

João Renato Carvalho Muniz

Estrutura Tridimensional da Bothropasina,  
uma Metaloprotease/Desintegrina do Veneno  
de *Bothrops jararaca*.

São Carlos – SP  
Agosto / 2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



João Renato Carvalho Muniz

Estrutura Tridimensional da Bothropasina,  
uma Metaloprotease/Desintegrina do Veneno  
de *Bothrops jararaca*.

Tese apresentada ao programa de pós-  
graduação do Instituto de Física de  
São Carlos, da Universidade de São  
Paulo, para obtenção do título de  
Doutor em Física Aplicada – área de  
concentração: Física Biomolecular

Orientadora: Profa. Dra. Dulce Helena Ferreira de Souza.

DOUTORADO EM CIÊNCIA – FÍSICA APLICADA  
INSTITUTO DE FÍSICA DE SÃO CARLOS  
UNIVERSIDADE DE SÃO PAULO

São Carlos – SP

Agosto / 2007

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

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Muniz, João Renato Carvalho  
Estrutura tridimensional da Bothropasina, uma metaloprotease/desintegrina do veneno de *Bothrops jararaca*. João Renato Carvalho Muniz; orientadora Dulce Helena Ferreira de Souza. São Carlos, 2007.  
153 f.

Tese (Doutorado - Área de concentração: Física Aplicada - Opção: Física Biomolecular) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Metaloprotease. 2. Desintegrina. 3. Cristalografia de proteínas. 4. Bothropasina. 5. Veneno de serpentes. I. Título.

## DEDICATÓRIA

*Ao meu pai, Wilson Renato Muniz. À minha mãe,  
Maria Lúcia P. C. Muniz. À minha tia, Maria  
Eugênia P. Carvalho. Às minhas irmãs,  
Marília e Luciana.*

## **AGRADECIMENTOS**

À Profa. Dra Dulce Helena Ferreira de Souza, pela orientação, compreensão, amizade, ricos ensinamentos, apoio, ajuda, paciência, liberdade de trabalho, respeito... Obrigado pela oportunidade que me foi dada e também por confiar esse projeto a mim.

Aos professores Richard C. Garratt, Glaucius Oliva, Otavio H. Thiemann, Adriano A., Eduardo Castellano, Igor Polikarpov e Heloísa Araújo (UFSCar) pelas discussões, dicas, apoio, incentivos imprescindíveis para grande parte das minhas realizações. Além da excelente convivência.

A TODOS os colegas e amigos (todos mesmo!) do grupo de cristalografia e de biofísica do IFSC, pelo companheirismo, atenção e sempre prestativos para boas conversas e ensinamentos, garantindo assim, um convívio alegre e hospitaleiro em todos os congressos, workshops, viagens, cursos, festas e churrascos. Grandes amigos e colegas que levarei para sempre...

Aos sempre companheiros “das antigas”: Wânus, Caqui, Castilho, Ney, Alécio, Diet, PH, Fredi, Artur, Sandrinha, Sandra K., Stefano, Bianca, Fábio, Portuga, Márcio Silva, Hamilton, Humberto, Susana, Celina, Marião, Marcão, Peninha, Zé Ricardo, Alexandre BB, Ricardo A., Adriana, Sauli, Elisandra, ...

Aos companheiros nem tão “antigos”: Amanda, Nico, Nádia, Napo, Alexandre C, Carols (F, C e H), Alessandro, Lucas B, Dani T., Dani C., Ana Letícia, Lívias (S. e M.), Renata, Maria, Tavinho, Sonoda, Tiagão, Mário, Leo, Alexander, Fernando B., Mamé, Jaque, Aline, Nathalia, Aderson, ...

Ao sempre prestativo amigo Andre Ambrósio, por toda ajuda, dicas, amizade e mão na massa!!!

A TODOS os meus colegas e amigos “EXTRA” cristalografia.

Aos funcionários do Instituto de Física de São Carlos. Em especial, gostaria de agradecer à Cristiane Gomes Lazarini Estella e Wladerez Aparecida Gounella Caiado

do Serviço de Pós-Graduação do Instituto de Física de São Carlos, por toda a atenção, coleguismo, apoio e incentivo. Gostaria de agradecer também à Samira Napolitano e Beth. Todas sempre prestativas e MUITO pacientes comigo!

Ao funcionário do CBME Fernando Falvo pelo coleguismo, por toda ajuda na compra de materiais de consumo, auxílios financeiros para viagens, posteres, boa vontade em ajudar e toda a paciência que demonstrar.

Gostaria muito de agradecer ao Prof. Sir Tom Blundell, por ter me acolhido em seu Departamento na Universidade de Cambridge durante inesquecíveis 7 meses. Ao Dr. Rinaldo Wander Montalvão e sua esposa Elisângela por todo apoio, ajuda, coleguismo, companheirismo e dicas de sobrevivência tanto em Cambridge, como na Inglaterra em geral ;). Aos colegas Vassily Bavro, Tammy, Tammy C, Dr. Dima Chirgadze, Dr. Nicholas Furnham, Dra. Eva Maria Priego, Dave, Wataro, Swannand, Vitor (e sua esposa) e Cris pelo companheirismo na hora de tomar mais um pint! (nunca half pint!).

Ao Prof. Randy Read e Dra. Airle McCoy (ambos do *MRC* – Cambridge), pelos ensinamentos e dicas imprescindíveis para o faseamento e construção do modelo cristalográfico da bothropasina.

À CAPES por todo o suporte que me foi dado para a realização do meu doutorado e também no Programa de Desenvolvimento de Estágio no Exterior (PDEE).

A todas as pessoas que de maneira direta ou indireta contribuíram para o andamento desse trabalho.

E à você que está lendo essa tese(ou pelo menos os agradecimentos!).

Meu mais sincero **MUITO OBRIGADO!!!**

Me desculpe se esqueci seu nome... juro que não foi intencional!

# ÍNDICE

DEDICATÓRIA.....	ii
AGRADECIMENTOS.....	iii
LISTA DE FIGURAS.....	vii
LISTA DE TABELAS.....	vii
RESUMO.....	viii
<i>ABSTRACT</i> .....	ix
ABREVIACÕES.....	x
<b>CAPÍTULO 1 – Introdução.....</b>	<b>1</b>
1.1 – A serpente Bothrops jararaca.....	1
1.2 – Envenenamento e quadro clínico.....	2
1.3 – Integrinas e desintegrinas.....	3
1.4 – Metaloproteases.....	7
1.5 – Proteínas humanas homólogas as SVMPs P-III.....	11
1.6 – Jararagina e Bothropasina.....	13
1.6.1 – Os domínios estruturais das moléculas de jararagina e bothropasina.....	14
1.6.2 – Atividades biológicas da jararagina.....	16
Jararagina e hemorragia.....	16
Efeitos da jararagina nos fatores de coagulação.....	16
Efeito da jararagina nas plaquetas.....	18
Interação com outros tipos de células.....	19
Atividade pró-inflamatória induzida pela jararagina.....	19
Jararagina e o câncer.....	21
Anticorpos e inibidores da jararagina.....	22
1.7 – Contextualização do projeto desenvolvido.....	23
<b>CAPÍTULO 2 – Objetivos.....</b>	<b>25</b>
<b>CAPÍTULO 3 – Purificação, Cristalização e Refinamento da bothropasina.....</b>	<b>26</b>
Obtenção da proteína bothropasina.....	26
Obtenção da bothropasina sob forma cristalina.....	29
Coleta de dados de difração.....	30
Faseamento e construção do modelo cristalográfico.....	32
Validação da estrutura da bothropasina.....	34
Determinação da estrutura 'nativa'.....	38
<b>CAPÍTULO 4 – Análise Estrutural da Bothropasina e Discussões.....</b>	<b>40</b>
Descrição do modelo cristalográfico.....	40
Contatos intermoleculares no cristal.....	42
Os domínios MDC.....	43
O domínio M: sítio ativo, sítio de ligação do inibidor, cálcio e açúcar.....	45
Região de ligação ao açúcar.....	49
O domínio tipo-desintegrina (D).....	50
O domínio rico em cisteína (C).....	53
Jararagina vs Bothropasina.....	56
Comparação estrutural com outras P-III SVMPs.....	58

Análise seqüencial comparativa entre bothropasina e outras P-III SVMPS.....	62
<b>CAPÍTULO 5 – Conclusões e Perspectivas.....</b>	<b>70</b>
<b>CAPÍTULO 6 – Considerações Finais.....</b>	<b>75</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>78</b>
<b>APÊNDICE I – Relatório Final - PDEE.....</b>	<b>87</b>
<b>APÊNDICE II – Artigos Publicados em Revistas Indexadas.....</b>	<b>100</b>

## LISTA DE FIGURAS

<b>Figura 1.1:</b> Principais áreas de incidência de serpentes da espécie <i>Bothrops jararacana</i> América do Sul.....	1
<b>Figura 1.2:</b> Classificação de metaloproteases de venenos de serpentes (SVMPs).....	10
<b>Figura 1.3:</b> Alinhamento entre as seqüências da jararagina e bothropasina.....	14
<b>Figura 3.1:</b> Purificação da bothropasina.....	28
<b>Figura 3.2:</b> Cristais da proteína bothropasina.....	30
<b>Figura 3.3:</b> Padrão de difração do cristal de bothropasina.....	31
<b>Figura 3.4:</b> Valores de Rfree para estruturas depositadas no PDB em função da resolução.....	36
<b>Figura 3.5:</b> Diagrama de Ramachandran do modelo cristalográfico final.....	36
<b>Figura 3.6:</b> Estatísticas geradas pelo programa PROCHECK.....	37
<b>Figura 4.1:</b> Composição da unidade assimétrica da célula unitária.....	41
<b>Figura 4.2:</b> Contatos intermoleculares do cristal.....	43
<b>Figura 4.3:</b> Estrutura da bothropasina complexada com o inibidor POL647.....	44
<b>Figura 4.4:</b> Domínio catalítico e seu subdomínio.....	46
<b>Figura 4.5:</b> Sítio de ligação ao zinco.....	47
<b>Figura 4.6:</b> Sítio de ligação ao cálcio do domínio M.....	48
<b>Figura 4.7:</b> Domínio catalítico evidenciando a cisteína livre 189.....	49
<b>Figura 4.8:</b> Molécula de NAG presente na estrutura cristalográfica.....	50
<b>Figura 4.9:</b> A estrutura do domínio tipo-desintegrina (D).....	51
<b>Figura 4.10:</b> Pontes dissulfeto no domínio rico em cisteína.....	54
<b>Figura 4.11:</b> Distribuição das pontes dissulfeto por toda a estrutura da bothropasina.....	55
<b>Figura 4.12:</b> Seqüência de aminoácidos da bothropasina mostrando os padrões de pontes dissulfeto para os domínios M, D e C respectivamente.....	55
<b>Figura 4.13:</b> A densidade eletrônica e os verdadeiros resíduos.....	57
<b>Figura 4.14:</b> Mobilidade dos domínios D e C em diferentes P-III SVMPs.....	59
<b>Figura 4.15:</b> Val215 e Leu220 como moduladores da movimentação estrutural.....	60
<b>Figura 4.16:</b> Papel articulador da Leu220.....	61
<b>Figura 4.17:</b> Alinhamento seqüencial de P-III SVMPs.....	65
<b>Figura 4.18:</b> Árvore filogenética de P-III SVMPs.....	66
<b>Figura 4.19:</b> Alinhamentos de SVMPs em grupos I e II.....	67

## LISTA DE TABELAS

<b>Tabela 3.1:</b> Parâmetros e estatísticas da coleta e processamento dos dados de difração de raios X para a proteína na presença de NaI (cristais crescidos a 40C).....	32
<b>Tabela 3.2:</b> Parâmetros e estatísticas da coleta e processamento dos dados de difração de raios X para a proteína 'nativa' (cristais crescidos a 40C).....	39



## RESUMO

A bothropasina é uma proteína hemorrágica de 48 kDa, pertencente à classe P-III das metaloproteases, isolada a partir do veneno bruto da serpente brasileira *Bothrops jararaca*, e que possui os domínios adesivos desintegrina (D) e rico em cisteína (C). Neste trabalho, nós apresentamos a estrutura cristalográfica da bothropasina complexada ao inibidor POL647. O domínio catalítico, metaloprotease (M), pode ser dividido em dois subdomínios, dispostos de maneira muito similar aos descritos para essa família de metaloproteases de venenos de serpentes (em inglês “SVMPs”), que inclui os sítios de ligação ao zinco e ao cálcio. A *cisteína livre*, resíduo Cys189, está localizado em um núcleo hidrofóbico e, sendo assim, impossibilitado de fazer pontes dissulfeto ou qualquer outra interação. O domínio D não apresenta estruturas secundárias bem definidas, sendo constituído, majoritariamente, por estruturas desordenadas como “*loops*”, porém estabilizados por 7 pontes dissulfeto e por dois íons cálcio. A região do motivo ECD está localizada em um “*loop*” e é estruturalmente relacionado à região RGD das desintegrinas-RGD, derivadas de SVMPs da classe P-II. O motivo ECD é estabilizado pela ponte dissulfeto Cys277-Cys310 (entre os domínios D e C), além de um íon cálcio. A cadeia lateral do Glu276 do motivo ECD está exposta ao solvente. Na bothropasina, a região hiper variada (em inglês HVR), descrita para outras P-III de SVMPs, presente no domínio C, de fato, é bastante conservada quando comparada a outros membros da classe P-III de diversas espécies. Nós propomos que esse subgrupo deva ser referido como PIII-HCR (região altamente conservada) SVMPs. Ainda é proposto que as diferenças estruturais dos domínios D, C ou DC possam estar envolvidas em uma melhor adaptação da estrutura na interação com diferentes alvos, além do reconhecimento e especificidade a um substrato para o domínio M.

## ABSTRACT

Bothropasin is a 48kDa hemorrhagic P-III metalloprotease isolated from the venom of the Brazilian snake *Bothrops jararaca*, which has the disintegrin (D) and cysteine-rich (C) adhesive domains. We present the crystal structure of the bothropasin complexed with the inhibitor POL647. The catalytic domain, metalloprotease (M), consists of two subdomains in a very similar scaffold to the ones described for other snake venom metalloproteases (SVMPs) including the zinc and calcium binding sites. The *free cysteine*, residue Cys189, is in a hydrophobic core and it is not available for disulfide bonding or other interactions. The D domain does not have a defined secondary structure, but instead is composed by mostly loops stabilized by seven disulfide bonds and by two calcium ions. The ECD region is in a loop and it is structurally related to the RGD region of RGD-disintegrins, which are derived from P-II SVMPs. The ECD motif is stabilized by the Cys277-Cys310 disulfide bond (between D and C domains) and by one calcium ion. The side chain of the Glu276 of the ECD motif is solvent exposed. In bothropasi, the HVR (hyper-variable region) described for other P-III SVMPs in the C domain in fact presents a well conserved sequence with respect to several other P-III members from different species. We propose that this subset be referred to as PIII-HCR (highly-conserved region) SVMPs. We further propose that the structural differences in the D, C or DC domains may be involved in selecting target binding which in turn could generate substrate diversity or specificity for the M domain.

## ABREVIACOES

$\alpha$ 2-PI –  *$\alpha$ 2-Plasmin Inhibitor*

ADAMs – *A Disintegrin And Metalloprotease*

ADP – Adenosina difosfato

ALT-C – Alternagina-C

CCP4 – *Collaborative Computational Project Number 4*

cDNA – *Complementary DNA*

DC – Desintegrina e rico em cisteína

DNA – *Deoxyribonucleic acid*

DMH – Dose M nima Hemorr gica

ECD – Amino cidos: glutamato, ciste na e aspartato

ECM – *Extracellular Matrix*

HCR – *Highly-conserved region*

HVR – *Hyper-variated region*

ICAM – *Intercellular Adhesion Molecule*

IL-1 $\beta$  – *InterLeukin-1 $\beta$*

IL-6 – *InterLeukin-6*

MADAM – *Mammalian A Disintegrin And Metalloprotease*

MD – Metaloprotease e desintegrina

MDC – Metaloprotease, desintegrina e rico em ciste na

MIDAS – *Metal Ion Dependent Adhesion Site*

MMP(s) – *Mammal Metalloproteinase(s)*

MPACs – *Murin Peritoneal Adherent Cells*

MR – *Molecular replacement*

mRNA – *Messenger ribonucleic acid*

NA – N mero de acesso

NCBI – *National center for biotechnology information*

NO –  xido n trico

PA1-I – *Type 1 Plasminogen Activator Inhibitor*

PDB – *Protein Data Bank*

PEG – *Poli-Etileno Glicol*

pI – Ponto isoel trico

RGD – Aminoácidos: arginina, glicina e aspartato

RHC – Região hiper conservada

RHV – Região hiper variada

SAD – *Single-wavelength Anomalous Dispersion*

SDS-PAGE – *Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis*

SM – Substituição Molecular

SVMP(s) – *Snake Venom Metalloproteinase(s)*

TACE – *Tumor necrosis factor Alfa Converting Enzyme*

TLS – *Translation liberation screw*

TNF- $\alpha$  – *Tumor necrosis factor  $\alpha$*

tPA – *Tissue-type Plasminogen Activator*

uPA – *Urokinase-type Plasminogen Activator*

VCAM – *Vascular Cell Adhesion Molecule*

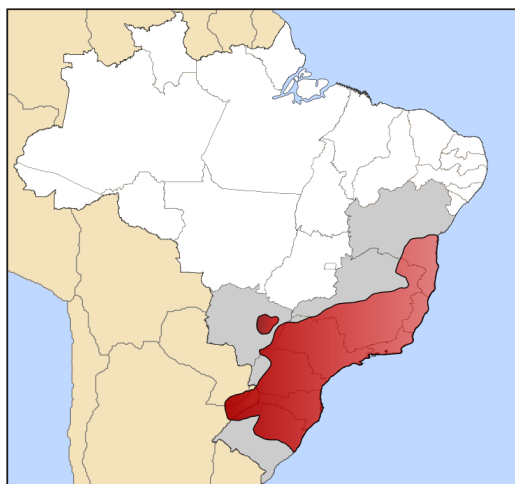
vWF – *von Willebrand Factor*



## INTRODUÇÃO

### 1.1 – A serpente *Bothrops jararaca*

O gênero *Bothrops* compreende cerca de 19 espécies de serpentes distribuídas por todo o território nacional. A espécie *Bothrops jararaca* é predominante no Brasil, sendo encontrada no sul da Bahia, Espírito Santo, Rio de Janeiro, Minas Gerais, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. Estende-se a oeste em direção ao extremo leste de Mato Grosso. No Paraguai ocorre no nordeste e na Argentina, ao norte (Hoge e Romano-Hoge 1981; Campbell e Lamar, 1989), onde é responsável por, aproximadamente, 90% dos casos de incidentes com serpentes venenosas (Figura 1.1).



**Figura 1.1: Principais áreas de incidência de serpentes da espécie *Bothrops jararaca* na América do Sul.** Os Estados que apresentam maior índice de *B. jararaca* estão destacados na cor cinza e as regiões de maior concentração estão destacadas em vermelho. Figura adaptada de Campbell e Lamar, 1989.

## 1.2 – Envenenamento e quadro clínico

O quadro clínico observado no envenenamento botrópico pode ser dividido em sistêmico e local (Manual de Diagnóstico e Tratamento de Acidentes por Animais Peçonhentos, 2003). As manifestações sistêmicas são caracterizadas principalmente por distúrbios na hemostasia, caracterizados por severa destruição da capacidade de coagulação sanguínea, alteração na agregação plaquetária e depleção de fibrinogênio (Hati *et al.*, 1999). A seqüência destes eventos culmina em sangramentos de ferimentos cutâneos pré-existentes e hemorragias à como, por exemplo, a gengivorragia.

Ainda, em nível sistêmico, são observados sintomas como sudorese, hipotensão arterial e hipotermia (Cardoso, 1997). As manifestações locais são caracterizadas por dor e edema, persistentes no local da picada, de intensidade variável e, em geral, de instalação precoce e caráter progressivo.

Equimoses e hemorragias, no ponto da picada, são freqüentes e bolhas podem aparecer na evolução do quadro clínico, acompanhadas ou não de necrose (Cardoso, 1997). Nos casos mais graves, o dano tecidual pode resultar em sérias seqüelas, tais como a perda funcional ou permanente do tecido ou do membro afetado, necrose, cicatrizes permanentes e deformidades (Rosenfeld, 1971; Cardoso *et al.*, 1993). A magnitude dos efeitos locais em envenenamentos botrópicos depende, basicamente, da quantidade de veneno inoculada, do tempo entre a picada e o início da soroterapia, do peso e idade do paciente e da região anatômica onde ocorreu a picada (França, 1997).

O tratamento utilizado, atualmente, no caso dos acidentes por *B. jararaca* é a soroterapia com antiveneno botrópico poliespecífico (Manual de Diagnóstico e Tratamento de Acidentes por Animais Peçonhentos, 2003). A quantidade de antiveneno administrada ao paciente varia conforme a gravidade do caso, que é classificado em

leve, moderado e grave, de acordo com as manifestações clínicas (Cardoso *et al.* 1993). Este tratamento mostra-se eficiente em neutralizar os efeitos sistêmicos observados, porém pouca neutralização é observada sobre os efeitos locais (Gutiérrez *et al.*, 1981). Dado o rápido desenvolvimento da lesão local, é extremamente difícil a sua completa neutralização por antivenenos, pois, freqüentemente, ocorre uma demora no início da aplicação da soroterapia. Como na maioria dos acidentes ofídicos no Brasil o início do tratamento ocorre em um tempo médio de quatro horas após a picada. Este quadro inflamatório local já se instalou e a soroterapia pouco contribui para revertê-lo. Em modelos experimentais, observa-se que a administração do antiveneno botrópico, 15 ou 30 minutos após a inoculação do veneno de *Bothrops asper* em camundongos também não se mostra eficiente na neutralização dos efeitos locais (hemorragia, edema e mionecrose), sendo parcialmente eficiente apenas quando o antiveneno é administrado imediatamente após a inoculação do veneno (Gutiérrez *et al.*, 1981). Entretanto, não se observa uma completa neutralização dos efeitos locais, mesmo quando o veneno botrópico é inoculado em animais experimentais, logo após a administração do antiveneno homólogo, ainda que este apresente alto título de anticorpos capazes de neutralizar as principais atividades do veneno botrópico *in vitro* (Otero *et al.*, 1997).

### 1.3 – Integrinas e desintegrinas

Integrinas são receptores extracelulares da matriz extracelular em animais e são glicoproteínas estruturadas como heterodímeros de duas cadeias ( $\alpha$  e  $\beta$ ) associados de forma não covalente. Baseando-se na similaridade seqüencial, o sistema de nomenclatura proposto para a designação das integrinas categoriza as subunidades individuais de cada heterodímero como  $\alpha$  (120 a 180 kDa) e  $\beta$  (90 a 110 kDa), ambas do tipo transmembrânicas, cujas combinações definem as propriedades adesiva e de

sinalização.

A maior parte da molécula de integrinas se encontra exposta no meio extracelular, enquanto que o citoesqueleto de actina e o maquinário de sinalização estão associados a um pequeno domínio citoplasmático (Dzamba *et al.*, 2001). Em mamíferos é constatada a existência de, ao menos, 18 subunidades  $\alpha$  e  $\beta$ , que, combinando-se entre si, garantem a grande diversidade de integrinas e suas diferenças de ligantes, afinidades e atividades biológicas (Marcinkiewicz *et al.*, 1999; Parise *et al.*, 2000; Cairns *et al.*, 2003). Os domínios extracelulares das subunidades  $\beta$  apresentam 56 resíduos de cisteínas, a maioria dos quais se localiza dentro de um agrupamento de 260 aminoácidos, separados da membrana plasmática por, aproximadamente, 50 resíduos. As subunidades  $\alpha$  contêm 7 seqüências repetitivas de, aproximadamente, 60 resíduos constituintes de motivos “*EF Hand*”, com seqüência consenso “DxDxDGxxD”, capazes de atrair cátions divalentes como  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  e  $\text{Mn}^{2+}$ . Estes motivos são importantes para manter a capacidade de adesão da integrina e as mudanças conformacionais que afetam o receptor pelo ligante (Dzamba *et al.*, 2001).

O termo integrina foi, inicialmente, utilizado, referindo-se ao papel de intermediador na comunicação entre o citoesqueleto celular e proteínas plasmáticas ou da matriz extra celular (ECM), por exemplo, laminina, fibrinogênio, fibronectina, colágeno, vitronectina, osteopontina, trombospondina e iC3b do sistema complemento. No entanto, essas proteínas de superfície celular também mediam a adesão célula-célula através de interações com outras proteínas de membrana como ICAM-1, 2 e 3, VCAM-1 e caderina E (Darribere *et al.*, 2000; Clegg *et al.*, 2003; Jin e Varner, 2004).

O principal determinante da especificidade da interação entre os ligantes é a composição das subunidades  $\alpha$  e  $\beta$  de cada tipo de molécula. Muitas das integrinas



possuem capacidade de reconhecimento da seqüência RGD em seus ligantes. Entretanto, essa característica não garante que proteínas, com essa seqüência, possam ser reconhecidas por um mesmo receptor, o que leva a considerar a probabilidade de que sejam seqüências vizinhas ao RGD e as diferentes conformações tanto dos sítios adesivos, como dos rearranjos conformacionais que definem tal especificidade (Dzamba *et al.*, 2001).

Ao se associarem a ligantes endógenos ou não, as integrinas dão início a um processo com diferentes estados de ativação, sendo definido como a troca na afinidade por seu ligante por um receptor individual. O processo regulatório, capaz de modificar a afinidade, se define como *Modulação da Afinidade*, que é disparado através de um sistema de sinais do tipo “*inside out*”, sendo então uma forma rápida para regular as funções adesivas em resposta a vários estímulos fisiológicos (Dzamba *et al.*, 2001). Entre os mecanismos envolvidos na regulação da interação com o ligante encontram-se íons metálicos, como sugerido pela presença dos motivos “*EF-hand*” e MIDAS (*Metal Ion Dependent Adhesion Site*), os quais podem modular positiva ou negativamente. Por exemplo, o  $Mn^{2+}$  induz ao aumento da afinidade do receptor por seu ligante, enquanto que o  $Ca^{2+}$  estabiliza conformações de baixa afinidade. Além do mais, são considerados os fenômenos “*clustering*” ou agrupamento de receptores como outra forma de aumentar a afinidade por seu ligante, o qual é independente do estado de ativação do receptor mediado por mudanças conformacionais. Neste modelo, a presença de ligantes multivalentes ou a associação de receptores com o citoesqueleto podem limitar a difusão das moléculas pela membrana, aumentando, desta maneira, a força adesiva da interação (Kim *et al.*, 2003).

As funções das integrinas como mediadoras de adesão e fusão celular (célula-célula e célula-matriz), pela ativação e inibição de sinais, se resumem em dois

tipos de sistemas de sinalização: (1) a origem é a própria célula portadora da integrina em sua parte exterior e (2) proveniente do meio externo para a célula, modulando assim sua função. O primeiro sistema é conhecido como sinalização “*inside-out*”, enquanto que o segundo é conhecido como sinalização “*outside-in*”. Na sinalização “*inside-out*”, o aumento de afinidade dos receptores, por mudanças conformacionais, é mediado pela ativação prévia de sistemas de sinalizações intracelulares próprios. Desta maneira, constitui-se uma sinalização bidirecional entre o ambiente extra e intracelular (Coppolino e Dedhar, 2000). Os eventos decorrentes da transdução de sinais a partir de integrinas são complexos e, até o momento, não foram completamente elucidados.

As cascatas de sinalizações, passíveis de serem disparadas por integrinas, podem ser moduladas, também, por proteínas de membrana como receptores de fatores de crescimento (Eliceiri, 2001; Ross, 2004), receptores de sobrevivência (Hanahan e Weinberg, 2000), receptores que se ligam à região constante de imunoglobulinas (Ortiz-Stern e Rosales, 2003), proteínas transmembrânicas (TM4SF, Woods e Couchman, 2000), receptor de ativador de plasminogênio tipo uroquinase (uPA, CD87; Xue *et al.*, 1997) e pela proteína associada a integrina (IAP, CD47; Lindberg *et al.*, 1996).

