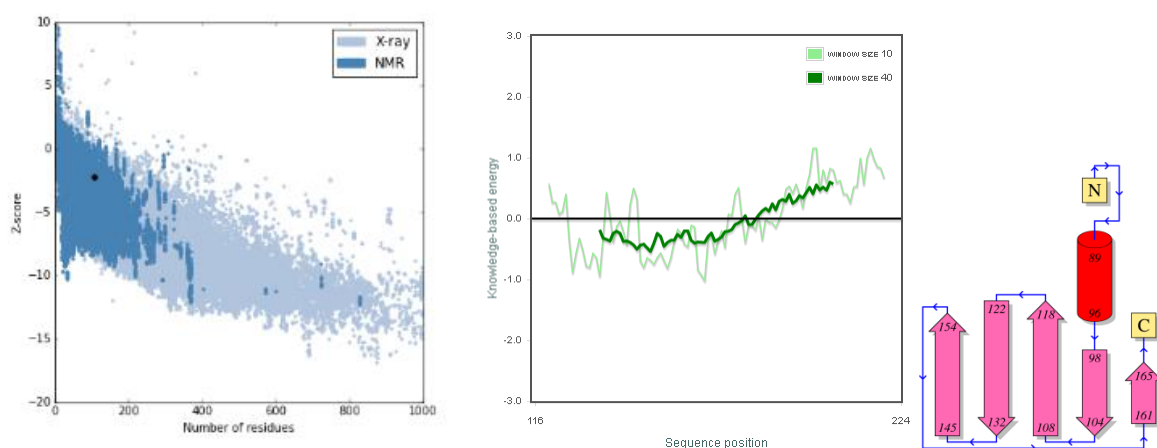
Molécula	Origem	Atuação
FR901464 Pladienolídeos Herboxidieno Tailanestatina FD-895	Metabólitos bacterianos.	- Têm como alvo a subunidade SF3b de U2 snRNP, atuando no estágio inicial de montagem do spliceossomo. - Apresentam ação anti-tumoral.
E7107 Spliceostatina Meayamicinas Sudemicina D6	Análogos sintéticos de metabólito bacteriano.	- Têm como alvo a subunidade SF3b de U2 snRNP, atuando no estágio inicial de montagem do spliceossomo. - Apresentam ação anti-tumoral.
N-palmitoil-L-leucina	Produto natural de origem vegetal.	- Inibe <i>splicing in vitro</i> , bloqueando a formação do complexo do spliceossomo pós-catálfico.
Isoginkgetina (biflavonoide)	Produto natural de origem vegetal.	- Inibe <i>splicing in vitro</i> e em cultura de células a partir da inibição da transição do complexo A para o complexo B do spliceossomo. - É conhecido como antagonista de invasão tumoral a partir da diminuição da expressão da metaloproteinase 9, porém não se conhece a relação deste mecanismo com a inibição do <i>splicing</i> .
Esteroides cardiotônicos		- Usados em tratamento clínico de doenças cardíacas e têm como alvo vias de sinalização celular. - Modulam o <i>splicing</i> alternativo de vários pré-mRNAs em cultura de células, em partes por levar ao controle negativo da expressão dos fatores reguladores de <i>splicing</i> : SRSF3 e TRA2B.
Derivados de indol		- Interagem diretamente com proteínas SR e inibem <i>splicing</i> de íntrons que são dependentes de sinais de sequências potencializadoras de <i>splicing</i> .
Ácido okadaico, tautomicina, microcistina- LR, vanadato de sódio e pseudocantaridinas.		- Inibidores de fosfatases PP1 e PP2A. - Inibem <i>splicing</i> e modulam seleção de sítios no <i>splicing</i> alternativo, por exemplo, do gene SMN2 relacionado à distrofia muscular espinhal.
NB-506 (indolcarbazol glicosilado)		- Inibe a atividade de fosforilação de proteínas SR por meio de topoisomerase I. Deste modo, acaba inibindo <i>splicing</i> ou afetando a regulação de <i>splicing</i> alternativo.
TG003 (benzotiazol)		- Inibe as cinases CLK1 e CLK4 que fosforilam proteínas SR. Assim, levam à inibição de <i>splicing</i> e alteração em <i>splicing</i> alternativo.
SRPIN340 (nicotinamida)		- Inibe as cinases SRPK1 e SRPK2 que fosforilam proteínas SR.
Amilorida		- Fármaco anti-hipertensivo que interfere na homeostase de sódio. - Afeta <i>splicing</i> de genes relevantes a câncer e diminui a fosforilação de SRSF1.
Clotrimazol		- Fármaco antifúngico que interfere na homeostase de cálcio. - Apresenta efeitos no <i>splicing</i> alternativo.
Flunarizina		- Bloqueador de canal de cálcio utilizado no tratamento de doenças neurológicas. - Apresenta efeitos no <i>splicing</i> alternativo.
Clorohexidrina		- Agente anti-infectivo. - Apresenta efeitos no <i>splicing</i> alternativo devido à inibição de proteínas cinases da família CLK.
1,4-naphtoquinonas e 1,4- quinonas heterocíclicas.		- Apresentam atividade anti-câncer. - Atuam na prevenção da formação do complexo spliceossomal anterior à segunda etapa de <i>splicing</i> .
Madrasina		- Retém o spliceossomo no complexo A, inibindo <i>splicing in vitro</i> e <i>in vivo</i> . - Adicionalmente, apresenta atividade citostática.
SMN-C1, SMN-C2, e SMN-C3		- Alteram <i>splicing</i> do transcrito do gene SMN2, ligado à SMA. - Possuem biodisponibilidade oral.

