PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA
UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA

ANÁLISE COMPUTACIONAL DE CANDIDATOS A HOMÓLOGOS A FATORES DE INICIAÇÃO DA TRADUÇÃO EM TRIPANOSSOMATÍDEOS

Rodolfo Katz

Recife, PE
Julho, 2006

RODOLFO KATZ

ANÁLISE COMPUTACIONAL DE CANDIDATOS A HOMÓLOGOS A FATORES DE INICIAÇÃO DA TRADUÇÃO EM TRIPANOSSOMATÍDEOS

Dissertação apresentada ao Programa de **Pós-Graduação em Genética da Universidade Federal de Pernambuco**, como parte dos requisitos necessários para a obtenção do grau de **Mestre em Genética**.

Orientador: **Dr Osvaldo Pompilio de Melo Neto**, Depto. de Microbiologia, Centro de Pesquisas Aggeu Magalhães.

Recife, PE
Julho, 2006

Katz, Rodolfo

Análise computacional de candidatos a homólogos a fatores de iniciação da tradução em tripanossomatídeos / Rodolfo Katz. – Recife : O Autor, 2006.

132 folhas : il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCB. Genética, 2006.

Inclui bibliografia, apêndices e anexos.

1. Genética – Biologia molecular. 2. Tripanossomatídeos – Iniciação da tradução. 3. Fatores da iniciação – Identificação e caracterização – Análise *in silico.* 4. Fator eIF4F – Modelagem molecular. I. Título.

577.122.5	**CDU (2.ed.)**	**UFPE**
572.645	**CDD (22.ed.)**	**BC2006 – 564**

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

PARECER DA COMISSÃO EXAMINADORA DA DEFESA DE DISSERTAÇÃO DO MESTRANDO
RODOLFO KATZ

"ANÁLISE COMPUTACIONAL DE CANDIDATOS A HOMÓLOGOS A FATORES DE INICIAÇÃO DA TRADUÇÃO EM TRIPANOSSOMATÍDEOS"

Área de Concentração: **BIOLOGIA MOLECULAR**

A comissão examinadora, composta pelos professores abaixo, sob a presidência do primeiro, considera o candidato Rodolfo Katz como aprovado.

Recife, 10 de julho de 2006

Orientador: Dr. Osvaldo Pompilio de Melo Neto (CPqAM/FIOCRUZ)

Titular Interno: Prof. Dr. Antonio Carlos de Freitas (Depto. Genética/UFPE)

Titular Externo: Prof. Dr. Valdir de Queiroz Balbino (Faculdade São Miguel)

Titular Externo: Prof Dr Laura Helena Vega Gonzales Gil (CPqAM/FIOCRUZ)

Suplente: Prof Dr Ederson Akio Kido (Depto. Genética/UFPE)

Suplente: Prof Dr Ana Maria Benko Iseppon (Depto. Genética/UFPE)

Coordenador: Prof. Dr. Marcos Antônio de Morais Jr. (Depto. Genética/UFPE)

AGRADECIMENTOS

Agradeço a todas as pessoas que, direta ou indiretamente, ajudaram o meu trabalho;

a Ewald, Dorothy, Excalibur, Watson (in memoriam), XTAL5, Shion, Saga e Akira;

e à CAPES, pela bolsa de estudos.

Muito Obrigado!

Sumário

LISTA DE TABELAS

LISTA DE FIGURAS

Anexo 1

LISTA DE ABREVIATURAS

3D	Tridimensional
4EBP	4E *Binding Protein* Proteína de ligação ao 4E
AA-box	Caixa de aminoácidos aromáticos/acídicos
ADP	Adenosina Di fosfato
ATP	Adenosina Tri fosfato
BLAST	*Basic Local Alignment Search Tool* Ferramenta Básica de Busca por Alinhamento Local
BLOSUM	*Blocks Substitution Matrix* Matriz de Substituição de Blocos
CDD	*Conserved Domain Database* Banco de Dados de Domínios Conservados
CTD	*Carboxi Terminal Domain* Domínio Carboxi-Terminal
DNA	*Desoxiribonucleic Acid* Ácido Desoxirribonucléico
EIF	*eukaryotic Initiation Factor* Fator de Iniciação Eucarioto
EJC	*Exon Junction Complex* Complexo de Junção de Éxon
EST	*Expressed Sequence Tag* Marcador de Seqüência Expressa
GeneDB	*Gene Data Base* Banco de Dados de Genes
GDP	Guanosina Di Fosfato
GTP	Guanosina Tri Fosfato
HEAT	*Huntingtin, Elongation factor 3*, *A subunit of protein phosphatase 2A [PP2A], and Target of rapamycin* Fator de Elongação Huntingtin 3, Subunidade A da proteína fosfatase 2A [PP2A], e Alvo de ripamicina

InfB	*Translation initiation factor 2 (IF-2; GTPase) [Translation, ribosomal structure and biogenesis]* Fator de iniciação da Tradução 2 (IF-2; GTPase) [Tradução, estrutura ribossomal e biogênese]
IRES	*Internal Ribosome Entry Site* Sítio Interno de Entrada Ribossomal
KDa	kilo Dalton
met	Metionina
Mnk1	*MAP-kinase-interacting kinase-1* Quinase-1 de interação MAP-quinase
MPN	*Mpr1-Pad1-N terminus*
mRNA	*Messenger* RNA RNA mensageiro
NCBI	*National Center for Biotechnology Information* Centro Nacional de Infomação Biotecnológica
NMD	*Non Sense Mediated Decay* Decaimento Mediado Sem Sentido
NTD	*Amino Terminal Domain* Domínio Amino Terminal
OB	*oligonucleotide/oligosaccharide binding fold* Dobramento de ligação a oligonucleotídeos/oligossacarídeos
PABP	*Poli-A Binding Protein* Proteína de Ligação a Poli-A
PAM	*Point Accepted Mutation* Mutação Aceita por Ponto
PARPS	*Procyclic Acid Repetitive Proteins* Proteínas Procíclicas Ácido Repetitivas
PCI	*Proteasome/COP9/Int6* Proteassomo/COP9/Int6
PCR	*Polimerase Chain Reaction* Reação de Cadeia de Polimerase

PDB	*Protein Data Base*
	Banco de Dados de Proteínas
Pfam	*Protein Familie*
	Família Proteica
PM	Peso molecular
RNA	*Ribonucleic Acid*
	Ácido Ribonucléico
RRM	RNA *Recognition Motif*
	Motivo de Reconhecimento de RNA
rRNA	*Ribosomal* RNA
	RNA ribossomal
tRNA	*Transporter* RNA
	RNA transportador
tRNAi	*Transporter* RNA *initiator*
	RNA transportador de iniciação
SUI	*Suppressor of Initiator codon mutations*
	Supressor de Mutações de Códons de Iniciação
TC	*Ternary Complex*
	Complexo Ternário

Resumo

A síntese protéica é um processo básico e essencial para a sobrevivência dos seres vivos. Um dos pontos chave deste processo é a etapa de iniciação da tradução que é regulada pela ação de ao menos doze fatores protéicos chamados eIFs (*eukaryotic Initiation Factor* – Fator de Iniciação de Eucariotos) perfazendo, aproximadamente, 30 polipeptídios em mamíferos. Os tripanossomatídeos, protozoários patogênicos de interesse médico e veterinário, apresentam características celulares próprias como a regulação da sua expressão gênica que ocorre em nível pós-transcricional. Nesse contexto a síntese de proteínas é um alvo em potencial para mecanismos de regulação, entretanto pouco se sabe sobre esse processo nos tripanossomatídeos. Em estudos prévios, foi iniciado nestes parasitas o estudo do fator eIF4F e observou-se a existência de múltiplos homólogos para cada uma de suas três subunidades. Neste trabalho utilizou-se ferramentas de bioinformática para identificar e caracterizar homólogos aos demais eIFs em *Leishmania major*, *Trypanosoma brucei* e *T. cruzi*. Foram identificados homólogos dos fatores eIF1, eIF1A, eIF5, eIF5A, eIF5B, eIF6 e sete subunidades do complexo eIF3 (b, c, d, e, f, i, k). Ao contrário do observado para as subunidades do eIF4F, e com a exceção da subunidade eIF3b, um único homólogo foi identificado para cada fator. A análise das seqüências protéicas mostrou que existe variabilidade no grau de conservação destes homólogos quando comparados com outros eucariotos (de 22% de identidade para o eIF3k até 58% para o eIF6). Em alguns casos foi possível mapear mutações exclusivas dos tripanossomatídeos. Também foram gerados modelos 3D de vários dos homólogos previamente identificados de subunidades do eIF4F facilitando sua caracterização funcional. Os resultados obtidos indicam que boa parte da iniciação da síntese protéica é conservada entre tripanossomatídeos e demais eucariotos. Todavia, diferenças significativas parecem ocorrer e merecem um estudo mais aprofundado.

Palavras-chaves: eIF; Iniciação da Tradução; *Leishamania major*; *Trypanosoma brucei*; *T. cruzi*

1- Introdução

Os tripanosomatídeos são protozoários flagelados que pertencem à ordem Kinetoplastida, de grande importância médica e veterinária, da qual podem-se ressaltar os gêneros *Leishmania* e *Trypanosoma*. Nestes gêneros situam-se os agentes causadores de patologias como a Doença de Chagas, a Doença do Sono e as diversas Leishmanioses. Estas enfermidades flagelam países em desenvolvimento, são de difícil controle por possuírem tratamentos extremamente tóxicos e as vacinas em sua maioria são ineficazes. Diante de tais fatos, busca-se um maior entendimento dos processos biológicos desses parasitas na expectativa de contribuir com o esforço da obtenção de profilaxias e agentes quimioterápicos mais eficazes.

A separação precoce dos tripanosomatídeos da linhagem que gerou os demais eucariotos levou esses parasitas a apresentar uma evolução paralela e características distintas das dos demais eucariotos, como a transcrição dos mRNAs em unidades policistrônicas; processamento em *trans* da unidade policistrônica do mRNA; e a aparente ausência de controle da expressão gênica durante etapas transcricionais e pré-transcricionais, o que sugere que esse controle da expressão ocorra, em sua maior parte, pós-transcricionalmente; entre outras.

Com o intuito de elucidar os mecanismos que regem a fisiologia dos tripanosomatídeos e, desta forma, determinar as semelhanças e diferenças com os demais eucariotos, representantes desta família tiveram seus genomas completamente seqüenciados. A disponibilidade destas seqüências tem permitido um avanço considerável nos estudos destes organismos assim como uma investigação mais eficiente e rápida quanto aos seus processos biológicos básicos. A síntese protéica ou tradução é um dos processos celulares que tem se beneficiado desta abordagem.

A etapa de iniciação da tradução, um dos pontos chave da síntese protéica, é exercida pela atuação de fatores protéicos num processo

complexo e auto-regulado. Em mamíferos, até o momento, foram identificados pelo menos doze fatores de iniciação da tradução denominados de eIFs (*eukaryotic Initiation Factor* – Fator de Iniciação de Eucariotos). Dentre eles, temos o eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4B, eIF4F, eIF4H, eIF5, eIF5A, eIF5B e eIF6, cada qual com seu papel, e a ausência de apenas um destes fatores é suficiente para perturbar todo o processo.

A descrição da iniciação da tradução em tripanossomátideos ainda é incipiente. Mais recentemente, com análises de bioinformática aliadas a técnicas bioquímicas, foi possível uma primeira caracterização de múltiplos homólogos às três subunidades do fator eIF4F. Ensaios bioquímicos preliminares demonstraram que ao menos alguns destes homólogos possuem propriedades compatíveis com a função na tradução, embora diferenças significativas pareçam existir ao se comparar estes fatores com seus homólogos já caracterizados em outros eucariotos. No entanto, além da anotação automática realizada pelos servidores que hospedam os genomas destes tripanosomatídeos, nada se sabe sobre os demais fatores de iniciação da tradução. Em vista disso, procura-se aqui aplicar as ferramentas de bioinformática nos fatores de iniciação da tradução com o intuito de aumentar o entendimento sobre o metabolismo desses parasitas e facilitar a utilização de outras abordagens experimentais.

2- Objetivos

2.1 Objetivo Geral

Analisar, por ferramentas de bioinformática, o conjunto de fatores de iniciação de tradução dos tripanossomatídeos.

2.2 Objetivos Específicos

1- Identificar e caracterizar com ferramentas de bioinformática candidatos a homólogos aos fatores de iniciação da tradução de *Leishmania major*.

2- Buscar em bancos de dados de *Trypanosoma brucei* e *T. cruzi* os ortólogos dos fatores identificados em *L. major*.

3- Analisar a conservação de domínios de ligação a proteínas nos fatores de iniciação da tradução de *L. major*, *T. brucei* e *T. cruzi*.

4- Realizar testes de alinhamento múltiplo com outros organismos para avaliar homologia e distâncias taxonômicas.

5- Criar e estudar estruturalmente modelos para homólogos selecionados de fatores de iniciação da tradução de *L. major* e *T. brucei*.

3- Revisão Bibliográfica

3.1- Tripanossomatídeos

A família *Trypanosomatidae*, pertencente à ordem Kinetoplastida, é compostas por parasitas uniflagelados e subdividida em dois grupos: parasitas monogenéticos (que apresentam um hospedeiro) dos gêneros *Blastocrithidia*, *Crithidia*, *Herpetomonas*, *Leptomonas*, *Rhyncoidomonas* e *Wallaceina*; e parasitas digenéticos (que possuem dois hospedeiros) dos gêneros *Endotrypanum*, *Phytomonas*, *Trypanosoma* e *Leishmania* (Stevens *et al*., 2001).

Morfologicamente estes protozoários apresentam o flagelo numa posição anterior ancorado por microtúbulos, um núcleo usualmente central e a sua divisão celular inicia-se próxima à zona de junção flagelar. Ao longo de suas diversas formas vegetativas, os tripanossomatídeos apresentam alterações morfológicas acentuadas podendo ir desde formas alongadas com grandes flagelos a conformações arredondadas com flagelos quase inexistentes (Figura 1 – Siqueira, 2000).

Esses organismos apresentaram uma evolução singular e tornaram-se parasitas obrigatórios de outros eucariotos após uma divergência precoce em suas linhagens evolutivas o que levou ao aparecimento de características impares. Como exemplo dessas peculiaridades tem-se uma única mitocôndria alterada, o cinetoplasto, que nomeia a ordem. O genoma do cinetoplasto representa cerca de 10-30% do DNA total dos tripanossomatídeos e este DNA “cinetoplasmático” passa pelo processo singular de editoramento (Borst *et al*., 1982; Benne, 1990).

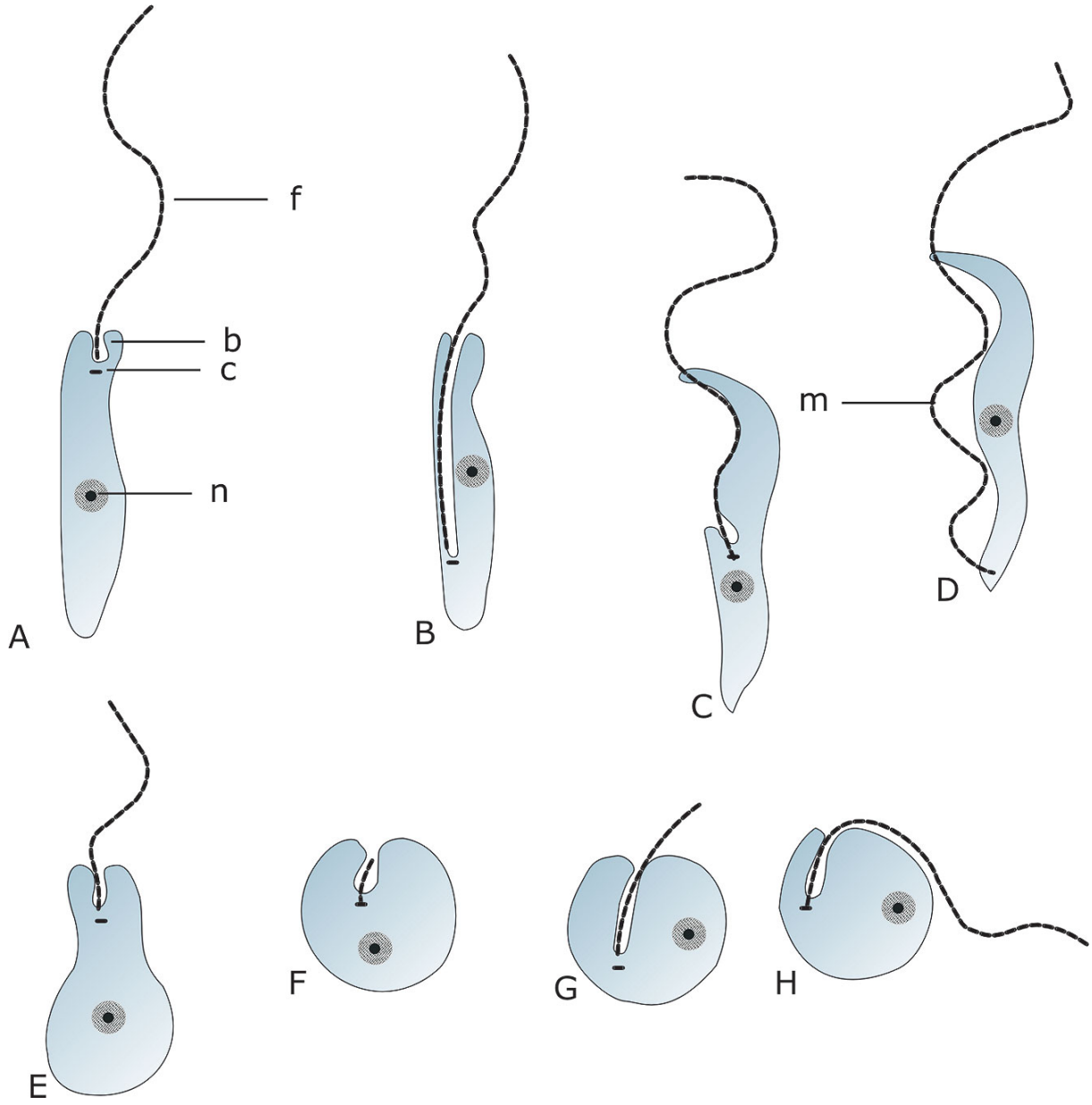


Figura 1: Morfologia dos tripanossomatídeos. A Promastigota; B Opistomastigota; C Epimastigota; D Tripomastigota; E Coanomastigota; F Amastigota; G Paramastigota; H Esferomastigota. b bolsa flagelar, c cinetoplasto, f flagelo, m membrana ondulante e n núcleo (adaptado de Siqueira, 2000).

3.1.1 Aspectos Biológicos e Epidemiológicos dos Tripanossomatídeos

3.1.1.1 *Trypanosoma brucei*

O *T. brucei*, agente causador da Doença do Sono, é transmitido ao homem pela *Glossina palpalis*, a mosca de Tse-Tse, que inocula os parasitas em sua forma procíclica durante o repasto sangüíneo. Ao atingir a corrente sangüínea, o *T.brucei* se diferencia na forma sangüínea delgada e se replica assexuadamente. Quando atinge altos níveis de viremia, diferencia-se numa forma mais curta e não replicativa capaz de infectar insetos reiniciando seu ciclo de vida (Grisard e Steindel, 2000).

O *T. brucei* consegue se ocultar do sistema imunológico dos hospedeiros vertebrados ao efetuar alterações nos antígenos glicoprotéicos de sua superfície (revisto por Borst *et al*., 1996). Já no inseto, esse parasita se furta da ação das proteases do trato digestivo ao expressar prociclinas que são proteínas de superfície protetoras, também chamadas de PARPS (*Procyclic Acid Repetitive Proteins* - Proteínas Procíclicas Ácido Repetitivas) (revisto por Roditi *et al*., 1998).

A Doença do Sono é exclusivamente africana atingindo 36 países subsaarianos e apresenta uma mortalidade anual em torno de 50 mil óbitos. O parasito apresenta duas subespécies que causam quadros clínicos distintos: *T. brucei gambiense* é o responsável pelos casos de infecção crônica, enquanto o *T. brucei rhodesiense* causa uma infecção mais aguda. O *T. brucei* se desenvolve no meio extracelular sanguíneo atingindo a rede linfática, órgãos periféricos e até mesmo o sistema nervoso central. Ao atingi-lo causa desordens neurológicas que usualmente levam à morte do paciente (El-Sayed *et al*., 2000).

3.1.1.2 *Trypanosoma cruzi*

O *T. cruzi* é o agente etiológico da Doença de Chagas, também conhecida como tripanossomíase americana ou esquizotripanose, que atinge principalmente a América Latina apresentando uma prevalência de 13 milhões de pessoas, com uma média de 200 mil novos casos anuais. A patologia apresenta uma fase inicial aguda, caracterizada pelo chagoma de inoculação e os sobreviventes evoluem para um quadro crônico assintomático, a forma indeterminada. É comum, após vários anos na forma assintomática, pacientes chagásicos desenvolverem lesões cardíacas e intestinais por causa da ação do *T. cruzi*, o que consiste na forma crônica sintomática cardíaca e intestinal, respectivamente. Este parasita é transmitido ao hospedeiro vertebrado através da inoculação involuntária de fezes de insetos infectados da subfamília Triatominae que após o repasto sanguíneo defecam sobre o local da picada (de Lana e Tafuri, 2000).

Ao invadir o hospedeiro vertebrado o *T. cruzi* se apresenta na forma de metacíclicos tripomastigotas e em contato com a corrente sanguínea invadem os macrófagos onde se diferenciam em amastigotas replicativos. Após alguns ciclos de replicação, os macrófagos são lisados liberando as formas amastigotas que podem invadir novos macrófagos ou permanecer no fluido extracelular sanguíneo se diferenciando novamente em tripomastigotas. O *T. cruzi* pode então invadir órgãos internos como o esôfago, intestino e principalmente o coração ou ser ingerido por insetos durante seu repasto sanguíneo. No interior do estômago dos triatomíneos, o *T. cruzi* se diferencia em formas epimastigostas replicativas. Algumas destas formas podem migrar para o reto do inseto onde se diferenciam em metacíclicos tripomastigotas que serão excretados junto com as fezes reiniciando o ciclo (de Lana e Tafuri, 2000).

3.1.1.3 O gênero *Leishmania*.

O gênero *Leishmania* é composto por pelo menos 20 agentes etiológicos causadores das diversas formas de leishmanioses, classificadas de acordo com as diferenças que apresentam no grau de invasão do corpo do hospedeiro. Fala-se em leishmaniose cutânea quando a infecção é caracterizada por lesões de pele múltiplas; leishmaniose muco-cutânea quando ocorre ulcerização das mucosas nasais e orais; leishmaniose visceral quando ocorre invasão do fígado e baço. Estes parasitas possuem uma distribuição cosmopolita sendo encontrada pelas Américas do Sul e Central, África Central, Oriente Médio, Paquistão, China e sudeste da Ásia. Uma população de aproximadamente 350 milhões de pessoas encontra-se em áreas de risco e cerca de 400 mil novos casos são confirmados anualmente (Genaro, 2000; Michalick, 2000; Desjeux, 2004). O agente transmissor destes parasitas no Velho Mundo são insetos dípteros do gênero *Phlebotomus*. Já no Novo Mundo a transmissão se dá por insetos pertencentes ao gênero *Lutzomyia* (Williams, 2000).

Durante o repasto sanguíneo dos insetos transmissores, formas promastigotas da *Leishmania* são inoculadas na corrente sanguínea do hospedeiro vertebrado. Após a inoculação estas infectam células mononucleares fagocitárias e lá se diferenciam em formas amastigotas replicativas. Após alguns ciclos de replicação, os monócitos terminam por ser lisados liberando formas amastigotas na corrente sanguínea. Nesse momento, podem invadir novos monócitos ou serem capturados pelo repasto sanguíneo de um inseto. Uma vez no interior do trato digestivo do inseto, as formas amastigotas se diferenciam em promastigotas replicativos fechando o ciclo de transmissão deste parasita (Genaro, 2000; Michalick, 2000; Desjeux, 2004).

3.1.2 Projetos Genoma de Tripanossomatídeos e Regulação da Expressão Gênica

Com o intuito de elucidar os mecanismos que regem a fisiologia dos tripanosomatídeos e, desta forma, determinar as semelhanças e diferenças com os demais eucariotos, representantes desta família tiveram seus genomas completamente seqüenciados. São eles *L. major*, *T. brucei* e *T. cruzi* (Berriman *et al*., 2005; El-Sayed *et al*., 2005a; Ivens *et al*., 2005). Outros integrantes destes gêneros estão em estado mais inicial de seqüenciamento, como é o caso de *L. brazilienses* e *L. infantum*. Conforme a progressão do seqüenciamento, as informações obtidas foram armazenadas em bancos de dados e se iniciou a montagem do genoma que pode ser analisado pelo uso de ferramentas de bioinformática (El-Sayed *et al*., 2005b). Tal metodologia permite o estudo do genoma destes parasitas, o que pode levar a um novo entendimento dos seus diversos aspectos celulares e, finalmente, a proposição de novas abordagens de pesquisa. Um dos processos celulares cujo estudo tem se beneficiado com tal abordagem é a síntese protéica ou tradução (Dhalia *et al*., 2005).

A partir do estudo do genoma dos tripanossomatídeos, observou-se que estes apresentam a quase totalidade dos seus genes codificantes de proteínas transcritos conjuntamente em blocos de unidades policistrônicas (Johnson *et al*., 1987; Muhich e Boothroyd, 1988; Alonso *et al*., 1992). Promotores que regulem a etapa de transcrição destes genes estão aparentemente ausentes (Clayton *et al*., 2002). É sabido que após a transcrição das unidades policistrônicas, estas são processadas através do mecanismo de *trans-splicing* que corta o transcrito primário, liberando diversos fragmentos de mRNA e adicionando a cada fragmento um “mini-éxon” em sua extremidade 5’, a seqüência *splice-leader*. Este “mini-éxon” apresenta na sua extremidade 5’ uma guanosina metilada adicionada em orientação reversa ao RNA, o cap dos demais eucariotos, que nestes organismos é seguido por mais quatro nucleotídeos metilados gerando o chamado *cap 4*. A maturação dos mRNAs prossegue com a adição de uma

cauda de poliadeninas na sua extremidade 3' e os mRNAs maduros são então transportados para o citoplasma para serem traduzidos (Clayton *et al*., 2002; Liang *et al*., 2003).

O que parece é que os tripanosomatídeos se destacam em relação aos demais eucariotos, e mesmo procariotos, em que a maior parte do controle da sua expressão gênica ocorre a nível pós-transcricional (Clayton *et al*., 2002). Nesse contexto, a síntese protéica ou tradução, e mais especificamente a iniciação da tradução, torna-se um alvo interessante de estudo uma vez que pode revelar mecanismos inéditos de controle da expressão gênica.

3.2 Biologia Computacional e Bioinformática

Os avanços tecnológicos estão dinamizando as pesquisas em biologia molecular como é o caso do desenvolvimento de técnicas como a PCR (*Polimerase Chain Reaction* – Reação em Cadeia de Polimerase) e de ferramentas como os seqüenciadores automatizados. De fato, tais aparatos permitem que a execução dos diversos projetos genoma gere um grande número de seqüências de DNA e de proteínas num tempo relativamente curto. Devido ao acúmulo dessas informações fez-se necessário o desenvolvimento de ferramentas que permitissem o armazenamento e posterior estudo destas seqüências. Surgiu, desta forma, uma nova ciência que uniu a computação e a biologia: a Bioinformática (o termo Biologia Computacional é utilizado de forma intercambiável). As ferramentas da bioinformática possuem um amplo espectro de aplicação, indo desde a organização das informações em bancos de dados, passando pela comparação de seqüências até o estudo de padrões estruturais de proteínas e a modelagem molecular das mesmas (Corpet, 1988; Higgins e Sharp, 1988; Altschul *et al*., 1990; Benson *et al*., 2000; Gibas e Jambeck, 2001; Forster, 2002).

3.2.1 Análise de seqüências

A comparação entre seqüências de proteínas e de nucleotídeos encontra-se, atualmente, contida no âmago da bioinformática e os bancos de dados de seqüências estão entre os maiores e mais importantes bancos de dados da atualidade. As ferramentas de análise de seqüências possibilitam uma série de estudos distintos e, através deles, é possível propor se determinado conjunto de proteínas apresenta funções similares ou uma estrutura de motivos compartilhada, bem como um histórico evolutivo comum ou mesmo se são ou não relacionadas (Gibas e Jambeck, 2001; Baxevanis, 2005).

Para um melhor entendimento destas análises se faz importante distinguir dois termos que usualmente são utilizados de forma intercambiável, mas que, na verdade, representam conceitos distintos: similaridade e homologia. Similaridade é a uma medida quantitativa, baseada numa observação, de quão similares são duas seqüências. A similaridade é aferida ao se alinhar duas seqüências e se fazer uma contagem dos aminoácidos ou nucleotídeos corretamente alinhados. Aminoácidos corretamente alinhados são aqueles que apresentam características equivalentes e os idênticos. Este número pode ser convertido em um valor percentual que representa a gradação de similaridade. Dentre os valores de similaridade, tem-se a identidade quando se conta apenas os resíduos identicamente alinhados. Outro valor de interesse é a positividade quando resíduos de propriedades físico-químicas similares também são considerados para a computação do percentual final de similaridade. Por outro lado, quando se fala em homologia refere-se a um parentesco mais intrínseco entre as seqüências. Quando as seqüências são homólogas, elas indicam uma origem ancestral comum podendo ou não apresentar atividades comuns, podendo ou não divergir quanto à similaridade seqüencial, portanto não existe gradação de homologia (Gibas e Jambeck, 2001; Baxevanis, 2005; Brinkman, 2005).

Os programas de bioinformática atuais conseguem comparar uma seqüência alvo contra um vasto banco de dados em poucos segundos com grande acuidade. Levando-se em consideração alguns parâmetros, pode-se determinar se o alinhamento ocorreu de forma aleatória ou se existe uma similaridade real entre as proteínas ou DNAs. Para se obter tais alinhamentos, foram criados diversos algoritmos que abordam a questão de formas diferentes. É possível agrupar as abordagens em dois grupos: métodos globais e locais. A primeira abordagem procura encontrar o melhor alinhamento possível entre duas seqüências ao longo de toda a sua extensão, sendo mais usada com seqüências altamente similares e de tamanhos parecidos. Já o método local de alinhamento diverge do primeiro por permitir que as seqüências “deslizem” uma sobre as outras, o que pode gerar blocos de alinhamento que permitem a busca por regiões isoladas de similaridade. Portanto, quando se realiza uma procura contra um banco de dados com muitas seqüências depositadas, entre seqüências de tamanhos divergentes e com uma distância evolutiva significativa, os algoritmos de alinhamento local costumam gerar melhores resultados uma vez que lacunas muito grandes tendem a levar aos algoritmos de alinhamento global a abortarem o processo (Gibas e Jambeck, 2001; Baxevanis, 2005).

Para a análise dos diversos alinhamentos de seqüências de proteínas foram criadas as chamadas matrizes de substituição. Estas fornecem valores para cada mutação de resíduo de aminoácido podendo levar em conta a conservação das posições, a freqüência observada das mutações em grupos protéicos específicos, aspectos químicos dos aminoácidos ou mesmo a estrutura terciária de proteínas determinadas experimentalmente (Gibas e Jambeck, 2001; Baxevanis, 2005). Dentre as diversas matrizes de substituição, pode-se citar como exemplo as famílias PAM (*Point Accepted Mutation* – Mutação Aceita por Ponto) (Dayhoff *et al*., 1978) e BLOSUM (*Blocks Substitution Matrix* - Matriz de Substituição de Blocos) (Henikoff e Henikoff, 1992). Matrizes diferentes possuem valores

diferentes e, portanto, alinhamentos distintos analisados por matrizes diferentes não são diretamente comparáveis.

Um dos programas mais usados para a análise entre seqüências é o BLAST (*Basic Local Alignment Search Tool* - Ferramenta Básica de Busca por Alinhamento Local). Esse programa utiliza a estratégia local de alinhamento e tem sua primeira versão datando de outubro de 1990 (Altschul *et al*., 1990). O BLAST dispõe de várias ferramentas e possui cinco programas básicos: o BLASTN, o BLASTP, o BLASTX, o TBLASTN e o TBLASTX. Cada programa é utilizado em situações específicas como mostra a Tabela 1.

Tabela 1: Relação dos algoritmos inclusos no BLAST e suas aplicações (adaptado de Baxevanis, 2005).

Programa	Seqüência pergunta	Banco de dados
BLASTN	Nucleotídeos	Nucleotídeos
BLASTP	Proteína	Proteína
BLASTX	Nucleotídeos traduzidos nas seis matrizes de leitura	Proteína
TBLASTN	Proteína	Nucleotídeos traduzidos nas seis matrizes de leitura
TBLASTX	Nucleotídeos traduzidos nas seis matrizes de leitura	Nucleotídeos traduzidos nas seis matrizes de leitura

Para a avaliação dos alinhamentos executados pelo BLAST estão disponíveis três parâmetros: o *raw score*, o *bit score* e o *E-value*. O primeiro, no caso de alinhamentos de seqüências protéicas, é o somatório do valor atribuído a cada par aminoácido-aminoácido e aminoácido-lacuna (em inglês *gap*) que compõe o alinhamento, obtido pelo uso de uma matriz de substituição. O BLAST utiliza como matriz padrão para alinhamentos entre proteínas a matriz BLOSUM 62 que pode ser vista na Figura 2. O *bit score* é, na realidade, o *raw score* normatizado por uma mudança de base logarítmica. Dessa forma os diferentes *bit scores* de alinhamentos distintos podem ser comparados. Por fim, o *E-value* é a probabilidade de ocorrer ao acaso um alinhamento com um *score* igual ou

maior ao observado num determinado banco de dados. Portanto, quanto maiores os valores do *raw score* e do *bit score* melhor é o alinhamento, já para o *E-value*, quanto menor o valor, melhor (Gibas e Jambeck, 2001; Baxevanis, 2005).

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													P
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-3	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

Figura 2: Matriz BLOSUM 62 (adaptado de Henikoff e Henikoff, 1992). Matriz padrão do BLAST para a obtenção do *score* de um alinhamento entre seqüências. Como por exemplo, a conservação de uma Leucina tem um valor de +4 (circulado em vermelho) enquanto que a substituição de uma leucina por uma Alanina tem um valor de –1 (circulado em azul).

Além da comparação de uma certa seqüência contra uma segunda outra análise muito importante da bioinformática é o alinhamento de três ou mais seqüências, os alinhamentos múltiplos. A utilização deste procedimento permite: inferência de função através da comparação seqüencial; a análise de domínios conservados numa dada proteína entre espécies distintas; o estudo de distâncias filogenéticas entre as espécies; além de ser um dos pontos chaves para a predição de estrutura secundária das proteínas (Barton, 2005). Também é interessante notar

que, em alguns casos, o alinhamento múltiplo pode até mesmo aumentar a qualidade de um alinhamento entre duas seqüências (Russell e Barton, 1992). Uma das abordagens utilizadas para se obter alinhamentos múltiplos é através de métodos hierárquicos (revisto por Barton, 2005) e, dentre os programas utilizados para a análise de múltiplas seqüências que se utilizam deste método, um dos mais conhecidos é o ClustalW (Higgins e Sharp, 1988; Barton, 2005).

3.2.2 Análise Filogenética

A análise filogenética é o estudo de características de determinados organismos com o intuito de se estabelecer o histórico da evolução das espécies. Com o advento da bioinformática, as seqüências protéicas e de nucleotídeos estão ultrapassando, em importância, o uso de características morfológicas nas análises filogenéticas. Vale ressaltar que toda análise filogenética se baseia no estudo de características e, no caso específico das análises baseadas em alinhamentos de seqüências, cada coluna representa uma característica. Com base nisso, é importante notar que a qualidade do alinhamento utilizado irá influir diretamente na análise, portanto, um alinhamento errôneo levará a inferências também erradas. Existem diversas abordagens para os estudos filogenéticos que permitem a criação de árvores filogenéticas as quais permitem uma rápida avaliação dos dados. Uma destas abordagens é o *Neighbour joining* que utiliza análise de matrizes de distâncias para a construção de árvores filogenéticas. Os resultados destas análises podem ser avaliados, para a verificação de sua acuidade, por testes de permutação como é o caso do *Bootstrap* que reconstrói a árvore do conjunto de dados inúmeras vezes e calcula a proporção de vezes que uma determinada topologia se forma (Gibas e Jambeck, 2001; Brinkman, 2005).

3.2.3 Análise Estrutural

Uma outra abordagem para o estudo de proteínas é a modelagem estrutural das mesmas. Uma vez determinada a estrutura protéica tridimensional é possível, por exemplo, o estudo funcional de proteínas, a proposição de sítios catalíticos e de superfícies de ligação a ligantes e outras proteínas. Também é possível inferir exatamente quais aminoácidos estão envolvidos nessas interações e como eles realizam seu papel. Dessa forma, a estrutura ainda permite a racionalização de ensaios que visem sítios específicos. Portanto, a obtenção da estrutura tridimensional de uma proteína é muito importante para sua caracterização (Gibas e Jambeck, 2001).

Atualmente, pode-se determinar, com altíssima definição, a estrutura de uma proteína experimentalmente através da realização de ensaios de espectroscopia por ressonância magnética nuclear ou através da resolução de estruturas cristalográficas oriundas da difração de raios X. No entanto, estas técnicas possuem o inconveniente de serem muito laboriosas, custosas, demoradas e algumas vezes impossíveis de serem realizadas. Por causa dessas dificuldades o banco de estrutura de proteínas está muito defasado, em termos numéricos, em comparação com os genes seqüenciados (Forster, 2002).

Diante desses impedimentos técnicos, surgiu o interesse de se buscar abordagens alternativas para a determinação da estrutura das proteínas. Uma das soluções foi o desenvolvimento de ferramentas de bioinformática que permitam a modelagem de proteínas *in silico*. Essas abordagens, que podem independer total ou parcialmente de ensaios experimentais prévios, representam alternativas rápidas, econômicas e viáveis para a obtenção de modelos tridimensionais de proteínas (Gibas e Jambeck, 2001; Forster, 2002).

Até o momento, ainda não foi possível a resolução do problema do enovelamento protéico. Desta forma, a aplicação de algoritmos de

predição por *ab initio*, que sejam capazes de predizer a estrutura tridimensional de uma proteína unicamente através da sua seqüência de aminoácidos, ainda é experimental, exige uma capacidade computacional muito grande e não é utilizada de forma cotidiana, sendo reservada para proteínas que não possuam informações de homólogos ou de estruturas similares conhecidas (Simons *et al*., 2001; Wishart, 2005). Por outro lado, métodos comparativos, como a modelagem molecular por homologia, têm obtido bons resultados e são mais acessíveis. Esse método prediz a estrutura de uma proteína enfatizando a semelhança seqüencial de aminoácidos em relação a uma outra proteína de estrutura tridimensional conhecida, baseado na premissa de que essa similaridade seqüencial implica em similaridade estrutural (Forster, 2002).

Além destas abordagens, têm sido desenvolvidos algoritmos para a modelagem de proteínas que não possuem homólogos com alto grau de identidade seqüencial com estrutura conhecida. Nestes métodos, o que é levado em consideração é a estrutura secundária da proteína em questão. A base desta homologia está na observação de que duas proteínas distintas, à luz de sua seqüência de aminoácidos, podem apresentar enovelamento semelhante. Expandindo o conceito para a estrutura terciária de proteínas, é possível que duas proteínas que não possuam identidade seqüencial elevada, possuam estrutura tridimensional e função semelhante (McGuffin e Jones, 2002; Gibas e Jambeck, 2001; Forster, 2002).

3.3. Iniciação da Tradução em Eucariotos

A síntese protéica constitui um dos processos essenciais para a sobrevivência de todos os organismos. Este processo possui diversas etapas que são controladas por fatores também protéicos que contribuem na regulação da expressão gênica de acordo com as necessidades celulares. Um dos pontos críticos do controle da síntese protéica ocorre no momento da iniciação da tradução. Nesse ponto de controle, os fatores de iniciação da tradução, os eIFs, têm como função promover a correta associação entre o mRNA e o ribossomo, assegurando que a síntese protéica se inicie no códon correto (Lewin, 1994).

Resumindo a iniciação da tradução em eucariotos, temos primeiro a dissociação do ribossomo 80S nas suas subunidades 40S e 60S. O tRNA de iniciação, contendo a metionina (Met-tRNAi), associa-se à subunidade ribossomal menor 40S. Desta forma, o complexo de pré-iniciação 43S é montado. Este complexo, por sua vez, se associa ao mRNA por um conjunto de fatores que reconhecem o cap (guanosina metilada adicionada à extremidade 5' do mRNA durante a maturação do transcrito primário). Em seguida, o complexo 43S vasculha a extremidade 5' do mRNA em busca do códon AUG de iniciação da tradução, em um processo denominado de *scanning*. Ao chegar no códon correto, ocorre a re-associação das subunidades ribossomais e o início da síntese protéica propriamente dita (Kapp e Lorsch, 2004).

Os estudos visando conhecer a iniciação da tradução em mamíferos conseguiram identificar, até o momento, ao menos doze fatores de iniciação da tradução que são o eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4F, eIF4B, eIF4H, eIF5, eIF5A, eIF5B e eIF6. Os fatores de iniciação podem se apresentar sob a forma de complexos protéicos, possuir diversas subunidades e podem, ainda, ter isoformas. Estes fatores participam de um processo altamente regulado onde a ausência de apenas um deles é suficiente para diminuir a eficácia de toda a síntese protéica ou até

mesmo impedi-la por completo (Kapp e Lorsch, 2004). A Figura 3 ilustra este processo.

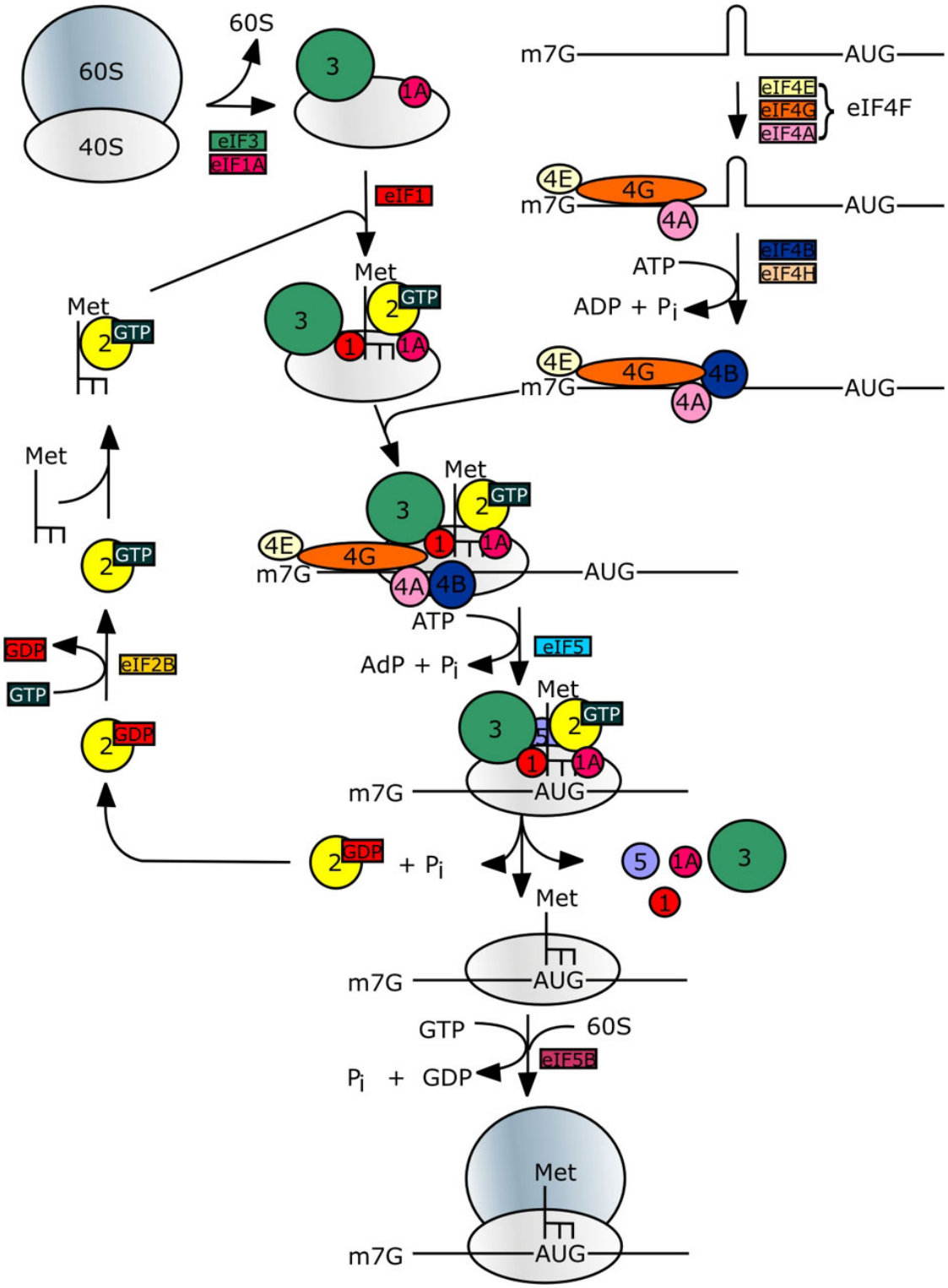


Figura 3: Esquema da iniciação da tradução em eucariotos. Os fatores de iniciação são mostrados como círculos coloridos e aparecem quando de sua primeira participação no processo. Os símbolos para o Met-tRNAi, mRNA e para as subunidades ribossomais são evidentes (adaptado de Hershey e Merrick, 2000).

Num primeiro momento o ribossomo é dissociado, sendo os fatores eIF3 e eIF6 os responsáveis por essa etapa. Estes promovem a dissociação do ribossomo 80S nas subunidades 40S e 60S que servem de sítios de ligação para o eIF3 e eIF6, respectivamente. Esses fatores também são responsáveis por impedir a re-associação precoce do ribossomo. Outra proteína relacionada à etapa de dissociação ribossomal é o fator eIF1A que também está envolvido com a ligação do Met-tRNAi, interação à subunidade ribossomal 40S pelo sítio A, ligação ao mRNA e com o *scanning* (Hershey e Merrick, 2000).

O fator responsável pelo recrutamento do tRNA que contém a metionina de iniciação (Met-tRNAi), é o eIF2. Esse fator se associa ao Met-tRNAi formando um complexo ternário (TC – *Ternary Complex*) com a presença de GTP. A esse complexo dá-se o nome de eIF2-GTP-Met-tRNAi e, após sua formação, o eIF2 se associa à subunidade ribossomal 40S, provavelmente no seu sítio P, pela ação de mais dois fatores de iniciação: o eIF1 e o eIF5. Assim, temos a formação de um intermediário de pré-iniciação na subunidade ribossomal 40S, denominado de complexo 43S, constituído pela associação dos fatores eIF1, eIF2 (associado a GTP e Met-tRNAi), eIF3, eIF5 e a própria subunidade 40S (Asano *et al*., 2000).

Paralelamente aos eventos envolvendo o ribossomo, ocorre o reconhecimento do mRNA por fatores que vão permitir a sua interação com o complexo 43S. Dentre os fatores envolvidos com a associação entre o mRNA e o ribossomo temos o fator eIF4F que possui ao menos três funções exercidas por subunidades distintas, são elas: o reconhecimento do cap (Gingras *et al*., 1999); a desnaturação ATP-dependente de estruturas secundárias no mRNA que podem dificultar a ligação do ribossomo e o *scanning* (Gingras *et al*., 1999; Hershey e Merrick, 2000); a ancoragem ou suporte de vários dos complexos protéicos envolvidos na iniciação da tradução. Este fator ainda participa da circularização do mRNA, interagindo com fatores associados a cauda de poliadenina na sua extremidade 3' (Gingras *et al*., 1999).

Após a montagem do complexo eIF4F no mRNA e do recrutamento do complexo ternário eIF2-GTP-Met-tRNAi para a subunidade ribossomal menor, ocorre a interação entre mRNA e ribossomo. Essa etapa se dá através da associação do eIF4F e do eIF3 que é seguida da busca pelo códon AUG que sinaliza o início da seqüência codificadora de proteína propriamente. Uma vez encontrado este códon de iniciação da tradução, o fator eIF5 estimula a hidrólise do GTP, associado ao eIF2, o que leva ao recrutamento da subunidade ribosomal 60S, e à liberação dos demais fatores de iniciação. Após este evento, com a restauração do ribossomo 80S, se inicia a fase de elongação da tradução ou síntese protéica propriamente dita (Hershey e Merrick, 2000). Nos proximos tópicos encontra-se uma descrição mais detalhada dos fatores protéicos individuais envolvidos na iniciação da tradução e que são objeto de estudo desta dissertação.

3.3.1 eIF1

O fator eIF1, conhecido como SUI1 (*suppressor of initiator codon mutations* – Supressor de Mutações de Códons de Iniciação) em leveduras, é o menor dos fatores de iniciação da tradução conhecidos, com 12.7 kDa em humanos, sendo essencial para a viabilidade celular (Yoon e Donahue, 1992). Juntamente com o eIF1A, está envolvido na formação do complexo 43S-mRNA e também com o processo de busca pelo códon de iniciação da tradução, assegurando o inicio da tradução no AUG correto. Já foi visto que mutações em determinados resíduos do eIF1 resultam em iniciação da tradução no códon UUG e outras relacionadas com mudanças na matriz de leitura. Dessa forma, o eIF1 possui um papel crítico no reconhecimento do códon AUG (Kyrpides e Woese, 1998; Fletcher *et al*., 1999; Hershey e Merrick, 2000; Maag *et al*., 2005).

3.3.2 eIF1A

O eIF1A é uma proteína pequena e estável com aproximadamente 20 kDA, essencial para a viabilidade celular em levedura, semelhante ao fator de iniciação IF1 de bactérias (apresenta 21% de identidade seqüencial com o fator IF1 de *Escherichia coli*). Apresenta dois domínios estruturais e suas extremidades são altamente polarizadas, sendo a amino-terminal (NTD, *amino terminal domain* – domínio amino terminal) básica e a carboxi-terminal (CTD, *carboxi terminal domain* – domínio carboxi terminal) ácida. A região central é constituída de uma estrutura de ligação denominada de OB (*oligonucleotide/oligosaccharide binding fold* – Dobramento de ligação a oligonucleotídeos/oligossacarídeos). Cada uma das regiões desta proteína possui uma função específica: o NTD é responsável por interagir com os fatores eIF3 e eIF2; o domínio OB com o ribossomo, provavelmente no sítio A da subunidade 40S; e o CTD interage com a extremidade carboxi-terminal do eIF5B e também com o TC no sítio P do ribossomo (Figura 4 - Roll-Mecak *et al*., 2001; Olsen *et al*., 2003). É possível que uma segunda ligação, de menor intensidade, ocorra envolvendo as extremidades amino-terminais do eIF1A e eIF5B (Olsen *et al*., 2003).

O eIF1A atua de forma pleiotrópica na iniciação da tradução. Inicialmente, em associação ao eIF3, este fator participa da dissociação do ribossomo 80S se ligando à subunidade 40S e impedindo a re-associação ribossomal. Posteriormente, participa no recrutamento do Met-tRNAi e interage com o eIF5B. Também é válido ressaltar que o eIF1A pode se ligar tanto a mRNA quanto a rRNA de forma não específica no que concerne à seqüência (Chaudhuri *et al*., 1997; Olsen *et al*., 2003).

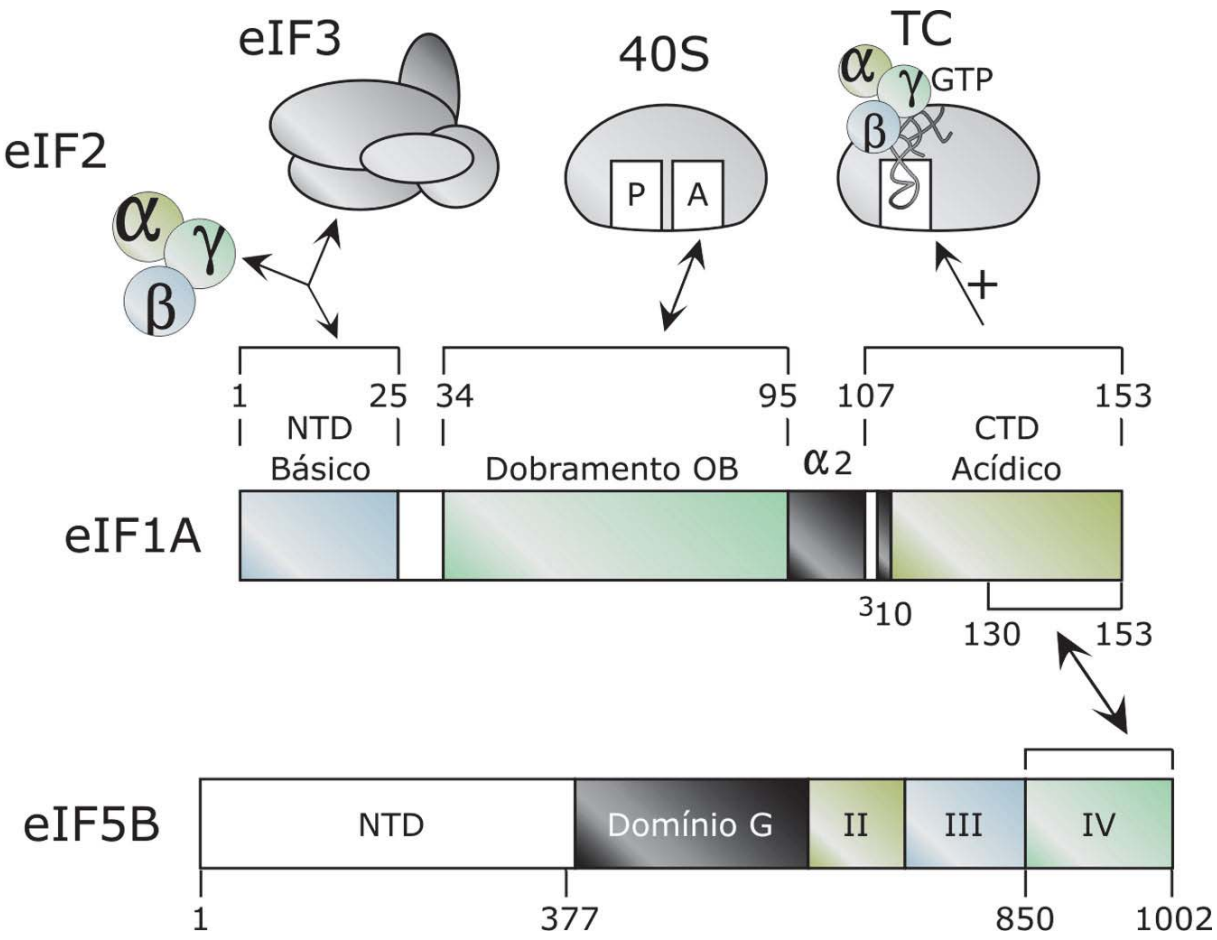


Figura 4: Desenho esquemático dos fatores eIF1A e eIF5B mostrando seus diversos domínios, amino terminal (NTD), carboxi-terminal (CTD), o local de interação entre eles e as regiões de ligação do eIF1A aos seus diversos parceiros. O eIF2 está representado pelo conjunto de suas subunidades α, β e γ. Já o eIF3 encontra-se com uma representação esquemática de seu núcleo central (ver a seguir). A interação entre o eIF1A e o ribossomo ocorre pelo sito A do último. O complexo ternário, composto pelo met-tRNAi associado a GTP e ao eIF2, está representado no sítio P do ribossomo (adaptado de Olsen *et al*., 2003).

3.3.3 eIF2 e eIF2B

O fator eIF2 é formado por três sub-unidades distintas (eIF2α, eIF2β e eIF2γ) e participa da formação do complexo ternário ao se ligar a GTP e ao met-tRNAi. O TC se liga à subunidade ribossomal 40S e após a re-associação ribossomal o eIF2 é responsável por clivar o GTP liberando a metionina. O eIF2 se desassocia do ribossomo como um complexo binário

carreando o GDP. Antes do eIF2 ser incorporado a outro ciclo de iniciação da tradução, o GDP ligado é trocado por um GTP pela ação de um fator heteropentamérico, o eIF2B (Hershey e Merrick, 2000). Devido a várias particularidades do eIF2 e eIF2B, eles não serão avaliados por este trabalho.

3.3.4 eIF3

Um outro complexo importante requerido para o início da tradução é o eIF3. Esse complexo promove a dissociação do ribossomo 80S ligando-se à subunidade ribossomal 40S. O complexo protéico eIF3 consiste de pelo menos 11 subunidades em mamíferos e possui um alto peso molecular. Em *S. cerevisiae* foram identificadas cinco subunidades deste complexo que são essenciais para o crescimento da levedura formando um núcleo central para o fator (Hershey e Merrick, 2000). As diversas subunidades do eIF3 foram nomeadas de forma independente em levedura, mamíferos e plantas utilizando-se o peso molecular ou o nome do gene o que levou a uma certa confusão ao se comparar os resultados dos diferentes organismos. Para facilitar as comparações entre espécies distintas, uma nomenclatura unificada foi proposta associando letras ao nome de cada subunidade (Browning *et al*., 2001). Dentre as suas funções estão: dissociar o ribossomo 80S; estabilizar a ligação do complexo ternário eIF2-GTP-Met-tRNA à subunidade 40S; estimular a ligação do mRNA à subunidade 40S por estar envolvido com ligações ao eIF4F. Um resumo das subunidades do eIF3, incluindo os seus motivos consensuais e o nome dos seus respectivos genes, pode ser visto na tabela 2.

Tabela 2: Resumo das subunidades do fator eIF3 (adaptado de Browning *et al*., 2001).

			Humano	Trigo	*A. thaliana*	*S. cerevisiae*
Nome	Motivo consenso	Nome do gene	PM	PM	PM	PM
eIF3a	PCI	*TIF32/RPG1*	170	116	114	110
eIF3b	RRM	*PRT1*	116	83	82	90
eIF3c	PCI	*NIP1*	110	107	105	93
eIF3d	Ausente		66	87	66	Ausente
eIF3e	PCI	(*INT6*)	48	45	51	Ausente
eIF3f	MPN		47	34	32	Ausente
eIF3g	RBD, *Zn finger*	*TIF35*	44	36	33	33
eIF3h	MPN		40	41b	38	Ausente
eIF3i	Repetições WD	*TIF34*(*TRIP1*)	36	41a	36	39
eIF3j	Ausente	*HCR1*	35	Ausente	Ausente	30
eIF3k	Ausente		28	28	25	Ausente
eIF3l	Ausente		Ausente	56	60	Ausente

O estudo das funções de cada subunidade do eIF3 revelou que ao menos três subunidades tem capacidade de se ligar a mRNA (eIF3a, eIF3d e eIF3g). Também foram mapeadas diversas interações entre determinadas subunidades e outros fatores de iniciação. Dentre as interações conhecidas temos: eIF3c ligando ao eIF1 e eIF5; eIF3g se ligando ao eIF4B. O eIF3 também interage com o complexo eIF4F, através da sua subunidade eIF4G, mas aparentemente, ao menos em levedura, esta ligação ocorre de forma indireta (Hershey e Merrick, 2000; Valasek *et al*., 2003; Kapp e Lorsch, 2004). Algumas subunidades do eIF3 compartilham domínios encontrados em grandes complexos protéicos, mais especificamente o PCI (*proteasome/COP9/Int6*) que pode ter um papel na montagem de complexos e o MPN (*Mpr1-Pad1-N terminus*) que aparentemente possui função catalítica. O primeiro é encontrado nas subunidades eIF3a, eIF3c e eIF3e, já o segundo está presente em eIF3f e eIF3h. Desta forma, sugere-se que o eIF3 possa servir como um ancoradouro para outras proteínas (Hershey e Merrick, 2000; Kim *et al*., 2004).

3.3.5 eIF4F

O fator eIF4F é um dos responsáveis pela associação entre o mRNA e o ribossomo e é formado pelas subunidades eIF4A, eIF4E, eIF4G (Haghighat *et al*., 1995).

O eIF4E é o responsável por reconhecer especificamente a extremidade 5' do mRNA, mais especificamente o nucleotídeo cap. Estruturalmente é constituído por um único domínio capaz de reconhecer o cap através do empacotamento do anel da guanosina por dois de seus triptofanos. Esta interação é conhecida como π-π *stacking* e serve de ancoradouro para o eIF4F sendo, desta forma, essencial para a tradução cap-dependente. Possui, também, um sítio de ligação ao eIF4G na face oposta ao sítio de ligação ao cap. Em mamíferos o eIF4E é alvo de regulação por uma classe de proteínas chamadas 4EBPs (4E *Binding Protein*) que competem com o eIF4G pelo sítio de ligação. A fosforilação do eIF4E inibe a ligação do 4EBP e favorece a ligação do eIF4G (Marcotrigiano *et al*., 1997; Gingras *et al*., 1999; Marcotrigiano *et al*., 1999).

O fator eIF4A apresenta atividade ATPase dependente de RNA e é uma helicase bidirecional ATP-dependente. Esta proteína é o protótipo da família protéica DEAD-*box*. Esse fator dissocia formações secundárias que poderiam dificultar o deslocamento do ribossomo ao longo do mRNA até o códon de iniciação da tradução (Gingras *et al.* 1999; Hershey e Merrick, 2000). Foram descritas em mamíferos três isoformas distintas do fator eIF4A nomeadas de eIF4AI, eIF4AII e eIF4AIII. As duas primeiras possuem uma similaridade sequencial de mais de 95% e provavelmente possuem função similar na iniciação da tradução uma vez que ambas conseguem reconstituir a subunidade do eIF4F (Nielsen e Trachsel, 1988; Yoder-Hill *et al*., 1993). Já o eIF4AIII é mais distante das outras duas proteínas, com uma similaridade sequencial de 80% com o eIF4A1, se localiza no núcleo e evidências o colocam como integrante do EJC (*Exon*

Junction Complex). Em mamíferos o eIF4AIII é essencial para o NMD (*Non Sense Mediated Decay*) e não participa na iniciação da tradução (Holzmann *et al*., 2000, Ferraiuolo *et al*., 2004; Chan *et al*., 2004; Palacios *et al*., 2004; Shibuya *et al*., 2004).

Estudos estruturais e bioquímicos do eIF4A demonstraram que esse fator possui nove domínios conservados em outras DNA e RNA helicases como ilustrado na figura 5. Esses domínios conservados são responsáveis pelas funções de ligação de ATP, hidrólise de ATP em ADP, ligação a RNA e atividade helicase (Tanner e Linder, 2001; Tanner *et al*., 2003).

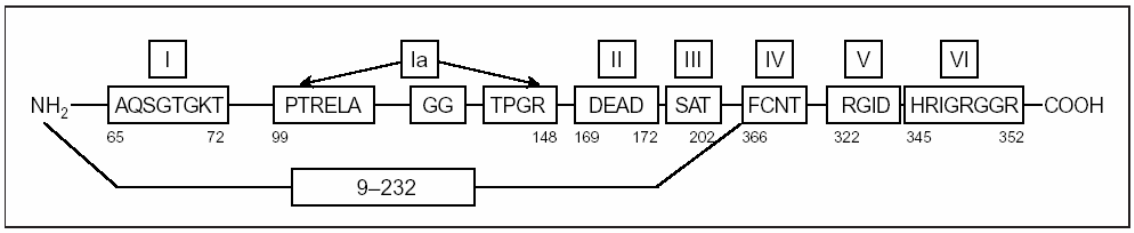


Figura 5: Representação esquemática dos motivos conservados em seqüências de RNA helicases da família DEAD e a localização dos mesmos no eIF4A (adaptado de Lorsch e Herschlag, 1998).