Como pode ser visto, as integrinas estão presentes em numerosos processos fisiológicos onde se faz vital o contato direto entre as células participantes, servindo não apenas como molécula ponte em interações intercelulares, mas também como molécula ponte entre a matriz extracelular e o citoesqueleto da célula, através da troca de sinais celulares. Em decorrência dessas atividades, as integrinas não estão associadas somente a processos fisiológicos, mas também a processos patológicos como a neoplasia, a metástase, a disfunção imune, as infecções virais, a osteoporose e a coagulopatia (Arnaout, 1990; Hynes, 1992).

A ligação reversível célula-matriz extracelular, mediada por integrina, pode ser interrompida por pequenos peptídeos chamados desintegrinas. O termo desintegrina foi utilizado, pela primeira vez, em 1989, para descrever um grupo de moléculas de baixa massa molecular (5-9 kDa), ricas em cisteínas, derivadas de venenos de serpentes, os quais interagem com receptores da superfície das células (integrinas) (Dennis *et al.*, 1989; Gould *et al.*, 1990; Niewiarowski *et al.*, 1994; Maclane *et al.*, 1998; Huang, 1998). Desintegrinas são potentes inibidores da agregação plaquetária por se ligarem especificamente ao receptor de fibrinogênio, à glicoproteína IIb/IIIa ( $\alpha_{IIb}\beta_3$ ) de membrana de plaquetas, através da sua seqüência adesiva RGD. O domínio desintegrina apresenta uma grande diversidade estrutural e funcional. As desintegrinas, derivadas das metaloproteases de venenos de serpentes (SVMPs) da classe P-II, são normalmente pequenas e geralmente contêm a seqüência adesiva RGD (Huang e Niewiarowski, 1994). Estudos demonstraram que desintegrinas, contendo a seqüência RGD, são capazes de inibir a adesão de células tumorais à matriz extracelular por competição específica pelos receptores de integrinas (Tripathi *et al.*, 1994; Sheu *et al.*, 1997; Chiang *et al.*, 1994).

Proteínas que não apresentam o motivo RGD, mas que apresentam atividade de inibição de ligação de ligantes a integrinas, são encontradas em venenos de serpentes. São as metaloproteases/desintegrinas ou tipo-desintegrinas, descritas no próximo item.

#### **1.4 – Metaloproteases**

As serpentes são as mais prodigiosas fontes de toxinas que podem chegar a ser úteis à humanidade. Seus venenos são compostos por complexas mesclas de polipeptídeos e proteínas que induzem a uma variedade de efeitos farmacológicos, os quais provêm uma útil ferramenta para a compreensão de muitos eventos moleculares

dentro dos processos fisiológicos normais. Os principais componentes encontrados em venenos de serpentes são lectinas, metaloproteinasas, serinoproteinasas, desintegrinas, fosfolipases e peptídeos que agem sobre a bradicinina e o sistema angiotensina. Além de proteínas, o veneno de serpentes também é caracterizado por uma rica mistura de compostos orgânicos e inorgânicos que variam de acordo com a espécie de serpente. Os mais abundantes compostos inorgânicos são o cálcio, cobre, ferro, potássio, magnésio, manganês, sódio, fósforo, cobalto e zinco. A função desses compostos está relacionada a mecanismos catalíticos de certos componentes enzimáticos presentes no veneno (Bjarnason e Fox, 1994). Em estudos proteômicos recentes, mais de 200 peptídeos foram revelados por eletroforese bi-dimensional do veneno de *B. jararaca*, confirmando, assim, a complexa natureza do veneno (Fox *et al.*, 2002).

Um grupo importante de enzimas proteolíticas, encontradas nos venenos, é formado pelas metaloproteases (SVMPs – *Snake Venom Metalloproteinases*).

As metaloproteases compreendem uma série de enzimas geralmente dependentes de zinco, também chamadas de zinco-metaloproteases, com variada massa molecular e responsáveis pelo efeito hemorrágico característico destes envenenamentos (Bjarnason e Fox, 1994). As metaloproteases são divididas de acordo com a estrutura primária de seus sítios catalíticos e incluem os grupos das zincinas (subdivididas em gluzincinas ou metzincinas, dependendo do quarto resíduo responsável pela coordenação do íon Zn, que pode ser um glutamato ou histidina, respectivamente), inverzincinas, carboxipeptidases e DD carboxipeptidases. Este último grupo ainda é subdividido em astacinas (encontradas em crustáceos), serralisinas (bactérias), matrixinas (metaloproteases presentes na matriz extracelular de mamíferos) e reprolisinas (venenos de serpentes) (Bode *et al.*, 1993; Hooper, 1994).

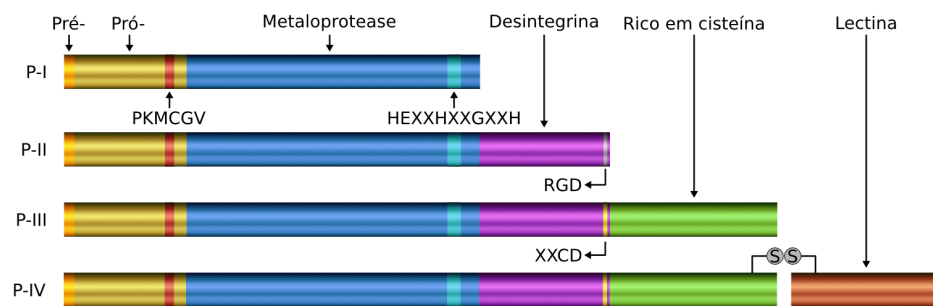
As metaloproteases hemorrágicas de venenos são classificadas, de acordo

com sua massa molecular, em pequenas (22-27 kDa), médias (33-48 kDa) e grandes (50-65 kDa), segundo Bjarnason e Fox (1994).

As SVMPs são produzidas na glândula de veneno como pré-pró-domínio, na forma de precursores inativos ou zimogênios, nos quais o mecanismo de “*cysteine-switch*” inibe sua atividade enzimática (Grams *et al.*, 1993; Gutiérrez *et al.*, 2000). O peptídeo sinal das SVMPs possui, aproximadamente, 18 aminoácidos altamente conservados, os quais servem como seqüência-sinal para determinar sua localização sub celular ou secreção (Sha *et al.*, 1997). Os zimogênios são processados proteoliticamente para serem ativados. Seu domínio de ligação ao  $Zn^{2+}$  é tetraédrico, com seqüência de aminoácidos **HExGHNLGxxHD** com 3 His coordenando o  $Zn^{2+}$  e uma molécula de água (Jia *et al.*, 1996). Entretanto, na presença de ligantes inibidores, o zinco também pode apresentar uma coordenação pentaédrica, através dos oxigênios do grupo carbonil desse ligante. A cadeia lateral da metionina do motivo “*Met-turn*” fornece uma base hidrofóbica abaixo das cadeias laterais das histidinas que coordenam o zinco (Bode *et al.*, 1993). O domínio catalítico ainda conta com um íon de cálcio localizado sobre a superfície da molécula, do lado oposto ao catalítico, importante na conformação estrutural.

Desde o ponto de vista genético, verifica-se que as SVMPs se encontram codificadas por quatro classes de DNA: classe nucleotídica I ou N-I, que codifica para domínios pró e metaloprotease, classe N-II que codifica para domínios pró, metaloprotease e desintegrina e classe N-III, que codifica para domínios pró, metaloprotease, tipo-desintegrina e rico em cisteína localizado na extremidade carboxi-terminal do domínio desintegrina e a última classe, a N-IV, que codifica para um domínio tipo lectina, na extremidade carboxi-terminal da proteína (Wolfsberg *et al.*, 1995; McLane *et al.*, 1998; Yamamoto *et al.*, 1999).

Com base em seus domínios estruturais, as metaloproteínas, quando isoladas do veneno bruto, podem ser classificadas em 4 classes distintas: P-I, P-II, P-III ou P-IV, de acordo com a presença de domínios adicionais na região carboxi-terminal do domínio metaloprotease (Bjarnason e Fox, 1994; Hite *et al.*, 1994) (figura 1.2). Assim, após o processamento, proteínas da classe P-I encontram-se somente com o domínio catalítico, as da classe P-II com domínios catalítico e desintegrina (contendo o motivo RGD), as da classe P-III com domínios catalítico, tipo-desintegrina e rico em cisteína, e proteínas da classe P-IV com a adição do domínio lectina na porção C-terminal do domínio rico em cisteína.



**Figura 1.2: Classificação de metaloproteases de venenos de serpentes (SVMPs).** A região altamente conservada “PKMCGV” no Pró-domínio, assim como o sítio de coordenação do zinco “HEXXHXXGXXH” estão em destaque. A ponte dissulfeto, que liga o domínio lectina ao rico em cisteína em metaloproteases da classe P-IV, está representada por dois “S” circunscritos. Diagrama adaptado de Bjarnason e Fox, 1994.

As metaloproteínas de venenos de serpentes (SVMPs), pertencentes à classe P-III, são denominadas tipo-desintegrina (não-RGD) por apresentarem o motivo E/DCD, substituindo o motivo RGD. Estudos têm mostrado que o motivo D/ECD está envolvido na ligação de integrinas (Myles *et al.*, 1994) e que domínio rico em cisteína também pode estar envolvido na interação (Bjarnason e Fox, 1994). As metaloproteínas tipo-desintegrinas (MDC) podem sofrer processamento proteolítico, liberando o domínio catalítico (metaloprotease) e os domínios tipo-desintegrina e rico

em cisteína (DC). Proteínas contendo somente os domínios DC processados mostram atividades de inibição de adesão celular e de agregação plaquetária. A ALT-C, uma DC isolada do veneno de *B. alternatus*, é capaz de inibir a adesão de células K562 transfectadas com a integrina  $\alpha 2\beta 1$  mediada pelo colágeno tipo I (Souza *et al.*, 2000), além de inibir a adesão de fibroblastos humanos (Cominetti *et al.*, 2004). A Jararagina-C, uma proteína DC de 28kDa originada do processamento da jararagina, possui a capacidade de inibir a agregação plaquetária induzida pelo colágeno e por adenosina difosfato (ADP) (Usami *et al.*, 1994; Jia *et al.*, 1997; Shimokawa *et al.*, 1997). A Jaracetina, um dímero de 60 kDa, que representa uma diferente forma processada da jararagina, interage com o domínio A do vWF e também bloqueia a adesão das plaquetas ao colágeno dependente de  $\alpha 2\beta 1$  (De Luca *et al.*, 1995).

### 1.5 – Proteínas humanas homólogas as SVMPs P-III

Os fenômenos de adesão célula-célula, célula-matriz e a proteólise da matriz extracelular são vitais para o desenvolvimento normal de processos, como a morfogênese dos tecidos, a cicatrização de feridas, e em processos patológicos como a invasão de células tumorais e metástase (Gould *et al.*, 1990; Wolfsberg *et al.*, 1995). Estes fenômenos se encontram mediados por proteínas de adesão de superfície celular, tais com as caderinas, a superfamília das imunoglobulinas, as selectinas, integrinas e sindecanas (Gould *et al.*, 1990). Outro grupo de proteínas envolvidas é o das moléculas ancoradas à superfície da membrana, as quais se encontram representadas pelas metaloproteases de matriz da membrana. Entre as proteases acopladas à matriz da membrana, são encontradas aquelas que contêm um domínio metaloprotease e um domínio tipo-desintegrina, codificadas pela família de genes conhecidos como ADAMs (*A Disintegrin And Metalloprotease*) (Wolfsberg *et al.*, 1995). As ADAMs são, portanto, proteínas homólogas às SVMPs da classe P-III encontradas em

mamíferos e outros animais e estão relacionadas a diversos processos fisiológicos como fertilização e diferenciação celular.

A estrutura básica das ADAMs está filogeneticamente bem conservada e as mudanças estruturais são produtos da adaptação necessária à sua respectiva função (Yamamoto *et al.*, 1999). Assim, por exemplo, as ADAM designadas desde 1 a 7 se expressam, de maneira principal, em órgãos reprodutores, enquanto que seu papel, durante a espermatogênese, é a fusão esperma-óvulo (Wolfsberg *et al.*, 1995; Yamamoto *et al.*, 1999; Blobel 2005), e cada uma é expressa em distintas etapas do processo espermatogênico. A ADAM 9, por sua vez, também chamada MDC9 (*Metalloprotease-like, Disintegrin-like and Cysteine rich*), tem sido encontrada em vários tipos de órgãos, entre os quais os pulmões e as glândulas mamárias, podendo ter relação com o desencadeamento de sinais de transdução. A ADAM 11 é encontrada em células de órgãos reprodutores e não reprodutores, enquanto que as ADAMs 12 e 19, também conhecidas como Meltrinas  $\alpha$  e  $\beta$  respectivamente, estão localizadas na musculatura embrionária (estágio neonatal) e tecido ósseo de embriões e adultos (Yamamoto *et al.*, 1999). Por outro lado, Yamamoto (1999) relaciona a ADAM 17, também chamada de TACE (*Tumor necrosis factor Alfa Converting Enzyme*), e ADAM 10, também conhecida por MADAM (*bovine Mammalian A Disintegrin And Metalloprotease*), que cumprem funções importantes no processamento da forma acoplada à superfície celular do precursor do fator de necrose tumoral  $\alpha$  (TNF- $\alpha$ ), na liberação da forma madura da citocina.

Todas as proteínas pertencentes à família das ADAMs contêm um domínio do tipo metaloprotease, o qual é similar ao encontrado nas SMVPs dependentes de  $Zn^{2+}$ . O domínio, tipo metaloprotease das ADAMs 1, 8, 9 e 10, apresenta resíduos conservados no sítio ativo, similares aos da SVMPs mantendo também sua propriedade



catalítica, enquanto que, no das ADAMs 2-7 e 11, o mesmo contém seqüências diferentes em seu sítio ativo. Assim, essas proteínas, apesar da presença do domínio metaloprotease, são cataliticamente inativas (Wolfsberg *et al.*, 1995).

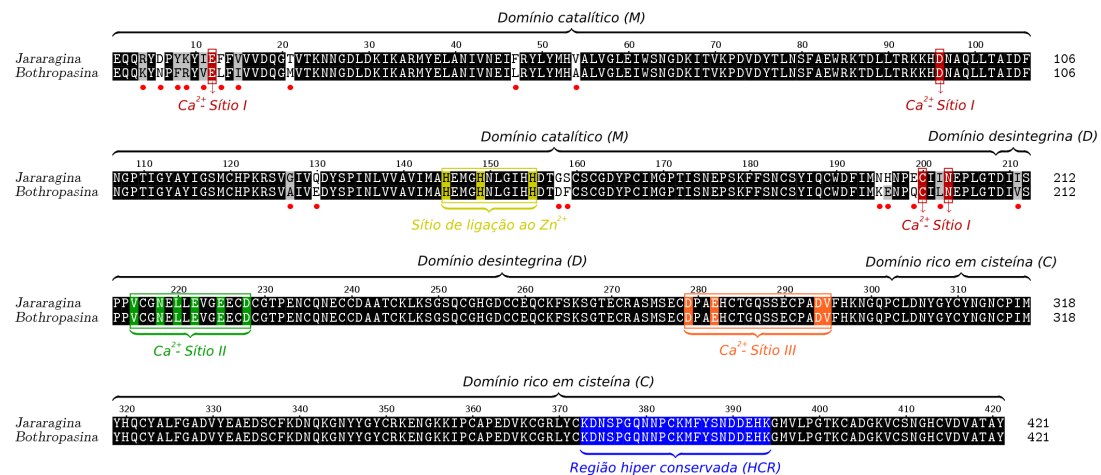
### 1.6 – Jararagina e Bothropasina

Jararagina e bothropasina são duas SVMPs da classe P-III isoladas do veneno de *B. jararaca*, e, portanto, pertencentes à família metalopeptidase, e à subfamília reprolisina, com atividade catalítica hidrolase do tipo endopeptidase, dependentes do zinco (Bjarnason e Fox, 1994). As duas proteínas apresentam elevado grau de identidade, aproximadamente 97%, o que as caracteriza como isoformas. Uma vez que o aminoácido N-terminal encontra-se bloqueado, não é possível caracterizar essas proteínas através de seqüenciamento da região N-terminal por degradação de Edman. Assim, segundo a literatura, as proteínas jararagina e bothropasina são distinguidas com base no protocolo de purificação, isto é, dependendo das colunas cromatográficas utilizadas, é isolada jararagina ou bothropasina.

A jararagina foi isolada pela primeira vez em 1992 (Paine *et al.*, 1992) e caracterizada como uma metaloproteinase/desintegrina de alta massa molecular (52 kDa) da classe P-III, com 421 aminoácidos na proteína madura, após a modificação pós-traducional. Possui ação hemorrágica (Paine *et al.*, 1992) e também a capacidade de inibir a agregação plaquetária, ao bloquear a ligação do colágeno à integrina  $\alpha 2\beta 1$  presente na membrana das plaquetas. A jararagina se liga ao domínio I da cadeia  $\alpha 2$  e/ou cliva a cadeia  $\beta 1$  dessa integrina (Kamiguti *et al.*, 1996). A jararagina interfere na função plaquetária de duas maneiras distintas: (1) por degradação dos diferentes receptores plaquetários e proteínas adesivas na hemostasia; e (2) interferência não enzimática (mediada pelo domínio desintegrina) com a função de receptores da adesão plaquetária (Kamiguti *et al.*, 1996b). Estudos recentes têm sugerido a existência de

formas alternativas de jararagina secretadas pela glândula venenífera. Uma das formas é proteoliticamente estável (que se apresenta no veneno na forma da jararagina) e outra proteoliticamente instável (processada rapidamente no veneno produzindo a jararagina-C) (Moura-da-Silva *et al.*, 2003).

A bothropasina foi isolada do veneno de *B. jararaca* pela primeira vez em 1982 (Mandelbaum *et al.*, 1982) e caracterizada como uma proteína da classe P-III das SVMPs. Assim como a jararagina, bothropasina também é constituída por 421 resíduos de aminoácidos e apresenta atividade hemorrágica (Mandelbaum *et al.*, 1982). Estudos recentes mostram a expressão dos domínios DC da bothropasina (Assakura *et al.*, 2003). Devido ao fato de não ser possível o seqüenciamento do N-terminal, as seqüências conhecidas de ambas as proteínas são as deduzidas do cDNA (mostradas na figura 1.3).



**Figura 1.3: Alinhamento entre as seqüências da jararagina e bothropasina.** Quando comparadas em um alinhamento seqüencial, 19 resíduos não idênticos são encontrados, todavia, 10 são similares. Seqüências deduzidas do DNA (Número de acesso no NCBI: P30431 e AAC61986.2, respectivamente).

### 1.6.1 – Os domínios estruturais das moléculas de jararagina e bothropasina

A jararagina e a bothropasina, duas proteínas da classe P-III das metaloproteases, como visto anteriormente, são compostas por uma estrutura de

multidomínios, representada esquematicamente nas figuras 1.2 e 1.3. Estes domínios são:

- **Peptídeo sinal:** tem como função garantir a correta localização subcelular ou secreção da proteína (Shah *et al.*, 2000).
- **Pró-domínio:** região com resíduos de aminoácidos altamente conservados, que regula a atividade enzimática, através da interação entre a sulfidrina do resíduo Cys, do motivo PKMCGVTQ das SMVPs, com o íon  $Zn^{2+}$  presente no sítio ativo. Esta interação resulta na inativação do domínio catalítico, um mecanismo chamado de “*cystein switch*”, similar ao proposto para as MMPs encontrado no precursor protéico não processado. Esta região permite que as metaloproteases sejam armazenadas na forma de zimogênios inativos, evitando, assim, sua autodigestão (Grams *et al.*, 1993).
- **Domínio catalítico – Metaloprotease (M):** é o domínio de ligação de íons metálicos. No caso da jararagina, o íon  $Zn^{2+}$ . Apresenta a seqüência de aminoácidos **HExxHxxGxxHD**, altamente conservada nessa família de enzimas e responsável pela coordenação do metal próximo à região CIMxP da volta de metionina ou “*met-turn*”, característica das metazincinas. Nesse domínio são encontrados 7 resíduos de cisteínas.
- **Domínio tipo-desintegrina (D):** este domínio apresenta a seqüência ECD que está associada com a especificidade das SVMPs pelas integrinas (Zhou *et al.*, 1995; Jia *et al.*, 2000). Nesse domínio são encontradas 14 cisteínas.
- **Domínio rico em cisteína (C):** Apesar de ser chamado de rico em cisteína, o domínio C possui uma “densidade de cisteínas” menor que o domínio D. Nele, são encontradas 13 cisteínas.

## 1.6.2 – Atividades biológicas da jararagina

### Jararagina e hemorragia

Metaloproteinasas de venenos de serpentes são freqüentemente responsáveis pela degradação de proteínas da matriz celular, resultando no rompimento da integridade das células endoteliais nas paredes dos vasos sanguíneos e subsequente hemorragia potencializada pelos distúrbios das funções plaquetárias (Bjarnason e Fox, 1994; Kamiguti *et al.*, 2003). As SVMPs afetam principalmente vasos capilares e pequenas veias (Moreira *et al.*, 1994).

A jararagina promove a homeostase, através da degradação das proteínas responsáveis pela coagulação, como, por exemplo, os fibrinogênios (Kamiguti *et al.*, 1994a; Markland, 1998) e pela inibição da agregação plaquetária (Kamiguti *et al.*, 1996a).

A hemorragia sistêmica induzida pela jararagina em ratos, parece ser dependente de sua atividade proteolítica, uma vez que sua atividade é completamente eliminada quando a jararagina é incubada com peptídeos inibidores de batimastato anteriormente à injeção (Escalante *et al.*, 2003). Além do mais, quando o batimastato e o quelante  $\text{CaNa}_2\text{EDTA}$  são administrados por alguns minutos que antecedem o envenenamento em ratos, os compostos apresentam clara eliminação das hemorragias locais e necroses dermais (Rucavado *et al.*, 2000). Sendo assim, pode-se constatar a importância da preservação da atividade catalítica das metaloproteases a fim de promover a quebra das proteínas de matriz, caracterizando-se como o principal mecanismo envolvido no sangramento.

### Efeitos da jararagina nos fatores de coagulação

A jararagina tem como característica clivar o fibrinogênio em sua porção C-terminal da cadeia  $\text{A}\alpha$ , resultando na remoção de um fragmento com

aproximadamente 23 kDa, mas não promove nenhum efeito tanto na cadeia  $\beta$ , quanto na  $\gamma$  (Kamiguti *et al.*, 1994a). A restante molécula de fibrinogênio purificada permanece funcional em ambas as respostas de agregação plaquetária para o ADP, adrenalina e sua habilidade coagulante plasmática pela ativação da trombina. A única consequência da atividade  $\alpha$ -fibrinogenase da jararagina foi a anormal polimerização de fibrina pela trombina (Kamiguti *et al.*, 1994a).

Tem-se mostrado que a jararagina aumenta a atividade fibrinolítica *in vitro* do plasma humano e de outros animais, através do aumento da atividade do tPA (*tissue-type Plasminogen Activator*), pela dissociação do complexo do tPA com seu inibidor PA1-I (*type 1 Plasminogen Activator Inhibitor*) e pela inativação do  $\alpha$ 2-PI ( $\alpha$  2-*Plasmin Inhibitor*) no plasma (Sugiki *et al.*, 1995). Além do mais, a ativação da uPA (*urokinase-type Plasminogen Activator*) foi descrita após interações *in vitro* entre o zimínogênio da uPA e jararafibrase I (Sugiki *et al.*, 1998).

Esses efeitos possivelmente ocorrem devido à atividade catalítica da jararagina permanecer inalterada pelos inibidores de protease plasmáticos como as  $\alpha$ <sub>2</sub>-macroglobulinas. Embora as P-I SVMPs sejam rapidamente inibidas, P-III SVMPs não apresentam inibição pelas  $\alpha$ <sub>2</sub>-macroglobulinas, além do fato da capacidade desse inibidor em fazer ligações covalentes com a enzima (Kamiguti *et al.*, 1994b). Os autores ainda sugerem que a resistência à inibição apresentada pelas P-III pode estar relacionada a restrições estruturais, uma vez que proteínas da classe P-III, possuem os domínios adicionais desintegrina e rico em cisteína. Esses domínios poderiam estar relacionados com interações entre SVMPs e os inibidores plasmáticos. Por outro lado, jararafibrase I é completamente inibida pelas macroglobulinas humanas e de ratos (Anai *et al.*, 1998).

### Efeito da jararagina nas plaquetas

A maioria dos estudos relacionados à ação da jararagina na hemostase envolve agregação plaquetária. Após a injúria vascular, plaquetas são ativadas pelas fibras de colágeno, mediadas pelo Fator de von Willebrand (vWF), resultando na adesão plaquetária ao subendotélio exposto. Os receptores, integrina  $\alpha_2\beta_1$  e o complexo GP VI-FcR cadeia  $\gamma$ , possuem um papel pivotal na ativação das plaquetas pelo colágeno (Watson e Gibbins, 1998). A adesão de plaquetas é mediada pela ligação do complexo GPIb-IX-V com o vWF (Weiss *et al.*, 1986; Atkinson *et al.*, 2003; Pinto *et al.*, 2007). A união desses receptores promove a ativação das plaquetas e agregação, via interação com a integrina  $\alpha_{IIb}\beta_3$  com os ligantes plasmáticos, proporcionando a formação de uma barreira plaquetária importante para que o sangramento cesse.

A jararagina inibe a agregação plaquetária induzida pela ristocetina e pelo colágeno, indicando que ambos os caminhos supracitados são afetados pela ação da toxina. A inibição da agregação plaquetária induzida pela ristocetina tem sido atribuída a um efeito direto da enzima sobre o vWF e não de sua ativação no receptor GP Ib-IX-V (Kamiguti *et al.*, 1996a; Serrano *et al.*, 2006). A jararagina cliva a subunidade vWF na metade N-terminal da molécula, que contém o sítio de ligação para o receptor GPIb (o domínio AI), dando origem a vários fragmentos. Além do mais, a hidrólise do vWF pela jararagina acarreta no desaparecimento das estruturas multiméricas de elevada massa molecular da proteína. A clivagem do vWF pela jararagina e, provavelmente por outras proteínas presentes no veneno de *B. jararaca*, contribui ao processo hemorrágico característico de pacientes com envenenamento.

A proteólise do vWF, induzida pela jararagina, somada aos efeitos sobre a interação plaqueta-colágeno, prejudica a coagulação que unida à coagulopatia de consumo, pode implicar em risco de morte da vítima (Kamiguti *et al.*, 1996a).

### **Interação com outros tipos de células**

A observação de que a jararagina interage com integrinas  $\alpha_2\beta_1$  e com o colágeno, sugere a possibilidade de interações com outros sistemas celulares que não plaquetas, interferindo em efeitos mediados pela  $\alpha_2\beta_1$ . Esses efeitos incluem mudanças no crescimento de fibroblastos em redes de colágenos compreendidos por colágenos fibrilares do tipo 1 (Zigrino *et al.*, 2002). Uma vez expostos à jararagina, ligações na superfície celular acontecem através da interação com integrinas  $\alpha_2\beta_1$  e, diferentemente das plaquetas, a ligação da jararagina ao  $\alpha_2\beta_1$  é suficiente para induzir as sinalizações essencialmente idênticas às observadas para o colágeno fibrilar e promove a expressão de MMP-1, MT1-MMP e integrinas  $\alpha_2\beta_1$  (Zigrino *et al.*, 2002).

Em fibroblastos, a jararagina atua como um substrato que mimetiza o colágeno, que se liga e ativa integrinas. Por outro lado, em agregação plaquetária induzida pelo colágeno, jararagina se liga ao domínio  $\alpha_2\beta_1$  da integrina, inibindo a sinalização intracelular mediada pelo colágeno. Kamiguti *et al.*, (2000) sugere que apenas a ligação jararagina/ $\alpha_2\beta_1$  seja suficiente para que a efetiva transdução do sinal em plaquetas aconteça.

Os mecanismos de ação das SVMPs em células endoteliais ainda estão fracamente elucidados. Entretanto, Masuda *et al.*, (1998) especula que as P-III SVMPs possam promover a apoptose das células endoteliais e não um efeito citotóxico.