Fonte: Adaptada de XARGAY-TORRENT et al<sup>269</sup>; BONNAL et al<sup>270</sup>; EFFENBERGER et al<sup>271</sup>; LAGISETTI et al<sup>272</sup>; PAWELLEK et al<sup>273</sup>; NARYSHKIN et al<sup>274</sup>.



## APÊNDICE A2

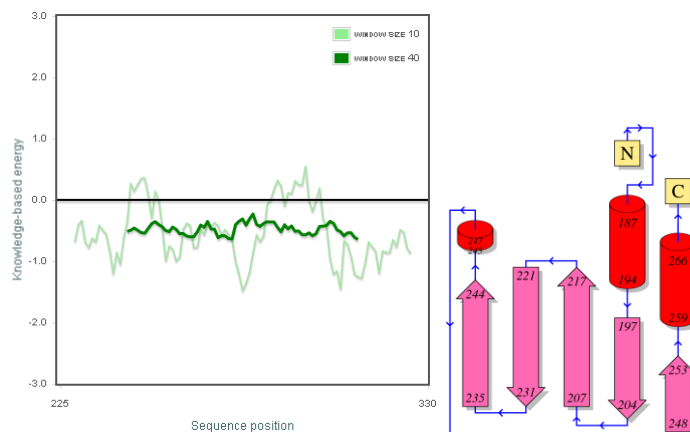
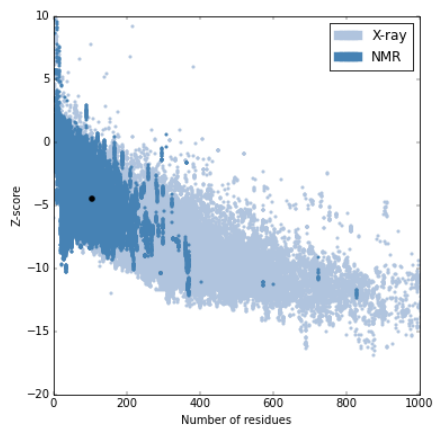
Os modelos de estrutura terciária de cada cadeia polipeptídica dos complexos Sm 1, 2 e 3 obtidos por modelagem comparativa, como descrito no Capítulo 2 (seção 2.4.7), foram avaliados utilizando os servidores *ProSa*<sup>230</sup> e *PDBSum*<sup>229</sup>, como se observa na Figura A2.1. São mostrados o *Z-score*, o gráfico de energia por resíduo e a topologia da estrutura.

Complexo 1 – cadeia A - Z = -3,25Complexo 1 – cadeia B - Z = -2,21

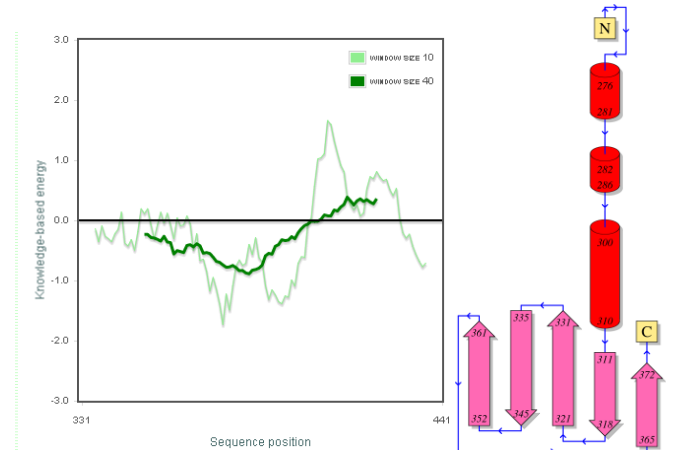
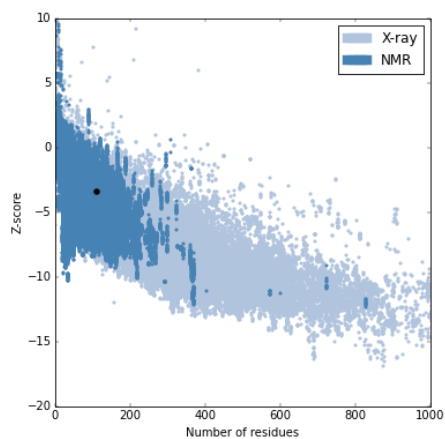
continua