O terceiro integrante do complexo eIF4F é a proteína eIF4G. Este fator apresenta duas isoformas em mamíferos (eIF4G1 e eIF4G2 de 171 kDa e 176 kDa respectivamente), duas em *Saccharomyces cerevisiae* e duas em plantas (Gingras *et al*., 1999). Esta proteína possui sítios de ligação para os outros integrantes do complexo eIF4F (eIF4A e eIF4E), desempenhando o papel de montagem deste complexo (Lewin, 1994). Além disso, o eIF4G também possui domínios de ligação para outras proteínas e fatores de iniciação, como é o caso do eIF3, PABP (*Poli-A Binding Protein*) e Mnk1 (*MAP-kinase-interacting kinase-1*). O eIF4G também consegue interagir diretamente com o mRNA. Dessa forma, o eIF4G possui um papel chave no processo de ligação entre os vários complexos de iniciação da tradução (Gingras *et al*., 1999). Um esquema da estrutura primária deste fator, evidenciando os seus domínios conservados, pode ser visto na figura 6. O domínio de ligação ao eIF4A

está no Pfam (*Protein families*) e no CDD (*Conserved Domain Database*), como um motivo denominado MIF4G, termo que foi utilizado neste trabalho para fins de normatização.

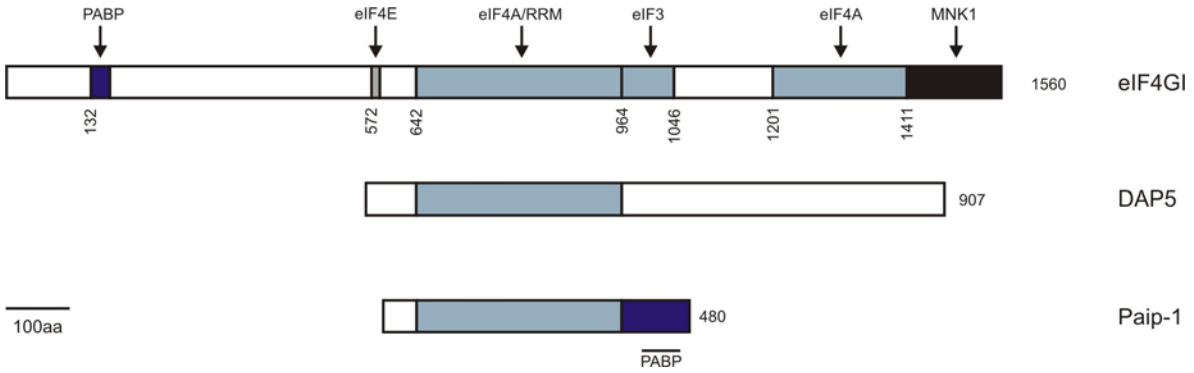


Figura 6: Desenho esquemático do fator eIF4GI de humano. As caixas coloridas representam os diversos domínios presentes nessas proteínas. No domínio de ligação ao eIF4E a região consensual é composta por YXXXXLΦ, onde Φ usualmente é uma leucina, mas pode também ser uma metionina ou uma fenilalanina (adaptado de Gingras *et al*., 1999).

Não existe consenso sobre a ordem de eventos que leva à formação do complexo eIF4F. Uma das hipóteses é a de que esse complexo seja formado antes da associação entre o eIF4E e o cap, uma vez que essa interação é mais estável na presença do fator eIF4G (Haghighat e Sonenberg, 1997). Outro modelo foi proposto, no qual o fator eIF4E ligar-se-ia ao mRNA e em seguida ao fator eIF4G, já associado à subunidade ribossomal 43S (Joshi *et al*., 1994).

Dentre as proteínas que interagem com o fator eIF4G destaca-se a PABP. Essa proteína é a responsável por reconhecer a cauda poli-A do mRNA e é essencial para o crescimento de levedura. É responsável, também, pela circularização do mRNA ao interagir diretamente com o fator eIF4G (Gingras *et al*., 1999). Essa interação ocorre através do domínio de ligação à PABP que os fatores eIF4G de mamífero, de *S. cerevisiae* e de plantas possuem na região amino-terminal. Entretanto, não há similaridade sequencial entre os domínios presentes em mamíferos

e de *S. cerevisiae* (Tarun *et al*., 1996; Le *et al*., 1997; Imataka *et al*., 1998).

3.3.6 eIF4B e eIF4H

A ação helicase do fator eIF4A é altamente estimulada pelo fator eIF4B (não pertencente ao complexo eIF4F). Este último apresenta função apenas na forma dimérica sendo que o domínio DRYG é o responsável por tal dimerização (Hershey e Merrick, 2000). Um fator relacionado com o eIF4B é o eIF4H. Estas proteínas apresentam homologia na sua região amino-terminal com uma identidade seqüencial de 39%. Todavia, o eIF4H diverge do eIF4B por não possuir o domínio DRYG, desta forma deve apresentar sua função na forma de monômero (Richter-Cook, *et al*. 1998; Hershey e Merrick, 2000).

3.3.7 eIF5

O eIF5 é um importante fator de iniciação da tradução. Apresenta um papel de hidrólise de GTP e também participa como um núcleo protéico para a formação do complexo de pré-iniciação ribossomal. Sua atividade GTPásica, que cliva o GTP ligado ao eIF2 levando ao desligamento dos demais fatores na fase final da iniciação da tradução, foi atribuída à sua extremidade amino-terminal. Ele ainda interage em sua extremidade carboxi-terminal com o eIF1, o eIF3c, o eIF4G e também com a subunidade beta do eIF2 (Hershey e Merrick, 2000; Singh *et al*, 2004). Próximo ao final da extremidade carboxi-terminal do eIF5 existe um motivo designado de “caixas de aminoácidos aromáticos/acídicos” (AA-boxes). Essas caixas também são encontradas no carboxi-terminal do eIF2Bε (subunidade catalítica do eIF2B) e no eIF4G de mamíferos (Singh *et al*., 2004).

3.3.8 eIF5A

O eIF5A foi nomeado desta forma pois os primeiros estudos desta proteína demonstraram seu efeito estimulatório sobre a iniciação da biossintese protéica em sistemas de células-livres. No entanto, a total depleção do eIF5A em leveduras não causou maiores modificações nas taxas de tradução. Portanto, argumenta-se que o fator eIF5A não seja um fator de iniciação da tradução propriamente dito, e teria, desta forma, um papel complementar no processo. Existem evidências que sugerem que o eIF5A participe da tradução de grupos específicos de mRNAs envolvidos na progressão do ciclo celular (transição G1/S) (Kim *et al*., 1998; Valentini *et al*., 2002; Li *et al*., 2004). Outra peculiaridade do eIF5A é a presença de uma lisina que sofre alterações pós-traducionais sendo modificada para o aminoácido pouco usual hipusina, uma das principais assinaturas desta proteína (Li *et al*., 2004).

3.3.9 eIF5B

O fator eIF5B, homólogo ao fator procarioto IF2, é bastante conservado ao longo da evolução e tem como função auxiliar na re-associação ribossomal, apresentando atividade GTPásica ribossomo dependente. Em leveduras este fator não é necessário para a viabilidade celular, mas sua deleção acarreta num fenótipo de crescimento celular extremamente lento (Hershey e Merrick, 2000; Pestova *et al*, 2000). A região central do eIF5B abriga o domínio G, responsável pela ligação a GTP e a região carboxi-terminal abriga três domínios bem conservados. São eles: domínio II, III e IV. Esses quatro domínios estão conservados em todos os reinos. Existe uma certa controvérsia sobre qual domínio é responsável pela ligação ao eIF1A onde aparentemente em eubactérias é

o domínio II e em eucariotos o domínio IV (figura 4). Além disso, os eucariotos e bactérias não-termófilas possuem uma extremidade amino-terminal de carga positiva (Roll-Mecak *et al*., 2000).

3.3.10 eIF6

O eIF6 é uma proteína de 25kDa, essencial em levedura. Possui cinco cópias quase idênticas de um domínio α/β com aproximadamente 45 resíduos cada. Esses domínios estão dispostos ao longo de um eixo de pseudo-simetria de ordem cinco (figura 7 - Groft *et al*., 2000). Ele está relacionado com a dissociação do ribossomo 80S e já foi demonstrado *in vivo* que sua depleção impede a biogênese da subunidade 60S. É teorizado que este fator se ligue à subunidade 60S e impeça a re-associação deste com a subunidade 40S. Desta forma, seria um fator de iniciação da tradução, mas este papel não está claro e sua classificação como fator de iniciação canônico também (Groft *et al*., 2000; Ceci *et al*., 2003).

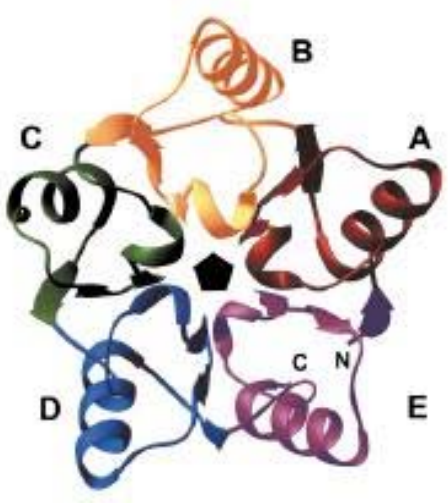


Figura 7: Estrutura tridimensional do eIF6. O pentágono no centro indica o eixo de pseudo-simetria de ordem cinco. Cada domínio quasi-idêntico está assinalado pelas letras maiúsculas A, B, C, D e E. As extremidades amino e carboxi-terminal também estão indicadas (adaptado de Groft *et al*., 2000).

3.4 Iniciação da Tradução em Tripanossomatídeos

A análise da iniciação da tradução nos tripanossomatídeos ainda se encontra num estado inicial. Estudos preliminares demonstraram a presença de pelo menos um homólogo funcional à PABP (Bates *et al*., 2000) e pelo menos um ao fator eIF4A (Skeiky *et al*., 1998) em tripanossomatídeos. Com a conclusão do projeto genoma da *L. major* foi identificado computacionalmente dois candidatos a homólogo ao eIF4A (*Lm*EIF4A1 e *Lm*EIF4A2), quatro ao eIF4E (*Lm*EIF4E1, *Lm*EIF4E2, *Lm*EIF4E3 e *Lm*EIF4E4) e cinco ao eIF4G (*Lm*EIF4G1, *Lm*EIF4G2, *Lm*EIF4G3, *Lm*EIF4G4 e LmEIF4G5) e verificou-se a conservação dessa multiplicidade de homólogos no genoma de *T. brucei* (Dhalia *et al*., 2005).

Ensaios bioquímicos demonstraram que ao menos alguns destes genes são expressos em quantidades concordantes com as observadas em leveduras e metazoários, embora estes estudos só tenham analisado a forma promastigota de *L. major*. Esses ensaios também sugerem que alguns desses candidatos podem interagir *in vitro* reconstituindo parcialmente o complexo eIF4F. Ao menos um dos homólogos ao eIF4E consegue se ligar ao cap de mamíferos. Estas evidências sugerem a participação de algumas dessas proteínas na iniciação da tradução dos tripanossomatídeos (Dhalia *et al*., 2005). Entretanto, sobre a existência dos demais fatores de iniciação descritos em outros eucariotos tem-se apenas as anotações automáticas geradas pelo projeto genoma (Ivens *et al*., 2005). Este projeto procurou confirmar a anotação automática realizada pelos bancos de dados que hospedam o genoma dos tripanossomatídeos buscando identificar artefatos de anotação e iniciar o mapeamento e estudo de domínios e mutações de interesse no conjunto dos fatores de iniciação da tradução.

4- Bibliografia

Alonso G, Guevara P and Ramirez JL (1992) *Trypanosomatidae* codon usage and GC distribution. Mem Inst Osv Cruz 87:517-523.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-10.

Asano K, Clayton J, Shalev A and Hinnebusch AG (2000) A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. Genes Dev 14:2534-46.

Barton GJ (2005) Creation and Analysis of Protein Multiple Sequence Alignments. In Baxevanis AD and Oullete BFF (eds) Bioinformatics. A practical guide to the analysis of genes and proteins, Wiley-Interscience, New Jersey, pp. 326-340.

Bates EJ, Knuepfer E and Smith DF (2000) Poly(A)-binding protein I of Leishmania: functional analysis and localisation in trypanosomatid parasites. Nucleic Acids Res 28:1211-20.

Baxevanis AD (2005) Assesing Pairwise Sequence Similarity: BLAST and FASTA. In Baxevanis AD and Oullete BFF (eds) Bioinformatics. A practical guide to the analysis of genes and proteins, Wiley-Interscience, New Jersey, pp. 296-324.

Benne R (1990) RNA editing in trypanosomes: is there a message? Trends Genet 177-181.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA and Wheeler DL (2000) GenBank. Nucleic Acids Res 28:15-8.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM (2005) The genome of the African trypanosome Trypanosoma brucei. Science 309:416-22.

Borst P, van der Ploeg M, van Hoek JFM, Tas J and James J (1982) On the DNA content of trypanosomes. Mol Biochem Parasitol 6:13.

Borst P, Rudenko G, Taylor MC, Blundell PA, van Leeuwen F, Bitter W, Cross M and McCulloch R (1996) Antigenic variation in Trypanosomes. Arch Med Res 27:379-388.

Brinkman FSL (2005) Phylogenetic Analysis. In Baxevanis AD and Oullette BFF (eds) Bioinformatics. A practical guide to the analysis of genes and proteins, Wiley-Interscience, New Jersey, pp. 366-392.

Browning KS, Gallie DR, Hershey JW, Hinnebusch AG, Maitra U, Merrick WC and Norbury C (2001) Unified nomenclature for the subunits of eukaryotic initiation factor 3. Trends Biochem Sci 26:284

Ceci M, Gaviraghi C, Gorrini C, Sala LA, Offenhauser N, Marchisio PC and Biffo S (2003) Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. Nature 426:579-84.

Chan CC, Dostie J, Diem MD, Feng W, Mann M, Rappsilber J and Dreyfuss G (2004) eIF4A3 is a novel component of the exon junction complex. RNA 10, 200-209.

Chaudhuri J, Si K and Maitra U (1997) Function of eukaryotic translation initiation factor 1A (eIF1A) (formerly called eIF-4C) in initiation of protein synthesis. J Biol Chem 272:7883-91.

Clayton CE (2002) Life without transcriptional control? From fly to man and back again. EMBO J 21:1881-88.

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881-90.

de Lana M and Tafuri WL (2000) *Trypanosoma cruzi* e Doença de Chagas. In Neves PD (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 73-96

Dhalia R, Reis CRS, Freire ER, Rocha PO, Katz R, Muniz JRC, Standart N and de Melo Neto OP (2005) Translation Initiation in *Leishmania major*: Characterisation of Multiple eIF4F Subunit Homologues. Mol Biochem Parasitol 140:23-41.

Desjeux P (2004) Leishmaniasis: current situation and perspectives. Comp Imun Microbiol Infec Dis 27:305-318.

El-Sayed NM, Hegde P, Quackenbush J, Melville SE and Donelson JE (2000). The African trypanosome genome. Int J Parasitol 30:329-345.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD and Andersson B. (2005a) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309:409-15.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N (2005b) Comparative genomics of trypanosomatid parasitic protozoa. Science. 309(5733):404-9

Ferraiuolo MA, Lee CS, Ler LW, Hsu JL, Costa-Mattioli M, Luo MJ, Reed R and Sonenberg N (2004) A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. Proc Natl Acad Sci U S A 101, 4118-4123.

Fletcher CM, Pestova TV, Hellen CU and Wagner G (1999) Structure and interactions of the translation initiation factor eIF1. EMBO J 18:2631-7.

Forster MJ (2002) Molecular modelling in structural biology. Micron 33:365-84.

Genaro O (2000) Leishmaniose Tegumentar Americana; Leishmaniose Tegumentar do Velho Mundo; Leishmaniose Visceral Americana. In Neves PD (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 36-53; 54-55; 56-72

Gibas C and Jambeck P (2001) Developing Bioinformatics Computer Skills. O'Reilly, Sebastopol, 427 pp.

Gingras AC, Raught B and Sonenberg N (1999) eIF4F Initiation Factors: Effectors of mRNA Recruitment to ribosomes and regulators of translation. Annu Rev Biochem 68:913-963.

Grisard EC and Steindel M (2000) Trypanosoma (Herpetosoma) rangeli. In Neves PD (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 97-100

Groft CM, Beckmann R, Sali A and Burley SK (2000) Crystal structures of ribosome anti-association factor IF6. Nat Struct Biol 7:1156-64.

Haghighat A, Mader S, Pause A and Sonenberg N (1995) Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. EMBO J 14, 5701-5709.

Haghighat A and Sonenberg N (1997) eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. J Biol Chem 272:21677-80..

Henikoff S and Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89:10915-9.

Hershey JWB and Merrick WC (2000) Pathway and mechanism of initiation of protein synthesis. In: Sonenberg N, *et al*. (eds) Translational control of gene expression. Cold Spring Harbor Laboratory Press, New York pp 33-88.

Higgins DG and Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237-44.

Holzmann K, Gerner C, Poltl A, Schafer R, Obrist P, Ensinger C, Grimm R and Sauermann G (2000) A human common nuclear matrix protein homologous to eukaryotic translation initiation factor 4A. Biochem Biophys Res Commun 267, 339-344.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra

M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B and Myler PJ (2005) The genome of the kinetoplastid parasite, *Leishmania major*. Science 309:436-42.

Johnson PJ, Kooter JM and Borst P (1987) Inactivation of transcription by UV *T. brucei* provides evidence for multicistronic transcription unit including VSG gene. Cell 51:273-281.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195-202.

Joshi B, Yan R and Rhoads RE (1994) In vitro synthesis of human protein synthesis initiation factor 4 gamma and its localization on 43 and 48 S initiation complexes. J Biol Chem 269:2048-55.

Kapp LD and Lorsch JR (2004) The molecular mechanics of eukaryotic translation. Annu Rev Biochem 73:657-704.

Kim KK, Hung LW, Yokota H, Kim R and Kim SH (1998) Crystal structures of eukaryotic translation initiation factor 5A from Methanococcus jannaschii at 1.8 A resolution. Proc Natl Acad Sci U S A 95:10419-24.

Kim TH, Kim BH, Yahalom A, Chamovitz DA and von Arnim AG (2004) Translational regulation via 5' mRNA leader sequences revealed by mutational analysis of the Arabidopsis translation initiation factor subunit eIF3h. Plant Cell 16:3341-56.

Kyrpides NC and Woese CR (1998) Universally conserved translation initiation factors. Proc Natl Acad Sci U S A 95:224-8.

Lewin B (1994) Genes V. Oxford University Press and Cell Press. 261-276.

Li AL, Li HY, Jin BF, Ye QN, Zhou T, Yu XD, Pan X, Man JH, He K, Yu M, Hu MR, Wang J, Yang SC, Shen BF and Zhang XM (2004) A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. J Biol Chem 279:49251-8.

Liang X-H, Haritan A, Uliel S and Michaeli S (2003) Trans and Cis Splicing in Tripanosomatids: Mechanisms, Factors, and Regulation. Euk Cell 2:830-840.

Lorsch JR and Herschlag D (1998) The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. Biochemistry 37:2194-206.

Maag D, Fekete CA, Gryczynski Z and Lorsch JR (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. Mol Cell 17:265-75.

Marcotrigiano J, Gingras AC, Sonenberg N and Burley SK (1997) Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell 89:951-61.

Marcotrigiano J, Gingras AC, Sonenberg N and Burley SK (1999) Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. Mol Cell 3:707-16.

McGuffin LJ and Jones DT (2002) Targeting novel folds for structural genomics. Proteins 48:44-52.

Michalick MSM (2000) Gênero Leishmania. In Neves DP (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 31-35

Muhich ML and Boothroyd JC (1988) Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Mol Cell Biol 8:3837-46.

Nielsen PJ and Trachsel H (1988) The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed. EMBO J 7:2097-2105.

Olsen DS, Savner EM, Mathew A, Zhang F, Krishnamoorthy T, Phan L and Hinnebusch AG (2003) Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. EMBO J 22:193-204.

Palacios IM, Gatfield D, St JD, and Izaurralde E (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. Nature 427:753-757.

Pestova TV, Lomakin IB, Lee JH, Choi SK, Dever TE and Hellen CU. (2000) The joining of ribosomal subunits in eukaryotes requires eIF5B. Nature 403:332-5.

Reed SG (1998) LeIF: a recombinant Leishmania protein that induces an IL-12-mediated Th1 cytokine profile. J Immunol 161:6171-9.

Richter-Cook NJ Dever TE Hensold JO and Merrick WC (1998) Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H. J Biol Chem 273:7579-87.

Roditi I, Furger A, Ruepp S, Schurch N and Butikofer P (1998) Unravelling the procyclin coat of Trypanosoma brucei. Mol Biochem Parasitol 91:117-30.

Roll-Mecak A, Cao C, Dever TE and Burley SK (2000) X-Ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. Cell 103:781-92.

Roll-Mecak A, Shin BS, Dever TE and Burley SK (2001) Engaging the ribosome: universal IFs of translation. Trends Biochem Sci 26:705-9.

Russell RB and Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins 14:309-23.

Shibuya T, Tange TO, Sonenberg N and Moore MJ (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. Nat Struct Mol Biol 11, 346-351.

Simons KT, Strauss C and Baker D (2001) Prospects for ab initio protein structural genomics. J Mol Biol 306:1191-9.

Singh CR, Yamamoto Y and Asano K (2004) Physical association of eukaryotic initiation factor (eIF) 5 carboxyl-terminal domain with the lysine-rich eIF2beta segment strongly enhances its binding to eIF3. J Biol Chem 279:49644-55.

Siqueira AM (2000) Subfilo Mastigophora. In Neves DP (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 27-30

Skeiky YA, Kennedy M, Kaufman D, Borges MM, Guderian JA, Scholler JK, Ovendale PJ, Picha KS, Morrissey PJ, Grabstein KH, Campos-Neto A and

Stevens JR, Noyes HA, Schofield CJ and Gibson W (2001) The Molecular evolution of Trypanosomatidae. Adv Parasit 48:1-55

Tanner NK and Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251-62. Review.

Tanner NK, Cordin O, Banroques J, Doere M and Linder P (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Mol Cell 11:127-38.

Valasek L, Mathew AA, Shin BS, Nielsen KH, Szamecz B and Hinnebusch AG. (2003) The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. Genes Dev 17:786-99.

Valentini SR, Casolari JM, Oliveira CC, Silver PA and McBride AE (2002) Genetic interactions of yeast eukaryotic translation initiation factor 5A (eIF5A) reveal connections to poly(A)-binding protein and protein kinase C signaling. Genetics 160:393-405.

Williams P (2000) Psychodidae. In Neves PD (ed) Parasitologia Humana, 10ª ed, Editora Atheneu, São Paulo, pp. 311-319

Wishart D (2005) Protein Structure Prediction and Analysis. In Baxevanis AD and Oullete BFF (eds) Bioinformatics. A practical guide to the analysis of genes and proteins, Wiley-Interscience, New Jersey, pp. 224-251.

Yoder-Hill J, Pause A, Sonenberg N and Merrick WC (1993) The p46 subunit of eukaryotic initiation factor (eIF)-4F exchanges with eIF-4A. J Biol Chem 268, 5566-5573.

Yoon HJ and Donahue TF (1992) The sui1 suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. Mol Cell Biol 12:248-60.

5- **Manuscrito de Artigo Científico**

Identificação e Caracterização *in silico* de Candidatos a Fatores de Iniciação da Tradução em Tripanossomatídeos

Manuscrito a ser encaminhado à revista
Eukaryotic Cell
ISSN: 1535-9778
(Washington, Estados Unidos da América)

Identificação e Caracterização *in silico* de Candidatos a Fatores de Iniciação da Tradução em Tripanossomatídeos

Rodolfo Katz.[1] e Osvaldo Pompilio de Melo Neto[1]*

[1]

*Correspondência: opmn@cpqam.fiocruz.br

Palavras-chave: eIF; Iniciação da Tradução; *Leishmania major*, *Trypanosoma brucei*; *T. cruzi*.

Resumo

A síntese protéica é um processo básico e essencial para a sobrevivência dos seres vivos. Um dos pontos chave deste processo é a etapa de iniciação da tradução que é regulada pela ação de ao menos doze fatores protéicos chamados eIFs (*Eukaryotic Initiation Factor*) perfazendo em torno de 30 polipeptídios em mamíferos. Os tripanossomatídeos, protozoários patogênicos de interesse médico e veterinário, apresentam características celulares próprias como a regulação da sua expressão gênica que ocorre a nível pós-transcricional. Nesse contexto a síntese de proteínas é um alvo em potencial para mecanismos de regulação, entretanto pouco se sabe sobre esse processo nos tripanossomatídeos. Em estudos prévios, foi iniciado nestes parasitas o estudo do fator eIF4F e observou-se a existência de múltiplos homólogos para cada uma de suas três subunidades. Neste trabalho utilizou-se ferramentas de bioinformática para identificar e caracterizar homólogos aos demais eIFs em *Leishmania major*, *Trypanosoma brucei* e *T. cruzi*. Foram identificados homólogos dos fatores eIF1, eIF1A, eIF5, eIF5A, eIF5B, eIF6 e a sete subunidades do complexo eIF3 (b, c, d, e, f, i, k). Ao contrário do observado para as subunidades do eIF4F, e com a exceção da subunidade eIF3b (com dois homólogos identificados), um único homólogo foi identificado para cada fator analisado. A análise das seqüências de aminoácidos mostrou que existe uma variação no grau de conservação destes homólogos quando comparados com outros eucariotos (de 22% de identidade para o eIF3k até 58% para o eIF6). Em alguns casos foi possível mapear mutações exclusivas dos tripanossomatídeos. Os resultados obtidos indicam que boa parte da iniciação da síntese protéica é conservada entre tripanossomatídeos e demais eucariotos. Todavia, diferenças significativas parecem ocorrer e merecem ser melhor estudadas.

Introdução

Os tripanosomatídeos são protozoários flagelados que compreendem parasitas do gênero *Leishmania* e *Trypanosoma* de grande importância médica e veterinária. Nestes gêneros estão contidos agentes causadores de várias patologias, dentre elas a doença de Chagas, doença do Sono e as diversas Leishmanioses que atingem principalmente países pobres da América Latina, África e sudeste da Ásia. Por terem divergido precocemente da linhagem que gerou os demais eucariotos, os tripanosomatídeos apresentam diversas características próprias como uma mitocôndria única modificada (cinetoplasto); editoramento do DNA do cinetoplasto; transcrição dos mRNAs em unidades policistrônicas; e processamento em *trans* da unidade policistrônica do mRNA (Borst *et al*., 1982; Johnson *et al*., 1987; Muhich e Boothroyd 1988; Benne R 1990; Alonso *et al*., 1992; Siqueira 2000). Outra característica marcante dos tripanossomatídeos é a aparente ausência de controle transcricional na expressão gênica, sugerindo que a regulação desse importante processo celular ocorra pós-transcricionalmente (Clayton *et al*., 2002). Com base nessa premissa, a iniciação da síntese de proteínas (ou tradução), sujeita a diferentes mecanismos de regulação em outros eucariotos, torna-se um interessante processo a ser estudado nos tripanossomatídeos.

Resumidamente, a iniciação da tradução começa com a dissociação do ribossomo 80S nas subunidades 40S e 60S. Em seguida, a subunidade ribossomal 40S (subunidade menor) recruta o tRNA iniciador contendo o aminoácido metionina formando o complexo de pré-iniciação 43S. A associação do complexo 43S ao mRNA é mediada por um conjunto de fatores que reconhecem o cap (guanosina metilada adicionada à extremidade 5’ do mRNA durante a maturação do transcrito primário). Em seguida, o complexo 43S vasculha a extremidade 5’ do mRNA em busca do códon AUG de iniciação num processo denominado de *scanning*. Ao chegar no códon correto, ocorre a re-associação das subunidades ribossomais e o início da síntese protéica propriamente dita (Hershey e

Merrick, 2000). Em todas as etapas citadas ocorre a atuação de fatores protéicos num processo complexo e regulado. Assim, até o momento, foram identificados pelo menos doze fatores de iniciação da tradução denominados de eIF (*eukaryotic Initiation Factor*). Os fatores são: eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4F, eIF4B, eIF4H, eIF5, eIF5A, eIF5B e eIF6, cada um atua de forma específica e essencial ao processo (Kapp e Lorsch, 2004).

Pouco se conhece a respeito da iniciação da tradução em tripanossomátideos. Inicialmente foi descrito um homólogo ao eIF4A, RNA helicase componente do eIF4F, em *L. brazilienses* como uma molécula antigênica, mas seu papel na tradução não foi avaliado (Skeiky *et al*., 1998). Mais recentemente, análises de bioinformática seguidas de ensaios bioquímicos permitiram a caracterização preliminar, em *L. major*, das outras subunidades de eIF4F. Assim, dois homólogos ao eIF4A, quatro homólogos ao eIF4E (a proteína de ligação ao cap) e cinco homólogos ao eIF4G (proteína que estrutura o eIF4F) foram identificados. Ensaios bioquímicos preliminares sugerem que alguns desses candidatos possam interagir *in vitro* reconstituindo o complexo eIF4F, embora não se entende, ainda, o motivo responsável pela ocorrência dos vários homólogos (Dhalia *et al*., 2005). No entanto, além da anotação automática realizada pelos servidores que hospedam os genomas destes tripanosomatídeos, nada se sabe sobre os demais fatores de iniciação da tradução. Em vista disso, procura-se aqui aplicar as ferramentas de bioinformática aos demais fatores de iniciação da tradução com o intuito de aumentar o entendimento sobre o metabolismo desses parasitas. Espera-se assim facilitar a aplicação de novas abordagens para o estudo dos seus processos biológicos básicos e, no futuro, o desenvolvimento de ferramentas de controle das suas diferentes patologias.

Metodologia

Busca por homólogos caracterizados de organismos específicos:

Inicialmente, foram obtidas as seqüências dos homólogos dos fatores de iniciação da tradução de *Homo sapiens* no banco de dados do NCBI (*National Center for Biotechnology Information* Centro - Nacional de Infomação Biotecnológica; http://www.ncbi.nlm.nih.gov). Foi dada preferência às seqüências de proteína devido à saturação de mutações nas seqüências de DNA de organismos filogeneticamente distantes. Com o intuito de complementar as análises abordando taxons com distâncias evolutivas representativas também foram selecionadas as seqüências dos homólogos de uma planta (*Arabidopsis thaliana* ou *Triticum aestivus* conforme a disponibilidade) e uma levedura (*Saccharomyces cerevisiae*). Nessas buscas foram selecionadas as seqüências dos fatores, e/ou suas subunidades, eIF1, eIF1A, eIF3, eIF5, eIF5A, eIF5B e eIF6. Uma vez que as subunidades do eIF4F, eIF2 e do eIF2B não foram alvos deste trabalho, eles ficaram de fora desta etapa. Todos os números de acessos das proteínas utilizadas encontram-se em uma tabela anexada ao artigo.

Rastreamento por candidatos a homólogos nos genomas de tripanossomatídeos:

As seqüências dos homólogos de humano a cada um dos fatores de iniciação da tradução selecionados foram utilizadas como entrada na busca realizada no GeneDB, banco de dados do Sanger Institute que hospeda os genomas completos dos tripanossomatídeos estudados (Hertz-Fowler C *et al*., 2004), utilizando-se a ferramenta omniBLAST. Esta ferramenta possibilita a comparação tanto de seqüências de nucleotídeos como de aminoácidos e seleciona o melhor algoritmo BLAST a ser

utilizado, além de permitir a busca contra diversas bibliotecas numa única análise tendo sido selecionadas as bibliotecas de *L. major*. Dos resultados, foram escolhidos aqueles com o menor *e-value* e com um ponto de corte mínimo de 1e-05. Em seguida buscou-se por homólogos em outros tripanossomatídeos de interesse, além de ter sido efetuada uma análise comparativa entre os candidatos a homólogos dos fatores de *L. major* com os de *T. brucei* e *T. cruzi*.

Testes para validação dos candidatos a homólogos

Para confirmar se as proteínas encontradas realmente se tratavam de candidatos a homólogos aos fatores de iniciação da tradução foram realizados testes complementares. O primeiro deles foi utilizar as seqüências obtidas numa busca contra o banco de dados de proteínas do NCBI através de um BLASTP. Dessa forma, caso as proteínas utilizadas forem de fato homólogos aos fatores de tradução, os melhores resultados tendem a ser os próprios fatores. Uma segunda abordagem foi verificar sua presença na biblioteca de ESTs (*Expressed Sequence Tag*) de *L. major* do GeneDB.

Estudo das seqüências de nucleotídeos e aminoácidos

Quando necessário, as seqüências brutas de DNA dos diversos genes foram retiradas do banco de dados do GeneDB e analisadas com o auxílio do programa Artemis, fornecido pelo próprio GeneDB (http://www.sanger.ac.uk/Software/Artemis). Já as seqüências protéicas foram analisadas com o auxílio do programa Bioedit (Hall, 1999).

Alinhamentos múltiplos e análises filogenéticas

Além destas metodologias, foram criados alinhamentos múltiplos através do CLUSTAL (Higgins, 1998) para a obtenção de análises filogenéticas com o programa MEGA (Kumar, 2004). Nestas análises filogenéticas optou-se por utilizar a metodologia de *NeighborJoining* com um *bootstrap* de mil reconstruções, lacunas foram consideradas como deleções completas. Quanto ao modelo de substituições foi utilizado a correção de Poisson com taxas homólogas entre os sítios. Também foi utilizada uma ferramenta do BLAST fornecida pelo NCBI (rpsBLAST) para um estudo da arquitetura básica das diversas seqüências obtidas comparando-as com os homólogos conhecidos dos demais eucariotos. Neste estudo, a biblioteca do CDD (*Conserved Domain Database*) foi a de escolha (Marchler-Bauer *et al*., 2005).

Resultados

Foi possível, através da metodologia proposta, encontrar candidatos a homólogos aos fatores eIF1, eIF1A, eIF5, eIF5A, eIF5B, eIF6 e de determinadas subunidades do eIF3 nos genomas de *L. major*, *T. brucei* e *T.cruzi*. A validação das seqüências obtidas pelo uso da biblioteca de ESTs não se mostrou muito eficaz uma vez que este banco do GeneDB está bastante defasado em relação ao seqüenciamento genômico deste tripanossomatídeo que está concluído. Uma compilação dos resultados do estudo das seqüências dos candidatos obtidos encontra-se resumida nas Tabelas 1 e 2. Os fatores eIF4B e eIF4H não tiveram candidatos identificados por este trabalho. Uma análise mais aprofundada sobre cada candidato foi realizada e para facilitar o entendimento dos resultados de cada fator, eles serão expostos individualmente e o fator eIF3 será deixado para o final por apresentar uma multiplicidade de subunidades e domínios.

Análise do fator eIF1

O fator eIF1, também conhecido como SUI1 (*suppressor of initiator codon mutations*), juntamente com o eIF1A, participa da formação do complexo 43S-mRNA. Ambos auxiliam no processo de busca pelo códon de iniciação e asseguram a iniciação da tradução no AUG correto. Já foi visto que mutações em determinados resíduos do eIF1 resultam em iniciação da tradução no códon UUG e outras estão relacionadas com mudanças na matriz de leitura. Dessa forma, o eIF1 possui um papel crítico na iniciação da tradução (Kyrpides e Woese, 1998; Fletcher *et al*., 1999; Hershey e Merrick, 2000; Maag *et al*., 2005).

Em *L. major* foi identificado um único candidato a homólogo ao eIF1 sob o número de acesso LmjF24.1210. O fator possui uma massa estimada de 12,3 kDa e seu gene se encontra no cromossomo 24. Uma

busca no genoma de *T. brucei* e *T. cruzi*, a partir da seqüência obtida de *L. major*, encontrou os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb11.02.3595 para o primeiro e Tc00.1047053508515.20 e Tc00.1047053508641.184 para o segundo. Numa análise de seqüência primária, comparando-a com dados da estrutura do eIF1 humano, os constituintes das folhas-β e α-helices estão bem conservados (dados não mostrados). Em seguida, o LmjF24.1210 foi utilizado numa busca por domínios conservados no CDD que teve como resultado a observação da conservação do domínio SUI1/eIF1 característico deste fator com um *e-value* de 3e-10. Vale lembrar que este domínio possui um enovelamento similar a outras proteínas de ligação a RNA.

A Figura 1 mostra o alinhamento das seqüências dos homólogos de eIF1 dos três tripanosomatídeos analisados com as seqüências de homólogos humano, de planta e de levedura. De uma forma geral o alinhamento mostra que boa parte das seqüências das proteínas de tripanosomatídeos é conservada. No entanto, chama a atenção um grupo de aminoácidos que não se encontram conservados. Estes resíduos na estrutura do eIF1 de humano se encontram na superfície da proteína. É interessante notar que estudos em levedura co-relacionaram estes aminoácidos com mudanças na matriz de leitura e iniciação da tradução em códons não AUG (Fletcher *et al*., 1999). Em mamíferos esses resíduos são: D88, Q89, R90 e G112 e nos tripanossomatídeos estão mutados por N, W, S e S respectivamente.

Com o intuito de verificar se as mutações encontradas representam uma característica específica dos tripanossomatídeos ou mutações conservadas ao longo da linhagem evolutiva, um segundo alinhamento múltiplo com um número maior de organismos foi construído. Dentre os organismos inseridos para esta segunda análise encontram-se outros protista como *Plasmodium falciparum*, *Dictyostelium discoideum* e *Entamoeba histolítica*. Também foi inserida a seqüência de uma archeabactéria, a *Methanocaldococcus jannaschii*. Todos os organismos

desse novo alinhamento apresentaram os resíduos conservados com o consenso de humano com exceção do resíduo Q89 de mamífero que em *P. falciparum* e em *D. discoideum* está mutado por um K e em *M. jannaschii* por um H. Outra peculiaridade do *M. jannaschii* é a ausência de um resíduo que alinhe com o G112 de mamífero, terminando antes de seu aparecimento. Com base nesse alinhamento múltiplo uma análise filogenética foi realizada na qual pode-se observar o agrupamento dos tripanossomatídeos num ramo separado dos demais organismos eucariotos (Figura 2).

Análise do fator eIF1A

O eIF1A atua de forma pleiotrópica na iniciação da tradução e foi inicialmente descrito em isolados de células de mamíferos e em extratos de germe de trigo. Inicialmente, em associação ao eIF3, este fator participa da dissociação do ribossomo 80S se ligando à subunidade 40S e impedindo a re-associação ribossomal. O eIF1A participa também no recrutamento do Met-tRNAi, interage com o eIF5B e provavelmente ocupa um sítio do ribossomo 40S. Tem-se ainda que o eIF1A pode se ligar tanto a mRNA quanto a rRNA de forma não específica no que concerne à seqüência (Chaudhuri *et al*., 1997; Olsen *et al*., 2003).

O eIF1A é uma proteína pequena e estável que apresenta dois domínios estruturais nas suas extremidades altamente polarizadas, sendo a amino-terminal básica e a carboxi-terminal ácida conferindo cargas positivas e negativas respectivamente. Em sua região central está presente um domínio de ligação a oligonucleotídeos e oligossacarídeos, OB (*oligonucleotide-oligosaccharide binding fold*). A região amino-terminal é responsável por interagir com os fatores eIF3 e eIF2; o domínio OB interage com o ribossomo, provavelmente com o sítio A da subunidade 40S; e o domínio carboxi-terminal do eIF1A interage com a extremidade carboxi-terminal do eIF5B (Roll-Mecak *et al*., 2001; Olsen *et al*., 2003).

Em *L. major* foi encontrado um homólogo com o número de acesso LmjF16.0140 com uma massa de 18,6 kDa e situado no cromossomo 16. Utilizando-se da seqüência de *L. major*, pode-se analisar o genoma de *T. brucei* e *T. cruzi* encontrando-se os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb08.11J15.1060 para o primeiro e Tc00.1047053503945.10 e Tc00.1047053506743.4 para o segundo. Um alinhamento múltiplo, com as seqüências dos tripanossomatídeos e de outros eucariotos, foi criado e pode-se perceber que a extremidade amino-terminal e o domínio central OB apresentam-se bem conservados. Também foi possível verificar que a polaridade específica das extremidades amino e carboxi-terminal estava presente (Figura 3). O resultado do domínio central OB foi confirmado pela análise através do CDD obtendo um *e-value* significativo de 5e-17. A extremidade carboxi-terminal apresentou a menor conservação em termos de conteúdo de resíduos, inclusive um alinhamento incluindo apenas esse trecho da seqüência de *L. major* e humanos não revela nenhuma similaridade significativa, resultado concordante com o comportamento das demais taxas.

Análise do fator eIF5

O eIF5 é um importante fator de iniciação da tradução. Ele interage com o eIF2, eIF3 e eIF1 através de sua extremidade carboxi-terminal. Essa proteína possui uma atividade GTPásica que cliva o GTP ligado ao eIF2 o que leva ao desligamento dos demais fatores de iniciação da tradução. A atividade GTPásica foi atribuída à extremidade amino-terminal e foi mapeado próximo ao fim da extremidade carboxi-terminal a presença de um motivo AA-box rico em resíduos acídicos e aromáticos (Hershey e Merrick, 2000; Singh *et al*, 2004).

O estudo em *L. major* forneceu um candidato a homólogo ao eIF5. Ele se encontra no GeneDB sob o número de acesso LmjF34.0350.

Apresenta uma massa de 42,9 kDa e está localizado no cromossomo 34. A partir da seqüência de *L. major*, pode-se analisar o genoma de *T. brucei* e *T. cruzi* encontrando-se os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb10.70.4880 para o primeiro e Tc00.1047053504119.10 e Tc00.1047053504105.20 para o segundo. O uso do CDD localizou a presença na extremidade amino-terminal de um domínio eIF2B_5, que está presente nos fatores eIF2B e eIF5, com um *e-value* de 2e-24. Por outro lado, o motivo AA-boxes, chamado de eIF5C no CDD, não foi localizado por esta busca. Com base nesses resultados, foi criado um alinhamento múltiplo no qual pode se constatar que a região amino-terminal da proteína é, de fato, mais conservada. Também foi possível verificar a conservação de aminoácidos aromáticos na extremidade carboxi-terminal dos tripanossomatídeos sugerindo a presença de um motivo AA-box alternativo (Figura 4).

Análise do fator eIF5A

O fator eIF5A apesar de inicialmente classificado como fator de iniciação da tradução, quando em total depleção em leveduras não causou maiores modificações nas taxas de síntese protéica. No entanto, evidências sugerem que o eIF5A participa da tradução de grupos específicos de mRNAs envolvidos na progressão do ciclo celular (transição G1/S - Kim *et al*., 1998; Valentini *et al*., 2002). O estudo deste fator em *L. major* revelou a presença de duas cópias idênticas e em *tandem* no genoma deste parasita. Os dois genes são idênticos e estão no GeneDB com o número de acesso LmjF25.0720 e LmjF25.0730. Eles se encontram no cromossomo 25 e a proteína possui uma massa predita de 17,8 kDa. Uma vez obtida a seqüência de *L. major*, pode-se analisar o genoma de *T. brucei* e *T. cruzi* encontrando-se os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb11.03.0410 para o primeiro e Tc00.1047053506925.120 e Tc00.1047053506925.130 para o segundo. É

interessante perceber que em *T. brucei* este fator não se encontra duplicado. Estudos de alinhamento múltiplo com os homólogos caracterizados de planta, mamífero e levedura revelaram que a lisina modificada em hipusina, uma peculiaridade do eIF5A (Li *et al*., 2004), se encontra conservada e na posição 53 em *L. major*. Da mesma forma, os doze aminoácidos que rodeiam esta mutação, em sua grande maioria, são estritamente conservados nos tripanossomatídeos (Figura 5).

Análise do fator eIF5B

O fator eIF5B é bastante conservado ao longo da evolução e tem como função auxiliar na re-associação ribossomal e atividade GTPásica (Pestova *et al*, 2000). A análise do genoma de *L. major* revelou a presença de um candidato a homólogo com *e-value* de 4,2e-159 que apresenta 816 aminoácidos e uma massa predita de 92,3 kDa. Ele se encontra no cromossomo 33 e está anotado no GeneDB pelo número de acesso LmjF33.2740. O CDD mostrou que a seqüência de *L. major* possui o domínio InfB (*Translation initiation factor 2 (IF-2; GTPase) [Translation, ribosomal structure and biogenesis]*) conservado com um *e-value* de 1e-103. Com base na seqüência de *L. major*, pode-se analisar o genoma de *T. brucei* e *T. cruzi* encontrando-se os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb927.2.3780 para o primeiro e Tc00.1047053506235.10 e Tc00.1047053511111.10 para o segundo. Um primeiro resultado obtido pela análise da seqüência primária dessa proteína é a que ela possui uma extremidade amino-terminal mais curta que os homólogos dos demais eucariotos. Em seguida passou-se para o alinhamento múltiplo dos homólogos encontrados em triponossomatídeos contra os de *A. thaliana*, *H. sapiens* e *S. cerevisie*. Os homólogos dos demais eucariotos apresentam uma conformidade em relação às regiões conservadas e divergentes onde a extremidade amino-terminal é mais divergente enquanto a região central e a carboxi-terminal são mais

conservadas. Segundo as análises feitas este padrão está presente nas seqüências de tripanossomatídeos.

Numa primeira análise, a extremidade amino-terminal estaria faltando nos tripanossomatídeos, portanto foi incluída uma seqüência de *archea* no alinhamento múltiplo. Com esse alinhamento, pode se constatar que apesar de curta, a extremidade n-terminal de tripanossomatídeos não é tão curta quanto a de *archea* (resultados não mostrados). Para se certificar de que o amino-terminal estava completo, a seqüência bruta do DNA foi retirada do geneDB e analisada com o programa Ártemis. Foi possível, então confirmar que a região amino-terminal estava completa pela existência de uma série de códons de parada em fase na extremidade 5' UTR do gene (resultados não mostrados).

O passo seguinte foi identificar a localização de cada domínio nas seqüências de tripanossomatídeos. A região central do eIF5B abriga o domínio G, responsável pela ligação a GTP enquanto que a região carboxi-terminal abriga três domínios bem conservados: domínio II, III e IV. (Roll-Mecak *et al*, 2000; Olsen *et al*., 2003). Dentro do domínio G, foram mapeados quatro motivos, são eles G1, G2, G3 e G4. Dentro desses motivos, a maioria dos aminoácidos apresenta-se conservada e os poucos mutados representam mutações presentes em outras taxas (figura 6). Os quatro domínios são bem conservados e sua conservação é decrescente no sentido G, II, III e IV, mesmo padrão apresentado ao longo de todos os reinos (Figura 7).

Análise do fator eIF6

O fator eIF6 está envolvido com a dissociação do ribossomo 80S e já foi demonstrado *in vivo* que sua depleção impede a biogênese da subunidade 60S. Como é uma proteína diretamente ligada com a tradução, o eIF6 também foi estudado nesse trabalho, embora não esteja clara a sua classificação como fator de iniciação canônico (Groft *et al*.,

2000; Ceci *et al*., 2003). Foi possível identificar um candidato a homólogo no genoma de *L. major*. Trata-se de uma proteína com 27,1 kDa de massa predita, com o gene codificado no cromossomo 36. Foi anotada automaticamente no GeneDB com o número de acesso LmjF36.0890. Com esta seqüência, pode-se analisar o genoma de *T. brucei* e *T. cruzi* encontrando-se os ortólogos correspondentes anotados no GeneDB com os números de acesso Tb10.70.1770 para o primeiro e Tc00.1047053506679.70 para o segundo. A seqüência de *T. brucei* apresenta uma longa inserção na extremidade amino-terminal que não se encontra conservada em nenhuma outra taxa, incluindo os outros tripanossomatídeos estudados, sugerindo tratar-se de um artefato de anotação. Com o auxílio do CDD foi possível verificar que a estrutura de "fechamento em velcro", responsável por manter a sua conformação fechada, teve seus constituintes identificados e em sua maioria apresentam-se conservados. Os poucos resíduos alterados da estrutura de "fechamente em velcro" representam alterações encontradas em outros organismos (Figura 8).

Análise do fator eIF3

O fator eIF3 trata-se de um complexo protéico constituído de pelo menos 11 subunidades em mamífero de elevado peso molecular. Em *S. cerevisiae* foram identificadas cinco subunidades deste complexo e elas são essenciais para o crescimento da levedura. Por se tratar de um complexo protéico, este fator teve as suas diversas subunidades estudadas isoladamente. Com o intuito de facilitar a exposição dos resultados, eles serão expostos por subunidade seguindo a nomenclatura unificada do fator (Browning *et al*., 2001).

A subunidade eIF3a tem como papel se ligar ao RNA e possui o domínio PCI. O primeiro omniBLAST realizado com a seqüência de humano de eIF3a revelou uma proteína hipotética não caracterizada. No

entanto, o BLAST reverso com essa seqüência revelou *hits* com *e-values* maiores a proteínas relacionadas à citocinese. É interessante comentar, também, que não foi encontrado nessa proteína o domínio PCI integrante das subunidades eIF3a. Em seguida foi feito o omniBLAST com a seqüência de planta que também revelou um hit que foi descartado no BLAST reverso. Os *hits* obtidos com a seqüência de levedura foram os mesmos que com humano e planta. Uma vez que as seqüências encontradas para esta subunidade em *L. major* foram descartadas, não se procurou em *T. brucei* e *T. cruzi* por candidatos a homólogos.

A subunidade eIF3b apresenta o domínio PCI e o motivo RRM podendo ter relação com ligação a RNA. Em relação ao eIF3b foi inicialmente obtido um candidato a homólogo no genoma de *L. major*: o LmjF17.1290. Esta proteína apresenta conservação em *T. brucei* e *T. cruzi* anotados no GeneDB com o acesso Tb927.5.2570 para o primeiro e para o segundo foram Tc00.1047053509177.68 e Tc00.1047053511303.60. O estudo de domínios conservados desta proteína, pelo banco de dados do CDD, revelou que o LmjF17.1290 possui apenas um domínio nomeado por este banco de COG5354 caracterizado pela presença de repetições de WD. O domínio RRM amino-terminal dessa proteína aparentemente está ausente. O alinhamento múltiplo criado para este candidato a homólogo, contendo o eIF3b caracterizado de plantas, mamíferos e leveduras, revelou que os tripanossomátideos possuem uma extremidade amino-terminal mais curta e inserções e deleções ao longo da seqüência (Figura 9).

Em seguida, foi identificada uma segunda proteína interessante, catalogada no geneDB como LmjF16.0690. Ela é mais distante dos homólogos de eIF3b dos demais eucariotos que o LmjF17.1290 mas apresenta o domínio COG5354. Assim como o LmjF17.1290, apresenta ortólogos em *T. brucei* e *T. cruzi* e a seqüência de *L. major* foi utilizado para enraizar a árvore filogenética. A árvore filogenética apresenta os tripanossomatídeos num ramo bem isolado quando comparado com os demais eucariotos (Figura 10).

A subunidade eIF3c tem como função se ligar ao eIF1 e ao eIF5. O eIF3c apresentou um candidato a homólogo em *L. major*. O homólogo de *L. major* foi submetido a uma análise pelo CDD onde revelou a presença de um domínio nomeado, por este banco de dados, de eIF3c-N característico dessa subunidade. Contudo, o domínio PCI carboxi-terminal, que é uma assinatura clássica deste fator, não foi encontrado por esta abordagem. O alinhamento com a seqüência de *H. sapiens* apresentou diversas lacunas na extremidade amino-terminal. Quando alinhada com a seqüência de *A. thaliana* a extensão das lacunas foi menor, mas ainda assim significantes para perturbar a produção de alinhamentos globais. Foi possível obter ortólogos em *T. brucei* e *T. cruzi* anotados no GeneDB com os acessos Tb10.6k15.2250 e Tb10.6k15.2220 para o primeiro e para o segundo foram Tc00.1047053507611.310 e Tc00.1047053507723.130. O alinhamento múltiplo mostrou que a região mais conservada entre os tripanossomatídeos e os demais eucariotos estava situada em torno da região central da proteína. Também foi possível pelo estudo do alinhamento múltiplo observar na extremidade amino-terminal dos tripanossomatídeos uma região acídica concordante com os demais eucariotos. Finalmente, o alinhamento múltiplo permitiu a observação de diversos resíduos de glicina, assim como resíduos de aminoácidos básicos, na extremidade carboxi-terminal dos tripanossomatídeos, característica também presente nos demais eucariotos (Figura 11).

O eIF3d é uma subunidade que está ausente em leveduras, portanto não participa do núcleo central do eIF3 e também não apresenta um domínio característico. Contudo, foi encontrado um candidato a homólogo em *L. major*, um em *T. brucei* e dois em *T. cruzi*. O alinhamento múltiplo revelou a existência de diversos resíduos aromáticos extremamente conservados ao longo da seqüência, assim como uma extremidade carboxi-terminal acídica também conservada (Figura 12).

A subunidade eIF3e possui o domínio PCI, sendo a assinatura deste fator. Foi encontrado um candidato à subunidade eIF3e no genoma de *L. major* que, de acordo com o CDD, possui o domínio PCI na extremidade

carboxi-terminal. Novamente, a busca em *T. brucei* e *T. cruzi* revelou a presença de ortólogos. Foi criado um alinhamento múltiplo incluindo as seqüências de planta, de humano e dos tripanossomatídeos que revelou a conservação de vários resíduos aromáticos ao longo da seqüência incluindo cinco W conservados em todos os organismos analisados (Figura 13).

Para o estudo do eIF3f, que em mamíferos apresenta o domínio MPN, a seqüência de humano foi utilizada no omniBLAST do geneDB. O melhor resultado que se obteve foi a proteína LmjF32.0390 que foi anotada automaticamente como um componente do proteassomo. O BLAST reverso também teve seus melhores resultados contra proteínas do proteassomo. No entanto, como se sabe da literatura que certas subunidades do eIF3 estão relacionadas com subunidades do proteassomo essa proteína foi analisada mesmo assim. Com a seqüência de *L. major*, fez-se uma busca nos bancos de *T. brucei* e *T. cruzi* encontrando no primeiro o ortólogo Tb10.61.2180 e no segundo Tc00.1047053509649.30 e Tc00.1047053511285.90 como ortólogos. Foi feita então uma busca no CDD com o LmjF32.0390, teve-se como resultado a identificação do domínio MPN no amino-terminal. Este domínio é encontrado em subunidades do proteassomo, subunidades do eIF3 e fatores de regulação da transcrição e também ocorre em procariotos. O passo seguinte foi criar um alinhamento múltiplo incluindo o eIF3f de diversas espécies e também subunidades do proteassomo de espécies selecionadas. Com base nesse alinhamento foi criada uma árvore filogenética na qual os tripanossomatídeos ficaram agrupados no ramo das subunidades do proteassomo (Figura 14). Um resultado similar foi obtido após a análise do eIF3h onde se encontrou um candidato a homólogo anotado, no geneDB, como LmjF34.0650 que no BLAST reverso também apresenta uma proximidade maior a componentes do proteassomo. A análise dessa proteína no CDD mostrou uma arquitetura similar à proteína de humano utilizada com a presença do domínio MPN. No entanto, as análises de

pairwise desta subunidade apresentaram *e-values* abaixo do ponto de corte pré-determinado.

A subunidade eIF3g tem a capacidade de se ligar ao eIF4B e a RNA. A busca pelo eIF3g no genoma de *L. major* utilizando o homólogo de *H. sapiens* e de planta levou a um omniBLAST dúbio. Como este fator possui um motivo RRM, as proteínas encontradas nessa busca aparentam ser um resultado artefatual uma vez que elas se alinham apenas neste motivo. Quando alinhadas as seqüências de humano e planta, a similaridade se estende por toda a seqüência. Assim como para o eIF3a, não foram feitas buscas nos bancos de dados de *T. brucei* e de *T. cruzi*.

A subunidade eIF3i apresenta o domínio WD40 que é caracterizado por apresentar repetições de um dipeptídeo GH e, cerca de 40 resíduos depois, um outro dipeptídeo: o WD. Em *L. major* foi possível encontrar um candidato a homólogo no cromossomo 32 com uma massa predita de 26.3kDa anotado automaticamente no GeneDB com o nome LmjF36.3880. Como resultado encontrou-se um ortólogo em *T. brucei* anotado com o número de acesso: Tb11.01.1370; e em *T. cruzi* foram encontrados dois ortólogos anotados com os números de acesso: Tc00.1047053511229.80 e Tc00.1047053511589.230. A seqüência de *L. major* apresentou uma extremidade amino-terminal mais longa que a de mamíferos e plantas com aproximadamente 70 resíduos a mais. Foi feito, então, um alinhamento múltiplo contendo os candidatos de *T. brucei* e *T. cruzi* e foi possível perceber que eles não apresentavam a extremidade amino-terminal mais longa. Esse resultado sugere que a seqüência de *L. major* inclui um artefato de anotação (Figura 15). De acordo com o CDD o candidato de *L. major* apresenta o domínio WD40, mas a análise do alinhamento múltiplo não permitiu a identificação dos dipeptídeos GH ou das repetições de WD.

O eIF3k também é ausente em leveduras, não possui um domínio característico e não faz parte do núcleo central do eIF3. Contudo, é a única subunidade do eIF3 com estrutura resolvida. O estudo do eIF3k, baseado no homólogo de mamíferos, não levou a resultados confiáveis,

uma vez que os *e-values* eram muito altos. Um novo omniBLAST, agora com a seqüência de planta foi efetuado e este gerou uma resposta mais confiável com um *e-value* mais significativo. Nas duas buscas a proteína LmjF32.2180 apresentou os melhores resultados, mas contra a seqüência de humano foi com um *e-value* de 0,00018, acima do ponto de corte pré-determinado, e contra a de planta de 8e-08. O BLAST reverso com a seqüência de *Leishmania* encontrou o homólogo de planta. Apesar do *e-value* apresentado pelo LmjF32.2180 ser baixo, essa proteína foi considerada como candidato a homólogo e necessita de estudos mais aprofundados para ser validada.

Discussão

Neste trabalho, foi possível aplicar ferramentas de bioinformática e obter candidatos a homólogos dos diversos fatores de iniciação da tradução nos genomas de *L. major*, *T. brucei* e *T. cruzi*. Estudos anteriores para o eIF4F observaram uma multiplicidade de candidatos a homólogos, mais especificamente dois ao eIF4A, quatro ao eIF4E e cinco ao eIF4G (Dhalia *et al.*, 2005). A proposta inicial era de que essa multiplicidade de fatores tivesse relação com as diversas fases do ciclo destes parasitas, no entanto não foi observada, para os fatores analisados neste trabalho, uma multiplicidade de candidatos a homólogos como a apresentada pelo eIF4G. Uma única proteína, a subunidade eIF3b, apresentou um segundo candidato a homólogo.

O fato do eIF4F se apresentar de forma tão distinta pode ter relação com particularidades do mRNA dos tripanossomatídeos que passa por um processo de maturação distinto dos demais eucariotos. Desta forma, necessitaria que os fatores que interagisse com eles apresentassem características próprias. Por outro lado, os demais fatores como estariam envolvidos com processos relacionados ao ribossomo se apresentam de forma mais conservada. Essa possibilidade é reforçada pelo exemplo do

eIF6 que apresentou o maior grau de conservação das proteínas aqui estudadas, possuindo 58% de identidade e 78% de positividade ao se comparar as seqüências de tripanossomatídeos com a de *H. sapiens* como pode ser visto na Tabela 1.

Um estudo anterior analisou três fatores de iniciação da tradução (eIF1, eIF1A e eIF5A) vistos como tendo um papel "periférico" na iniciação da tradução como sendo universalmente conservados nas principais linhagens evolutivas (Kyrpides e Woese, 1997). Estas três proteínas foram encontradas em tripanossomatídeos, com altos níveis de similaridade, como esperado. No caso específico do eIF1 foram mapeados em tripanossomatídeos mutações singulares de resíduos importantes. Peculiarmente, essas mutações ocorrem em resíduos relacionados com a identificação do códon AUG de iniciação correto e foram identificadas apenas nos tripanossomatídeos. Não está claro o porquê destas mutações. Uma hipótese é que elas participem de algum processo de controle da tradução ainda desconhecido.

O fator eIF3 funciona como um ancoradouro central para diversos fatores de iniciação da tradução incluindo o eIF4F (Hershey e Merrick, 2000). As subunidades do eIF3 apresentaram a menor conservação dentre os candidatos a homólogos de tripanossomatídeos aqui estudados. Duas subunidades que fazem parte do núcleo central deste fator em leveduras, sendo essenciais neste organismo, não foram localizadas por este trabalho, mais especificamente o eIF3a e o eIF3g. É interessante notar que um trabalho anterior em leveduras revelou que as subunidades eIF3a e eIF3c e o fator eIF5 formam uma unidade mínima suficiente para ligação à subunidade ribossomal 40S e também foi demonstrado que a extremidade carboxi terminal do eIF3a interage com o fator eIF2 (Valasek *et al*., 2003). Portanto, a aparente ausência dessa subunidade em tripanossomatídeos sugere que o fator eIF3 destes organismos apresenta particularidades próprias. Um trabalho paralelo identificou uma proteína em *T. brucei* nomeada de EIF3D que seria um homólogo ao eIF3g (De Gaudenzi *et al*., 2005). Esta proteína EIF3D não foi identificada por este

trabalho durante a busca por candidatos ao eIF3g e a comparação entre o EIF3D e o eIF3g de mamíferos não revelou uma homólogia clara.

As subunidades eIF3d, eIF3e, eIF3f e eIF3k do eIF3 foram identificadas em tripanossomatídeos. Das subunidades pertencentes ao núcleo central de leveduras foram identificadas a eIF3b, eIF3c e eIF3i e as duas primeiras apresentaram diversas deleções após a análise de alinhamento múltiplo com outros eucariotos. As diferenças encontradas neste fator podem refletir sua natureza de complexo protéico responsável por interações proteína-proteína tendo, desta forma, a se adaptar à maquinaria dos organismos em estudo.

Os resultados obtidos por este trabalho em sua maioria confirmaram a análise preliminar dos fatores de iniciação da tradução realizada automaticamente pelo GeneDB. Em pelo menos dois casos, a seqüência de *T. cruzi* do eIF6 e a seqüência de *L. major* do eIF3i, observou-se longas extremidades amino-terminais não consistente com os demais organismos, o que sugerem que sejam artefatos da anotação automática. Também foi possível realizar o estudo de domínios e motivos dos diversos fatores assinalando mutações e conservações relevantes assim como hipóteses para o entendimento destas. Esses resultados mostram que existem muitas características conservadas entre candidatos a fatores de iniciação dos tripanossomatídeos e os dos demais eucariotos. No entanto, foi possível verificar a existência de alterações interessantes, algumas delas únicas, nos tripanossomatídeos que necessitam de uma análise mais aprofundada.

Figuras e Tabelas

Tabela 1: Tabela dos resultados obtidos através da busca no banco de dados do GeneDB. O *Score* (em *bits*), *e-value*, Identidade e Similaridade foram calculados contra o homólogo de humanos. * Os valores de Identidade e Similaridade são dados em percentuais. [1] Fatores para os quais não foram obtidos candidatos a homólogos por este trabalho. [2] Candidatos que tiveram seus *Scores* (em *bits*), *e-values*, Identidades e Similaridades calculados contra o homólogo de *A. thaliana*. [3] A seqüência de *L. major* apresenta uma duplicação em tandem idêntica (LmjF25.0730) que foi omitida da tabela.