### **Atividade pró-inflamatória induzida pela jararagina**

A jararagina atua em uma variedade de diferentes mecanismos envolvidos na homeostase e é diretamente relacionada na hemorragia local e sistêmica observadas em pacientes envenenados (Kamiguti *et al.*, 1991; Paine *et al.*, 1992) e *in vitro* (Kamiguti *et al.*, 1991; Gallagher *et al.*, 2005). Outro importante papel desenvolvido pela jararagina é o de promover uma forte atividade pró-inflamatória, acarretando em

extensos danos ao tecido local (Costa *et al.*, 2002; Laing *et al.*, 2003) e prejudicando a proliferação de células dos sistemas responsáveis pela reconstituição do tecido (Ho *et al.*, 2002).

Apesar dos recentes estudos realizados para se compreender melhor os efeitos locais desencadeados, após acidentes botrópicos, em relação às toxinas envolvidas no processo de ativação dos componentes pró-inflamatórios endógenos e o mecanismo desencadeado imediatamente após o contato destas com os tecidos, ainda não estão bem esclarecidos.

A exposição das células MPACs (*Murin Peritoneal Adherent Cells*) à jararagina estimula, após um período de aproximadamente 4 horas, a produção de mediadores imunes de amplo espectro como: o TNF- $\alpha$ , a IL-1 $\beta$  (*InterLeukin-1 $\beta$* ) e a IL-6 mRNAs (*InterLeukin-6*). Entretanto, quando recombinantes solúveis de TNF- $\alpha$  e IL-6 foram encubados na presença de jararagina, clivagens proteolíticas foram observadas.

A toxina nativa também estimula a subsequente degradação das citocinas produzidas por clivagem proteolítica. O TNF- $\alpha$  e a IL-6 foram rapidamente degradados, enquanto que a atividade sobre a IL-1 $\beta$  foi mínima (Clissa *et al.*, 2001).

A susceptibilidade do TNF- $\alpha$  e da IL-6 ao efeito catalítico da Jararagina, não é clara, mas, provavelmente, tenha correlação com a presença de pontes dissulfeto, presentes no TNF- $\alpha$  e na IL-6 (Callard e Gearing, 1994) e ausentes na IL1 $\beta$ . Possivelmente, as pontes dissulfeto geram características estruturais, que expõem resíduos hidrofóbicos na superfície das moléculas e que são necessários para a atividade proteolítica da jararagina. Deste modo, a jararagina, na sua forma nativa, pode atuar como ativador do mecanismo endógeno da resposta inflamatória (Mourada-Silva *et al.*, 1996a), induzindo à produção de citocinas pela estimulação de células



inflamatórias (Clissa *et al.*, 2001). Células endoteliais também são ativadas e liberam óxido nítrico (NO) e IL-6 em resposta à ação da jararagina (Schattner *et al.*, 2005).

As citocinas, por possuírem um papel de grande relevância na regulação de mecanismos inflamatórios, têm se mostrado como fortes candidatas a serem os possíveis agentes endógenos pró-inflamatórios, liberados em resposta ao envenenamento. A produção de citocinas tem sido descrita em modelos experimentais e em vítimas de acidentes botrópicos e crotálicos (Lomonte *et al.* 1993; Barravieira *et al.* 1995; Barros *et al.* 1998; Petricevich *et al.* 2000). Porém, poucas evidências sobre o envolvimento direto das citocinas, na inflamação ocasionada por envenenamentos ofídicos, estão disponíveis na literatura.

### **Jararagina e o câncer**

Recentes estudos foram realizados com o intuito de investigar a habilidade da jararagina em interferir em mecanismos relacionados ao câncer. O tratamento de cultura de células de melanoma humano com jararagina levou a alterações morfológicas celulares, variabilidade e adesão a proteínas da matriz extracelular, resultando em uma significativa redução da metástase pulmonar quando comparadas aos controles (Correa *et al.*, 2002). Em uma diferente metodologia, polipeptídeos similares aos da angiostatina foram gerados a partir da hidrólise do plasminogênio pela jararagina *in vitro*. Os fragmentos resultantes isolados inibiram a proliferação de células endoteliais de maneira similar à angiostatina (Ho *et al.*, 2002). Esses dados, aliados a resultados previamente encontrados na literatura, mostram que a jararagina é capaz de liberar componentes endógenos bioativos e fatores anti-angiogênicos, o que contribuiria para uma resposta regenerativa de danos locais em tecidos após a mordida da cobra. Entretanto, a principal implicação desses estudos é o de que a jararagina parece possuir propriedades anti-tumorais (Correa *et al.*, 2002; Ho *et al.*, 2002).

### Anticorpos e inibidores da jararagina

Dada a importância da jararagina, assim como a de outras SVMPs, em patologia de venenos, diversas estratégias têm sido elaboradas para inibir ou neutralizar essa família de proteínas, com a finalidade de desenvolver um tratamento mais efetivo para o combate dos efeitos locais provenientes da mordida, assim como a procura por inibidores endógenos às SMVPs presentes no soro de certos mamíferos e répteis, naturalmente resistentes ao veneno de serpentes (Domont *et al.*, 1991; Fox e Bjarnason *et al.*, 1998). Esses inibidores, por exemplo, poderiam ser retirados do próprio soro de *B. jararaca* ou de gambás (Rocha *et al.*, 2002; Jurgilas *et al.*, 2003).

O inibidor de SVMPs, BJ46a, similar aos membros da superfamília de proteínas cistatinas, isolado a partir do soro da *B. jararaca*, dá origem a um complexo formado por ligações não covalentes com a jararagina, acarretando na inibição de suas atividades hemorrágicas e catalíticas (Valente *et al.*, 2002). Similarmente, o DM 43 (Neves-Ferreira *et al.*, 2002) e o PO41 (Jurgilas *et al.*, 2003), inibidores de SVMPs, são incluídos na superfamília das imunoglobulinas, isoladas a partir dos soros de *Didelphis marsupialis* e *Philander opossum*, respectivamente. Ambos formam complexos estáveis com a jararagina e inibem sua atividade hemorrágica, fibrinolítica e proteolítica.

Outros exemplos de inibidores da jararagina são antivenenos que vêm se mostrando bastante eficazes na inibição da atividade hemorrágica, induzida por jararagina. Um exemplo, soros extraídos de coelhos, imunizados com veneno de *B. jararaca*, foram capazes de neutralizar as toxinas responsáveis pela atividade hemorrágica de venenos de outras espécies de serpentes *Bothrops* (Lopes-Ferreira *et al.*, 1992). Entretanto, sabe-se que o desenvolvimento de antivenenos com anticorpos mais específicos pudessem atuar diretamente nas toxinas mais importantes, o que

contribuiria, de maneira muito significativa, na eficácia da soroterapia (Theakston e Reid, 1993; Harrison *et al.*, 2002). Um total de sete anticorpos monoclonais, cultivados contra a jararagina, reagiram preferencialmente com a jararagina-C. O anticorpo monoclonal (MAJar3) inibiu as interações jararagina/colágeno, atividades proteolíticas para peptídeos e substratos, e, ainda, a atividade hemorrágica induzida pela jararagina (Tanjoni *et al.*, 2003a).

Ao estabelecer a ligação de anticorpos anti-jararagina-C aos domínios desintegrina e rico em cisteína, espera-se que as interações da molécula com as integrinas  $\alpha_2\beta_1$  fiquem comprometidas (Clissa *et al.*, 2006). Esse fato contribuiria para a observada inibição da hemorragia dermal (Harrison *et al.*, 2000). Entretanto, a evidência de que a ligação da MAJar3 ao domínio desintegrina neutraliza ambas as atividades hemorrágicas e de ligação ao colágeno; sugere que sítios no domínio desintegrina sejam importantes na interação com substratos presentes na matriz extracelular, permitindo, assim, que as atividades enzimáticas aconteçam (Laing e Moura-da-Silva, 2005).

Como pode ser observado, jararagina tem sido amplamente estudada e explorada em seus mais diversos aspectos medicinais, nos casos de acidentes ofídicos, assim como ferramenta para o entendimento do mecanismo molecular de ação para essa classe de toxinas. Apesar de toda a informação acumulada para as atividades biológicas de SVMPs da classe P-III, pouco se sabe sobre os mecanismos moleculares de adesão e propriedades dos domínios D e C.

### **1.7 – Contextualização do projeto desenvolvido**

Embora seja de conhecimento as importantes atividades biológicas das SVMPs P-III, o envolvimento de cada domínio nessas atividades ainda não é totalmente conhecido e o estudo estrutural dessas moléculas pode trazer importantes

informações sobre a relação estrutura-função dessas proteínas. Nesse sentido, foram iniciados estudos estruturais de uma proteína da classe P-III, isolada de veneno de *B. jararaca*, em colaboração com a Dra. Ana Maria Moura-da-Silva, do Instituto Butantan, que nos forneceu amostra purificada de acordo com Paine (1992). Segundo a literatura, a proteína purificada de acordo com Paine é a jararagina. Testes de cristalização foram realizados com a proteína e os cristais obtidos foram coletados (Souza *et al.*, 2000). Através de substituição molecular foi possível determinar a estrutura do domínio catalítico, porém o refinamento da estrutura ficou prejudicado em função de não ser possível determinar a estrutura dos domínios desintegrina e rico em cisteína. Foi dada continuidade ao estudo, e, para solucionar o problema da degradação proteolítica da amostra, a purificação da proteína começou a ser realizada no laboratório, com um protocolo diferente daquele utilizado anteriormente descrito por Paine. A proteína isolada apresenta a mesma massa molecular aparente da jararagina e também apresenta atividade hemorrágica. Cristais desta proteína foram obtidos e a estrutura da molécula inteira foi determinada, observando-se, pela estrutura, que se trata da bothropasina, uma proteína homóloga à jararagina. Com a estrutura da bothropasina elucidada foi possível determinar a estrutura da proteína anteriormente cristalizada. Constatou-se que as duas estruturas determinadas, com proteínas purificadas de maneiras diferentes, são de bothropasina, fato que sugere que os diferentes protocolos de purificação não são suficientes para a obtenção específica de jararagina ou bothropasina.

Um fato interessante é que todos os estudos de atividade da jararagina, descritos nos tópicos anteriores, foram realizados com proteína isolada segundo Paine (1992).

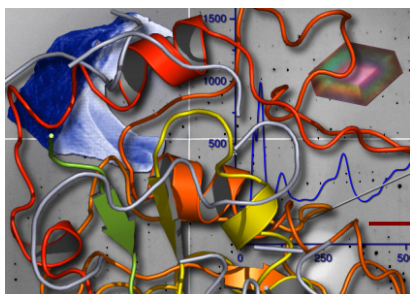


## OBJETIVOS

O principal objetivo do presente trabalho foi o estudo estrutural de uma metaloproteinase/desintegrina, proteína da classe P-III das SVMs, isolada do veneno de *B. jararaca*.

Objetivos específicos:

- Desenvolver protocolos de cristalização da proteína isolada de veneno bruto, da serpente *Bothrops jararaca*.
- Construir e analisar a estrutura tridimensional da proteína.
- Comparar e correlacionar a estrutura da proteína com proteínas homólogas.
- Estabelecer um mecanismo de ação para essa classe de proteínas multimodulares.



## **PURIFICAÇÃO, CRISTALIZAÇÃO E REFINAMENTO DA BOTHROPASINA**

### **Obtenção da proteína bothropasina**

O veneno bruto de *Bothrops jararaca*, na forma liofilizada, gentilmente cedido pelo Laboratório de Herpetologia do Instituto Butantan, na cidade de São Paulo, foi obtido, a partir de diversas amostras extraídas de diferentes exemplares de serpentes. A amostra liofilizada foi mantida a 4°C até o uso.

A bothropasina foi isolada a partir de 50 mg do veneno liofilizado de *B. jararaca* ressuspenso em 1 mL da solução tampão 20 mM Tris-HCl, pH 8,0, centrifugados a 12000 x g durante 15 minutos e submetidos a cromatografia de troca iônica. Para tanto, a coluna DEAE-*Sepharose Fast Flow* (Pharmacia, volume de 70 mL) foi previamente equilibrada com a mesma solução tamponante, acoplada a um sistema ÄKTA Explorer 10 (GE).

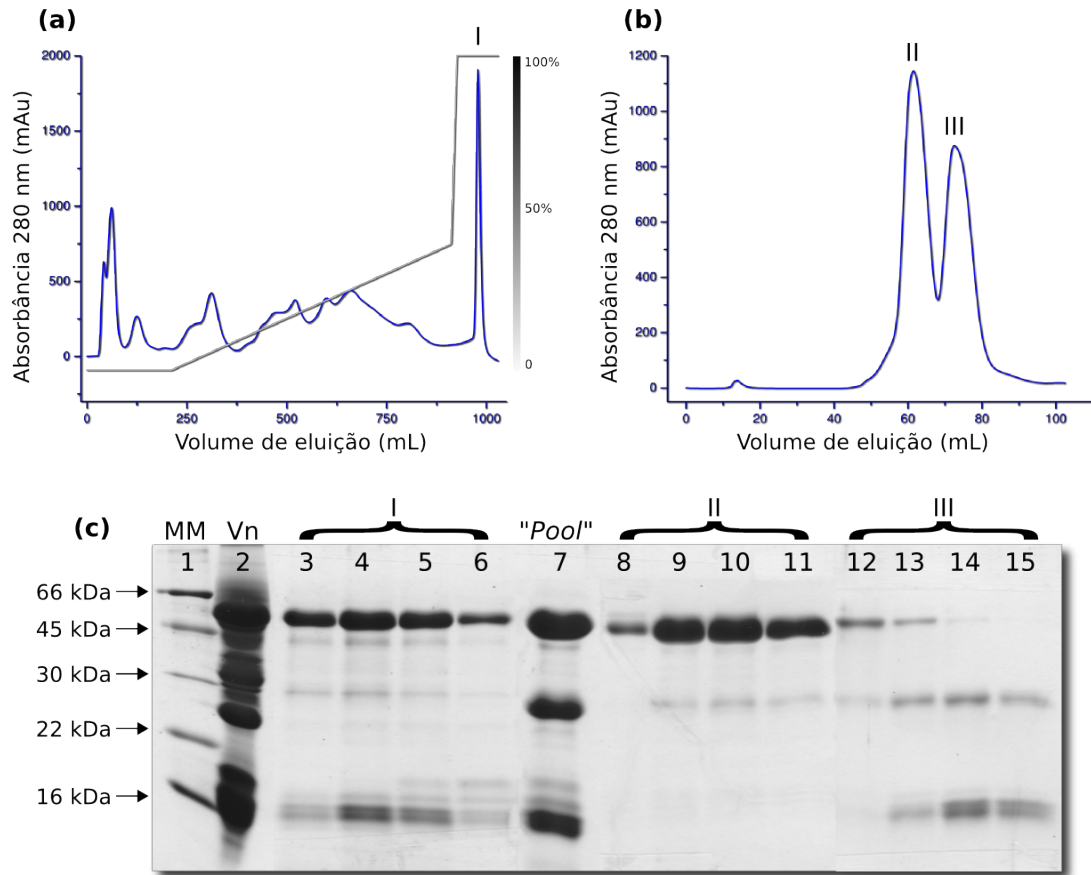
Após a remoção das proteínas não aderidas à resina da coluna com, aproximadamente, 210 mL do tampão de equilíbrio, foi efetuado o processo de eluição. A fase eluente utilizada foi um gradiente composto por três estágios: (1) gradiente linear de 10 vezes o volume da coluna (total de 700 mL) a 40% de NaCl 0,5 M em mesmo tampão; (2) rápido gradiente linear de 40 a 100% de NaCl 0,5 M em 20 mM

Tris-HCl, pH 8,0 e (3) 120 mL de solução a 100% de NaCl 0,5 M em 20 mM Tris-HCl, pH 8,0 (figura 3.1 (a)).

A cromatografia foi realizada em um fluxo constante de 3 mL.min<sup>-1</sup> e frações contendo proteína com massa molecular aproximada de 55 kDa (estimada por SDS-PAGE, figura 3.1 (c)), foram reunidas e a amostra foi dialisada durante 18 horas a 4°C contra tampão Tris-HCl 20 mM, pH 8,0 com 250 mM NaCl, concentrada a 1 mL e aplicada em uma coluna de exclusão por massa molecular Superdex75 (Pharmacia, 1,6 cm x 60 cm), equilibrada com o mesmo tampão da diálise. A eluição das proteínas foi realizada a um fluxo constante de 0,5 mL.min<sup>-1</sup> e foram coletadas frações de 1mL. As cromatografias foram monitoradas pela leitura da absorbância a 280 nm e todas as etapas foram efetuadas a baixa temperatura (4°C). As figuras 3.1 (a), (b) e (c), ilustram, de maneira resumida, as etapas envolvidas no processo de obtenção da bothropasina

As frações obtidas de bothropasina foram analisadas quanto ao seu grau de pureza, por eletroforese em gel de poliacrilamida, na presença de SDS, conforme descrito por Laemmli (Laemmli, 1970) (figura 3.1 (c)). A atividade biológica foi estudada através de ensaios de atividade hemorrágica que foram realizados em colaboração com a Profa. Dra. Heloisa Selistre-de-Araújo, do Departamento de Ciências Fisiológicas da Universidade Federal de São Carlos. A atividade hemorrágica foi determinada pela Dose Mínima Hemorrágica (DMH), que é definida como a quantidade mínima de enzima necessária para produzir um halo hemorrágico de 1cm de diâmetro na pele de animais como coelho ou camundongos, duas horas após a injeção (Kondo *et al.*, 1960). Solução de bothropasina (100 µL) foi injetada intradermicamente em camundongos de cerca de 25 g. Duas horas após a injeção, os camundongos foram sacrificados e as peles removidas para determinação do halo

hemorrágico. A DMH para a bothropasina foi de 20  $\mu\text{g}$  o que está de acordo com o relatado na literatura (Mandelbaum, 1982).



**Figura 3.1: Purificação da bothropasina.** (a) 50 mg de veneno bruto ressuspenso em 1mL de solução tampão 20 mM Tris-HCl, pH 8,0 e aplicado a cromatografia de troca iônica em coluna DEAE-*Sepharose FF*. A curva cinza representa o gradiente salino. (b) Cromatografia das frações eluídas em (a) "I" em coluna de exclusão por massa molecular Superdex75. A bothropasina elui com um grau de pureza de aproximadamente 95% (frações "II"), suficientes para os testes de cristalização. (c) Gel de poliacrilamida 15% indicando as etapas mais relevantes do processo de purificação da bothropasina. Poço 1: marcador de massa molecular; Poço 2: veneno bruto; Poços 3-6: frações eluídas com 100% NaCl 0,5 M 20 mM Tris-HCl, pH 8,0; Poço 7: combinação das amostras 3-6 aplicado na coluna Superdex75; Poços 8-11: frações eluídas da coluna (pico II) com bothropasina utilizada nos testes de cristalização; Poços 12-15: frações com contaminantes eluídos no pico III da coluna Superdex75.

A concentração protéica das amostras provenientes da cromatografia foi estimada através do método descrito inicialmente por Bradford (Bradford, 1976). A



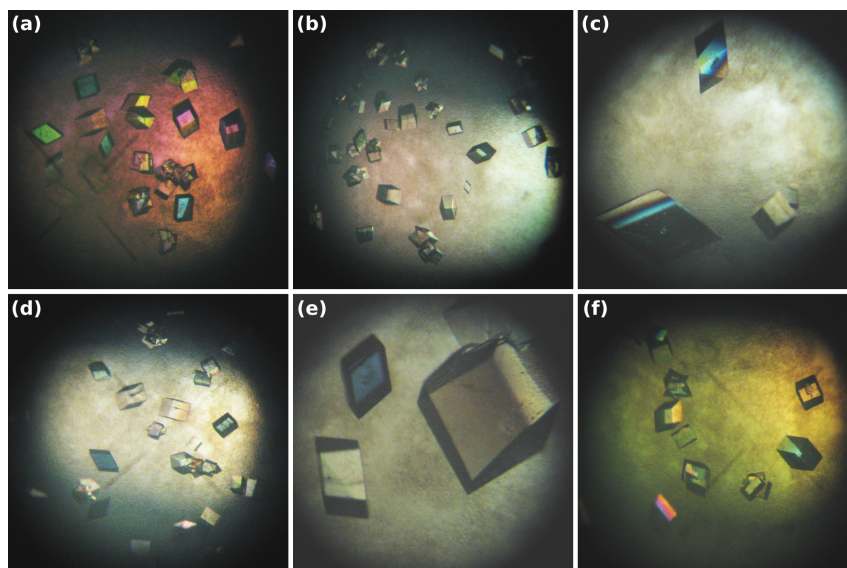
amostra passou por mais um processo de diálise (aproximadamente 18 horas) contra a solução tamponante fosfato de sódio 50 mM em pH 6,0.

### **Obtenção da bothropasina sob forma cristalina**

Visando ao aumento da estabilidade protéica através da inibição da atividade enzimática e, dessa forma, contribuindo para a qualidade na formação da rede cristalina, o inibidor não específico POL647 (Gomis-Ruth *et al.*, 1998) foi encubado em uma razão molar proteína/inibidor igual a 1:50 (POL647 8 mM) por cerca 120 min à temperatura de 4°C. Posteriormente, a proteína foi concentrada para aproximadamente 8 mg.mL<sup>-1</sup> (160 µM) e utilizada em experimentos de cristalização, com o emprego do método de difusão de vapor com gotas suspensas (Hampel *et al.*, 1968).

Em uma tentativa inicial para obtenção de cristais, foi realizada uma triagem, através de conjuntos de soluções comerciais de cristalização do tipo matrizes esparsas “*Crystal Screen*” I e II (Hampton Research) a 4°C e 18°C. Os ensaios consistiam de gotas suspensas de volume igual a 8 µL (50% proteína e 50% solução mãe) e reservatório contendo 1 mL de solução mãe.

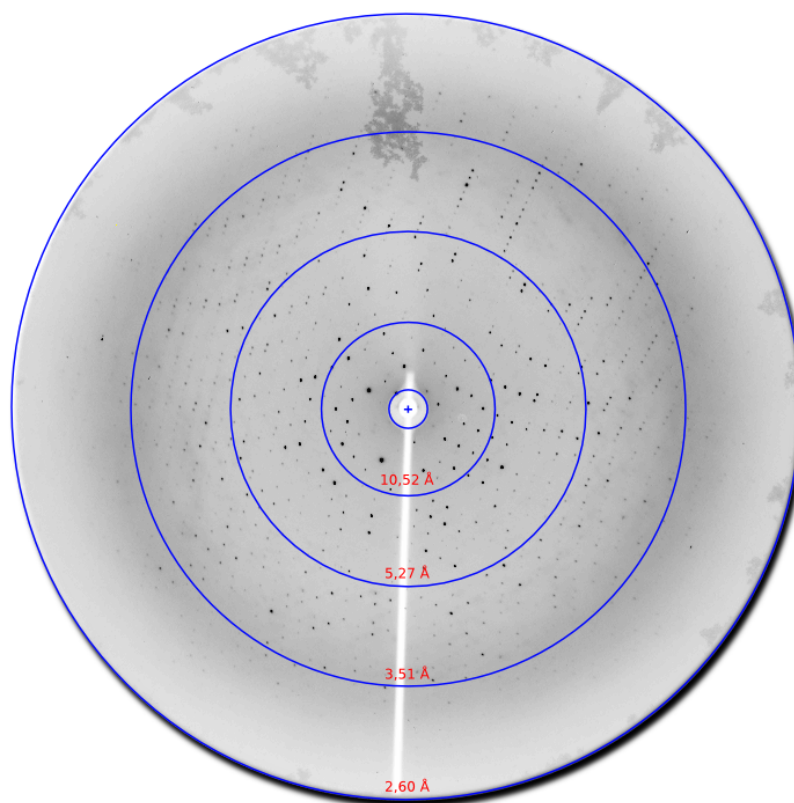
Após um período de 5 dias, foram observados cristais protéicos em diversas condições nas matrizes esparsas. Varreduras foram feitas nas condições que se mostraram mais promissoras. A condição ideal de cristalização foi estabelecida como sendo: acetato de amônia 40 mM, solução tamponante de acetato de sódio 100 mM, pH 4.6 e 29% PEG 4000 como agente precipitante a temperatura de 4°C (figura 3.2).



**Figura 3.2: Cristais da proteína bothropasina.** As imagens de (a)-(f) mostram as melhores condições para a obtenção de cristais de bothropasina. Os cristais apresentam uma morfologia semelhante, porém algumas formas como trapézio e cúbica também são observadas. Algumas condições com volume menor de agente precipitante, figuras (c) e (e), os cristais apresentaram um crescimento acima do esperado, mas com baixo poder de difração. O tamanho médio dos cristais de bothropasina é de 0,8 x 0,5 x 0,6 mm.

### Coleta de dados de difração

Vários cristais foram testados, sempre em temperatura criogênica (100 K, crioprotetidos com adição de glicerol 5% (v/v) à solução mãe), e 8 conjuntos de dados foram coletados em uma fonte de raios X anôdo rotatório, disponível no laboratório, modelo ultraX 18 (RIGAKU/MSU) com detector do tipo placa de imagem, modelo Mar345dtb (MAR Research). Um conjunto de dados de mais alta resolução foi coletado na linha MX1 (Polikarpov *et al.*, 1998), no Laboratório Nacional de Luz Síncrotron (LNLS/Campinas), equipado com placa MARccd (MAR Research) utilizando um comprimento de onda igual a 1,4027 Å. Um exemplo da imagem de difração medida no LNLS é mostrada na figura 3.3.



**Figura 3.3: Padrão de difração do cristal de bothropasina.** Imagem característica obtida de um cristal congelado após imersão por, aproximadamente, 2 min em solução de NaI 0,5 M. Imagem com 01 grau de oscilação com tempo de 90 s de exposição.

Todos os conjuntos de imagens de difração foram processados no programa MOSFLM (Leslie, 1992) e as intensidades foram escalonadas e reduzidas no programa SCALA (Evans, 1993), que inclui o programa TRUNCATE (French e Wilson, 1978), ambos do pacote de programas CCP4 (Collaborative Computational Project Number 4 (1994)). Os parâmetros e estatísticas do conjunto de dados processados são mostrados na tabela 3.1.

**Tabela 3.1:** Parâmetros e estatísticas da coleta e processamento dos dados de difração de raios X para a proteína na presença de NaI (cristais crescidos a 4°C).

Valores em parênteses são referentes à camada mais externa de resolução	
Grupo espacial	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Parâmetros da rede	a= 72,5 Å, b= 101,2 Å, c= 133,6 Å
Limites de resolução (Å)	20,0 – 2,85 (3,00 – 2.85)
Número de imagens	200
$\Delta\phi$ (°)	1
Mosaicidade (°)	0,7
Reflexões únicas	23552 (3384)
Multiplicidade	5,7 (5,8)
R <sub>sym</sub> (%)	7,3 (32,4)
$\langle I/\sigma(I) \rangle$	17,9 (4,5)
Completeza (%)	99,7 (100,0)

O número de moléculas presentes na unidade assimétrica foi estimado por meio do programa Matthews\_coef (parte do pacote CCP4), que faz uso da metodologia proposta por Matthews (Matthews, 1968).

Considerando-se a massa molecular da bothropasina, como sendo aproximadamente 52 kDa, foram estimadas duas moléculas como constituintes da unidade assimétrica cristalina, resultando em um volume de Matthews equivalente a 2,6 Å<sup>3</sup>/Da e conteúdo de solvente no cristal de 49,8%. O fator de temperatura (B) do cristal encontrado por regressão linear no gráfico de Wilson foi de 80,7 Å<sup>2</sup>.

### Faseamento e construção do modelo cristalográfico

A dificuldade na obtenção de cristais de metaloproteinases/desintegrinas (MD) da classe P-III, provavelmente devido à sua complexidade e flexibilidade em seus multidomínios, tem como consequência a pouca informação estrutural de proteínas dessa classe. Somente em 2006 foi determinada a primeira estrutura de uma MD isolada do veneno de *Crotalus atrox* (Takeda *et al.*, 2006). O elevado número de

pontes dissulfeto, que torna a produção de formas recombinantes da proteína um processo bastante laborioso, a presença de poucas formas homogêneas no veneno (Moura-da-Silva *et al.*, 2003) e a necessidade do uso de um inibidor específico, capaz de reduzir a auto-degradação protéica, também colaboram para a dificuldade de se obter cristais.