continuação

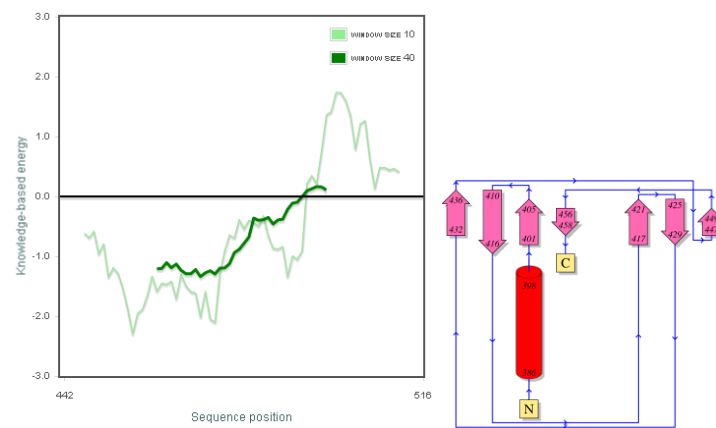
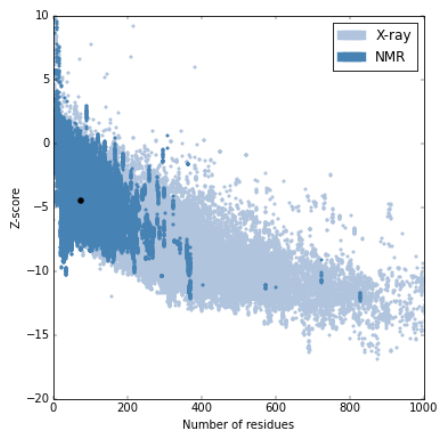
Complexo 1 – cadeia C - Z = -4,48



Complexo 1 – cadeia D - Z = -3,38



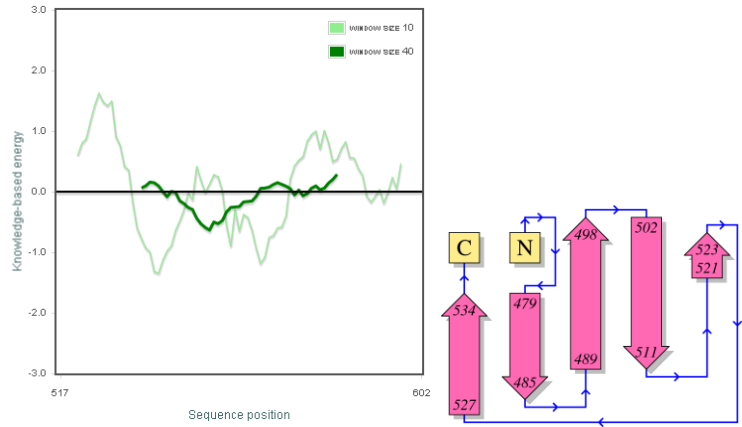
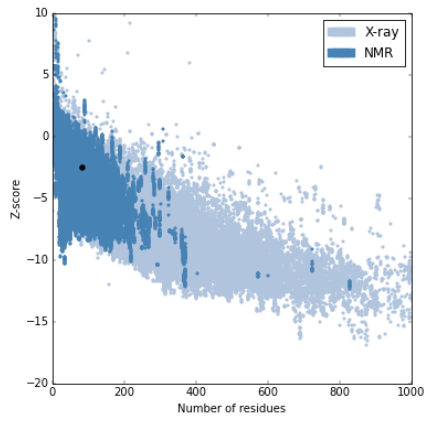
Complexo 1 – cadeia E - Z = -4,49



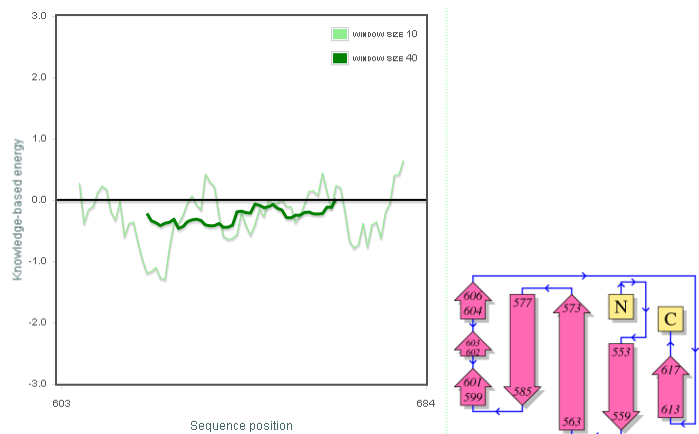
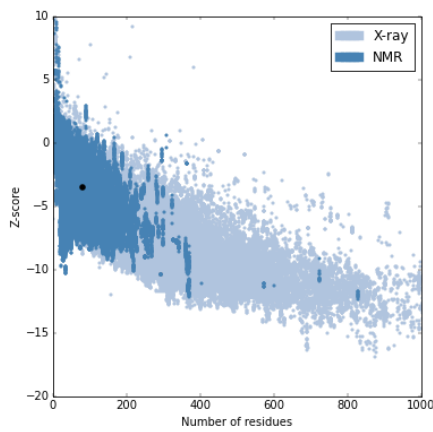
continua