Fator	Accession(GeneDB)	Tamanho	Score	e-value	Identidade*	Similaridade*
eIF1	LmjF24.1210	107 aa	72	4e-12	43%	62%
	LinJ24.0880	107 aa	72.4	4e-12	43%	62%
	Tb11.02.3595	109 aa	68.9	3e-11	43%	59%
eIF1A	LmjF16.0140	168 aa	60.1	2e-08	39%	57%
	LinJ16.0150	168 aa	60.1	2e-08	39%	57%
	Tb08.11J15.1060	172 aa	64.3	9e-10	45%	61%
eIF3a[1]						
eIF3b1	LmjF17.1290	709 aa	124	1e-26	25%	41%
	LinJ17.1070	709 aa	122	8e-26	24%	40%
	Tb05.26K5.1010	696 aa	149	6e-34	23%	40%
eIF3b2	Lm16.0690	551 aa	56.6	6e-06	23%	36%
	LinJ16.0720	551 aa	56.2	8e-06	24%	36%
	Tb927.5.3450	527 aa	54.7	2e-05	21%	36%
eIF3c	LmjF36.6980	731 aa	85.9	8e-15	22%	43%
	LinJ36.6090	731 aa	85.9	8e-15	22%	42%
	Tb10.6k15.2250	740 aa	129	5e-28	22%	41%
eIF3d	LmjF30.3040	531 aa	97.4	1e-18	24%	39%
	LinJ30.3100	531 aa	109	4e-22	24%	39%
	Tb06.26G9.950	536 aa	127	2e-27	28%	45%
eIF3e	LmjF28.2310	405 aa	380	1e-34	27%	47%
	LinJ28.2420	405 aa	394	3e-36	27%	47%
	Tb11.01.3420	413 aa	300	2e-25	27%	46%
eIF3f	LmjF32.0390	359 aa	108	4e-22	27%	50%
	LinJ32.0400	359 aa	104	5e-21	26%	49%
	Tb10.61.2180	381 aa	88.6	4e-16	24%	46%
eIF3g[1]						
eIF3h[1]						
eIF3i[2]	LmjF36.3880	419 aa	166	2e-39	33%	51%
	LinJ36.3460	407 aa	167	6e-40	33%	51%
	Tb11.01.1370	342 aa	215	3e-54	37%	55%
eIF3k[2]	LmjF32.2180	233 aa	148	3e-08	22%	47%
	LinJ32.2240	208 aa	148	3e-08	22%	47%
	Tb11.01.7070	205 aa	146	5e-08	26%	46%
eIF5	LmjF34.0350	378 aa	124	6e-27	44%	60%
	LinJ34.0320	378 aa	124	8e-27	44%	60%
	Tb10.70.4880	382 aa	115	3e-24	44%	56%
eIF5A	LmjF25.0720[3]	166 aa	142	3e-33	44%	64%
	LinJ25.0640	166 aa	141	6e-33	44%	64%
	Tb11.03.0410	166 aa	142	4e-33	42%	66%
eIF5B	LmjF33.2740	816 aa	547	e-154	47%	67%
	LinJ33.2360	817 aa	548	e-154	47%	67%
	Tb927.2.3780	833 aa	548	e-154	44%	60%
eIF6	LmjF36.0890	249 aa	784	6e-82	58%	78%
	LinJ36.1740	249 aa	785	5e-82	58%	78%
	Tb10.70.1770	248 aa	796	3e-83	58%	78%

Tabela 2: Comparação entre os candidatos a homólogos de *L. major* contra *T. brucei* e *T. cruzi*. Foram realizados alinhamentos de *pairwise* com a seqüência de *L. major* contra as seqüências de *T. brucei* e *T. cruzi* isoladamente. As colunas da tabela representam os valores de *e-value*, percentual de identidade e similaridade, respectivamente, para cada alinhamento.

	T. brucei			*T. cruzi*		
	e-value	Similaridade	Identidade	*e-value*	Similaridade	Identidade
*Lm*EIF1	4e-40	75%	88%	7e-41	81%	90%
*Lm*EIF1A	3e-59	69%	77%	9e-59	67%	77%
*Lm*EIF3b1	e-173	44%	63%	e-177	45%	63%
*Lm*EIF3b2	e-148	50%	64%	e-144	50%	62%
*Lm*EIF3c	e-114	36%	53%	e-114	36%	55%
*Lm*EIF3d	6e-73	34%	51%	3e-82	35%	52%
*Lm*EIF3e	e-102	50%	64%	e-119	56%	70%
*Lm*EIF3f	e-112	59%	69%	e-110	60%	70%
*Lm*EIF3i	5e-91	51%	63%	e-90	50%	64%
*Lm*EIF3k	3e-30	39%	64%	6e-39	44%	66%
*Lm*EIF5	1e-90	44%	61%	9e-93	44%	61%
*Lm*EIF5A	7e-61	71%	81%	2e-62	71%	82%
*Lm*EIF5B	0.0	63%	75%	0.0	64%	75%
*Lm*EIF6	e-126	92%	96%	e-125	91%	95%

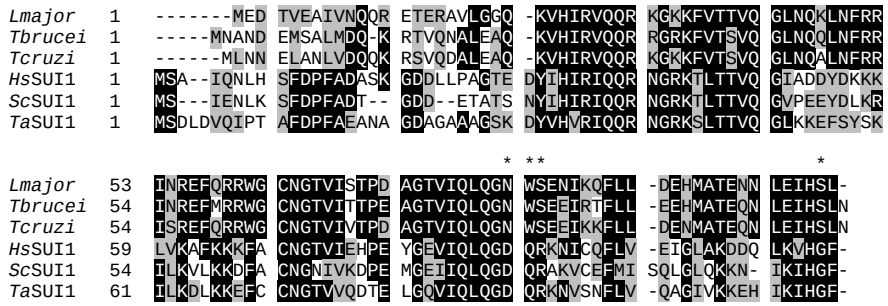


Figura 1: Alinhamento múltiplo dos candidatos a homólogos ao fator eIF1 de *L. major*, *T. brucei* e *T. cruzi*. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens*, *S. cerevisie*, *T. aestivum*. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. Estrelas (*) indicam mutações específicas de tripanossomatídeos em posições que nos demais eucariotos levam a mudanças na matriz de leitura e iniciação em códons alternativos.

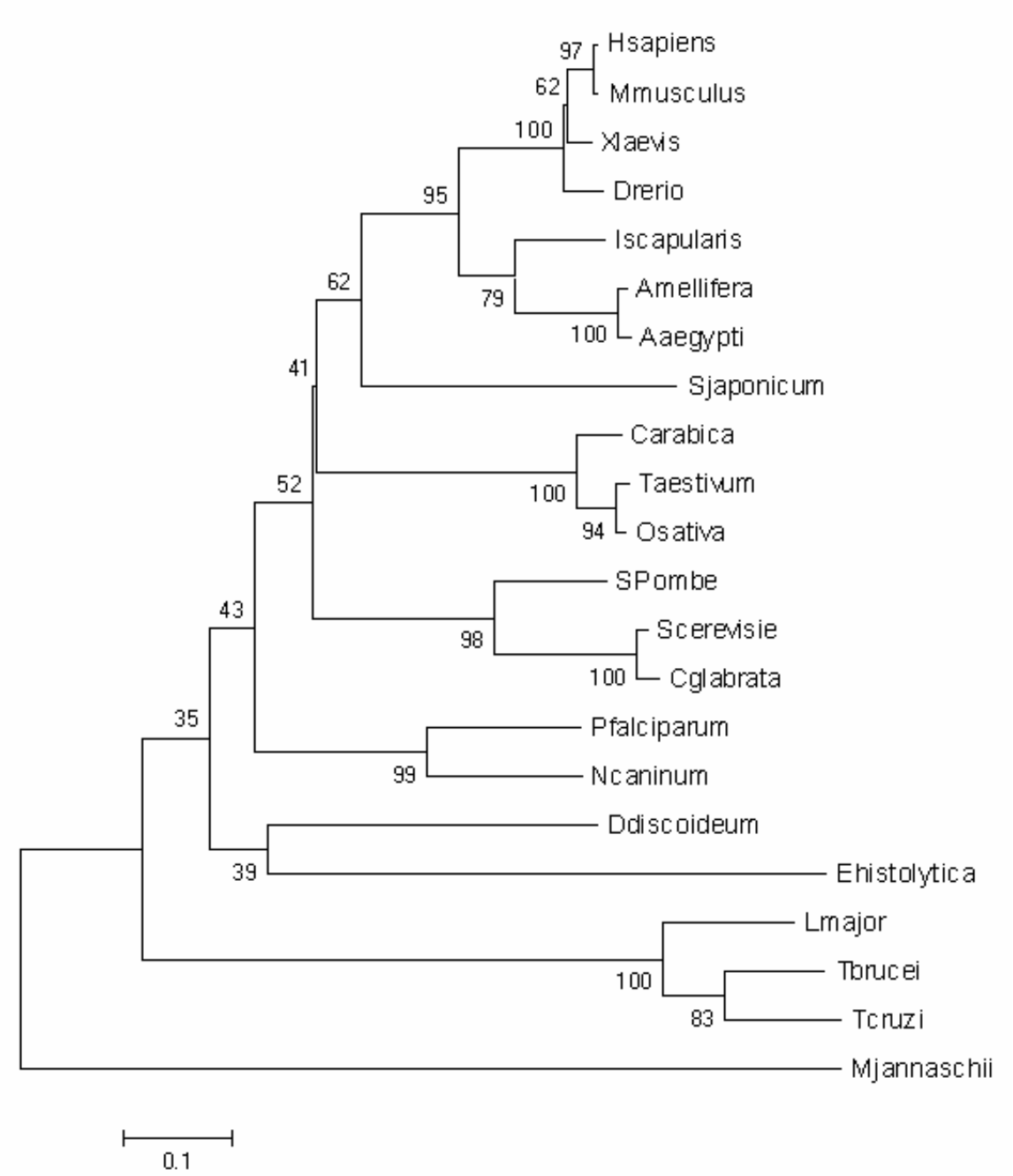


Figura 2: Árvore filogenética obtida através do alinhamento múltiplo dos homólogos do eIF1 de diversos organismos. Foi criada por *neighbourjoining* com um *bootstrap* de 1000 reconstruções. Os tripanossomatídeos ficam num ramo bem isolado e a seqüência de *M. jannaschii* forma um grupo externo.

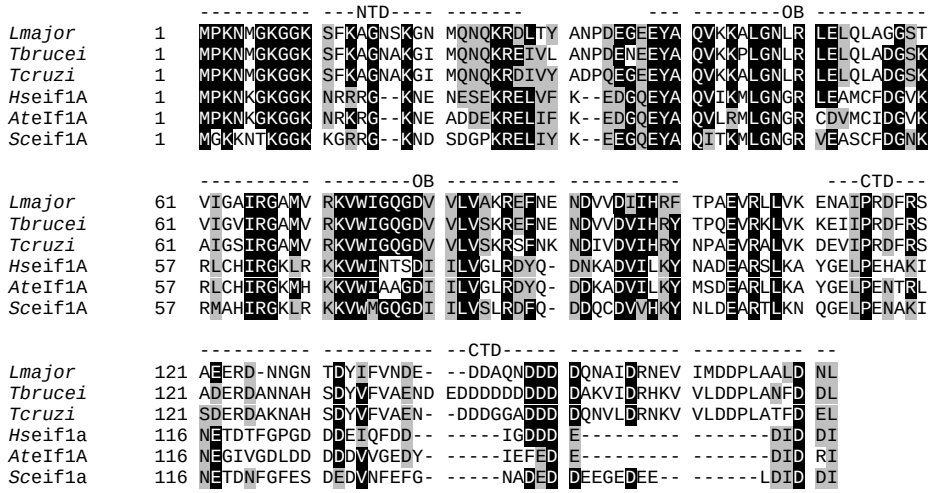


Figura 3: Alinhamento múltiplo dos candidatos a homólogos ao fator eIF1A de *L. major*, *T. brucei* e *T. cruzi*. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens*, *S. cerevisie* e *A. thaliana*. do fator eIF1A. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. As extremidades amino e carboxi terminal estão assinaladas, assim como o dobramento OB central.

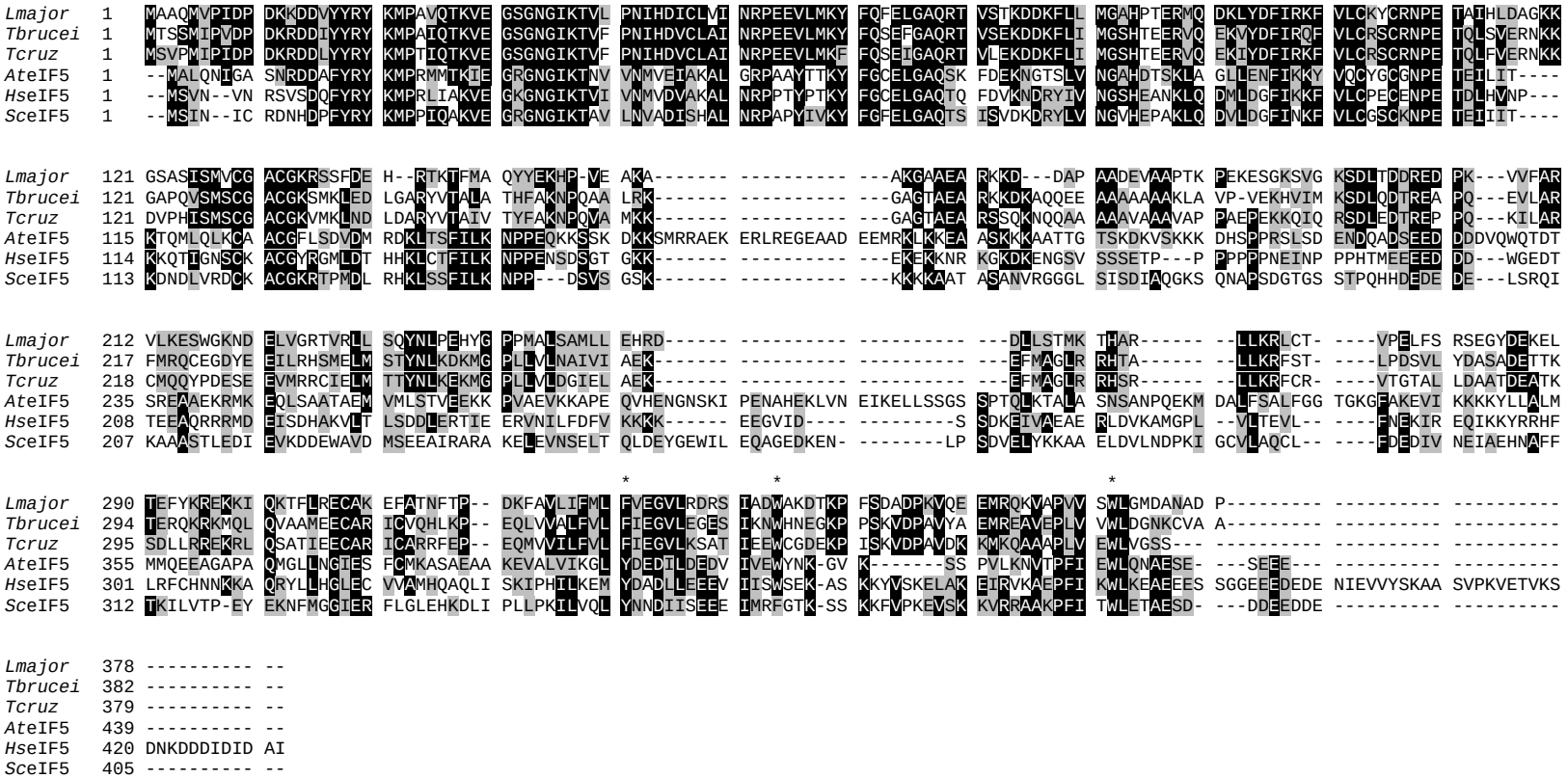


Figura 4: Alinhamento múltiplo dos candidatos a homólogos ao fator eIF5 de *L. major*, *T. brucei* e *T. cruzi*. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens*, *S. cerevisie* e *A. thaliana*. do fator eIF5. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. Observa-se a extremidade amino-terminal conservada. Estrelas (*) representam resíduos aromáticos conservados na extremidade carboxi-terminal.

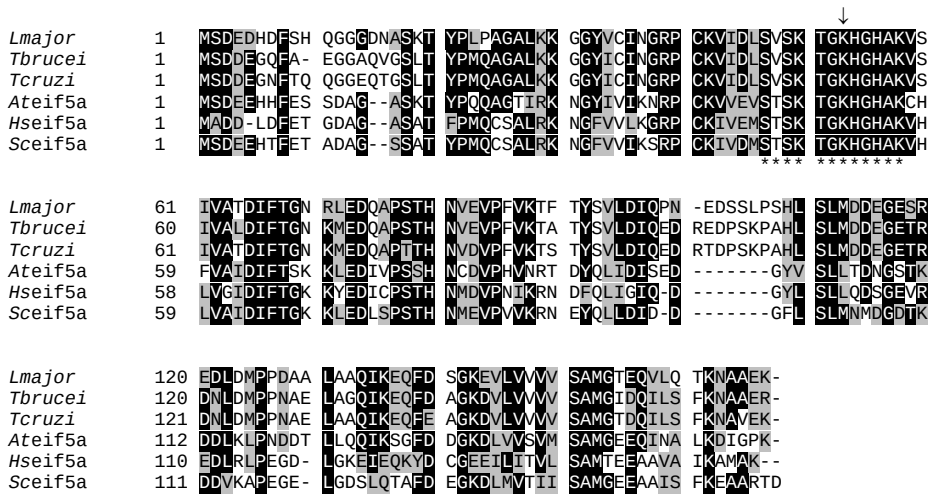


Figura 5: Alinhamento múltiplo dos candidatos a homólogos ao fator eIF5A de *L. major*, *T. brucei* e *T. cruzi*. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens*, *S. cerevisie* e *A. thaliana*. do fator eIF5A. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. A seta (↓) indica a lisina que é alterada em hipusina e as estrelas (*) os resíduos extremamente conservados que circundam a hipusina.

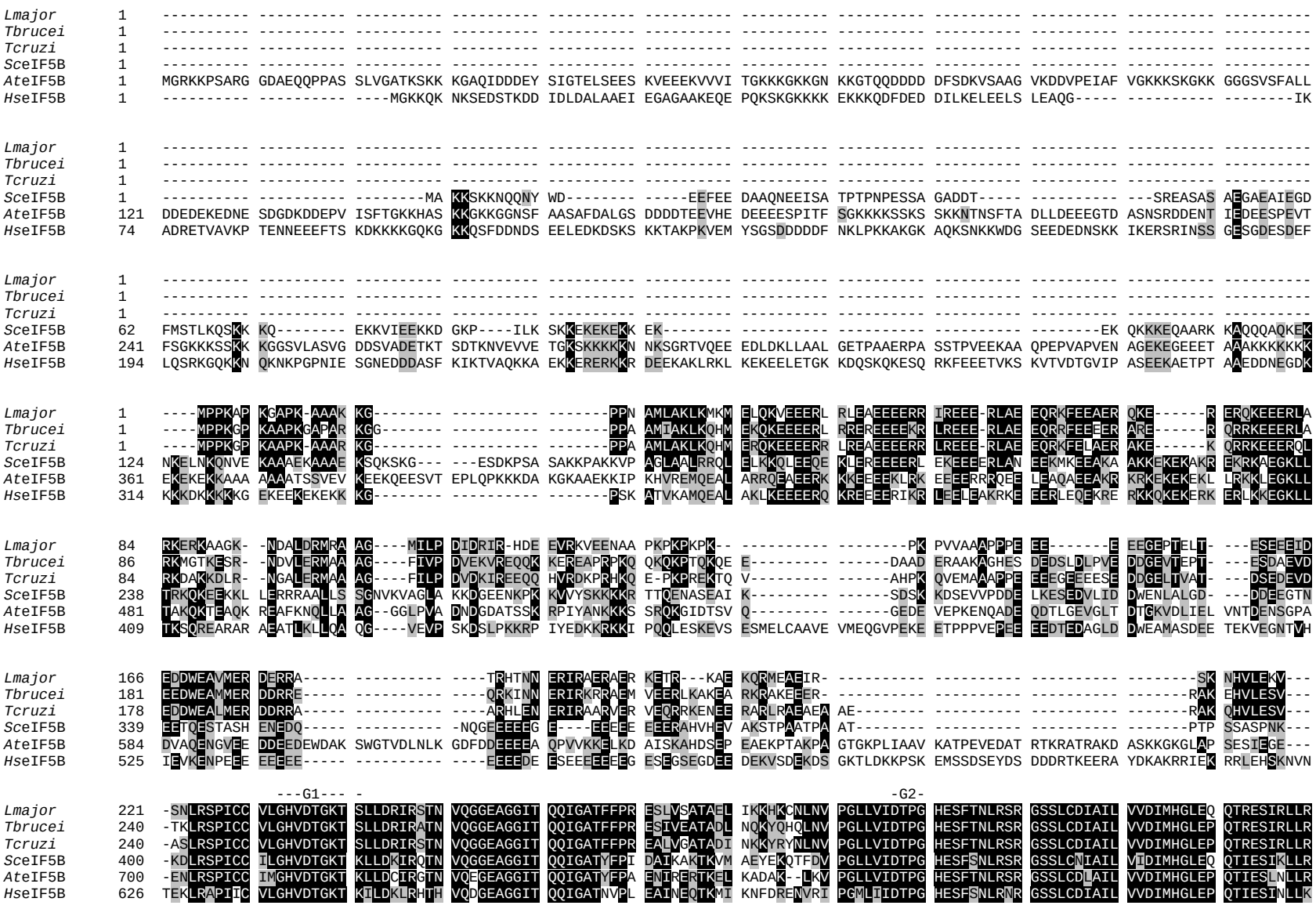

Lmajor 1
Tbrucei 1
Tcruzi 1
SceIF5B 1
AteIF5B 1 MGRKKPSARG GDAEQQPPAS SLVGATKSKK KGAQIDDDEY SIGTELSEES KVEEEKVVVI TGKKKGKKGN KKGTQQDDDD DFSDKVSAAG VKDDVPEIAF VGKKKSKGKK GGGSVSFALL
HseIF5B 1 ---MGKKQK NKSEDSTKDD IDLDALAAEI EGAGAAKEQE PQKSKGKKKK EKKKQDFDED DILKELEELS LEAQG----- ---------- --------IK
SceIF5B 1 MA KKSKKNQQNY WD-------- -----EEFEE DAAQNEEISA TPTPNPESSA GADDT----- ---------- ---SREASAS AEGAEAIEGD
AteIF5B 121 DDEDEKEDNE SDGDKDDEPV ISFTGKKHAS KKGKKGGNSF AASAFDALGS DDDDTEEVHE DEEEESPITF SGKKKKSSKS SKKNTNSFTA DLLDEEEGTD ASNSRDDENT IEDEESPEVT
HseIF5B 74 ADRETVAVKP TENNEEEFTS KDKKKKGQKG KKQSFDDNDS EELEDKDSKS KKTAKPKVEM YSGSDDDDDF NKLPKKAKGK AQKSNKKWDG SEEDEDNSKK IKERSRINSS GESGDESDEF
SceIF5B 62 FMSTLKQSKK KQ-------- EKKVIEEKKD GKP----ILK SKKEKEKEKK EK-------- ---------- ---------- ---------- --------EK QKKKEQAARK KAQQQAQKEK
AteIF5B 241 FSGKKKSSKK KGGSVLASVG DDSVADETKT SDTKNVEVVE TGKSKKKKKN NKSGRTVQEE EDLDKLLAAL GETPAAERPA SSTPVEEKAA QPEPVAPVEN AGEKEGEEET AAAKKKKKKK
HseIF5B 194 LQSRKGQKKN QKNKPGPNIE SGNEDDDASF KIKTVAQKKA EKKERERKKR DEEKAKLRKL KEKEELETGK KDQSKQKESQ RKFEEETVKS KVTVDTGVIP ASEEKAETPT AAEDDNEGDK
Lmajor 1 ----MPPKAP KGAPK-AAAK KG-------- ---------- -------PPN AMLAKLKMKM ELQKVEEERL RLEAEEEERR IREEE-RLAE EQRKFEEAER QKE------R ERQKEEERLA
Tbrucei 1 ----MPPKGP KAAPKGAPAR KGG------- ---------- -------PPA AMIAKLKQHM EKQKEEEERL RREREEEEKR LREEE-RLAE EQRRFEEEER ARE------R QRRKEEERLA
Tcruzi 1 ----MPPKGP KAAPK-AAAR KG-------- ---------- -------PPA AMLAKLKQHM ERQKEEEERR LREAEEEERR LREEE-RLAE EQRKFELAER AKE------K QRRKEEERQL
SceIF5B 124 NKELNKQNVE KAAAEKAAAE KSQKSKG--- ---ESDKPSA SAKKPAKKVP AGLAALRRQL ELKKQLEEQE KLERFEEEERL EKEEEERLAN EEKMKEEAKA AKKEKEKAKR EKRKAEGKLL
AteIF5B 361 EKEKEKKAAA AAAATSSVEV KEEKQEESVT EPLQPKKKDA KGKAAEKKIP KHVREMQEAL ARRQEAEERK KKEEEEKLRK EEEERRRQEE LEAQAEEAKR KRKEKEKEKL LRKKLEGKLL
HseIF5B 314 KKKDKKKKKG EKEEKEKEKK KG-------- ---------- -------PSK ATVKAMQEAL AKLKEEEERQ KREEEERIKR LEELEAKRKE EERLEQEKRE RKKQKEKERK ERLKKEGKLL
Lmajor 84 RKERKAAGK- -NDALDRMRA AG----MILP DIDRIR-HDE EVRKVEENAA PKPKPKPK-- ---------- --------PK PVVAAAPPPE EE-------E EEGEPTELI- ---ESEEEID
Tbrucei 86 RKMGTKESR- -NDVLERMAA AG----FIVP DVEKVREQQK KEREAPRPKQ QKQKPTQKQE E--------- ------DAAD ERAAKAGHES DEDSLDLPVE DDGEVTEPT- ---ESDAEVD
Tcruzi 84 RKDAKKDLR- -NGALERMAA AG----FILP DVDKIREEQQ HVRDKPRHKQ E-PKPREKTQ V--------- ------AHPK QVEMAAAPPE EEEGEEEESE DDGELTVAT- ---DSEDEVD
SceIF5B 238 TRKQKEEKKL LERRRAALLS SGNVKVAGLA KKDGEENKPK KVVYSKKKKR TTQENASEAI K--------- ------SDSK KDSEVVPDDE LKESEDVLID DWENLALGD- ---DDEEGTN
AteIF5B 481 TAKQKTEAQK REAFKNQLLA AG--GGLPVA DNDGDATSSK RPIYANKKKS SRQKGIDTSV Q--------- ------GEDE VEPKENQADE QDTLGEVGLT DTGKVDLIEL VNTDENSGPA
HseIF5B 409 TKSQREARAR AEATLKLLQA QG----VEVP SKDSLPKKRP IYEDKKRKKI PQQLESKEVS ESMELCAAVE VMEQGVPEKE ETPPPVEPEE EEDTEDAGLD DWEAMASDEE TEKVEGNTVH
Lmajor 166 EDDWEAVMER DERRA----- ---------- ----TRHTNN ERIRAERAER KEIR---KAE KQRMEAEIR- ---------- ---------- ---------- --------SK NHVLEKV---
Tbrucei 181 EEDWEAMMER DDRRE----- ---------- ----QRKINN ERIRKRRAEM VEERLKAKEA RKRAKEEER- ---------- ---------- -------RAK EHVLESV---
Tcruzi 178 EDDWEALMER DDRRA----- ---------- ----ARHLEN ERIRAARVER VEQRRKENEE RARLRAEAEA AE-------- ---------- ---------- -------RAK QHVLESV---
SceIF5B 339 EETQESTASH ENEDQ----- ---------- -NQGEEEEEG E----EEEEE EEERAHVHEV AKSTPAATPA AT-------- ---------- ---------- -------PTP SSASPNK---
AteIF5B 584 DVAQENGVEE DDEEDEWDAK SWGTVDLNLK GDFDDEEEEA QPVVKKELKD AISKAHDSEP EAEKPTAKPA GTGKPLIAAV KATPEVEDAT RTKRATRAKD ASKKGKGLAP SESIEGE---
HseIF5B 525 IEVKENPEEE EEEEE----- ---------- ----EEEEDE ESEEEEEEEG ESEGSEGDEE DEKVSDEKDS GKTLDKKPSK EMSSDSEYDS DDDRTKEERA YDKAKRRIEK RRLEHSKNVN
---G1--- - -G2-
Lmajor 221 -SNLRSPICC VLGHVDTGKT SLLDRIRSTN VQGGEAGGIT QQIGATFFPR ESLVSATAEL IKKHKCNLNV PGLLVIDTPG HESFTNLRSR GSSLCDIAIL VVDIMHGLEQ QTRESIRLLR
Tbrucei 240 -TKLRSPICC VLGHVDTGKT SLLDRIRATN VQGGEAGGIT QQIGATFFPR ESIVEATADL NQKYQHQLNV PGLLVIDTPG HESFTNLRSR GSSLCDIAIL VVDIMHGLEP QTRESIRLLR
Tcruzi 240 -ASLRSPICC VLGHVDTGKT SLLDRIRSTN VQGGEAGGIT QQIGATFFPR EALVGATADI NKKYRYNLNV PGLLVIDTPG HESFTNLRSR GSSLCDIAIL VVDIMHGLEP QTRESIRLLR
SceIF5B 400 -KDLRSPICC ILGHVDTGKT KLLDKIRQTN VQGGEAGGIT QQIGATYFPI DAIKAKTKVM AEYEKQTFDV PGLLVIDTPG HESFSNLRSR GSSLCNIAIL VIDIMHGLEQ QTIESIKLLR
AteIF5B 700 -ENLRSPICC IMGHVDTGKT KLLDCIRGTN VQEGEAGGIT QQIGATYFPA ENIRERTKEL KADAK--LKV PGLLVIDTPG HESFTNLRSR GSSLCDLAIL VVDIMHGLEP QTIESLNLLR
HseIF5B 626 TEKLRAPIIC VLGHVDTGKT KILDKLRHTH VQDGEAGGIT QQIGATNVPL EAINEQTKMI KNFDRENVRI PGMLTIDTPG HESFSNLRNR GSSLCDIAIL VVDIMHGLEP QTIESINLLK

Figura 6: Alinhamento múltiplo da extremidade amino-terminal e do domínio G do eIF5B incluindo os candidatos a homólogos de *L. major*, *T. cruzi* e *T. brucei* e as seqüências caracterizadas de *H. sapiens*, *A. thaliana* e *S. cerevisie*. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. Os quatro motivos do domínio G estão indicados e a seta marca a mutação R421 de *L. major* situada no motivo G4.

Figura 7: Desenho esquemático do eIF5B mostrando o percentual de identidade de cada domínio entre *L. major* e *H. sapiens*. É possível perceber que o percentual de identidade é decrescente no sentido dos domínios G, II, III e IV. Em branco está representado o domínio amino terminal (NTD) que é a porção menos conservada da proteína.

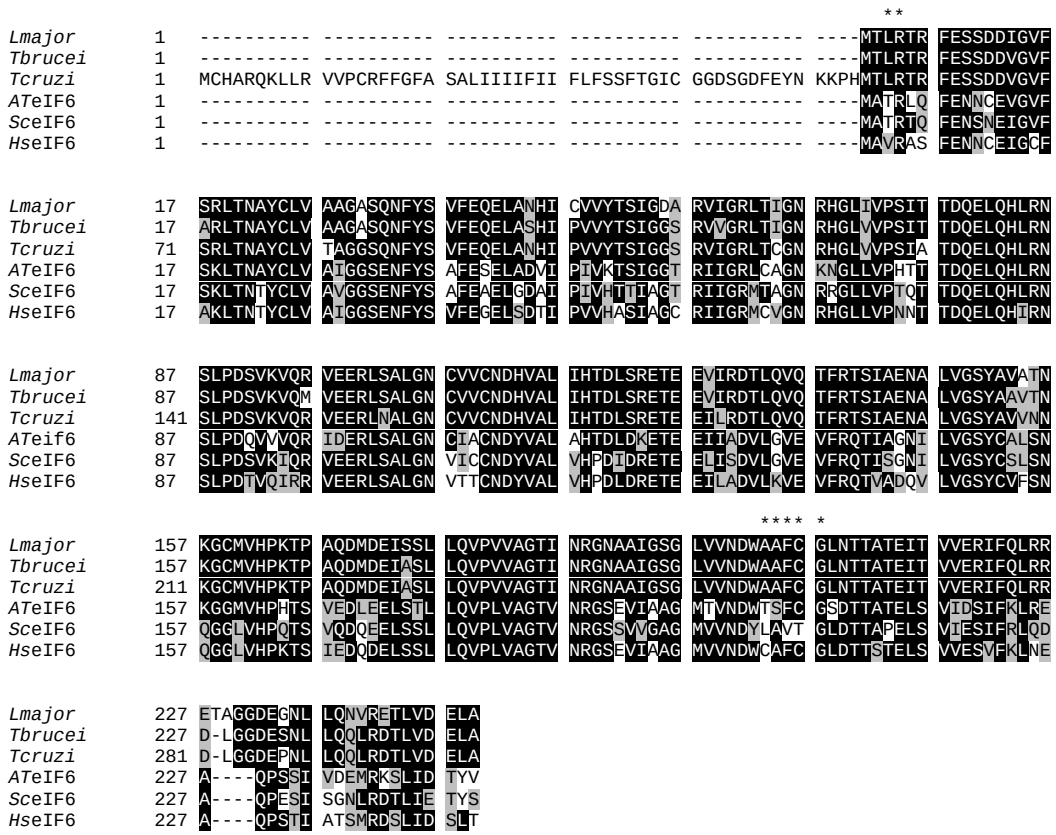


Figura 8: Alinhamento múltiplo do eIF6 mostrando sua alta conservação ao nível de seqüência. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens*, *S. cerevisie* e *A. thaliana*. do fator eIF6. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. Estrelas (*) indicam os resíduos que compõe o fechamento em velcro. A região amino-terminal de *T. cruzi* parece ser um artefato de anotação.

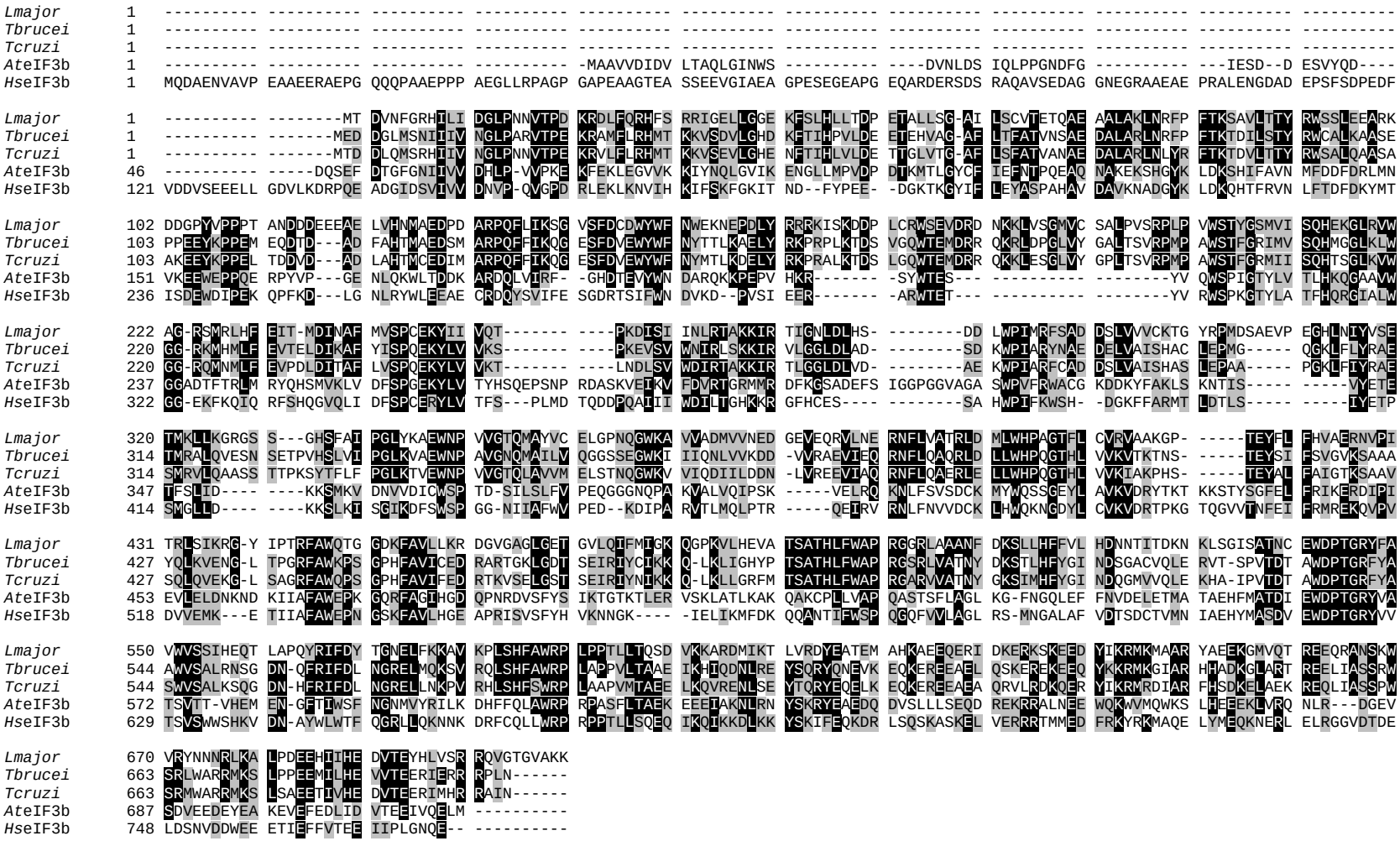


Figura 9: Alinhamento múltiplo do eIF3b. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens* e *A. thaliana* da subunidade eIF3b. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. É possível verificar que a extremidade amino-terminal dos tripanossomatídeos é mais curta que a dos demais eucariotos. Também são visíveis diversas inserções e deleções ao longo do alinhamento.

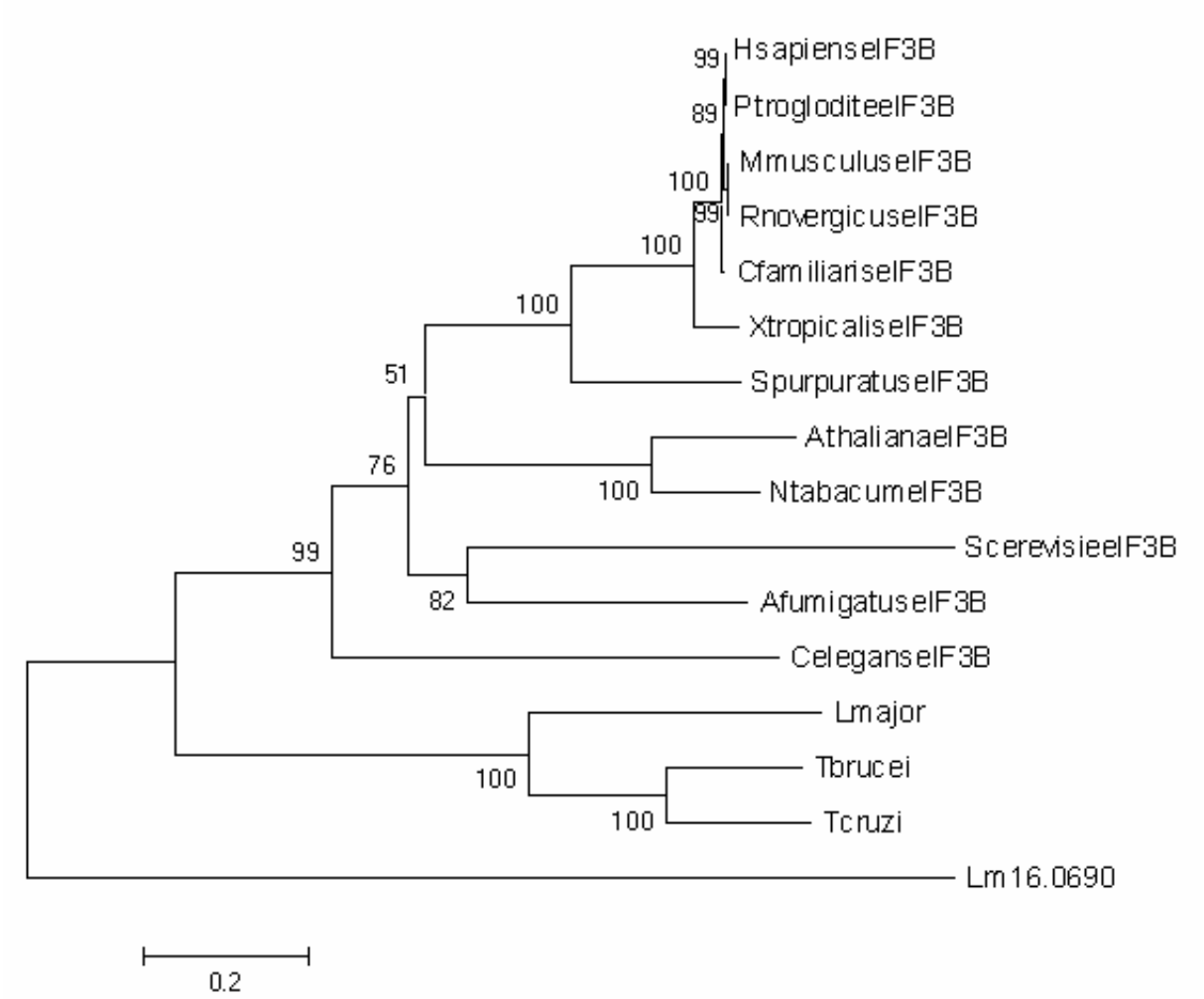


Figura 10: Árvore filogenética do eIF3b onde os tripanossomatídeos formam um ramo isolado dos demais eucariotos. A seqüência do Lm16.0690 foi incluída para gerar um grupo externo.

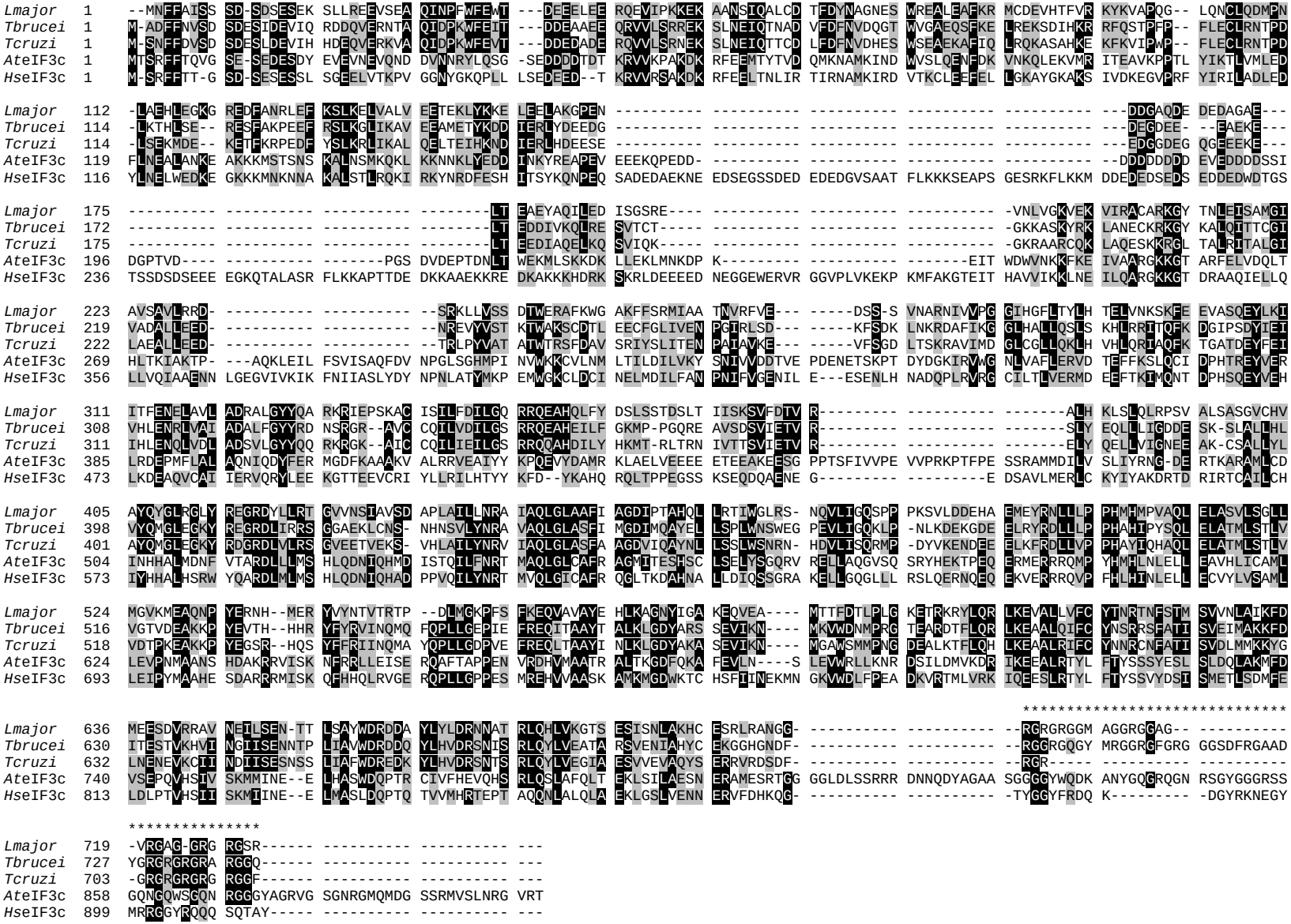


Figura 11 Alinhamento múltiplo do eIF3c. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens* e *A. thaliana*. do fator eIF3c. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. As estrelas (*) demarcam a região carboxi terminal rica em resíduos de glicina e em resíduos básicos.

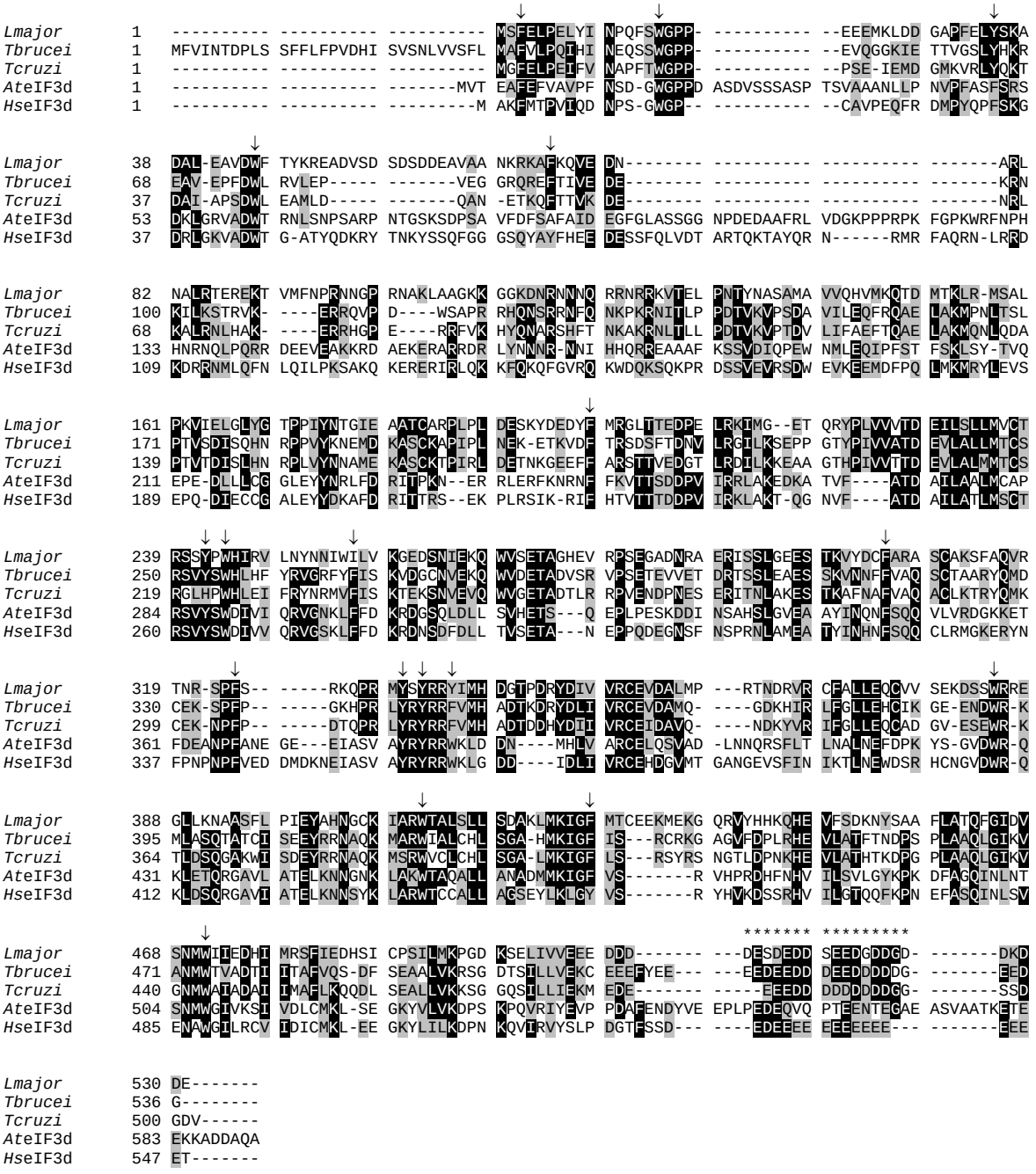


Figura 12: Alinhamento múltiplo do eIF3d. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens* e *A. thaliana* da subunidade eIF3d. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. Setas (↓) representam resíduos aromáticos extremamente conservados e estrelas (*) marcam a região acídica-amino terminal.

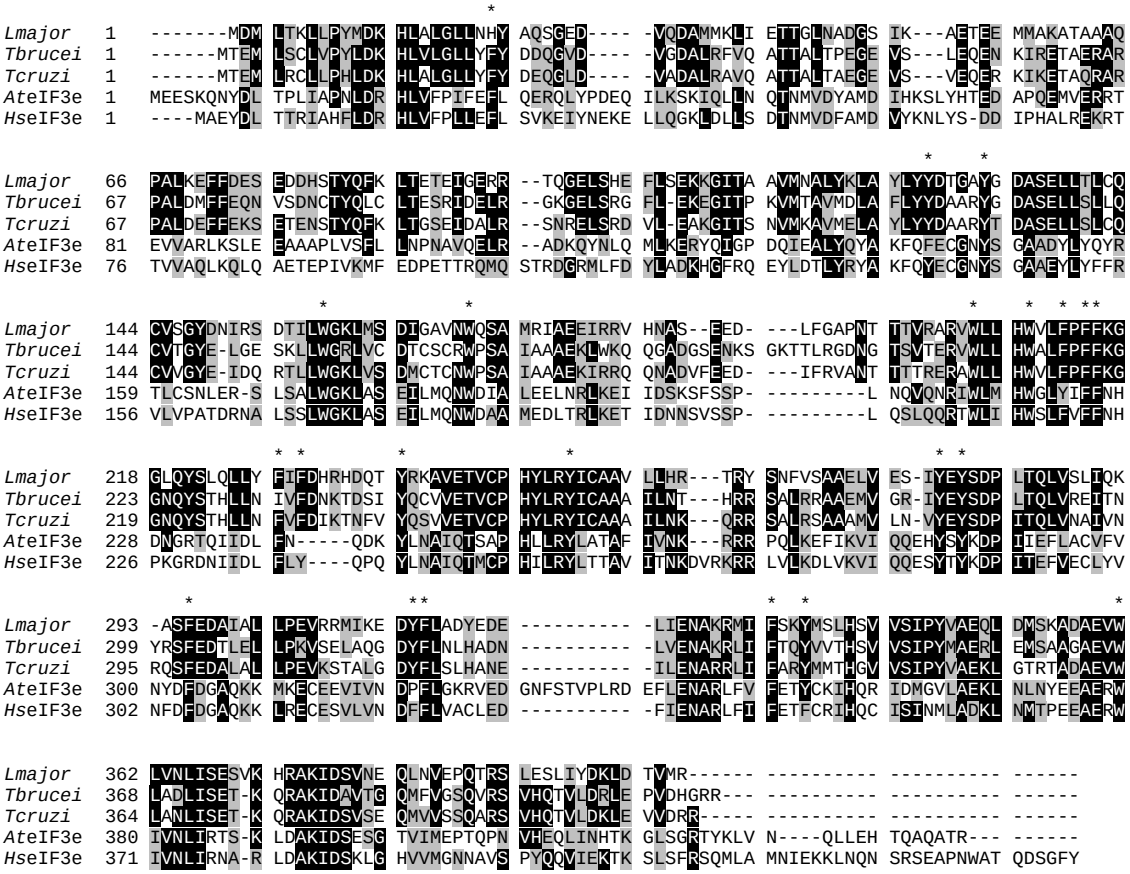


Figura 13: Alinhamento múltiplo do eIF3e. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens* e *A. thaliana* da subunidade eIF3e. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. As estrelas (*) representam resíduos aromáticos extremamente conservados nos organismos alinhados.

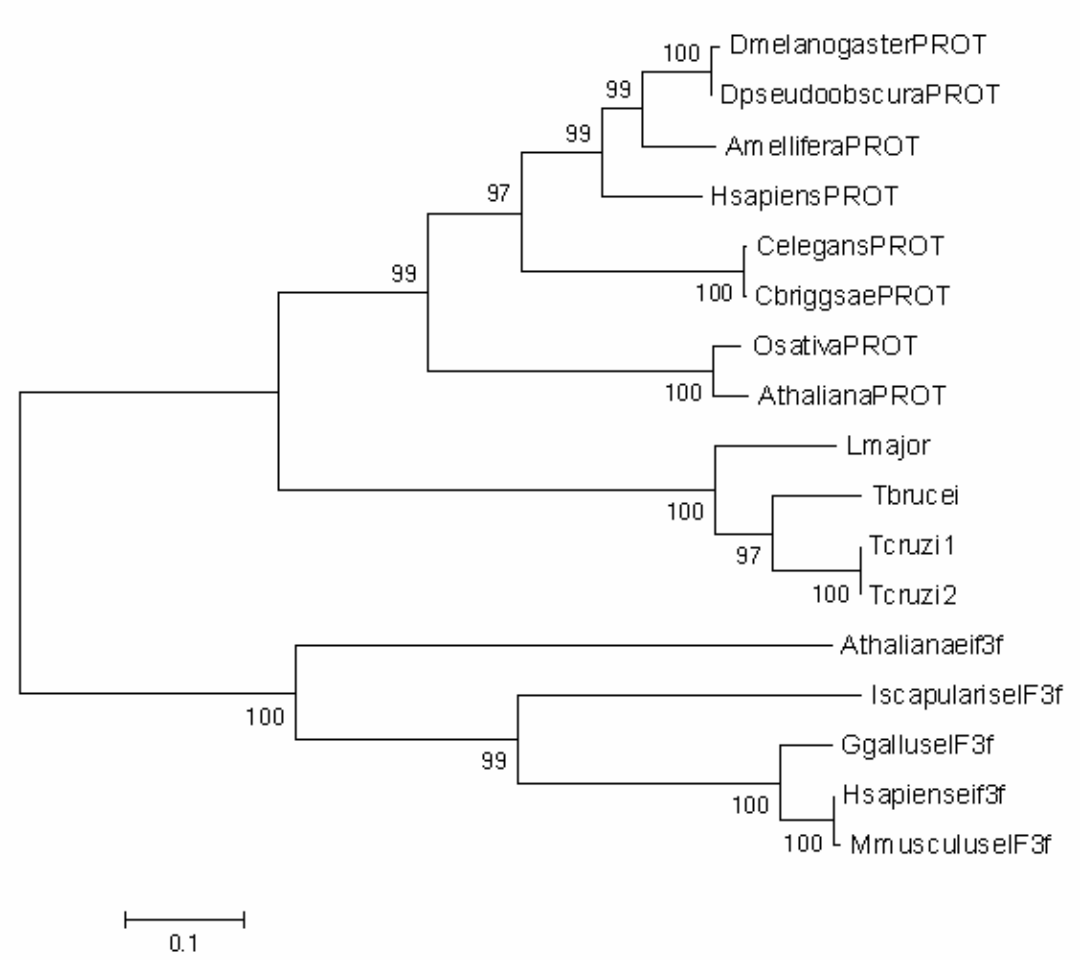


Figura 14: Árvore filogenética do eIF3f. Percebe-se a formação de dois blocos, o superior concentra as proteínas do proteassomo enquanto que o inferior as subunidades do eIF3f. As proteínas de tripanossomatídeos ficaram incluídas no bloco dos componentes do proteassomo estando, desta forma, mais próximas destas.

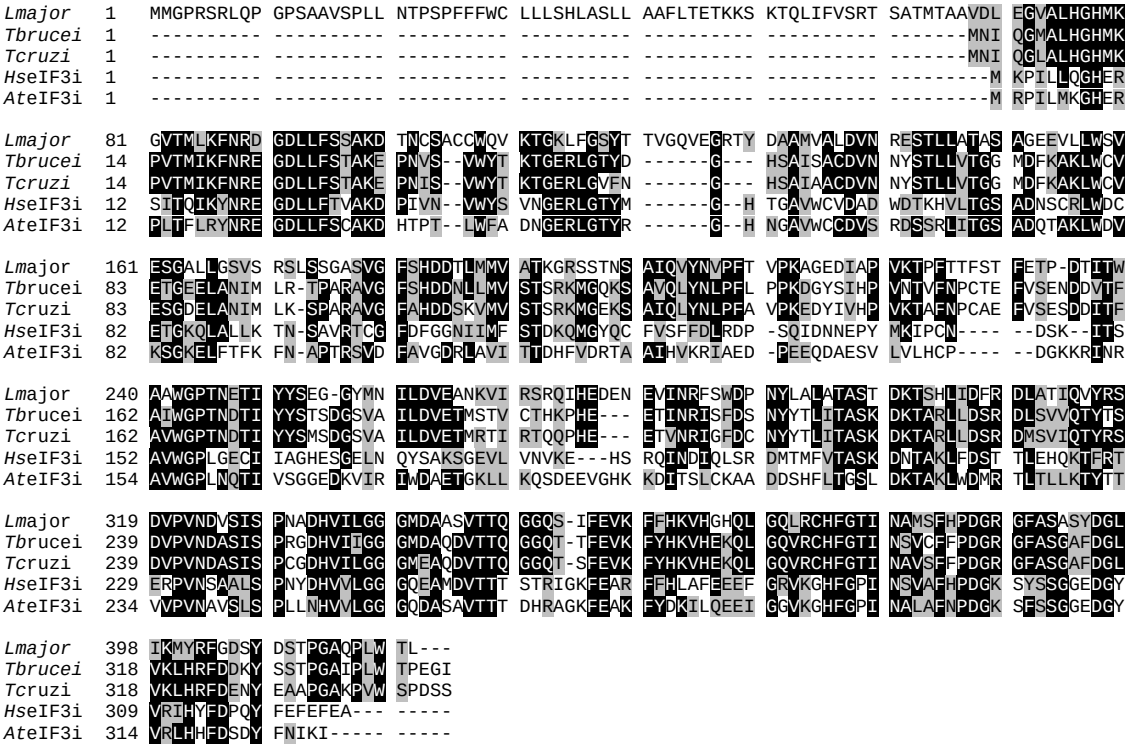


Figura 15: Alinhamento múltiplo do eIF3i mostrando a conservação ao longo de toda a seqüência. Os tripanossomatídeos foram alinhados contra os homólogos caracterizados de *H. sapiens* e *A. thaliana* da subunidade eIF3i. Utilizou-se um limiar de 50% para o sombreamento dos resíduos sendo os identicamente alinhados de preto e os similares de cinza. A extremidade amino-terminal de *L. major* anotada automaticamente aparenta ser um artefato.

Bibliografia

Alonso G, Guevara P and Ramirez JL (1992) Trypanosomatidae codon usage and GC distribution. Mem Inst Osv Cruz 87:517-523.

Benne R (1990) RNA editing in trypanosomes: is there a message? Trends Genet V.6: 177-181.

Borst P, van der Ploeg M, van Hoek JFM, Tas J and James J (1982) On the DNA content of trypanosomes. Mol Biochem Parasitol 6:13.

Browning KS, Gallie DR, Hershey JW, Hinnebusch AG, Maitra U, Merrick WC and Norbury C (2001) Unified nomenclature for the subunits of eukaryotic initiation factor 3. Trends Biochem Sci 26:284

Ceci M, Gaviraghi C, Gorrini C, Sala LA, Offenhauser N, Marchisio PC and Biffo S (2003) Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. Nature 426:579-84.

Clayton CE (2002) Life without transcriptional control? From fly to man and back again. EMBO J 21:1881-88.

Chaudhuri J, Si K and Maitra U (1997) Function of eukaryotic translation initiation factor 1A (eIF1A) (formerly called eIF-4C) in initiation of protein synthesis. J Biol Chem 272:7883-91.

De Gaudenzi J, Frasch AC and Clayton C (2005) RNA-binding domain proteins in Kinetoplastids: a comparative analysis. Eukaryot Cell 4:2106-14.

Dhalia R, Reis CRS, Freire ER, Rocha PO, Katz R, Muniz JRC, Standart N and de Melo Neto OP (2005) Translation Initiation in Leishmania major: Characterisation of Multiple eIF4F Subunit Homologues. Mol Biochem Parasitol 140:23-41.

Fletcher CM, Pestova TV, Hellen CU and Wagner G (1999) Structure and interactions of the translation initiation factor eIF1. EMBO J 18:2631-7.

Groft CM, Beckmann R, Sali A and Burley SK (2000) Crystal structures of ribosome anti-association factor IF6. Nat Struct Biol 7:1156-64.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

Hershey JWB and Merrick WC (2000) Pathway and mechanism of initiation of protein synthesis. In: Sonenberg N, *et al*. (eds) Translational control of gene expression. Cold Spring Harbor Laboratory Press, New York pp 33-88.

Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, Parkhill J, Ivens AC, Rajandream MA and Barrell B. (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Res 32:D339-43.

Higgins DG and Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237-44.

Johnson PJ, Kooter JM and Borst P (1987) Inactivation of transcription by UV T. brucei provides evidence for multicistronic transcription unit including VSG gene. Cell 51:273-281.

Kapp LD and Lorsch JR (2004) The molecular mechanics of eukaryotic translation. Annu Rev Biochem 73:657-704.

Kim KK, Hung LW, Yokota H, Kim R and Kim SH (1998) Crystal structures of eukaryotic translation initiation factor 5A from Methanococcus jannaschii at 1.8 A resolution. Proc Natl Acad Sci U S A 95:10419-24.

Kumar S, Tamura K, Nei M (2004) MEGA 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. Briefings in Bioinformatics 5:150-163.

Kyrpides NC and Woese CR (1998) Universally conserved translation initiation factors. Proc Natl Acad Sci U S A 95:224-8.

Li AL, Li HY, Jin BF, Ye QN, Zhou T, Yu XD, Pan X, Man JH, He K, Yu M, Hu MR, Wang J, Yang SC, Shen BF and Zhang XM (2004) A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. J Biol Chem 279:49251-8.

Maag D, Fekete CA, Gryczynski Z and Lorsch JR (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. Mol Cell 17:265-75.

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH 2005 *CDD: a Conserved Domain Database for protein classification.* Nucleic Acids Res 33:D192-6.

Muhich ML and Boothroyd JC (1988) Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Mol Cell Biol 8:3837-46.

Olsen DS, Savner EM, Mathew A, Zhang F, Krishnamoorthy T, Phan L and Hinnebusch AG (2003) Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. EMBO J 22:193-204.