Através do método de “*quick cryo-soaking*” (Dauter *et al.*, 2000; Nagem *et al.*, 2001; 2003), cristais protéicos foram imersos em solução de 0,5 M de NaI (dissolvidos em 100 mM, pH 4.6 e 29% PEG 4000), por um período de aproximadamente 3 min, visando à obtenção de melhores fases experimentais

As fases foram determinadas com a utilização de um protocolo misto de faseamento, implementado no programa PHASER<sup>1</sup> (McCoy *et al.*, 2004; 2005), que faz uso das técnicas de Substituição Molecular (SM), combinado com faseamento por dispersão anômala. No presente trabalho, a dispersão anômala foi a um único comprimento de onda (SAD).

Os modelos de busca para a SM foram encontrados através de alinhamento seqüencial entre a bothropasina e as proteínas depositadas no PDB “*Protein Data Bank*”. Os melhores alinhamentos foram obtidos para a acutolisina (PDB 1BSW) uma metaloprotease de classe P-I e trimestatina (PDB 1J2L) uma desintegrina RGD.

A posição dos sítios de iodo foi determinada com auxílio do programa PHASER. Um total de 8 sítios de iodo, com ocupação acima de 80%, foi encontrado no conjunto de dados. As fases iniciais foram então calculadas utilizando-se as coordenadas dos átomos de iodo, assim como as moléculas homólogas em um processo misto de faseamento. Os mapas iniciais calculados não permitiram o traçado

---

1 A versão utilizada do programa PHASER no presente trabalho foi gentilmente cedida e também teve a participação efetiva, durante todo o processo de faseamento, pelo Prof. Randy Read do “*Medical Research Council*” – Universidade de Cambridge, UK, durante o desenvolvimento do meu Programa de Doutorado no País com Estágio no Exterior (PDEE) na mesma universidade. Maiores informações podem ser encontradas no apêndice I.

automático nem mesmo manual para os átomos das cadeias principais e laterais do domínio rico em cisteína. Através de ciclos de modificação de densidade eletrônica e achatamento de solvente utilizando o programa RESOLVE (Terwilliger, 2003), fragmentos do domínio C puderam ser traçados para as duas moléculas presentes na unidade assimétrica.

Após aplicação da solução por SM, foram feitos refinamento de corpo rígido, anelamento simulado, com o pacote de programas CNS (Brunger *et al.*, 1998) e PHENIX (Adams *et al.*, 2004). Foram feitos ciclos de refinamento no espaço real (modificações manuais do modelo, através da inspeção dos mapas de densidade eletrônica usando o programa gráfico Coot (Emsley e Cowtan, 2004)), alternados com cálculos intermediários no espaço recíproco para refinamento de posições atômicas, restrições geométricas e de contato, refinamento de fator de temperatura isotrópico individual para todos os átomos. Além de minimização de funções de energia e TLS. Foram utilizados os programas Refmac 5.0 (*Collaborative Computational Project, Number 4*) e PHENIX\_refine (Afonine *et al.*, 2005).

Ciclos exaustivos de construção e reconstrução do modelo estrutural, aliados à estrutura primária da bothropasina (Mandelbaum *et al.*, 1982) e a descrição das pontes dissulfeto na literatura para a catrocolastatina (uma SVMP de *C. atrox* (Calvete *et al.*, 2000)), tornaram possível a construção de 100% da estrutura da bothropasina. O monitoramento da coerência do modelo com os dados experimentais foi conduzido através da análise da variação do fator R, juntamente com o fator  $R_{\text{free}}$  (Validação Cruzada, Brunger *et al.*, 1998).

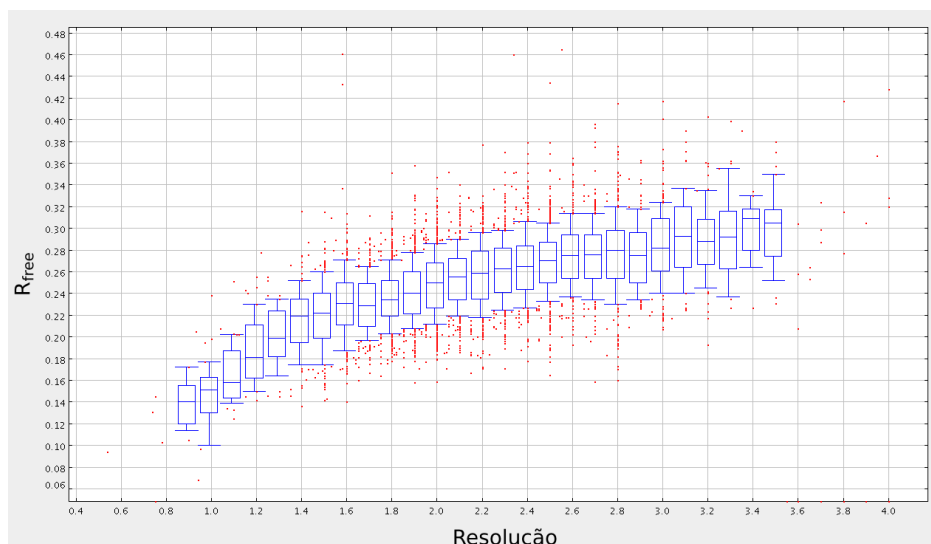
### **Validação da estrutura da bothropasina**

Foram utilizadas 23552 reflexões no cálculo do  $R_{\text{factor}}$  para o “conjunto de trabalho” e 1201 reflexões para o “conjunto teste”, ou seja, 5,1% do conjunto total,

para o cálculo do  $R_{\text{free}}$ , compreendendo uma faixa de resolução de 20,0 a 2,8 Å. Os valores finais desses índices de discrepância ficaram em 24,9% ( $R_{\text{factor}}$ ) e 29,3% ( $R_{\text{free}}$ ). A introdução de moléculas de água foi feita manualmente, com o auxílio do programa Coot, sob critérios de distância mínima e máxima de 2,3 Å a 3,5 Å, fator de temperatura menor que 50 Å<sup>2</sup> e em mapa de densidade eletrônica com nível de contorno de  $1\sigma$ . O refinamento foi dado como encerrado quando não foi mais possível a melhora dos índices  $R_{\text{factor}}$  e  $R_{\text{free}}$ . O modelo final é constituído por 2 moléculas na unidade assimétrica (com um total de 833 resíduos de aminoácidos), 2 moléculas de POL647 e 103 moléculas de água.

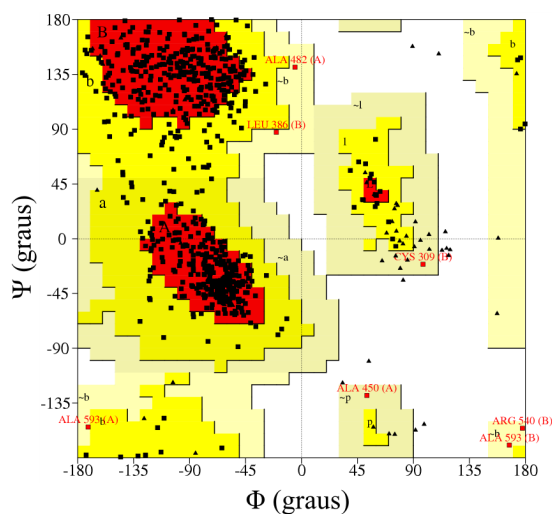
Os valores finais para os desvios médios quadráticos para comprimentos e ângulos de ligações interatômicas da estrutura foram 0,015 Å e 1,815° respectivamente. O fator de temperatura médio para todos os átomos das cadeias protéicas foi de 57,2 Å<sup>2</sup>.

De forma a ilustrar a qualidade dos valores finais de  $R$  e  $R_{\text{free}}$  da bothropasina, a figura 3.4 mostra a média dos valores observados para estruturas refinadas à mesma resolução, ou seja, 2,8 Å. Dadas as características da proteína, como ausência de estruturas secundárias bem definidas para uma boa parte da estrutura, resolução e ainda por ter um de seus domínios construídos manualmente, as estatísticas de  $R$  e  $R_{\text{free}}$  podem ser consideradas como bastante satisfatórias e, como mostrado na figura 3.4, dentro de valores aceitáveis.



**Figura 3.4: Valores de  $R_{\text{free}}$  para estruturas depositadas no PDB em função da resolução.** O valor de  $R_{\text{free}}$  esperado para uma estrutura refinada a 2,8 Å, de acordo com estatísticas do PDB, deveria estar entre aproximadamente 26 e 30%. A bothropasina teve valor final de  $R_{\text{free}}$  igual a 29,3%, portanto, dentro das estatísticas do PDB. Figura adaptada de “Harry Plotter”<sup>2</sup>.

A análise do diagrama de Ramachandran (figura 3.5), gerado pelo PROCHECK (Laskowski *et al.*, 1993), mostra que 671 resíduos (80,0%) encontram-se nas regiões mais favoráveis, 158 (18,9%) estão em regiões adicionais permitidas, 9 resíduos (1,1%) em regiões generosamente permitidas e nenhum resíduo na região desfavorável.

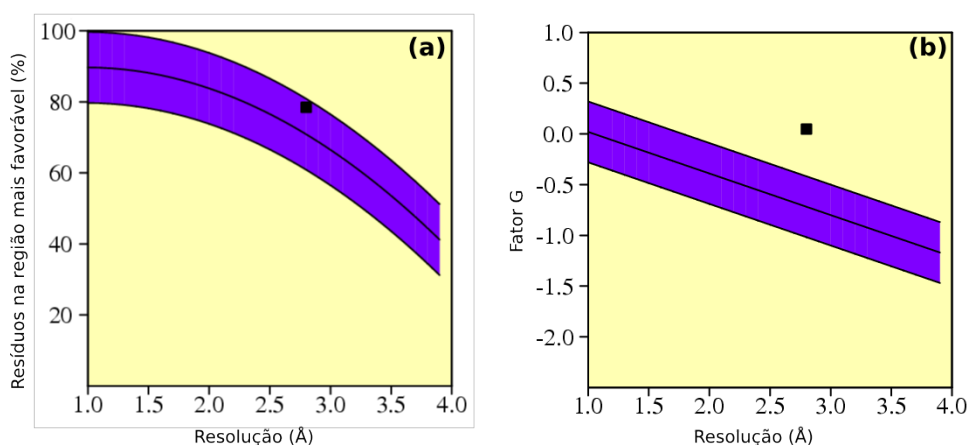


**Figura 3.5.: Diagrama de Ramachandran do modelo cristalográfico final.** É avaliada a qualidade da molécula através da distribuição (e combinação) dos ângulos torsionais diedrais ( $\Phi$  e  $\Psi$ ). O resultado indica que 80% dos resíduos se encontram nas regiões mais favoráveis (em vermelho), 18,9% estão em regiões adicionalmente permitidas (em amarelo), 1,1% em regiões generosamente permitidas (em bege) e nenhuma ocupação na região desfavorável (em branco).

<sup>2</sup> <http://xray.bmc.uu.se/gerard/supmat/rfree2000/plotter.html>. Acesso em 14 de julho de 2007.



Os parâmetros estereoquímicos da cadeia principal foram comparados pelo PROCHECK a estruturas cristalográficas de mesma resolução (figura 3.6). A figura 3.6 (a) mostra a porcentagem de resíduos que ocupam a região mais favorável do diagrama de Ramachandran, em função da resolução de seu conjunto de dados. Já o gráfico mostrado na figura 3.6 (b), mostra o valor médio do fator G (medida da normalidade estereoquímica da estrutura), de cada resíduo na estrutura. Nesses gráficos, a região destacada em roxo corresponde a faixa de resultados para essas estruturas e a linha central representa o mínimo quadrático para a inclinação média em função da resolução, e a correspondente largura da banda indica a variação de um desvio sobre a média.



**Figura 3.6: Estatísticas geradas pelo programa PROCHECK.** As análises indicam que o modelo cristalográfico apresentado neste trabalho apresenta qualidade estereoquímica superior quando comparado a outras estruturas determinadas à mesma resolução. Em (a) observa-se que a quantidade obtida de, aproximadamente, 80% de resíduos na região mais favorável do diagrama de Ramachandran está acima da média e do desvio padrão na resolução de 2,8 Å. O fator G de normalidade estereoquímica, de acordo com a figura (b), apresenta-se muito superior, em relação à média das estruturas na mesma faixa de resolução.

Através das figuras 3.6 (a) e (b), pode-se observar que, de maneira geral, os resultados pós-refinamento obtidos são condizentes e dentro do esperado. Dada a dificuldade de interpretação dos mapas, o processo de traçar as cadeias principais e

laterais dos resíduos, dentro das densidades eletrônicas, é mais laborioso. Com uma média resolução de 2,8 Å, o número de distorções nos ângulos  $\Phi$  e  $\Psi$ , quando comparado a outras estruturas determinadas na mesma resolução, encontra-se acima da média (próximo dos 80%). O mesmo vale para a qualidade estereoquímica do modelo. O fator G apresenta-se significativamente superior à grande maioria, ilustrando a possibilidade de se obter valores aceitáveis para R e  $R_{\text{free}}$ , mantendo-se uma boa estereoquímica.

### **Determinação da estrutura 'nativa'**

O projeto inicial visando à determinação da estrutura cristalográfica de uma P-III de *Bothrops jararaca* (no caso, a jararagina) foi iniciado pela doutora Dulce Helena Ferreira de Souza, supervisora do atual projeto, no início da corrente década. A partir do fornecimento do veneno bruto pelo mesmo laboratório e seguindo o protocolo de purificação de Paine (1992), foi obtido somente um único cristal da proteína nativa.

Dados de difração de raios X foram coletados na linha CPr (hoje MX1) do LNLS, cujas estatísticas se encontram descritas na tabela 3.2.

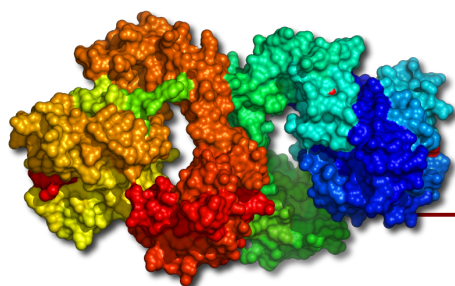
Como auxílio da SM e utilizando a estrutura cristalográfica da acutolisina como molécula molde, o domínio M foi determinado. No entanto, devido à ausência de modelos cristalográficos de proteínas homólogas aos domínios D e C, e da impossibilidade de se conseguir um conjunto de dados a partir de cristais derivados, a construção de 100% da estrutura foi impossibilitada. Como restava metade da molécula para ser construída, os mapas de densidade eletrônica para as cadeias laterais do domínio M não eram bons o suficiente para a completa elucidação da estrutura deste domínio.

**Tabela 3.2:** Parâmetros e estatísticas da coleta e processamento dos dados de difração de raios X para a proteína 'nativa' (cristais crescidos a 4°C).

Valores em parênteses são referentes à camada mais externa de resolução	
Grupo espacial	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Parâmetros da rede	a= 72,5 Å, b= 101,2 Å, c= 133,6 Å
Limites de resolução (Å)	40,1 - 2,70 (2,85 - 2,70)
Número de imagens	105
$\Delta\phi$ (°)	0,7
Mosaicidade (°)	0,6
Reflexões únicas	25611 (3731)
Multiplicidade	2,9 (2,9)
R <sub>sym</sub> (%)	8,5 (35,0)
$\langle I/\sigma(I) \rangle$	12,5 (2,5)
Completeza (%)	93,4 (94,0)

De posse da estrutura obtida pelo método da derivatização, descrito nesta tese, foi possível a determinação do modelo 'nativo' usando o conjunto nativo, através da técnica de substituição molecular, com auxílio do programa PHASER.

Como pode ser observado para os dois conjuntos de dados ('nativo' e derivado), parâmetros do cristal, tais como a simetria e tamanho da célula unitária, são muito parecidos.

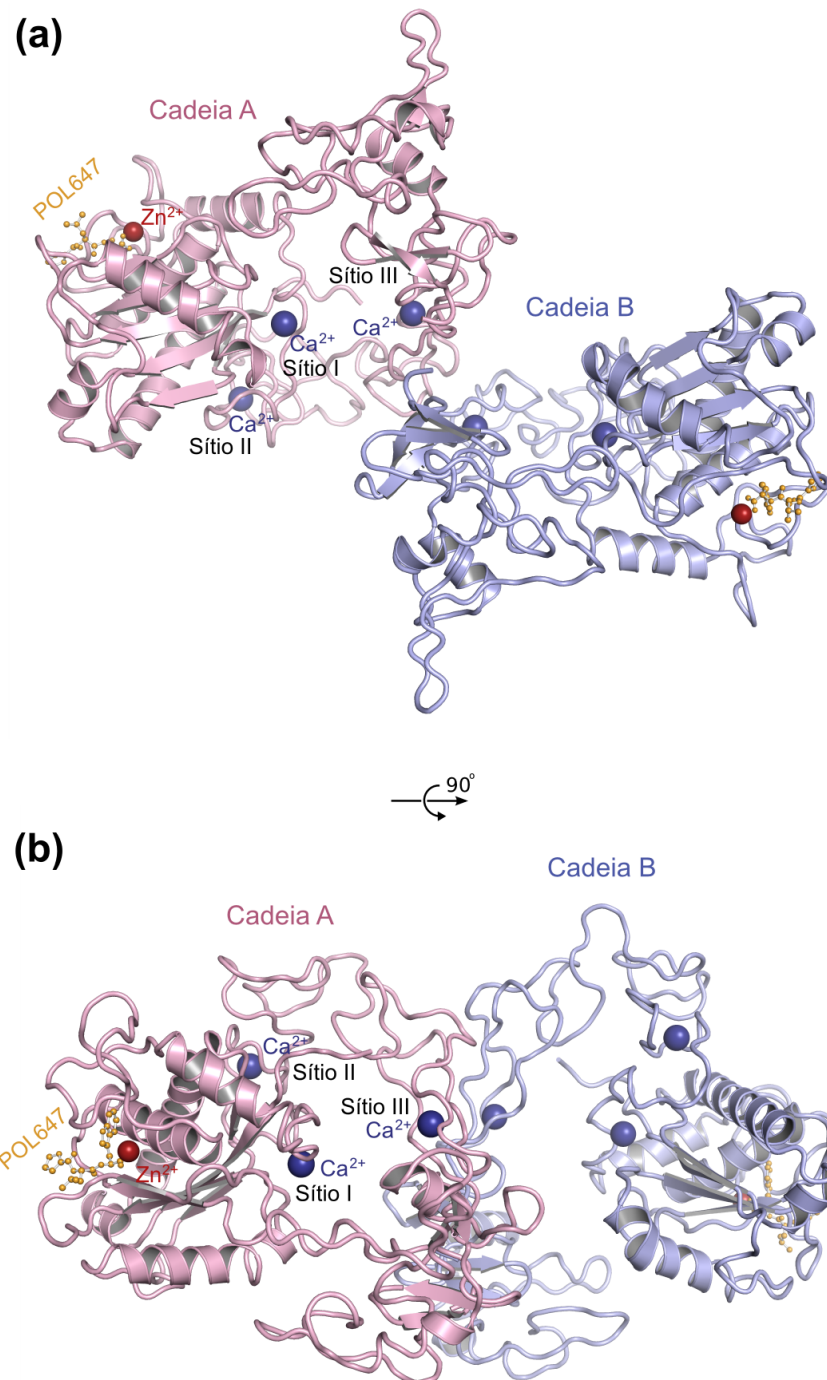


## **ANÁLISE ESTRUTURAL DA BOTHROPASINA E DISCUSSÕES**

### **Descrição do modelo cristalográfico**

Neste trabalho é apresentada a estrutura cristalográfica da bothropasina, presente no veneno da serpente *Bothrops jararaca*, em complexo com seu inibidor POL647, ilustrada na figura 4.1. O modelo final, determinado a 2,8 Å de resolução, contém 2 moléculas de bothropasina, em sua unidade assimétrica, o que totaliza 833 aminoácidos, 2 inibidores da auto-catálise POL647, 2 átomos de zinco e 6 íons cálcio ligados em seus respectivos sítios catalíticos e 103 moléculas de água (figura 4.1).

Dos 421 resíduos que compõem o monômero da bothropasina, apenas 2, um na extremidade N-terminal e outro na extremidade C-terminal, não puderam ser modelados devido à ausência de densidade eletrônica. Todos os átomos presentes foram modelados dentro das densidades eletrônicas bem interpretadas, com exceção de algumas poucas cadeias laterais de resíduos, na maioria das vezes presentes na superfície dos monômeros e expostos ao solvente, cujas conformações não puderam ser modeladas. Nestes casos, a ocupação para tais átomos foi igualada a zero.



**Figura 4.1: Composição da unidade assimétrica da célula unitária.**

(a) O cristal de bothropasina tem, em sua unidade assimétrica, dois monômeros de bothropasina, formando um dímero cristalográfico. O eixo de simetria de ordem 2 passa perpendicularmente ao plano da folha, entre as cadeias A e B, no meio da figura. O inibidor POL647 está representado em “balls and sticks”. Os íons de Ca<sup>2+</sup> e Zn<sup>2+</sup> estão representados por esferas azuis e vermelhas respectivamente. (b) Rotação de 90° no eixo indicado pela seta em relação à figura (a).

Devido à baixa resolução do conjunto de dados, o modelo foi refinado, inicialmente, com imposição de alta correlação por simetria não-cristalográfica (NCS) para todos os átomos, que foi liberado ao longo do processo de refinamento.

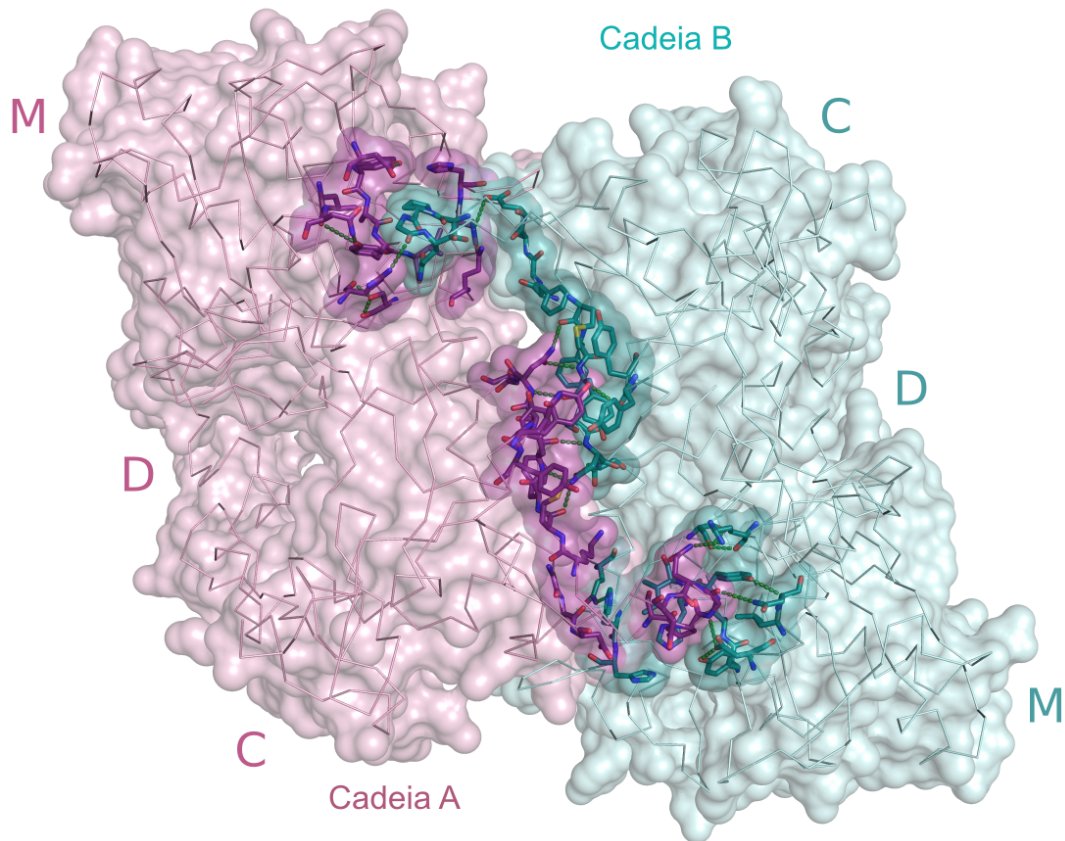
### **Contatos intermoleculares no cristal**

A presença de duas moléculas, na unidade assimétrica, implica em uma grande quantidade de solvente no cristal (aproximadamente 50% do volume total), o que contribui para o baixo poder de difração dos cristais, possibilitando uma maior desordem estática das moléculas.

A determinação e a descrição dos contatos intermoleculares foram feitas com o auxílio do servidor PISA<sup>3</sup> (Krissinel e Henrick, 2005). Por esse servidor, pode-se calcular a área de contato entre os dois monômeros, descrever quais resíduos realizam contatos polares (ligações de hidrogênio e pontes salinas) e hidrofóbicos, além de se determinar a acessibilidade ao solvente de cada um dos resíduos presentes na superfície da estrutura. Também é calculada a probabilidade, que varia de 0 a 1, de a interface analisada ter relevância na atividade biológica da proteína. Resumidamente, os contatos intermoleculares no cristal de bothropasina compartilham áreas pequenas (valor máximo de 973 Å<sup>2</sup> e mínimo de 92 Å<sup>2</sup>) com baixos números de contatos polares (em média 8 por interface; figura 4.2). Esses valores de área de contato estão muito abaixo daqueles previstos como sendo funcionalmente importantes (em média 1400 Å<sup>2</sup>, Lo Conte *et al.*, 1999), e não devem representar um papel biológico.

---

3 [http://www.ebi.ac.uk/msd-svd/prot\\_int/pistart.html](http://www.ebi.ac.uk/msd-svd/prot_int/pistart.html). Acesso em 21 de junho de 2007.