continuação

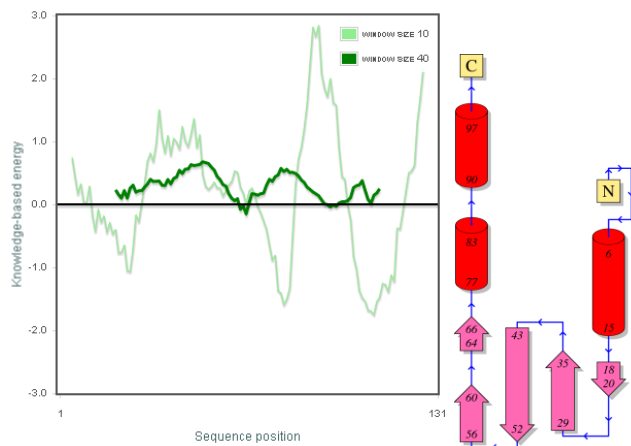
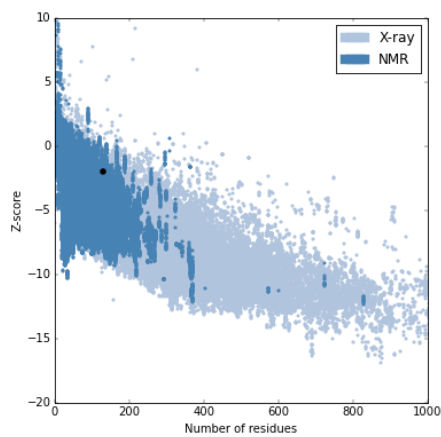
Complexo 1 – cadeia F - Z = -2,50



Complexo 1 – cadeia G – Z = -3,49



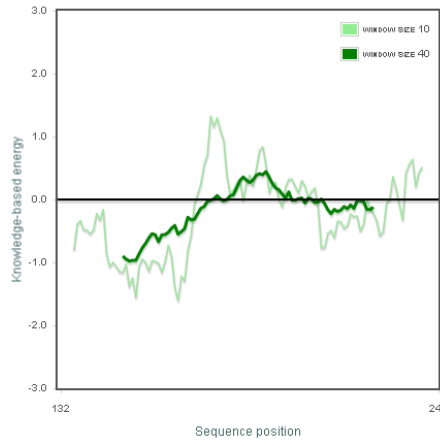
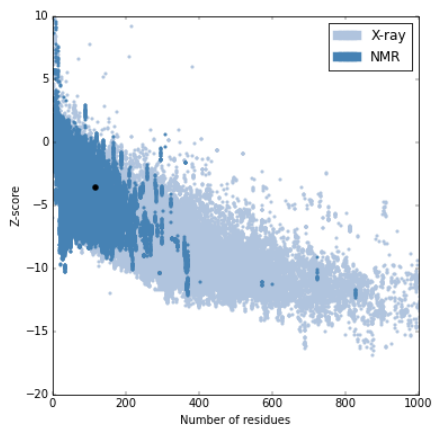
Complexo 2 – cadeia A - Z = -2,00



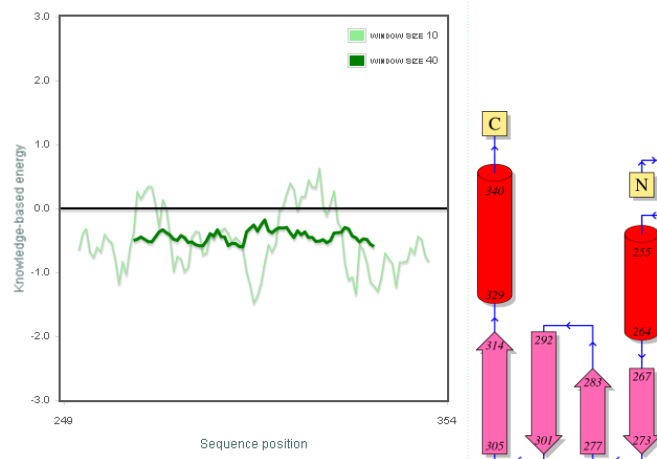
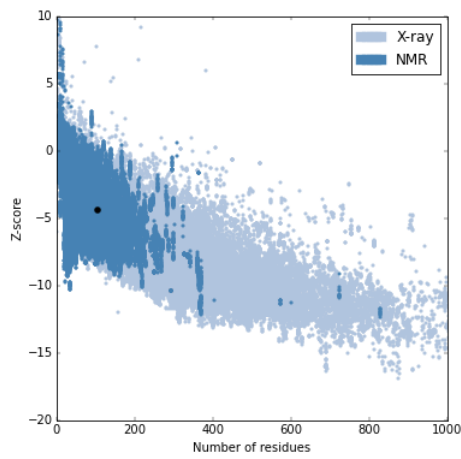
continua