Pestova TV, Lomakin IB, Lee JH, Choi SK, Dever TE and Hellen CU. (2000) The joining of ribosomal subunits in eukaryotes requires eIF5B. Nature 403:332-5.

Roll-Mecak A, Cao C, Dever TE and Burley SK (2000) X-Ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. Cell 103:781-92.

Roll-Mecak A, Shin BS, Dever TE and Burley SK (2001) Engaging the ribosome: universal IFs of translation. Trends Biochem Sci 26:705-9.

Skeiky YA, Kennedy M, Kaufman D, Borges MM, Guderian JA, Scholler JK, Ovendale PJ, Picha KS, Morrissey PJ, Grabstein KH, Campos-Neto A and Reed SG (1998) LeIF: a recombinant Leishmania protein that induces an IL-12-mediated Th1 cytokine profile. J Immunol 161:6171-9.

Siqueira AM (2000) Subfilo Mastigophora. In Neves PD (ed) Parasitologia Humana, 10a ed, Editora Atheneu, São Paulo, pp. 27-30

Singh CR, Yamamoto Y and Asano K (2004) Physical association of eukaryotic initiation factor (eIF) 5 carboxyl-terminal domain with the

lysine-rich eIF2beta segment strongly enhances its binding to eIF3. J Biol Chem 279:49644-55.

Valasek L, Mathew AA, Shin BS, Nielsen KH, Szamecz B and Hinnebusch AG. (2003) The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. Genes Dev 17:786-99.

Valentini SR, Casolari JM, Oliveira CC, Silver PA and McBride AE (2002) Genetic interactions of yeast eukaryotic translation initiation factor 5A (eIF5A) reveal connections to poly(A)-binding protein and protein kinase C signaling. Genetics 160:393-405.

6- Anexos

O fator eIF4F, que em tripanossomatídeos apresenta múltiplos homólogos a cada subunidade, também foi estudado por este trabalho, mas a abordagem utilizada foi um estudo tridimensional de seus constituintes. Fez-se necessário, para avaliar a participação de cada subunidade do complexo eIF4F, procurar entender o porque desta multiplicidade e como eles atuam nos tripanossomatídeos. O conjunto destes resultados de modelagem, por si só, não constituem um trabalho à parte, no entanto, juntamente com resultados bioquímicos, integram trabalhos maiores e se encontram em um artigo publicado e um outro aceito para publicação (Dhalia *et al.,* 2005; Dhalia *et al*., 2006, no prelo) incluídos no final desta dissertação como apêndices. Estes resultados estão descritos a seguir.

6.1. Modelagem Molecular de Componentes do fator eIF4F em Tripanossomatídeos

Introdução

Os estudos que abordaram o fator eIF4F em tripanossomatídeos encontram uma multiplicidade de candidatos a homólogos às subunidades eIF4A, eIF4E e eIF4G (Em *L. major Lm*EIF4A1-2; *Lm*EIF4E1-4; *Lm*EIF4G1-5 e em *T. brucei* TbEIF4A1-2). Ensaios bioquímicos com várias dessas proteínas puderam verificar diferenças nos seus níveis de expressão, onde algumas eram expressas em quantidades concordantes com os demais homólogos dos eucariotos enquanto que outras eram expressas em níveis menores. Também foram realizados ensaios de ligação aos respectivos parceiros e esses ensaios também apresentaram diferenças na dinâmica das diversas proteínas (Dhalia R, 2005). No entanto, neste momento, não foi viável realizar ensaios bioquímicos estruturais com essas diversas proteínas. A alternativa encontrada para se obter informações estruturais sobre essas proteínas foi a construção de modelos computacionais daquelas proteínas que apresentaram melhores resultados em estudos anteriores (Dhalia R *et al*., 2005; Dhalia *et al*., no prelo).

Metodologia

A primeira etapa para a criação dos modelos foi a obtenção de homólogos em bancos de dados. Inicialmente foi utilizada a ferramenta BLASTP do NCBI (*National Center for Biotechnology Information*). O BLASTP realiza uma procura nos bancos de dados de proteínas por seqüências de aminoácidos que possuam homologia seqüencial com a proteína alvo. Uma vez obtida as seqüências homólogas, analisa-se quais

possuem melhores alinhamentos e quais possuem estrutura resolvida. Dentre as diversas seqüências que o BLASTP fornece como saída, é escolhida como molde aquela que possuir maior similaridade com a proteína alvo e que possua estrutura resolvida. Se a identidade seqüencial encontrada for superior a 30% pode-se iniciar a modelagem. Porém, quando a identidade seqüencial encontrada mostra-se inferior a 30%, é necessário o uso de programas que levem em consideração a predição de estrutura secundária para obter um melhor alinhamento. Para esse alinhamento estrutural utiliza-se o PSI-BLAST do NCBI e o GenThreader (Jones, 1999).

Uma vez escolhida a seqüência molde, o banco de dados do PDB (*Protein Data Base*) (Westbrook *et al*., 2002) foi utilizado para a obtenção do arquivo com as coordenadas da estrutura resolvida. O alinhamento escolhido foi então refinado e utilizado como entrada no programa Modeller6a (Sali e Blundell, 1993) para a obtenção dos modelos. Para este trabalho foram feitos 50 modelos inicialmente.

Em seguida, os modelos foram validados seguindo alguns parâmetros. O primeiro parâmetro utilizado foi a avaliação das pseudo-energias de cada modelo fornecidas pelo programa Modeller6a. Por esse parâmetro, mantêm-se os cinco modelos de menor pseudo-energia e os demais são descartados. Em seguida foram utilizados programas específicos para validar o modelo, utilizando-se como parâmetros o ambiente químico, a estereoquímica e contato atômico. Os programas usados nesse passo, respectivamente, foram: VERIFY 3D (Luthy *et al*., 1992), PROCHECK (Laskowski *et al*., 1998) e WHATIF (Vriend 1990; Hooft *et al.* 1996). Nos casos em que os modelos obtidos não passaram pelos parâmetros estabelecidos por cada programa, realizou-se a construção de novos modelos. Esses novos modelos passaram pelas mesmas etapas de validação. Até a obtenção de um modelo que respeitasse os parâmetros definidos.

Com o modelo validado, foi possível estudar a sua estrutura e compará-la com a do molde. Esse estudo foi realizado com o auxílio do programa O (Jones *et al*., 1991). Dentre as diversas funções do O, temos a de calcular as distâncias entre átomos e a de rotacionar estruturas. Quando necessário, utilizou-se uma biblioteca de rotâmeros para melhorar o direcionamento de determinados aminoácidos em relação ao molde e a função por ele exercida. Outro programa utilizado foi o GRASP (Nicholls *et al*., 1991), que permite o cálculo do potencial eletrostático da superfície de uma proteína. Com base nas análises realizadas com auxílio desses programas, foram criadas figuras dos modelos através do programa PyMol (DeLano WL 2002).

Resultados

Foi possível a obtenção de um modelo do *Lm*EIF4E1, um do *Lm*EIF4A1, dois do *Tb*EIF4A1, dois do *Tb*EIF4A2 e um do domínio central do *Lm*EIF4G3. O modelo do *Lm*EIF4E1 teve como molde o eIF4E de *Mus musculus* (Marcotrigiano *et al*., 1997). O modelo do domínio central do *Lm*EIF4G3 foi obtido usando-se como molde o domínio central do eIF4G2 de *Homo sapiens* (Marcotrigiano *et al*., 2001). Os modelos do *Lm*EIF4A1, *Tb*EIF4A1 e *Tb*EIF4A2 foram obtidos usando-se como molde o eIF4A de *Saccharomyces cerevisiae* (Caruthers *et al*., 2000). Os dois homólogos de *Trypanosoma brucei* foram modelados uma segunda vez utilizando como molde a proteína Dhh1, pertencente a mesma família de RNA helicases do tipo DEAD Box que o eIF4A, de *Saccharomyces cerevisiae* (Cheng *et al*., 2005).

Modelagem do *Lm*EIF4E1

Estudos anteriores revelaram que a estrutura tri-dimensional do eIF4E possui uma conformação de "mão em concha". Ele é formado por três α-hélices longas, uma α-hélice curta e por oito folhas β-antiparalelas. O reconhecimento do cap se dá através da interação de dois triptofanos (no modelo do *Lm*EIF4E1 os amino ácidos W37 e W83) do eIF4E que fazem um "sanduíche" da base nitrogenada do cap. Essa interação recebe o nome de π-π *stacking*. Além desses dois triptofanos, uma série de outros aminoácidos interagem com o cap formando ligações de hidrogênio, contatos de van der Waals, interações mediadas por moléculas de água e pontes salinas. Dentre esses aminoácidos temos, no modelo, W87, E84, K93, R167, R172 e W176. O Q71 (em mamíferos um D) participa indiretamente da ligação ao cap ancorando R167 (Marcotrigiano *et al*. 1997). Foi possível obter um modelo do *Lm*EIF4E1 que pode ser visualizado na figura 1.

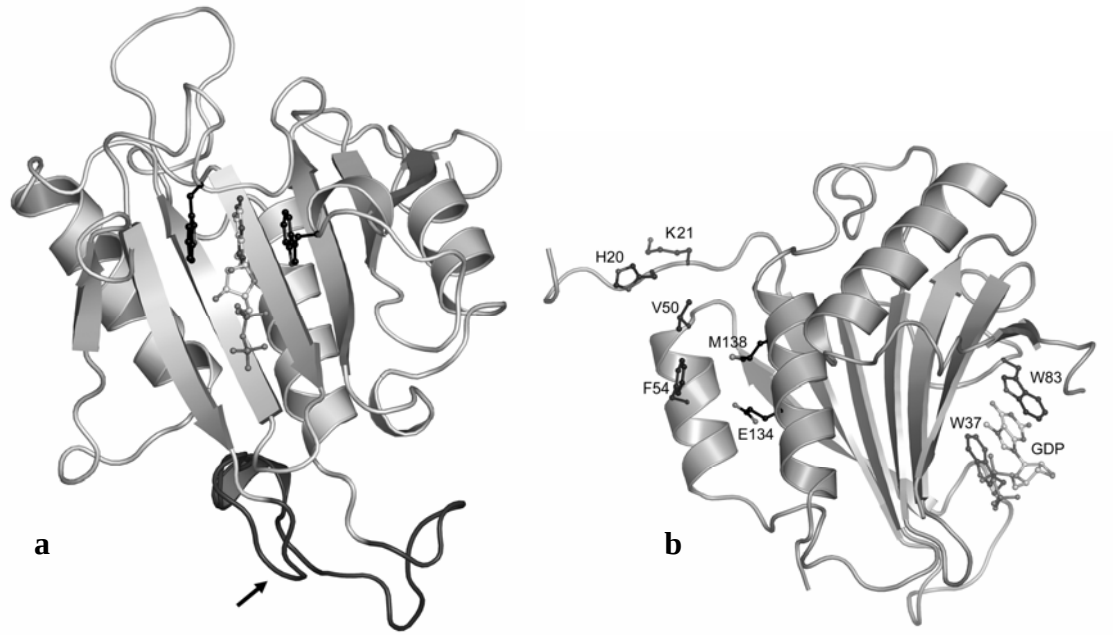


Figura 1: Visões gerais do modelo obtido para o *Lm*EIF4E1. a: A seta indica uma alça que não teve um resultado satisfatório na modelagem, sendo omitida da análise. É possível visualizar os triptofanos emparelhando a base nitrogenada do GDP. b: Visão rotacionada em 90°. À direita se encontram os triptofanos emparelhando o GDP e à esquerda estão indicados os resíduos que interagem com o eIF4G.

O modelo do *Lm*EIF4E1 apresentou a maioria dos aminoácidos relacionados com a interação ao cap conservados (W56, W102, E103, R157 em mamíferos). As mutações de R112 de mamíferos para K93 no *Lm*EIF4E1 e de K162 em *M. musculus* para R172 em *L. major* também ocorrem no homólogo do eIF4E de *Schizosaccharomyces pombe*. A única mutação não conservada em outros organismos é o D90 em mamíferos para Q71 no homólogo de *L. major* estudado. Entretanto, mesmo nesse caso, as interações atômicas demonstraram-se possíveis, uma vez que se trata de uma ponte salina entre o oxigênio do Q e o NH2 do R (Figura 02).

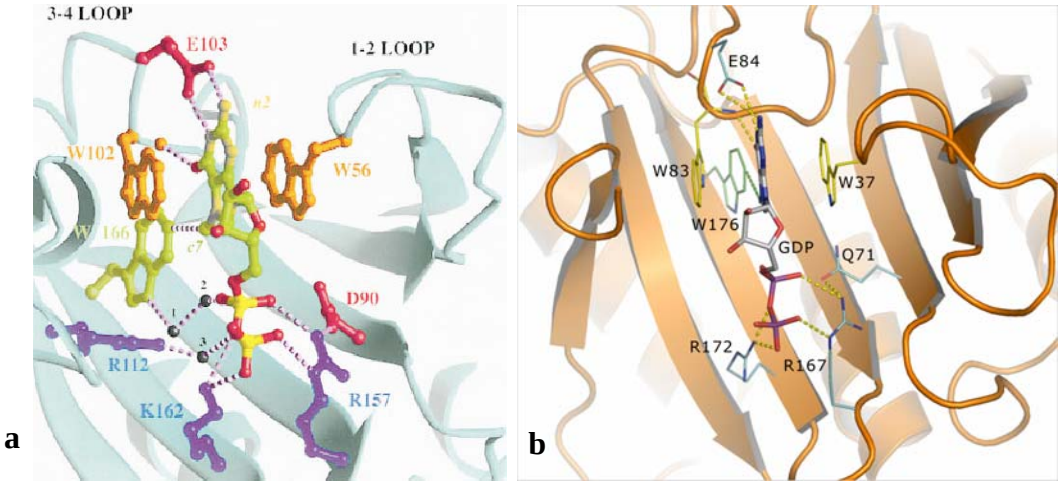


Figura 02: Comparação entre o sítio de ligação da estrutura modelada (direita) e o da estrutura resolvida usada como molde (esquerda). A partir desta figura pode-se perceber que as interações exercidas por W56, W102, E103, R157, K162, W166 e D90 do molde estão conservadas nos aminoácidos W37, W83, E84, R167, R172, W176 e Q71 da estrutura modelada. O As interações realizadas através de moléculas de água foram omitidas. Dessa forma, o K93, que corresponde a R112, não está visível no molde (PAinel A retirado de Marcotrigiano *et al*., 1997).

O sítio de ligação do cap é uma fenda na superfície do eIF4E. O estudo do potencial eletrostático de superfície da fenda do *Lm*EIF4E1, revelou uma marcante complementaridade de cargas com o cap (no caso, foi usado um análogo ao cap, o 7-metil-GDP). Este resultado é concordante com o potencial eletrostático observado na mesma região da

superfície da estrutura resolvida do eIF4E de *M. musculus* como ilustrado na figura 03 (Marcotrigiano *et al*., 1997).

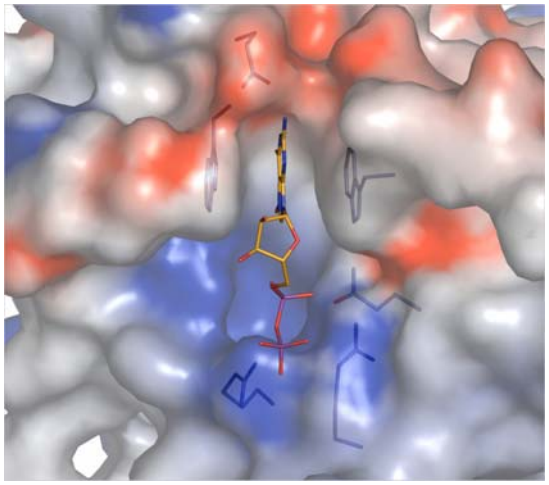


Figura 03: Potencial eletrostático do sítio de ligação ao cap do *Lm*EIF4E1. Os aminoácidos destacados são os mesmos dos da figura 02 com a exceção do W166. Como é possível observar, a complementaridade de cargas é marcante entre o sítio e o cap. Em vermelho temos regiões de potencial eletrostático negativo, em azul positivo e em branco neutro.

Além da ligação ao cap, o eIF4E interage, em sua face oposta, com o eIF4G, onde também ocorrem aminoácidos conservados e, dentre eles, tem-se H37, P38, Q40, V69, W73, L131, E132 e L135 em mamíferos. A família 4EBP (4E *Binding Protein*) é uma inibidora da iniciação da tradução que impede a interação do eIF4E com o eIF4G. Eles disputam com o eIF4G o sítio de ligação no eIF4E e sua ligação é um pouco mais intensa, pois o 4EBP interage com dois resíduos a mais do eIF4E que são E140 e D147. O 4EBP se desliga do eIF4E ao ser fosforilado uma vez que surge uma repulsão eletrostática com um círculo acídico do eIF4E que possui como centro os aminoácidos E70 e D71 (Marcotrigiano *et al*., 1999).

No modelo do *Lm*EIF4E1 temos que os resíduos H16, Q19, V50, E51 e E143 estão conservados quando comparados com o molde. Já o K17, E52 e M138 são, em mamífero P, D e L, respectivamente. Essas mutações

já foram encontradas em outras espécies (gérmen de trigo para o mutante P17K; *S. cerevisiae*, *S. pombe* e *Triticum aestivus* para o mutante D51E; *Xenopus laevis* e *T. aestivus* para o mutante L138M).

As mutações F54 que deveria ser um W e E134 que deveria ser um L poderiam, numa primeira análise, dificultar a ligação do eIF4G e do 4EBP. Todavia é interessante notar que algumas das ligações entre o eIF4E e o eIF4G se dão por interações intermoleculares, onde o volume total poderia estar envolvido. Se for o caso, a perda de volume na mutação da tríade WLL (uma vez que o W tornou-se um F) poderia ser compensada pela tríade FEM (pois o segundo L estaria mutado por um E). Mesmo que não seja o caso, essas mutações poderiam refletir mutações existentes no sítio de ligação ao eIF4E dos eIF4Gs de *L. major*, o que representaria um motivo de ligação ao eIF4E não canônico. Essa hipótese tem como respaldo o fato de que o motivo YXXXXLϕ não estar conservado nos candidatos a homólogos a eIF4Gs de *L. major* identificados até o momento (figura 04).

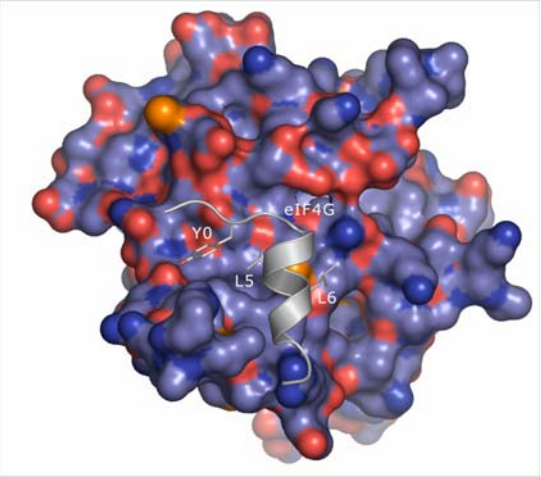


Figura 04: Representação do raio de Van der Waals da superfície do *Lm*EIF4E1. Em azul claro temos átomos de carbono, em vermelho os de oxigênio, em azul escuro os de nitrogênio e em laranja os de enxofre. Um fragmento do eIF4G, mais especificamente seu sitio de ligação ao eIF4E, está ilustrado em cinza com os resíduos de tirosina (Y) e leucina (L) do motivo YXXXXLL de mamíferos destacados.

O estudo do potencial eletrostático de superfície do *Lm*EIF4E1, leva a crer que a regulação deste por uma proteína similar ao 4EBP seja possível. A repulsão eletrostática gerada pelo círculo acídico E70 e D71 em mamíferos, quando a 4EBP é fosforilada, pode ocorrer no eIF4E de *L. major* devido à presença da dupla E51 e E52 (figura 05). No entanto, não existe indícios da existência de homólogos ao 4EBP em tripanossomatídeos. Finalmente, temos que o *Lm*EIF4E1 possui uma longa inserção (22 aminoácidos) que não pode ser modelada pelos atuais programas de modelagem. Faz-se necessário um estudo mais refinado dessa região com o intuito de avaliar um possível papel na iniciação da tradução de *L. major*.

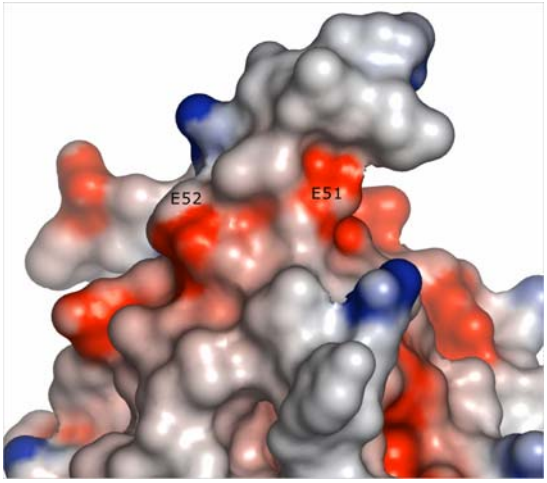


Figura 05: Representação da superfície eletrostática do *Lm*EIF4E1. Os E51 e E52 estão destacados. Em vermelho temos regiões de potencial eletrostático negativo, em azul positivo e em branco neutro.

Além do modelo do *Lm*EIF4E1, iniciou-se a produção de modelos para mais dois homólogos do eIF4E de *L. major*, mais especificamente o *Lm*EIF4E2 e o *Lm*EIF4E3. No entanto, características destes homólogos desencorajaram uma abordagem de modelagem. O *Lm*EIF4E2 possui uma inserção de aproximadamente vinte aminoácidos dentro de uma folha beta. O *Lm*EIF4E3 possui uma longa extremidade amino-terminal, com

mais de 100 aminoácidos e possui um dos tritofanos responsáveis pelo π-π *stacking* do GTP mutado por uma metionina.

Modelagem do *Lm*EIF4G3

Até o momento, não foi possível obter a estrutura completa do fator eIF4G, mas determinados domínios, como o domínio central de ligação ao eIF4A (MIF4G), foram estruturalmente resolvidos. A estrutura do domínio central demonstrou que ele pertencia à família HEAT (*Huntingtin, Elongation factor 3*, *A subunit of protein phosphatase 2A [PP2A], and Target of rapamycin*). A família HEAT consiste de proteínas que possuem um conjunto de repetições de pares de α-hélices antiparalelas que apesar de possuir um arranjo tridimensional conservado, não possuem seqüências de aminoácidos conservadas (Marcotrigiano *et al*., 2001).

A partir deste molde, o *Lm*EIF4G3, que foi o único na ocasião que através de ensaios bioquímicos se ligava ao eIF4A de mamíferos e de *L. major*, teve sua metade inicial modelada. Isso se deve ao fato de apenas o domínio de ligação ao eIF4A deste fator de mamíferos possuir estrutura cristalográfica resolvida. Duas regiões do *Lm*EIF4G3, referentes às alças compreendidas entre os aminoácidos 128~140 e 161~178, não puderam ser modeladas, pois a estrutura usada como molde não teve uma densidade eletrônica boa o suficiente para resolver a estrutura nessas posições (figura 06).

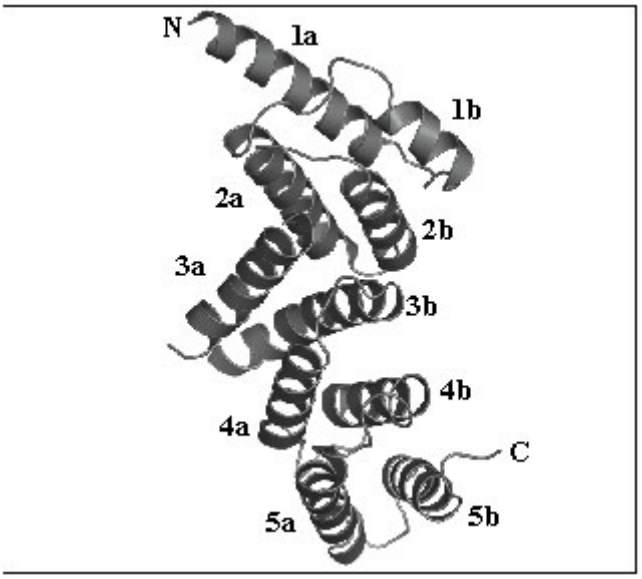


Figura 06: Vista geral do modelo do *Lm*EIF4G3. A superfície côncava do crescente se encontra à direita e a convexa à esquerda. Os pares de α-hélices estão demarcados, assim como a extremidade amino e carboxi-terminal.

O modelo adquiriu a estrutura em crescente do molde, com cinco pares de α-hélices antiparalelas que formam a dupla camada côncava e convexa pelo conjunto de hélices A e B, respectivamente. Uma diferença marcante entre o modelo de *Leishmania* e o molde de mamíferos foi percebida pela análise do potencial eletrostático de superfície dessas moléculas ao ser calculado pelo programa GRASP (figura 07). Esta diferença se concentra na região que na proteína humana foi relacionada com a interação a um sítio de ligação independente ao ribossomo utilizado por vírus. Na estrutura de humano essa região encontra-se carregada positivamente, enquanto que no *Lm*EIF4G3 a região equivalente apresenta-se carregada negativamente. Por outro lado, a região que envolve a interação com o eIF4A mostrou um potencial eletrostático de superfície similar ao molde e todos os aminoácidos que participam da interação (R61, H64, K69, L70, R270 e F273) encontram-se posicionados adequadamente para exercerem suas funções.

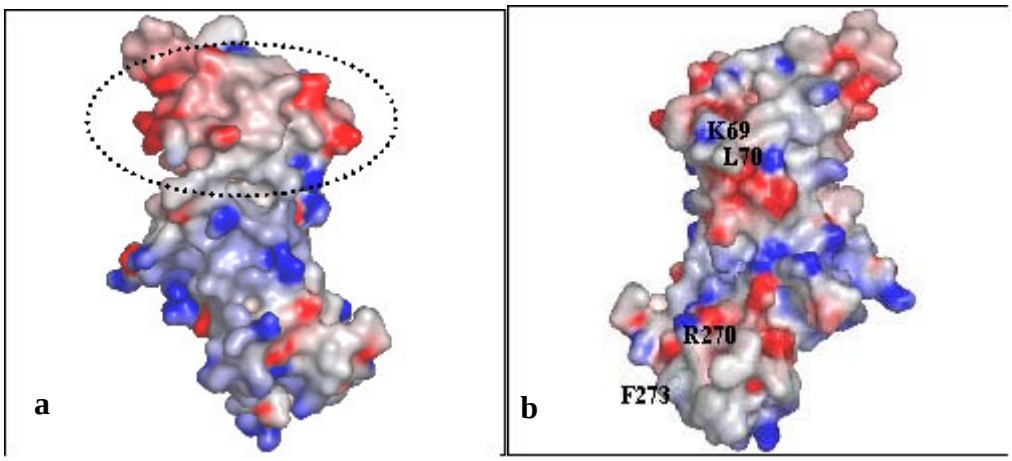


Figura 07: Potencial eletrostático do modelo do *Lm*EIF4G3. O painel (A) e a figura 07 representam vistas idênticas do modelo enquanto que o painel (B) mostra uma vista rotacionada em 180°. As regiões demarcadas em vermelho são negativas, as azuis positivas e as brancas neutras. No painel (A) o circulo pontilhado delimita a região do LmEif4G3 que difere do molde de H. sapiens. No painel (B) estão demarcados os resíduos R61, H64, K69, L70, R270 e F273 que participam da ligação ao eIF4A.

Modelagem do *Lm*EIF4A1

A estrutura tridimensional do eIF4A revela dois domínios estruturais compactos, que representam a porção amino e carboxi terminal da proteína. Os dois domínios possuem uma organização com folhas-β e α-hélices paralelas conectados por um *link* flexível. Dos nove motivos presentes no eIF4A seis (Q, I, Ia, Ib, II e III) se encontram na porção amino-terminal e os outros três (IV, V e VI) na carboxi-terminal. Foi possível encontrar todos os motivos no molde do *Lm*EIF4A1 (figura 08).

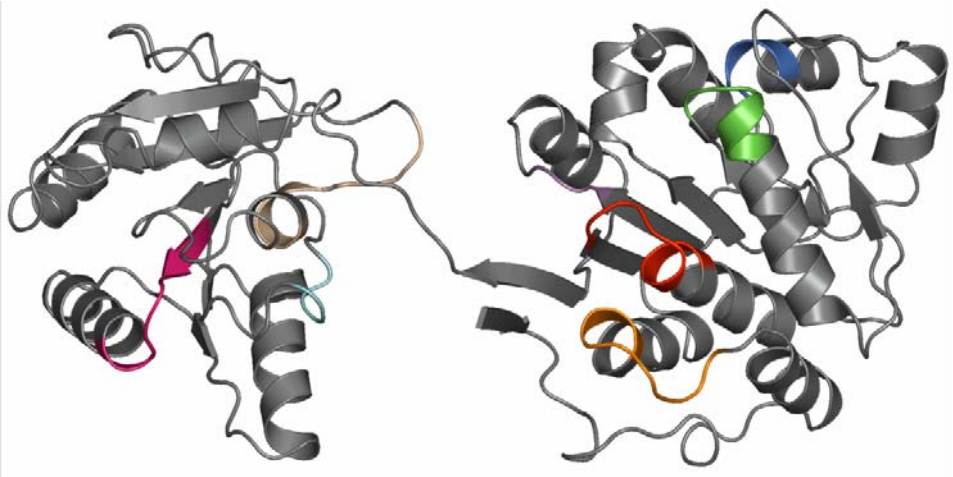
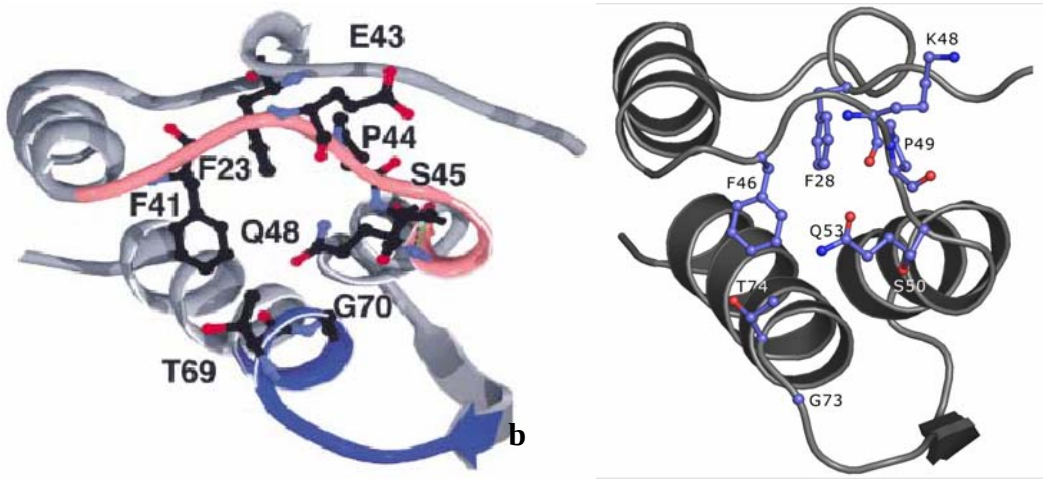


Figura 08: Modelo do *Lm*EIF4A1 e seus diversos motivos. O motivo Q está de laranja; o I (Walker A) de vermelho; o Ia de verde; o Ib em azul; o motivo II (Walker B) situa-se entre o Walker A e o Ia (na alça entre a folha-β e a α-hélice); o motivo III em violeta, acima do Walker A; o motivo IV está em rosa; o motivo V em ciano; o motivo VI em bege.

O motivo I é responsável pela ligação do ATP e sua atividade é regulada pelo motivo Q. Dentre os aminoácidos envolvidos nessa interação temos F46, S50, Q53, T74, G75 conservados com o de mamífero. O K48 em mamíferos é um E. Todavia, como esse resíduo interage através de sua cadeia principal, portanto, a mutação não interfere na ligação ao ATP (figura 09). O motivo II (caixa DEAD) está envolvido com a hidrólise do ATP em ADP. Foi proposto que o motivo Q regule a ligação ao ATP e que a alça onde se encontram T74 e G75 tenha uma certa mobilidade. Quando fechado, a ligação do ATP não seria possível. Quando este abrisse haveria uma mudança conformacional do sítio, com a mudança de direcionamento de F46, e a ligação ao ATP se tornaria possível. Os demais domínios participam também dessas funções e também nas de ligação a RNA e atividade helicase. As funções de ligação e hidrólise do ATP estão mais relacionadas aos domínios da extremidade amino-terminal enquanto que a ligação a RNA e a atividade helicase estão associadas com a extremidade carboxi-terminal. No entanto, mutações em qualquer uma das extremidades são capazes de

perturbar qualquer um dos papeis do eIF4A (Benz *et al*., 1999; Caruthers *et al*., 2000; Tanner *et al*., 2003).



a **b**

Figura 09: Comparação entre o sítio de ligação ao ATP do molde (A) e do LmeIF4A1 (B). Nesta figura, a alça do T74 encontra-se fechada e o F46 está desorientado. (Painel A - Tanner *et al*., 2003).

Prosseguindo com a avaliação do sítio de ligação ao ATP, calculou-se o potencial eletrostático da superfície do LmeIF4A1. Pode-se constatar que o sítio se encontrava numa região muito negativa do modelo. Algumas poucas áreas positivas encontravam-se nas proximidades como se vê na figura 10. Esse padrão permite que a adenina se ancore fortemente na região positiva. Numa primeira análise, a falta de regiões positivas para ancorar os fosfatos desestabilizaria essa ligação. Todavia, deve-se levar em consideração que a ligação entre o eIF4A e o ATP é breve (apenas para a hidrolise em ADP). Outro fato importante é que o *Lm*EIF4A1, assim como ocorre com o seu homólogo de mamíferos, deve passar por mudanças conformacionais quando o ATP e o RNA estão ligados a eles.

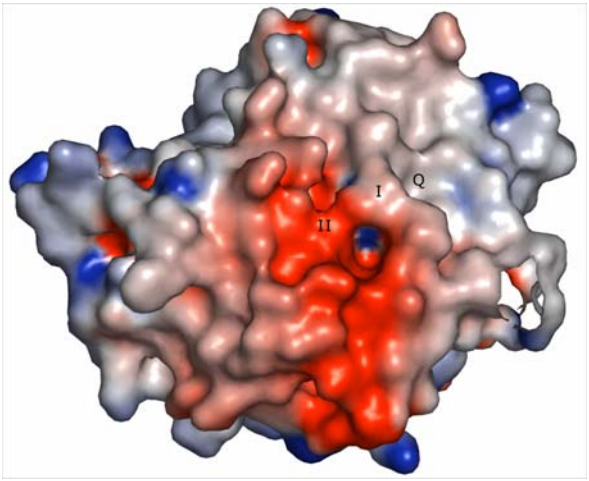


Figura 10: Representação do potencial eletrostático da superfície amino-terminal do *Lm*EIF4A1. Estão destacados os motivos Q, Walker A (I), e Walker B (II). Em vermelho temos regiões de potencial eletrostático negativo, em azul positivo e em branco neutro.

Modelagem dos *Tb*EIF4A1-2

Evidências bioquímicas revelaram que o *Tb*EIF4A1 e o *Tb*EIF4A2 apresentam localização intracelular distintas e estão presentes em níveis celulares distintos. Estes resultados sugerem que possuam funções celulares também distintas. Aliando os dados bioquímicos com estudos de aminoácidos diagnósticos, obtidos pela comparação da seqüência primária dessas duas proteínas contra o eIF4AI e eIF4AIII de outros organismos, o *Tb*EIF4A1 estaria mais próximo do eIF4AI enquanto que o *Tb*EIF4A2 estaria mais próximo do eIF4AIII. Com o intuito de ajudar na avaliação desta hipótese foi criado um segundo modelo estrutural para as proteínas *Tb*EIF4A1 e *Tb*EIF4A2 tendo como molde a Dhh1. Esta proteína foi escolhida para essa segunda análise por apresentar, em sua estrutura tridimensional, uma compactação maior dos dois domínios quando comparada com a estrutura do eIF4A de levedura que possui os dois domínios bem segregados. Uma segunda diferença entre os dois moldes é que o primeiro apresenta o sítio de ligação ao ATP com a alça fechada e a fenilalanina desorientada enquanto que o cristal da Dhh1 possui este sítio

com a alça aberta e com a fenilalanina orientada. Apesar dessas diferenças importantes, a Dhh1, assim como o eIF4A, é uma proteína da família DEADBox e possue um grau de similaridade seqüencial de mais de 50% com os demais contituintes desta família protéica, o que permite seu uso como molde. De acordo com a literatura, esta conformação mais compacta, propiciando a interação entre os dois domínios da proteína, com o sítio de ligação ao ATP apresentando alça aberta são características da forma ativa desta família de helicases.

Ao analisar os dois modelos obtidos, observou-se que na interface entre os domínios amino e carboxi-terminal tem-se uma diferença importante entre *Tb*EIF4A1 e *Tb*EIF4A2. No domínio carboxi-terminal do segundo temos um triptofano, enquanto que no primeiro temos uma valina. Esses dois aminoácidos, que pertencem ao motivo V, apontam para um mesmo bolsão no domínio carboxi-terminal. Para cada átomo livre da valina e do triptofano, calculou-se uma esfera de raio de quatro Angstrons que representaria a distância máxima para interações fracas. O triptofano possui um número muito maior de interações putativas que a valina, isso poderia estar relacionado com uma maior estabilidade da estrutura compactada no homólogo *Tb*EIF4A2, enquanto que no *Tb*EIF4A1, esta conformação compactada seria mais lábil (figura 11). Isso é concordante com a proposta inicial do *Tb*EIF4A1 ser homólogo ao fator de iniciação e o *Tb*EIF4A2 ser homólogo ao componente do EJC. É sabido que ao longo do *scanning*, o eIF4A do eIF4F é reciclado diversas vezes e, para realizar seu papel, teria de alternar entre a conformação aberta e fechada; já no EJC, a ligação ao mRNA seria muito mais estável sendo importante a manutenção da conformação fechada. Ainda em relação a interface dos domínios, o *Tb*EIF4A2 possui nessa região uma preponderância de cargas positivas, ao contrário do *Tb*EIF4A1 que possui um maior número de cargas neutras. Esse excesso de cargas positivas no *Tb*EIF4A2 pode servir para ancorar os fosfatos do mRNA de forma muito mais estável que no *Tb*EIF4A1.

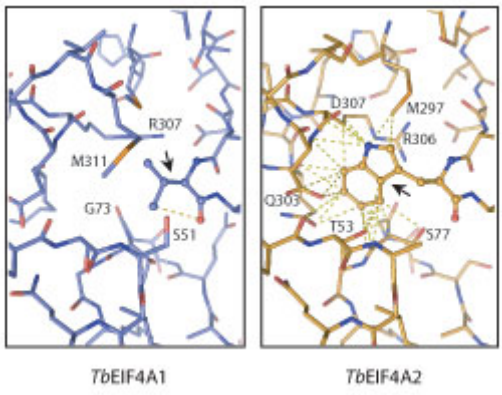


Figura 11: Vista da interface entre os domínios carboxi e amino terminal dos modelos *Tb*EIF4A1 e *Tb*EIF4A2. A valina do *Tb*EIF4A1 e o triptofano do *Tb*EIF4A2 estão indicados pelas setas. As ligações putativas num raio de quatro Angstrons estão representadas pelas linhas pontilhadas e os resíduos envolvidos estão destacados.

Além destas diferenças encontradas na interface entre os dois domínios estruturais, outras mutações interessantes no modelo do *Tb*EIF4A2 também foram mapeadas próximas às hélices 5 (N/KV93R, Q/R139E/G, Q/A146D/E) e 10 (L256F, E264D), nas extremidades amino e carboxi-terminal respectivamente. Esses resíduos apresentaram-se em sua maioria expostos ao solvente e são, em sua maioria, mais polares. O calculo do potencial eletrostático dos modelos revela que o *Tb*EIF4A2 possui uma superfície mais carregada que a do *Tb*EIF4A1. É interessante notar que a hélice 10 está relacionada com a ligação ao eIF4G e mutações nessa hélice impedem a ligação entre o eIF4AI e o eIF4G. De modo similar, é possível que as mutações na extremidade amino-terminal, mapeadas na hélice 5, estejam envolvidas em interações proteína-proteína, uma vez que se sabe que tanto o eIF4AI e o eIF4AIII de mamíferos possuem outros parceiros que não possuem sítios de interação definidos (figura 12).

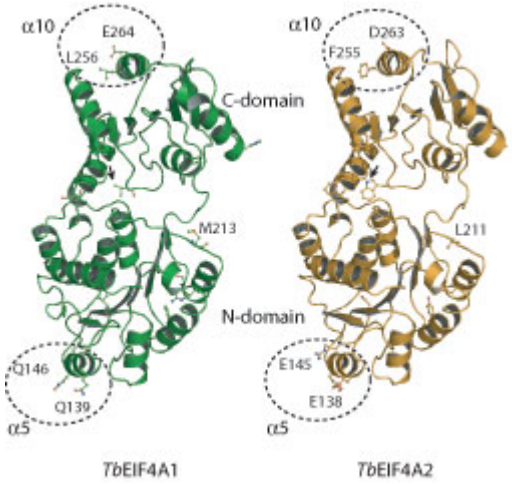


Figura 12: Visão geral dos modelos do *Tb*EIF4A1 e do *Tb*EIF4A2. Os resíduos com mutações de intesse estão destacados, assim como a região de ligação ao eIF4G demarcada pelo círculo pontilhado próximo à hélice α10. Um segundo sítio de interação proteína-proteína foi proposto para a região envolvendo a hélice α5 e também está demarcada por um círculo pontilhado.

Conclusão

Foi possível a criação de um modelo para o *Lm*EIFA1, um para o *Lm*EIF4E1, um para o domínio HEAT do *Lm*EIF4G3 e dois modelos para o *Tb*EIF4A1 e *Tb*EIF4A2. O modelo do *Lm*EIF4E1 demonstra que ele possui todas as características necessárias para um eIF4E funcional. Os resultados obtidos para o *Lm*EIF4G3 dão suporte ao seu papel na iniciação da tradução, no entanto a falta de um sítio de ligação ao eIF4E e as diferenças puntuais do sítio de ligação ao RNA indicam que esta proteína ou possua um papel regulatório na síntese protéica ou exerça sua função por mecanismos distintos aos observados nos demais eucariotos. Os resultados dos modelos do *Lm*EIFA1, *Tb*EIF4A1 e *Tb*EIF4A2 Sugerem que o homólogo dos tripanossomatídeos ao eIF4AI de mamíferos seja o *Lm*EIFeA1 e o *Tb*EIF4A1, enquanto que o *Tb*EIF4A2 estaria relacionado com o eIF4AIII de mamíferos. Aliando os resultados dos ensaios bioquímicos foi possível propor que o eIF4AIII é uma proteína bastante conservada ao longo da linhagem evolutiva dos eucariotos. (Dhalia *et al*., no prelo).

Bibliografia

Benz J, Trachsel H and Baumann U (1999) Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae--the prototype of the DEAD box protein family. Structure Fold Des. 7(6):671-9.

Caruthers JM, Johnson ER and McKay DB (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proc Natl Acad Sci U S A. 97(24):13080-5.

Cheng Z, Coller J, Parker R and Song H (2005) Crystal structure and functional analysis of DEAD-box protein Dhh1p. RNA, 11,1258-1270.

DeLano WL (2002) The PyMOL Molecular Graphics System. Available from World Wide Web http://www.pymol.org

Dhalia R, Reis CRS, Freire ER, Rocha PO, Katz R, Muniz JRC, Standart N and de Melo Neto OP (2005) Translation Initiation in *Leishmania major*: Characterisation of Multiple eIF4F Subunit Homologues. Mol Biochem Parasitol. 140(1):23-41.

Dhalia R, Marinsek N, Reis CRS, Katz R, Muniz JRC, Standart N, Carrington M and de Melo Neto OP (2006) The two eIF4A helicases in Trypanosoma brucei are functionally distinct. Manuscrito aceito para publicação na revista Nucleic Acid Research.

Hooft RW, Sander C and Vriend G (1996) Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins. 26(4):363-76.

Jones, TA, Zou JY, Cowan SW and Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models.” Acta Cryst A47, 110-119.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 292(2):195-202.

Laskowski RA, MacArthur MW and Thornton JM (1998) Validation of protein models derived from experiment. Curr Opin Struct Biol. 8(5):631-9.

Luthy R, Bowie JU and Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature. 356(6364):83-5.

Marcotrigiano J, Gingras AC, Sonenberg N and Burley SK (1997) Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell. 89(6):951-61.

Marcotrigiano J, Gingras AC, Sonenberg N and Burley SK (1999) Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. Mol Cell. 3(6):707-16.

Marcotrigiano J, Lomakin IB, Sonenberg N, Pestova TV, Hellen CUT and Burley SK (2001) A Conserved HEAT Domain within eIF4G Directs Assembly of the Translation Initiation Machinery. Mol Cell. 7:193-203.

Nicholls A, Sharp KA and Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins.11(4):281-96.

Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

Tanner NK and Linder P (2001) DExD/H Box RNA Helicases: From Generic Motors to Specific Dissociation Functions. Molecular Cell. 8: 251-262.

Tanner NK, Cordin O, Banroques J, Doère M and Linder P (2003) The Q Motif: A Newly Identified Motif in DEAD Box Helicases May Regulate ATP Binding and Hydrolysis. Molecular Cell. 11: 127-138.

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph. 8(1):52-6, 29.

Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, Ravichandran V, Gilliland GL, Bluhm W, Weissig H, Greer DS, Bourne PE and Berman HM (2002) The Protein Data Bank: unifying the archive. Nucleic Acids Res. 30(1):245-8.

6.2. Números de acesso utilizados.

Proteína	Organismo	Accesión
eIF1	L*eishmania* major	CAJ04877
	Trypanosom brucei	EAN79547
	Trypanosom cruzi	EAN97939
	Plasmodium falciparum	AAN36503
	Neospora caninum	AAF76883
	Dictyostelium discoideum	EAL70012
	Entamoeba histolytica	EAL45610
	Homo sapiens	AAH08710
	Mus musculus	P48024
	Saccharomyces cerevisiae	NP_014155
	Schizosaccharomyces pombe	CAA22621
	Candida glabrata	CAG60992
	Triticum aestivum	AAM34279
	Coffea arabica	CAD58629
	Oryza sativa	XP_478516
	Schistosoma japonicum	AAW25113
	Methanocaldococcus jannaschii	NP_247438
	Xenopus laevis	AAL78005
	Danio rerio	AAQ97785
	Ixodes scapularis	AAY66832
	Apis mellifera	XP_392601
	Aedes aegypti	AAV69394
eIF1A	*Leishmania major*	CAJ03431
	Trypanosoma brucei	AAX69318
	Trypanosoma cruzi	EAN94505
	Arabidopsis thaliana	NP_851095
	Homo sapiens	AAH67851
	Saccharomyces cerevisiae	AAS56290
eIF5	*Leishmania major*	CAJ07736

	Trypanosoma brucei	EAN77702
	Trypanosoma cruzi	EAN83191
	Arabidopsis thaliana	AAL07057
	Homo sapiens	NP_892116
	Saccharomyces cerevisiae	NP_015366
eIF5A	*Leishmania major*	CAJ04915
	Trypanosoma brucei	EAN79055
	Trypanosoma cruzi	EAN99705
	Arabidopsis thaliana	F86272
	Homo sapiens	AAH80196
	Saccharomyces cerevisiae	NP_010880
eIF5B	*Leishmania major*	CAJ06767
	Trypanosoma brucei	AAX79619
	Trypanosoma cruzi	EAN83255
	Arabidopsis thaliana	NP_177807
	Homo sapiens	O60841
	Saccharomyces cerevisiae	P39730
eIF6	*Leishmania major*	CAJ09042
	Trypanosoma brucei	EAN77947
	Trypanosoma cruzi	EAN98181
	Arabidopsis thaliana	AAP75806
	Homo sapiens	P56537
	Saccharomyces cerevisiae	AAT92935
eIF3a	*Arabidopsis thaliana*	NP_192881
	Homo sapiens	Q14152
	Saccharomyces cerevisiae	NP_009635
eIF3b	*Leishmania major* (17)	CAJ04010
	Leishmania major (16)	CAJ03612
	Trypanosoma brucei	AAX79391
	Trypanosoma cruzi	EAN95506
	Arabidopsis thaliana	AAF67758

	Nicotiana tabacum	P56821
	Homo sapiens	EAL23952
	Pan troglodytes	XP_527644
	Mus musculus	AAH31704
	Rattus norvegicus	AAH98728
	Canis familiaris	XP_862108
	Xenopus tropicalis	NP_001016724
	Caenorhabditis elegans	NP_001022469
	Strongylocentrotus purpuratus	XP_796053
	Saccharomyces cerevisiae	NP_015006
	Aspergillus fumigatus	XP_749953
eIF3c	*Leishmania major*	CAJ09680
	Trypanosoma brucei	EAN78235
	Trypanosoma cruzi	EAN99318
	Arabidopsis thaliana	O49160
	Homo sapiens	AAH71705
	Saccharomyces cerevisiae	NP_014040
eIF3d	*Leishmania major*	CAJ06710
	Trypanosoma brucei	AAX79274
	Trypanosoma cruzi	EAN91953
	Arabidopsis thaliana	P56820
	Homo sapiens	CAG30375
eIF3e	*Leishmania major*	CAJ05687
	Trypanosoma brucei	EAN80110
	Trypanosoma cruzi	EAN90796
	Arabidopsis thaliana	NP_567047
	Homo sapiens	CAG33310
eIF3f	*Leishmania major*	CAJ08505
	Trypanosoma brucei	EAN78821
	Trypanosoma cruzi	EAN85256
	Arabidopsis thaliana	NP_181528

	Homo sapiens	CAG33240
	Mus musculus	AAH83190
	Gallus gallus	XP_421624
	Ixodes scapularis	AAY66837
Proteassomo	*Homo sapiens*	BAA08780
	Oryza sativa	BAB78487
	Arabidopsis thaliana	AAG50979
	Drosophila melanogaster	P26270
	Drosophila pseudoobscura	EAL26566
	Apis mellifera	XP_391960
	Caenorhabditis elegans	NP_491319
	Caenorhabditis briggsae	CAE66740
eIF3g	*Arabidopsis thaliana*	CAC01929
	Homo sapiens	CAG33415
eIF3h	*Homo sapiens*	AAC84044
eIF3i	*Leishmania major*	CAJ09354
	Trypanosoma brucei	EAN79913
	Trypanosoma cruzi	EAN94173
	Arabidopsis thaliana	NP_850450
	Homo sapiens	AAC97144
eIF3k	*Leishmania major*	CAJ08745
	Trypanosoma brucei	EAN80488
	Trypanosoma cruzi	EAN90599
	Arabidopsis thaliana	NP_195051
	Homo sapiens	1RZ4_A
	Mus musculus	AAH91749
eIF3j	*Homo sapiens*	O75822
eIF4B	*Homo sapiens*	P23588
eIF4H	*Homo sapiens*	NP_071496

7- Abstract

Protein synthesis or translation is a basic and essential process for the survival of all living beings. One of its key points is its initiation stage which is regulated by the action of at least twelve protein factors called eIFs (eukaryotic Initiation Factor), summing about 30 polypeptides in mammals. The trypanosomatids, pathogenic protozoa of medical interest, display unique cellular characteristics such as its regulation of gene expression which occurs mainly at the post-transcriptional level. In this context the protein synthesis is a potential target for regulatory mechanisms, however little is known about this process in trypanosomatids. In previous studies, the eIF4F complex was investigated in these parasites and multiples homologues for each one of its three subunits were observed. In this work, bioinformatic tools were used to identify and characterize proteins homologous to the others eIFs in *Leishmania major*, *Trypanosoma brucei* and *T. cruzi*. Homologues to the factors eIF1, eIF1A, eIF5, eIF5A, eIF5B, eIF6 and the seven subunities of the eIF3 complex (b, c, d, e, f, i, k) have been identified. In contrast to what was observed for the subunits of eIF4F, and with the exception of the eIF3b subunit (with two identified homologues), only one homologue was identified for each factor studied. The analysis of the amino acid sequences showed a variation in the degree of conservation of these homologues when compared to that of other eukariotes (from 22% identity for eIF3k up to 58% for eIF6). Also, in some cases it was possible to map mutations unique to the trypanosomatids. In another approach, 3D models have been generated for several of the eIF4F subunit homologues previously identified. Those models have contributed to the characterization of possible eIF4F functions. Overall, the results obtained indicate that the initiation of protein synthesis is conserved between trypanosomatids and other eukariotes. However, significant differences seem to occur and deserve to be studied in these parasites.

Key-words: eIF; Translation Initiation; *Leishmania major*; *Trypanosoma cruzi*; *T. brucei*

8.1.1. Apêndice 1
INSTRUÇÕES PARA AUTORES

Revista
GENETICS AND MOLECULAR BIOLOGY

ISSN 1415-4757
Ribeirão Preto, Brasil

Genetics and Molecular Biology - NOTICE TO CONTRIBUTORS

Scope and policy

Genetics and Molecular Biology (formerly named Revista Brasileira de Genética/Brazilian Journal of Genetics - ISSN 0100-8455) is published quarterly by the Sociedade Brasileira de Genética (Brazilian Society of Genetics).

The Journal considers contributions that present the results of original research in genetics, evolution and related scientific disciplines.

Although **Genetics and Molecular Biology** is an official publication of the Brazilian Society of Genetics, contributors are not required to be members of the Society.

It is a fundamental condition that submitted manuscripts have not been and will not be published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries.

The use of registered names and trademarks does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Submission of papers

1. Manuscripts should be submitted to:

Fábio de Melo Sene, Editor-in-Chief
Genetics and Molecular Biology
Rua Capitão Adelmio Norberto da Silva, 736
14025-670 Ribeirão Preto, SP - Brasil

2. A submission package sent to the Editorial Office must contain:

a. A cover letter signed by all authors stating that they have approved the submission of the manuscript and that the findings have not been published or are not under consideration for publication elsewhere;
b. Three copies of the manuscript and figures.
c. Two copies of any unpublished or in-press companion articles referred to in the submission.
d. A copy of the text, tables and figures on a disk. Be sure that the disk is adequately protected; if a disk arrives damaged, a new disk will be requested, causing delays in publication. Formats for text are Word or RTF, in Windows platform. Images in TIF or JPG formats should be sent in separate files (For Figures, see detailed instructions in 3.1.g). Disk must be labeled with the first author's last name, platform and software. (See detailed instructions below). Failure to adhere to these guidelines can delay the handling of your contribution, and manuscripts may be returned before being reviewed.

3. Categories of Contribution:

3.1.Research Articles

Manuscripts must be written in English in double-spaced, 12-point type throughout, including the References Cited section, appendices, tables and legends; printed on one side only of A4 paper with 2.5 cm margins; marked with consecutive page numbers, beginning with the cover page.

The following elements must start on a new page and be ordered as they are listed below:

a) The title page must contain: a concise and informative title; the authors' names (first name at full length); the authors' institutional affiliation, including department, institution, and city, state or province, and country; different affiliations indicated with superscript numbers; a short running title of about 35 characters, including spaces; up to five key words; the corresponding author's name, postal address, phone and fax numbers and email address. The corresponding author is the person responsible for checking the page proofs, and arranging for the payment of color illustrations and author alterations charges.

b) The Abstract must be a single paragraph that does not exceed 200 words and summarizes the main results and conclusions of the study. It should not contain references.

c) The text: must be as succinct as possible. Text citations: articles should be referred to by authors' surnames and date of publication; citations with two authors must include both names; in citations with three or more authors, name the first author and use "et al". Only articles that are published or in press should be cited. In the case of personal communications or unpublished results, all contributors must be listed by initials and last name ("et al" should not be used). Numbers: In the text, numbers nine or less must be written out except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numerals for numbers larger than nine. Avoid starting a sentence with a number. Binomial Names: Latin names of genera, species and intraspecific taxa in the text must be printed in italics; names of orders and families should be in the Title.

The text includes the following elements:

Introduction – Description of the background that led to the study.

Material (or Subjects) and Methods – Details relevant to the conduct of the study. Statistical methods should be explained at the end of this section.

Results – Undue repetition in text and tables should be avoided. Comment on significance of results is appropriate but broader discussion should be part of the Discussion section.

Discussion – The findings of the study should be placed in context of relevant published data. Ideas presented in other publications should not be discussed solely to make an exhaustive presentation.

Some manuscripts may require different formats appropriate to their content.

d) The Acknowledgments must be a single paragraph that immediately follows the discussion and includes references to grant support.

e) The References Section: citations must be ordered alphabetically by the first author; only articles that are published or in press should be included; personal communications must be cited within the text; journal titles must be abbreviated according to Medline (http://www.ncbi.nlm.nih.gov/entrez/jrbrowser.cgi).

Genetics and Molecular Biology

Sample journal article citation:

Breuer ME and Pavan C (1955) Behaviour of polytene chromosomes of Rhynchosciara angelae at different stages of larval development. Chromosoma 7:371-386.

Bertollo LAC, Takahashi CS and Moreira-Filho O (1978) Cytotaxonomic consideration on Hoplias lacerdae (Pisces, Erythrinidae). Rev Bras Genet 1:103-120.

Sample book citation

Salzano FM and Freire-Maia N (1967) Populações Brasileiras. Companhia Editora Nacional and EDUSP, São Paulo, 178 pp.

Dobzhansky T (1951) Genetics and Origin of Species. 3rd edition. Columbia University Press, New York, 364 pp.

Sample chapter-in-book citation:

Carvalho A, Monaco LC and Krug CA (1966) Melhoramento genético das plantas e sua repercussão econômica. In: Pavan C and da Cunha AB (eds) Elementos de Genética. 2nd ed. EDUSP and Companhia Editora Nacional, São Paulo, pp 587-653.

Sample abstracts in meeting citation:

Basile R (1973) Cromossomos Politênicos em células nutritivas de ovócitos de ovário atrofiado de Rhyncosciara. Ciênc e Cult 25 (suppl): 248. XXV Reunião Anual da SBPC, Rio de Janeiro, Brazil.

Sample Thesis/Dissertation citation:

Frota-Pessoa O (1953) Revision of the Tripunctata group of Drosophila with description of fifteen new species. PhD Thesis, Universidade do Brasil, Rio de Janeiro.

f) Tables each table must start on a new page. A concise title should be provided above the table. Tables must be numbered consecutively in Arabic numerals. Each column must have a title in the box head. Footnotes, typed directly below the table, should be indicated in lowercase superscript numbers.

g) Figures must be numbered consecutively in Arabic numerals. Legends should be typed on a separate sheet. Three sets of illustrations of the highest quality must be provided, one original and two copies in glossy paper. If you have created figures electronically, submit them also as hard copies. Scanned figures should not be submitted. Images should be in TIF or JPG format and provided in separate files. Figures in Word format cannot be published. Journal quality reproduction will require grayscale and color at resolution yielding 300 ppi. Authors should submit bitmapped line art at resolution yielding 600–1200 ppi. These resolutions refer to the out put size of the file; if it is anticipated that images will be enlarged or reduced, the resolutions should be adjusted accordingly. Identify each illustration by affixing on the back a label containing: the number of the figure, the name of the first author, and an arrow indicating top of illustration. Illustrations supplied on disks must follow instructions in item 2 (Submission package). Color illustration can be accepted, but authors are asked to defray the cost. For costs of color figures, check with the Editorial Office.

h) Nomenclature: current standard international nomenclature should be adhered to.

i) Sequences may appear in text or in figure. DNA must be sequenced on both strands. DNA, RNA , or protein sequences equal to or greater than 50 units must be entered into appropriate data bank and the accession number must be provided before publication of the article. Long sequences requiring more than two pages to reproduce will not be published unless the Editorial decision is that the publication is necessary. Complete mtDNA sequence will not be published.

j) Data access: reference should be made to availability of detailed data and materials used for reported studies.

k) Ethical issues: Reports of experiments on live vertebrates must include a brief statement that the work was approved by the institutional review board. For experiments involving human subjects, authors must also include a statement that informed consent was obtained from all subjects. If photos or any other identifiable data are included, a copy of the signed consent must accompany the manuscript.

3.2 Short Communications present brief observations that do not warrant full-length articles. They should not be considered preliminary communications. Their format is that of full-length article. The text must be kept to a minimum.

3.3 Letters to the Editor relate or respond to recent published items in the journal. Discussions of political, social and ethical issues of interest to geneticists are also welcome in this form.

3.4 Review Articles are welcome.

3.5 Book Reviews: publishers are invited to submit books on Genetics, Evolution and related disciplines, for review in the journal.

3.6 History, Story and Memories: accounts on historical aspects of Genetics relating to Brazil.

4. Proofs: Page proofs will be sent to the corresponding author. Changes made to page proofs, apart from printer's errors, will be charged to the authors. Notes added in proof require Editorial approval.

8.1.2. Apêndice 2
INSTRUÇÕES PARA AUTORES

Revista
EUKARYOTIC CELL

ISSN 1535-9778
Washington, Estados Unidos da América

EUKARYOTIC CELL, Jan. 2006, p. 1–17 Vol. 5, No. 1
1535-9778/06/$08.00+0 doi:10.1128/EC.5.1.1–17.2006

EUKARYOTIC CELL

2006 INSTRUCTIONS TO AUTHORS*

SCOPE

Eukaryotic Cell (EC) publishes reports of basic research on eukaryotic microorganisms such as yeasts, fungi, algae, protozoa, and social amoebae. Topics include but are not limited to basic biology; molecular and cellular biology; mechanisms, and control, of developmental pathways; structure and form inherent in basic biological processes; cellular architecture; metabolic physiology; comparative genomics, biochemistry, and evolution; ecology; and population dynamics.

The journal will consider manuscripts reporting results from the use of genome-, transcriptome-, or proteome-wide screening approaches when the experiments address a specific question or working hypothesis and when the results are used to illuminate mechanisms of gene regulation or interactions of signal transduction pathways via additional experiments. Studies that only catalog differences and similarities between genotypes or responses to stimuli are not likely to be reviewed favorably.

In addition, EC will consider manuscripts dealing with the viruses of these organisms and their organelles and with interactions with other living systems, where the focus is clearly on the eukaryotic cell.

Questions about these guidelines may be directed to the editor in chief.

ASM publishes a number of different journals covering various aspects of microbiology. Each journal has a prescribed scope that must be considered in determining the most appropriate journal for each manuscript. If transfer to another ASM journal is recommended by an editor, the corresponding author will be contacted.

Note that a manuscript rejected by one ASM journal on scientific grounds or on the basis of its general suitability for publication is considered rejected by all other ASM journals.

EDITORIAL POLICY

Use of Microbiological Information

The Council Policy Committee (CPC) of the American Society for Microbiology affirms the long-standing position of the Society that microbiologists will work for the proper and beneficent application of science and will call to the attention of the public or the appropriate authorities misuses of microbiology or of information derived from microbiology. ASM members are obligated to discourage any use of microbiology contrary to the welfare of humankind, including the use of microbes as biological weapons. Bioterrorism violates the fundamental principles expressed in the Code of Ethics of the Society and is abhorrent to ASM and its members.

ASM recognizes that there are valid concerns regarding the publication of information in scientific journals that could be put to inappropriate use as described in the CPC resolution mentioned above. Members of the ASM Publications Board will evaluate the rare manuscript that might raise such issues during the review process. However, as indicated elsewhere in these Instructions, research articles must contain sufficient detail, and material/information must be made available, to permit the work to be repeated by others. Supply of materials should be in accordance with laws and regulations governing the shipment, transfer, possession, and use of biological materials and must be for legitimate, bona fide research needs. Links to, and information regarding, these laws and regulations can be found at http://www.asm.org/Policy/index.asp.