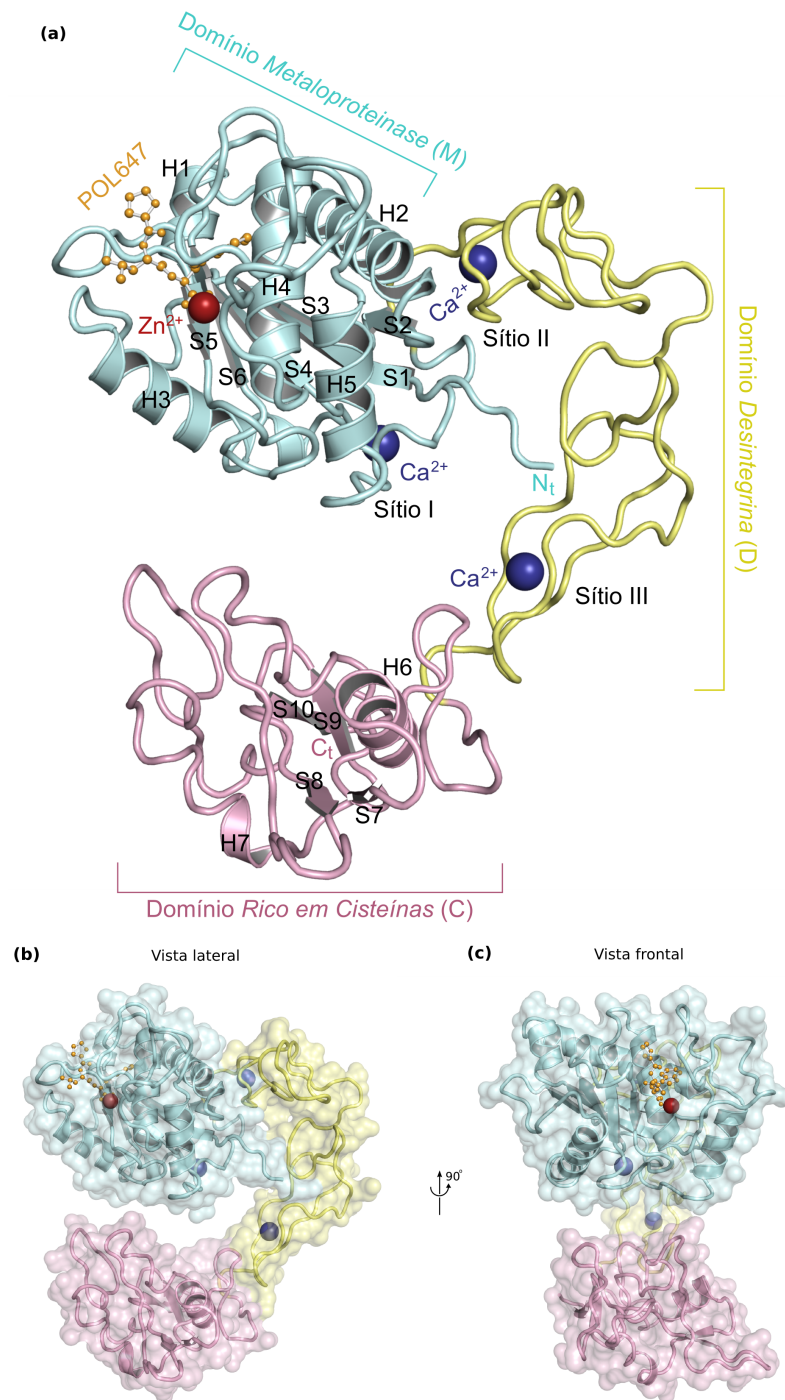


**Figura 4.2: Contatos intermoleculares do cristal.** As interações entre as duas moléculas se dão através de contatos polares (8 ligações de hidrogênio e uma ponte salina destacadas por linhas tracejadas na cor verde) e hidrofóbicos. As superfícies de van der Waals na região de interface estão destacadas em roxo (cadeia A) e azul (cadeia B) respectivamente. As cadeias A e B estão representadas por traços apenas para os átomos da cadeia principal e destacadas nas cores rosa e azul respectivamente. A superfície de van der Waals para os dois monômeros está em branco. M, D e C representam os domínios metaloprotease, desintegrina e rico em cisteína, respectivamente.

### Os domínios *MDC*

A estrutura da bothropasina é composta pelos domínios: N-terminal *Metaloproteínase* (M), o domínio *Desintegrina* (D) e pelo domínio C-terminal *Rico em Cisteína* (C), como esperado para proteínas pertencentes à classe P-III de SVMPs (figura 4.3). Devido à grande semelhança existente entre as cadeias A e B, por simples convenção, a cadeia A foi adotada como referência no decorrer deste trabalho.





**Figura 4.3: Estrutura da bothropasina complexada com o inibidor POL647.** (a) Visão lateral da estrutura revelando a presença de três domínios: o domínio catalítico (M); o domínio desintegrina (D) e o domínio rico em cisteína (C). (b) Bothropasina disposta da mesma maneira que em (a) e superfície de van der Waals para ressaltar a formação de um bolsão no interior da estrutura. (c) Representação da estrutura da bothropasina rotacionada em 90° em relação a (b) evidenciando, através da utilização da superfície de van der Waals a não interação existente entre o domínios (M) e (C), apesar de próximos. Os domínios M, D e C estão representados nas cores azul ciano, amarela e rosa, respectivamente.



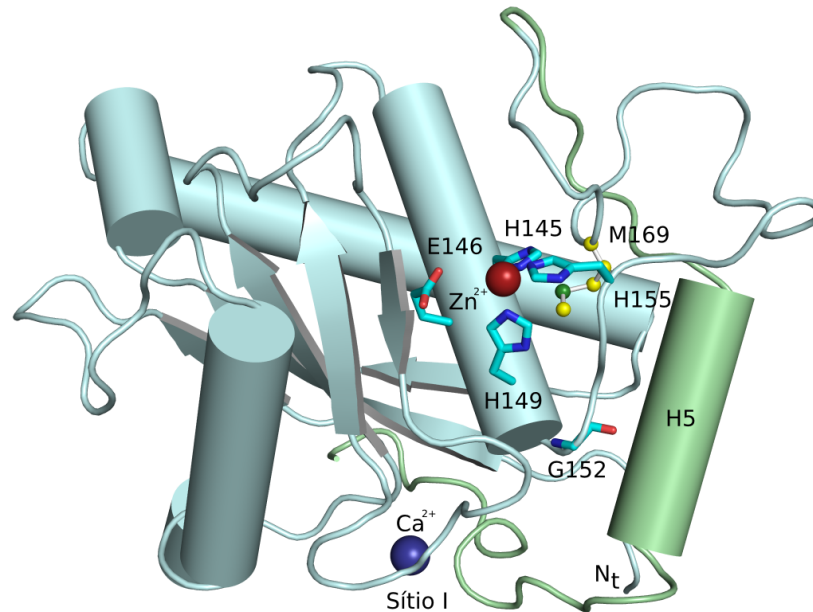
O domínio D tem início nas proximidades da região de ligação do íon  $\text{Ca}^{2+}$  do domínio catalítico M (“Sítio I” na figura 4.3 (a)), ou seja, na parte oposta à região de ligação do  $\text{Zn}^{2+}$  e, conseqüentemente, do inibidor POL647 (figura 4.3 (a)). O domínio D, aliado ao domínio localizado na porção carbóxi-terminal da estrutura, o domínio C, formam uma estrutura similar a um braço, em direção ao domínio M. Apesar de próximos, os domínios M e C não realizam nenhum tipo de interação direta entre as cadeias laterais dos resíduos que os compõem (figura 4.3 (b) e (c)). Também merece ser destacada a presença de um significativo espaço no interior (cavidade) na estrutura. A presença desses espaços, aliados à falta de interação entre os domínios M e C garantem uma maior flexibilidade à estrutura, além de aumentar a porcentagem de solvente presente no empacotamento cristalino, colaborando com o baixo poder de difração e aumento da desordem e mosaicidade no cristal.

Cada monômero contém 35 resíduos de cisteína, sendo que 34 deles fazem pontes dissulfeto e possuem espaçamento estritamente conservado dentro da classe P-III de SVMP. O único resíduo que não forma ponte dissulfeto é a Cys189 que será apresentado em maiores detalhes no decorrer deste trabalho.

#### **O domínio M: sítio ativo, sítio de ligação do inibidor, cálcio e açúcar.**

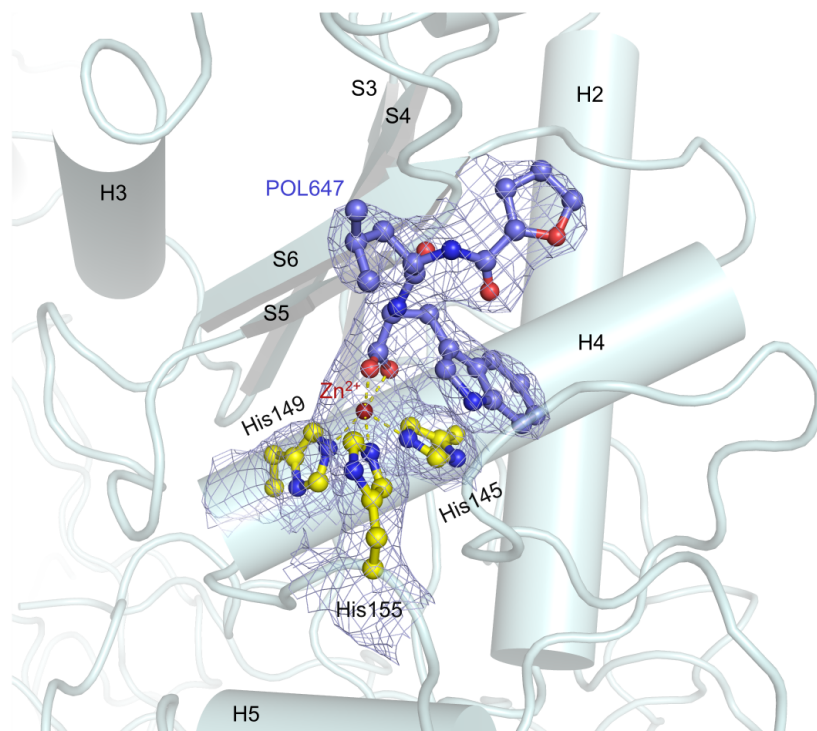
O domínio M consiste em dois subdomínios em um arranjo muito similar àqueles descritos SVMPs da classe P-I como a adamalisina II e acutolisina (Gomis-Ruth *et al.*, 1993a; 1993b). O principal subdomínio apresenta uma arquitetura com motivo  $\alpha/\beta$ , formando uma folha  $\beta$ , composta por 5 fitas paralelas (fitas S1,2, 3, 4 e 6), e uma fita anti-paralela (fita S5). Além das fitas, o domínio também apresenta 4 hélices- $\alpha$  (H1, 2, 3 e 4). O menor subdomínio consiste em uma hélice- $\alpha$  (H5) e “loops” (figura 4.4). O motivo altamente conservado HExxHxxGxxH e a metionina conservada da “volta de metionina” Met169, característicos de membros das metazincinas (Gomis-

Ruth *et al.*, 1998), são encontrados no subdomínio principal (figura 4.4).



**Figura 4.4: Domínio catalítico e seu subdomínio.** Detalhe da região catalítica mostrando os resíduos altamente conservados na classe P-III de SVMPs “**H**<sub>145</sub>**E**<sub>146</sub>**XXH**<sub>149</sub>**XXG**<sub>152</sub>**XXH**<sub>155</sub>”, além da “volta de metionina” representada pelo resíduo M169 e a porção C-terminal do domínio M formado pela hélice- $\alpha$  H5 destacada em verde que se estende desde a região próxima ao sítio de coordenação do zinco, até a parte posterior do domínio, próximo ao sítio de ligação do cálcio.

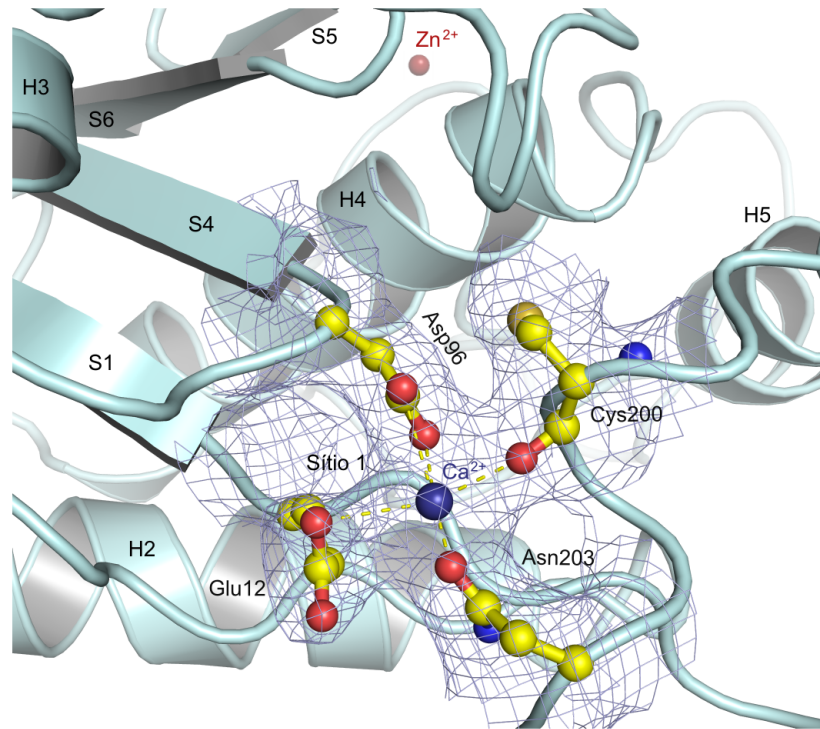
O sítio ativo do domínio catalítico da bothropasina é muito similar aos observados em SVMPs da classe P-I (figura 4.5). O íon zinco apresenta uma coordenação pentaédrica formada pelos átomos N $\epsilon$ 2 das histidinas 145, 149 e 155 do motivo **H**<sub>145</sub>**E**<sub>146</sub>**XXH**<sub>149</sub>**XXG**<sub>152</sub>**XXH**<sub>155</sub> e por dois oxigênios presentes no grupo carboxílico do inibidor POL647 (figura 4.5) de maneira similar ao observado na adamalisina II complexada com o mesmo inibidor (Gomis-Ruth *et al.*, 1998).



**Figura 4.5: Sítio de ligação ao zinco.** Detalhe mostrando os resíduos envolvidos na coordenação pentaédrica do íon zinco (esfera vermelha) e o inibidor não específico POL647 (representado em “balls and sticks” na cor azul) presente no sítio catalítico. O mapa de densidade eletrônica ( $2F_{\text{obs}}-F_{\text{calc}}$ ) é traçado e contornado a  $1\sigma$ .

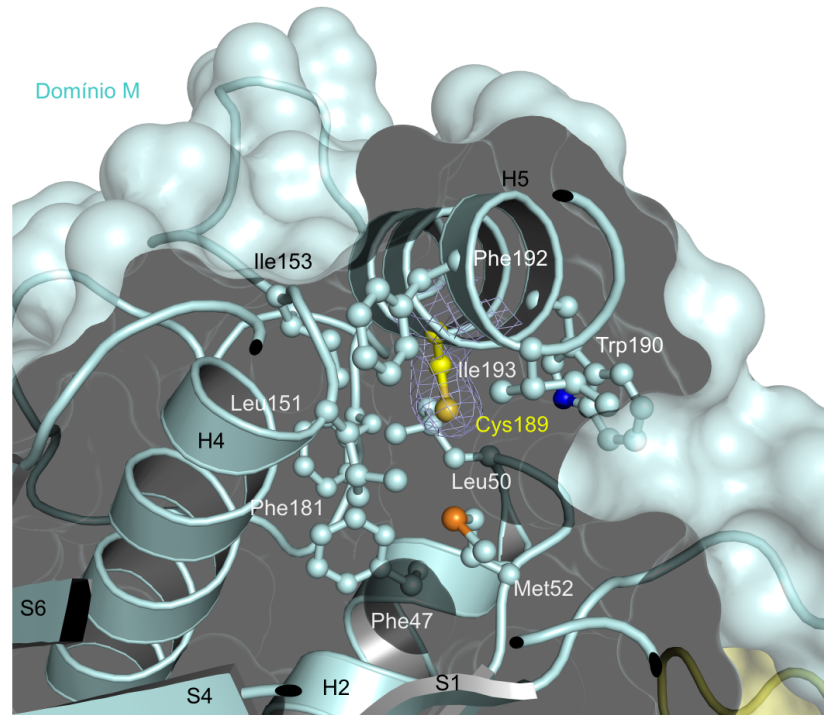
Estruturas de proteínas da classe P-I de SVMPs, obtidas na ausência do inibidor, mostram o íon zinco coordenado por três histidinas e uma molécula de água (Gomis-Ruth *et al.*, 1993b; Gong *et al.*, 1998) e a cadeia lateral de uma tirosina fazendo o papel do quinto ligante (Bode *et al.*, 1992).

O mapa de densidade eletrônica mostra a presença de um íon cálcio na superfície do domínio M, localizado ao lado oposto do sítio ativo. Esse íon, tem sido relacionado à estabilização da folha  $\beta$  que contém o N-terminal. Na bothropasina, o íon cálcio está coordenado por um átomo de oxigênio, proveniente da cadeia lateral do Glu12 e Asn203, dois átomos de oxigênio da Asp96 e pelo grupo carbonil da Cys200 (figura 4.6).



**Figura 4.6:** Sítio de ligação ao cálcio do domínio M. Detalhe da coordenação do íon  $\text{Ca}^{2+}$  (esfera azul) pelos resíduos representados em “balls and sticks” localizados na parte oposta ao sítio de ligação ao zinco (parte superior central da figura).

O domínio catalítico da bothropasina ainda é provido de 7 resíduos de cisteína, sendo que 6 deles realizam pontes dissulfeto: Cys120-Cys200, Cys160-Cys184 e Cys162-Cys167 (figuras 4.11 e 4.12). A possibilidade do resíduo de cisteína livre, no domínio catalítico, ser responsável por interações inter-domínios ou inter-moleculares foi afastada pela observação de que o resíduo de cisteína livre, Cys189, é circundado por resíduos hidrofóbicos e livre de qualquer possibilidade de interações com outras moléculas ou domínios (figura 4.7). Uma observação semelhante foi feita na estrutura da acutolisina (Gong *et al.*, 1998) e para a VAP2B (Igarashi *et al.*, 2007).



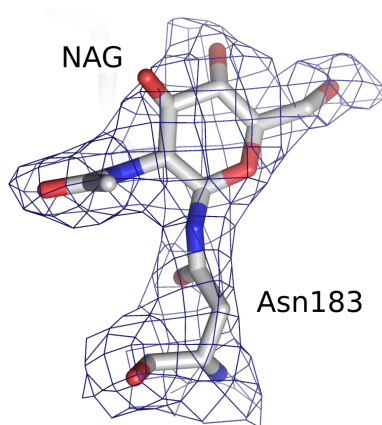
**Figura 4.7: Domínio catalítico evidenciando a cisteína livre 189.** O domínio catalítico M está representado em “cartoon” envolto por um gráfico de superfície de van der Waals para enfatizar a impossibilidade de interação do resíduo Cys189 em uma possível formação de ponte dissulfeto. A Cys189 é circundada por resíduos hidrofóbicos formando uma espécie de barreira e tornando-a inacessível ao meio externo.

### Região de ligação ao açúcar

A glicosilação de proteínas para formar glicoproteínas é um processo importante, pois promove alterações nas propriedades protéicas como estabilidade, solubilidade e volume físico. Os carboidratos em glicoproteínas também atuam como sinais de reconhecimento, essenciais no processo de direcionamento protéico e para o reconhecimento celular de proteínas e de outras células. Há dois tipos de glicosilação: a nitrogênio-glicosilação (N-glicosilação), que ocorre no nitrogênio da amida de cadeias laterais de asparagina e a oxigênio-glicosilação (O-glicosilação), que ocorre no hidróxi-oxigênio de serina e treonina. O processo da N-glicosilação é concomitante à tradução, isto é, ocorre enquanto a proteína está sendo sintetizada e, assim, pode afetar

o enovelamento da proteína. O processo de O-glicosilação ocorre após a tradução e apenas sobre proteínas totalmente enoveladas.

É descrito na literatura que metaloproteases de venenos são moléculas glicosiladas. Na estrutura da bothropasina foi encontrada densidade eletrônica suficiente para acomodar uma molécula de açúcar, provavelmente o N-acetil glucosamina (NAG), ligada ao resíduo Asn183 (figura 4.8), caracterizando, portanto, a ocorrência de N-glicosilação.



**Figura 4.8: Molécula de NAG presente na estrutura cristalográfica.** Uma molécula de NAG foi modelada na densidade eletrônica da bothropasina na região de glicosilação.

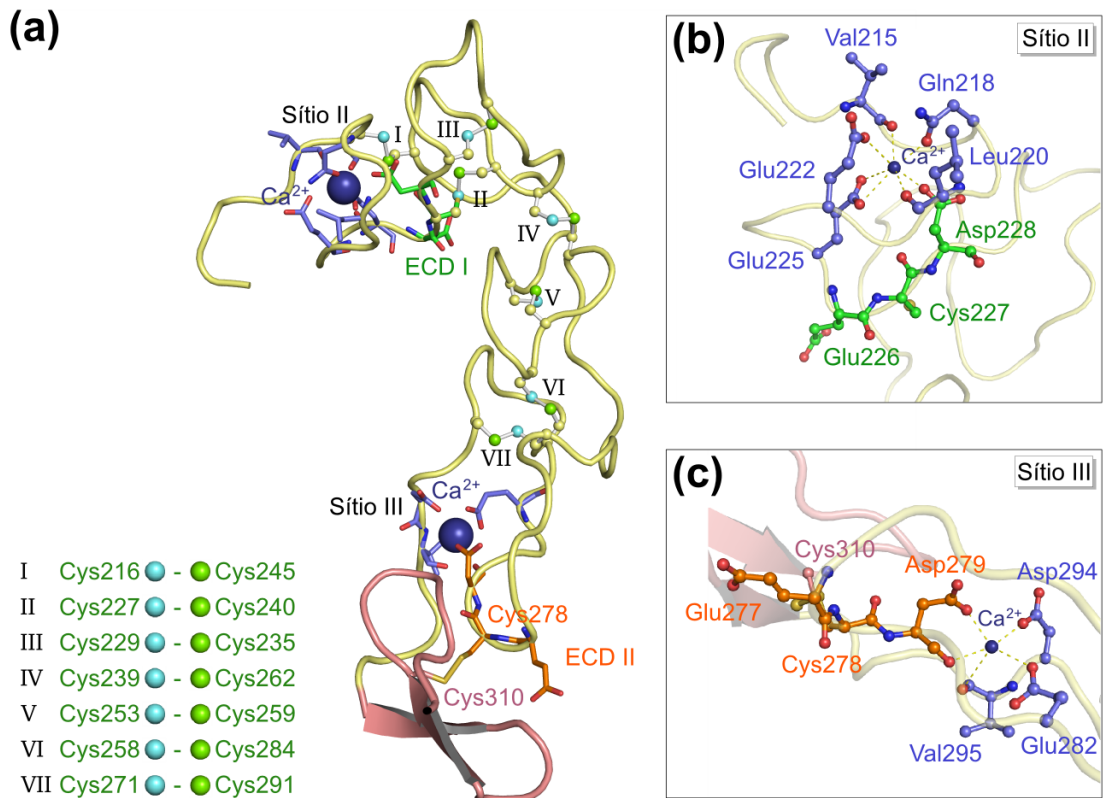
Foi observada uma densidade eletrônica de maior tamanho, no sítio de ligação ao NAG, indicando que outros açúcares poderiam ser adicionados à molécula de NAG como, por exemplo, manoses e glicoses, porém, dada a qualidade da resolução, não foi possível a identificação dessas moléculas.

### **O domínio tipo-desintegrina (D)**

O início do domínio D foi definido de acordo com evidências experimentais, a partir de dois trabalhos independentes. O sequenciamento da região N-terminal da jararagina-C nativa (Usami *et al.*, 1994) e ALT-C (Souza *et al.*, 2000), duas proteínas derivadas do processo proteolítico, contendo apenas os domínios DC,

evidenciaram exatamente a mesma seqüência em seu N-terminal: IISPPVCGNELLEV. A primeira isoleucina, corresponde a Ile210, presente também na região espaçadora.

O domínio desintegrina não apresenta estruturas secundárias bem definidas, sendo formado, exclusivamente, por “*loops*” estabilizados por pontes dissulfeto e também pela presença de dois íons de cálcio encontrados nos sítios de ligação I e II (figura 4.9 (a)). O cálcio no sítio II, localizado no início do domínio D, é coordenado por átomos de oxigênio dos resíduos Val215, Leu220, Glu222 e Asp228 (figura 4.9 (b)). O Asp228 está incluído no primeiro motivo ECD altamente conservado na classe P-III das SVMPs, e cuja função nunca foi investigada.



**Figura 4.9: A estrutura do domínio tipo-desintegrina (D).** (a) Resíduos envolvidos na coordenação do íon  $\text{Ca}^{2+}$  e as pontes dissulfeto representadas por esferas azul ciano e verde, respectivamente, e numeradas de acordo com a legenda no canto inferior esquerdo da figura (a). Os íons de cálcio estão representados como esferas azuis. Os “*loops*” desintegrina (ECD I e II) estão em verde e alaranjado respectivamente. O início do domínio rico em cisteína (C) aparece na cor rosa. Representações detalhadas dos resíduos responsáveis pela coordenação dos íons  $\text{Ca}^{2+}$  aparecem nas figuras (b) e (c).



O domínio desintegrina apresenta um alto conteúdo de resíduos de cisteínas, e 7 pontes dissulfeto são encontradas: Cys216-Cys245, Cys227-Cys240, Cys229-Cys235, Cys239-Cys262, Cys253-Cys259, Cys258-Cys284, Cys271-Cys291 (Figura 4.9 (a)). Uma surpreendente observação ficou por conta de uma ponte dissulfeto inter-domínios entre a Cys278 e Cys310, conectando os domínios D e C. A Cys278 pertence ao segundo motivo ECD da estrutura, o qual tem sido demonstrado estar envolvido na ligação à integrina (Jia *et al.*, 1997). Este motivo ECD está localizado em um “loop” e é estruturalmente relacionado à região RGD de desintegrinas-RGD, derivadas das P-II SVMPS. Nessas proteínas, o “loop” é qualificado como importante para que a ligação com as integrinas seja estabelecida (McLane *et al.*, 1998). Em bothropasina, o motivo ECD é também encontrado em uma região estruturalmente desorganizada, a qual é estabilizada pela ponte dissulfeto Cys278-Cys310 e também por um íon  $\text{Ca}^{2+}$  presente no sítio III (figura 4.9 (c)). A cadeia lateral do Glu277 do motivo ECD está exposta ao solvente e, portanto, poderia ser importante em um processo de reconhecimento e/ou garantir especificidade a ligantes.

O cálcio no sítio III é coordenado por dois átomos de oxigênio do Asp279, do domínio ECD, um átomo de oxigênio do Asp294 e um átomo de oxigênio do Glu282 e da cadeia principal da Val279 (figura 4.9 (c)).

Recentemente, a estrutura de uma metaloprotease/desintegrina de veneno de *Crotalus atrox* foi determinada (Takeda *et al.*, 2006). Os autores sugerem que o domínio D poderia ser dividido em mais dois subdomínios, o “ombro” (domínio D<sub>s</sub>) e o domínio “braço” (domínio D<sub>a</sub>) que constitui a região estrutural rígida (Takeda *et al.*, 2006). Os autores sugerem que o “loop” desintegrina, que contém o motivo ECD, é inacessível à ligação da proteína. Embora a exata função do motivo ECD ainda



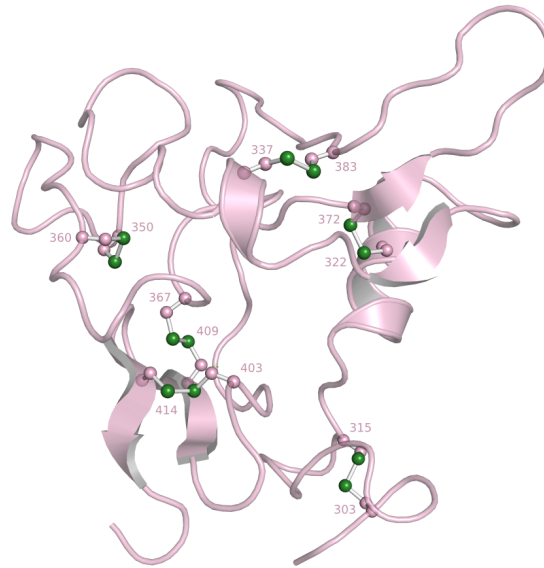
permaneça controversa, alguns trabalhos têm mostrado interações do motivo ECD das P-III SVMPs e ADAMs com algumas integrinas (Jia *et al.*, 1997; Gichuhi *et al.*, 1997; Ramos *et al.*, 2007).

### O domínio rico em cisteína (C)

Apesar da notação de “rico em cisteína”, o domínio C possui um número menor de cisteínas que o domínio D, assim como uma razão número de aminoácidos/número de cisteínas também inferior ao domínio D. A denominação “rico em cisteína” pode ser explicada pelo fato de diferentes autores tentarem dividir os domínios embasados na seqüência de aminoácidos e não a partir da estrutura cristalográfica.

Estabeleceu-se o início do domínio C como sendo a partir da seqüência de aminoácidos QPCLDN, baseado no C-terminal de desintegrinas RGD desprovidas do domínio C (Calvete *et al.*, 2000; Kamiguti *et al.*, 2003; Fox e Serrano, 2005; Pinto *et al.*, 2007). Apesar da ausência natural do domínio C, formas ativas recombinantes têm sido produzidas para homólogas à bothropasina, como, por exemplo, a atrolisina A (Jia *et al.*, 2000; Serrano *et al.*, 2005). Atrolisina A é capaz de inibir a agregação plaquetária simulada por colágeno e interagir com células MG-63, a fim de inibir a adesão ao colágeno. Esses dados sugerem um papel funcional ao domínio rico em cisteína da classe P-III.