continuação

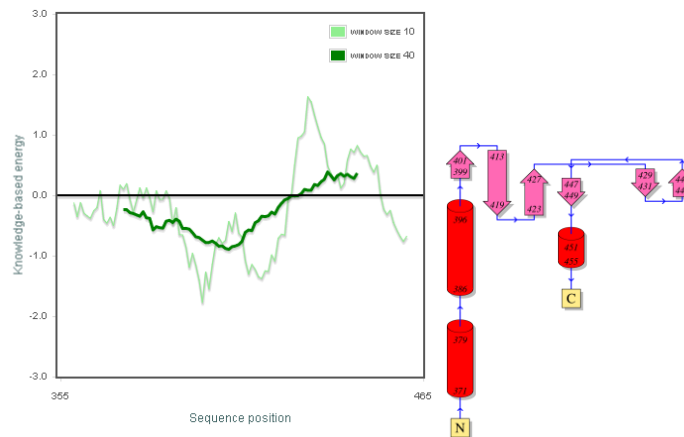
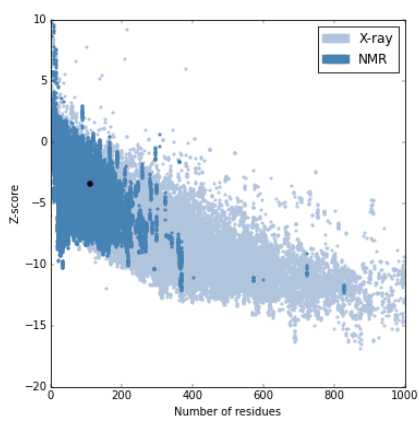
Complexo 2 – cadeia B - Z = -3,53



Complexo 2 – cadeia C - Z = -4,62



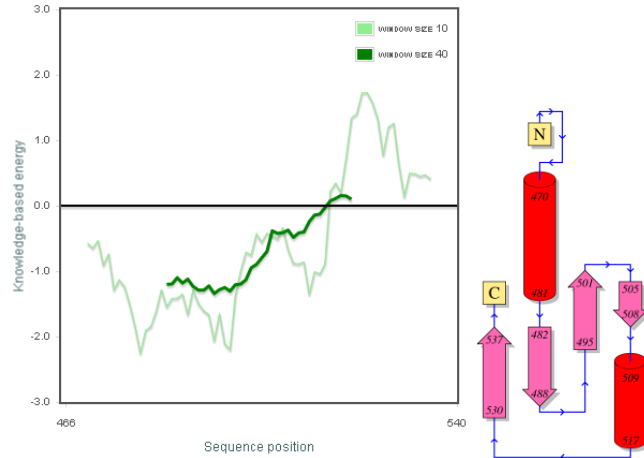
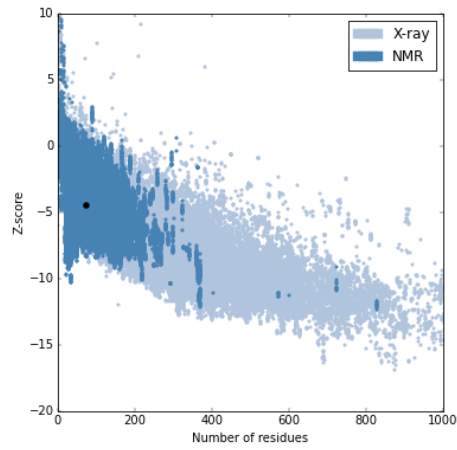
Complexo 2 – cadeia D - Z = -3,39



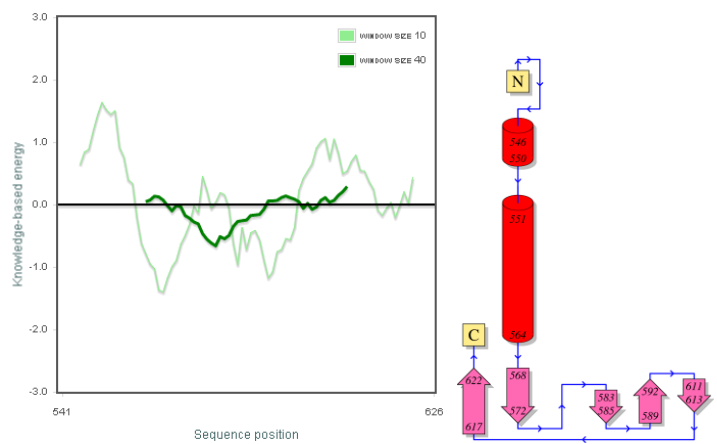
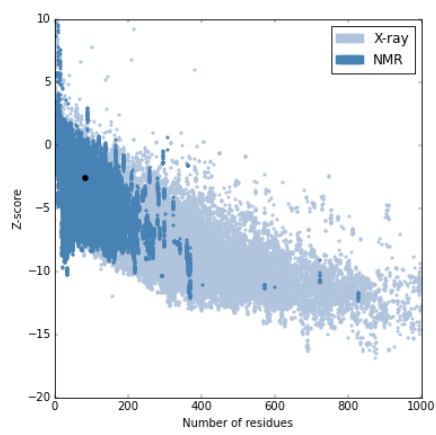
continua