General Requirements

Manuscripts submitted to the journal must represent reports of original research, and the original data must be available for review by the editor if necessary.

All authors of a manuscript must have agreed to its submission and are responsible for its content (initial submission and any subsequent versions), including appropriate citations and acknowledgments, and must also have agreed that the corresponding author has the authority to act on their behalf in all matters pertaining to publication of the manuscript. The corresponding author is responsible for obtaining such agreements and for informing the coauthors of the manuscript's status throughout the submission, review, and publication process. For Authors' Corrections and Retractions, signed letters of agreement from all of the authors must be submitted (see p. 10–11).

By submission of a manuscript to the journal, **the authors guarantee that they have the authority to publish the work and that the manuscript, or one with substantially the same content, was not published previously, is not being considered or published elsewhere, and was not rejected on scientific grounds by another ASM journal.**

It is expected that the authors will provide written assurance that permission to cite unpublished data or personal communications has been granted.

By publishing in the journal, the authors agree that any DNAs, viruses, microbial strains, mutant animal strains, cell lines, antibodies, and similar materials newly described in the article are available from a national collection or will be made available in a timely fashion, at reasonable cost, and in limited quantities to members of the scientific community for noncommercial

*Shading indicates material that has been added or significantly updated.

purposes. The authors guarantee that they have the authority to comply with this policy either directly or by means of material transfer agreements through the owner.

Similarly, the authors agree to make available computer programs, originating in the authors' laboratory, that are the only means of confirming the conclusions reported in the article but that are not available commercially. The program(s) and suitable documentation regarding its (their) use may be provided by any of the following means: (i) as a program transmitted via the Internet, (ii) as an Internet server-based tool, or (iii) as a compiled or assembled form on a suitable medium (e.g., magnetic or optical). It is expected that the material will be provided in a timely fashion and at reasonable cost to members of the scientific community for noncommercial purposes. The authors guarantee that they have the authority to comply with this policy either directly or by means of material transfer agreements through the owner.

Primary Publication

A scientific paper *or its substance* published in a serial, periodical, book, conference report, symposium proceeding, or technical bulletin, posted on a nonpersonal website, or made available through any other retrievable source, including CD-ROM and other electronic forms, is unacceptable for submission to an ASM journal on grounds of prior publication.

Posting of a method/protocol on a nonpersonal website should not interfere with the author's ability to have a manuscript utilizing that technique considered for publication in an ASM journal; however, ultimately, it is an editorial decision whether the method constitutes the substance of a paper.

Posting of a limited amount of original data on a personal/university/company website or websites of small collaborative groups working on a problem does not preclude subsequent submission to, and publication by, an ASM journal. The posted data, however, may not constitute the substance of the submission. Specific questions about this policy may be referred to the Publications Board chairman on a case-by-case basis. Posting of theses and dissertations on a personal/university-hosted website does not preclude subsequent submission to, and publication by, an ASM journal.

Posting of unpublished sequence data on the Internet is usually not considered prior publication; however, the address (URL) of the source of the sequence should be included in the text.

Preliminary disclosures of research findings webcast as meeting presentations or published in abstract form as adjuncts to a meeting, e.g., part of a program, are not considered prior publication.

It is incumbent upon the author to acknowledge any prior publication, including his own articles, of the data contained in a manuscript submitted to an ASM journal. A copy of the relevant work should be submitted with the paper as supporting material.

Ultimately, it is an editorial decision whether the material constitutes the substance of a paper.

Permissions

The corresponding author is responsible for obtaining permission from both the original author and the original publisher (i.e., the copyright owner) to reproduce or modify figures and tables and to reproduce text (in whole or in part) from previous publications.

The original **signed** permission(s) must be submitted directly to the editor, outside the Rapid Review system, no later than the modification stage and should be identified as to the relevant item in the ASM manuscript (e.g., "permissions for Fig. 1 in EC00123-06"). In addition, a statement indicating that the material is being reprinted with permission must be included in the relevant figure legend or table footnote of the manuscript. Reprinted text must be enclosed in quotation marks, and the permission statement must be included as running text or indicated parenthetically.

For supplemental material intended for posting by ASM (see p. 4), if the authors of the EC manuscript are not also the owner of the supplemental material, the corresponding author **must** send to ASM **signed** permission from the copyright owner that allows posting of the material, as a supplement to the article, by ASM. The corresponding author is also responsible for incorporating in the supplemental material any copyright notices required by the owner.

Authorship

An author is one who made a substantial contribution to the overall design and execution of the experiments; therefore, **ASM considers all authors responsible for the entire paper**. Individuals who provided assistance, e.g., supplied strains or reagents or critiqued the paper, need not be listed as authors but may be recognized in the Acknowledgments section.

A study group, surveillance team, working group, consortium, or the like (e.g., the Active Bacterial Core Surveillance Team) may be listed as a coauthor in the byline if its contributing members satisfy the requirements for authorship and accountability as described in these Instructions. The names (and institutional affiliations if desired) of the contributing members only may be given in a footnote keyed to the study group name in the byline or as a separate paragraph in the Acknowledgments section.

If the contributing members of the group associated with the work do not fulfill the criteria of substantial contribution to and responsibility for the paper, the group may not be listed in the author byline. Instead, it and the names of its contributing members may be listed in the Acknowledgments section.

All authors must agree to the order in which their names are listed in the byline. Statements regarding equal contributions by two or more authors (e.g., X.J. and Y.S. contributed equally to . . .) are permitted as footnotes to bylines. Other statements of attribution may

be included in the Acknowledgments section.

A change in authorship (order of listing or addition or deletion of a name) **after submission of the manuscript will be implemented only after receipt of signed statements of agreement from all parties involved.** Disputes about authorship may delay review and/or publication of the manuscript.

Conflict of Interest

All authors are expected to disclose, in the manuscript submittal letter, any commercial affiliations as well as consultancies, stock or equity interests, and patent-licensing arrangements that could be considered to pose a conflict of interest regarding the submitted article. (Inclusion of a company name in the author address lines of the manuscript does not constitute disclosure.) Details of the disclosure to the editor will remain confidential. However, it is the responsibility of authors to provide, in the Acknowledgments section, a *general* statement disclosing financial or other relationships that are relevant to the study. Examples of potentially conflicting interests that should be disclosed include relationships that might detract from an author's objectivity in presentation of study results, and interests whose value would be enhanced by the results presented. All funding sources for the project, institutional and corporate, should be credited in the Acknowledgments section, as described below. In addition, if a manuscript concerns a commercial product, the manufacturer's name must be indicated in the Materials and Methods section or elsewhere in the text, as appropriate, in an obvious manner.

Copyright

To maintain and protect the Society's ownership and rights and to continue to afford scientists the opportunity to publish in high-quality journals, ASM requires the corresponding author to sign a copyright transfer agreement on behalf of all the authors. This agreement is sent to the corresponding author when the manuscript is accepted and scheduled for publication. Unless this agreement is executed (*without changes and/or addenda*), ASM will not publish the manuscript.

In the copyright transfer agreement signed by an author, ASM grants to that author (and coauthors) the right to republish *discrete portions* of his (their) article in any other publication (print, CD-ROM, and other electronic forms) of which he is (they are) the author(s) or editor(s), *on the condition that appropriate credit is given to the original ASM publication*. This republication right also extends to posting on a host computer to which there is access via the Internet. Except as indicated below, significant portions of the article may *not* be reprinted/posted without ASM's prior written permission, however, as this would constitute duplicate publication.

Authors may post their own published articles on their *personal or university-hosted* (but not corporate, government, or similar) websites without ASM's prior written permission provided that appropriate credit is given (i.e., either the copyright lines shown on the top of the first page of the PDF version or "Copyright © American Society for Microbiology, [insert journal name, volume number, page numbers, and year]" for the HTML version).

The copyright transfer agreement asks that authors who were U.S. Government employees and who wrote the article as part of their employment duties be identified. This is because works authored solely by such U.S. Government employees are not subject to copyright protection, so there is no copyright to be transferred. The other provisions of the copyright transfer agreement, such as author representations of originality and authority to enter into the agreement, apply to U.S. Government employee-authors as well as to other authors.

ASM also requires that copyright transfer agreements be signed for cover artwork/photographs.

Copyright for supplemental material (see p. 5) remains with the author, but a license permitting the posting by ASM will be sent, along with the article copyright transfer agreement, to the corresponding author for signing at the acceptance stage. (If the author of the article is not also the copyright owner of the supplemental material, the corresponding author **must** send to ASM **signed** permission from the owner that allows posting of the material, as a supplement to the article, by ASM. The corresponding author is also responsible for incorporating into the supplemental material any copyright notices required by the owner.)

Funding Agency Repositories

The National Institutes of Health (NIH) requests that its grantee and intramural authors provide copies of their accepted manuscripts to PubMed Central (PMC) for posting in the PMC Public Access Repository. ASM allows such EC authors to do so. ASM also allows EC authors whose work was supported by similar funding agencies that have public access requirements like those of NIH (e.g., the Wellcome Trust) to post their **accepted manuscripts** in publicly accessible electronic repositories maintained by those funding agencies. If a funding agency does not itself maintain such a site, then ASM allows the author to fulfill that requirement by depositing the manuscript (**not** the typeset article) in an appropriate institutional or subject-based open repository established by a government or noncommercial entity.

Since ASM makes the final, typeset articles from its primary-research journals available free of charge on the ASM Journals and PMC websites 6 months after final publication, ASM recommends that when submitting the accepted manuscript to PMC or similar public access site, the author specify that the **posting release date for the manuscript be no earlier than 6 months after publication of the typeset article by ASM.**

Use of Human Subjects or Animals in Research

The use of human subjects or other animals for research purposes is regulated by the federal government and individual institutions. Manuscripts containing in-

formation related to human or animal use should clearly state that the research has complied with all relevant federal guidelines and institutional policies. Copies of these guidelines and policy statements must be available for review by the editor if necessary.

Patient Identification

When isolates are derived from patients in clinical studies, do not identify them by using the patients' initials, even as part of a strain designation. Change the initials to numerals or use randomly chosen letters. Do not give hospital unit numbers; if a designation is needed, use only the last two digits of the unit. (Note: Established designations of some viruses and cell lines, although they consist of initials, are acceptable [e.g., JC virus, BK virus, and HeLa cells].)

Nucleotide and Amino Acid Sequences

It is expected that newly determined nucleotide and/or amino acid sequence data will be deposited and GenBank/EMBL/DDBJ accession numbers will be included in the manuscript no later than the modification stage of the review process. It is also expected that the sequence data will be released to the public no later than the publication date of the article. The accession numbers should be included in a separate paragraph at the end of the Materials and Methods section for full-length papers or at the end of the text for Notes. If conclusions in a manuscript are based on the analysis of sequences and a GenBank/EMBL/DDBJ accession number is not provided at the time of the review, authors should provide the sequence data as supplemental material.

It is expected that, when previously published sequence accession numbers are cited in a manuscript, the original citations (e.g., journal articles) will be included in the References section when possible and reasonable.

Authors are also expected to do elementary searches and comparisons of nucleotide and amino acid sequences against the sequences in standard databases (e.g., GenBank) immediately before manuscripts are submitted and again at the proof stage.

Analyses should specify the database, and the date of each analysis should be indicated in the format MM/YY. If relevant, the version of the software used should be specified.

See p. 13 for nucleic acid sequence formatting instructions.

The URLs of the databases mentioned above are as follows: DNA Data Bank of Japan (DDBJ), http://www.ddbj.nig.ac.jp; EMBL Nucleotide Sequence Submission (EMBL), http://www.ebi.ac.uk; and GenBank, National Center for Biotechnology Information (GenBank), http://www.ncbi.nlm.nih.gov.

Structural Determinations

It is expected that coordinates for new structures of macromolecules will be deposited in the Protein Data Bank and that assigned identification codes will be included in the manuscript no later than the modification stage of the review process. It is also expected that the coordinates will be released to the public no later than the publication date of the article. Authors are encouraged to send coordinates with their original submission, however, so that reviewers can examine them along with the manuscript. The accession number(s) should be listed in a separate paragraph at the end of the Materials and Methods section for full-length papers or at the end of the text for Notes.

The URLs for coordinate deposition are http://pdb.rutgers.edu and http://pdbdep.protein.osaka-u.ac.jp.

Microarray Data

It is expected that the entire set of supporting microarray data will be deposited in the appropriate public database (e.g., GEO, ArrayExpress, or CIBEX) and that the assigned accession number(s) will be included in the manuscript no later than the modification stage of the review process. It is also expected that the data will be released to the public no later than 6 months after publication of the typeset article. Authors are encouraged to send the relevant data with their original submission, however, so that reviewers can examine them along with the manuscript. The accession number(s) should be listed in a separate paragraph at the end of the Materials and Methods section for full-length papers or at the end of the text for Notes.

The URLs of the databases mentioned above are as follows: Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo; ArrayExpress, http://www.ebi.ac.uk/arrayexpress; and Center for Information Biology Gene Expression Database (CIBEX), http://cibex.nig.ac.jp.

Culture Deposition

EC expects authors to deposit important strains in publicly accessible culture collections and to refer to the collections and strain numbers in the text. Since the authenticity of subcultures of culture collection specimens that are distributed by individuals cannot be ensured, authors should indicate laboratory strain designations and **donor sources** as well as original culture collection identification numbers.

Supplemental Material

Supplemental material intended for posting by ASM may not include additional figures or tables that simply support the authors' conclusions. It must be restricted to large or complex data sets or results that cannot be readily displayed in printed form because of space or technical limitations. Such material may include data from microarray, structural, biochemical, or video imaging analyses. In such cases, the manuscript submitted for review should include a distillation of the results so that the principal conclusions are fully supported without

referral to the supplemental material.

Supplemental material intended for posting by ASM must be uploaded in Rapid Review and will be reviewed along with the manuscript. The decision to publish (i.e., post online only) the material with the article if it is accepted will be made by the editor and conveyed to the corresponding author in the acceptance e-mail. Note, therefore, it is possible that an article will be accepted but that the supplemental material will not be.

If the software required for users to view/use the supplemental material is not embedded in the file, you are urged to use shareware or generally available/easily accessible programs.

Unlike the article, supplemental material will **not** be edited by the ASM Journals staff and proofs will not be made available.

Supplemental material will always remain associated with its article and is **not** subject to any modifications after publication.

Material that has been published previously (print or online) is not acceptable for posting as supplemental data. Instead, the appropriate reference(s) to the original publication should be made in the article text.

Copyright for the supplemental material remains with the author, but a license permitting the posting by ASM will be sent, along with the article copyright transfer agreement, to the corresponding author for signing. If you are not the copyright owner, you **must** provide to ASM **signed** permission from the owner that allows posting of the material, as a supplement to your article, by ASM. You are responsible for including in the supplemental material any copyright notices required by the owner.

A one-time charge (amount not yet determined) may be levied for posting of supplemental material. When instituted, the charge will be indicated in the ASM acceptance letter.

Compliance

Failure to comply with the policies described in these Instructions may result in a letter of reprimand, a suspension of publishing privileges in ASM journals, and/or notification of the authors' institutions.

Warranties and Exclusions

Articles published in this journal represent the opinions of the authors and do not necessarily represent the opinions of ASM. ASM does not warrant the fitness or suitability, for any purpose, of any methodology, kit, product, or device described or identified in an article. The use of trade names is for identification purposes only and does not constitute endorsement by ASM.

Page Charges

Authors whose research was supported by grants, special funds (including departmental and institutional), or contracts (including governmental) or whose research was done as part of their official duties (government, corporate, etc.) are required to pay page charges.

For a **corresponding author who is an ASM member,** page charges are currently $65 per page for the first six pages and $200 per page for each page in excess of six (subject to change without notice). To obtain the member rate, the *corresponding author* must be an ASM For a **nonmember corresponding author,** page charges are currently $75 per page for the first six pages and $250 for each page in excess of six. A corresponding author who is not an ASM member may join ASM to obtain the member rate.

If the research was not supported by any of the means described above, a request to waive the charges may be mailed (Journals Department, ASM, 1752 N St., N.W., Washington, DC 20036-2904, USA) or faxed (202-942-9355) to the Journals Department. This request must indicate how the work was supported and should be accompanied by copies of the title page and Acknowledgments section.

Minireviews are not subject to page charges.member.

Editorial Style

The editorial style of ASM journals conforms to the *ASM Style Manual for Journals* (American Society for Microbiology, 2006, in-house document) and *How To Write and Publish a Scientific Paper*, 5th ed. (Oryx Press, 1998), as interpreted and modified by the editors and the ASM Journals Department.

The editors and the Journals Department reserve the privilege of editing manuscripts to conform with the stylistic conventions set forth in the aforesaid publications and in these Instructions.

Review Process

All manuscripts are considered to be confidential and are reviewed by the editors, members of the editorial board, or qualified ad hoc reviewers. To expedite the review process, authors should recommend at least two reviewers who are not members of their institution(s) and have never been associated with them or their laboratory(ies); please provide their contact information where indicated on the submission form.

Copies of in-press and submitted manuscripts that are important for judgment of the present manuscript should be included as supporting material to facilitate the review.

When a manuscript is submitted to the journal, it is given a number (e.g., EC00047-06 version 1) and assigned to one of the editors. **(Always refer to this number in communications with the editor and the Journals Department.)** *It is the responsibility of the corresponding author to inform the coauthors of the manuscript's status throughout the submission, review, and publication processes*. The reviewers operate under strict guidelines set forth in "Guidelines for Reviewers" (http://www.journals.asm.org/misc/reviewguide.shtml) and are expected to complete their reviews expeditiously.

The corresponding author is notified, generally within 4 to 6 weeks after submission, of the editor's decision to

accept, reject, or require modification. When modification is requested, the corresponding author must either submit the modified version within 2 months or withdraw the manuscript. A point-by-point response to the reviews must be provided in the Rebuttal section of the Rapid Review submission form for the revised manuscript, and a compare copy of the manuscript (without figures) should be included as supporting material if the editor requested one.

Manuscripts that have been rejected, or withdrawn after being returned for modification, may be resubmitted if the major criticisms have been addressed. **(Note: A manuscript rejected by one ASM journal on scientific grounds or on the basis of its general suitability for publication is considered rejected by all other ASM journals.)** The cover letter must state that the manuscript is a resubmission, and the *former manuscript number should be provided* in the appropriate field on the submission form. A **point-by-point response** to the reviews and a compare copy of the revised manuscript showing the changes **must** be included as supporting material (the Rebuttal section appears in the submission form only if the manuscript is a modification). Resubmitted manuscripts are normally handled by the original editor.

Rejected manuscripts may be resubmitted only once unless permission has been obtained from the original editor or from the editor in chief.

Notification of Acceptance

When an editor has decided that a manuscript is acceptable for publication on the basis of scientific merit, the author and the Journals Department are notified. The text files undergo an automated preediting, cleanup, and tagging process specific to the particular article type, and the illustrations are examined. If all files have been prepared according to the criteria set forth in these Instructions and those in Rapid Review, the acceptance procedure will be completed successfully. If there are problems that would cause extensive corrections to be made at the copyediting stage or if the files are not acceptable for production, ASM Journals staff will contact the corresponding author.

Once all the material intended for publication has been determined to be adequate, the manuscript is scheduled for the next available issue and an acceptance letter indicating the month of publication, approximate page proof dates, and table of contents section is mailed to the corresponding author; a copyright transfer agreement is also included, as is a license to permit posting of supplemental material (if applicable). The editorial staff of the ASM Journals Department completes the editing of the manuscript to bring it into conformity with prescribed standards.

Page Proofs

Page proofs, together with a query sheet and instructions for handling proofs, will be made available to the corresponding author electronically via a PDF file that can be accessed through a unique password. Since corresponding authors will be notified of the availability of their PDF proofs, instructed how to access information about page charges, reprints, and color figure charges (if applicable), and assigned their unique password via e-mail, an e-mail address **must** be supplied in the correspondent footnote. Failure to do so may result in a delay in publication. **The PDF page proofs must be printed out, and corrections must be written on the hard copy.** Queries must be answered on the query page or on a separate sheet of paper, and any changes related to the queries must be indicated on the proofs. Note that the copy editor does not query at every instance where a change has been made. Queries are written only to request necessary information or clarification of an unclear passage or to draw attention to edits that may have altered the sense. It is the author's responsibility to read the entire text, tables, and figure legends, not just items queried. As soon as the page proofs are corrected and signed by the person who proofread them (within 48 h), **they should be mailed or sent by a courier service** such as FedEx, **not** faxed or sent as an e-mail attachment, to the ASM Journals Department.

The proof stage is not the time to make extensive corrections, additions, or deletions. Important new information that has become available between acceptance of the manuscript and receipt of the proofs may be inserted as an addendum in proof with the permission of the editor. If references to unpublished data or personal communications are added, it is expected that written assurance granting permission for the citation will be included. Limit changes to correction of spelling errors, incorrect data, and grammatical errors and updated information for references to articles that have been submitted or are in press. If URLs have been provided in the article, recheck the sites to ensure that the addresses are still accurate and the material that you expect the reader to find is indeed there.

Questions about *late proofs and problems in the proofs* should be directed to the ASM Journals Department (telephone, 202-942-9231). Questions about *accessing or viewing your PDF proofs* should be directed to Katie Gay of Cadmus Professional Communications at 804-261-3155 or gayk@cadmus.com.

Reprints

Reprints (in multiples of 100) may be purchased by all coauthors. In the proof notification e-mail, the corresponding author will be instructed how to access information about reprints.

The corresponding authors of Minireviews and Guest Commentaries will receive 100 free reprints of their contribution; additional reprints (in multiples of 100) may be purchased if desired. As for regular articles, the corresponding author will be instructed, in the proof notification e-mail, how to access information about reprints.

PDF Files

A corresponding author who has included an e-mail address in his "corresponding author" footnote will have limited access (10 downloads, total) to the PDF file of his published article. An e-mail alert will automatically be sent to him on the day the issue is posted. It will provide a URL, which will be required to obtain access, and instructions. An article may be viewed, printed, or stored, provided that it is for the author's own use.

Should coauthors or colleagues be interested in viewing the paper for their own use, the corresponding author may provide them with the URL; a copy of the article may not be forwarded electronically. However, they must be made aware of the terms and conditions of the ASM copyright. (For details, go to http://www.journals.asm.org/misc/terms.shtml.) Note that each such download will count toward the corresponding author's total of 10. After 10 downloads, access will be denied and can be obtained only through a subscription to the journal (either individual or institutional) or after the standard access control has been lifted (i.e., 6 months after publication).

HOW TO SUBMIT MANUSCRIPTS

All submissions to EC must be made electronically via the Rapid Review online submission and peer review system at the following URL: www.rapidreview.com/ASM2/author.html. (E-mailed submissions will not be accepted.) First-time users must create an Author account, which may be used for submitting to all ASM journals. Instructions for creating an Author account are available at the above URL under the Create Account button. The person from whose account a submission is made will be recorded as the corresponding author, and any correspondence will be sent to the e-mail address in that account. Step-by-step instructions for submitting a manuscript via Rapid Review are available from the account holder's My Manuscripts page. Information on file types acceptable for electronic submission can be found under the More About File Formats button.

ORGANIZATION AND FORMAT

On receipt at ASM, an accepted manuscript undergoes an automated preediting, cleanup, and tagging process specific to the particular article type. To optimize this process, manuscripts must be supplied in the correct format and with the appropriate sections and headings.

Type every portion of the manuscript double spaced (a minimum of 6 mm between lines), including figure legends, table footnotes, and References, and number all pages in sequence, including the abstract, figure legends, and tables. Place the last two items after the References section. **Manuscript pages should have line numbers; manuscripts without line numbers may be editorially rejected by the editor, with a suggestion of resubmission after line numbers are added.** The font size should be no smaller than 12 points. It is recommended that the following sets of characters be easily distinguishable in the manuscript: the numeral zero (0) and the letter "oh" (O); the numeral one (1), the letter "el" (l), and the letter "eye" (I); and a multiplication sign (×) and the letter "ex" (x). Do not create symbols as graphics or use special fonts that are external to your word processing program; use the "insert symbol" function. Set the page size to 8½ by 11 inches (ca. 21.6 by 28 cm). Italicize or underline any words that should appear in italics, and indicate paragraph lead-ins in bold type.

Authors who are unsure of proper English usage should have their manuscripts checked by someone proficient in the English language.

Manuscripts may be editorially rejected, without review, on the basis of poor English or lack of conformity to the standards set forth in these Instructions.

Full-Length Papers

Full-length papers should include the elements described in this section.

Title, running title, and byline. Each manuscript should present the results of an independent, cohesive study; thus, numbered series titles are not allowed. Avoid the main title/subtitle arrangement, complete sentences, and unnecessary articles. On the title page, include the title, running title (not to exceed 54 characters and spaces), name of each author, address(es) of the institution(s) at which the work was performed, each author's affiliation, and a footnote indicating the present address of any author no longer at the institution where the work was performed. Place an asterisk after the name of the author to whom inquiries regarding the paper should be directed (see "Correspondent footnote" below).

Study group in byline. A study group, surveillance team, working group, consortium, or the like (e.g., the Active Bacterial Core Surveillance Team) may be listed as a coauthor in the byline if its contributing members satisfy the requirements for authorship and accountability as described in these Instructions. The names (and institutional affiliations if desired) of the contributing members only may be given in a footnote keyed to the study group name in the byline or a separate paragraph in Acknowledgments.

If the contributing members of the group associated with the work do not fulfill the criteria of substantial contribution to and responsibility for the paper, the group may not be listed in the author byline. Instead, it and the names of its contributing members may be listed in the Acknowledgments section.

Correspondent footnote. The complete mailing address, a single telephone number, a single fax number, and a single e-mail address for the corresponding author should be included on the title page of the manuscript. This information will be published in the article as a footnote to facilitate communication, and the e-mail address will be used to notify the corresponding author of availability of proofs and, later, of the PDF file of the published article.

Abstract. Limit the abstract to **250 words or fewer** and concisely summarize the basic content of the paper without presenting extensive experimental details. Avoid abbreviations and references, and do not include diagrams. When it is essential to include a reference, use the same format as shown for the References section but omit the article title. Because the abstract will be published separately by abstracting services, it must be complete and understandable without reference to the text.

Introduction. The introduction should supply sufficient background information to allow the reader to understand and evaluate the results of the present study without referring to previous publications on the topic. The introduction should also provide the hypothesis that was addressed and the rationale for the present study. Use only those references required to provide the most salient background rather than an exhaustive review of the topic.

Materials and Methods. The Materials and Methods section should include sufficient technical information to allow the experiments to be repeated. When centrifugation conditions are critical, give enough information to enable another investigator to repeat the procedure: make of centrifuge, model of rotor, temperature, time at maximum speed, and centrifugal force ($\times g$ rather than revolutions per minute). For commonly used materials and methods (e.g., media and protein concentration determinations), a simple reference is sufficient. If several alternative methods are commonly used, it is helpful to identify the method briefly as well as to cite the reference. For example, it is preferable to state "cells were broken by ultrasonic treatment as previously described (9)" rather than to state "cells were broken as previously described (9)." The reader should be allowed to assess the method without constant reference to previous publications. Describe new methods completely and give sources of unusual chemicals, equipment, or microbial strains. When large numbers of microbial strains or mutants are used in a study, include tables identifying the immediate sources (i.e., sources from whom the strains were obtained) and properties of the strains, mutants, bacteriophages, plasmids, etc.

A method, strain, etc., used in only one of several experiments reported in the paper may be described in the Results section or very briefly (one or two sentences) in a table footnote or figure legend. It is expected that the sources from whom the strains were obtained will be identified.

Results. The Results section should include the results of the experiments. Reserve extensive interpretation of the results for the Discussion section. Present the results as concisely as possible in **one** of the following: text, table(s), or figure(s). Avoid extensive use of graphs to present data that might be more concisely presented in the text or tables. For example, except in unusual cases, double-reciprocal plots used to determine apparent K_m values should not be presented as graphs; instead, the values should be stated in the text. Similarly, graphs illustrating other methods commonly used to derive kinetic or physical constants (e.g., reduced-viscosity plots and plots used to determine sedimentation velocity) need not be shown except in unusual circumstances. Limit photographs (particularly photomicrographs and electron micrographs) to those that are absolutely necessary to show the experimental findings. Number figures and tables in the order in which they are cited in the text, and be sure to cite all figures and tables.

Discussion. The Discussion should provide an interpretation of the results in relation to previously published work and to the experimental system at hand and should not contain extensive repetition of the Results section or reiteration of the introduction. In short papers, the Results and Discussion sections may be combined.

Acknowledgments. The source of any financial support received for the work being published must be indicated in the Acknowledgments section. (It will be assumed that the absence of such an acknowledgment is a statement by the authors that no support was received.) The usual format is as follows: "This work was supported by Public Health Service grant CA-01234 from the National Cancer Institute."

Recognition of personal assistance should be given as a separate paragraph, as should any statements disclaiming endorsement or approval of the views reflected in the paper or of a product mentioned therein.

Appendixes. Appendixes, which contain additional material to aid the reader, are permitted. Titles, authors, and References sections that are distinct from those of the primary article are not allowed. If it is not feasible to list the author(s) of the appendix in the byline or the Acknowledgments section of the primary article, rewrite the appendix so that it can be considered for publication as an independent article, either full-length or Note style. Equations, tables, and figures should be labeled with the letter "A" preceding the numeral to distinguish them from those cited in the main body of the text.

References. (i) Works listed in References. The References section must include all journal articles (both print and online), books and book chapters (both print

and online), patents, theses and dissertations, and published conference proceedings (not abstracts; see below), as well as in-press journal articles, book chapters, and books (publication title must be given). Arrange the citations in **alphabetical order** (letter by letter, ignoring spaces and punctuation) by first author and **number consecutively.** Provide the names of **all** the authors for each reference. All listed references **must** be cited parenthetically by number in the text. Since title and byline information that is downloaded from PubMed does not show accents, italics, or special characters, authors should refer to the PDF files or hard-copy versions of the articles and incorporate the necessary corrections in the submitted manuscript. Abbreviate journal names according to *BIOSIS Serial Sources* (BIOSIS, Philadelphia, Pa., 2005).

Follow the styles shown in the examples below.

Print references:

1. **Arendsen, A. F., M. Q. Solimar, and S. W. Ragsdale.** 1999. Nitrate-dependent regulation of acetate biosynthesis and nitrate respiration by *Clostridium thermoaceticum*. J. Bacteriol. **181:**1489–1495.
2. **Cox, C. S., B. R. Brown, and J. C. Smith.** J. Gen. Genet., in press.* {*Article title is optional; journal title is mandatory.*}
3. **da Costa, M. S., M. F. Nobre, and F. A. Rainey.** 2001. Genus I. *Thermus* Brock and Freeze 1969, 295,[AL] emend. Nobre, Trüper and da Costa 1996b, 605, p. 404–414. *In* D. R. Boone, R. W. Castenholz, and G. M. Garrity (ed.), Bergey's manual of systematic bacteriology, 2nd ed., vol. 1. Springer, New York, N.Y.
4. **Elder, B. L., and S. E. Sharp.** 2003. Cumitech 39, Competency assessment in the clinical laboratory. Coordinating ed., S. E. Sharp. ASM Press, Washington, D.C.
5. **Fitzgerald, G., and D. Shaw.** *In* A. E. Waters (ed.), Clinical microbiology, in press. EFH Publishing Co., Boston, Mass.* {*Chapter title is optional.*}
6. **Forman, M. S., and A. Valsamakis.** 2003. Specimen collection, transport, and processing: virology, p. 1227–1241. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, J. H. Jorgensen, and R. H. Yolken (ed.), Manual of clinical microbiology, 8th ed. ASM Press, Washington, D.C.
7. **Green, P. N., D. Hood, and C. S. Dow.** 1984. Taxonomic status of some methylotrophic bacteria, p. 251–254. *In* R. L. Crawford and R. S. Hanson (ed.), Microbial growth on C_1 compounds. Proceedings of the 4th International Symposium. American Society for Microbiology, Washington, D.C.
8. **Odell, J. C.** April 1970. Process for batch culturing. U.S. patent 484,363,770. {*Include the name of the patented item/process if possible.*}
9. **O'Malley, D. R.** 1998. Ph.D. thesis. University of California, Los Angeles. {*Title is optional.*}

*A reference to an in-press ASM publication should state the control number (e.g., EC00577-06) if it is a journal article or the name of the publication if it is a book.

Online references:

1. **Charlier, D., and N. Glansdorff.** September 2004, posting date. Biosynthesis of arginine and polyamines. *In* R. Curtiss III et al. (ed.), EcoSal—*Escherichia coli* and *Salmonella*: cellular and molecular biology, chapter 3.6.1.10. [Online.] http://www.ecosal.org. ASM Press, Washington, D.C. {*For online-only books or continually updated Web resources [for the latter, posting or accession date is required, but publisher's name and location are optional].*}
2. **Dimick, J. B., H. G. Welch, and J. D. Birkmeyer.** 18 August 2004, posting {*or* revision} date. Surgical mortality as an indicator of hospital quality. JAMA **292**. [Online.] http://jama.ama-assn.org/cgi/content/short/292/7/847. {*For online journals; page numbers may not be available.*}
3. **Sullivan, C. J. (ed.).** 1999–2001. Fungi: an evolving electronic resource for the microbiological community. ASM Press. [Online.] http://link.asmusa.de/link/service/books/91090. Accessed 7 September 2001. {*For online-only books.*}
4. **Zellnitz, F., and P. M. Foley.** 2 October 1998, posting {*or* revision} date. History of virology. Am. Virol. J. **1:**30–50. [Online.] http://www.avj.html. {*For online-only journals; page numbers may not be available.*}
5. **Zheng, Z., and J. Zou.** 5 September 2001. The initial step of the glycerolipid pathway: identification of glycerol-3-phosphate/dihydroxyacetone phosphate dual substrate acyltransferases in Saccharomyces cerevisiae. J. Biol. Chem. doi:10.1074/jbc.M104749200. {*For papers published online in manuscript form.*}

NOTE: A URL or DOI is necessary for each online-only reference; a posting or accession date is required for any online reference that is periodically updated or changed.

(ii) Items cited in the text. References to unpublished data, articles submitted for publication, meeting abstracts (including those published in journal supplements), personal communications, letters (irrespective of type) and authors' replies to letters, company publications, patent applications and patents pending, computer software, databases, and websites should be made parenthetically in the text as follows.

. . . similar results (R. B. Layton and C. C. Weathers, unpublished data).

. . . system was used (J. L. McInerney, A. F. Holden, and P. N. Brighton, submitted for publication).

. . . in mitochondria (S. De Wit, C. Thioux, and N. Clumeck, Abstr. 34th Intersci. Conf. Antimicrob. Agents Chemother., abstr. 114, 1994).

. . . for other bacteria (A. X. Jones, personal communication).

. . . discussed previously (L. B. Jensen, A. M. Hammerum, R. L. Poulsen, and H. Westh, Letter, Antimicrob. Agents Chemother. **43:**724–725, 1999).

. . . discussed previously (S. L. W. On and P. A. R. Vandamme, Authors' Reply to Letter, J. Clin. Microbiol. **39:**2751–2752, 2001).

. . . the manufacturer (Sigma manual, Sigma Chemical Co., St. Louis, Mo.).

. . . this process (V. R. Smoll, 20 June 1999, Australian Patent Office). {*For non-U.S. patent applications, give the date of publication of the application.*}

. . . information found at the XYZ website (http://cbx_iou.pgr).

. . . the ABC program (version 2.2; Department of Microbiology, State University [http://www.stu.micro]).

URLs for companies that produce any of the products mentioned in your study or for products being sold may NOT be included in the article. However, company URLs that permit access to scientific data related to the study or to shareware used in the study are permitted.

Notes

The Note format is intended for the presentation of brief observations that do not warrant full-length papers. Submit Notes in the same way as full-length papers. *They receive the same review, they are not published more rapidly than full-length papers, and they are not considered preliminary communications.*

Each Note must have an **abstract of no more than 50 words**. Do not use section headings in the body of the Note; combine methods, results, and discussion in a single section. Paragraph lead-ins are permissible. The text should be kept to a minimum and if possible **should not exceed 1,000 words**; the number of figures and tables should also be kept to a minimum. **Materials and methods should be described in the text, not in the figure legends or table footnotes.** Present acknowledgments as in full-length papers, but do not use a heading. The References section is identical to that of full-length papers.

Minireviews

Minireviews are brief (**limit of 6 printed pages exclusive of references**) biographical profiles, historical perspectives, or summaries of developments in fast-moving areas. They must be based on published articles; they may address any subject within the scope of the journal.

Minireviews may be either solicited or proffered by authors responding to a recognized need. Irrespective of origin, Minireviews are subject to review and should be submitted via Rapid Review. The cover letter should state whether the article was solicited and by whom.

Minireviews do not have abstracts. In the Abstract section of the submission form, put "Not applicable." The body of the Minireview may either have section headings or be set up like a Note (see above).

Guest Commentaries

Guest Commentaries are communications written in response to invitations issued by the editors and concern relevant topics in eukaryotic microbiology that are not necessarily covered by Minireviews. They should raise issues of interest to the scholarly community, initiate or focus discussion, and propose needed position or consensus statements by leadership groups in research and education. Reviews of the literature, methods and other how-to-papers, and responses targeted at a specific published paper are not appropriate. Guest Commentaries are subject to review.

The length may not exceed 4 printed pages, and the format is like that of a Minireview (see above). Commentaries should be submitted via Rapid Review.

Errata

The Erratum section provides a means of correcting errors that occurred during the writing, typing, editing, or printing (e.g., a misspelling, a dropped word or line, or mislabeling in a figure) of a published article. Send Errata directly to the ASM Journals Department (1752 N St., N.W., Washington, DC 20036-2904, USA), both on disk and in hard copy **(only one hard copy is necessary). Please see a recent issue for correct formatting.**

Authors' Corrections

The Author's Correction section provides a means of correcting errors of omission (e.g., author names or citations) and errors of a scientific nature that do not alter the overall basic results or conclusions of a published article.

For omission of an author's name, the authors of the article and author whose name was inadvertently omitted must agree, in writing, to publication of the Correction. For other issues involving authorship, including contributions and use or ownership of data and/or materials, all disputing parties must agree, in writing, to publication of the Correction. Copies of the agreement letters must accompany the Correction and be sent directly to the Journals Department. Send the Correction both on disk and in hard copy **(only one hard copy is necessary). Please see a recent issue for correct formatting.**

Corrections of a scientific nature (e.g., an incorrect unit of measurement or order of magnitude used throughout; contamination of one of numerous cultures; or misidentification of a mutant strain, causing erroneous data for only a portion [noncritical] of the study) must be sent, both on disk and in hard copy, directly to the editor who handled the article and must be accompanied by *signed letters of agreement* from all of the authors of the article. If the editor believes that publication is warranted, he will send the

Correction to the Journals Department for publication. *Note that the addition of new data is not permitted*.

Retractions

Retractions are reserved for major errors or breaches of ethics that, for example, may call into question the source of the data or the validity of the results and conclusions of an article. Send a Retraction and an accompanying explanatory letter *signed by all of the authors* directly to the editor in chief of the journal. The editor who handled the paper and the chairman of the ASM Publications Board will be consulted. If all parties agree to the publication and content of the Retraction, it will be sent to the Journals Department for publication.

ILLUSTRATIONS AND TABLES

Digital files that are acceptable for production (see below) must be provided for all illustrations on return of the modified manuscript. (On initial submission, the entire paper may be submitted in PDF format.)

We strongly recommend that before returning their modified manuscripts, authors check the acceptability of their digital images for production by running their files through Rapid Inspector, a tool provided at the following URL: http://rapidinspector.cadmus.com/mw/. Rapid Inspector is an easy-to-use Web-based application that identifies file characteristics that may render the image unusable for production.

Illustrations may be continuous-tone images, line drawings, or composites. Color graphics may be submitted, but the cost of printing in color must be borne by the author. Suggestions about how to reduce costs and ensure accurate color reproduction are given below.

The preferred format for tables is MS Word; however, WordPerfect and Acrobat PDF are also acceptable (see the section on Tables below).

Macintosh

Application	File type	
	Black and white	Color (CMYK)[a]
Adobe Illustrator 6.0, 7.0, 8.0, 9.0, 10.0, 11.0 CS	EPS	EPS
Adobe InDesign 1.0	EPS	EPS
Adobe PageMaker 6.5	EPS	EPS
Adobe Photoshop 4.0, 5.0, 5.5, 6.0, 7.0, 8.0 CS	TIFF	TIFF
Adobe Photoshop 5.0 LE	TIFF	N/A[b]
ChemDraw Pro 5.0	EPS/TIFF	EPS/TIFF
Corel Photo-Paint 8.0	TIFF	EPS
CorelDRAW 6.0, 8.0	EPS/TIFF	EPS
Deneba Canvas 6.0, 7.0, 8.0	EPS/TIFF	EPS
Macromedia FreeHand 7.0, 8.0, 9.0	EPS	EPS
PowerPoint 98, 2001	PPT[c]	N/A[b]
Prism 3 by GraphPad	TIFF	N/A[b]
Synergy Kaleidagraph 3.08, 3.51	EPS	N/A[b]

[a] Color graphics must be saved and printed in the CMYK mode, *not* RGB.
[b] ASM accepts only black-and-white, not color, graphics created with Kaleidagraph, Adobe Photoshop 5.0 LE, Prism 3 by GraphPad, and PowerPoint.
[c] For instructions on saving PowerPoint files, refer to the Cadmus digital art website at http://cjs.cadmus.com/da/index.asp.

Windows

Application	File type	
	Black and white	Color (CMYK)[a]
Adobe Illustrator 7.0, 8.0, 9.0, 10.0, 11.0 CS	EPS	EPS
Adobe InDesign 1.0	EPS	EPS
Adobe PageMaker 6.5	EPS	EPS
Adobe Photoshop 4.0, 5.0, 5.5, 6.0, 7.0, 8.0 CS	TIFF	TIFF
Adobe Photoshop 5.0 LE	TIFF	N/A[b]
ChemDraw Pro 5.0	EPS/TIFF	EPS/TIFF
Corel Photo-Paint 8.0, 9.0	TIFF	EPS
CorelDRAW 7.0, 8.0, 9.0	EPS/TIFF	EPS
Deneba Canvas 6.0, 7.0	EPS/TIFF	EPS
Macromedia FreeHand 7.0, 8.0, 9.0	EPS	EPS
PowerPoint 97, 2000, XP	PPT[c]	N/A[b]
Prism 3 by GraphPad	TIFF	N/A[b]
SigmaPlot 8.01	EPS	EPS

[a] Color graphics must be saved and printed in the CMYK mode, *not* RGB.
[b] ASM accepts only black-and-white, not color, graphics created with Adobe Photoshop 5.0 LE, Prism 3 by GraphPad, and PowerPoint.
[c] For instructions on saving PowerPoint files, refer to the Cadmus digital art website at http://cjs.cadmus.com/da/index.asp.

Since the contents of computer-generated images can be manipulated for better clarity, the Publications Board at its May 1992 meeting mandated that a description of the software/hardware used should be put in the figure legend(s).

Illustrations

File types and formats. As mentioned above, **illustrations may be supplied as PDF files for reviewing purposes only on initial submission; in fact, we recommend this option to minimize file upload time. At the modification stage, production quality digital files must be submitted:** TIFF or EPS files from supported applications or PowerPoint files (black and white only). Except for figures produced in PowerPoint, all graphics submitted with modified manuscripts must be bitmap, grayscale, or CMYK (*not* RGB). Acceptable file types and formats for production are given in the charts above. More-detailed instructions for preparing illustrations are available on the World Wide Web at http://cjs.cadmus.com/da. Please review this information before preparing your files. If you require additional information, please send an e-mail inquiry to digitalart@cadmus.com.

Minimum resolution. It is extremely important that a high enough resolution is used. Any imported images must be at the correct resolution before they are placed. Note, however, that the higher the resolution, the larger the file and the longer the upload time. Publication quality will *not* be improved by using a resolution higher than the minimum. Minimum resolutions are as follows:

300 dpi for grayscale and color
600 dpi for lettering
1,200 dpi for line art
600 dpi for combination art (lettering and images)

Size. All graphics **MUST be submitted at their intended publication size;** that is, the image uploaded should be 100% of its print dimensions so that no reduction or enlargement is necessary. Resolution must be at the required level at the submitted size. Include only the significant portion of an illustration. White space must be cropped from the image, and excess space between panel labels and the image must be eliminated.

Maximum width for a 1-column figure: 3 5/16 inches (ca. 8.4 cm)
Maximum width for a 2-column figure: 6 7/8 inches (ca. 17.4 cm)
Minimum width for a 2-column figure: 4 1/4 inches (10.8 cm)
Maximum height: 9 1/16 inches (23.0 cm)

Contrast. Illustrations must contain sufficient contrast to withstand the inevitable loss of contrast and detail inherent in the printing process. See also the section on color illustrations below.

Labeling and assembly. All final lettering, labeling, tooling, etc., MUST be incorporated into the figures. It cannot be added at a later date. If a figure number is included, it **must** appear well outside the boundaries of the image itself. (Numbering may need to be changed at the copyediting stage.) Each figure must be uploaded as a separate file, and any multipanel figures must be assembled into one file; i.e., rather than uploading a separate file for each panel in a figure, assemble all panels in one piece and supply them as one file.

Fonts. To avoid font problems, set all type in one of the following fonts: Helvetica, Times Roman, European PI, Mathematical PI, or Symbol. All fonts other than these five must be converted to paths (or outlines) in the application with which they were created. For font use in PowerPoint images, refer to the Cadmus digital art website, http://cjs.cadmus.com/da.

Compression. Images created with Macintosh applications may be compressed with Stuffit. Images created with Windows applications may be compressed with WINZIP or PKZIP.

Color illustrations. Because the process of placing ink on paper by using printing presses is different from that used to produce a photo print or a laser print and the color rendition on images viewed on a monitor depends to some extent on monitor resolution, some differences in color and contrast between the image you submit and the image printed in the journal or published online will be evident. (Figures showing red or green fluorescence and those with a significant range of colors may be difficult or impossible to reproduce exactly.) Color illustrations must be saved as either TIFF or EPS files, according to the application used (see charts above). The mode of the TIFF or EPS file must be CMYK, *not* RGB. Graphics in the RGB color space are intended for display on a monitor only and will not separate correctly for printing.

The cost of printing in color must be borne by the author. The current color costs may be accessed from the submission form in Rapid Review and, for accepted manuscripts, will be included in the acceptance letter sent out by ASM. Adherence to the following guidelines, in addition to the general ones above, will help to minimize costs and to ensure color reproduction that is as accurate as possible.

Include only the significant portions of illustrations so that the number of printed pages containing color figures is minimized. The individual panels of a single figure must be assembled in a single file, including any necessary labels. Optimal color reproduction will be obtained if the composites comprise panels containing similar colors of similar lightness or darkness. If necessary, make unlike panels into separate figures/files; this will increase the cost, but the color rendition will be more accurate since the two panels will be "scanned" separately.

Drawings

Submit graphs, charts, complicated chemical or mathematical formulas, diagrams, and other drawings as finished products not requiring additional artwork or typesetting. No part of the graph or drawing may be handwritten. *All* elements, including letters, numbers, and symbols, *must* be easily readable, and both axes of a graph must be labeled. Keep in mind that the journal is published both in print and online and that the same electronic files submitted by the authors are used to produce both.

When creating line art, please use the following guidelines:

1. **All art MUST be submitted at its intended publication size.** For acceptable dimensions, see the Size section on p. 11.
2. **Avoid using screens (i.e., shading)** in line art. It can be difficult and time-consuming to reproduce these images without moiré patterns. Various pattern backgrounds are preferable to screens as long as the fill patterns are not imported from another application. If you must use images containing screens,
   - Generate the image at line screens of 85 lines per inch or lower.
   - When applying multiple shades of gray, differentiate the gray levels by at least 20%.
   - Never use levels of gray below 20% or above 70% as they will fade out or become totally black upon scanning and reduction.
3. Use thick, solid lines that are no finer than 1 point in thickness.
4. No type should be smaller than 6 points at the final publication size.

5. Avoid layering type directly over shaded or textured areas.
6. Avoid the use of reversed type (white lettering on a black background).
7. Avoid heavy letters, which tend to close up, and unusual symbols, which the printer may not be able to reproduce in the legend.
8. If colors are used, avoid using similar shades of the same color and avoid very light colors.

In figure ordinate and abscissa scales (as well as table column headings), **avoid the ambiguous use of numbers with exponents.** Usually, it is preferable to use the appropriate Système International d'Unités (SI) symbols (μ for 10^{-6}, m for 10^{-3}, k for 10^{3}, M for 10^{6}, etc.). A complete listing of SI symbols can be found in the International Union of Pure and Applied Chemistry (IUPAC) "Manual of Symbols and Terminology for Physico-chemical Quantities and Units" (Pure Appl. Chem. **21:**3–44, 1970). Thus, representation of 20,000 cpm on a figure ordinate should be made by the number 20 accompanied by the label kcpm.

Where powers of 10 must be used, the journal requires that the exponent power be associated with the number shown. In representing 20,000 cells per ml, the numeral on the ordinate would be "2" and the label would be "10^4 cells per ml" (not "cells per ml $\times$ 10^{-4}"). Likewise, an enzyme activity of 0.06 U/ml would be shown as 6 accompanied by the label 10^{-2} U/ml. The preferred designation would be 60 mU/ml (milliunits per milliliter).

Presentation of Nucleic Acid Sequences

Nucleic acid sequences of limited length which are the primary subject of a study may be presented freestyle in the most effective format. Longer nucleic acid sequences must be presented as figures in the following format to conserve space. Print the sequence in lines of approximately 100 to 120 nucleotides in a nonproportional (monospace) font that is easily legible when published with a line length of 6 inches (ca. 15.2 cm). If possible, lines of nucleic acid sequence should be further subdivided into blocks of 10 or 20 nucleotides by spaces within the sequence or by marks above it. Uppercase and lowercase letters may be used to designate the exon-intron structure, transcribed regions, etc., if the lowercase letters remain legible at a 6-inch (ca. 15.2-cm) line length. Number the sequence line by line; place numerals, representing the first base of each line, to the left of the lines. **Minimize spacing between lines of sequence, leaving room only for annotation of the sequence.** Annotation may include boldface, underlining, brackets, boxes, etc. Encoded amino acid sequences may be presented, if necessary, immediately above or below the first nucleotide of each codon, by using the single-letter amino acid symbols. Comparisons of multiple nucleic acid sequences should conform as nearly as possible to the same format.

Figure Legends

Legends should provide enough information so that the figure is understandable without frequent reference to the text. However, detailed experimental methods must be described in the Materials and Methods section, not in a figure legend. A method that is unique to one of several experiments may be reported in a legend only if the discussion is very brief (one or two sentences). Define all symbols used in the figure and define all abbreviations that are not used in the text.

Tables

Tables that contain artwork, chemical structures, or shading must be submitted as illustrations in an acceptable format at the modification stage. The preferred format for regular tables is MS Word; however, WordPerfect and Acrobat PDF are also acceptable. Note that a straight Excel file is *not* currently an acceptable format. Excel files must be either embedded in a Word or WordPerfect document or converted to PDF *before* being uploaded. **If your modified manuscript contains PDF tables, select "for reviewing purposes only" at the beginning of the file upload process.**

Tables should be formatted as follows. Arrange the data so that **columns of like material read down, not across.** The headings should be sufficiently clear so that the meaning of the data is understandable without reference to the text. See the Abbreviations section (p. 16) of these Instructions for those that should be used in tables. Explanatory footnotes are acceptable, but more extensive table "legends" are not. Footnotes should not include detailed descriptions of the experiment. Tables must include enough information to warrant table format; those with fewer than six pieces of data will be incorporated into the text by the copy editor. Table 1 is an example of a well-constructed table.

TABLE 1. Effect of glucose on levels of catabolic enzymes and morphology in *M. rouxii*

Cell type	Enzyme activity				
	Pyruvate kinase[a]		Phospho-fructo-kinase[b]	Glutamate dehydro-genase[c]	Pyruvate decarboxy-lase[c]
	1 min	5 min			
Mycelium					
−Glucose	1,056	2	1.7	4.3	0.05
+Glucose	2,930	10	8.04	0.53	1.3
Yeast					
−Glucose	1,145	4	23.2	4.1	0.04
+Glucose	4,380	30	63.6	0.03	1.7

[a] Nanomoles of pyruvate formed per milligram of protein in time shown (3, 9).
[b] Millimoles of fructose 1,6-diphosphate produced per minute per milligram of protein (7).
[c] Micromoles of NADH oxidized per minute per milligram of protein (10).

Cover Photographs and Drawings

EC publishes photographs and drawings on the front cover. Invitations are issued to authors whose manuscripts

are returned for modification or whose manuscripts have been accepted for publication in EC; material should be related to the work presented in the EC manuscript. Unsolicited photos will be considered in hard-copy format (two copies) only; if an unsolicited photo is chosen for the cover, the author may be asked to submit digital files. No material submitted for consideration will be returned to the author. Authors will be notified only if their cover art is selected. Copyright for the chosen material must be transferred to ASM. A short description of the cover material will be included at the end of the table of contents or the author index of the issue. Technical specifications for submission are available from the cover editor, C. C. Wang (e-mail: ccwang@cgl.ucsf.edu).

NOMENCLATURE

Chemical and Biochemical Nomenclature

The recognized authority for the names of chemical compounds is *Chemical Abstracts* (CAS, Columbus, Ohio) and its indexes. *The Merck Index*, 13th ed. (Merck & Co., Inc., Whitehouse Station, N.J., 2001), is also an excellent source. For guidelines to the use of biochemical terminology, consult *Biochemical Nomenclature and Related Documents* (1978; reprinted for The Biochemical Society, London, England) and the instructions to authors of the *Journal of Biological Chemistry* and the *Archives of Biochemistry and Biophysics* (first issues of each year).

Do not express molecular weight in daltons; molecular weight is a unitless ratio. Molecular mass is expressed in daltons.

For enzymes, use the recommended (trivial) name assigned by the Nomenclature Committee of the International Union of Biochemistry (IUB) as described in *Enzyme Nomenclature* (Academic Press, Inc., New York, N.Y., 1992) and at http://www.chem.qmul.ac.uk/iubmb/enzyme/. If a nonrecommended name is used, place the proper (trivial) name in parentheses at first use in the abstract and text. Use the EC number when one has been assigned, and express enzyme activity either in katals (preferred) or in the older system of micromoles per minute.

Nomenclature of Mice

For mouse strain and genetic nomenclature, ASM encourages authors to refer to the guidelines set forth by the International Committee on Standardized Genetic Nomenclature for Mice, available on the Mouse Genome Database home page at http://www.informatics.jax.org and in *Genetic Variants and Strains of the Laboratory Mouse*, 3rd ed. (M. F. Lyon et al., ed., Oxford University Press, Oxford, England, 1996).

Nomenclature of Microorganisms

Binary names, consisting of a generic name and a specific epithet (e.g., *Saccharomyces cerevisiae*), must be used for all microorganisms. Names of categories at or above the genus level may be used alone, but specific and subspecific epithets may not. A specific epithet must be preceded by a generic name, written out in full the first time it is used in a paper. Thereafter, the generic name should be abbreviated to the initial capital letter (e.g., *S. cerevisiae*), provided there can be no confusion with other genera used in the paper. Names of all taxa (kingdoms, phyla, classes, orders, families, genera, species, and subspecies) are printed in italics and should be underlined (or italicized) in the manuscript; strain designations and numbers are not.

The spelling of bacterial names should follow the *Approved Lists of Bacterial Names (Amended) & Index of the Bacterial and Yeast Nomenclatural Changes* (V. B. D. Skerman et al., ed., ASM Press, Washington, D.C., 1989) and the validation lists and notification lists published in the *International Journal of Systematic and Evolutionary Microbiology* (formerly the *International Journal of Systematic Bacteriology*) since January 1989. In addition, two sites on the World Wide Web list current approved bacterial names: Bacterial Nomenclature Up-to-Date (http://www.dsmz.de/microorganisms/main.php?contentleft_id=14) and List of Prokaryotic Names with Standing in Nomenclature (http://www.bacterio.cict.fr).

Since the classification of fungi is far from complete, it is the responsibility of the author to determine the accepted binomial for a given organism. Sources for these names include *The Yeasts: a Taxonomic Study*, 4th ed. (C. P. Kurtzman and J. W. Fell, ed., Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1998), and *Ainsworth and Bisby's Dictionary of the Fungi*, 9th ed. (P. M. Kirk, P. F. Cannon, J. C. David, and J. A. Stalpers, ed., CABI Publishing, Wallingford, Oxfordshire, United Kingdom, 2001).

Names used for viruses should be those approved by the International Committee on Taxonomy of Viruses (ICTV) and published in *Virus Taxonomy: Classification and Nomenclature of Viruses, Seventh Report of the International Committee on Taxonomy of Viruses* (M. H. V. van Regenmortel et al., ed., Academic Press, San Diego, Calif., 2000). In addition, the recommendations of the ICTV regarding the use of species names should generally be followed: when the entire species is discussed as a taxonomic entity, the species name, like other taxa, is italic and has the first letter and any proper nouns capitalized (e.g., *Tobacco mosaic virus*, *Murray Valley encephalitis virus*). When the behavior or manipulation of individual viruses is discussed, the vernacular (e.g., tobacco mosaic virus, Murray Valley encephalitis virus) should be used. If desired, synonyms may be added parenthetically when the name is first mentioned. Approved generic (or group) and family names may also be used.

Microbial strains, viruses, and plasmids should be given individual designations consisting of letters and serial numbers. It is generally advisable to include a worker's initials or a descriptive symbol of locale, laboratory, etc., in the designation. Each new strain, mutant, isolate, or derivative should be given a new (serial) designation. This designation should be distinct from those of the

genotype and phenotype, and genotypic and phenotypic symbols should not be included.

Genetic Nomenclature

To facilitate accurate communication, **it is important that standard genetic nomenclature be used whenever possible and that deviations or proposals for new naming systems be endorsed by an appropriate authoritative body**. Review and/or publication of submitted manuscripts that contain new or nonstandard nomenclature may be delayed by the editor or the Journals Department so that they may be reviewed by the Genetics and Genomics Committee of the ASM Publications Board.

Before submission of manuscripts, authors may direct questions on genetic nomenclature to the committee's chairman: Maria Costanzo (e-mail: maria@genome.stanford.edu). Such a consultation should be mentioned in the manuscript submission letter.

Eukaryotes. The nomenclature used for the genetics of lower eukaryotic microorganisms has not been as well formalized as that for bacteria and bacteriophages. Generally, authors should conform to current practices in identifying mutants and their genotypes. For organisms not mentioned below, it is advisable to consult the *Handbook of Microbiology* (A. I. Laskin and H. A. Lechevalier, ed., CRC Press, 1974) or the *Handbook of Genetics*, vol. 1, *Bacteria, Bacteriophages, and Fungi* (R. C. King, ed., Plenum Publishing Corp., 1974).

The genetic nomenclature of *Dictyostelium* is summarized in the *Trends in Genetics* "Genetic Nomenclature Guide" (p. S.5–S.6; Elsevier Science Ltd., Cambridge, United Kingdom, 1998; out of print). The most recent modifications can be found at http://dictybase.org/GeneNames.html.

For *Saccharomyces cerevisiae*, a gene name should always be indicated in italics and, for the wild-type locus (or dominant alleles), capital letters (e.g., *URA3*). Loss-of-function (hypomorphic) or altered-function (neomorphic) alleles of the same locus should always be indicated in italics and lowercase letters (e.g., *ura3*Δ). The product of a gene (i.e., a protein) should be indicated in Roman type with an initial capital letter (e.g., Ura3). There is generally no need to add the suffix "p" to the symbol for a protein; however, in rare instances, where it may be deemed necessary to indicate unambiguously that the symbol refers to a protein, the "p" suffix may be added (e.g., Ura3p). For the most recent information on *S. cerevisiae* gene names, consult the *Saccharomyces* Genome Database (SGD) at http://www.yeastgenome.org. Details on the format of *S. cerevisiae* locus and allele designations are provided in the SGD and are also described by Cherry (Trends Genet. **March:**11–12, 1995) (available for download as a PDF file at SGD, http://www.yeastgenome.org/sgdpub/Saccharomyces_cerevisiae.pdf). **Authors should use standard *S. cerevisiae* gene names, as listed in the SGD, in their submitted manuscripts and register new gene names with the SGD no later than the modification stage.**

The most recent information for *Neurospora crassa* can be found in *The Neurospora Compendium: Chromosomal Loci* (D. R. Perkins et al., Academic Press, San Diego, Calif., 2001), and that for *Aspergillus* spp. can be found at http://www.fgsc.net/nomenclature.htm. The 1998 *Trends in Genetics* "Genetic Nomenclature Guide" (Elsevier Science Ltd., Cambridge, United Kingdom; out of print) contains nomenclature guidelines for several eukaryotic microbes: *Schizosaccharomyces pombe* (p. S.7–S.9), *Chlamydomonas reinhardtii* (p. S.18–S.19), *Neurospora crassa* (p. S.14–S.15), and *Aspergillus nidulans* (p. S.12–S.13). In addition, for *S. pombe*, the websites http://www.sanger.ac.uk/Projects/S_pombe/SP_Name_FAQ.shtml and http://www-rcf.usc.edu/~forsburg/plasmids.html#nomenclature may be helpful, and for *C. reinhardtii*, use http://www.biology.duke.edu/chlamydb/.

For *Trypanosoma* and *Leishmania*, consult the article by Clayton et al. (Mol. Biochem. Parasitol. **97:**221–224, 1998).

For the most recent information on *Candida albicans*, consult the *Candida* Genome Database (CGD) at http://www.candidagenome.org. Details on the format of *C. albicans* gene nomenclature are described at http://www.candidagenome.org/Nomenclature.shtml. **Authors should use standard *C. albicans* gene names, as listed in the CGD, in their submitted manuscripts and should register new gene names with the CGD no later than the modification stage.**

Prokaryotes. The genetic properties of prokaryotes are described in terms of phenotypes and genotypes. The phenotype describes the observable properties of an organism. The genotype refers to the genetic constitution of an organism, usually in reference to some standard wild type. Use the recommendations of Demerec et al. (Genetics **54:**61–76, 1966) as a guide to the use of these terms. If your manuscript contains genetic nomenclature, please refer to the Instructions to Authors in the January issue of the *Journal of Bacteriology.*

Viruses. In most cases, viruses have no phenotype, since they have no metabolism outside host cells. Therefore, distinctions between phenotype and genotype are not made. Superscripts are used to indicate hybrid genomes. Genetic symbols may be one, two, or three letters. For example, a mutant strain of lambda may be designated λ *c*I857 *int*2 *red*114 *A*am11; this strain carries mutations in genes *c*I, *int*, and *red* and an amber-suppressible (am) mutation in gene *A*. Host DNA insertions into viruses should be delineated by square brackets, and the genetic symbols and designations for such inserted DNA should conform to those used for the host genome.

Conventions for naming genes. It is recommended that (entirely) new genes be given names that are mnemonics of their function, avoiding names that are already assigned and earlier or alternative gene names, irrespective of the bacterium for which such assignments have been made. Similarly, it is recommended that, whenever possible, homologous genes present in different organisms receive the same name. When homology

is not apparent or the function of a new gene has not been established, a provisional name may be given by one of the following methods. (i) The gene may be named on the basis of its map location in the style *yaaA*, analogous to the style used for recording transposon insertions (*zef*) as discussed below. (ii) A provisional name may be given in the style described by Demerec et al. (e.g., *usg*, gene upstream of *folC*). Such names should be unique, and names such as *orf* or *genX* should not be used.