O domínio C é constituído por 118 resíduos e é composto por duas hélices- $\alpha$  (Met<sub>318</sub>-Phe<sub>326</sub> e Asp<sub>335</sub>-Phe<sub>338</sub>), 4 pequenas fitas  $\beta$  (Gli<sub>300</sub>-Cys<sub>303</sub>; Tyr<sub>307</sub>-Cys<sub>310</sub>, Lys<sub>407</sub>-Ser<sub>410</sub>; His<sub>413</sub>-Asp<sub>416</sub>) e o restante da estrutura é composto por “loops” (figura 4.10). A presença de pontes dissulfeto é importante para garantir a estabilização desses “loops”. Um total de 6 pontes dissulfeto é encontrada neste domínio (Cys<sub>303</sub>-Cys<sub>315</sub>, Cys<sub>322</sub>-Cys<sub>372</sub>, Cys<sub>337</sub>-Cys<sub>383</sub>, Cys<sub>350</sub>-Cys<sub>360</sub>, Cys<sub>367</sub>-Cys<sub>409</sub>, Cys<sub>403</sub>-Cys<sub>413</sub>; figura 4.10).

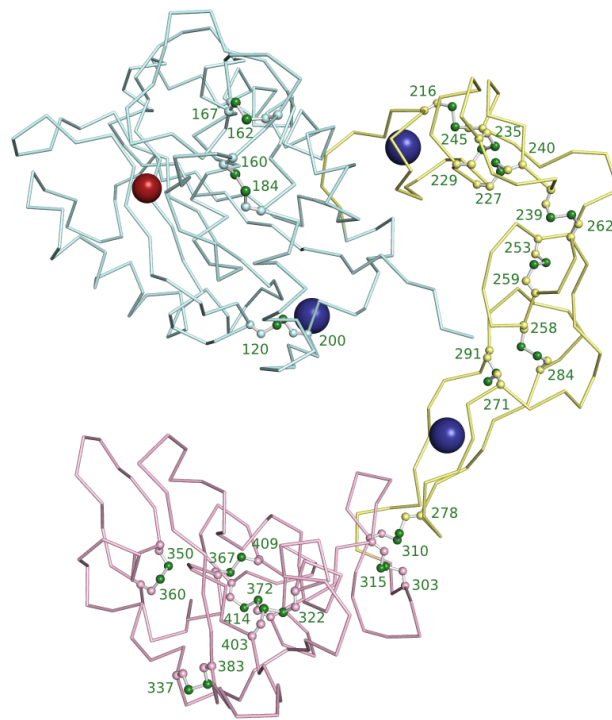


**Figura 4.10: Pontes dissulfeto no domínio rico em cisteína.** Representação em “cartoon” do domínio rico em cisteína e suas pontes dissulfeto representadas em “balls and sticks”. Os 12 resíduos de Cys formadores de 6 pontes dissulfetos estão numerados.

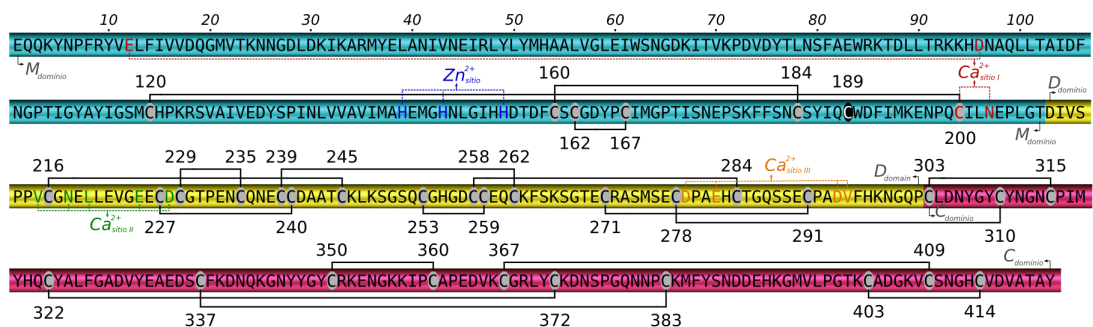
As pontes dissulfeto da catrocolastatina-C, uma proteína tipo-desintegrina, derivada da catrocolastatina, uma SVMP PIII de *Crotalus atrox*, foram determinadas bioquimicamente (Calvete *et al.*, 2000). Foi estabelecido que o domínio tipo-desintegrina apresenta 8 pontes dissulfeto e que o domínio rico em cisteína apresenta 6, e nenhuma ponte dissulfeto inter-domínios foi relatada. A estrutura da bothropasina apresenta um padrão de pontes dissulfeto diferente do relatado para a catrocolastatina, sendo verificada uma ponte dissulfeto entre os domínios D e C, constituída pelos resíduos Cys278-Cys310. Este resultado pode justificar o fato do processo proteolítico liberar somente uma molécula contendo os domínios DC e não ser encontrada uma molécula constituída somente do domínio C. Interessantemente, a estrutura da catrocolastatina (Igarashi *et al.*, 2007) mostra um padrão de pontes dissulfeto semelhante ao encontrado para bothropasina, porém os autores não discutem esses resultados.

As figuras 4.11 e 4.12 ilustram as pontes dissulfeto na estrutura tridimensional,

e na estrutura primária da bothropasina.



**Figura 4.11: Distribuição das pontes dissulfeto por toda a estrutura da bothropasina.** As 17 pontes dissulfeto estão representadas em “balls and sticks” e numeradas. O zinco e o cálcio ligados ao domínio M, assim como os cálcios ligados ao domínio D estão representados por esferas nas cores vermelha para o zinco e azul para o cálcio respectivamente. Os domínios M, D e C estão representados nas cores azul ciano, amarela e rosa, respectivamente.



**Figura 4.12: Seqüência de aminoácidos da bothropasina mostrando os padrões de pontes dissulfeto para os domínios M, D e C respectivamente.** Os resíduos formadores de pontes dissulfeto estão destacados em cinza, numerados e conectados por linhas contínuas. A cisteína livre 189 está destacada em preto. Os sítios de ligação ao cálcio I, II e III assim como o de ligação ao zinco estão indicados.

### Jararagina vs Bothropasina

Durante décadas, acreditava-se que a jararagina fosse a proteína predominante no veneno de *Bothrops jararaca* e que, por um método de purificação estabelecido por Paine (1992), a obtenção da proteína P-III tratava-se sempre da jararagina.

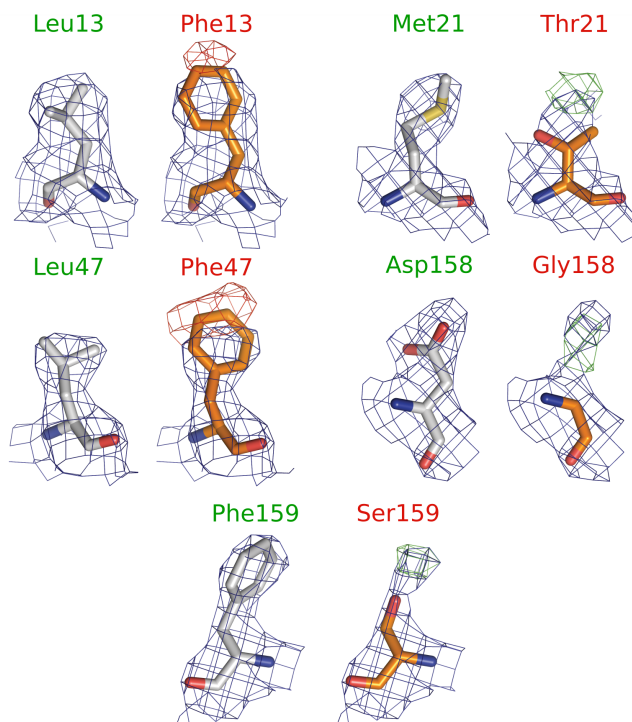
Graças à aplicação da técnica de cristalografia de proteínas, empregada neste trabalho, foi possível uma comparação entre as estruturas tridimensionais das proteínas obtidas pelo método de Paine (1992) e pelo inédito protocolo de purificação desenvolvido e estabelecido nos experimentos já apresentados nesta tese.

Como descrito no Capítulo 3, de posse da estrutura refinada da bothropasina, foi possível o tratamento de um conjunto de imagens de difração de raios X coletado em meados do ano 2000 de um cristal supostamente formado pela proteína jararagina a fim de se checar a estrutura da proteína. A análise inicial mostra um modelo com grande similaridade estrutural à bothropasina descrita nesta tese.

Durante o processo de refinamento da estrutura com dados coletados em 2000, diversas regiões apresentaram as mesmas diferenças pontuais da bothropasina em relação aos resíduos a serem modelados, quando comparados aos da jararagina. Essas diferenças eram fortemente acusadas, tanto pelas densidades eletrônicas do mapa diferença ( $F_{\text{obs}} - F_{\text{calc}}$ ), quanto pelo mapa pesado ( $2F_{\text{obs}} - F_{\text{calc}}$ ). Como pode ser visualizado na figura 4.13, o primeiro cristal era constituído puramente de bothropasina.

Além da indubitável certeza da correta seqüência através da inspeção dos mapas de densidades eletrônicas, outro fato que merece atenção é a quase inexistente diferença do ponto isoelétrico teórico (pI) e massa molecular entre as duas proteínas. Os pontos isoelétricos calculados para as proteínas são: 5,08 e 5,05 para jararagina e bothropasina, respectivamente. Suas massas moleculares são de 48,85 e 48,74 kDa

para jararagina e bothropasina, respectivamente.



**Figura 4.13: A densidade eletrônica e os verdadeiros resíduos.** Dentre as 19 diferenças sugeridas no seqüenciamento através do mapa de densidade eletrônica, os 5 melhores representantes são mostrados nesta figura. À esquerda, em verde, estão os resíduos, de bothropasina, corretamente atribuídos dentro das densidades e, à direita, em vermelho, os resíduos supostamente atribuídos, baseados na seqüência da jararagina. O mapa  $2F_{obs}-F_{calc}$  está em azul e o mapa diferença, contornado com valores para sigma iguais a 3 e -3 estão representados em verde e vermelho respectivamente.

Com base nas informações supracitadas, verifica-se a dificuldade de se obter e garantir que a proteína de interesse seja jararagina ou bothropasina, independente do método de purificação empregado. Uma das maneiras de se aferir qual proteína foi purificada, seria através do seqüenciamento da proteína por degradação de Edman ou através de espectrometria de massas. Entretanto, o emprego dessas técnicas ao final de todo o processo de purificação seria muito laborioso e inviável. Além disso, o seqüenciamento da região N-terminal é impossibilitado por apresentar-se bloqueado.

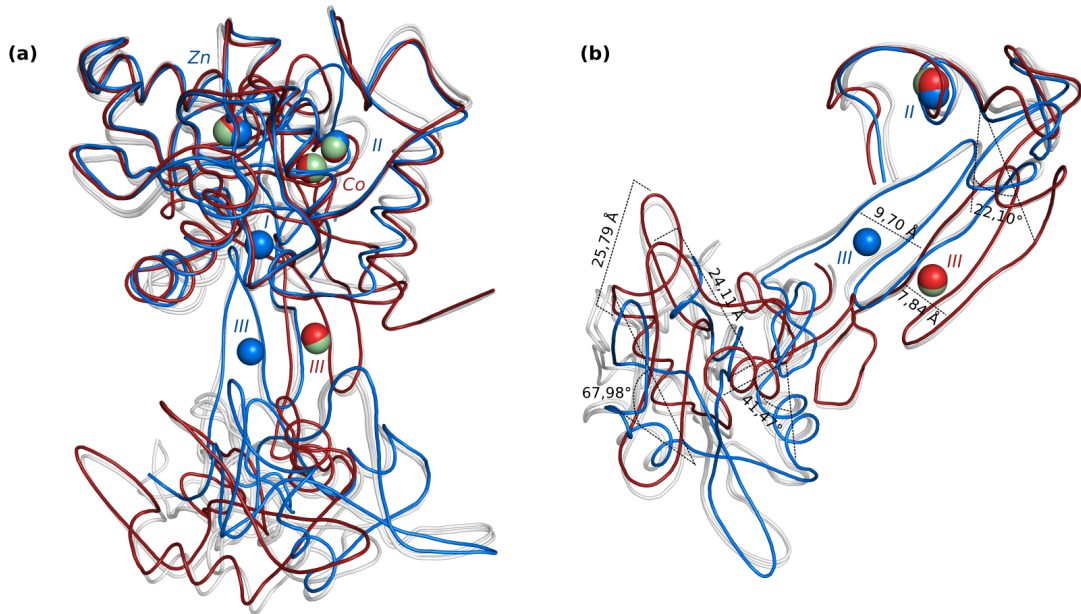
Uma vez que dois experimentos, baseados em protocolos totalmente distintos de purificação e cristalização, resultaram na obtenção da estrutura cristalográfica da mesma proteína, pode-se então considerar que a forma protéica de maior concentração encontrada no extrato bruto do veneno da serpente *B. jararaca* é a bothropasina e não a jararagina.

### Comparação estrutural com outras P-III SVMPs

Até 2006 não havia nenhuma estrutura de SVMPs da classe P-III conhecida. O grupo que determinou a primeira estrutura tridimensional publicou, recentemente, a estrutura de mais uma SVMP, ambas do veneno de *Crotalus atrox* (Takeda *et al.*, 2006; Igarashi *et al.*, 2007). A VAP1 é uma metaloprotease/desinegrina dimérica e a VAP2B1 é monomérica.

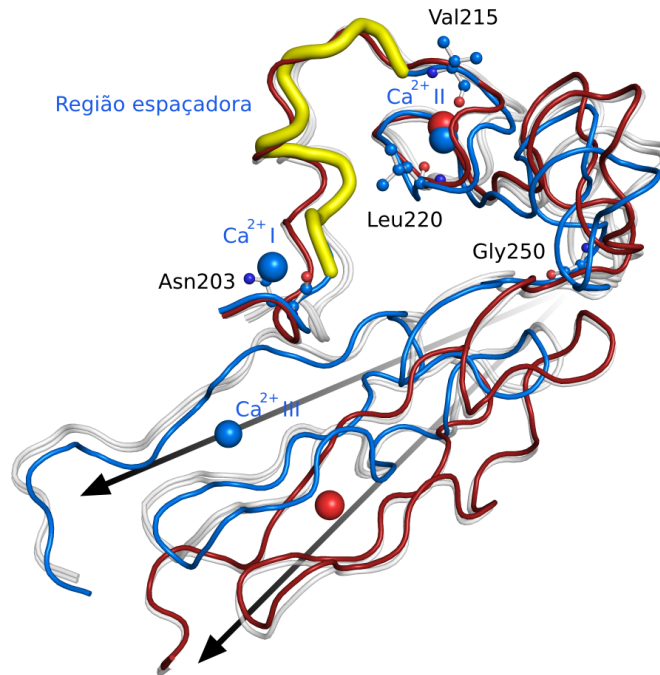
Neste trabalho, estruturas que apresentam os domínios M, D, DC e MDC (um total de 14 estruturas), foram comparadas à molécula de bothropasina com a finalidade de se observar onde as maiores mudanças conformacionais estão localizadas. As sobreposições estruturais foram realizadas com o auxílio do programa LSQMAN (Kleywegt e Jones, 1994).

As orientações relativas, de alguns subdomínios, mostraram-se bastante divergentes quando alinhados estruturalmente. As maiores diferenças foram observadas quando o domínio M das SVMPs PIII VAP1 e VAP2B são sobrepostos. As partes N- e C-terminais dos domínios D e C da bothropasina mostram-se rotacionadas em, aproximadamente, 20°, 40° e chegando até a 70° nas extremidades do domínio C, resultando em um deslocamento de até 26 Å em sua terminação distal (figura 4.14 (a) e (b)).



**Figura 4.14: Mobilidade dos domínios D e C em diferentes P-III SVMPs.** (a) Os domínios M, D e C de duas P-III SVMPs estão destacados nas cores azul e vermelho: bothropasina e VAP1 respectivamente. Outros representantes de SVMPs estão na cor cinza claro. Os íons  $\text{Ca}^{2+}$  e  $\text{Zn}^{2+}$  estão representados por esferas para um melhor discernimento das regiões em questão. Esferas azuis, vermelhas e verdes representam os íons para os sítios de ligação das proteínas: bothropasina, VAP1 e VAP2B respectivamente. Nota-se que tanto a VAP1 como VAP2B (uma das estruturas em cinza) possuem um íon  $\text{Co}^{+2}$  no domínio M, próximo à região de ligação ao cálcio na bothropasina. (b) Superposição dos domínios D e C de P-III SMVPs. A partir do sítio II de ligação ao  $\text{Ca}^{2+}$ , os domínios D e C, apesar de estruturalmente equivalentes, adotam uma bifurcação. O reflexo de tal bifurcação fica ainda mais evidenciado no domínio C, com ângulos de rotação de até  $70^\circ$ .

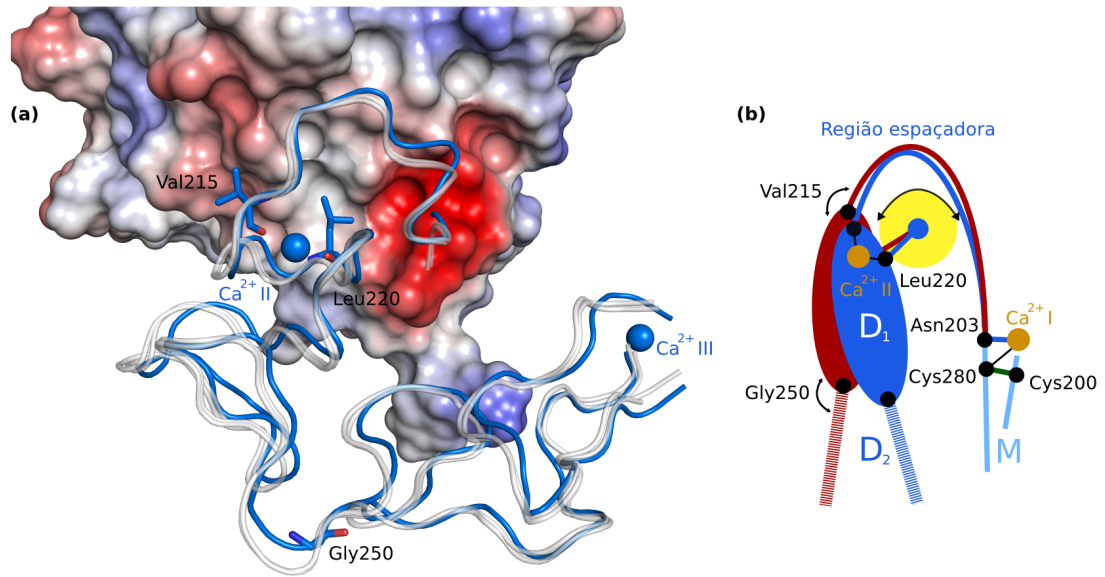
Na sobreposição do domínio C das estruturas, verifica-se o início de deslocamentos a partir da Leu220 que se estende até o resíduo Gly250 (figura 4.15), que é conservado em todas as P-III SVMPs analisadas. Toda a região entre os sítios de ligação aos íons cálcio (sítio I e II) apresenta uma leve movimentação em consequência da rotação dos ângulos da cadeia principal do resíduo Val215 (figura 4.15). A partir do resíduo Gly250, a sobreposição dos domínios mostra que ocorre uma ‘bifurcação’, ou seja, que a posição C-terminal do domínio D é deslocada ao longo da cadeia (figura 4.15).



**Figura 4.15: Val215 e Leu220 como moduladores da movimentação estrutural.** A Val215, que participa da coordenação do cálcio no sítio II, separa a região espaçadora destacada em amarelo da região onde ocorrem as maiores torções nas cadeias principais na porção N-terminal do domínio D. A Gly250, por se tratar de um resíduo bastante flexível, é o responsável pelo início da divergência estrutural adotada por essa classe de proteínas. As setas indicam que, a partir da Gly250, as estruturas adotam diferentes direções, mostrando-se fundamental para efetivação desse processo. As cores das esferas azuis e vermelhas representam os íons para os sítios de ligação das proteínas bothropasina e VAP1 respectivamente, assim como suas estruturas tridimensionais.

Durante o processo de movimentação da região espaçadora (Asn<sub>203</sub>-Val<sub>215</sub>), a cadeia lateral do resíduo Leu220 no domínio D aparece posicionada em um ponto articulador (figuras 4.16 (a) e (b)). O átomo oxigênio da cadeia principal do resíduo Leu220 é um dos que coordenam o íon Ca<sup>2+</sup> no sítio II, e sua cadeia lateral proeminente na porção N-terminal do domínio D interage com uma pequena cavidade hidrofóbica na superfície do domínio M (figura 4.16 (a)).





**Figura 4.16: Papel articulador da Leu220.** (a) Detalhe da região hidrofóbica entre a cadeia lateral da Leu220 e o domínio catalítico. A superfície molecular do domínio M está colorida de acordo com sua superfície de potencial eletrostático (do vermelho (-) ao azul (+)). A estrutura da bothropasina está destacada na cor azul e duas catrocolastatina estão na cor cinza claro. Os cálcios são representados por esferas azuis e os resíduos chaves estão representados em “stick”. (b) Diagrama esquemático da movimentação tipo dobradiça na junção da região espaçadora e porção N-terminal do domínio desintegrina. O domínio D foi subdividido em D<sub>1</sub> e D<sub>2</sub> com a finalidade didática. D<sub>1</sub> representa a região bastante flexível Cys<sub>216</sub>-Ser<sub>249</sub>. D<sub>2</sub> é a região C-terminal do domínio desintegrina. Esquema adaptado de Igarashi *et al.*, 2007.

Outras cadeias laterais de aminoácidos hidrofóbicos como Leu, Phe, Ile ou Val, são encontrados nessa mesma posição articuladora, nas famílias das ADAM e SVMPS (figuras 4.17 e 4.19). A cadeia lateral desses resíduos poderia atuar como uma junta, permitindo a parte superior do domínio D adotar várias orientações em relação ao domínio M. A região espaçadora realiza poucas interações específicas com a parte superior do domínio D, garantindo, assim, um certo grau de liberdade para movimentação.

A região espaçadora (E<sub>204</sub>PLxTDIVSPP<sub>214</sub>) é altamente conservada nas mais diversas famílias de metaloproteases, como pode ser observado na figura 4.17. A região atua primariamente como elo entre os domínios M e D e poderia exercer o papel

de modulador dos movimentos e orientação do domínio M em relação ao resto da molécula, visando a uma melhor adaptação da estrutura na interação com diferentes alvos. Com esta vasta quantidade de regiões flexíveis espalhadas por toda a molécula da bothropasina, assim como em estruturas homólogas, pode-se especular que toda essa flexibilidade intrínseca seja importante para um melhor ajuste no reconhecimento e especificidade a um substrato.

A verdadeira função dessas regiões flexíveis, assim como todo o processo de reconhecimento de substratos e adesão ainda necessitam ser elucidados com o auxílio, por exemplo, de técnicas espectroscópicas, microscopia eletrônica de complexos protéicos e estruturas de complexos proteína-ligantes.

#### **Análise seqüencial comparativa entre bothropasina e outras P-III SVMs**

É sugerido na literatura que um “loop” no domínio D da catrocolastatina, denotado como “*Hiper Variable Region*”, HVR, seja a região de reconhecimento/especificidade dos ligantes e, portanto, importante para a atividade da proteína (Takeda *et al.*, 2006).

Uma análise comparativa de 41 seqüências de aminoácidos de proteínas da classe P-III de SVMs, disponíveis no banco de dados NCBI<sup>4</sup>, foi feita com a finalidade de se obter informações adicionais a partir da estrutura primária dessas proteínas (figura 4.17). As seqüências foram alinhadas com auxílio do programa MULTALIGN (Corpet, 1988) e árvores filogenéticas foram construídas utilizando-se o pacote de programas PHYLIP (Felsenstien, 1989).

O Domínio pró-enzima mostrou-se altamente conservado, assim como o sítio de ligação ao zinco no domínio catalítico. Em estudos anteriores, foi demonstrado que os domínios ativos foram mais susceptíveis a uma acelerada evolução que os domínios

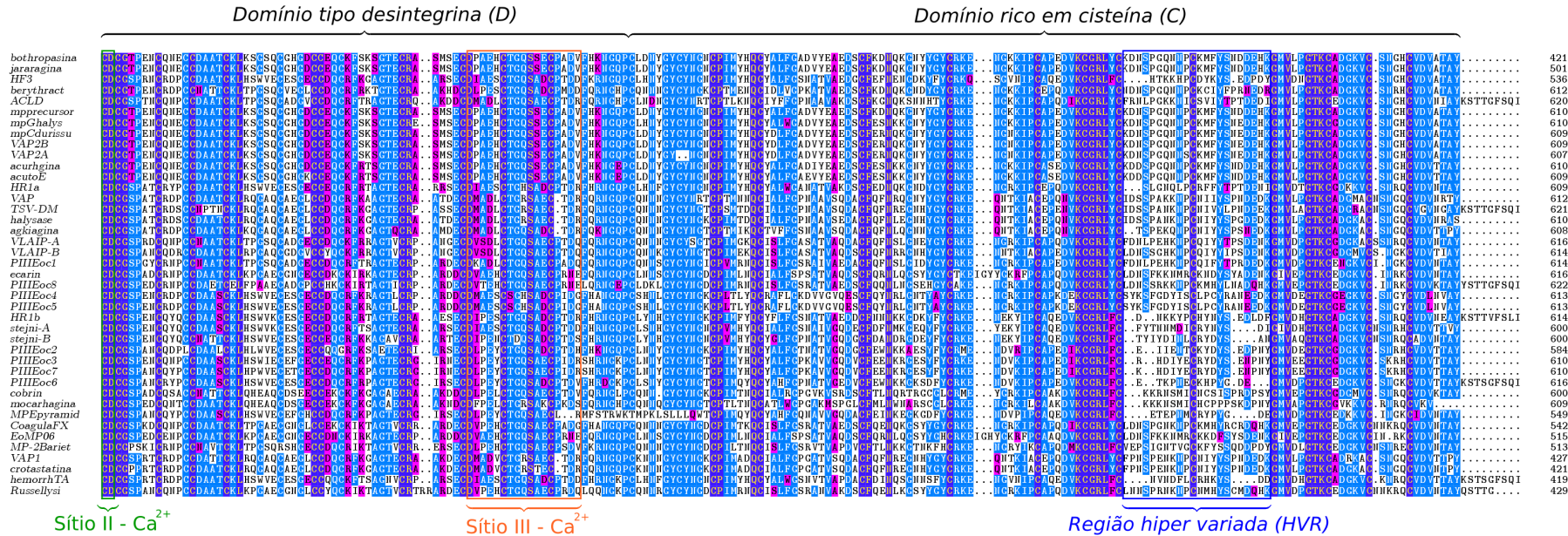
---

4 [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Acesso em 03 de junho de 2007.