continuação

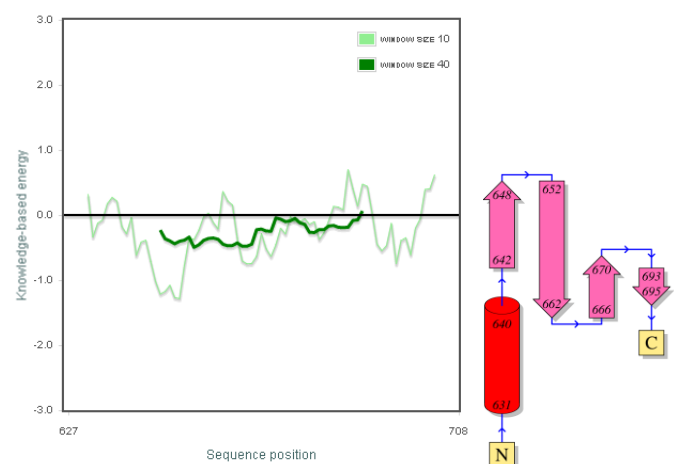
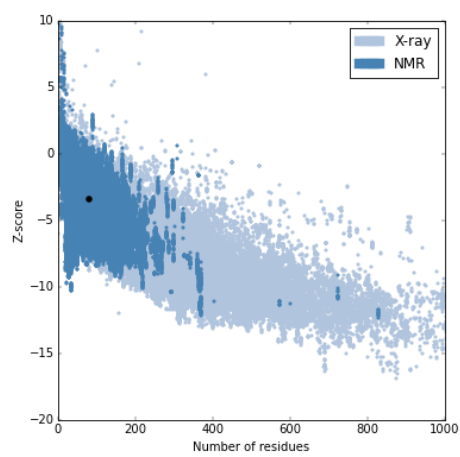
Complexo 2 – cadeia E - Z = -4,49