"Homology" versus "similarity." For use of terms that describe relationships between genes, consult the articles by Theissen (Nature **415:**741, 2002) and Fitch (Trends Genet. **16:**227–231, 2000). "Homology" implies a relationship between genes that share a common evolutionary origin; partial homology is not recognized. When sequence comparisons are discussed, it is more appropriate to use the term "percent sequence similarity" or "percent sequence identity," as appropriate.

"Mutant" versus "mutation." Keep in mind the distinction between a *mutation* (an alteration of the primary sequence of the genetic material) and a *mutant* (a strain carrying one or more mutations). One may speak about the mapping of a mutation, but one cannot map a mutant. Likewise, a mutant has no genetic locus, only a phenotype.

Transposable elements, plasmids, and restriction enzymes. Nomenclature of transposable elements (insertion sequences, transposons, phage Mu, etc.) should follow the recommendations of Campbell et al. (Gene **5:**197–206, 1979), with the modifications referred to in the Instructions to Authors in the *Journal of Bacteriology*.

The system of designating transposon insertions at sites where there are no known loci, e.g., *zef-123*::Tn*5*, has been described by Chumley et al. (Genetics **91:**639–655, 1979). Use the nomenclature recommendations of Novick et al. (Bacteriol. Rev. **40:**168–189, 1976) for plasmids and plasmid-specified activities, of Low (Bacteriol. Rev. **36:**587–607, 1972) for F′ factors, and of Roberts et al. (Nucleic Acids Res. **31:**1805–1812, 2003) for restriction enzymes, DNA methyltransferases, homing endonucleases, and their genes. The nomenclature for recombinant DNA molecules constructed in vitro follows the nomenclature for insertions in general. DNA inserted into recombinant DNA molecules should be described by using the gene symbols and conventions for the organism from which the DNA was obtained.

ABBREVIATIONS AND CONVENTIONS

Verb Tense

ASM strongly recommends that for clarity you use the **past** tense to narrate particular events in the past, including the procedures, observations, and data of the study that you are reporting. Use the present tense for your own general conclusions, the conclusions of previous researchers, and generally accepted facts. Thus, most of the abstract, Materials and Methods, and Results will be in the past tense, and most of the introduction and some of the Discussion will be in the present tense.

Be aware that it may be necessary to vary the tense in a single sentence. For example, it is correct to say "White (30) demonstrat*ed* that XYZ cells *grow* at pH 6.8," "Figure 2 show*s* that ABC cells fail*ed* to grow at room temperature," and "Air *was* removed from the chamber and the mice *died*, which *proves* that mice *require* air." In reporting statistics and calculations, it is correct to say "The values for the ABC cells *are* statistically significant, indicating that the drug inhibit*ed* . . . ."

For an in-depth discussion of tense in scientific writing, see p. 207–209 in *How To Write and Publish a Scientific Paper*, 5th ed.

Abbreviations

General. Abbreviations should be used as an aid to the reader, rather than as a convenience for the author, and therefore their **use should be limited**. Abbreviations other than those recommended by the IUPAC-IUB (*Biochemical Nomenclature and Related Documents*, 1978) should be used only when a case can be made for necessity, such as in tables and figures.

It is often possible to use pronouns or to paraphrase a long word after its first use (e.g., "the drug" or "the substrate"). Standard chemical symbols and trivial names or their symbols (folate, Ala, Leu, etc.) may also be used.

It is strongly recommended that all abbreviations except those listed below be introduced in the first paragraph in Materials and Methods. Alternatively, define each abbreviation and introduce it in parentheses the first time it is used; e.g., "cultures were grown in Eagle minimal essential medium (MEM)." Generally, eliminate abbreviations that are not used at least three times in the text (including tables and figure legends).

Not requiring introduction. In addition to abbreviations for Système International d'Unités (SI) units of measurement, other common units (e.g., bp, kb, and Da), and chemical symbols for the elements, the following should be used without definition in the title, abstract, text, figure legends, and tables: DNA (deoxyribonucleic acid); cDNA (complementary DNA); RNA (ribonucleic acid); cRNA (complementary RNA); RNase (ribonuclease); DNase (deoxyribonuclease); rRNA (ribosomal RNA); mRNA (messenger RNA); tRNA (transfer RNA); AMP, ADP, ATP, dAMP, ddATP, GTP, etc. (for the respective 5′ phosphates of adenosine and other nucleosides) (add 2′-, 3′-, or 5′- when needed for contrast); ATPase, dGTPase, etc. (adenosine triphosphatase, deoxyguanosine triphosphatase, etc.); NAD (nicotinamide adenine dinucleotide); NAD^+ (nicotinamide adenine dinucleotide, oxidized); NADH (nicotinamide adenine dinucleotide, reduced); NADP (nicotinamide adenine dinucleotide phosphate); NADPH (nicotinamide adenine dinucleotide phosphate, reduced); $NADP^+$ (nicotinamide ade-

nine dinucleotide phosphate, oxidized); poly(A), poly (dT), etc. (polyadenylic acid, polydeoxythymidylic acid, etc.); oligo(dT), etc. (oligodeoxythymidylic acid, etc.); UV (ultraviolet); PFU (plaque-forming units); CFU (colony-forming units); MIC (minimal inhibitory concentration); Tris [tris(hydroxymethyl)aminomethane]; DEAE (diethylaminoethyl); EDTA (ethylenediaminetetraacetic acid); EGTA [ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid]; HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid); PCR (polymerase chain reaction); and AIDS (acquired immunodeficiency syndrome). Abbreviations for cell lines (e.g., HeLa) also need not be defined.

The following abbreviations should be used without definition in tables:

amt (amount)
approx (approximately)
avg (average)
concn (concentration)
diam (diameter)
expt (experiment)
exptl (experimental)
ht (height)
mo (month)
mol wt (molecular weight)
no. (number)
prepn (preparation)
SD (standard deviation)
SE (standard error)
SEM (standard error of the mean)
sp act (specific activity)
sp gr (specific gravity)
temp (temperature)
tr (trace)
vol (volume)
vs (versus)
wk (week)
wt (weight)
yr (year)

Reporting Numerical Data

Standard metric units are used for reporting length, weight, and volume. For these units and for molarity, use the prefixes m, μ, n, and p for 10^{-3}, 10^{-6}, 10^{-9}, and 10^{-12}, respectively. Likewise, use the prefix k for 10^{3}. Avoid compound prefixes such as mμ or μμ. Use μg/ml or μg/g in place of the ambiguous ppm. Units of temperature are presented as follows: 37°C or 324 K.

When fractions are used to express units such as enzymatic activities, it is preferable to use whole units, such as "g" or "min," in the denominator instead of fractional or multiple units, such as μg or 10 min. For example, "pmol/min" is preferable to "nmol/10 min," and "μmol/g" is preferable to "nmol/μg." It is also preferable that an unambiguous form such as exponential notation be used; for example, "μmol g^{-1} min^{-1}," is preferable to "μmol/g/min." Always report numerical data in the appropriate SI units.

For a review of some common errors associated with statistical analyses and reports, plus guidelines on how to avoid them, see the article by Olsen (Infect. Immun. **71:**6689–6692, 2003).

For a review of basic statistical considerations for virology experiments, see the article by Richardson and Overbaugh (J. Virol. **79:**669–676, 2005).

Isotopically Labeled Compounds

For simple molecules, isotopic labeling is indicated in the chemical formula (e.g., $^{14}CO_2$, $^{3}H_2O$, and $H_2{}^{35}SO_4$). Brackets are not used when the isotopic symbol is attached to the name of a compound that in its natural state does not contain the element (e.g., ^{32}S-ATP) or to a word which is not a specific chemical name (e.g., ^{131}I-labeled protein, ^{14}C-amino acids, and ^{3}H-ligands).

For specific chemicals, the symbol for the isotope introduced is placed in square brackets directly preceding the part of the name that describes the labeled entity. Note that configuration symbols and modifiers precede the isotopic symbol. The following examples illustrate correct usage:

[^{14}C]urea
L-[*methyl*-^{14}C]methionine
[2,3-^{3}H]serine
[α-^{14}C]lysine
[γ-^{32}P]ATP
UDP-[U-^{14}C]glucose
E. coli [^{32}P]DNA
fructose 1,6-[1-^{32}P]bisphosphate

EC follows the same conventions for isotopic labeling as the *Journal of Biological Chemistry*, and more-detailed information can be found in the instructions to authors of that journal (first issue of each year).

8.2. Apêndice 3
PARTICIPAÇÃO DO DISCENTE EM ARTIGO PUBLICADO

Translation initiation in Leishmania major: characterisation of multiple eIF4F subunit homologues

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 140 (2005) 23–41
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues

Rafael Dhalia[a,e], Christian R.S. Reis[b,e], Eden R. Freire[b,e], Pollyanna O. Rocha[b,e], Rodolfo Katz[b,e], João R.C. Muniz[c], Nancy Standart[d], Osvaldo P. de Melo Neto[e,*]

[a] *Departamento de Biologia Celular, Universidade de Brasilia, Brasilia 70910-900, D.F., Brazil*
[b] *Departamento de Genética, Universidade Federal de Pernambuco, Avenida Professor Moraes Rego s/n, Cidade Universitária, Recife 50732-970, PE, Brazil*
[c] *Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos 13560-970, SP, Brazil*
[d] *Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB21GA, UK*
[e] *Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Avenida Professor Moraes Rego s/n, Cidade Universitária, Recife 50670-420, PE, Brazil*

Received 14 August 2004; received in revised form 25 November 2004; accepted 7 December 2004
Available online 29 December 2004

Abstract

In eukaryotes protein synthesis initiates with the binding of the multimeric translation initiation complex eIF4F – eIF4E, eIF4A and eIF4G – to the monomethylated cap present on the 5′ end of mRNAs. eIF4E interacts directly with the cap nucleotide, while eIF4A is a highly conserved RNA helicase and eIF4G acts as a scaffold for the complex with binding sites for both eIF4E and eIF4A. eIF4F binding to the mRNA recruits the small ribosomal subunit to its 5′ end. Little is known in detail of protein synthesis in the protozoan parasites belonging to the family *Trypanosomatidae*. However, the presence of the highly modified cap structure, cap4, and the spliced leader sequence on the 5′ ends of all mRNAs suggests possible differences in mRNA recruitment by ribosomes. We identified several potential eIF4F homologues by searching *Leishmania major* databases: four eIF4Es (*Lm*EIF4E1-4), two eIF4As (*Lm*EIF4A1-2) and five eIF4Gs (*Lm*EIF4G1-5). We report the initial characterisation of *Lm*EIF4E1-3, *Lm*EIF4A1-2 and *Lm*EIF4G3. First, the expression of these proteins in *L. major* promastigotes was quantitated by Western blotting using isoform specific antibodies. *Lm*EIF4A1 and *Lm*EIF4E3 are very abundant, *Lm*EIF4G3 is moderately abundant and *Lm*EIF4E1/*Lm*EIF4E2/*Lm*EIF4A2 are rare or not detected. In cap-binding assays, only *Lm*EIF4E1 bound to the 7-methyl-GTP-Sepharose resin. Molecular modelling confirmed that *Lm*EIF4E1 has all the structural features of a cap-binding protein. Finally, pull-down assays were used to investigate the potential interaction between the eIF4A (*Lm*EIF4A1/*Lm*EIF4A2) and eIF4G (*Lm*EIF4G1-3) homologues. Only *Lm*EIF4G3, via the HEAT domain, bound specifically both to *Lm*EIF4A1 as well as to human eIF4A. Therefore for each factor, one of the *L. major* forms seems to fulfil, in part at least, the expected characteristics of a translational initiation factor.

Keywords: eIF4F; Translation initiation; *Leishmania major*; Protein–protein interaction

1. Introduction

Flagellate protozoans belonging to the order *Kinetoplastida*, family *Trypanosomatidae*, are responsible for a series of important diseases, such as the Leishmaniasis, Sleeping Sickness and Chagas' Disease, which affect millions of people world-wide and put a great number of lives at risk (http://www.who.int/tdr). These are unique eukaryotes which are characterised by unusual processes in the formation of mature mRNAs, including polycistronic transcription and *trans*-splicing [1,2] (for recent reviews see [3,4]). These parasites are also distinguished by the lack of known mechanisms for the transcriptional control of the expression of

* Corresponding author. Tel.: +55 81 2101 2636; fax: +55 81 3453 2449.
E-mail address: opmn@cpqam.fiocruz.br (O.P. de Melo Neto).

0166-6851/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2004.12.001

their protein coding genes (reviewed in [5]). Processing of the polycistronic precursor mRNAs by *trans*-splicing adds the capped 39 nt long spliced leader (SL or mini-exon sequence) to the 5′ end of each monocistronic mature mRNA. The first 4 nt of the SL sequence are subject to various modifications and constitute, with the 7-methyl-GTP nucleotide at its 5′ end, the structure called cap4 [6,7]. Roles for the capped SL sequence in mRNA metabolism have been proposed, based largely on cap functions in metazoans, in mRNA transport, stability and/or translation, but no firm conclusions have been reached yet. Experiments using antisense oligonucleotides complementary to the SL sequence in heterologous cell-free translation systems are suggestive of a role for this sequence in translation [8,9]. Indirect evidence from nematodes, using message-dependent homologous protein synthesis systems, also reinforces a role in translation [10]. Indeed recent data suggests that either the SL sequence or the cap4 is responsible for polysome association of mRNA in *L. tarentolae* [11]. Since the 5′ end of most eukaryotic mRNAs is required for translation initiation, it is expected that their influence would be exerted at this stage of the translation process.

In animals, yeast and plants**,** the highly regulated initiation stage of protein synthesis starts with the binding of eIF4F (formed by the translation initiation factors eIF4E, eIF4A and eIF4G) to the monomethylated cap present on the 5′ end of the mRNAs [12]. With the help of other initiation factors, such as eIF3, eIF4F allows the recognition of the mRNAs by the 40S ribosomal subunit and the start of the translation process. Its interaction with the poly(A)-binding protein (PABP), via eIF4G, may also promote the circularization of the mRNA and enhance translation reinitiation (reviewed in [13–15]). eIF4E is a 24 kDa polypeptide responsible for specific cap recognition. It binds to both the cap and eIF4G and its activity can be regulated by phosphorylation or through the eIF4E interacting proteins (reviewed in [16–18]). eIF4A is an ATP-dependent RNA helicase which has been classified as a member of the DEAD box family of RNA helicases (reviewed in [19–21]). During translation initiation eIF4A binds eIF4G and, with the help of another translation initiation factor eIF4B, is responsible for melting secondary structures along the mRNA 5′UTR which allows the recruitment and scanning of the small ribosomal subunit to the translation initiation codon. The large subunit eIF4G (approximately 200 kDa) provides the scaffold for the eIF4F complex (reviewed in [22]). Through its N-terminus, eIF4G binds both eIF4E [23] and PABP [24,25], providing the link between the 5′and 3′ ends of the mRNA. eIF4G's central core region includes the conserved HEAT repeat domain (also known as MIF4G/NIC1) [26], which is responsible for the interaction with eIF4A and RNA-binding, and also mediates the interaction with eIF3 [27,28], which recruits the 40S ribosomal subunit. Mammalian eIF4G contain a C-terminal domain with a second eIF4A binding site as well as the binding site for the eIF4E kinase Mnk. Yeast eIF4G lacks this C-terminal domain [29–31].

Knowledge about the process of protein synthesis in trypanosomatid protozoans is inferred by indirect evidence such as sequence similarities between individual translation factors with homologues from higher eukaryotes. Translation factors characterised so far from these organisms include subunits of elongation factors such as eEF1 and eEF2 [32] and the PABP from *T. cruzi*, *T. brucei* and *Leishmania major* [33–35]. The eIF4A component of the eIF4F complex has also been identified both in *L. major* [36] and in *L. braziliensis*, as an antigen that elicits IL-12 production and Th1-type response in humans [37]. However, little is known regarding the role of these factors in translation.

To understand translation initiation in the trypanosomatids, we set out to characterise the various eIF4F subunits from *L. major*. Multiple potential homologues for the three eIF4F components were identified. These homologues seem to vary in different aspects such as cap binding affinity of the eIF4Es, levels of expression and interaction with other components of eIF4F. Our results suggest a high degree of complexity in translation initiation in these parasites, which may reflect an adaptation to their complex life cycle.

2. Materials and methods

2.1. Parasites

Promastigotes of *L. major* (MHOM/IL/81/Friedlin) were maintained at 26 °C in modified LIT medium pH 7.2 (0.2% sucrose (w/v); 0.36% liver broth (w/v); 0.1% tryptose (w/v); 0.002% haemin (w/v)) containing 20% foetal bovine serum, ampicillin (10 U/ml) and streptomycin (10 μg/ml). Parasites were kept on continuous log phase growth by changing the culture media every 3–4 days. Total protein lysates were obtained from log phase haemocytometer quantified parasite cell pellets resuspended directly in SDS-PAGE sample buffer.

2.2. Sequence analysis and modelling

BLAST searches were carried out for possible *L. major* eIF4F subunit homologues at the GenBank databases (http://www.ncbi.nlm.nih.gov) using as query the sequences of the human and *Saccharomyces cerevisae* proteins. Similar searches were also initially performed with the non-annotated draft *L. major* and *T. brucei* genome databases (http://www.sanger.org), and later with those available at the Gene DB website of the Sanger Institute Pathogen Sequencing Unit (http://www.genedb.org). Sequences were aligned with Clustal W (http://www.cmbi.kun.nl/bioinf/tools/clustalw.shtml); occasionally manual refinement of the alignments was performed. For the eIF4A sequence analysis, the alignment produced with CLUSTAL W was followed by phylogenetic and molecular evolutionary analyses conducted with the program MEGA version 2.1 [38] using the Neighbour-Joining method [39]. Gaps were treated as miss-

ing data. Bootstrap analysis (10,000 replicates) was done by interior branch test.

For the molecular modelling of *Lm*EIF4E1 and the *Lm*EIF4G3 HEAT domain, structural alignments were performed with the GenTHREADER program [40] comparing their secondary structure with that of available proteins with resolved structure. Best matches were the mouse eIF4E bound to 7-methyl-GDP [41] and the human eIF4GII HEAT domain [26], respectively. The atomic coordinates from these structures were submitted with the alignment results to the program MODELLER [42] in order to produce the models, which were then validated using the programs PROCHECK [43], Verify 3D [44] and WHATIF [45].

2.3. PCR and cloning methods for the various L. major factors

The *LmEIF4E1*, *LmEIF4E2*, *LmEIF4E2b* (same as *LmEIF4E2* but missing the 5′ end coding for the first seven amino acids of the protein), *LmEIF4E3* and *LmEIF4A1* coding sequences were amplified using primers flanked by sites for the restriction enzymes *Kpn*I/*Xba*I (*LmEIF4E2*) or *Bam*HI/*Xba*I (all others) and cloned into the corresponding restriction sites of the pGEM3zf+ vector obtained from Promega (the sequences of all the oligonucleotides described in this work are listed on the supplementary table). The resulting full length sequences were then reamplified flanked by sites for *Afl*III/*Not*I (*LmEIF4E1*), *Nco*I/*Not*I (*LmEIF4E2* and *LmEIF4E3*—in the latter a N to D mutation was introduced in the second codon to accommodate the *Nco*I site) or *Nco*I/*Xho*I (*LmEIF4A1*) and cloned into the *Nco*I/*Not*I or *Nco*I/*Xho*I sites of the expression vector pET21D (Novagen) for the expression of recombinant C-terminal His-tagged proteins. For the expression of N-terminal Glutathione *S*-transferase (GST) fusions, fragments *LmEIF4E1* and *LmEIF4E2*b were recovered from the pGEM derived plasmids and subcloned into the *Bam*HI/*Sal*I sites of the vector pGEX4T3 (Amersham Biosciences). For the same reason *LmEIF4E3* and *LmEIF4A1* were reamplified flanked by sites for *Bam*HI/*Not*I and *Bam*HI/*Xho*I, respectively and cloned into the same sites of the pGEX4T3 vector. *LmEIF4A1* was further subcloned into the same sites of the pRSETA plasmid (Invitrogen) for the expression of N-terminal His-tagged proteins. The *LmEIF4A2* gene was amplified flanked by sites for *Bam*HI/*Xho*I and cloned directly into the same sites of the vectors pRSETA and pGEX4T3.

Fragments coding for the HEAT repeats domain from the *Lm*EIF4G1-3 homologues, $LmEIF4G1_{128–475}$ (the numbers indicate the amino acids remaining from the wild type protein), $LmEIF4G2_{387–705}$ (missing the last 22 amino acids from the HEAT domain in Fig. 7) and $LmEIF4G3_{26–310}$, flanked by sites for *Afl*III/*Not*I ($LmEIF4G1_{128–475}$) or *Nco*I/*Not*I (the others), were amplified and cloned into the *Nco*I/*Not*I sites of the pET21D plasmid for the expression of C-terminal His-tagged proteins. For both $LmEIF4G2_{387–705}$ and $LmEIF4G3_{26–310}$, an AUG start codon was introduced in the constructs with the *Nco*I site. The full length *LmEIF4G3* ($LmEIF4G3_{1–636}$), as well as its first half ($LmEIF4G3_{1–343}$) were also amplified and cloned into the same sites of pET21D. The latter two fragments were then reamplified and cloned into the *Bam*HI/*Not*I sites of pGEX4T3 to have both proteins as GST fusions.

All of the amplified fragments above were sequenced and the resulting sequences confirmed by comparison with those from the *L. major* genome sequencing project.

2.4. Cap binding assay

^{35}S-Met-labelled *L. major* eIF4E proteins were synthesised in the TnT T7 Coupled Reticulocyte Lysate System (Promega) using the pET derived plasmids as templates. As a positive control, a *Xenopus laevis* eIF4E cDNA, in pSP64TEN, was recovered by digestion with *Hin*DIII/*Bgl*II and subcloned into the *Hin*DIII/*Bam*HI sites of pGEM2 (Promega) under control of the T7 promoter. For the cap binding assay, 35 μl of 7-methyl-GTP Sepharose 4B beads (Amersham Biosciences), previously equilibrated with buffer A (50 mM Hepes; 1 mM EDTA; 0.1 mM GTP; 14 mM 2-mercaptoethanol; 100 mM KCl; pH 7.4), was mixed with 25 μl of the translation products (in 200 μl final volume—completed with buffer A) and incubated for 30 min in ice. The beads were then washed three times with 500 μl buffer A and, for the elution, three times with 50 μl buffer A containing 50 μM cap analogue ($m^7G(5')ppp(5')G$ (New England Biolabs) instead of the GTP. Any labelled protein still remaining on the beads was recovered with two final washes with 2 M KCl and SDS-PAGE sample buffer, respectively. Aliquots of all fractions were mixed with SDS sample buffer and analysed by SDS-PAGE and autoradiography.

2.5. Expression and purification of recombinant proteins

For the expression of either His or GST-tagged recombinant proteins, plasmids were transformed into *Escherichia coli* BLR or BL21 cells. The transformed bacteria were grown in LB medium and induced with IPTG. Induced cells were sedimented, resuspended in PBS and lysed by sonication or French press. Protein purification was performed as described [46] with either Ni-NTA Agarose (Qiagen) or glutathione-Sepharose (Amersham Biosciences). Protein products were analysed in 15% SDS-PAGE stained with Coomassie Blue R-250. For the quantification of the recombinant proteins, serial dilutions were compared in Coomassie stained gels with serial dilutions of known concentrations of BSA.

2.6. Antibody production and Western blotting

Rabbit antisera were raised against *Lm*EIF4A1-2, *Lm*EIF4E1-2 and $LmEIF4G3_{26–310}$ by immunising adult New Zealand White rabbits with the His-tagged recombi-

nant forms. For *Lm*EIF4E3, the GST-fusion was first cleaved with thrombin and the fragment corresponding to the protein excised from SDS-PAGE gels and used for the immunization. The various antibodies were affinity purified using the protocol described in [47], with minor modifications, stored at 4 °C with 0.03% sodium azide and used in a dilution of 1/500. Western blots were performed with the Immobilon-P PVDF membrane (Millipore), using as second antibody peroxidase conjugated goat anti-rabbit IgG serum (Jackson ImmunoResearch Laboratories) diluted 1/15,000. The reactions were detected by enhanced chemiluminescence (ECL). Densitometric scanning of the various results was performed with the Kodak 1D Image Analysis Software, version 3.5 for Windows.

2.7. Pull-down assays

The Ni-NTA Agarose or glutathione-Sepharose beads used for the pull-down assays were initially equilibrated with binding buffer B (100 mM KCl, 1 mM $MgCl_2$, 50 mM Hepes pH 7.2, 0.2% NP-40, 5% glycerol). Buffer B was supplemented with 5 mM of imidazole during all the steps of the His-tagged pull-downs. For the GST-tagged protein experiments, the glutathione-Sepharose beads were initially saturated with 10 mg/ml of BSA, followed by two washes with Buffer B prior to their use in the assays. Approximately 10 μl of the equilibrated beads was then incubated with 2 μg of the His or GST tagged constructs in a final volume of 200 μl (completed with Buffer B), for 1 h rotating at 4 °C. The beads were washed twice with Buffer B and incubated with 10 μl of translation lysate containing the ^{35}S-Met-labelled proteins, in Buffer B in a final volume of 200 μl, for 2 h at 4 °C. After three more washes, the bound proteins were eluted by addition of SDS sample buffer and the samples ran on 15% SDS-PAGE. The gels were stained with Coomassie Blue (to visualize the recombinant proteins), and subjected to autoradiography. Labelled proteins were obtained through the linearizations of the various plasmids described previously (with *Not*I for pET-*LmEIF4G1*$_{128–475}$ and the *LmEIF4G3* variants and with *Xho*I for pET-*LmEIF4G2*$_{387–705}$ and pRSET-*LmEIF4A1-2*), followed by transcription with T7 RNA polymerase in the presence of the cap analogue and translation in the rabbit reticulocyte lysate supplemented with ^{35}S-methionine. To obtain the labelled human eIF4A, its cDNA in plasmid pET(His$_6$-eIF4A) [48] was first recovered by digestion with *Nde*I/*Not*I and subcloned into pET21A. The resulting plasmid was linearized with *Not*I and the eIF4A cDNA transcribed and translated as above. Labelled human eIF4G was obtained as described [49].

3. Results

3.1. Sequence analysis of four eIF4E homologues from L. major

To characterise *L. major* eIF4F homologues, we initially focused on four available eIF4E sequences which we named as *Lm*EIF4E1 (GeneDB ID—LmjF27.1620), *Lm*EIF4E2 (LmjF19.1500/LmjF19.1480), *Lm*EIF4E3 (LmjF28.2500) and *Lm*EIF4E4 (LmjF30.0450) (see Table 1). Naming of the various proteins described in this word followed the proposed nomenclature for *Leishmania* and *Trypanosoma* proteins [50]. The potential eIF4E sequences code for proteins ranging in size from 214 to 281, 349 and 447 amino acids, respectively, for *Lm*EIF4E1-4, with homologies varying from 41 to 45% similarity to the human eIF4E homologue. Single genes present on chromosomes 27, 28 and 30 encode

Table 1
Summary of the sequence analysis of the various *L. major* eIF4F homologues, depicting identities (similarities) between the *Leishmania*/human sequences and between the *L. major*/*T. brucei* orthologues

L major sequence	GeneDB ID	Predicted molecular weight (kDa)	Chromosome localization	Identity (similarity) to Human homologue[a]	Identity (similarity) to the *T. brucei* orthologue
*Lm*EIF4E1[b]	LmjF27.1620	24	27	22% (42)	48% (61)
*Lm*EIF4E2[b]	LmjF19.1500/LmjF19.1480	31.5	19 (2 genes)	27% (41)	45% (57)
*Lm*EIF4E3	LmjF28.2500	38	28	27% (43)	49% (61)
*Lm*EIF4E4	LmjF30.0450[c]	33.8	30	28% (45)	37% (50)
*Lm*EIF4A1[b]	LmjF01.0780/LmjF01.0770	45.3	1 (2 genes)	56% (74)	85% (91)
*Lm*EIF4A2	LmjF28.1530	43.9	28	50% (71)	79% (89)
*Lm*DHH1	LmjF35.0370	46.4	35	30% (50)	86% (92)
*Lm*EIF4G1	LmjF15.0060	114	15	25% (43)[d]	31% (47)
*Lm*EIF4G2	LmjF15.1320	145.9	15	21% (37)[d]	33% (49)
*Lm*EIF4G3	LmjF16.1600	71.2	16	26% (39)[d]	38% (55)
*Lm*EIF4G4	LmjF36.6060	84.6	36	22% (38)[d]	26% (45)
*Lm*EIF4G5	LmjF10.1080	88.8	10	21% (37)[d]	43% (60)

Results obtained using the GenBank BLAST with the BLOSUM 62 Matrix.

[a] Human eIF4GI. GenBank accession: Q04637.

[b] GenBank accessions available: *Lm*EIF4E1—CAB77676; *Lm*EIF4E2—CAB94109 and CAB94111; *Lm*EIF4A1—NP_047099.

[c] Here the sequence used in our analysis includes the N-terminal extension, present in the *T. brucei* orthologue, which has been omitted from the *L. major* GeneDB annotation.

[d] These sequences show similarity to the human sequence only at the level of the central HEAT domain.

*Lm*EIF4E1, 3 and 4, respectively. In contrast, *Lm*EIF4E2 is encoded by two identical genes contained within a small duplicated region on chromosome 19, which also includes an unidentified trans-membrane protein ORF, separated by approximately 4.5 kb.

In order to compare conserved features in the four *Leishmania* eIF4E homologues we aligned these sequences with characterised human, yeast and plant eIF4E proteins (Fig. 1). Two distinct *Caenorhabditis elegans* homologues with contrasting cap binding affinities – IFE-1 which binds both monomethylated and trimethylated caps and IFE-3 which only binds a monomethylated cap [51] – were also included in the alignment. For simplicity, all the numbers mentioned in Fig. 1 will refer to the amino acids' position in the human eIF4E sequence which is representative of the mammalian protein (only three substitutions in the N-terminus distinguish the human and mouse sequences). eIF4E is characterised by eight tryptophan residues located at conserved positions along the protein [41,12]. *Lm*EIF4E2 contains all eight of these tryptophan residues, *Lm*EIF4E1 contains seven, *Lm*EIF4E4 five and *Lm*EIF4E3 only four. Three-dimensional structures of mouse eIF4E and its yeast homologue, both bound to 7-methyl-GDP, were solved by X-ray crystallography [41], and by solution NMR spectroscopy [52], respectively. The complex is shaped like a cupped hand, with the cap analogue located in a narrow cap-binding slot on the concave side of the protein. Recognition of the 7-methylguanine moiety is mediated by base sandwich-stacking between W56 and W102, formation of three Watson-Crick-like hydrogen bonds with a side-chain carboxylate of a conserved E103 and a backbone NH of W102, and a van der Waals contact of the N(7)-methyl group with W166. The four residues making contacts with 7-methylguanine are conserved among most known eIF4E proteins (see Ref. [41] and Fig. 1). All four are present in both *Lm*EIF4E1 and *Lm*EIF4E2, while in *Lm*EIF4E3-4 W56 is replaced by a methionine and a glutamate, respectively. Of the three residues that interact with the two phosphate groups, R157 is absolutely conserved in eIF4E sequences, and the remaining two residues (R112 and K162) make either direct or water mediated contacts with phosphates and are either arginine or lysine [41]. In the *Leishmania* proteins, *Lm*EIF4E1 and 2 contain K112, R157 and R/K162, respectively, while both *Lm*EIF4E3-4 contain only R157, and lack basic residues at 112 and 162. On this basis, *Lm*EIF4E1 and 2 possess all the features expected of a 7-methyl-GDP-binding protein, while *Lm*EIF4E3-4 have some, but not all, the predicted features.

eIF4G binds the convex dorsal side of eIF4E primarily due to the interaction between W73 of eIF4E and three amino acid side chains of the eIF4E binding peptide, YXXXXLphi (where X can be any amino acid and phi is usually a hydrophobic amino acid such as L, M or F [23]), also present in eIF4E-binding proteins [53,54]. Mutation of W73 to A or R, respectively, in the mouse and yeast eIF4Es abolishes binding to eIF4G, while the yeast F73 eIF4E mutant maintains binding to eIF4G [55,31]. *Lm*EIF4E2, 3 and 4 contain W73, while *Lm*EIF4E1 contains F73 (Fig. 1); thus on this basis all four proteins would be predicted to be able to interact with eIF4G.

The most striking differences between the *Leishmania* proteins and known eIF4Es include the *Lm*EIF4E3-4 N-terminal extensions of about 80 and 190 amino acids, respectively, a shorter C-terminus for *Lm*EIF4E1 and the significant insertions of variable number of amino acids at specific positions in the first three *Leishmania* proteins. *Lm*EIF4E1 is distinguished by the presence of a 21 amino acids long insertion not observed in other eIF4E sequences. *Lm*EIF4E2 has two short insertions, one of them also present in *Lm*EIF4E3, which are located next to amino acids relevant to cap binding, and a third long insertion on its C-terminus. Comparison with the two nematode eIF4E sequences included in Fig. 1 does not indicate any motifs that may be involved in differential cap recognition and indeed it seems that the multiple eIF4E homologues evolved independently in both systems. Nevertheless, a comparison between the *L. major* and *T. brucei* eIF4E sequences (Table 1; see Section 4) confirm that all the specific features mentioned above for the various *L. major* proteins, with minor variations, are also conserved in the four *T. brucei* homologues. Such conservation indicates that the multiple eIF4E proteins are conserved within the family *Trypanosomatidae* and play significant roles during the parasites' life cycle.

3.2. Expression analysis and quantitation of *LmEIF4E1-3 in L. major promastigotes*

*Lm*EIF4E1-3 (*Lm*EIF4E4 was only identified later in this study) were expressed in *E. coli* His-tagged or as GST-fusions and isoform-specific antibodies generated in rabbit against the three proteins. These antibodies were affinity purified and tested in Western blots against the recombinant GST fusions to confirm their specificity. No cross-reaction was observed to the other recombinant *L. major* eIF4Es (data not shown). They were then used to analyse the expression and to estimate the cellular levels of *Lm*EIF4E1-3 in extracts of exponentially growing *L. major* promastigotes (Fig. 2). All three proteins were detected in these extracts although their levels varied significantly, with *Lm*EIF4E3 being very abundant ($>5 \times 10^4$ molecules/cell) and both *Lm*EIF4E1 and 2 present at much lower levels (about $2–4 \times 10^3$ and 10^3 molecules/cell, respectively—Table 2). When compared to yeast eIF4E ($3.2–3.6 \times 10^5$ molecules/cell [56]), even *Lm*eIF4E3 is expressed at levels lower than those obtained for the yeast protein. However, considering that in yeast, a significant fraction of the eIF4E pool is not complexed to eIF4G (present at only $(1.5–2) \times 10^4$ molecules/cell) these levels could account for the production of enough eIF4F for the survival of the organism. Nevertheless it seems unlikely that the levels of either *Lm*EIF4E1 or *Lm*EIF4E2 alone would be sufficient to support translation in these parasites, at least in the promastigote stage.

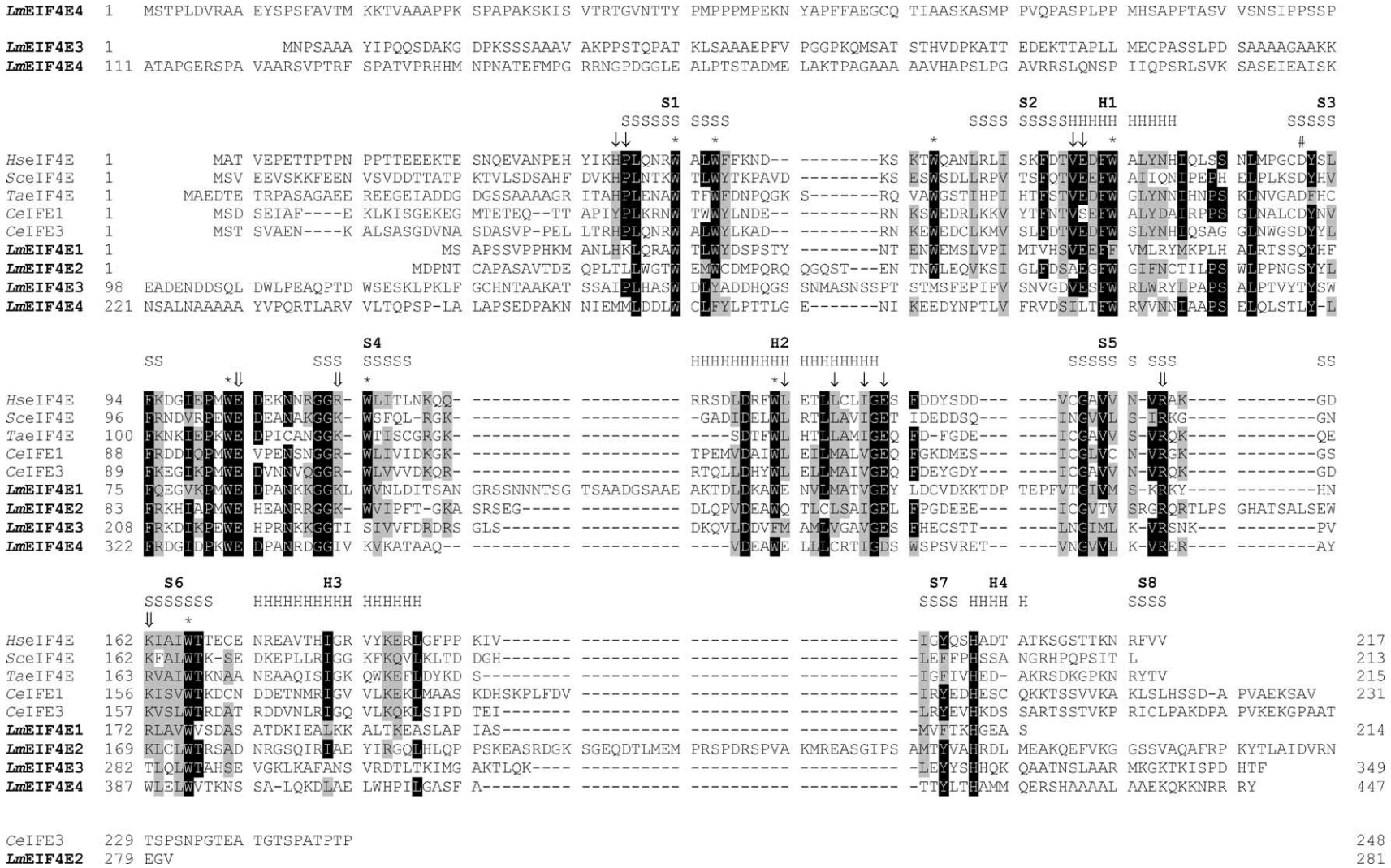


Fig. 1. Sequence comparison of the putative *L. major* eIF4E homologues with the human, yeast, plant and nematode sequences. Clustal W alignment of different eIF4E homologues from selected organisms. Amino acids identical in more than 60% of the sequences are highlighted in dark gray, while amino acids defined as similar, based on the BLOSUM 62 Matrix, on more than 60% of the sequences, are shown in pale gray. When necessary, spaces were inserted within the various sequences (dashes) to allow better alignment. The structural elements of the mammalian protein are shown numbered S1–S8 and H1–H4 (from [41]). * indicates the conserved tryptophan residues. Double arrows highlight amino acids required for the interaction with the cap structure. Single arrows indicate conserved non-tryptophan residues shown to be involved in eIF4G binding [53]. # indicates the amino acid D90, replaced by a Q in *Lm*EIF4E1 (Q71) and shown in Fig. 4. Relevant GenBank accession numbers: human (*Hs*)—P06730; yeast (*S. cerevisae*, *Sc*)—P07260; plant (*Triticum aestivum*, *Ta*) P29557; nematode (*C. elegans*, *Ce*) IF4E1 and IF4E3—NP_499751 and NP_503124, respectively.

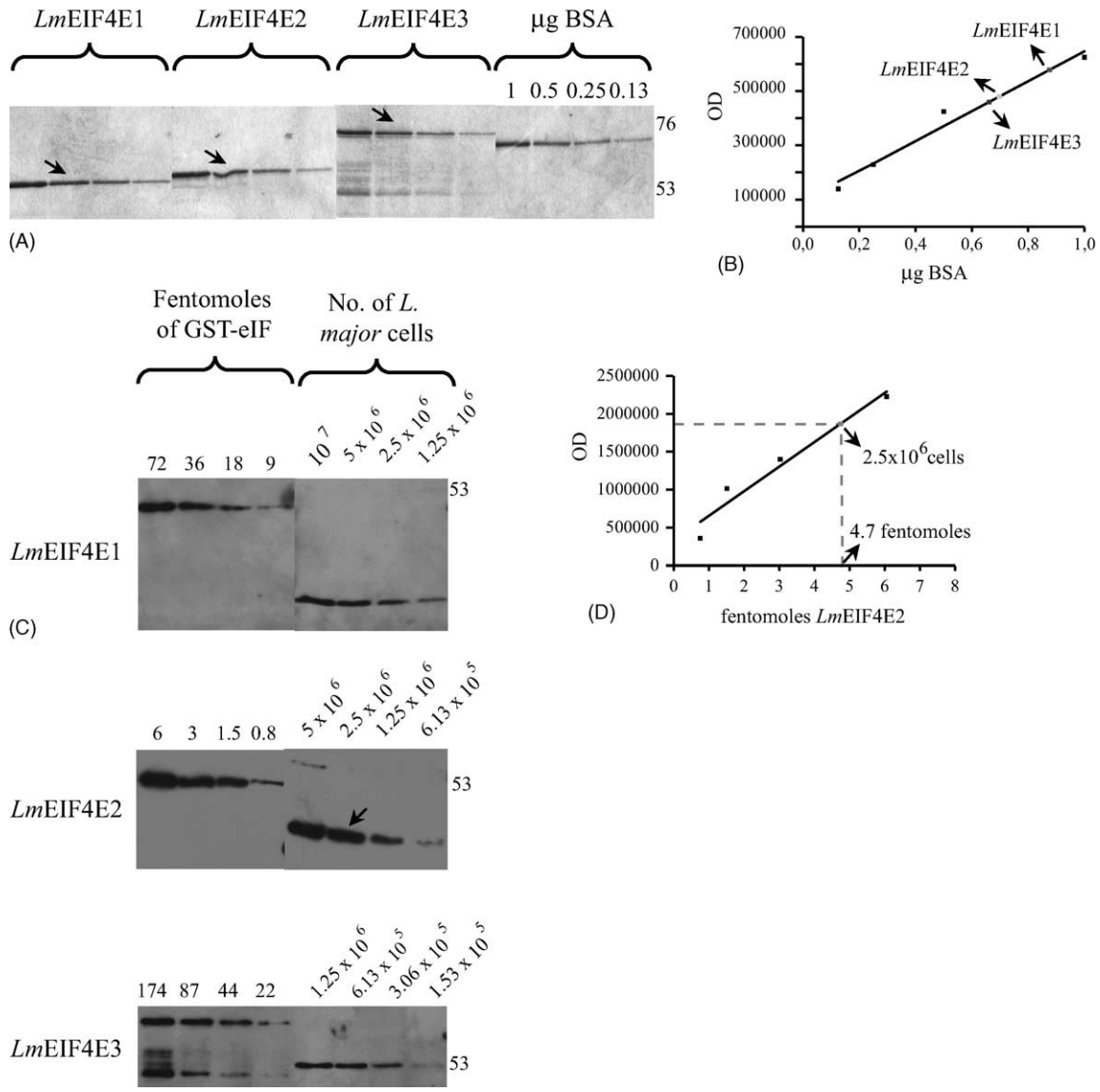


Fig. 2. Expression analysis and quantitation of *Lm*EIF4E1-3 in exponentially grown *L. major* promastigotes. (A) Serial dilutions of recombinant GST-tagged *Lm*EIF4E1-3 compared in Coomassie Blue stained gels with known concentrations of BSA. (B) Quantitation of the recombinant proteins (a representative example is shown). The bands from the BSA curve shown in A were quantified by densitometric scanning and plotted as a function of the BSA concentration. To calculate the concentration of recombinant *Lm*EIF4E1-3, the optical density of representative bands from these proteins (indicated by arrows in the figure) was determined to derive approximate quantities using the BSA curve. (C) Expression analysis of *Lm*EIF4E1-3 in extracts of *L. major* promastigotes. The different recombinant GST-fusions, diluted to appropriate concentrations (in fentomoles), and whole parasite extracts were fractionated by 15% SDS-PAGE and analysed by western blotting with the isoform specific sera. (D) Procedure for the quantitation of the endogenous levels of the various factors using *Lm*EIF4E2 as an example. The Western blot result shown in C for GST-*Lm*EIF4E2 was quantitated by densitometry and plotted as a function of fentomoles of recombinant protein. The optical density obtained for the endogenous factor at 2.5×10^6 cells (arrow) was then used to calculate its concentration in fentomoles (4.7 fentomoles in the example). The same procedure was used for the different factors including *Lm*EIF4A1 and *Lm*EIF4G3. All the results presented are representative of at least six different experiments using a minimum of three independently grown cell cultures. The data obtained from the various experiments were processed as described above and used to calculate the values shown in Table 2. On the right of the various panels in A and C are shown the sizes in kDa of protein molecular weight markers.

3.3. Cap binding assay of the putative LmEIF4Es

*Lm*EIF4E1-3 were then labelled with ^{35}S-Met by in vitro translation and tested for their ability to recognise the mammalian monomethylated cap. As a positive control, *Xenopus* eIF4E was also labelled and used in the same assay. All four proteins were tested in binding assays with the resin 7-methyl-GTP Sepharose 4B, followed by washes with GTP and elution with the cap analogue. The *Xenopus* eIF4E bound to the resin as expected, and was eluted with the cap analogue. In contrast, of the three *L. major* proteins, only *Lm*EIF4E1 was capable of binding to the resin (Fig. 3). Binding was specific since GTP did not release the protein and in parallel experiments performed with CL6B Sepharose no bind-

Table 2
Summary of the quantitation of the various *L. major* eIF4F homologues in promastigotes

L major sequence	Fentomoles/10^6 cell	No. of molecules/cell	No. of molecules/cell in yeast[a]
*Lm*EIF4E1	5.2 ± 2.5	$3.2 \pm 1.5 \times 10^3$	$(3.2–3.6) \times 10^5$
*Lm*EIF4E2	1.8 ± 0.5	$1 \pm 0.3 \times 10^3$	
*Lm*EIF4E3	117 ± 56	$7.1 \pm 3.4 \times 10^4$	
*Lm*EIF4A1	594 ± 218	$3.6 \pm 1.3 \times 10^5$	$(7–9) \times 10^5$
*Lm*EIF4A2	ND	ND	
*Lm*EIF4G3	10.6 ± 4.1	$6.4 \pm 2.5 \times 10^3$	$(1.5–2) \times 10^4$

[a] Numbers based on Ref. [56] for yeast eIF4E, eIF4A and eIF4G. ND: not determined.

ing was observed (not shown). *Lm*EIF4E2 and 3 lacked any affinity with the resin, both labelled proteins eluted in the flow-through. These results confirm that at least one of the *Leishmania* putative eIF4E homologues (*Lm*EIF4E1) has all the hallmarks of a typical eIF4E, including size, sequence and cap-binding ability.

3.4. Modelling of LmEIF4E1

In order to better define the structure/function conservation of the three *L. major* eIF4E homologues (*Lm*EIF4E1-3), we attempted to model their structure based on the mammalian eIF4E structure [41]. However, adequate modelling of both *Lm*EIF4E2 and *Lm*EIF4E3 was prevented due to the insertions between the first two β-strands (S1 and S2 in Fig. 1) which are essential to define the cap-binding slot. In contrast, the 21 amino acids long insertion in *Lm*EIF4E1, which is located in a loop between a β-strand and an α-helix (S4 and H2), did not prevent its modelling since it does not cause major interferences with the β-sheet or the general structure. In addition to the MODELLER pseudo-energy term, the quality of the models generated for *Lm*EIF4E1 was also independently evaluated by the programs PROCHECK, Verify 3D and WHATIF [43-45]. The representative model chosen was found to have self-consistency in terms of sequence-structure compatibility and to be of good overall quality (Fig. 4).

The final *Lm*EIF4E1 model confirms the predictions derived from the alignment in that it is possible to observe that the interactions mediated by the amino acids W56, D90, W102, E103, R157, K162, R112 and W166 in mammalian eIF4E, which are required for specific binding to the 7-methyl-GDP, are taken over by the amino acids W37, Q71, W83, E84, K93, R167, R172 and W176 (Fig. 4B). The potential to bind eIF4G was also investigated through the analysis of the convex side of the molecule. The amino acid F54, replacing the crucial W73 in mammalian eIF4E, is placed in the expected position for the eIF4E/eIF4G interaction (Fig. 4C). Also shown in Fig. 4C are the amino acids H20 and V50 equivalent to H37 and V69 in the mammalian protein, as well as amino acids E134 and M138 substituting for L128 and L135 (Fig. 1). In mammalian eIF4E, these amino acids have been shown to be involved in the eIF4E/eIF4G interaction [53]. Their positioning in the *Lm*EIF4E1 model confirms that, despite the minor variations in sequence, not only *Lm*EIF4E1 but the other *L. major* eIF4E homologues too may bind eIF4G. Finally, an analysis of the electrostatic potential of both the cap binding slot as well as the eIF4G binding side in the *Lm*EIF4E1 model, using the program GRASP [57], does not show any significant differences in overall charges

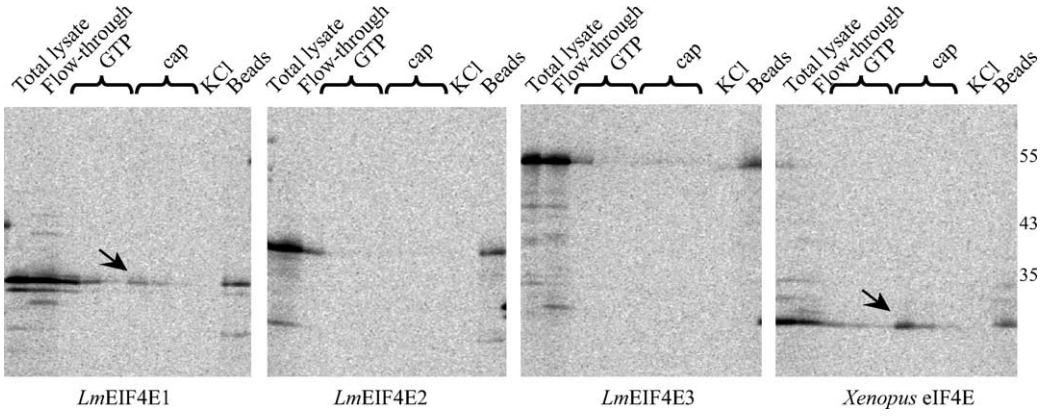


Fig. 3. *Lm*EIF4E1, but not *Lm*EIF4E2 or 3, binds specifically to the eukaryotic cap analogue. The three *L. major* eIF4E homologues (*Lm*EIF4E1-3), as well as the *X. laevis* eIF4E (*Xenopus* in the figure) used as positive control, were labelled with ^{35}S-methionine by in vitro translation and tested for their ability to bind the resin 7-methyl-GTP Sepharose. Non-specific binding was removed by washes with GTP and specific elution was achieved with the cap analogue. Aliquots of the various washes were ran on SDS-PAGE and compared with samples from the original translation reaction (Total lysate) as well as the non-bound fraction (Flow-through) and any protein remaining bound to the beads after washes with 2 M KCl and SDS-PAGE sample buffer (Beads). Arrows indicate proteins eluted by cap analogue. On the right are indicated sizes in kDa of protein molecular weight markers.

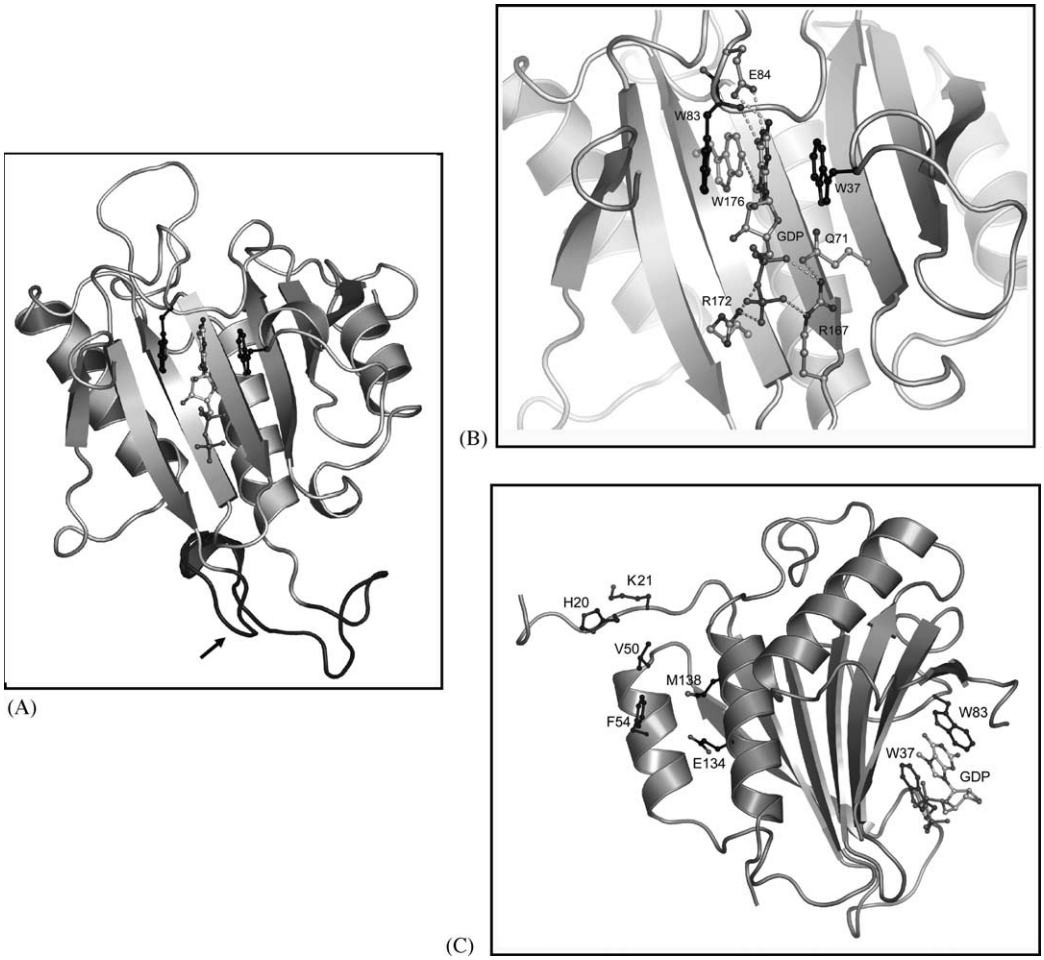


Fig. 4. Model of the predicted *Lm*EIF4E1 structure bound to 7-methyl GDP. Ribbon diagrams of the overall predicted *Lm*EIF4E1 structure and detail of the cap binding slot created using the program PyMol (http://www.pymol.org). (A) Model of the structure of *Lm*EIF4E1 bound to 7-methyl GDP, based on the mouse eIF4E structure [41]. W37 and W83 (equivalent to W56 and W102 in mammalian eIF4E) are shown binding the cap. The arrow indicates the loop specific to *Lm*EIF4E1 which could not be modelled adequately. (B) The cap binding slot in the predicted structure, highlighting the interactions between amino acids W37, Q71, W83, E84, R167, R172 and W176 and the cap nucleotide. The interactions mediated through water molecules are not shown so that the K93 interaction (corresponding to R112 in mammals) is omitted. (C) Side view of the *Lm*EIF4E1 model showing details of the predicted eIF4G binding surface. Amino acids H20, K21, V50, F54, E134 and M138 equivalent to H37, P38, V69, W73, L128 and L135 in the mammalian protein are highlighted. The cap binding slot with the W37 and W83 amino acids is also shown.

when compared with the mouse homologue (not shown). Once again these results confirm that *Lm*EIF4E1 has all the features required for a functional eIF4E homologue.

3.5. Sequence analysis of the two L. major EIF4A proteins

The RNA helicase subunit of the eIF4F complex, eIF4A, is a highly conserved eukaryote protein. The similarity between the described *L. major* eIF4A (LeiF) [36] here called *Lm*EIF4A1, and the human eIF4A1 homologue is 74% (Table 1). *Lm*EIF4A1 (GeneDB IDs—LmjF01.0780/LmjF01.0770) is a 403 amino acids long protein encoded by two identical genes placed *in tandem* roughly 1.8 kb apart within chromosome 1 (Table 1). Blast searches with the human eIF4A as query using the raw sequence data from the *L. major* genome yielded both *Lm*EIF4A1 as well as a possible second eIF4A homologue from *L. major* (*Lm*EIF4A2—LmjF28.1530), with 71% similarity to the human protein. The *LmEIF4A2* gene is located on chromosome 28 and codes for a 389 amino acids long protein. Orthologues to both *Leishmania* eIF4A proteins were also identified in *T. brucei*, indicating that they are conserved within the *Trypanosomatids* (Table 1).

Fig. 5 shows a sequence alignment comparing both *L. major* eIF4A sequences with those of human, yeast and plant eIF4A homologues. We have also included in the alignment human eIF4AIII (a negative regulator of translation [58]), a related Eubacteria protein (*Ec*W2) which may also be involved in translation initiation [59], an Archae DEAD-box protein (*Mj*DEAD) whose structure has been solved [60], as well as two unrelated eukaryotic RNA helicases, *Lm*DHH1 and *Spisula* p47 (see below). eIF4A and other related DEAD-box RNA helicases are characterised by nine motifs (motifs I,

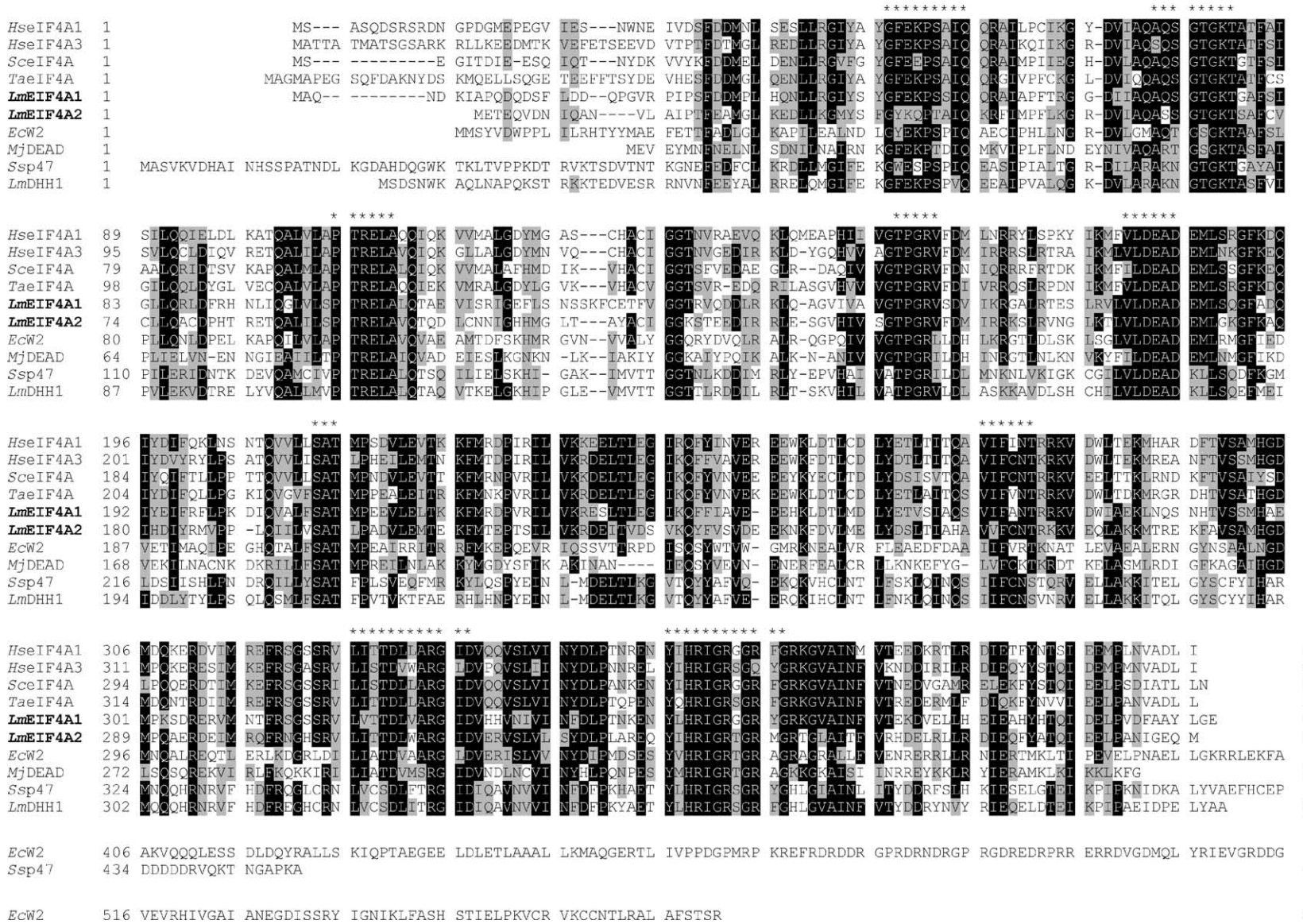


Fig. 5. Sequence comparison of the *L. major* eIF4A homologues with various related protein sequences. Clustal W alignment performed as described in Fig. 1 comparing the sequences of the *L. major* eIF4A homologues with selected sequences from several different organisms. * indicates the nine motifs typical of DEAD box RNA helicases [20,21]. Relevant GenBank accession numbers: human (*Hs*) eIF4A1—P04765; human eIF4AIII (*Hs*eIF4A3)—P38919; yeast (*S. cerevisae*, *Sc*)—NP_012397; plant (*T. aestivum*, *Ta*)—P41378; Eubacteria (*E. coli*, *Ec*) W2—AAA23674; Archae (*M. jannaschii*, *Mj*) DEAD box protein—NP_247653; Clam (*S. solidissima*, *Ss*) p47—AAK85400.

Ia, Ib, II, III, IV, V and VI plus the recently identified Q motif), conserved among the various members of this protein family, with roles in ATP binding, and hydrolysis, and RNA-binding [61,20]. The crystal structure of yeast eIF4A [62] indicates that it assumes a "dumbbell" shape with two globular domains connected by a flexible linker and it implicates several conserved arginine residues as important for eIF4A/helicase function. It is thought that these domains interact so that the binding and hydrolysis of ATP influences RNA binding, and comparison with the structure of the related protein from the Archae *Methanococcus jannaschii Mj*DEAD supports this model [60,21].