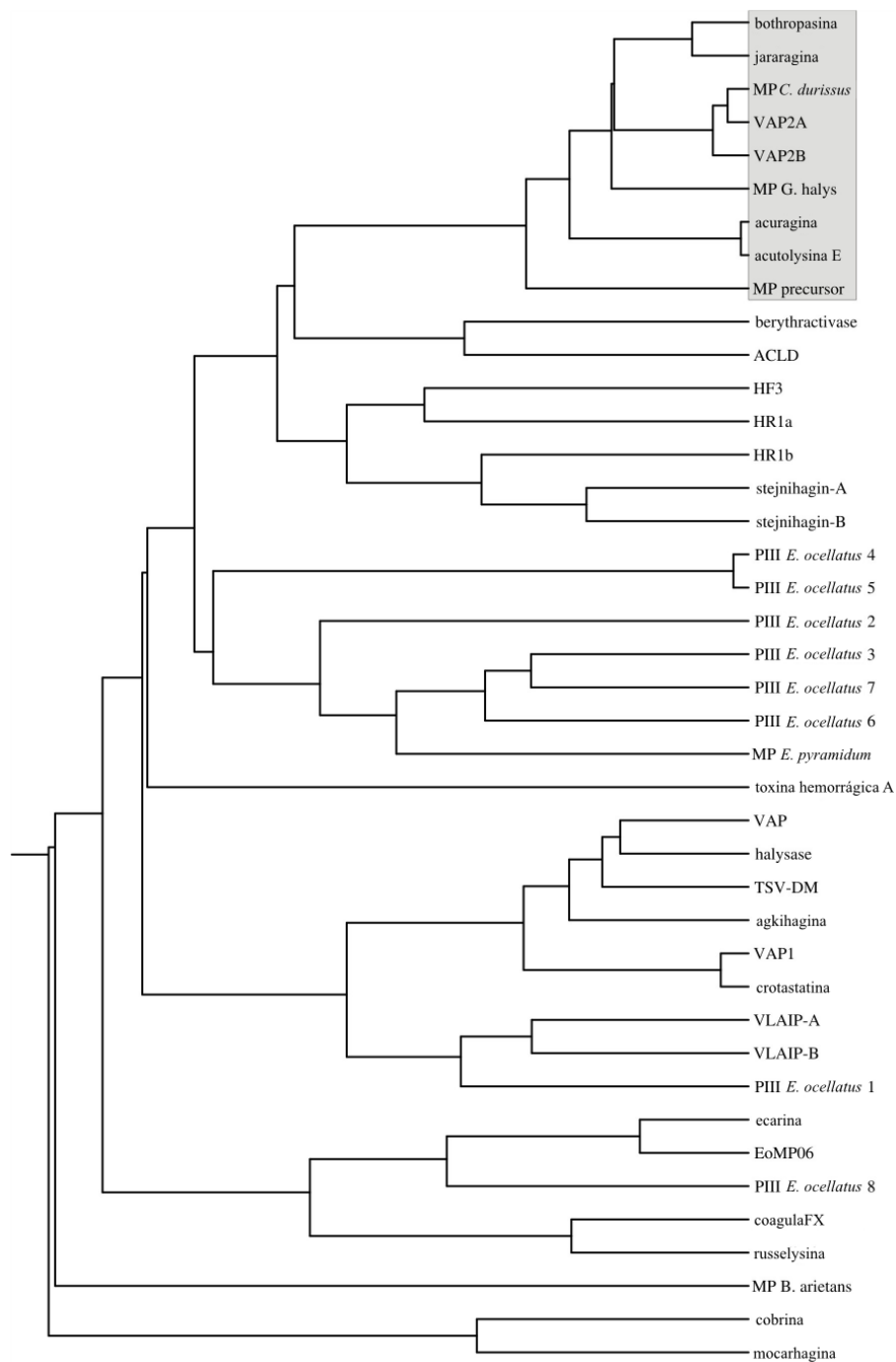
regulatórios, os quais, provavelmente, geraram subfamílias distintas de proteínas com diferentes atividades biológicas ou especificidade (Moura-da-Silva *et al.*, 1996b). Análises filogenéticas mostraram que as classes P-III de SVMPs podem ser subdivididas em diversos agrupamentos protéicos (figura 4.18).





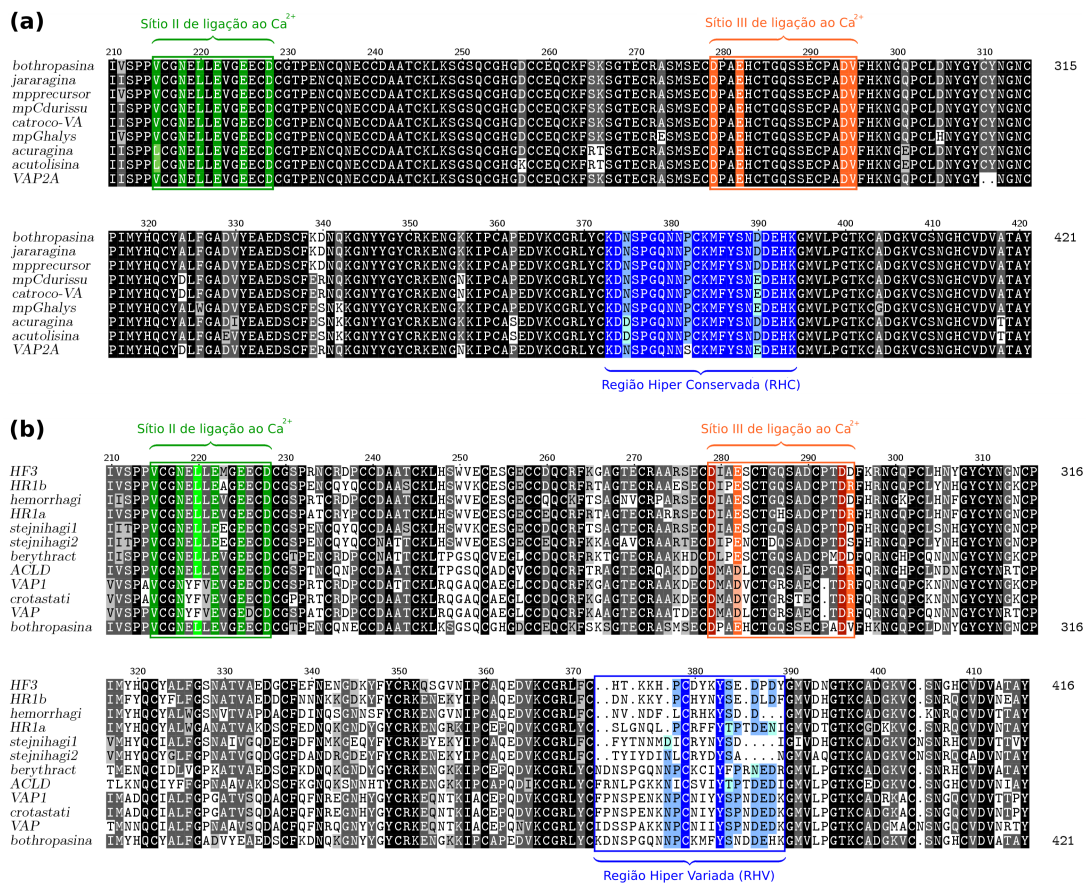


**Figura 4.17: Alinhamento seqüencial de P-III SVMPs.** Um total de 41 P-III SVMPs foram alinhadas. O alinhamento destaca uma boa similaridade entre todas as seqüências. O domínio pró-enzima é altamente conservado, assim como o número e espaçamento entre as cisteínas, sítios de ligação ao cálcio e de ligação ao zinco. As seqüências foram obtidas no banco de dados NCBI e são listadas da seguinte maneira: jararagina (NA: P30431); bothropasina (NA: AAC61986.2); metaloproteínase precursor de *B. insularis* (NA: AAM09693.1); metaloproteínase P-III de *C. durissus durissus* (NA: ABA42117.1); VAP2B de *C. atrox* (NA: AAC59672.1); metaloprotease de *G. halys* (NA: AAD02652.1); acuragina precursor de *D. acutus* (NA: AAS57937.1); acutolysina-E precursor de *D. acutus* (NA: Q9W6M5); VAP2A de *C. atrox* (NA: BAF56420.1); HF3 de *B. jararaca* (NA: AAG48931.5); HR1b de *T. flavoviridis* (NA: P20164); toxina hemorrágica de *C. atrox* (NA: AAA03326.1); HR1a de *T. flavoviridis* (Q8JIR2); stejnihagin-A de *V. stejneri* (NA: ABA40760.1); stejnihagin-B de *V. stejneri* (NA: ABA40759.1); berythrativase de *B. erythromelas* (NA: AAL47169.1); ACLD de *A. controtrix laticinctus* (NA: AAC18911.1); VAP1 de *C. atrox* (NA: BAB18307.1); crotastatina de *C. durissus terrificus* (NA: ABB42830.1); VAP de *T. flavoviridis* (NA: BAB60682.1); TSV-DM de *V. stejneri* (NA: ABC73079.1); halsyase de *G. halys* (JC8056); agkihagina de *D. acutus* (NA: ABB79955.1); VLAIP-A de *M. lebetina* (NA: AAX38181.1); VLAIP-B de *M. lebetina* (AAX38182.1); P-III SVMP de *E. ocellatus* 1 (NA: CAJ01679.1); Ecarina precursor de *E. carinatus* (NA: Q90495); P-III SVMP de *E. ocellatus* 8 (NA: CAJ01680.1); P-III SVMP de *E. ocellatus* 4 (NA: CAJ01685.1); P-III SVMP de *E. ocellatus* 5 (CAJ01684.1); P-III SVMP de *E. ocellatus* 2 (NA: CAJ01688.1); P-III SVMP de *E. ocellatus* 3 (NA: CAJ01683.1); P-III SVMP de *E. ocellatus* 7 (NA: CAJ01682.1); P-III SVMP de *E. ocellatus* 6 (NA: CAJ01687.1); cobrina precursor de *N. naja* (NA: AAF00693.1); mocarhagina 1 de *N. mossambica mossambica* (NA: AAM51550.1); metaloprotease de *E. pyramidum* (NA: CAA55566.1); CoagulaFX de *M. lebetina* (NA: Q7T046); EoMP06 de *E. ocellatus* (NA: AAP92424.1); MP-2 de *B. arietans* (NA: AAX86634.1) and Russellysina de *D. russellii siamensis* (NA: Q7LZ61).



**Figura 4.18: Árvore filogenética de P-III SVMPs.** As seqüências de 41 P-III SVMPs foram obtidas como descrito na legenda da figura 4.17. O alinhamento sequencial foi submetido ao programa ProDist (parte do pacote de programas PHYLIP) para o cálculo das matrizes de distância a partir da estrutura primária e análises do tipo “Neighbor-joining” (NJ) através do método UPGMA de agrupamento com o intuito de inferir as relações entre as seqüências. O subgrupo formado pela bothropasina, jararagina, metaloprotease de *Crotalus durissus*, VAP2A, VAP2B, metaloprotease de *Gloydius halys*, acuragina, acutolisina E e metaloprotease de *B. insularis* estão destacadas.

A análise dos alinhamentos das seqüências mostra que a HVR não é encontrada em todas as proteínas. Ao contrário, observamos que as proteínas podem ser divididas em dois grupos, o constituído por proteínas com o HVR e o grupo de proteínas que apresentam esta região altamente variável, região denominada, neste trabalho, de “*Highly Conserved Region*”, HCR. Exemplos de proteínas dos dois grupos são mostrados na figura 4.19.



**Figura 4.19: Alinhamentos de SVMPs em grupos I e II.** (a) Os domínios DC de bothropasina alinhados a 8 proteínas homólogas e seus fortemente conservados sítios de ligação ao cálcio II e III e a região altamente conservada destacados nas cores verde, alaranjada, azul e suas nuances respectivamente compõem o grupo I. (b) Alinhamento entre proteínas provenientes de diferentes grupos mostram uma baixa conservação do domínio RHV quando comparadas com proteínas pertencentes ao mesmo grupo. Os sítios de ligação ao cálcio II e III e a região RHV seguem o mesmo padrão de cores de (a) e compõem o grupo II. Os números de acesso (NA) para as seqüências estão descritos na legenda da figura 4.17.

A bothropasina é agrupada com outras 8 homólogas, incluindo a

jararagina, VAP2A e VAP2B mas não VAP1 (figura 4.19 (a)). Pode ser observado que a região conhecida como “*Hyper Variable Region*” (HVR; resíduos 373-394), é altamente conservada nesse grupo da bothropasina e muito diferente do grupo da VAP1 (figura 4.19 (b)).

Baseado nessas observações, pode-se concluir que a chamada *região hiper variada* é variada apenas entre os grupos I e II, e não dentro do mesmo grupo. Entretanto, se essas proteínas que compartilham alto grau de similaridade seqüencial também compartilham especificidades biológicas similares, diferentemente dos outros subgrupos, ainda necessita ser elucidado.

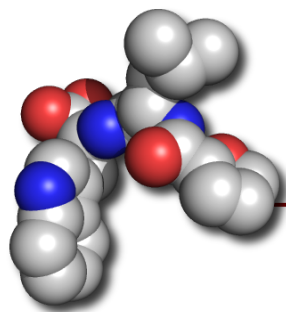
Tem sido proposto que o HVR está envolvido no reconhecimento do alvo em um processo de ataque proteolítico do domínio M (Takeda *et al.*, 2006). Essa hipótese, embora muito provável, ainda necessita de uma comprovação experimental. Por essa razão, aqui é proposto a classificação das SVMPS em dois subgrupos: PIII-HCR e PIII-HVR, de acordo com a identidade dos domínios DC. Esta classificação introduzirá uma maior complexidade a essa classe de proteínas desde a primeira proposta de classificação, baseada em processamento dos domínios DC (Fox e Serrano, 2005).

Uma importante observação diz respeito ao alto grau de conservação dos resíduos envolvidos na coordenação dos íons  $\text{Ca}^{2+}$  nos sítios II e III presentes no domínio DC (figuras 4.19 (a) e (b)). Apesar de pequenas variações em outras regiões ao longo das seqüências, o alto nível de conservação dos sítios de ligação ao cálcio indica, fortemente, que os dois íons  $\text{Ca}^{2+}$  são necessários para assegurar função à desintegrina. Interessantemente, um motivo ECD é encontrado em cada um dos sítios de coordenação de cálcio. O segundo motivo ECD tem sido demonstrado estar envolvido na ligação à integrina que dispara uma resposta de sinalização intracelular,



mediada pela integrina (Kamiguti *et al.*, 1999; Mariano-Oliveira *et al.*, 2003).

As proteínas do subgrupo PIII-HCR, que não apresentam a HVR, incluindo a bothropasina, devem reconhecer os ligantes, através de outro mecanismo que não seja por esse “*loop*” do domínio C, podendo ser, como já comentado, através da cadeia lateral do resíduo glutâmico do motivo ECD do domínio desintegrina.



## **CONCLUSÕES E PERSPECTIVAS**

O presente projeto de doutorado teve como objetivo o estudo estrutural de uma metaloproteinase/desintegrina, oriunda do veneno da serpente (SVMP) *Bothrops jararaca*.

Foram realizados estudos estruturais de amostras obtidas por dois procedimentos diferentes de purificação e verificado que as proteínas purificadas e cristalizadas, em ambos os processos, tratam-se da bothropasina. Esses resultados foram surpreendentes, pois, a princípio, acreditava-se que a amostra supostamente era de jararagina, uma proteína isoforma à bothropasina.

Tanto a bothropasina, quanto a jararagina são obtidas, segundo a literatura, através do isolamento do veneno bruto de *B. jararaca* utilizando colunas cromatográficas diferentes. No entanto, a cristalização mostra que, possivelmente, a fração de maior proporção no veneno seja de bothropasina. Além disso, haja vista que as duas proteínas apresentam massa molecular e ponto isoelétrico praticamente idênticos, a separação das duas isomorfias torna-se extremamente difícil e o emprego de outras técnicas, como espectrometria de massas e/ou seqüenciamento protéico, fazem-se necessárias para o discernimento entre essas duas isoformas.

Uma característica importante dessa proteína é a capacidade de sofrer autólise com produção dos domínios metaloprotease e o domínio desintegrina/rico em cisteína. Para diminuir a ocorrência desse processo e conseqüentemente melhorar o rendimento da obtenção de cristais de bothropasina, o inibidor sintético POL647 mostrou-se fundamental. Centenas de cristais de bothropasina também foram obtidos na ausência do inibidor, contudo, nenhum dos cristais testados apresentou um bom poder de difração, a ponto de ajudar no procedimento de obtenção da estrutura cristalográfica.

A condição ideal de cristalização foi bothropasina concentrada a 8 mg/mL em acetato de amônia 40 mM, solução tamponante de acetato de sódio 100 mM, pH 4.6 e 29% de PEG 4000 como agente precipitante à temperatura de 4°C. Os cristais obtidos foram submetidos a experimentos de difração de raios X no Laboratório Nacional de Luz Síncrotron em Campinas, no “*Brookhaven National Synchrotron Light Source*”, Upton, NY, EUA e no Instituto de Física de São Carlos – USP, em um gerador ânodo rotatório de raios X. Todas as tentativas objetivaram a busca por melhores conjuntos de imagens de difração.

Devido a características intrínsecas da proteína e dos cristais de bothropasina, o melhor conjunto de dados obtido foi de, aproximadamente, 2,80 Å de resolução. Por não possuir estruturas homólogas para o domínio rico em cisteína, diversas tentativas de incorporação de metais pesados aos cristais protéicos foram executadas e os melhores resultados foram obtidos em ensaios de “*quick-soaking*” do cristal em 0,5 M de Iodeto de Sódio (NaI).

A partir de uma combinação de experimentos de substituição molecular e contribuição dos espalhadores anômalos (iodo), fases iniciais foram encontradas, permitindo, assim, a caracterização do domínio metaloprotease (o que representa,

aproximadamente, metade da estrutura) através de um experimento de SAD/MR. A solução das fases possibilitou o emprego de métodos de reconhecimento de padrões em densidades eletrônicas, com o intuito de traçar, inicialmente, as posições dos átomos que compõem a cadeia principal dos domínios, ainda desconhecidos, da proteína. Todavia, a correta interpretação dos mapas de densidade eletrônica foi impossibilitada devido à presença do alto conteúdo de pontes dissulfeto nos domínios ricos desintegrina e rico em cisteína (6 pontes) e à baixa resolução dos mapas de densidade eletrônica.

Através de predição de estruturas secundárias, foi constatado que o domínio, ainda inédito, era predominantemente formado por estruturas aleatórias, também chamadas de “*loops*”, além de hélices- $\alpha$  e também fitas  $\beta$ . Os “*loops*” e bifurcações dificultaram a interpretação adequada dessas regiões, sendo necessária a utilização de métodos mais eficazes de atribuição de átomos nessas densidades.

A molécula de bothropasina é formada por três diferentes domínios: (1) metaloprotease e (2) tipo-desintegrina, associados às atividades de inibição da agregação plaquetária e de adesão celular e (3) rico em cisteína, de função ainda não esclarecida.

Na literatura, um “*loop*” no domínio C, cuja seqüência de aminoácidos é altamente variável (HVR) em P-III SVMPS, tem sido descrito como possível região de reconhecimento e/ou de ligação. Uma análise realizada, neste trabalho, a partir da estrutura primária de 41 proteínas representantes da classe P-III das SVMPS mostrou que, a despeito de sua grande similaridade, as proteínas podem ser classificadas como subgrupo PIII-HCR, como sendo aquelas que não apresentam a HVR em um dos “*loops*” presentes no domínio rico em cisteína (C) e subgrupo PIII-HVR, caracterizado por aquelas proteínas que apresentam o “*loop*” com seqüências altamente variáveis

(HVR).

Embora seja realmente factível que a região HVR tenha a função de reconhecer e ligar-se a ligantes, isso só seria possível nas proteínas do grupo II, uma vez que as proteínas do grupo I têm essa região muito conservada e, sendo assim, não garantiria especificidade ao reconhecimento de ligantes. Todavia, neste trabalho, é sugerido que outras regiões da molécula possam ter essa função, como, por exemplo, o motivo ECD do domínio tipo-desintegrina que apresenta o ácido glutâmico exposto ao solvente e livre para realizar diversos tipos de interações.

Nenhuma estrutura de SVMP era conhecida até meados de 2006, fato esse que acarretou na utilização de técnicas cristalográficas específicas para a determinação da estrutura completa da bothropasina. Até o momento outras duas SVMPs de veneno de *Cotralus atrox* tiveram suas estruturas elucidadas, mas a análise realizada nesses trabalhos levou à proposição de mecanismos que mostram-se restritos a um grupo de proteínas e não para o grupo no qual a bothropasina e a jararagina se classificam.

A obtenção da estrutura da bothropasina abriu novas perspectivas de estudo que podem esclarecer os papéis dos domínios tipo-desintegrina e rico em cisteína nas SVMPs. A cristalização da bothropasina complexada com ligantes como a integrina e a determinação da estrutura tridimensional desse complexo podem mostrar as regiões envolvidas na ligação e elucidar o envolvimento de cada um dos domínios nas atividades biológicas da proteína.

As SVMPs são as principais toxinas envolvidas em patogenicidades induzidas por veneno e, sendo assim, importantes alvos na produção de antivenenos ainda mais eficazes. Contudo, os alvos fisiológicos das SVMPs, assim como seu mecanismo molecular de reconhecimento dos mesmos, ainda é pouco conhecido. A estrutura da bothropasina, apresentada neste trabalho, revela uma dinâmica e uma

arquitetura modular dos domínios MDC para esta classe de proteínas. Essa flexibilidade intrínseca pode ser importante para um ajuste fino no reconhecimento do substrato, proporcionando um alinhamento espacial dos sítios de adesão e catalítico.

## **CONSIDERAÇÕES FINAIS**

Desde o início do meu doutorado, tive ciência da importância e da dimensão do trabalho que me foi proposto pela minha orientadora Profa. Dra. Dulce Helena Ferreira de Souza, assim como pelos professores Glaucius Oliva e Richard Garratt.

Meu trabalho, que iniciou-se em abril de 2003, deu continuidade à busca pela elucidação da estrutura da bothropasina (na época “jararagina”) que iniciou-se no fim de 1999, portanto, foram quase 8 anos para que esse trabalho fosse concretizado e, independente das adversidades, fico muito contente com o desfecho desta história de muita luta e perseverança.

Durante meu doutoramento, também conquistei a oportunidade de participar do Programa de Doutorado com Estágio no Exterior – PDEE, mantido pela CAPES, no Laboratório do Prof. Tom Blundell, Coordenador do Grupo de Cristalografia e Biocomputação no Departamento de Bioquímica da Universidade de Cambridge, Reino Unido. O Prof. Blundell é um renomado especialista em cristalografia de proteínas e modelagem molecular, tendo mais de 300 publicações em revistas de elevado impacto na literatura acadêmica. No Laboratório de

Biocomputação, coordenado por ele, diversos programas foram desenvolvidos e testados em diferentes sistemas, demonstrando um excelente desempenho, mas ainda indisponíveis para uso público em geral.

Tive a oportunidade de aprender técnicas que poderiam ser empregadas na construção e refinamento de estruturas cristalográficas, principalmente em casos considerados difíceis de baixa resolução, assim como dar continuidade ao aprimoramento e aprendizado de técnicas para a construção de modelos tridimensionais por modelagem molecular comparativa. A modelagem molecular comparativa está presente em minha formação, desde meu mestrado, que foi desenvolvido nesta mesma instituição.

Na área de cristalografia de proteínas, tive a oportunidade de participar de congressos, debates e discussões com renomados cristalógrafos como o Prof. Randy Read (Professor do “*Medical Research Council*” (MRC), Universidade de Cambridge) e até então, sua aluna de pós-doc Dra. Airlie McCoy que, gentilmente, empregaram o programa Phaser (versão 2.0), que ainda estava em fase de testes e/ou aprimoramentos, na resolução da estrutura da bothropasina.

Particpei, também, da construção de modelos tridimensionais de proteínas de diversos organismos, explorando análises comparativas de funcionalidade entre proteínas relacionadas com enovelamentos similares, porém, divergentes em relação às seqüências primárias, através da modelagem molecular por comparação durante minha participação na reunião bienal CASP 7 (“*Critical Assessment of Techniques for Protein Structure Prediction*” 7) como um dos integrantes do grupo do Prof. Blundell, representando a Universidade de Cambridge.

Além da relevância para o desenvolvimento desta tese de doutoramento, o estágio no exterior foi de grande importância para o aprendizado de novos conceitos,



para a integração com um grupo estrangeiro de excelência e para o aperfeiçoamento profissional e formação acadêmica do doutorando. Ademais, esta experiência agregou conhecimentos que poderão auxiliar na resolução de problemas relacionados ao refinamento, resolução e modelagens de proteínas dentro do laboratório do CBME do IFSC onde o atual projeto de doutoramento se desenvolveu.

Maiores informações quanto ao que foi desenvolvido durante o estágio no exterior, são mostradas no Apêndice I.

Tive a oportunidade de participar, como colaborador, em trabalhos científicos diversificados. No Apêndice II, apresento esses trabalhos em seu texto original e na íntegra que foram contemplados com a publicação em revistas internacionais indexadas. Os trabalhos se resumem na obtenção de modelos estruturais, através do emprego da modelagem molecular comparativa e adequação dos modelos propostos a resultados experimentais.

Por fim, deixo claro que meus 6 anos de convivência com funcionários, alunos, professores, colegas e amigos no Instituto de Física de São Carlos, desde o meu mestrado, foram muito importantes para que atingisse maturidade científica, fundamental para meu aprendizado, além de criar laços de amizades duradouras e o mais importante de tudo: aprender a não desistir!

## REFERÊNCIAS BIBLIOGRÁFICAS

- Adams, P.D., Gopal, K., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C. and Terwilliger, T.C. (2004) *J. Synchrotron Rad.* **11**, 53-55.
- Afonine, P.V., Grosse-Kunstleve, R.W. and Adams, P.D. (2005) *CCP4 Newsletter July, Contribution* **8**.
- Anai, K, Sugiki, M., Yoshida, E. and Maruyama, M. (1998) *Toxicon*, **36**, 1127-1139.
- Arnaout, M.A. (1990) *Immunol. Rev.*, **114**, 145-180.
- Assakura M.T., Silva, C.A., Mentele, R., Camargo, A.C. and Serrano, S.M. (2003) *Toxicon*, **41**, 217-227.
- Atkinson, B.T., Jarvis, G.E. and Watson, S.P. (2003) *J. Thromb. Hemost.*, **1**, 1278-1287.
- Barravieira, B., Lomonte, B., Tarkowski, A., Hanson, L.A. and Meira, D.A. (1995) *J. Venom. Anim. Toxins*, **1**, 1-11.
- Barros, S.F., Friedlanskaia, I., Petricevich, V.L. and Kipnis, T.L. (1998) *Mediat. Inflamm.*, **7**, 339-346.
- Bjarnason, J.B. and Fox, J.W. (1994) *Pharmacol. Ther.*, **62**, 325-372.
- Blobel, C.P. (2005) *Nat. Rev. Mol. Cell Biol.*, **6**, 32-43.
- Bode, W., Gomis-Ruth, F.X., Huber, R., Zwillig, R. and Stöcker (1992) *Nature*, **358**, 164-167.
- Bode, W., Gomis-Ruth, F.X. and Stocker, W. (1993) *FEBS Lett.*, **331**, 134-140.
- Bradford, M.M. (1976) *Analytical Biochemistry*, **72**, 248-254.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G. (1998) *Acta Cryst. D*, **54**, 905-921.
- Cairns, R.A., Khokha, R. and Hill, R.P. (2003) *Curr. Mol. Med.*, **3**, 659-671.
- Callard, R.E. and Gearing, A.J.H. (1994) *The cytokine facts book*. Academic Press, pp. 31.
- Calvete, J.J., Moreno-Marciano, M.P., Sauz, L., Jurgens, M., Schader, M., Raida, M., Benjamín, D.C. and Fox, J.W. (2000) *Protein Sci.*, **9(7)**, 1365-1373.