Complexo 2 – cadeia F - Z=-2,56



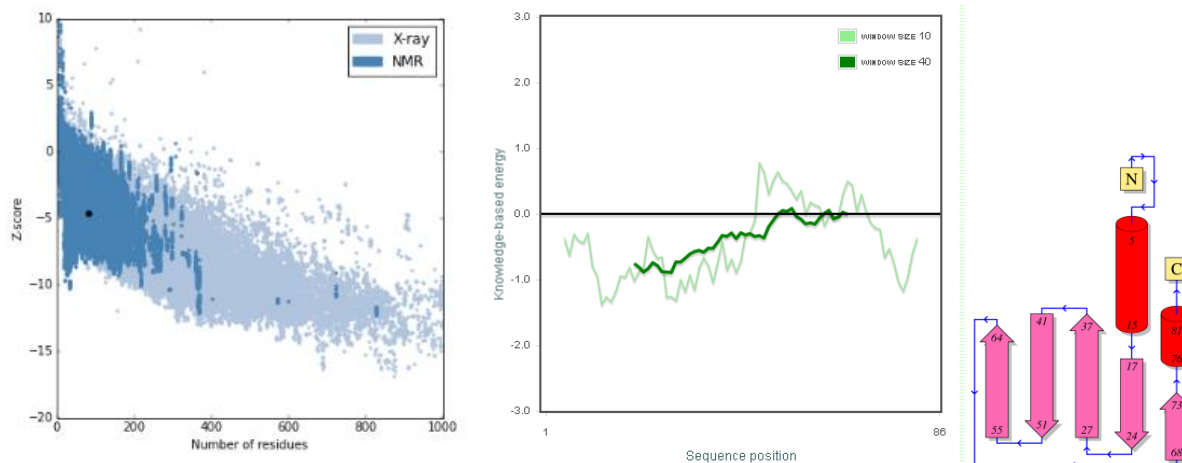
Complexo 2 – cadeia G - Z = -3,37



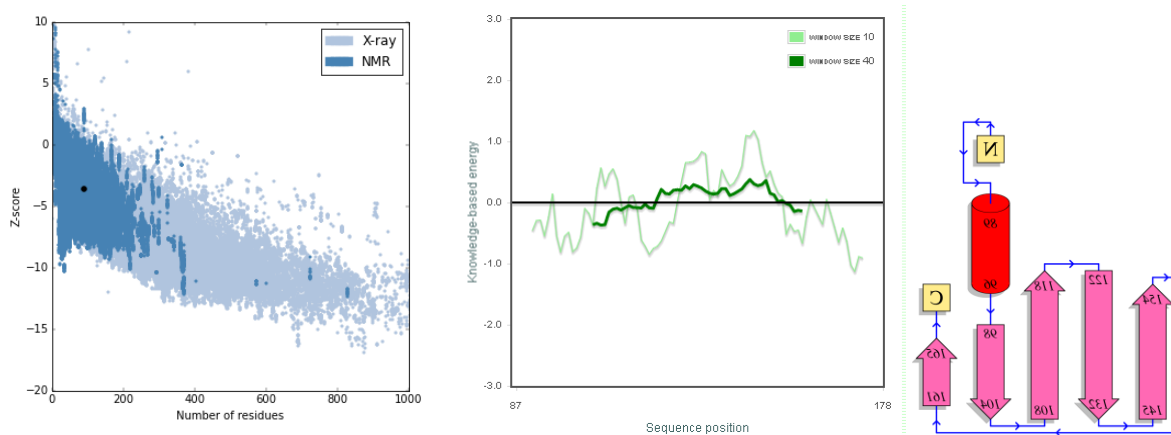
continua

continuação

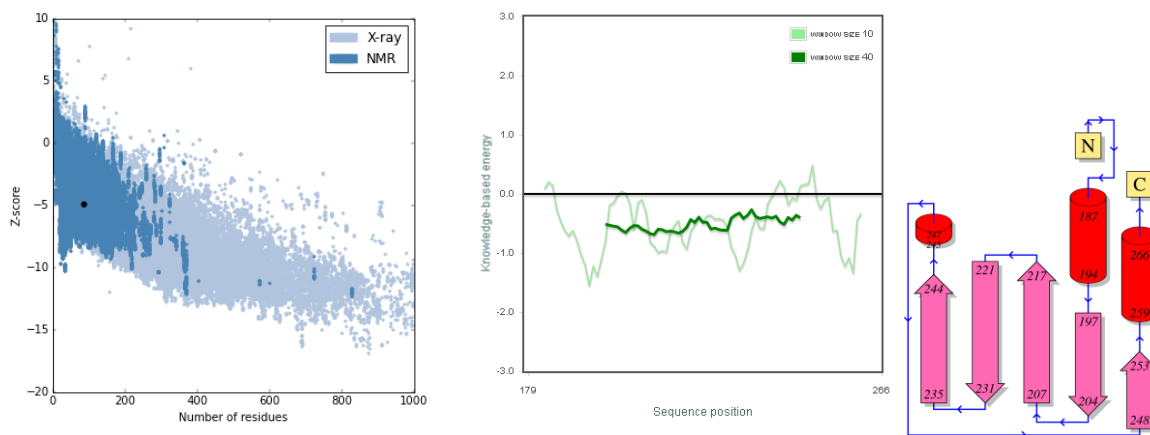
Complexo 3 – cadeia A - Z = -4,62



Complexo 3 – cadeia B - Z = 3,56



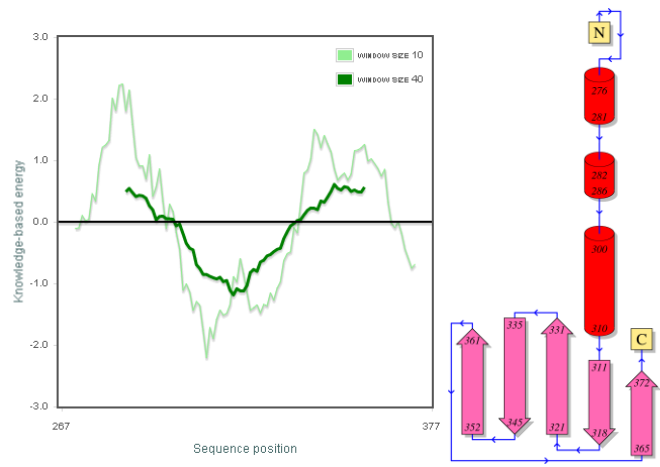
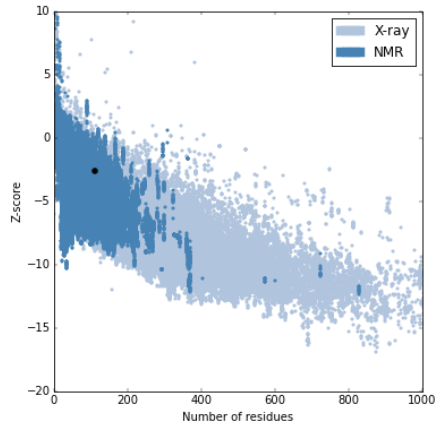
Complexo 3 – cadeia C - Z = -4,94



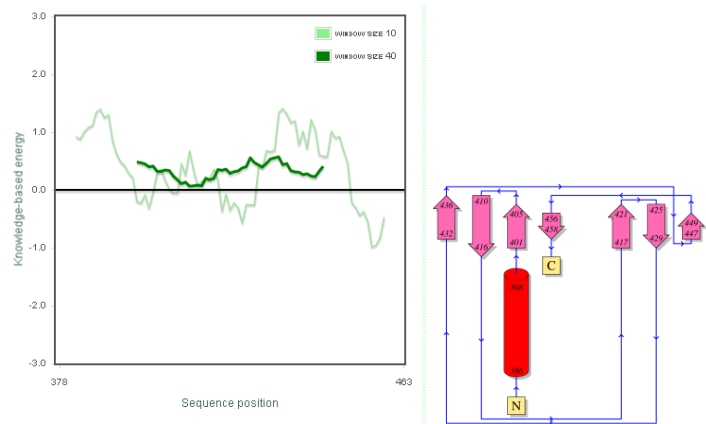
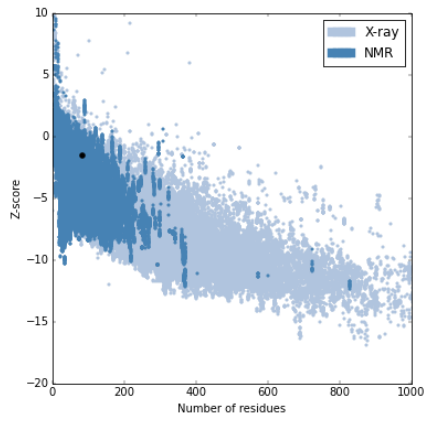
continua