Overall, the alignment in Fig. 5 indicates that both *L. major* proteins share all the conserved motifs from RNA helicases, as well as several minor sequences typical of eIF4A. To support our identification of *Lm*EIF4A1 and 2 as possible eIF4A homologues, we searched the *L. major* sequence databases for any related RNA helicases. The sequence with the nearest match (*Lm*DHH1 in Fig. 5) contains all the motifs which characterise the DEAD-box RNA helicases yet it shares a similarity of only 50% with the human eIF4A1 sequence (Table 1). Remarkably this protein has a much higher degree of similarity (82%) to the clam *Spisula solidissima* p47 RNA helicase [63], a member of the Dhh1/p54 family of RNA helicases, with roles in decapping and translational repression [63,64], suggesting that it is a Dhh1 homologue. The sequences aligned in Fig. 5, plus a few others which we judged relevant, were used in a neighbour-joining analysis to compare their phylogenetic relationships. The resulting tree (Fig. 6A) confirms that both *L. major* eIF4A homologues are more closely related to each other and to other eIF4A proteins involved in translation than to unrelated RNA helicases including *Lm*DHH1 and bacterial proteins, and also highlights the close

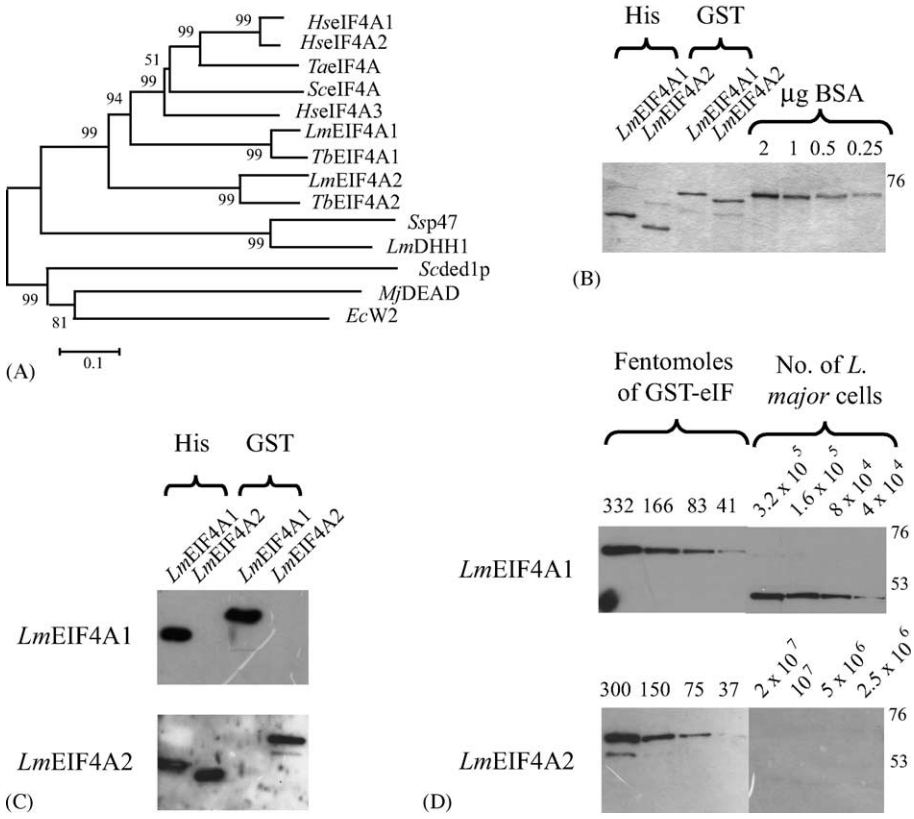


Fig. 6. *Lm*EIF4A1, but not *Lm*EIF4A2, is expressed as an abundant protein in *L. major* promastigotes. (A) Evolutionary relationship of the *L. major* eIF4A sequences. Neighbour-joining tree based on the alignment of various eIF4A sequences including *Lm*EIF4A1-2. Boostrap values are shown next to the respective branches (10,000 replicates). When compared with Fig. 5 we have included for this tree the sequences from: human eIF4A2 (GenBank accession AAH12547); *S. cerevisae* ded1p (accession—NP_014847); *T. brucei* eIF4A1 (GeneDB ID—Tb09.160.3270); *T. brucei* eIF4A2 (GeneDB ID—Tb11.12.0011). (B)–(D) Expression analysis of the two putative *L. major* eIF4As done as described in Fig. 2 using recombinant *Lm*EIF4A1-2 as well as isoform specific purified antibodies. (B) Recombinant proteins used for antibody production (His) and quantitation assays (GST). Approximately 0.5 (His-*Lm*EIF4A2 and both GST proteins) or 1 μg (His-*Lm*EIF4A1) of each recombinant protein was loaded on the gels and compared with known concentrations of BSA. (C) Analysis of the specificity of the antibody obtained against the recombinant proteins. Both GST and His-tagged fusions of *Lm*EIF4A1-2 were tested with the two antisera. For the *Lm*EIF4A1 assay, 12.5 ng of each recombinant protein were loaded on the gel whilst for *Lm*EIF4A2 6.25 ng of the His and 25 ng of the GST fusions were used. Note that the antibody produced against His-*Lm*EIF4A2 cross-reacts with His-*Lm*EIF4A1 since both recombinant proteins share conserved epitopes introduced by the plasmid vector. These epitopes are absent from the GST constructs and from the protein used to produce the anti-*Lm*EIF4A1 antibody. (D) Analysis of the expression of both *Lm*EIF4A1 and *Lm*EIF4A2 in total *L. major* extracts. Quantitation of *Lm*EIF4A1 in promastigotes was performed as described for the *Lm*EIF4E homologues in Fig. 2.

conservation between the *Leishmania* and *T. brucei* orthologues.

3.6. LmEIF4As quantitation and expression analysis in L. major

Both the *Lm*EIF4A1 and the *Lm*EIF4A2 coding sequences were amplified from *L. major* genomic DNA, expressed in *E. coli* and purified in His- or GST-tagged forms. Fig. 6B shows a sample of the representative recombinant proteins. His-tagged *Lm*EIF4A1 and 2 were then used to immunize rabbits to obtain specific antisera. The resulting antibodies were affinity purified and used to confirm their specificity as well as to determine their cellular levels in total *L. major* extracts as described for the *Lm*EIF4Es (Fig. 6C and D). *Lm*EIF4A1, readily detected in promastigotes, is a very abundant protein with about (3–5) $\times 10^5$ molecules per cell (Fig. 6D, Table 2). This concentration is not very different from that of yeast eIF4A [56], and the eIF4A/eIF4E ratio in yeast is similar to that of *Lm*eIF4A1/*Lm*EIF4E3. These values are also consistent with results in HeLa and reticulocyte lysate where eIF4A has been found to be a very abundant translation initiation factor present at levels approximately 10-fold higher than eIF4E [65,66]. In contrast we could not detect *Lm*EIF4A2, even in 2×10^7 promastigote cells (Fig. 6D). We estimate from the sensitivity of the *Lm*EIF4A2 antibody that levels above 3×10^3 molecules/cell would have been detected. However in this case, its abundance would be at least 60–150-fold lower than that of *Lm*EIF4A1. We conclude then that *Lm*EIF4A1 is the functional homologue of eIF4A in *L. major* promastigotes and that *Lm*EIF4A2 is either expressed in other stages of the parasite life cycle, or is a rare factor which is unlikely to have a prominent role in general translation.

3.7. Identification of candidate LmEIF4G homologues

In contrast to eIF4A, the eIF4G subunit of eIF4F is a much less conserved protein at the sequence level. The two functionally equivalent eIF4G homologues in mammals (eIF4GI and II) and in yeast (Tif4631 and Tif4632) are only ∼45% identical. However, the central eIF4A/RNA binding domain is conserved in the eIF4G sequences from divergent organisms (reviewed in [12]). The recently solved structure of this domain from human eIF4GII consists mostly of alpha-helices organised into HEAT repeats [26]. The eIF4G HEAT domain not only mediates the interaction with eIF4A and RNA [26], but in yeast was also implicated, together with flanking sequences, in the binding to eIF5 and eIF1 to bridge the interaction between eIF3 and eIF4G [67,68]. A second motif conserved in eIF4G sequences from distantly related organisms is the eIF4E binding peptide [23], whilst the PABP binding region does not appear to be conserved in sequence from yeast to mammals [69,25].

Using the human eIF4GI as a BLAST query, we found five *L. major* ORFs containing the conserved central eIF4G domain (*Lm*EIF4G1, GeneDB ID—LmjF15.0060; *Lm*EIF4G2—LmjF15.1320; *Lm*EIF4G3—LmjF16.1600; *Lm*EIF4G4—LmjF36.6060; *Lm*EIF4G5—LmjF10.1080). The five proteins vary significantly in size (ranging from 1016 to 1425, 635, 765 and 782 amino acids for *Lm*EIF4G1-5, respectively), overall charge (positive in *Lm*IEF4G1 and negative in the other four) and in the relative location of the conserved eIF4G domain (central in *Lm*EIF4G1, 2, 5 and N-terminal in *Lm*EIF4G3-4). Both the *Lm*EIF4G1 and *Lm*EIF4G2 genes are located on chromosome 15 while the *Lm*EIF4G3-5 genes lie on chromosomes 16, 36 and 10, respectively (see Table 1). No additional similarities with human, yeast or plant homologues can be seen within their sequences and surprisingly, no obvious eIF3, eIF4E or PABP binding domains can be identified based on homology analysis alone (not shown, see Section 4). Apart from the HEAT domain, the *Leishmania* sequences appear unrelated, with the exception of *Lm*EIF4G3 and 4. These two proteins share a short similar N-terminus and a conserved region (approximately 120 amino acids long) located about 180 amino acids downstream of the HEAT domain, indicating that they may be functionally related.

Fig. 7 shows an alignment comparing the sequences of the HEAT domain from the five putative *L. major* eIF4G proteins with the equivalent sequence from human eIF4GI, *S. cerevisae* Tif4631p, and wheat eIFiso4G as well as the translation regulator PAIP1. Overall, the domain from *Lm*EIF4G1 is the most similar to human eIF4GI whilst the ones from *Lm*EIF4G2 and 5 are the least similar (Table 1), but the differences are small. All five *L. major* domains contain most of the conserved amino acids which in mammalian eIF4G homologues are required for binding to eIF4A [29,30,26]. Residues which have been shown not to be required for eIF4A binding (such as F737, K765, E769, P770, F812 and R855—human eIF4GI numbering) are also well conserved in most *Leishmania* proteins and presumably have other conserved functions. It may be noteworthy that the doublet E769/P770 (strictly conserved in *Lm*EIF4G1, 3 and 4) has been implicated in 43S recruitment and formation of the 48S complex [26]. Nevertheless, it is not possible to identify which of these proteins are true eIF4G orthologues with major roles in translation initiation. Strikingly, orthologues to all five proteins, with significant conservation outside the HEAT domain, have been identified in both *T. brucei* (Table 1) and in *T. cruzi* (data not shown) confirming their importance for the survival of these organisms as a group.

3.8. Analysis of the interaction between L. major eIF4A/eIF4G through pull-down assays

In animals, yeast and plants, the specific binding between the eIF4A and eIF4G subunits of eIF4F needs to occur so that both proteins can function in translation. In order to functionally demonstrate a possible role in translation for the various *L. major* candidate eIF4A/eIF4G proteins, we investigated their interaction through pull-down assays. First, to validate the assay, we immobilised N-terminally His-tagged

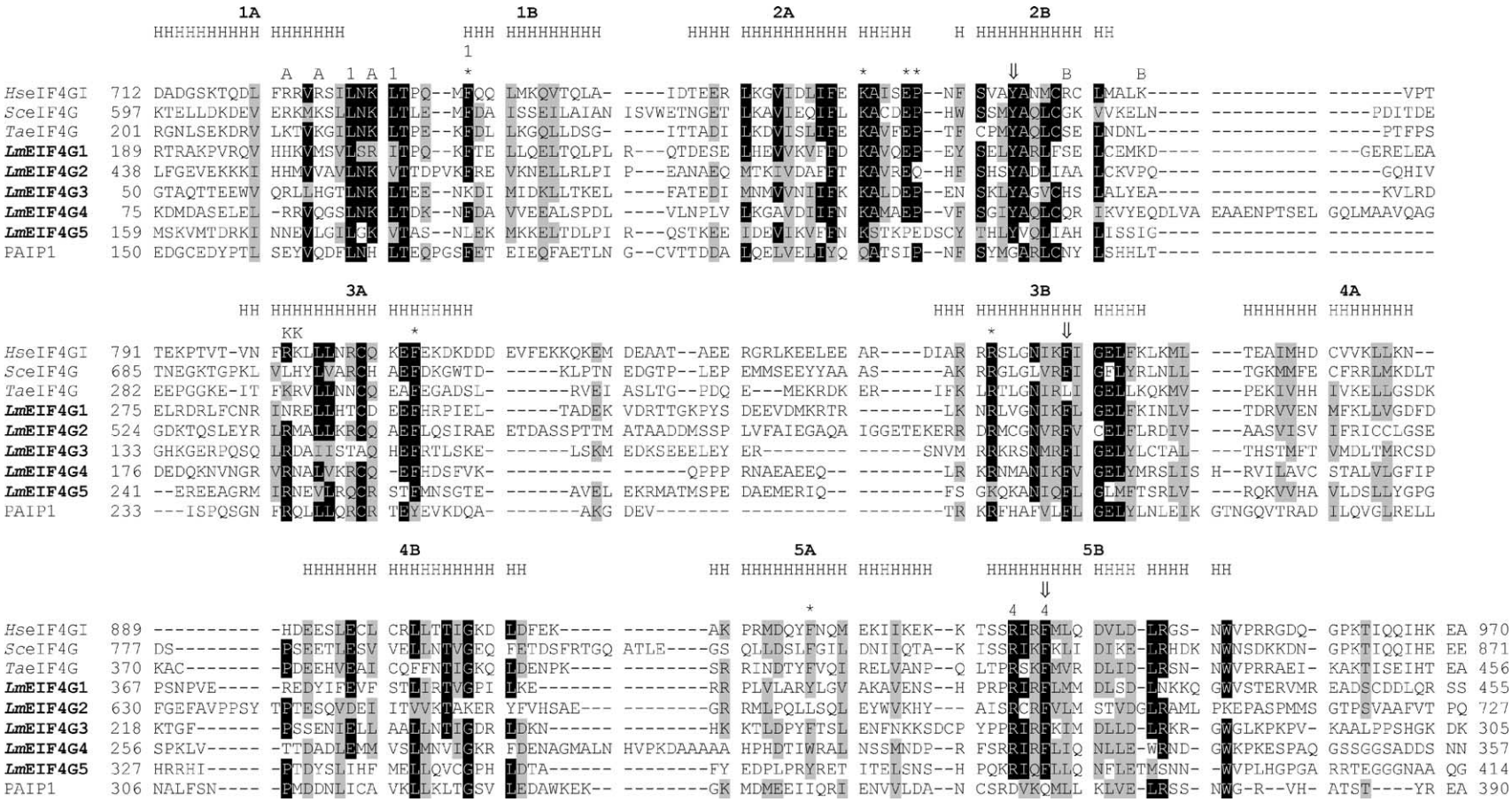


Fig. 7. Sequence comparison of the HEAT domain from the putative *L. major* eIF4G homologues with selected sequences from various organisms. Clustal W alignment performed as described in Fig. 1 comparing the sequences of the HEAT domain from the putative *L. major* eIF4G homologues (*Lm*EIF4G1-5) with the equivalent domain from human, yeast and plant eIF4G homologues as well as the human PAIP1 protein. The predicted five antiparallel alfa helical pairs (1–5, A and B) are indicated [26]. Selected mutations in mammalian eIF4G which have been shown to reduce the binding to eIF4A are shown as follows—A: R723D, R726D and K731D [26]; B: R781D and K787D [26]; 1 (M-1): L729A, L732A and F737A [29]; and 4 (M-4): R935A and F938A [29]. The R801D/K802D mutant (K in the figure) does not prevent eIF4A binding although it does abolish IRES binding [26]. Double arrows highlight amino acids which abolish the binding to eIF4A when individually mutated to alanine [30]. * indicates conserved amino acids which do not interfere with eIF4A binding when mutated to alanine [30,26]. Relevant GenBank accession numbers: human eIF4GI (*Hs*)—Q04637; yeast Tif4631p (*S. cerevisae*, *Sc*)—NP_011678; plant eIFiso4G (*T. aestivum*, *Ta*)—Q03387; human PAIP1—NP_006442.

recombinant *Lm*EIF4A1-2 and human eIF4A (His_6-eIF4A [48]) on Ni-NTA beads (Fig. 8A—left panel), and incubated the beads with ^{35}S-labelled human eIF4G [70]. Human eIF4A efficiently bound to human eIF4G, however neither *Lm*EIFA1 nor *Lm*EIFA2 was capable of binding to the human protein (Fig. 8A—right panel). Next, fragments coding for the HEAT domain from three of the *L. major* candidate eIF4G proteins (*Lm*EIF4G1$_{128–475}$, *Lm*EIF4G2$_{387–705}$, *Lm*EIF4G3$_{26–310}$), were synthesised in the presence of ^{35}S methionine and assayed for their ability to bind eIF4A. All three fragments include the region equivalent to amino acids 722–949 from human eIF4G1 sufficient to promote 48S

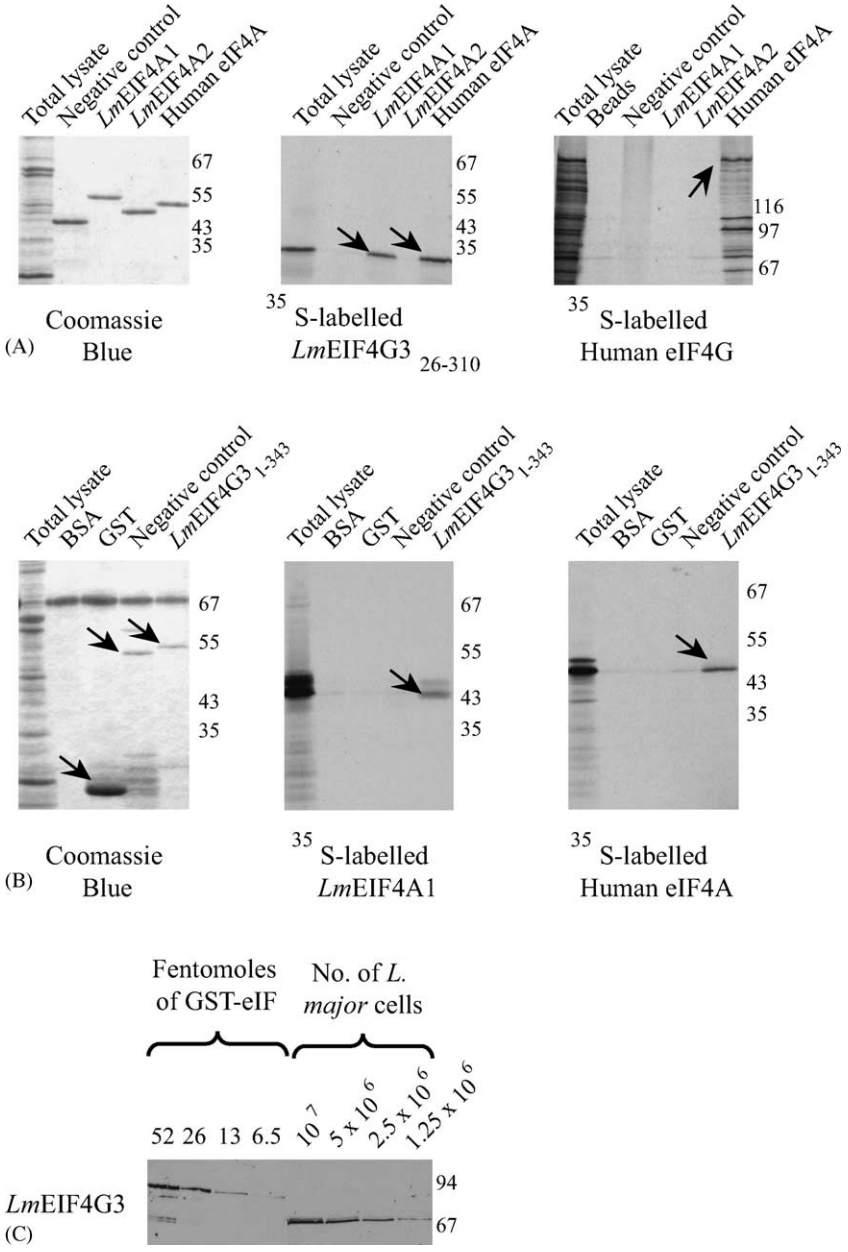


Fig. 8. Analysis of the interaction between human and *L. major* eIF4A homologues with the *Lm*EIF4G3 protein. (A) Pull-down assay using recombinant His-tagged *Lm*EIF4A1-2 and human eIF4A incubated with either ^{35}S-labelled human eIF4G or *Lm*EIF4G3$_{26–310}$. Proteins bound to the Ni-NTA beads were analysed through 15% (*Lm*EIF4G3$_{26–310}$) or 10% (human eIF4G) SDS-PAGE. Left panel: Coomassie Blue stained gel showing total translation extract (Total lysate) as well as the recombinant His-tagged proteins. Middle panel: autoradiography showing specific binding (arrows) between *Lm*EIF4A1-His/labelled *Lm*EIF4G3$_{26–310}$ and human eIF4A-His/labelled *Lm*EIF4G3$_{26–310}$. Right panel: positive control of a 10% SDS-PAGE showing the specific binding between human eIF4A and labelled human eIF4G. (B) Reverse pull-down assay using GST-tagged *Lm*EIF4G3$_{1–343}$ and ^{35}S-labelled *Lm*EIF4A1/human eIF4A. Left panel: 15% SDS-PAGE showing total translation extract and recombinant GST-tagged proteins. Middle and right panels: autoradiography showing specific binding (arrows) between *Lm*EIF4G3$_{1–343}$-GST/labelled *Lm*EIF4A1 and *Lm*EIF4G3$_{1–343}$-GST/labelled human eIF4A. (C) Quantitation of *Lm*EIF4G3 in *L. major* promastigotes. The quantitation of the endogenous protein levels was performed as described in Fig. 2 with recombinant GST-*Lm*EIF4G3$_{1–636}$ and isoform specific polyclonal sera against *Lm*EIF4G3.

complex formation on the EMCV IRES [28]. Only labelled *Lm*EIF4G3$_{26–310}$ was able to bind efficiently to *Lm*EIF4A1 and to human eIF4A as shown in Fig. 8A (middle panel). No specific binding by the *Lm*EIF4G1-2 constructs to the various recombinant proteins was observed (data not shown). Moreover, even though *Lm*EIF4A1 and 2 share a similarity of 71%, no significant binding by any of the three *Lm*EIF4G proteins to *Lm*EIF4A2 was observed (Fig. 8A and data not shown).

To confirm the specific interaction between *Lm*EIF4G3$_{26–310}$ with *Lm*EIF4A1, we expressed the N-terminal half of *Lm*EIF4G3 containing the HEAT domain (*Lm*EIF4G3$_{1–310}$) in *E. coli* fused to GST and tested it in a reverse pull-down assay. Here, the GST fusion was immobilised on glutathione-Sepharose and incubated with ^{35}S-labelled *Lm*EIF4A1 and human eIF4A (Fig. 8B). As negative controls, we used GST on its own and a murine GST-cdc2 fusion protein of a similar size to the GST-*Lm*EIF4G3$_{1–310}$ protein. This reverse assay confirmed the interaction between the GST-*Lm*EIF4G3$_{1–310}$ fusion protein and both labelled *Lm*EIF4A1 and human eIF4A (Fig. 8B—middle and right panels). In summary, the pull-down experiments indicate that of the three *L. major* putative eIF4Gs tested, only *Lm*EIF4G3 interacts specifically with *Lm*EIF4A1 as well as with human eIF4A.

Thus, our results are consistent with roles for both *Lm*EIF4A1 and *Lm*EIF4G3 in the process of parasite RNA translation. In order to compare the ratio of the eIF4A/4G homologues in *L. major* with those described for other eukaryotes, an antibody was raised against the His-*Lm*EIF4G3$_{26–310}$

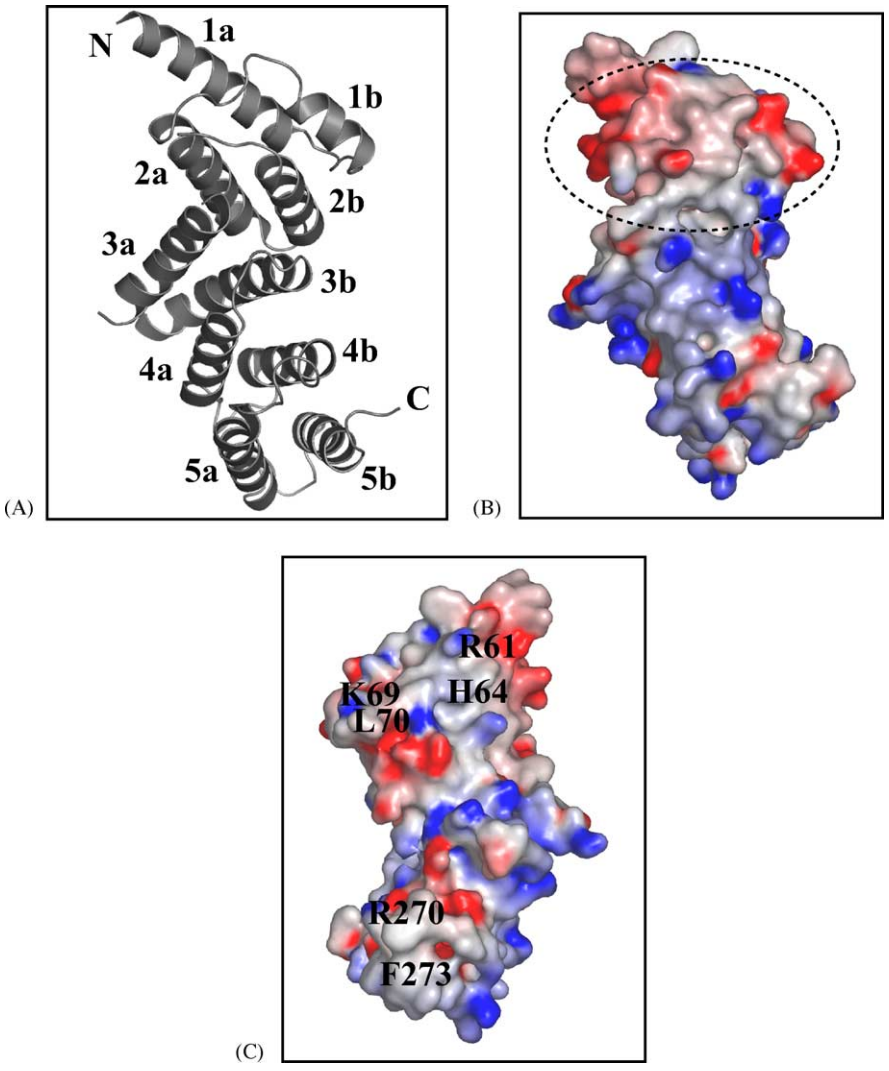


Fig. 9. Model of the structure of the *Lm*EIF4G3 HEAT domain. (A) Ribbon diagram of the structure of the predicted *Lm*EIF4G3 HEAT domain with the concave surface on the right and the convex surface on the left. The α helices are labelled as described for human eIF4GII [26] and as shown in Fig. 7. (B)–(C) Grasp representation of the surface of the domain colored coded for electrostatic potential (red <−8kBT; blue >8kBT). A and B represent identical views of the molecule whilst in C the opposite side of the molecule (rotated 180° about the long axis relative to (A) and (B)), predicted to be involved in the interaction with eIF4A, is shown. In (B) the dotted circle delimits the region in *Lm*EIF4G3 which differs from the human eIF4GII HEAT domain. Highlighted in (C) are the amino acids R61, H64, K69, L70, R270, F273, equivalent to amino acids implicated in the binding to eIF4A in the mammalian protein.

to enable quantitation of the endogenous *L. major* *Lm*EIF4G3 levels, as previously done for *Lm*EIF4E1-3 and *Lm*EIF4A1-2, using full length GST-*Lm*EIF4G3$_{1–636}$ for comparison. This quantitation (Fig. 8C and Table 2) indicates that, at about (4–8) $\times$ 10^3 molecules per cell, *Lm*EIF4G3 is present at about twice the level observed for *Lm*EIF4E1 in *L. major* promastigotes, and is about 10- and 50-fold less abundant than *Lm*EIF4E3 and *Lm*EIF4A1, respectively. The large differences in concentration between the eIF4E/eIF4G or eIF4A/eIF4G homologues are similar to those reported in yeast, where the eIF4G levels are about 20- and 45-fold less abundant than those of eIF4E and eIF4A, respectively [56]. However, considering that the number of *Lm*EIF4G3 molecules in *L. major* is about one-third to one half of the levels in yeast, it is unclear whether the intracellular levels of *Lm*EIF4G3 would be sufficient to support translation in *L. major*. While it would appear to function in translation or translational control, based on its sequence and the *Lm*EIF4A binding data, we cannot rule out contributions from the other eIF4G homologues in these processes.

3.9. Molecular modelling of the LmEIF4G3 HEAT domain

So far, the full length eIF4G has been refractory to structural studies. Only recently have individual domains within this protein, such as the central HEAT domain from human eIF4GII [26] and the eIF4E interacting region from yeast eIF4GI [71], been solved at the structural level. Given the evidence obtained above implicating *Lm*EIF4G3 in translation in *L. major*, we decided to attempt the molecular modelling of *Lm*eIF4G3. However, its very short N-terminus, only 50 amino acids long, plus the apparent lack of the consensus eIF4E binding motif (see Section 4), indicates that it may not bind eIF4E factors in the same manner as shown for the mammalian and yeast proteins [53,54,71]. Nevertheless, the *Lm*eIF4G3 HEAT domain could be modelled and its structural features associated with eIF4A binding were compared with the structure of the human eIF4GII HEAT domain.

Fig. 9A shows the ribbon drawing of the predicted three-dimensional structure of the *Lm*EIF4G3 HEAT domain. Absent from the drawings are the two segments corresponding to two loops (amino acids 128–140 and 161–178) which were not structured in the original human protein. As for the *Lm*EIF4E1 model, the overall quality of the final model was confirmed by the programs PROCHECK, Verify 3D and WHATIF. The *Lm*EIF4G3 HEAT domain folds into the same crescent-shaped molecule described for the human protein, with the five pairs of antiparallel α helices forming the same double layer with the convex and concave surfaces formed by the A and B helices, respectively (Fig. 9A). A major difference between the *Leishmania* and mammalian protein arises from the analysis of the surface electrostatic potential of the model performed with the program GRASP (Fig. 9B and C). This difference is concentrated in the region which in the human protein has been implicated in the interaction with the EMCV IRES [26]. In human eIF4GII this region has an overall positive charge but in the *Lm*EIF4G3 HEAT domain the equivalent region is negatively charged (Fig. 9B). In contrast, the region involved in the eIF4A interaction presents a surface electrostatic potential similar to the human model and all the conserved amino acids implicated in the interaction with eIF4A (R61, H64, K69, L70, R270 and F273 equivalent to R756, R759, K764, L765, R968 and F971 in the mammalian protein) are positioned adequately for the interaction to occur (Fig. 9C and [26]). Again these results support a role for *Lm*EIF4G3 in translation, although the lack of a possible eIF4E interacting region and the differences in the putative RNA binding side of the protein's HEAT domain might indicate a regulatory role in protein synthesis or a different mechanism for its function in translation initiation.

4. Discussion

The three eIF4F initiation factors have multiple isoforms in *L. major* protozoa. This complexity is reflected in *T. brucei* (which have similar isoforms) and indicates that they all have conserved roles in the parasites. Our results show that for each factor, one of the *L. major* forms seems to fulfil, in part at least, the expected characteristics of a translation initiation factor. However we believe that under more stringent assays, at least some of the other isoforms may also have specific roles in translation. For instance, it is possible that the various eIF4E homologues have different affinities for the parasite cap4 structure than those observed for the cap-Sepharose resin. Recently the chemical synthesis of the trypanosomatid cap4 has been achieved and the binding of a *Leishmania* eIF4E homologue, LeishIF4E-1 (equivalent to the *Lm*EIF4E1 protein described here), to this cap4 structure investigated through fluorescence titration measurements [72,73]. LeishIF4E-1 bound to both the cap4 and m^7GTP with very similar affinities, confirming the results obtained from the simple cap binding assay shown in Fig. 3 as well as our sequence and structural analysis. It remains to be seen whether either *Lm*EIF4E2 or 3, which do not bind cap-Sepharose, as well as *Lm*EIF4E4, not yet investigated experimentally, would be able to bind the cap4 structure. It is also possible that any one of these proteins may require association to the eIF4G homologues in order to bind efficiently to the cap, since in other eukaryotes the complex eIF4E/eIF4G binds with a higher affinity to the cap structure than eIF4E alone [71]. Alternatively, one last possibility would be for some of these other eIF4E homologues to act as regulators of *Lm*EIF4E1 function by competing with it for binding to eIF4G.

The existence of multiple isoforms for the eIF4F subunits in other eukaryotes, especially pluricellular organisms, may be associated with different patterns of tissue expression and during development. Insights into the role of the multiple isoforms come from the nematode system which in many aspects regarding mRNA metabolism can be compared to the trypanosomatids. In *Ascaris lumbricoides*, a 22 nt SL se-

quence and its associated trimethylated cap, *trans*-spliced to about 80–90% of all mRNAs, has been shown to functionally collaborate to enhance translation, very likely at the level of initiation [10]. It is not known how these features influence protein synthesis, but it seems likely that they do so via *trans*-acting factors. The presence of multiple eIF4E homologues in *C. elegans* is also reminiscent of what we see in the parasite system [51]. The five nematode homologues differ in cap binding affinity, requirement for viability [74] and possible roles in development [75]. In the case of the trypanosomatid protozoan unicellular organisms, the multiple eIF4F isoforms could be associated with their different life stages or be required for the translation of different classes of mRNAs. Indirect evidence for the second hypothesis comes from wheat germ, where two different eIF4F isoforms have been described, with distinct eIF4E and eIF4G subunits, which differ in their ability to translate mRNAs containing structured regions in their 5′UTR as well as uncapped mRNAs and dicistronic messages [76].

Our own results tend to indicate a stage specific expression for at least some of the eIF4E orthologues, since in *T. brucei* we have observed that the *Tb*EIF4E3 protein is expressed at high levels in the procyclic stage and it is absent from bloodstream forms (Dhalia et al., unpublished results). The mRNA encoding *Lm*EIF4E1 contains in its 3′UTR a sequence similar (68% identity) to the regulatory element found in the *Leishmania* amastin mRNA [77,78]. This element is found in a number of *Leishmania* mRNAs, several of which are differentially expressed in *L. donovani* amastigotes, and can confer amastigote-specific expression to a reporter mRNA possibly by regulating translation instead of stability [78]. It seems plausible that the expression of *Lm*eIF4E1 may be enhanced at the amastigote stage specially considering that, assuming a similar behaviour to the *T. brucei* orthologue, *Lm*eIF4E3 (the most abundant of the three eIF4E proteins in promastigotes), may be absent from the *Leishmania* mammalian form as well. Additional evidence for the stage specific expression of the eIF4E homologues was provided recently by microarray analysis, where the *Lm*EIF4E2 coding sequence was found within a subset of genes preferentially expressed in metacyclic populations of *L. major* [79]. Regarding the eIF4A subunit, the differences in levels between the two *Leishmania* eIF4A orthologues leave no doubt as to the relevant role for *Lm*EIF4A1 in translation in promastigotes. Since equivalent levels of expression have been reported for *Lm*EIF4A1 in both promastigote and amastigote stages of the *Leishmania* life cycle [37], it is unlikely that *Lm*EIF4A2 would replace it to a significant extent in the mammalian stage.

The multiple candidate eIF4G homologues identified in *L. major* add yet a new level of complexity to the study of translation initiation in this parasite. None of the identified homologues can be unambiguously assigned the role of a translation factor. Although *Lm*EIF4G3 (and possibly the related *Lm*EIF4G4) seems to be clearly involved in translation, its very short N-terminus may not accommodate binding sites for both eIF4E and PABP, as in other eukaryotes. It is possible that both *Lm*EIF4G3-4 function as a translational regulator/inhibitor as proposed for the mammalian protein p97/DAP-5/NAT1 (reviewed in [12]), which is homologous to the C-terminal two-thirds of mammalian eIF4G but lacks the N-terminal one-third including the eIF4E- and PABP-binding sites. However none of the other eIF4G homologues from *L. major* share any homology to *Lm*EIF4G3-4 outside the HEAT domain and despite having much longer N-termini, they do not have well defined eIF4E-binding motifs.

To continue the comparison with the nematode system, we searched for eIF4G homologues in the *C. elegans* genome. Strikingly, we found only one clear homologue which contains what seems to be a modified eIF4E binding sequence, **F**GR**D**F**M**V (GenBank accession NP495729). Similar sequences **F**SL**D**E**VV** and **F**SL**E**R**VL** are present in the short N-termini of *Lm*EIF4G3 and *Lm*EIF4G4 respectively and variations of it can also be found in possible *T. brucei* and *T. cruzi* orthologues. Their similarity to the nematode sequence suggests that all three sequences may bind eIF4E proteins, in which case both *Lm*EIF4G3 and *Lm*EIF4G4 could be true eIF4G orthologues. However considering the possible deviation from the eIF4E binding consensus the potential eIF4E binding sites need to be determined experimentally.

In summary, the results reported here indicate that *Lm*EIF4E1, *Lm*EIF4A1 and *Lm*EIF4G3 are the factors that functionally interact with the cap structure or each other, suggesting that they perform essential roles in protein synthesis. In order to assess their importance for cellular viability, as well as of the other homologues identified, we are currently performing RNA interference experiments in *T. brucei*. To address their function we will continue the protein–protein pull down assays and investigate possible eIF4G/eIF4E interactions. Finally expression analysis will be carried out so as to identify patterns of expression during the parasite life cycle for the various homologues. We expect with these studies to define, within the homologues already identified, which ones are required for translation initiation in trypanosomatids and whether their expression/activity can be regulated during the different stages of differentiation of these unique organisms.

Acknowledgements

This work was supported by grants/studentships from CNPq, CAPES, FIOCRUZ, FACEPE and the Wellcome Trust. Technical assistance to grow the *Leishmania* cells was provided by L. Rocha. We thank A. Ivens for the gift of cosmid DNAs and help with some preliminary sequence data. The human eIF4A cDNA in plasmid pET(His$_6$-eIF4A) and the *X. laevis* eIF4E cDNA in pSP64TEN were kind gifts from I. Ali, R. Jackson and S. Morley, respectively. We would also like to acknowledge E. Cupolillo for the *L. major* Friedlin strain, M. Carrington for access to preliminary *T. brucei* eIF4E sequences and A. Git for the GST-cdc2 clone. Finally, we are deeply indebted to O. Thiemann and D.H.F Souza for

allowing us access to the computing facilities necessary to the modelling work. Sequencing of *L. major* was accomplished as part of the *Leishmania* genome network and was financially supported by the Wellcome Trust, Burroughs Wellcome Fund, the EU, WHO/TDR and NIAID. This work could not be accomplished as such without the free availability of the sequences produced by this network.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.molbiopara.2004.12.001.

References

[1] Muhich ML, Boothroyd JC. Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Mol Cell Biol 1988;8:3837–46.
[2] Ullu E, Tschudi C. Trans splicing in trypanosomes requires methylation of the 5′ end of the spliced leader RNA. Proc Natl Acad Sci USA 1991;88:10074–8.
[3] Campbell DA, Thomas S, Sturm NR. Transcription in kinetoplastid protozoa: why be normal? Microbes Infect 2003;5:1231–40.
[4] Liang XH, Haritan A, Uliel S, Michaeli S. *Trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell 2003;2:830–40.
[5] Clayton CE. Life without transcriptional control? From fly to man and back again. EMBO J 2002;21:1881–8.
[6] Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J Biol Chem 1992;267:9805–15.
[7] Nilsen TW. *Trans*-splicing: an update. Mol Biochem Parasitol 1995;73:1–6.
[8] Pascolo E, Blonski C, Shire D, Toulme JJ. Antisense effect of oligodeoxynucleotides complementary to the mini-exon sequence of the protozoan parasite *Leishmania amazonensis*. Biochimie 1993;75:43–7.
[9] Ramazeilles C, Mishra RK, Moreau S, Pascolo E, Toulme JJ. Antisense phosphorothioate oligonucleotides: selective killing of the intracellular parasite *Leishmania amazonensis*. Proc Natl Acad Sci USA 1994;91:7859–63.
[10] Maroney PA, Denker JA, Darzynkiewicz E, Laneve R, Nilsen TW. Most mRNAs in the nematode *Ascaris lumbricoides* are *trans*-spliced: a role for spliced leader addition in translational efficiency. RNA 1995;1:714–23.
[11] Zeiner GM, Sturm NR, Campbell DA. The *Leishmania tarentolae* spliced leader contains determinants for association with polysomes. J Biol Chem 2003;278:38269–75.
[12] Gingras AC, Raught B, Sonenberg N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem 1999;68:913–63.
[13] Hershey JWB, Merrick WC. Pathway and mechanism of initiation of protein synthesis. In: Sonenberg N, et al., editors. Translational control of gene expression. New York: Cold Spring Harbor Laboratory Press; 2000. p. 33–88.
[14] Pestova TV, Kolupaeva VG, Lomakin IB, et al. Molecular mechanisms of translation initiation in eukaryotes. Proc Natl Acad Sci USA 2001;98:7029–36.
[15] Sonenberg N, Dever TE. Eukaryotic translation initiation factors and regulators. Curr Opin Struct Biol 2003;13:56–63.
[16] Sonenberg N, Gingras AC. The mRNA 5′ cap-binding protein eIF4E and control of cell growth. Curr Opin Cell Biol 1998;10:268–75.
[17] Raught B, Gingras AC. eIF4E activity is regulated at multiple levels. Int J Biochem Cell Biol 1999;31:43–57.
[18] Scheper GC, Proud CG. Does phosphorylation of the cap-binding protein eIF4E play a role in translation initiation? Eur J Biochem 2002;269:5350–9.
[19] Luking A, Stahl U, Schmidt U. The protein family of RNA helicases. Crit Rev Biochem Mol Biol 1998;33:259–96.
[20] Linder P. Yeast RNA helicases of the DEAD-box family involved in translation initiation. Biol Cell 2003;95:157–67.
[21] Rocak S, Linder P. DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol 2004;5:232–41.
[22] Prevot D, Darlix JL, Ohlmann T. Conducting the initiation of protein synthesis: the role of eIF4G. Biol Cell 2003;95:141–56.
[23] Mader S, Lee H, Pause A, Sonenberg N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. Mol Cell Biol 1995;15:4990–7.
[24] Tarun Jr SZ, Sachs AB. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. EMBO J 1996;15:7168–77.
[25] Imataka H, Gradi A, Sonenberg N. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO J 1998;17:7480–9.
[26] Marcotrigiano J, Lomakin IB, Sonenberg N, Pestova TV, Hellen CU, Burley SK. A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. Mol Cell 2001;7:193–203.
[27] Korneeva NL, Lamphear BJ, Hennigan FL, Rhoads RE. Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. J Biol Chem 2000;275:41369–76.
[28] Lomakin IB, Hellen CU, Pestova TV. Physical association of eukaryotic initiation factor 4G (eIF4G) with eIF4A strongly enhances binding of eIF4G to the internal ribosomal entry site of encephalomyocarditis virus and is required for internal initiation of translation. Mol Cell Biol 2000;20:6019–29.
[29] Imataka H, Sonenberg N. Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. Mol Cell Biol 1997;17:6940–7.
[30] Morino S, Imataka H, Svitkin YV, Pestova TV, Sonenberg N. Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. Mol Cell Biol 2000;20:468–77.
[31] Pyronnet S, Imataka H, Gingras AC, Fukunaga R, Hunter T, Sonenberg N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. EMBO J 1999;18:270–9.
[32] Nakamura Y, Hashimoto T, Kamaishi T, et al. Phylogenetic position of kinetoplastid protozoa inferred from the protein phylogenies of elongation factors 1alpha and 2. J Biochem (Tokyo) 1996;119:70–9.
[33] Batista JA, Teixeira SM, Donelson JE, Kirchhoff LV, de Sa CM. Characterization of a Trypanosoma cruzi poly(A)-binding protein and its genes. Mol Biochem Parasitol 1994;67:301–12.
[34] Hotchkiss TL, Nerantzakis GE, Dills SC, Shang L, Read LK. *Trypanosoma brucei* poly(A) binding protein I cDNA cloning, expression, and binding to 5′ untranslated region sequence elements. Mol Biochem Parasitol 1999;98:117–29.
[35] Bates EJ, Knuepfer E, Smith DF. Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites. Nucl Acids Res 2000;28:1211–20.
[36] Skeiky YA, Kennedy M, Kaufman D, et al. LeIF: a recombinant *Leishmania* protein that induces an IL-12-mediated Th1 cytokine profile. J Immunol 1998;161:6171–9.

[37] Skeiky YA, Guderian JA, Benson DR, et al. A recombinant *Leishmania* antigen that stimulates human peripheral blood mononuclear cells to express a Th1-type cytokine profile and to produce interleukin 12. J Exp Med 1995;181:1527–37.
[38] Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 2001;17:1244–5.
[39] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–25.
[40] Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815.
[41] Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell 1997;89:951–61.
[42] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
[43] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.
[44] Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83–5.
[45] Vriend G. WHATIF: a molecular modelling and drug design program. J Mol Graph 1990;8:52–6.
[46] de Melo Neto OP, Standart N, de Sa CM. Autoregulation of poly(A)-binding protein synthesis in vitro. Nucl Acids Res 1995;23:2198–205.
[47] Minshall N, Thom G, Standart N. A conserved role of a DEAD box helicase in mRNA masking. RNA 2001;7:1728–42.
[48] Pestova TV, Hellen CU, Shatsky IN. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. Mol Cell Biol 1996;16:6859–69.
[49] Melo EO, Dhalia R, de Sa CM, Standart N, de Melo Neto OP. Identification of a C-terminal PABP-PABP interaction domain: role in cooperative binding to poly(A) and efficient cap distal translational repression. J Biol Chem 2003;278:46357–68.
[50] Clayton C, Adams M, Almeida R, et al. Genetic nomenclature for *Trypanosoma* and *Leishmania*. Mol Biochem Parasitol 1998;97:221–4.
[51] Jankowska-Anyszka M, Lamphear BJ, Aamodt EJ, et al., Rhoads RE. Multiple isoforms of eukaryotic protein synthesis initiation factor 4E in *Caenorhabditis elegans* can distinguish between mono-and trimethylated mRNA cap structures. J Biol Chem 1998;273:10538–42.
[52] Matsuo H, Li H, McGuire AM, et al. Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. Nat Struct Biol 1997;4:717–24.
[53] Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. Mol Cell 1999;3:707–16.
[54] Niedzwiecka A, Marcotrigiano J, Stepinski J, et al. Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5′ cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. J Mol Biol 2002;319:615–35.
[55] Ptushkina M, von der Haar T, Vasilescu S, Frank R, Birkenhager R, McCarthy JE. Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5′ cap in yeast involves a site partially shared by p20. EMBO J 1998;17:4798–808.
[56] von der Haar T, McCarthy JE. Intracellular translation initiation factor levels in *Saccharomyces cerevisiae* and their role in cap-complex function. Mol Microbiol 2002;46:531–44.
[57] Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–96.
[58] Li Q, Imataka H, Morino S, et al. Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. Mol Cell Biol 1999;19:7336–46.
[59] Lu J, Aoki H, Ganoza MC. Molecular characterization of a prokaryotic translation factor homologous to the eukaryotic initiation factor eIF4A. Int J Biochem Cell Biol 1999;31:215–9.
[60] Story RM, Li H, Abelson JN. Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. Proc Natl Acad Sci USA 2001;98:1465–70.
[61] Tanner NK, Cordin O, Banroques J, Doere M, Linder P. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Mol Cell 2003;11:127–38.
[62] Caruthers JM, Johnson ER, McKay DB. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proc Natl Acad Sci USA 2000;97:13080–5.
[63] Minshall N, Thom G, Standart N. A conserved role of a DEAD box helicase in mRNA masking. RNA 2001;7:1728–42.
[64] Coller JM, Tucker M, Sheth U, Valencia-Sanchez MA, Parker R. The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA 2001;7:1717–27.
[65] Duncan R, Milburn SC, Hershey JW. Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. J Biol Chem 1987;262:380–8.
[66] Rau M, Ohlmann T, Morley SJ, Pain VM. A reevaluation of the cap-binding protein, eIF4E, as a rate-limiting factor for initiation of translation in reticulocyte lysate. J Biol Chem 1996;271:8983–90.
[67] Asano K, Shalev A, Phan L, et al. Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. EMBO J 2001;20:2326–37.
[68] He H, von der Haar T, Singh CR, et al. The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection. Mol Cell Biol 2003;23:5431–45.
[69] Tarun Jr SZ, Wells SE, Deardorff JA, Sachs AB. Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc Natl Acad Sci USA 1997;94:9046–51.
[70] Joshi B, Yan R, Rhoads RE. In vitro synthesis of human protein synthesis initiation factor 4 gamma and its localization on 43 and 48 S initiation complexes. J Biol Chem 1994;269:2048–55.
[71] Gross JD, Moerke NJ, von der Haar T, et al. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell 2003;115:739–50.
[72] Lewdorowicz M, Yoffe Y, Zuberek J, et al. Chemical synthesis and binding activity of the trypanosomatid cap-4 structure. RNA 2004;10:1–10.
[73] Yoffe Y, Zuberek J, Lewdorowicz M, et al. Cap-binding activity of an eIF4E homolog from *Leishmania*. RNA 2004;10:1764–75.
[74] Keiper BD, Lamphear BJ, Deshpande AM, et al. Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*. J Biol Chem 2000;275:10590–6.
[75] Amiri A, Keiper BD, Kawasaki I, et al. An isoform of eIF4E is a component of germ granules and is required for spermatogenesis in *C. elegans*. Development 2001;128:3899–912.
[76] Gallie DR, Browning KS. eIF4G functionally differs from eIFiso4G in promoting internal initiation, cap-independent translation, and translation of structured mRNAs. J Biol Chem 2001;276:36951–60.
[77] Wu Y, El Fakhry Y, Sereno D, Tamar S, Papadopoulou B. A new developmentally regulated gene family in *Leishmania amastigotes* encoding a homolog of amastin surface proteins. Mol Biochem Parasitol 2000;110:345–57.
[78] Boucher N, Wu Y, Dumas C, et al. A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element. J Biol Chem 2002;277:19511–20.
[79] Almeida R, Gilmartin BJ, McCann SH, et al. Expression profiling of the *Leishmania* life cycle: cDNA arrays identify developmentally regulated genes present but not annotated in the genome. Mol Biochem Parasitol 2004;136:87–100.

8.3. Apêndice 4
PARTICIPAÇÃO DO DISCENTE EM ARTIGO PUBLICADO

The two eIF4A helicases in Trypanosoma brucei are functionally distinct

Published online May 10, 2006

Nucleic Acids Research, 2006, Vol. 34, No. 9 ***2495–2507***
doi:10.1093/nar/gkl290

The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct

Rafael Dhalia, Nina Marinsek[1], Christian R. S. Reis, Rodolfo Katz, João R. C. Muniz[2], Nancy Standart[1], Mark Carrington[1] and Osvaldo P. de Melo Neto*

Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Avenue Moraes Rego s/n, Campus UFPE, Recife PE 50670-420, Brazil, [1]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK and [2]Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos SP 13560-970, Brazil

Received February 1, 2006; Revised March 4, 2006; Accepted April 6, 2006

ABSTRACT

Protozoan parasites belonging to the family *Trypanosomatidae* are characterized by an unusual pathway for the production of mRNAs via polycistronic transcription and *trans*-splicing of a 5′ capped mini-exon which is linked to the 3′ cleavage and polyadenylation of the upstream transcript. However, little is known of the mechanism of protein synthesis in these organisms, despite their importance as agents of a number of human diseases. Here we have investigated the role of two *Trypanosoma brucei* homologues of the translation initiation factor eIF4A (in the light of subsequent experiments these were named as *Tb*EIF4AI and *Tb*EIF4AIII). eIF4A, a DEAD-box RNA helicase, is a subunit of the translation initiation complex eIF4F which binds to the cap structure of eukaryotic mRNA and recruits the small ribosomal subunit. *Tb*EIF4AI is a very abundant predominantly cytoplasmic protein (over 1×10^5 molecules/cell) and depletion to ~10% of normal levels through RNA interference dramatically reduces protein synthesis one cell cycle following double-stranded RNA induction and stops cell proliferation. In contrast, *Tb*EIF4AIII is a nuclear, moderately expressed protein (~$1–2 \times 10^4$ molecules/cell), and its depletion stops cellular proliferation after approximately four cell cycles. Ectopic expression of a dominant negative mutant of *Tb*EIF4AI, but not of *Tb*EIF4AIII, induced a slow growth phenotype in transfected cells. Overall, our results suggest that only *Tb*EIF4AI is involved in protein synthesis while the properties and sequence of *Tb*EIF4AIII indicate that it may be the orthologue of eIF4AIII, a component of the exon junction complex in mammalian cells.

INTRODUCTION

The flagellate protozoan parasites belonging to the family *Trypanosomatidae* include a number of important pathogens responsible for diseases of worldwide impact such as the Sleeping Sickness (*Trypanosoma brucei*), Chagas' Disease (*Trypanosoma cruzi*) and the various forms of Leishmaniasis (*Leishmania* sp.) (www.who.int/tdr). These organisms are unusual in a number of processes necessary for mRNA synthesis and maturation; transcription is polycistronic and monocistronic mRNAs arise after *trans*-splicing of a capped short exon on to the 5′ end and cleavage and polyadenylation at the 3′ end [reviewed in (1,2)]. As a result of *trans*-splicing, the 5′ ends of mature trypanosomatid mRNAs all share the same 39 nt leader sequence with a modified cap 4 structure (3). To date, little is known about how these mRNAs are translated, if major differences exist within the process of protein synthesis when compared with other eukaryotes and whether the common leader sequence influences how the mRNAs are recruited for translation.

In eukaryotes, protein synthesis is a complex process which requires a myriad of different macromolecules including RNAs and proteins. The critical initiation step requires a number of translation initiation factors (eIFs) whose activity can be highly regulated [for reviews see (4–7)]. Paramount within these factors is the heterotrimeric eIF4F complex, which is required for the recruitment of the small ribosomal subunit to the 5′ end of the mRNA. eIF4F is composed of the RNA helicase eIF4A, the cap-binding protein eIF4E and the large scaffolding protein eIF4G which mediates interactions between eIF4F and other translation factors as well as the small ribosomal subunit [reviewed in (8)].

*To whom correspondence should be addressed. Tel: 55 81 2101 2636; Fax: 55 81 3453 2449; Email: opmn@cpqam.fiocruz.br

eIF4A is the prototype member of the DEAD-box family of RNA helicases which includes several proteins mainly involved in RNA metabolism. These proteins are classified within the superfamily II of a much larger group of related RNA and DNA helicases (9). The RNA helicases couple the hydrolysis of ATP to various activities relevant for RNA function such as rearrangement of inter- or intra-molecular RNA structures, dissociation of RNA–protein complexes and RNA unwinding. The DEAD-box family members are characterized by nine sequence motifs (I, Ia, Ib, II, III, IV, V and VI and the Q motif), as well as several individual amino acids, conserved among the various proteins assigned to this family [for reviews see (10,11)]. Structurally, eIF4A assumes a 'dumbbell' shape with two globular domains connected by a flexible linker (12). Comparison with the structure of related RNA and DNA helicases and the *Methanococcus jannaschii* DEAD-box protein, similar in size to eIF4A (13), have confirmed the basic overall structure of the core helicase domains. The various conserved motifs are positioned in the interface between the two domains and have been implicated in RNA binding and ATP binding and hydrolysis. However, little is known about the molecular basis for RNA specificity and helicase function [reviewed in (11,14)]. In translation initiation, eIF4A binds to the central region of eIF4G, via the eIF4G HEAT domain (15) and, in mammals at least, also to the eIF4G C-terminus (16,17). eIF4A seems to be responsible for melting secondary structures along the mRNA 5′-untranslated region (5′-UTR), facilitating the binding of the small ribosomal subunit and the scanning of the leader region to locate the initiation codon (18,19) [reviewed in (4,6)].

In mammals three different isoforms of eIF4A have been described. Both eIF4AI and II (90% identity between the two proteins) are able to reconstitute the eIF4F subunit and presumably have similar roles in translation (20,21). In contrast, eIF4AIII, only 66% identical to mammalian eIF4AI, is functionally distinct. While eIF4AIII exhibits RNA-dependent ATPase activity and ATP-dependent RNA helicase activity, it does not support binding of the small ribosomal subunit to the mRNA, and inhibits translation *in vitro* (22). eIF4AIII localizes to the nucleus (23) and recent reports indicate that it may act as an anchoring factor for the exon junction complex (EJC), and is essential for nonsense-mediated decay (NMD) in mammals (24–30).

The mechanisms of translation initiation are virtually unknown in trypanosomatids. A *Leishmania* eIF4A homologue (called LeiF) was first described in *Leishmania braziliensis* and *Leishmania major* as a 45.3 kDa antigen, expressed in both insect and mammalian stages of the parasite life cycle, but its role in translation was not investigated (31,32). Recently, our group has identified multiple *L.major* homologues for the three eIF4F subunits, all of which are conserved in *T.brucei* (33). We characterized two putative *L.major* eIF4A homologues, *Lm*EIF4A1 (LeiF) and *Lm*EIF4A2, with 59 and 52% identities to human eIF4AI, respectively. When assayed with isoform specific antibodies these two factors differ significantly in abundance in *L.major* promastigotes. *Lm*EIF4A1 is very abundant with over 10^5 molecules/cell whilst *Lm*EIF4A2 is either absent or present at levels below 10^4 molecules/cell. Furthermore, only *Lm*EIF4A1 was found to bind specifically to the HEAT domain of one of the *Leishmania* eIF4G homologues (33).

In this paper we take advantage of the genetic tools available for the study of gene function in *T.brucei* to extend this analysis of the two trypanosomatid eIF4A homologues. Initially, the mRNA and protein levels of the two *T.brucei* eIF4A orthologues were analysed during the life cycle. Their intracellular localization was identified through overexpression of enhanced yellow fluorescent protein (EYFP) fusions and their role for parasite viability investigated through RNA interference and overexpression of dominant negative mutants. Our results show that the *T.brucei* orthologue of *Lm*EIF4A1 (named as *Tb*EIF4AI) is the functional homologue of the eIF4A present in eIF4F. As for the orthologue of *Lm*EIF4A2, it seems to be the functional homologue of the nuclear eIF4AIII present in higher eukaryotes and has been named here as *Tb*EIF4AIII.

MATERIALS AND METHODS

Sequence analysis and molecular modeling

BLAST searches were carried out with the *T.brucei* genome sequences available at the Gene DB website of the Sanger Institute Pathogen Sequencing Unit (www.genedb.org). Further sequence searches, Clustal W alignments and molecular modeling were done as described previously (33).

PCR and cloning methods

The *Tb*EIF4AI coding sequence was amplified from *T.brucei* Lister 427 genomic DNA (5′primer, AAG CTT CCG CCA CCA TGG CCC AAC AAG GAA AG; and 3′primer, GGA TCC AGA ACC CTC ACC AAG GTA GGC AGC; added restriction sites used in cloning are underlined) resulting in the entire open reading frame (ORF) flanked by sites for the enzymes HindIII and BamHI. The same strategy was used for the amplification of the *Tb*EIF4AIII sequence (5′primer, AAG CTT CCG CCA CCA TGA CAG CAA CCG CAA GG; and 3′primer, GGA TCC AGA ACC GAA CTG TTC ACC GAC GTT TG). The amplified fragments were then cloned into the vector pGEM-T Easy (Promega) and sequenced. In order to express N-terminal His-tagged fusion proteins both fragments were then recovered by digestion with HindIII and BamHI and subcloned into the same sites of a modified pET15b vector. To generate the *Tb*EIF4AI-EYFP and *Tb*EIF4AIII-EYFP constructs, the two eIF4A fragments were cloned into the HindIII and BamHI sites of p2215, a modified form of pLEW82 (34). To make p2215, the EYFP ORF (Clontech) was obtained as a BamHI/BglII fragment and inserted into the BamHI site of pLEW82. On expression, the resultant fusion protein had the sequence: eIF4A C-terminal residue-GSGSGGG-EYFP. For the RNAi experiments the same two eIF4A DNA fragments were also subcloned into the HindIII/BamHI sites of the transfection vector p2T7-177 (35). Dominant negative mutants were made by altering the sequence of the DEAD box of the helicase (motif II in Figure 1) to DQAD (11,36). Tetracycline-inducible expression of wild type and dominant negative forms of eIF4A was performed using p2280, a derivative of pLEW100 made by introducing a BamHI/BglII DNA segment encoding three tandem myc epitope tags to its BamHI site. The two HindIII/BamHI *T.brucei* eIF4A fragments were cloned into the same sites of p2280 resulting in the expression of fusion proteins with the myc epitope tags on their C-terminus giving the sequence

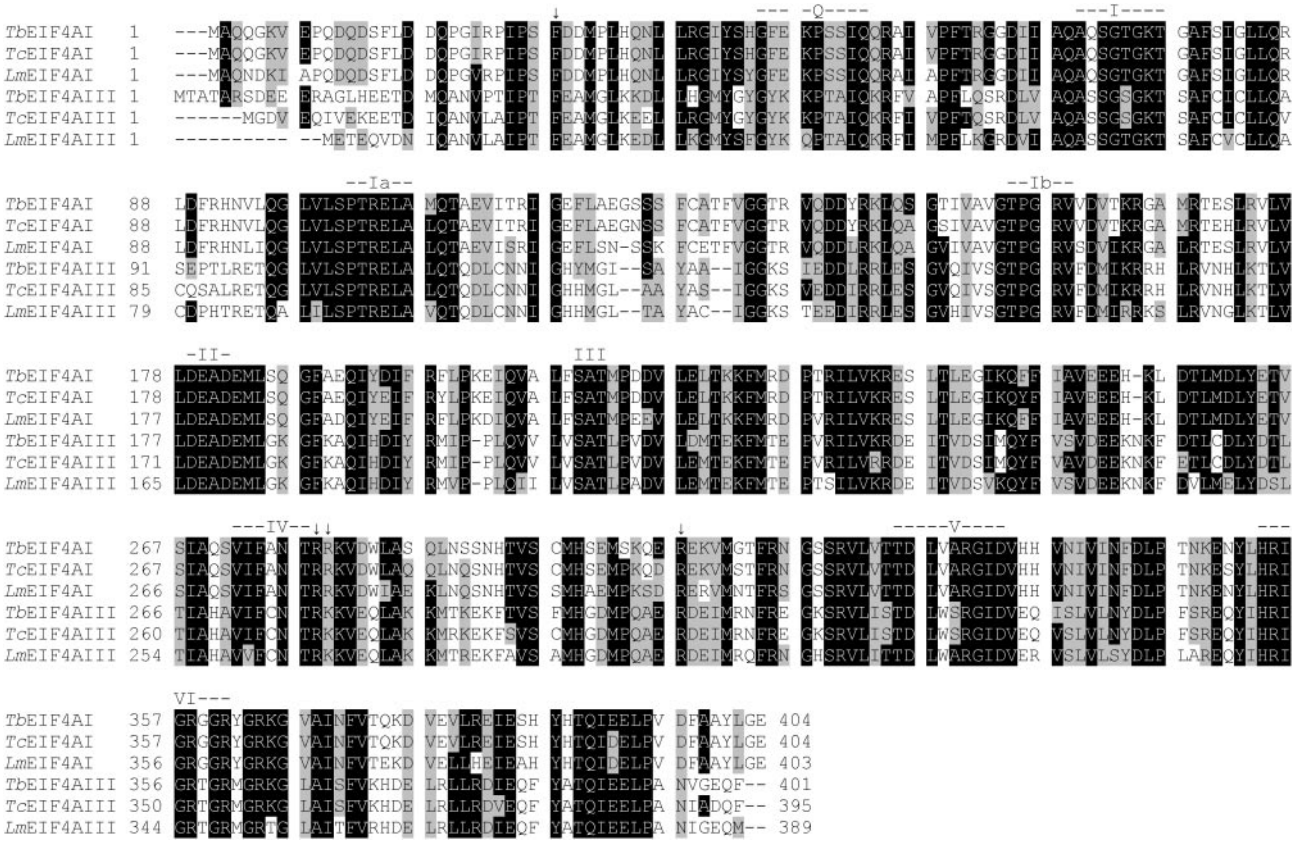


Figure 1. Sequence alignment comparing the *T.brucei*, *T.cruzi* and *L.major* eIF4A homologues. Sequences were aligned with the Clustal W program, from the Centre for Molecular and Biomolecular Informatics (http://www.cmbi.kun.nl/bioinf/tools/clustalw.shtml). Amino acids identical in >60% of the sequences are highlighted in dark gray, while amino acids defined as similar, based on the BLOSUM 62 Matrix, on >60% of the sequences, are shown in pale gray. When necessary, gaps were inserted within the various sequences (dashes) to allow better alignment. The nine motifs typical of DEAD-box RNA helicases (10,11) are highlighted. The single arrows indicate other individual amino acids which seems to be relevant for eIF4A function or RNA binding (12,42). Relevant GenBank accession numbers: *Lm*EIF4AI, AAC24684/AAC24685; *Lm*EIF4AIII, CAJ05468; *Tb*EIF4AI, EAN76544; *Tb*EIF4AIII, EAN79829; *Tc*EIF4AI, EAN98527; *Tc*EIF4AIII, EAN88971.

eIF4A-GSGSGPREQKLISEEDLPREQKLISEEDLPREQK-LISEEDLPR.

Parasite growth, transfection and RNAi

Procyclic form *T.brucei* Lister 427 cells were used throughout. RNAi and ectopic expression of eIF4A were performed using *T.brucei* Lister 427 29-13, containing integrated copies of pLEW 29 and pLEW13 (34). Procyclic *T.brucei* forms were propagated in SDM-79 medium at 27°C, supplemented with 10% feotal calf serum (FCS). For the 29–13 cell line, cultures were also supplemented with G418 (15 μg/ml) and hygromycin (25 μg/ml). Parasite growth was monitored microscopically every 24 h. Mid-log phase cultures (10^6–10^7 cells/ml) were then used for transfection and total protein extract production. Bloodstream forms (Lister 427) were cultivated in HMI-9 medium (37) at 37°C, 5% CO_2, supplemented with 10% FCS. Cultures grown to mid-log phase cultures (10^5–10^6 cells/ml) were also harvested for the production of total protein extract.

Plasmids were linearized with NotI prior to electroporation and stable DNA integration was selected using phleomycin (2.5 μg/ml). For the RNAi experiments 1 μg/ml of tetracycline was added to mid-log phase cultures of transfected cells.

RNA analysis

RNA extraction and Northern blots were performed using standard methods (38,39). DNA fragments containing complete ORFs were used as probes for *Tb*EIF4As and EP procyclin. A genomic repeat containing both α- and β-tubulin genes was used to detect tubulin mRNA.

Recombinant protein expression, antibody production and western blots

His-tagged *Tb*EIF4AI and *Tb*EIF4AIII were expressed in *Escherichia coli* BL21 Star (DE3) using pET15b derived plasmids. The recombinant polypeptides were insoluble after lysing the cells using a French Press. The polypeptides were purified by preparative SDS–PAGE and the bands corresponding to the recombinant proteins were then excised and sent for the production of polyclonal serum (CovalAb). Prior to their use, both antibodies were first affinity purified as described elsewhere (40) with their respective recombinant proteins. Cross-reacting antibodies were eliminated by previous incubation of the anti-*Tb*EIF4AI antisera with *Tb*EIF4AIII recombinant protein and vice versa. To estimate the levels of the eIF4A proteins, first the recombinant proteins were quantified by serial dilutions in SDS–PAGE by comparison

with known concentrations of BSA (data not shown). After quantification they were then used in western blots with the respective antisera and compared with serial dilutions of total protein extract from both procyclic and bloodstream forms of *T.brucei*. The endogenous protein levels were then estimated by the densitometric analysis of the western blot results as described elsewhere (33).

Fluorescence microscopy

For the indirect immunofluorescence assay, wild-type procyclic cells grown to mid-log phase (5×10^6/ml) were harvested, washed with phosphate-buffered saline (PBS)/10 mM glucose and adsorbed to polylysine coated slides. The cells were then fixed in 100% methanol at −20°C/15 min. Antibody detection of *Tb*EIF4AI and III followed standard procedures. DNA was stained using Hoechst 33258. For the analysis of the cells expressing *Tb*EIF4AI-EYFP and *Tb*EIF4AIII-EYFP, aliquots of 5×10^6 cells were harvested, washed with PBS/10 mM glucose and fixed in 0.1% formaldehyde for 5 min. In this case, DNA was stained using Hoechst 33342.

Metabolic labelling

To measure the rate of protein synthesis, [^{35}S]methionine (10 μCi/ml) was added to mid-log cultures which were incubated for 1 h prior to the determination of trichloroacetic acid precipitable incorporation into protein. Parallel incubations in the presence of 50 μg/ml cycloheximide were used to estimate incorporation of radiolabel by processes other than cytoplasmic protein synthesis. For metabolic labeling, cultures were washed twice with methionine-free RPMI 1640 medium and then resuspended at 1×10^7 cells/ml in methionine-free RPMI 1640 containing 50 μCi/ml [^{35}S]methionine and incubated for 1 h at 28°C prior to harvesting and analysis by SDS–PAGE and autoradiography.

RESULTS

Identification of the *T.brucei* eIF4A homologues

The *T.brucei* homologues of eIF4A were identified in searches of the genome sequence using human eIF4AI as well as the two *Leishmania* eIF4A sequences. At the amino acid level, the two *T.brucei* proteins, *Tb*EIF4AI and *Tb*EIF4AIII, are very similar to their *L.major* orthologues with identities of 88% for the eIF4AI and 82% for the eIF4AIII pair (to avoid confusion and in view of the data presented below the *L.major* eIF4A homologues, previously called *Lm*EIF4A1 and 2, have also been renamed to *Lm*EIF4AI and *Lm*EIF4AIII and this nomenclature will be used when needed). The assignment as eIF4A homologues reflects the fact that the two sequences are the closest matches in the two parasite protein databases to human eIF4AI and both share identities of over 50% with the human protein. The third nearest eIF4A homologue in both *T. brucei* and *L. major* databases has been assigned to another group of RNA helicases, Dhh1 (33), with an identity of only ∼40% to human eIF4AI. In order to analyse the conservation of the putative eIF4A homologues within an additional member of the family *Trypanosomatidae,* we performed similar searches using the *T.cruzi* genome database. Again, orthologues to both proteins could be found in *T.cruzi* with the third nearest match to human eIF4AI being Dhh1.

Figure 1 shows a sequence alignment comparing the two eIF4A sequences from *T.brucei* with those from *T.cruzi* and *L.major*. Highlighted in the figure are the various conserved motifs typical of eIF4A and related proteins which have been shown to be required for different aspects of the RNA helicase activity. Motifs I, II, VI and the recently identified Q motif (41) have been implicated in ATP binding and hydrolysis; motif III may link nucleotide hydrolysis to helicase function; motifs Ia, Ib, IV and V may be involved in RNA binding [reviewed in (11,14)]. Several conserved arginine residues, which have also been implicated as important for eIF4A/helicase function in yeast eIF4A (12), as well as a conserved N-terminal phenylalanine residue are also shown (42). Overall the alignment confirms the close similarity between the various homologues. In general the N-terminal half of the protein is less conserved than the C-terminal half but only in the very N-terminus are significant differences in the sequences observed. A few conserved differences between the three eIF4AI and three eIF4AIII homologues with potential significance for their function can be identified, such as the replacement of a conserved F46 E47 doublet within the Q motif of eIF4AI by YK in eIF4AIII proteins. Other individual substitutions conserved between the eIF4AI and eIF4AIII homologues can be seen within motifs Q (S50T, S51A), I (Q71S), IV (A275C), V (V328W) and VI (G359T). So far however the functional significance of these substitutions is unknown.

Expression of *Tb*EIF4A mRNAs in *T.brucei* bloodstream and procyclic forms

To begin the functional characterization of *T.brucei* eIF4A homologues and assay their expression at the mRNA level, the two genes were amplified, cloned and used as probes in northern blots of RNA from *T.brucei* procyclic and bloodstream forms (Figure 2A). The membranes were also probed for the constitutively expressed tubulin, to confirm that equal amounts of mRNA were loaded in each lane, and for the procyclic-specific EP procyclin mRNA to verify the stage specificity of both sets of mRNAs (43).

The two *T.brucei* eIF4A mRNAs were readily detected and found to be at constant levels throughout the parasite life cycle. However, according to the northern blot, the *Tb*EIF4AI mRNA produces a much stronger signal than *Tb*EIF4AIII. Since both probes used were of similar specific activity, and the exposures times for the films were similar as well, it seems that the *Tb*EIF4AIII mRNA is far less abundant than *Tb*EIF4AI. Remarkably, although the ORFs for both proteins are similar, 1215 versus 1206 bp for *Tb*EIF4AI and *Tb*EIF4AIII, respectively, their mRNAs differ significantly, with the *Tb*EIF4AI message, at ∼3 kb, being nearly twice the length of *Tb*EIF4AIII (∼1.6 kb), probably reflecting a considerable difference in the length of the 3′-UTR.

*Tb*EIF4AI is about 10-fold more abundant than *Tb*EIF4AIII in both procyclic and bloodstream forms

Recombinant His-*Tb*EIF4AI and III were expressed in *E.coli*, purified from inclusion bodies and used to produce antisera. Affinity purification and depletion was used to produce specific antibodies for each protein (see below, Figure 6A). The

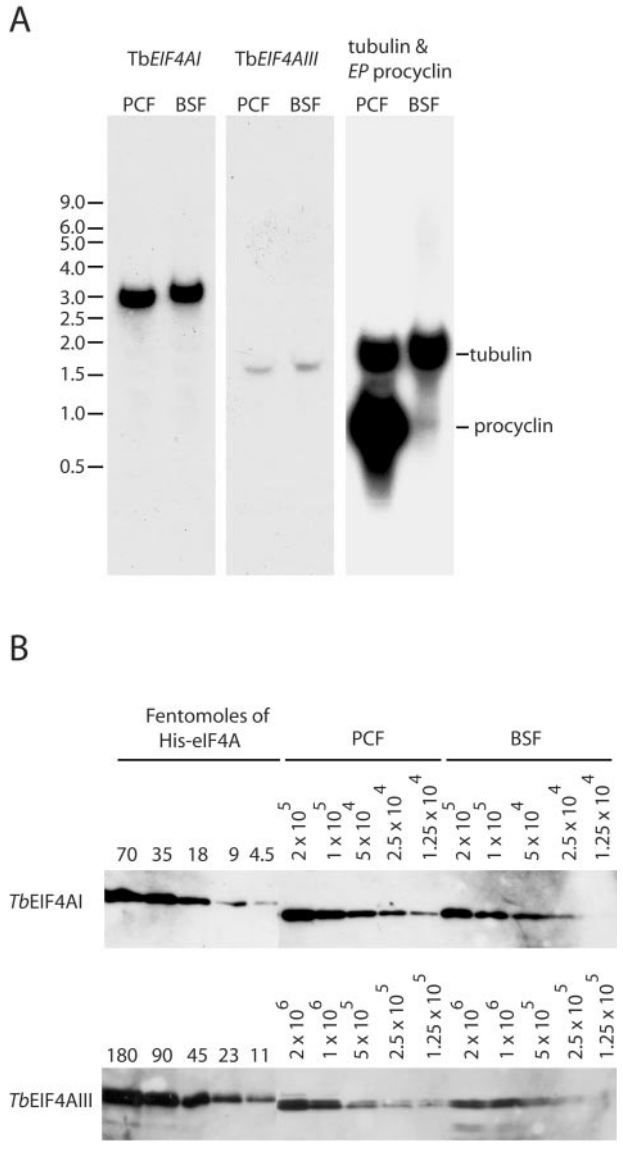


Figure 2. Expression analysis of *Tb*EIF4AI and III. (**A**) Total RNA from both procyclic (PCF) and bloodstream (BSF) *T.brucei* forms was separated on denaturing gels and used in northern blot assays to detect the expression of *Tb*EIF4AI and III. One of the blots was overprobed with tubulin (ubiquitously expressed) and EP procyclin (expressed in procyclics only) as controls. The migration of RNA size markers is indicated on the left in kilobases. (**B**) Quantification of *Tb*EIF4AI and *Tb*EIF4AIII in the procyclic and bloodstream forms of *T.brucei*. Recombinant His-tagged *Tb*EIF4AI and III were quantified, diluted to defined concentrations (in fmol) and ran on SDS–PAGE gels with whole parasite extract obtained from known number of cells from both procyclic and bloodstream forms (1.25×10^4–2×10^5 for *Tb*EIF4AI and 1.25×10^5–2×10^6 for *Tb*EIF4AIII). The proteins samples were then transferred to Immobilon-P membranes followed by incubation with the affinity purified isoform specific antisera and goat anti-rabbit IgG conjugated with peroxidase, and detection by ECL. The values obtained for the abundance of the two proteins in fentomoles/10^5 or 10^6 cells were then converted in number of molecules/cell.

antibodies were then used in western blots to analyse the expression of both proteins as well as to estimate their intracellular levels. *Tb*EIF4AI is very abundant (Figure 2B) and although the quantification is only approximate, its levels were estimated at $\sim$2–5 $\times$ 10^5 and 0.8–1.5 $\times$ 10^5 molecules/cell in procyclic and bloodstream forms, respectively, this difference being a reflection of the relative volumes of the two cell types. These levels are compatible with what has been observed with the *L.major* orthologue (33) as well as yeast eIF4A (44). In contrast, *Tb*EIF4AIII levels were estimated at $\sim 2 \times 10^4$ and 1×10^4 molecules/cell in procyclic and bloodstream forms, respectively (Figure 2B). These data indicate that *Tb*EIF4AI is present at levels at least 10-fold higher than *Tb*EIF4AIII, a difference which is reminiscent of the situation with the *L.major* orthologues (33). Since there are estimated to be $\sim$50 000 mRNAs per procyclic cell (Supplementary Data), *Tb*EIF4AI is in excess relative to mRNA, in contrast to *Tb*EIF4AIII. Overall we conclude that both proteins are expressed constitutively and that only the *Tb*EIF4AI levels are compatible with a role in translation.

Subcellular localization of *Tb*EIF4AI and III

To determine the subcellular localization of the *T.brucei* eIF4A homologues we used two different experimental approaches. First the *Tb*EIF4AI and *Tb*EIF4AIII ORFs were cloned into the vector p2215 and the construct integrated into the non-transcribed spacer of a ribosomal RNA gene locus in the procyclic cell line Lister 427 29-13. This resulted in a tetracycline-inducible transgene encoding the eIF4A fused at the C-terminus to EYFP. Expression of both constructs was first verified by western blotting and similar levels of expression were observed for both *Tb*EIF4AI and III-EYFP fusion proteins (data not shown). The fluorescent proteins were visualized by microscopy (Figure 3) and strikingly, the two proteins localize differentially within cells. *Tb*EIF4AI-EYFP is found predominantly in the cytoplasm, whilst *Tb*EIF4AIII-EYFP is only found in the nucleus. These results were confirmed for the endogenous proteins through indirect immunofluorescence using isoform specific antibodies: again *Tb*EIF4AI was mainly found in the cytoplasm whilst *Tb*EIF4AIII was only detected in the nucleus (Figure 3).

RNAi of *Tb*EIF4AI and *Tb*EIF4AIII

The function of the two eIF4A homologues was then investigated by knock down of expression through RNA interference. First, both ORFs were subcloned into the vector p2T7-177 vector (35) and the constructs integrated into the procyclic cell line Lister 427 29-13 resulting in cell lines with tetracycline-inducible expression of double-stranded RNA.

Cell proliferation was reduced within 24 h and ceased around 48 h after induction of *Tb*EIF4AI RNAi (Figure 4A) and the cell density increased by $\sim$3-fold during this time. Western blotting over a time course after addition of tetracycline showed that the level of the protein decreased to <10% of the starting level but expression was not completely ablated (Figure 4B). Protein synthesis after induction of *Tb*EIF4AI RNAi was monitored in two ways: (i) metabolic labelling to identify any alterations in the complement of polypeptides synthesized, and (ii) the rate of total protein synthesis was measured. There were no substantial changes in the profile of proteins synthesized although a small number of polypeptides appear to be relatively less affected by *Tb*EIF4AI depletion. The overall rate of protein synthesis had halved by $\sim$22 h, the time at which cell proliferation ceased and was reduced to <20% of the uninduced control by 48 h as shown in Figure 4C.

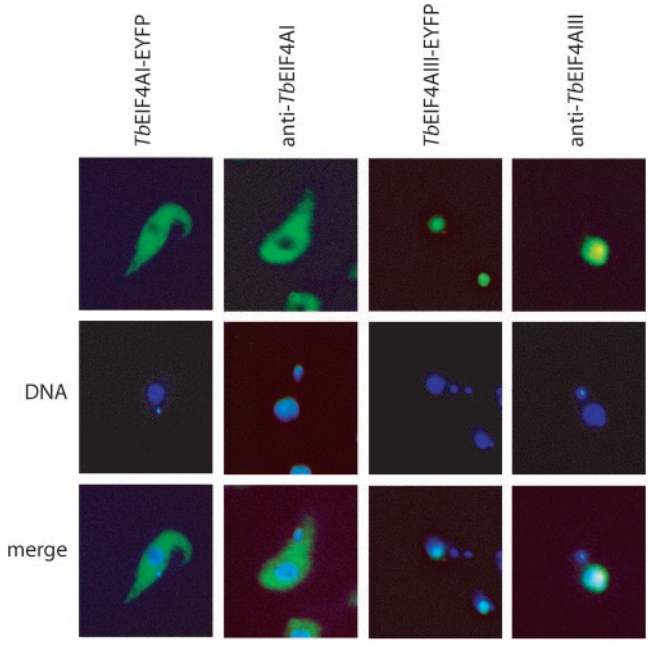
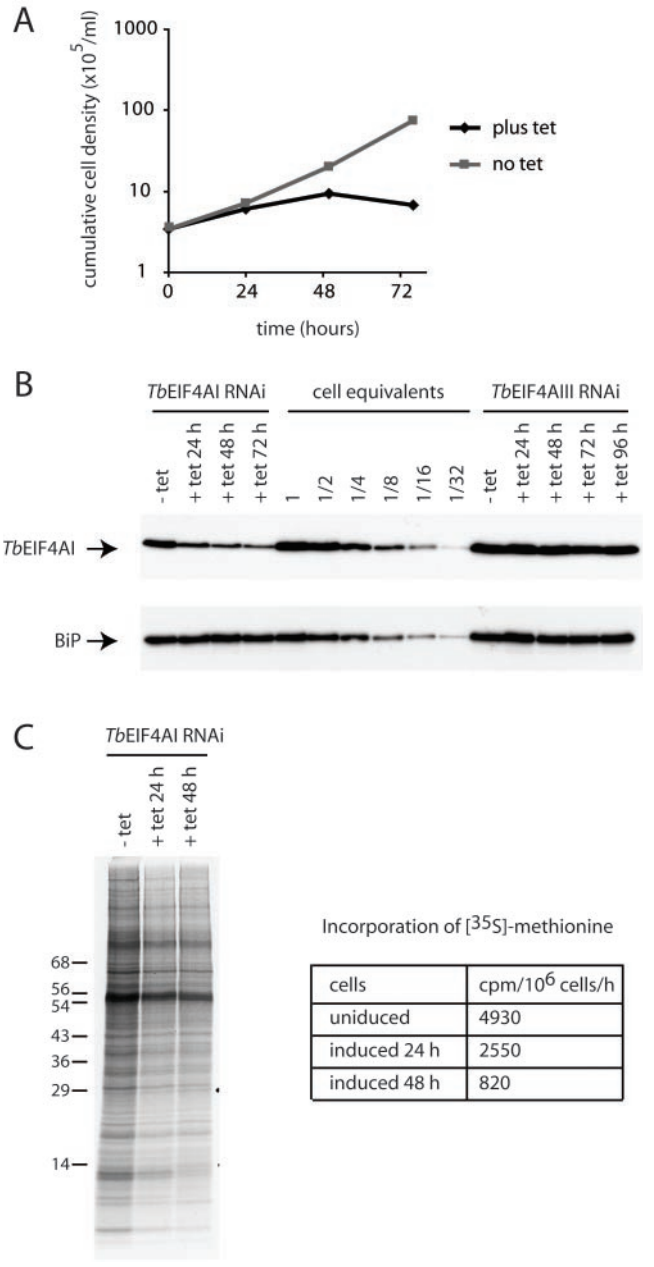


Figure 3. Subcellular localization of *Tb*EIF4AI and III in *T.brucei* procyclic forms. Subcellular localization of the *Tb*EIF4AI and III /EYFP fusion proteins in transfected *T.brucei* cells was examined with a fluorescence microscope. The localization of native *Tb*EIF4AI and III was also confirmed in wild-type procyclic cells (WT 427) by indirect immunofluorescence using the *Tb*EIF4AI or *Tb*EIF4AIII specific antibodies followed by incubation with the fluorescein-conjugated secondary antibody. Where indicated, the cells were counterstained to locate the nuclear and kinetoplast DNA. Note lack of *Tb*EIF4AIII staining of the kinetoplast.

The phenotype of cells after RNAi ablation of *Tb*EIF4AIII was different. These cells only showed a dramatic reduction in the rate of proliferation 3 days after induction of RNAi, during which time the cell density increased ~20-fold (Figure 5A). Levels of *Tb*EIF4AIII fell dramatically during the first 24 h of RNAi (Figure 5B), and the protein was only just detectable in extracts derived from cells at the 48 h time point. It is possible that, owing to its low abundance even in wild-type cells, residuals levels of *Tb*EIF4AIII persist longer than 48 h in the cells after RNAi although they are not detected by the western blotting assay. These residual levels would be responsible for the delayed onset of the growth phenotype. The western blotting results also confirm that lack of *Tb*EIF4AIII is not involved in the phenotype induced by the depletion of *Tb*EIF4AI since no reduction in levels of *Tb*EIF4AIII was observed in cells submitted to the *Tb*EIF4AI RNAi procedure (Figure 5B). Likewise the *Tb*EIF4AIII RNAi does not lead to any reduction in the levels of *Tb*EIF4AI (Figure 4B). These results are compatible with *Tb*EIF4AIII being required only at very low levels so that many cell cycles are required after addition of tetracycline to impair cell growth. In contrast, the levels of *Tb*EIF4AI, despite its abundance in wild-type cells, are much more sensitive to RNAi mediated depletion, consistent with a role in overall protein synthesis.

Expression of dominant negative mutants of *Tb*EIF4AI and III in transfected procyclic cells

The helicase activity of eIF4A is essential for protein synthesis and viability and some mutations that abolish its activity can



cells	cpm/10^6 cells/h
uniduced	4930
induced 24 h	2550
induced 48 h	820

Figure 4. RNAi of *TbEIF4AI*. Procyclic *T.brucei* cells were transfected with the p2T7-177 derived plasmid containing the *TbEIF4AI* gene. Transfected cells were selected after growth in the presence of phleomycin and RNA interference induced after tetracycline addition. At regular intervals, cellular growth was monitored by counting the number of viable cells, expression of *Tb*EIF4AI assayed and total protein synthesis investigated by [^{35}S]methionine incorporation. (**A**) Cell density of transfected cultures with and without tetracycline addition. (**B**) Western blot analysis of the time course. Note the various dilutions of total cell extract for comparison (1–1/32 cell equivalent—1 cell equivalent equals to 10^6 cells and was used in the various RNAi lanes). *Tb*EIF4AI was detected with the affinity purified antisera and anti-BiP was used as a loading control. The same blot was probed with both antibodies. Equivalent extracts of cells transfected with the p2T7-177/*Tb*EIF4AIII construct (see also Figure 5) were also used in the blot to monitor for *Tb*EIF4AI levels. (**C**) [^{35}S]methionine incorporation profile in transfected cells grown without tetracycline or 24 and 48 h after its addition. Total protein synthesis was estimated after RNAi for *Tb*EIF4AI by incubating aliquots of the cells in the presence of [^{35}S]methionine for 1 h followed by TCA precipitation, quantitation of the incorporated radioactivity or SDS–PAGE followed by autoradiography of the selected samples.

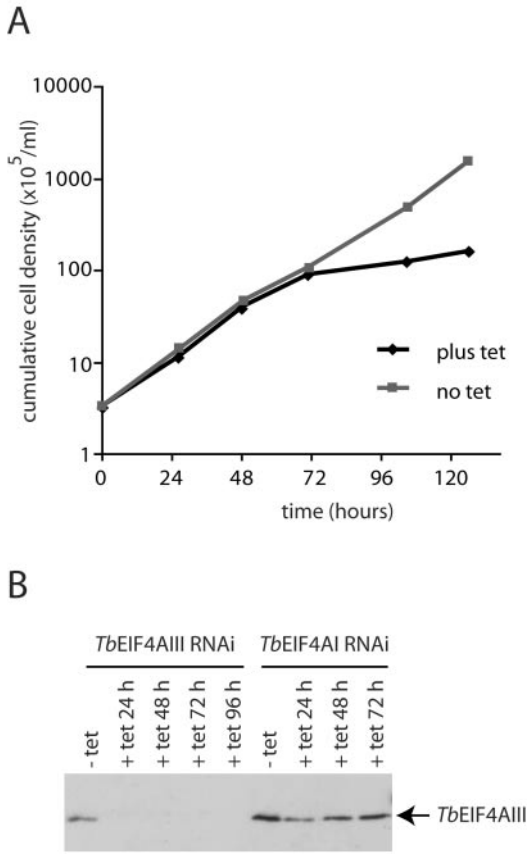


Figure 5. RNAi of *Tb*EIF4AIII. Procyclic *T.brucei* cells were transfected with the p2T7-177/ *Tb*EIF4AIII construct as described for Figure 4, monitored for cellular growth and assayed for expression of *Tb*EIF4AIII. (**A**) Cell density of transfected cultures at different time points with and without tetracycline addition. (**B**) Western blot analysis of the time course for both the *Tb*EIF4AIII and *Tb*EIF4AI RNAi experiments using the *Tb*EIF4AIII antibodies. Samples from the same experiment assayed in Figure 4B were assayed for *Tb*EIF4AIII expression.

act as dominant negative mutants. Wild-type *T.brucei* helicases and equivalent DEAD-box mutant transgenes, in which the glutamic acid residue in the DEAD motif II (Figure 1) was substituted with a glutamine (DEAD→DQAD), DQAD), were expressed using a tetracycline-inducible promoter. This mutation induces a dominant negative phenotype in mammalian eIF4AI, resulting in potent inhibition of protein synthesis and is widely used to abrogate the function of DEAD-box proteins (11,36). The transgenes encoded a C-terminal triple myc tag to distinguish the expression of the transgene from the endogenous protein. Wild-type and mutant versions of the two proteins were then expressed in procyclic Lister 427 29-13 cells and analysed by western blotting (Figure 6). The tetracycline regulation of expression was effective and the wild type and mutant proteins were expressed at similar levels (Figure 6A). The level of expression relative to the endogenous protein varied; the expression from the *Tb*EIF4AI transgenes was lower than expression from the endogenous gene whereas expression from the *Tb*EIF4AIII transgenes was several fold higher than the endogenous protein. All the myc-tagged transgenes localized correctly (data not shown).

Cell growth and transgene expression was monitored over a time course (Figure 6B). Expression from the *Tb*EIF4AI transgene reduced over the time course and was barely detectable by 104 h. We have observed this diminution of expression over time with other, but not all, transgenes expressed from vectors derived from pLEW100 and are unsure of the cause. The only transgene that had any effect on growth was the mutant form of *Tb*EIF4AI (Figure 6C), all others grew at the same rates as the control cultures without tetracycline (data not shown). At 18–51 h after the addition of tetracycline, the expression levels of the *Tb*EIF4AI transgenes were readily detectable and the mutant, but not the wild type, produced a slowing of growth. As the expression of the transgene reduced, the culture returned to the same rate of growth as the no tetracycline control. In contrast, the significantly overexpressed *Tb*EIF4AIII mutant transgene had no effect on growth of the culture. Interestingly, the expression of the *Tb*EIF4AIII transgenes, but not the *Tb*EIF4AI transgenes, resulted in increased levels of the endogenous protein. Overall, the results are compatible with the RNA helicase activity of *Tb*EIF4AI being strictly required for growth. As for *Tb*EIF4AIII, the lack of a slow growth phenotype when the dominant negative mutant is expressed suggests that either its RNA helicase activity is not required for the protein function or it is not affected by the DEAD→DQAD mutation. Either option strongly indicates that *Tb*EIF4AIII is not active in translation.

Mapping of isoform specific amino acids

The results described above for *Tb*EIF4AIII are reminiscent of what is known of mammalian eIF4AIII (Discussion). Human eIF4AIII, identified previously as a negative regulator of translation (22), has been shown to be a component of the EJC, with roles in mRNA export, cytoplasmic RNA localization and NMD (24–27). Pairwise sequence comparisons between *Tb*EIF4AI and III (or their orthologues in *T.cruzi* and *L.major*) and the functionally divergent human eIF4AI/eIF4AIII do not show a clear match between either of the parasite homologues with the two human sequences. Indeed, the overall identity between human eIF4AI/eIF4AIII (66%) is greater than that between either protein and the two trypanosomatid eIF4As (∼55–60%).

The kinetoplastid eIF4AI and eIF4AIII sequences were then aligned with putative eIF4AI and eIF4AIII homologues from the major lines of eukaryotic evolution (Figure 7). The homologues from *Arabidopsis thaliana* and *Schizosaccharomyces pombe* were identified using BLAST searches of non-redundant sequence databases using the human eIF4AI or eIF4AIII sequences as queries. The alignment in Figure 7 does not show any continuous sequence of amino acids that distinguish between all putative eIF4AI or eIF4AIII homologues. However, at various positions, interspersed within the sequences common to both sets of proteins, individual amino acids can be identified which are conserved and unique either to the eIF4AI or eIF4AIII proteins. Table 1 lists 13 positions where a clear difference could be found between the two sets of sequences. Several, but not all, of these amino acid substitutions are also shared by an eIF4AIII-related protein from *Saccharomyces cerevisae*, Fal1p, a nucleolar protein shown to be required for 40S ribosomal subunit formation (45). Fal1p, however, does not seem to be involved in EJC

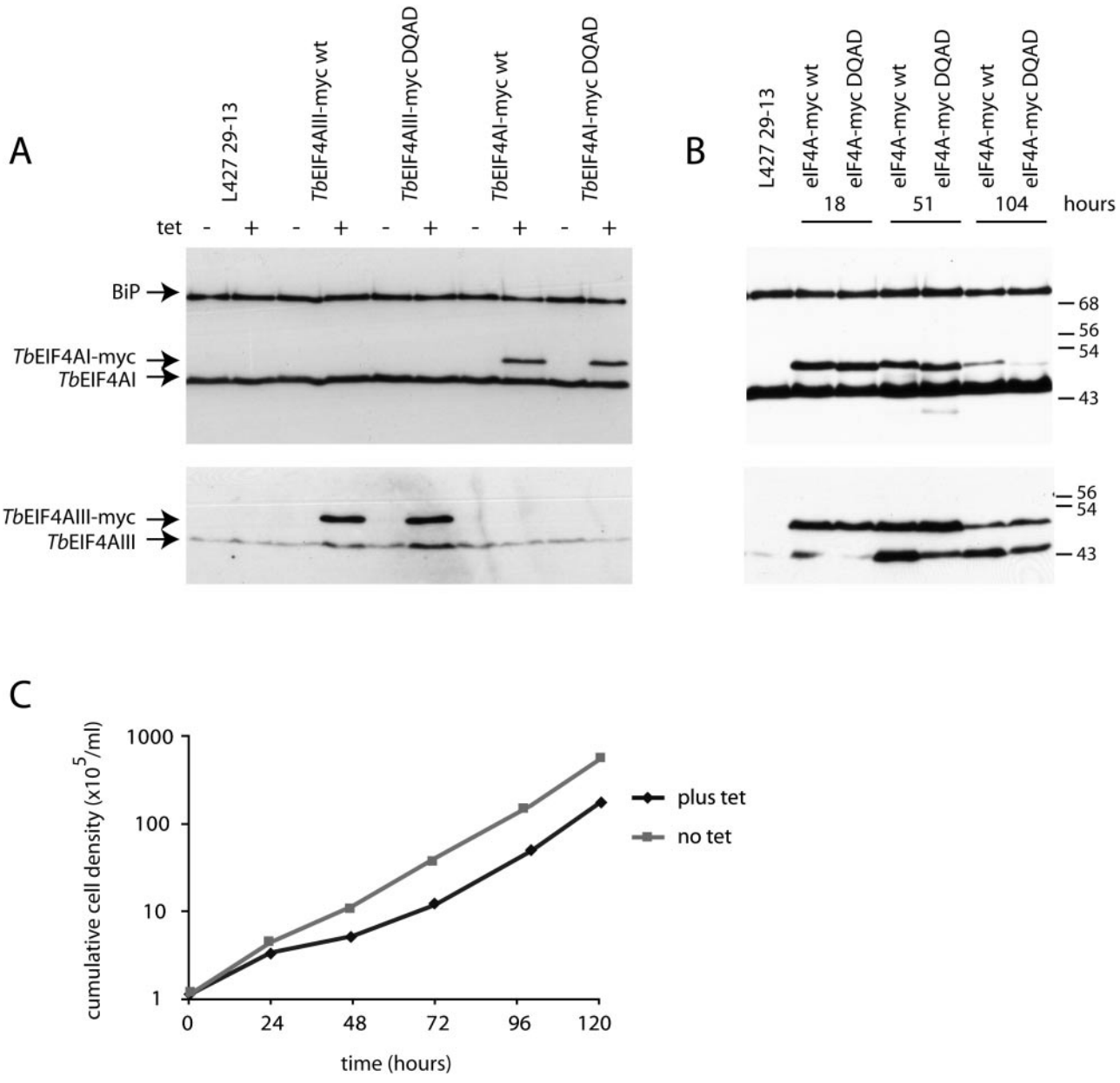


Figure 6. Expression of myc-tagged dominant negative mutants of *Tb*EIF4AI and III in procyclic cells. (**A**) Western blot analysis of the expression of the various *Tb*EIF4AI and III/myc fusions in transfected cells in the absence or after exposure to tetracycline for 18 h. In each case the expression was detected using antibodies specific to each of the eIF4A homologues. The *Tb*EIF4AI western blot was simultaneously probed with anti-BiP as a loading control. (**B**) Time course expression of the different versions of *Tb*EIF4A-*myc* after tetracycline addition to the culture. The *Tb*EIF4AI western blot was simultaneously probed with anti-BiP as a loading control. (**C**) Effect of the expression of the dominant negative form of *Tb*EIF4AI-myc on the growth of the transfected cells in culture.

formation since a search in *S.cerevisae* for similar EJC constituents, conserved in other fungi and in plants, such as Magoh or Y14, did not produce any clear homologues.

The various amino acid substitutions listed in Table 1 (*Tb*EIF4A I numbering), indicated by a star in the alignment in Figure 7, discriminate between all putative eIF4AI and eIF4AIII homologues compared, including the two trypanosomatid proteins. These substitutions are located in the two globular domains present in eIF4A and related DEAD-box helicases (11,14). Both the N- and the C-terminal domains have been shown to participate in the binding to RNA and ATP required for the helicase/ATPase activities, but few roles have been postulated for them regarding specific protein functions. The alignment results clearly show that candidate eIF4AIII homologues are present throughout the various eukaryotic lineages, although it has only been functionally characterized in metazoans. The unique substitutions are also indicative of amino acids involved in specific aspects of eIF4AI/III function in general (see below).

Molecular modelling of *Tb*EIF4AI and *Tb*EIF4AIII

To understand the functional implications of the observed amino acids substitutions to eIF4A function, not only in

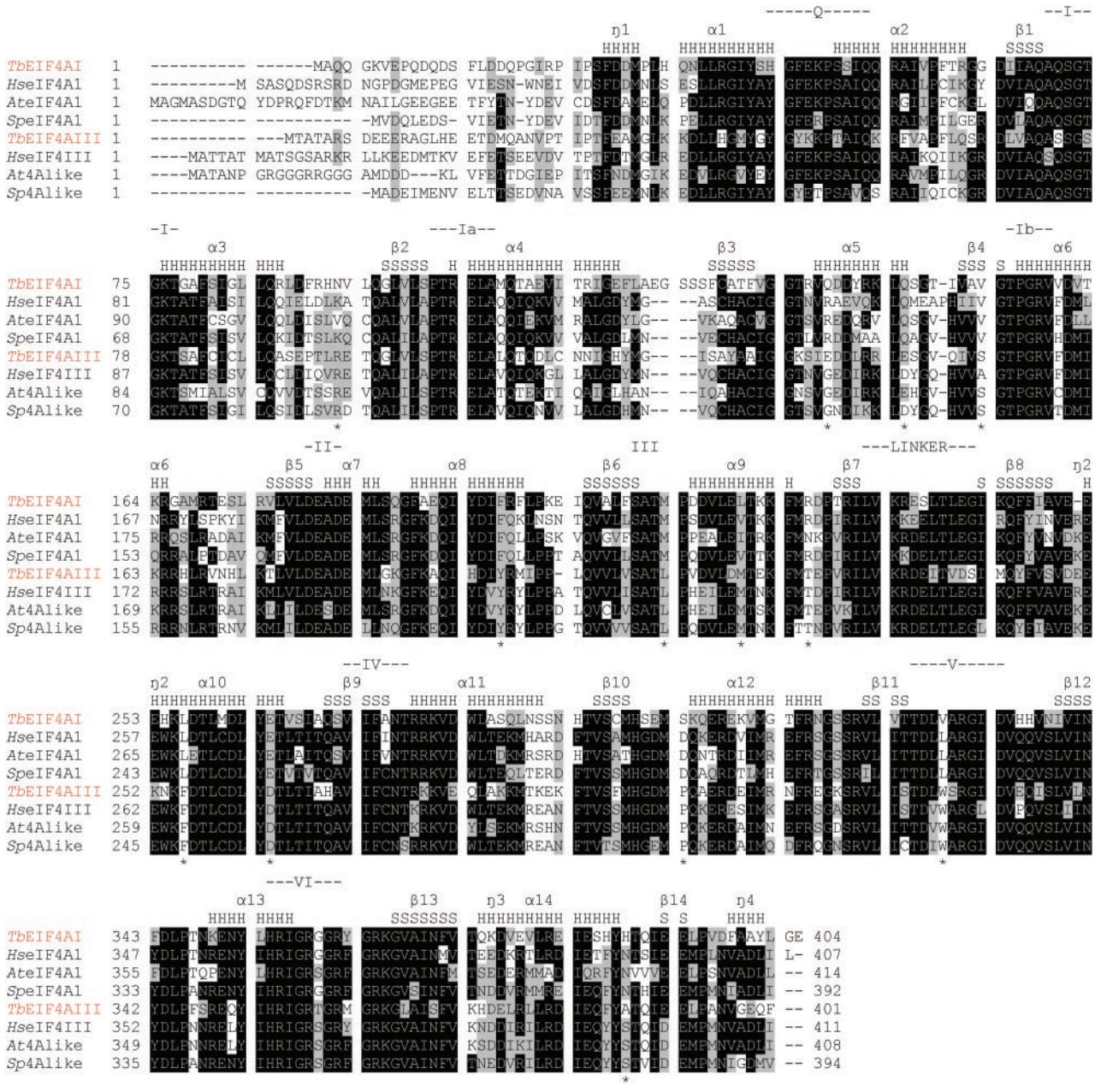


Figure 7. Sequence alignment comparing *Tb*EIF4AI and III with the putative eIF4AI and eIF4AIII from selected organisms. (**A**) Sequences were aligned as described in Figure 1 and the various DEAD-box motifs are shown as indicated previously. The predicted secondary structural elements derived from the modelling shown in Figure 8 and from Ref. (46) are indicated numbered α1–α13/η1–η4 (alpha-helices—H) and β1–β14 (beta-strands—S). Asterisk indicates amino acids which distinguish between the eIF4AI and eIF4AIII homologues. Further relevant GenBank accession numbers: human (*Hs*) eIF4AI, AAX43035; human eIF4AIII (*Hs*eIF4A3), P38919; *S.pombe* (*Sp*) eIF4A1, CAA56772; *S.pombe* eIF4A-like protein (*Sp*4Alike), CAA92238; *A.thaliana* (*At*) eIF4A1, NP_177417; *A.thaliana* eIF4A-like protein (*At*4Alike), NP_188610.

trypanosomatids but also in eukaryotes in general, we modeled the structures of both *Tb*EIF4AI and III based on the solved structure of either the yeast *S.cerevisae* eIF4A (12) or the related DEAD-box protein Dhh1p (46). The structure of yeast eIF4A is in an open conformation with the two globular domains positioned apart and non-interacting. In contrast, Dhh1p is in a closed conformation with the two domains facing each other. Most of the conserved motifs in Dhh1p are positioned in close spatial proximity facing the cleft between the two domains. Both sets of models were validated as described (33) and found to have self-consistency in terms of sequence–structure compatibility and to be of good overall quality. For our analysis we favored the closed conformation structure since the two domains need to interact in order to fully form the ATP- and RNA-binding sites (12,46,47).

Figure 8A shows the ribbon drawing for the predicted structures of *Tb*EIF4AI and *Tb*EIF4AIII. Highlighted in the figure are several of the diagnostic amino acid substitutions identified in the eIF4A alignment (Figure 7). Of special interest is the V/L328W substitution in motif V, in the C-terminal domain.

Table 1. Summary of the amino acid substitutions identified between the putative eIF4AI/eIF4AIII homologues from the main lineages of eukaryotic evolution

Position: T*b*EIFAI	Substitution: 4AI > 4AIII	Secondary structure	Domain	Overall position in predicted tertiary structure
93	K/N/V > R	Loop	N-terminal	Near α5/exposed
139	Q/R > E/G	α 5	N-terminal	Exposed
146	Q/A > E/D	α 5	N-terminal	Exposed
153	V > S/A[a]	β4	N-terminal	Buried
197	F > Y	α8	N-terminal	Partially exposed
213	M > L	Loop	N-terminal	Interface/next to Motif III/partially buried
220	L/V/I > M	α9	N-terminal	Buried
226	R/N > T	Loop	C-terminal	Exposed
256	L > F[a]	α 10	C-terminal	Partially buried
264	E > D[a]	α 10	C-terminal	Exposed
303	S/D > P[a]	Loop	C-terminal	Exposed
328	V/L > W[a]	Loop	C-terminal	Interface/Motif V/Buried
388	H/N > A/S[a]	Loop	C-terminal	Exposed

[a]These substitutions but not the others are present in the *S.cerevisae* nucleolar protein Fal1p [may be related to the eIF4AIII proteins (45)].

Motif V lies in a loop positioned in the interface between the two domains and, in Dhh1p, several amino acids in this motif are seen to make direct interactions with specific amino acids in motifs I and Q, positioned in the N-terminal domain (46). In the models shown here both the V and W residues in *Tb*EIF4AI and III, respectively, are protruding from the main polypeptide backbone in the direction of a cleft in the proteins' N-terminal domain. To investigate the likelihood of either amino acid interacting with neighboring chains, atoms in these chains were first identified which are positioned within a radius of 4 Å from the two residues. These are the only ones capable of forming non-covalent interactions to atoms in either amino acid and the full set of potential interactions are shown in Figure 8B as dotted lines. The substantially larger W residue in *Tb*EIF4AIII is capable of making a number of interactions with neighboring amino acids in both the N- and C-terminal domains, as well as with the polypeptide backbone. In contrast, the V residue in *Tb*EIF4AI is very limited in the number of interactions it can establish. It is possible then that the presence of the W residue in *Tb*EIF4AIII, and other eIF4AIII homologues, can enhance the interaction between the helicase's two domains as compared to the V/L residue in the eIF4AI and even Dhh1p proteins.

Other potentially interesting eIF4AIII-like substitutions map in helices 5 (Q/R139E/G, Q/A146D/E) and 10 (L256F, E264D), on the N- and C-terminal domains, respectively. These helices are largely exposed to the solvent on the external side of the proteins (Table 1) and thus the amino acids involved could mediate eIF4A binding to functional partners. Indeed, recent evidence strongly supports such a hypothesis. First, the binding surface for eIF4GII has been mapped to the C-terminal domain of eIF4AI (47). A double mutation in human eIF4AI which prevents binding to eIF4GII maps to helix 10 and targets the same glutamate residue (E264) found to be unique to the eIF4AI sequences. Likewise, the two substitutions in helix 5 are included within a proposed eIF4AIII specific motif (motif C) which has just been found to constitute part of the binding site for the EJC component MNL51 (30). Moreover, a further unique substitution identified in Figure 7, which lies in an exposed loop near the C-terminal end of the eIF4A proteins (H/N388A/S—also shown in Figure 8), lies within another proposed motif (motif H) found to be required for eIF4AIII to bind spliced mRNA and to rescue NMD in eIF4AIII depleted cells (30). In summary, we have identified several individual amino acids conserved in either eIF4AI or eIF4AIII sequences which may play significant roles in these proteins' functions not only in trypanosomatids but also in eukaryotes in general.

DISCUSSION

The results presented here provide strong support that only one of the two eIF4A homologues identified in trypanosomatids is involved in the initiation of translation. The abundance of the *Tb*EIF4AI protein, its constitutive expression during the parasite life cycle as well as its cytoplasmic localization, the effect of RNAi depletion and the dominant negative phenotype of the DEAD→DQAD mutation are all compatible with what is expected of this protein. In contrast, *Tb*EIF4AIII does not seem to play an obvious role in protein synthesis. The nuclear localization of *Tb*EIF4AIII, its low abundance, longer response to the RNAi induced phenotype and lack of inhibition by the dominant negative mutant all indicate an essential role in RNA metabolism in the nucleus unrelated to eIF4A function in translation. These results are also compatible with what is known of the *L.major* orthologues; *Lm*EIF4AI binds strongly to at least two eIF4G homologues whereas *Lm*EIF4AIII has a reduced binding activity [(33) and C. R. S. Reis, unpublished data]. Mammalian eIF4AIII localizes mainly to the nucleus (23), is present in levels ∼10-fold lower than eIF4AI in HeLa cells and does not function in protein synthesis (22). An unusual feature of human eIF4AIII is that the DEAD→DQAD DQAD mutation has no effect on its activity in EJC formation and NMD (30). Thus, *Tb*EIF4AIII behaves similarly to human eIF4AIII in several important aspects and, coupled with the sequence analysis data, our results are consistent with it being an eIF4AIII orthologue with functions possibly conserved along most major lines of eukaryotic organisms.

As part of the EJC, eIF4AIII binds directly to the core proteins Magoh, Y14 and MLN51 (25–30) and also to other proteins required for EJC function such as the TAP and Aly/REF proteins involved in nuclear mRNA export (25). Magoh homologues have been clearly identified in the three trypanosomatid genomes finished to date, *T.brucei* (GenBank, AAZ12053), *T.cruzi* (EAN97132) and *L.major* (CAJ06870) and possible TAP homologues can also be found. We have also tried to identify candidate Y14, MLN51 or Aly/REF

homologues but so far without success. However, both Y14 and Aly/REF are small RNA-binding proteins with single RRMs, a category which includes many proteins with unassigned functions in those three genomes (48). It may be possible that, due to the degree of evolutionary distance between trypanosomatids and animals, homologues to these two proteins cannot be clearly identified by sequence analysis alone. As for MLN51 it is poorly conserved outside the metazoans so it is unlikely also for homologues to be identified in trypanosomatids only by sequence analysis. Nevertheless the strong conservation of the Magoh sequences between the human and parasite homologues (over 50% identity) is an indication that the EJC may be present throughout the major groups of eukaryotes and that eIF4AIII-like proteins may be active within this complex.

In a very recent study eight eIF4AIII specific motifs (named A to H) were identified in an alignment comparing various eIF4AIII homologues with the human eIF4AI and II proteins. Selected amino acids in some of these motifs, as well as in the canonical eIF4A motifs I, Ia and VI, were then mutated in recombinant or *in vivo* overexpressed eIF4AIII to investigate their requirement for eIF4AIII function (30). In the alignment provided here, which includes both *T.brucei* eIF4A homologues, as well as eIF4AI sequences from divergent organisms, no continuous set of amino acids were found to be typical of either eIF4AI or eIF4AIII proteins. However, unique amino acid substitutions were identified which distinguish eIF4AIII-like proteins from eIF4AI homologues in all sequences investigated so far. Several of these substitutions not only coincide with some of the proposed eIF4AIII specific motifs (motifs C, E and H), but also are included in two of those motifs found to be involved in specific eIF4AIII functions such as binding to the EJC partner MLN51 (motif C) and requirements for binding to spliced mRNA and for NMD (motif H) (30). However some of the unique eIF4AI/eIF4AIII substitutions identified here do not coincide with the remaining proposed motifs. These might be involved in mediating other aspects of eIF4A function and should be considered as targets for further investigation.

Very few protein coding genes in trypanosomatids contains a *cis*-intron (49,50). However, every cytoplasmic mRNA is *trans*-spliced to form the mature 5′ end of the mRNA and this splice site is possibly the location of EJC binding. The function of the EJC in these organisms remains obscure specially considering that the splice site is always to the 5′ side of the ORF. The EJC-mediated mechanism of NMD seems to be restricted to mammalian cells [reviewed in (51,52)] and indeed there is strong evidence that NMD does not occur in trypanosomatids (38). In mammals, both the EJC and the nuclear cap-binding complex (CBC, composed of two subunits CBP20 and CBP80)

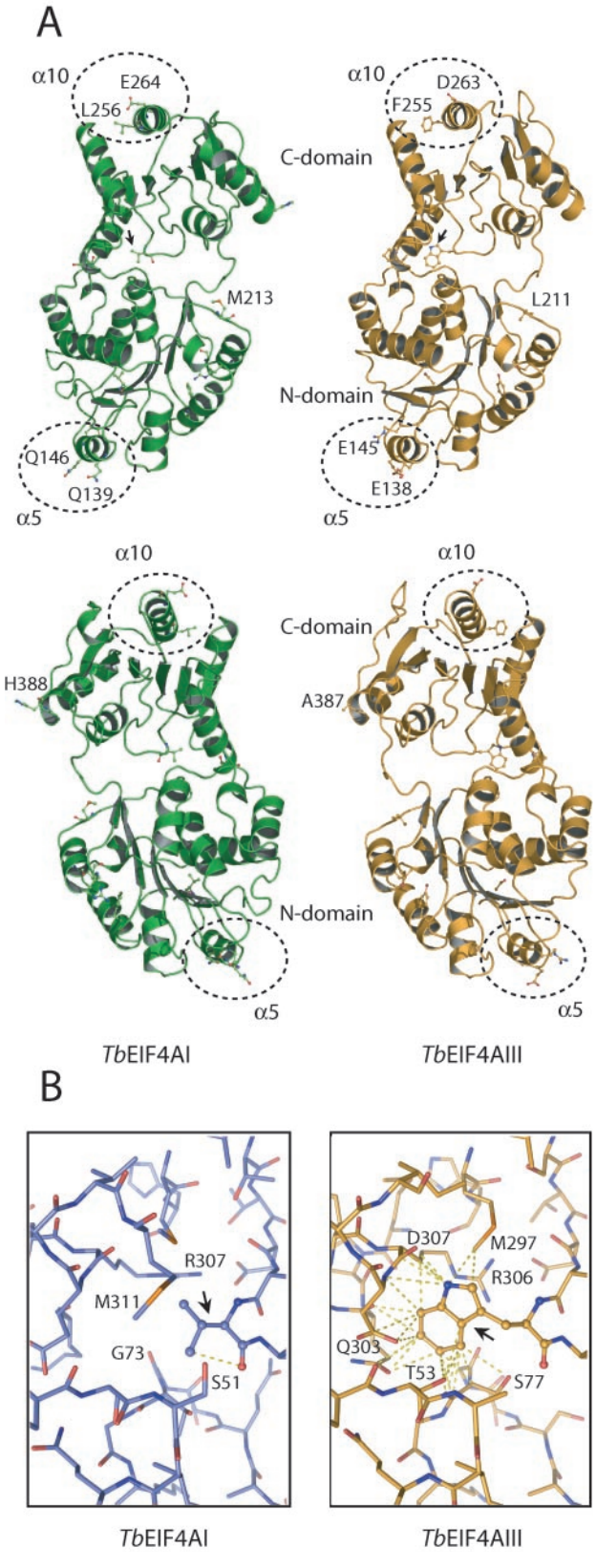


Figure 8. Molecular modelling of *Tb*EIF4AI and III highlighting the position of amino acids unique to the eIF4AI or eIF4AIII homologues. Diagrams were created with the program PyMol (http://www.pymol.org). (**A**) Ribbon diagrams of the overall structure of both *Tb*EIF4AI and III viewed as in (46) (upper panel) or rotated 180° about its long axis (lower panel). The structure is in a closed conformation where the two, N- and C-terminal, domains are facing each other. The arrows indicate the position of the L328W substitution which lies in the loop containing Motif V and is positioned in the interface between the two domains. The dotted circles delimit the two helices discussed in the text, α5 and α10. The H/N388A/S and M213L substitutions are also indicated (their numbering differ however from the eIF4AI/eIF4AIII sequences—for instance, H388 in *Tb*EIF4AI is equivalent to A387 in *Tb*EIF4AIII and so on). (**B**) Balls and sticks representation showing the neighbourhood of the L328W substitution in both *Tb*EIF4AI and III. The dotted lines indicate the atoms in the neighbouring amino acid chains which are positioned within a radius of 4 Å from the atoms in either the L or W residues. In both (A and B), the relevant amino acids are listed.

bind to precursor mRNAs in the nucleus, prior to or during the splicing event, and remain bound to the mRNAs until they are transported to the cytoplasm and/or translated for the first time (53,54). In *T.brucei* a novel CBC has been described which consists of a CBP20 subunit (also present in yeast and humans) plus four other polypeptides, one of which is importin-α (known to associate with CBC in other eukaryotes) and three novel proteins only present in trypanosomatids. The parasite CBC has been implicated in the early steps of mRNA maturation, prior to the *trans*-splicing event whereas the polycistronic precursor mRNA is cleaved into mature monocistronic units (55). At this stage it still remains to be determined whether *Tb*EIF4AIII and other components of the putative EJC are also necessary for mRNA processing, export from the nucleus or even translation in trypanosomatids.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank M. Moore for sharing unpublished data and J. Bangs for the gift of the BiP antibody. O. Thiemann and D. Souza allowed us access to the computing facilities necessary for the modelling work. The work in Cambridge was funded by the Wellcome Trust and by a Nuffield Foundation summer studentship to N. Marinsek. The Brazilian scientists were supported by grants/studentships from CNPq and CAPES. Funding for the laboratory in Recife came in part from a grant from FACEPE/CNPq/CT INFRA.

Conflict of interest statement. None declared.

REFERENCES

1. Campbell,D.A., Thomas,S. and Sturm,N.R. (2003) Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect*., **5**, 1231–1240.
2. Liang,X.H., Haritan,A., Uliel,S. and Michaeli,S. (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell*, **2**, 830–840.
3. Perry,K.L., Watkins,K.P. and Agabian,N. (1987) Trypanosome mRNAs have unusual 'cap 4' structures acquired by addition of a spliced leader. *Proc. Natl Acad. Sci. USA*, **84**, 8190–8194.
4. Hershey,J.W.B. and Merrick,W.C. (2000) Pathway and mechanism of initiation of protein synthesis. In Sonenberg,N., Hershey,J.W.B. and Mathews,M.B. (eds), *Translational Control Of Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 33–88.
5. Pestova,T.V., Kolupaeva,V.G., Lomakin,I.B., Pilipenko,E.V., Shatsky,I.N., Agol,V.I. and Hellen,C.U. (2001) Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl Acad. Sci. USA*, **98**, 7029–7036.
6. Preiss,T. and Hentze,M.W. (2003) Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays*, **25**, 1201–1211.
7. Sonenberg,N. and Dever,T.E. (2003) Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol*., **13**, 56–63.
8. Gingras,A.C., Raught,B. and Sonenberg,N. (1999) eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem*., **68**, 913–963.
9. Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases-amino-acid sequence comparisons and structure–function relationships. *Curr. Opin. Struct. Biol*., **3**, 419–429.
10. Linder,P. (2003) Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol. Cell*, **95**, 157–167.
11. Rocak,S. and Linder,P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell Biol*., **5**, 232–241.
12. Caruthers,J.M., Johnson,E.R. and McKay,D.B. (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
13. Story,R.M., Li,H. and Abelson,J.N. (2001) Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
14. Caruthers,J.M. and McKay,D.B. (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol*., **12**, 123–133.
15. Marcotrigiano,J., Lomakin,I.B., Sonenberg,N., Pestova,T.V., Hellen,C.U. and Burley,S.K. (2001) A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol. Cell*, **7**, 193–203.
16. Imataka,H. and Sonenberg,N. (1997) Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol. Cell. Biol*., **17**, 6940–6947.
17. Yang,H.S., Cho,M.H., Zakowicz,H., Hegamyer,G., Sonenberg,N. and Colburn,N.H. (2004) A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A. *Mol. Cell. Biol*., **24**, 3894–3906.
18. Svitkin,Y.V., Pause,A., Haghighat,A., Pyronnet,S., Witherell,G., Belsham,G.J. and Sonenberg,N. (2001) The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. *RNA*, **7**, 382–394.
19. Pestova,T.V. and Kolupaeva,V.G. (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev*., **16**, 2906–2922.
20. Nielsen,P.J. and Trachsel,H. (1988) The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed. *EMBO J*., **7**, 2097–2105.
21. Yoder-Hill,J., Pause,A., Sonenberg,N. and Merrick,W.C. (1993) The p46 subunit of eukaryotic initiation factor (eIF)-4F exchanges with eIF-4A. *J. Biol. Chem*., **268**, 5566–5573.
22. Li,Q., Imataka,H., Morino,S., Rogers,G.W.,Jr, Richter-Cook,N.J., Merrick,W.C. and Sonenberg,N. (1999) Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell. Biol*., **19**, 7336–7346.
23. Holzmann,K., Gerner,C., Poltl,A., Schafer,R., Obrist,P., Ensinger,C., Grimm,R. and Sauermann,G. (2000) A human common nuclear matrix protein homologous to eukaryotic translation initiation factor 4A. *Biochem. Biophys. Res. Commun*., **267**, 339–344.
24. Ferraiuolo,M.A., Lee,C.S., Ler,L.W., Hsu,J.L., Costa-Mattioli,M., Luo,M.J., Reed,R. and Sonenberg,N. (2004) A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc. Natl Acad. Sci. USA*, **101**, 4118–4123.
25. Chan,C.C., Dostie,J., Diem,M.D., Feng,W., Mann,M., Rappsilber,J. and Dreyfuss,G. (2004) eIF4A3 is a novel component of the exon junction complex. *RNA*, **10**, 200–209.
26. Palacios,I.M., Gatfield,D., St,J.D. and Izaurralde,E. (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*, **427**, 753–757.
27. Shibuya,T., Tange,T.O., Sonenberg,N. and Moore,M.J. (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nature Struct. Mol. Biol*., **11**, 346–351.
28. Ballut,L., Marchadier,B., Baguet,A., Tomasetto,C., Seraphin,B. and Le,H.H. (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nature Struct. Mol. Biol*., **12**, 861–869.
29. Tange,T.O., Shibuya,T., Jurica,M.S. and Moore,M.J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA*, **11**, 1869–1883.
30. Shibuya,T., Tange,T.O., Stroupe,M.E. and Moore,M.J. (2006) Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA*, **12**, 360–374.
31. Skeiky,Y.A., Guderian,J.A., Benson,D.R., Bacelar,O., Carvalho,E.M., Kubin,M., Badaro,R., Trinchieri,G. and Reed,S.G. (1995) A recombinant *Leishmania* antigen that stimulates human peripheral blood mononuclear cells to express a Th1-type cytokine profile and to produce interleukin 12. *J. Exp. Med*., **181**, 1527–1537.

32. Skeiky,Y.A., Kennedy,M., Kaufman,D., Borges,M.M., Guderian,J.A., Scholler,J.K., Ovendale,P.J., Picha,K.S., Morrissey,P.J., Grabstein,K.H. *et al*. (1998) LeIF: a recombinant *Leishmania* protein that induces an IL-12-mediated Th1 cytokine profile. *J. Immunol*., **161**, 6171–6179.
33. Dhalia,R., Reis,C.R., Freire,E.R., Rocha,P.O., Katz,R., Muniz,J.R., Standart,N. and de Melo Neto,O.P. (2005) Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. *Mol. Biochem. Parasitol*., **140**, 23–41.
34. Wirtz,E., Leal,S., Ochatt,C. and Cross,G.A. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol*., **99**, 89–101.
35. Wickstead,B., Ersfeld,K. and Gull,K. (2002) Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Mol. Biochem. Parasitol*., **125**, 211–216.
36. Pause,A., Methot,N., Svitkin,Y., Merrick,W.C. and Sonenberg,N. (1994) Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. *EMBO J*., **13**, 1205–1215.
37. Hirumi,H. and Hirumi,K. (1989) Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol*., **75**, 985–989.
38. Webb,H., Burns,R., Ellis,L., Kimblin,N. and Carrington,M. (2005) Developmentally regulated instability of the GPI-PLC mRNA is dependent on a short-lived protein factor. *Nucleic Acids Res*., **33**, 1503–1512.
39. Carrington,M., Roditi,I. and Williams,R.O. (1987) The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*. *Nucleic Acids Res*., **15**, 10179–10198.
40. Minshall,N., Thom,G. and Standart,N. (2001) A conserved role of a DEAD box helicase in mRNA masking. *RNA*, **7**, 1728–1742.
41. Cordin,O., Tanner,N.K., Doere,M., Linder,P. and Banroques,J. (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J*., **23**, 2478–2487.
42. Tanner,N.K., Cordin,O., Banroques,J., Doere,M. and Linder,P. (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell*, **11**, 127–138.
43. Roditi,I., Carrington,M. and Turner,M. (1987) Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature*, **325**, 272–274.
44. von der Haar,T. and McCarthy,J.E. (2002) Intracellular translation initiation factor levels in *Saccharomyces cerevisiae* and their role in cap-complex function. *Mol. Microbiol*., **46**, 531–544.
45. Kressler,D., de la,C.J., Rojo,M. and Linder,P. (1997) Fal1p is an essential DEAD-box protein involved in 40S-ribosomal-subunit biogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol*., **17**, 7283–7294.
46. Cheng,Z., Coller,J., Parker,R. and Song,H. (2005) Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA*, **11**, 1258–1270.
47. Oberer,M., Marintchev,A. and Wagner,G. (2005) Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev*., **19**, 2212–2223.
48. De Gaudenzi,J., Frasch,A.C. and Clayton,C. (2005) RNA-binding domain proteins in kinetoplastids: a comparative analysis. *Eukaryot. Cell*, **4**, 2106–2114.
49. Mair,G., Shi,H., Li,H., Djikeng,A., Aviles,H.O., Bishop,J.R., Falcone,F.H., Gavrilescu,C., Montgomery,J.L., Santori,M.I. *et al*. (2000) A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. *RNA*, **6**, 163–169.
50. Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R. *et al*. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436–442.
51. Conti,E. and Izaurralde,E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol*., **17**, 316–325.
52. Lejeune,F. and Maquat,L.E. (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol*., **17**, 309–315.
53. Lejeune,F., Ishigaki,Y., Li,X. and Maquat,L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J*., **21**, 3536–3545.
54. Ishigaki,Y., Li,X., Serin,G. and Maquat,L.E. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell*, **106**, 607–617.
55. Li,H. and Tschudi,C. (2005) Novel and essential subunits in the 300-kilodalton nuclear cap binding complex of *Trypanosoma brucei*. *Mol. Cell Biol*., **25**, 2216–2226.