continuação

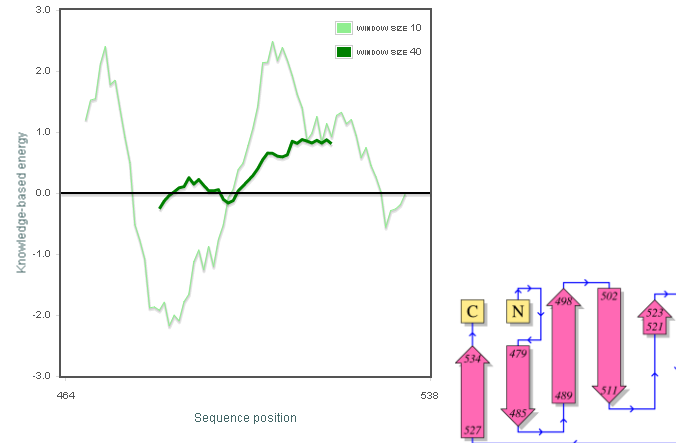
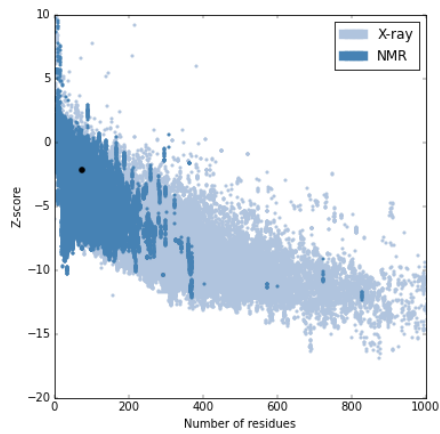
Complexo 3 – cadeia D - Z = -2,62



Complexo 3 – cadeia E - Z = -1,52



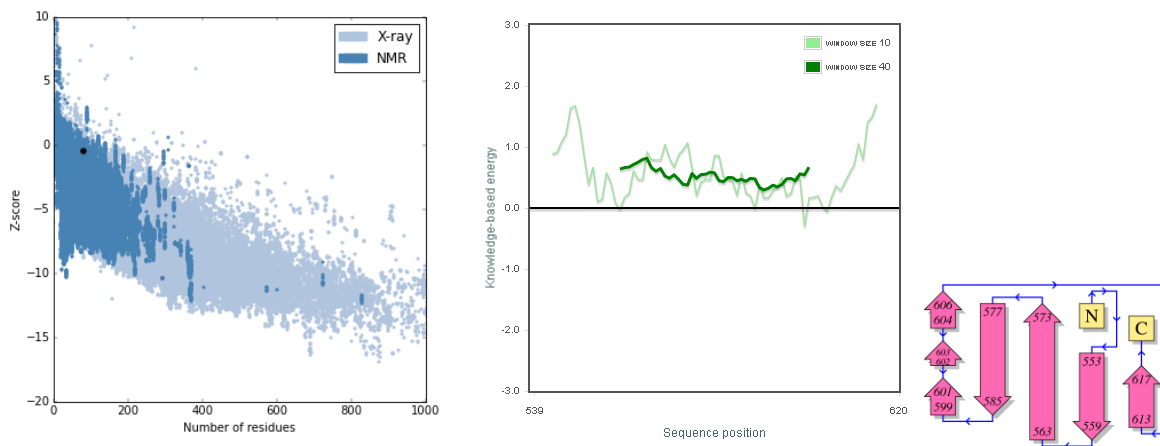
Complexo 3 – cadeia F - Z = -2,14



continua

continuação

Complexo 3 – cadeia G - Z = -0,62



**Figura A2.1 – Avaliação da qualidade dos modelos obtidos por modelagem comparativa dos complexos Sm de *T. brucei*.** São mostrados, da esquerda para a direita, para cada cadeia polipeptídica nos três complexos, a distribuição dos valores de pontuação Z para estruturas cristalográficas e de ressonância magnética nuclear depositadas no PDB e a pontuação Z para o modelo em análise, seguida do gráfico de energia por resíduo de aminoácido medida em janelas de 10 (verde claro) e 40 resíduos (verde escuro), e da topologia da cadeia polipeptídica.

Fonte: Elaborada pelo autor a partir dos resultados dos servidores *ProSa*<sup>230</sup> e *PDBSum*<sup>229</sup>.