- Campbell, J.A. and Lamar, W. (1989) *The Venomous Reptiles of Latin America*. Cornell Univ. Press, Ithaca, USA.
- Cardoso, J.L.C. (1997) Ofidios – *Bothrops (jararaca, cruzeira)*. In: Nicolela, A., Barros, E., Torres, J.B. et al. eds. *Acidentes com Animais Peçonhentos - Consulta Rápida - Porto Alegre - Ministério da Saúde*.
- Cardoso, J.L.C., Fan, H.W., França, F.O.S., Jorge, M.T., Leite, R.P., Nishioka, S.A. et al. (1993) *Quat. J. Med.*, **86**, 315-325.
- Chiang, H.S., Swaim, M.W. and Huang, T.F. (1994) *Br. J. Haematol.*, **87**, 325-331.
- Clegg, D.O., Wingerd, K.L., Hikita, S.T. and Tolhurst, E.C. (2003) *Front Biosci.*, **8**, d723-d750.
- Clissa, P.B., Laing, G.D., Theakston, R.G.D., Mota, I., Taylor, M.J. and Moura-da-Silva, A.M. (2001) *Toxicon*, **39**, 1567-1573.
- Clissa, P.B., Lopes-Ferreira, M., Della-Casa, M.S., Farsky, S.H.P. and Moura-da-Silva, A.M. (2006) *Toxicon*, **47**, 591-596.
- Collaborative Computational Project Number 4 (1994) *Acta Cryst. D*, **50**, 760-763.
- Cominetti, M.R., Terruggi, C.H.B., Ramos, O.H.P., Fox, J.W., Mariano-Oliveira, A., De Freitas, M.S., Figueiredo, C.C., Morandi, V. and Selistre-de-Araujo, H.S. (2004) *Jol. Biol. Chem*, **279**, 18247-18255.
- Coppolino, M.G. and Dedhar, S. (2000) *Int. J. Biochem. Cell Biol.*, **32**, 171-188.
- Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881-10890.
- Correa Jr., M.C., Maria, D.A., Moura-da-Silva, A.M., Pizzocaro, K.F. and Ruiz, I. (2002) *Toxicon*, **40**, 739-748.
- Costa, E.P., Clissa, P.B., Teixeira, C.F.P. and Moura-da-Silva, A.M. (2002) *Inflammation*, **26**, 13-17.
- Darribere, T., Skalski, M., Cousin, H.L., Gaultier, A., Montmory, C. and Alfandari, D. (2000) *Biol. Cell*, **92**, 5-25.
- Dauter, Z., Dauter, M. and Rajashankar, K. R. (2000) *Acta Cryst. Section D*, **56**, 232-237.
- Nagem, R.A.P., Polikarpov, I. and Dauter, Z. (2003) *Meth. Enzymol.– Macrom. Cryst.*, **374**, 120-137.
- De Luca, M., Ward, C.M., Ohmori, K., Andrews, R.K. and Berndt, M.C. (1995) *Biochem. Biophys. Res. Commun.*, **206**, 570-576.

- Dennis, M.S. *et al.* (1989) *Proc. Natl. Acad. Sci. (USA)*, **87**, 2471-2475.
- Domont, G.B., Perales, J. and Moussatche, H. (1991) *Toxicon*, **29**, 1183-1194.
- Dzamba, B., Bolton, M. and Desimone, D. (2001) *Frontiers in Molecular Biology, Cell Adhesion*, Oxford University Press. Chapter 4, pp. 100-154. Oxford, UK.
- Eliceiri, B.P. (2001) *Circ. Res.*, **89**, 1104-1110.
- Emsley, P. and Cowtan, K. (2004) *Acta. Cryst. D*, **60**, 2126-2132.
- Escalante, T., Núñez, J., Moura-da-Silva, A.M., Rucuvado, A., Theakston, R.G.D. and Gutiérrez, J.M. (2003) *Toxicol. Appl. Pharmacol.*, **193**, 17-28.
- Evans, P.R. (1993) *Proceedings of CCP4 Study Weekend on Data Collection & Processing*, January 9-10, Daresbury Laboratory, Daresbury, UK.
- Felsenstein, J. (1989) *Cladistics*, **5**, 164-166.
- Fox, J.W. and Bjarnason, J.B. (1998) *Enzymes from snake venom*. Alasken, Inc., Colorado, pp. 599-632.
- Fox, J.W., Shannon, J.D., Stefansson, B., Kamiguti, A.S., Theakston, R.G.D., Serrano, S.M.T., Camargo, A.C.M. and Sherman, N. (2002) *Role of Discovery Science in Toxinology: Examples in Venom Proteomics*. In: Menez, A. (Ed.). Wiley, New York, pp. 97-106.
- Fox, J.W. and Serrano, S.M. (2005) *Toxicon*, **45(8)**, 969-985.
- França, F.O.S. (1997) Tese (Doutorado) - Faculdade de Medicina da Universidade de São Paulo.
- French, G.S. and Wilson, K.S. (1978) *Acta Cryst. A*, **34**, 517-525.
- Gallagher, P., Bao, Y., Serrano, S.M., Laing, G.D., Theakston, R.D., Gutierrez, J.M., Escalante, T., Zigrino, P., Moura-da-Silva, A.M., Nischt, R., Mauch, C., Moskaluk, C. and Fox, J.W. (2005) *Arch.Biochem. Biophys.*, **441(1)**, 1-15.
- Gichuhi, P.M., Ford, W.C. and Hall, L. (1997) *Int. J. Androl.*, **20(3)**, 165-170.
- Gomis-Ruth, F.X., Kress, L.F. and Bode, W. (1993a) *EMBO*, **1**, 4151-4157.
- Gomis-Ruth, F.X., Stöcker, W., Huber, R., Zwilling, R. and Bode, W. (1993b) *J. Mol. Biol.*, **229(4)**, 945-968.
- Gomis-Ruth, F.X., Meyer, E.F., Kress and L.F., Politi, V. (1998) *Protein Sci.*, **7**, 283-287.
- Gong, W., Zhu, X., Liu, S., Teng, M. and Niu, L.J. (1998) *Mol. Biol.*, **283**, 657-668.

- Gould, R.J., Polokoff, M.A., Friedman, P.A., Huang, T.F., Holt, J.C., Cook, J.J. *et al.* (1990) *Proc. Soc. Exp. Biol. Med.*, **195**, 168-171.
- Grams, F., Huber, R., Kress, L.F., Moroder, L. and Bode, W. (1993) *FEBS Lett.*, **335**, 76-80.
- Gutiérrez, J. and Rucavado, A. (2000) *Biochimie*, **82**, 841-850.
- Gutierrez, J.M., Chaves, F., Bolaños, R., Cerdas, L., Rojas, E., Arroyo, O. *et al.* (1981) *Toxicon*, **19**, 493-500.
- Hampel, A., Labanaskas, M., Conners, P.G., Kirkegard, L., Raj Bhandary, U.L., Sigler, P.B. and Bock, R.M. (1968) *Science*, **162**, 1384-1386.
- Hanahan, D. and Weinberg, R.A. (2000) *Cell*, **100**, 57-70.
- Harrison, R.A., Richards, A., Laing, G.D. and Theakston, R.G.D. (2002) *Vaccine*, **15**, 1702-1706.
- Harrison, R.A., Moura-da-Silva, A.M., Laing, G.D., Wu, Y., Richards, A., Broadhead, A., Bianco, A.E. and Theakston, R.G.D. (2000) *Clin. Exp. Immunol.*, **121**, 358-363.
- Hati, R., Mitra, P., Sarker, S. and Bhattacharyya, K.K. (1999) *Crit. Rev. Toxicol.*, **29**, 1-19.
- Hite, L.A., Jia, L.G., Bjarnason, J.B. and Fox, J.W. (1994) *Arch. Biochem. Biophys.*, **308**, 182-191.
- Ho, P.L., Serrano, S.M.T., Chudzinski-Tavassi, A.M., Moura-da-Silva, A.M., Mentele, R., Caldas, C., Oliva, M.L.V., Batista, I.F.C. and Sarno de Oliveira, M.L. (2002) *Biochem. Biophys. Res. Commun.*, **294**, 879-885.
- Hoge, A.R. and Romano-Hoge, S.A.R.W.L. (1981) *Mem. Inst. Butantan*, **42/43**, 179-309.
- Hooper, N.M. (1994) *FEBS Lett.*, **354**, 1-6.
- Huang, T. F. (1998) *Cell Mol. Life Sci.*, **54**, 527-540.
- Huang, T.F. and Niewiarowski, S. (1994) *J. Toxicol*, **13**, 253-273.
- Hynes, R.O. (1992) *Cell*, **110**, 673-687.
- Igarashi, T., Araki, S. and Takeda, S. (2007) *FEBS Letter*, **581(13)**, 2416-2422.
- Jia, L.G., Shimokawa, K., Bjarnason, J.B. and Fox, J.W. (1996) *Toxicon*, **34**, 1269-1276.

- Jia, L.G., Wang, X.M., Shannon, J.D., Bjarnason, J.B. and Fox, J.W. (1997) *Biol. Biochem.*, **272**, 13094-13102.
- Jia, L.G., Wang, X.M., Shannon, J.D., Bjarnason, J.B. and Fox, J.W. (2000) *Arch. Biochem. Biophys.*, **373**, 281-286.
- Jin, H. and Varner, J. (2004) *Br. J. Cancer*, **90**, 561-565.
- Jurgilas, P.B., Neves-Ferreira, A.G., Domont, G.B. and Perales, J. (2003) *Toxicon*, **42**, 621-628.
- Kamiguti, A.S., Theakston, R.D.G., Desmond, H.P. and Hutton, R.A. (1991) *Toxicon*, **29**, 1097-1105.
- Kamiguti, A.S., Slupsky, J.R., Zuzel, M. and Hay, C.R.M. (1994a) *Thromb. Haemost.*, **72**, 244-249.
- Kamiguti, A.S., Desmond, H.P., Theakston, R.D.G., Hay, C.R.M. and Zuzel, M. (1994b) *Biochim. Biophys. Acta*, **1200**, 307-314.
- Kamiguti, A.S., Hay, C.R., Theakston, R.D.G. and Zuzel, M. (1996a) *Toxicon*, **34**, 627-642.
- Kamiguti, A.S., Hay, C.R. and Zuzel, M. (1996b) *Biochem. J.*, **320**, 635-641.
- Kamiguti, A.S., Moura-da-Silva, A.M., Laing, G.D., Knapp, T., Zuzel, M., Crampton, J.M. and Theakston, R.D.G. (1999) *Arch. Biochem. Biophys.*, **369(2)**, 295-301.
- Kamiguti, A.S., Theakston, R.D.G., Watson, S.P., Bon, C., Laing, G.D. and Zuzel, M. (2000) *Arch. Biochem. Biophys.*, **374**, 356-362.
- Kamiguti, A.S., Gallagher, P., Marcinkiewicz, C., Theakston, R.D.G., Zuzel, M. and Fox, J.W. (2003) *FEBS Letter*, **549(1-3)**, 129-134.
- Kim, M., Carman, C.V. and Springer, T.A. (2003) *Science*, **301**, 1720-1725.
- Kleywegt, G.J. and Jones, T.A. (1994) *CCP4/ESF-EACBM Newsletter on Protein Crystallography*, **31**, 9-14.
- Kondo, H., Kondo, S., Ikezawa, H. and Murata, R. (1960) *Jpn. J. Med. Sci. Biol.*, **13**, 43-51.
- Krissinel, E. and Henrick, K. (2005) In: Berthold, M.R. *et al.* (Eds.): *CompLife 2005*, LNBI 3695, 163-174. Springer-Verlag, Berlin, Heidelberg.
- Laemmli, U.K. (1970) *Nature*, **227**, 680-685.
- Laing, G.D., Clissa, P.B., Theakston, R.D.G., Moura-da-Silva, A.M. and Taylor, M.J. (2003) *Eur. J. Immunol.*, **33**, 3458-3463.

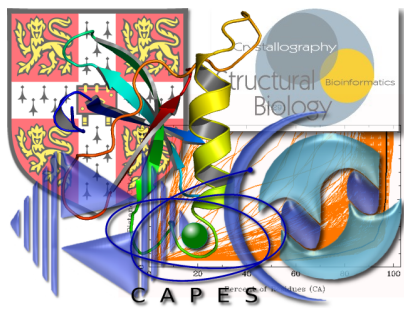
- Laing, G.D. and Moura-da-Silva, A.M. (2005) *Toxicon*, **45**, 987-996.
- Laskowski, R.A., McArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Apply. Cryst.*, **26**, 283-291.
- Leslie, A.G.W. (1992) *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No 26.
- Lindberg, F.P., Gresham, H.D., Reinhold, M.I. and Brown, E.J. (1996) *J. Cell Biol.*, **134**, 1313-1322.
- Lo Conte, L., Chothia, C. and Janin, J. (1999) *J. Mol. Biol.*, **285**, 2177-2198.
- Lomonte, B., Gutierrez, J.M., Borkow, G., Ovadia, M., Tarkowski, A. and Hanson, L.A. (1994) *Toxicon*, **32**, 505-510.
- Lopes-Ferreira, M., Moura-da-Silva, A.M. and Mota, I. (1992) *Toxicon*, **30**, 1591-1602.
- Manual de Diagnóstico e Tratamento de Acidentes por Animais Peçonhentos. (2003) Ministério da Saúde, Fundação Nacional de Saúde.
- Mandelbaum, F.R., Reichel, A.P. and Assakura, M.T. (1982) *Toxicon* **45(8)**, 969-985.
- Marcinkiewicz, C., Calvete, J.J., Vijay-Kumar, S., Marcinkiewicz, M.M., Raida, M., Schick, P. et al. (1999) *Biochemistry*, **38**, 13302-13309.
- Mariano-Oliveira, A., Coelho, A.L., Terruggi, C.H., Selistre-de-Araujo, H.S., Barja-Fidalgo, C. and Freitas, M.S. (2003) *Eur. J. Biochem.*, **270(24)**, 4799-4808.
- Markland, F.S. (1998) *Toxicon*, **36**, 1749-1800.
- Masuda, S., Hayashi, H. and Araki, S. (1998) *Eur. J. Biochem.*, **253**, 36-41.
- Matthews, B.W. (1968) *J. Mol. Biol.*, **33**, 491-497.
- McCoy, A.J., Storoni, L.C. and Read, R.J. (2004) *Acta Cryst. D.*, **60**, 1220-1228.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. (2005) *Acta Cryst. D.*, **61**, 458-464.
- McLane, M.A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I. and Niewiarowski, S. (1998) *Proc. Soc. Exp. Biol. Med.*, **219**, 109-119.
- Moreira, L., Borkow, G., Ovadia, M. and Guitiérrez, J.M. (1994) *Toxicon*, **32**, 977-987.
- Moura-da-Silva, A.M., Della-Casa, M.S., David, A.S, Assakura, M.T., Butera, D., Lebrun, I., Shannon, J.D., Serrano, S.M. and Fox, J.W. (2003) *Arch. Biochem.*

- Biophys.*, **409**, 395-401.
- Moura-da-Silva, A.M., Laing, G.D., Paine, M.J., Dennison, J.M.T., Politi, V., Crampton, J.M. and Theakston, R.D.G. (1996a) *Eur. J. Immunol.*, **26**, 2000-2005.
- Moura-da-Silva, A.M., Theakston, R.D.G. and Crampton, J.M. (1996b) *J. Mol. Biol.*, **43(3)**, 263-269.
- Myles, D.G., Kimmel, L.H., Blobel, C.P., White, J.M. and Primakoff, P. (1994) *Proc. Natl. Acad. Sci. (USA)*, **91**, 4195-4198.
- Nagem, R.A.P., Dauter, Z. and Polikarpov, I. (2001) *Acta Crystallographica Section D*, **57**, 996-1002.
- Neves-Ferreira, A.G., Perales, J., Fox, J.W., Shannon, J.D., Makino, D.L., Garratt, R.C. and Domont, G.B. (2002) *J. Biol. Chem.*, **277**, 13129-13137.
- Niewiarowski, S., Mclane, M.A., Kloczewiak, M. and Stewart, G.J. (1994) *Semin. Hematol.*, **31**, 289-300.
- Ortiz-Stern, A. and Rosales, C. (2003) *Immunol. Lett.*, **90**, 137-143.
- Otero, R.; Nunez, V.; Gutierrez, J.M.; Robles, A.; Estrada, R.; Osorio, R.G. et al. (1997) *Braz. J. Med. Biol. Res.*, **30**, 375-379.
- Paine, M.J., Desmond, H.P., Theakston, R.D. and Crampton, J.M. (1992) *J. Biol. Chem.*, **267**, 22869-22876.
- Parise, L.V., Lee, J. and Juliano, R.L. (2000) *Semin. Cancer Biol.*, **10**, 407-414.
- Petricevich, V.L., Teixeira, C.F.T., Tambourgi, D.V. and Gutiérrez (2000) *Toxicon*, **38**, 1253-1266.
- Pinto, A.F., Terra, R.M., Guimaraes, J.A. and Fox, J.W. (2007) *Arch. Biochem. Biophys.*, **457**, 41-46.
- Polikarpov, I., Perles, L.A., de Oliveira, R.T., Oliva, G., Castellano, E.E., Garratt, R.C. and Craievich, A. (1998) *J. Synchro. Radiat.*, **5**, 72-76.
- Ramos, O.H., Terruggi, C.H., Ribeiro, J.U., Cominetti, M.R., Figueiredo, C.C., Berard, M., Crepin, M., Morandi, V. and Selistre-de-Araujo, H.S. (2007) *Arch. Biochem. Biophys.*, **461(1)**, 1-6.
- Rocha, S.L.G., Lomonte, B., Neves-Ferreira, A.G.C., Trugilho, M.R.O., Junqueira-de-Azevedo, I., de L, M., Ho, P.L., Domont, G.B., Gutierrez, J.M. and Perlaes, J. (2002) *Eur. J. Biochem.*, **269**, 6052-6062.
- Rosenfeld, G. (1971) *Venomous Animal and Their Venoms*, **II**, 346-381.



- Ross, R.S. (2004) *Cardiovasc. Res.*, **63**, 381-390.
- Rucuvado, A., Escalante, T., Franceschi, A., Chaves, F., Leon, G., Cury, Y., Ovadia, M. and Gutierrez, J.M. (2000) *Am. J. Trop. Med. Hyg.*, **63**, 313-319.
- Schattner, M., Fritzen, M., Ventura, J.S., Albuquerque Modesto, J.C., Pozner, R.G., Moura-da-Silva, A.M. and Chusziński-Tavassi, A.M. (2005) *Biol. Chem.*, **386(4)**, 369-374.
- Serrano, S.M., Jia, L.G., Wang, D., Shannon, J.D. and Fox, J.W. (2005) *Biochem. J.*, **391**, 69-76.
- Serrano, S.M., Kim, J., Wang, D., Draguley, B., Shannon, J.D., Mann, H.H., Veit, G., Wagenes, R., Koch, M. and Fox, J.W. (2006) *J. Biol. Chem.*, **281(52)**, 39746-39756.
- Shah, N.K., Rejto, P.A. and Verkhivker, G.M. (1997) *Proteins*, **28**, 421-433.
- Shah, Z.H., Hakkaart, A.J.G., Arku, B., Jong, L., Spek, H.V.D., Grivell, L.A. and Jacobs, H.T. (2000) *FEBS Lett.*, **478**, 267-270.
- Sheu, J.R., Yen, M.H., Kan, Y.C., Hung, W.C., Chang, P.T. and Luk, H.N. (1997) *Biochem. Biophys. Acta*, **1336**, 445-454.
- Shimokawa, K., Shannon, J.D., Jia, L.G. and Fox, J.W. (1997) *Arch. Biochim. Biophys.*, **343**, 35-43.
- Souza, D.H.F., Iemma, M.R.C., Ferreira, L.L., Faria, J.P., Oliva, M.L.V., Zingali, R.B., Niewiarowski and Selistre-de-araujo, H.S. (2000) *Arch. Biochem. Biophys.*, **384**, 341-350.
- Sugiki, M., Maruyama, M., Yoshida, E., Mihara, H., Kamiguti, A.S. and Theakston, D.G. (1995) *Toxicon*, **33**, 1605-1617.
- Sugiki, M., Yoshida, E., Anai, K. and Maruyama, M. (1998) *Toxicon*, **36**, 993-1000.
- Takeda, S., Igarashi, T., Mori, H. and Araki, S. (2006) *EMBO J.*, **11**, 2388-2396
- Tanjoni, I., Butera, D., Bento, L., Della-Casa, M.S., Marques-Porto, R., Takehara, H.A., Gutierrez, J.M., Fernandes, I. and Moura-da-Silva, A.M. (2003a) *Toxicon*, **42**, 801-808.
- Terwilliger, T.C. (2003) *Methods Enzymol.*, **374**, 22-37.
- Theakston, R.G.D. and Reid, H.A. (1983) *Bull. WHO*, **61(6)**, 949-956.
- Trikha, M., de Clerck, Y.A. and Markland, F.S. (1994) *Cancer Res.*, **54**, 4993-4998.
- Usami, Y., Fugimura, Y., Miura, S., Shima, H., Yoshida, E., Yoshika, A., Hirano, K.,

- Suzuki, M. and Titani, K.A. (1994) *Biochem. Res. Commun.*, **201**, 331-339.
- Valente, R.H., Dragulev, B., Perales, J., Fox, J.W. and Domont, G.B. (2001) *Eur. J. Biochem.*, **268**, 3042-3052.
- Watson, S.P. and Gibbins, J.M. (1998) *Immunol. Today*, **19**, 260-264.
- Weiss, H.J., Turitto, V.T. and Baumgartner, H.R. (1986) *Blood*, **67**, 322-330.
- Wolfsberg, T., Primakoff, P., Myles, D. and White, J. (1995) *J. Cell Biol.*, **131**, 275-278.
- Woods, A. and Couchman, J.R. (1998) *Trends Cell Biol.*, **8**, 189-192.
- Xue, W., Mizukami, I., Todd, R.F. and Petty, H.R. (1997) *Cancer Res.*, **57**, 1682-1689.
- Yamamoto, S., Higuchi, Y., Yoshiyama, K., Shimizu, E., Kataoka, M., Hijiya, N. and Matsuura, K. (1999) *Immunol. Today*, **20**, 278-284.
- Zhou, O., Smith, J.B. and Grossman, M.H. (1995) *Biochem.*, **307**, 411-417.
- Zigrino, P., Kamiguti, A.S., Eble, J., Drescher, C., Nischt, R., Fox, J.W. and Mauch, C. (2002) *J. Biol. Chem.*, **277**, 40528-40535.



Apêndice

**RELATÓRIO FINAL - PDEE**

RELATÓRIO FINAL DE ESTÁGIO DE DOUTORADO NO EXTERIOR  
PROGRAMA – PDEE – CAPES

JOÃO RENATO CARVALHO MUNIZ

NÚMERO DO PROCESSO: 3777-05-7

SÃO CARLOS – SP

## 1 – Introdução e Objetivos Propostos

A execução do Programa de Desenvolvimento de Estágio no Exterior (PDEE) pode ser dividido em duas etapas: Desenvolvimento dos estudos cristalográficos – relacionado com o estudo e determinação da estrutura tridimensional da Jararagina a partir de sua forma cristalina e estudos relacionados ao aprendizado e desenvolvimento de novas ferramentas relacionadas a modelagem molecular comparativa por homologia e/ou cálculos diretos (primeiros princípios, *ab initio*) e serão discutidos nos itens 3.1 e 3.2 respectivamente.

Outro aspecto abordado no presente relatório, diz respeito à formação cultural e intelectual do doutorando durante o processo de aprendizagem como idioma, disponibilidade de infra-estrutura, adaptação cultural, etc, seguidos de uma avaliação pessoal sob forma de críticas e sugestões, como sugerido pelas normas do programa PDEE.

### 1.1 – Desenvolvimento dos Estudos Cristalográficos

O projeto de doutorado tem como objetivo o estudo estrutural da bothropasina, uma metaloprotease/desintegrina oriunda do veneno da serpente *Bothrops jararaca*. Esta proteína é uma toxina bastante estudada do ponto de vista bioquímico, devido a suas atividades de inibição da agregação plaquetária e de adesão celular (Kamiguti *et al.*, 1996; Kamiguti *et al.*, 1997; Moura-da-Silva *et al.*, 2001). O estudo estrutural da bothropasina é de grande interesse devido a sua potencial aplicação no desenvolvimento de drogas anti-trombótica e anti-metastática. A molécula de bothropasina é formada por três diferentes domínios: os domínios “metaloprotease” e “desintegrina” estão associados às atividades de inibição da agregação plaquetária e de adesão celular enquanto que o domínio chamado “rico em cisteína” não possui uma função ainda esclarecida.

Nesta pesquisa, a proteína foi isolada do veneno bruto da *B. Jararaca* fornecido pelo Instituto Butantan, com a utilização de cromatografia de troca iônica em coluna DEAE-SefaroseFF, seguida de cromatografia por exclusão de tamanho em coluna Superdex75. A proteína apresenta um adequado nível de pureza e o rendimento do processo de purificação é de aproximadamente 2% da massa inicial do veneno.

Uma característica importante dessa proteína é a capacidade de sofrer

autólise, com produção dos domínios metaloprotease e o domínio desintegrina/rico em cisteína. Para diminuir a ocorrência de autólise e conseqüentemente melhorar o rendimento do processo de obtenção da bothropasina, o inibidor sintético POL647 (Gomis-Ruth *et al.*, 1998) foi utilizado. Testes de cristalização da bothropasina foram realizados primeiramente com as soluções dos Fatoriais 1 e 2 da Hampton. Algumas condições provenientes dessa matriz esparsa mostraram-se mais propícias a uma precipitação sob forma de cristais. Após a obtenção de pequenos cristais essas condições foram refinadas e a condição ideal final foi a de 60 mM de acetato de amônio, 200 mM de acetato de sódio em pH de 5,6 e 30% de precipitante PEG 4000 com a proteína a aproximadamente 8 mg/mL. Os cristais obtidos foram submetidos a experimentos de difração de raios-X no Laboratório Nacional de Luz Síncrotron em Campinas, no “*Brookhaven National Synchrotron Light Source*”, Upton, NY, EUA e no Instituto de Física de São Carlos – USP, em um gerador ânodo rotatório de raios X. Todas as tentativas objetivaram a busca por melhores conjuntos de imagens de difração.

Devido a características intrínsecas da proteína e dos cristais de Jararagina, o melhor conjunto de dados obtido foi a aproximadamente 2,75Å de resolução. Por não possuir estruturas homólogas para os domínios desintegrina e rico em cisteínas, diversas tentativas de incorporação de metais pesados aos cristais protéicos foram executadas e os melhores resultados foram obtidos em ensaios de *soaking* do cristal em 0,5 M de Iodeto de Sódio (NaI). A partir de uma combinação de experimentos de substituição molecular e contribuição dos espalhadores anômalos (Iodo) fases iniciais foram encontradas, permitindo assim a caracterização do domínio metaloprotease (o que representa aproximadamente metade da estrutura) através de um experimento de SAD. A solução das fases possibilitou o emprego de métodos de reconhecimento de padrões em densidades eletrônicas, com o intuito de traçar inicialmente as posições dos átomos que compõem a cadeia principal dos domínios ainda desconhecidos da proteína. Todavia, a interpretação dos mapas de densidade eletrônica tem sido impossibilitada devido a presença do alto conteúdo de pontes dissulfeto (14) nos domínios desintegrina e rico em cisteína e à baixa resolução dos mapas de densidade eletrônica. Através de predição de estruturas secundárias constatamos que os dois domínios ainda inéditos são predominantemente formados por estruturas aleatórias, também chamadas de “*loops*”, laços e voltas. Os “*loops*” e bifurcações dificultam a interpretação adequada dessas regiões, sendo necessária a utilização de métodos mais

eficazes de atribuição de átomos nessas densidades.

## 1.2 – Modelagem Molecular Comparativa

O avanço nos estudos de seqüenciamento de genomas de diversos organismos tem contribuído de maneira significativa para o desenvolvimento de novos métodos capazes de compilar grandes volumes de informações, seja diretamente através de estudos de genômica funcional, estrutural, ou comparativa, seja indiretamente a partir do proteoma, transcriptoma, ou metaboloma (Skolnick *et al.*, 2000; Baker e Sali, 2001). Diversas pesquisas comprovaram que muitas proteínas das mais diferentes famílias compartilham sub-estruturas/motivos e enovelamentos semelhantes (Orengo *et al.*, 1997; Yang e Honing, 2000; Martí-Renom, 2000; Harrison *et al.*, 2002; Pazos e Sternberg, 2004).

A principal motivação desta segunda parte do trabalho consiste em solucionar alguns problemas comumente encontrados durante o processo de modelagem molecular por comparação/“*threading*” protéico e métodos *ab initio*, tais como identificação da(s) proteína(s) que será(ão) utilizada(s) como molde(s) (baseado em um alinhamento seqüencial/estrutural), modelagem das regiões alinhadas e não alinhadas (neste caso por métodos diretos), minimização energética de todo o modelo obtido e subsequente validação do mesmo. Para a realização dessa tarefa, uma nova metodologia de modelagem molecular foi desenvolvida no Laboratório do Prof Blundell: O pacote de programas Orchestra<sup>5</sup>

## 2 – Resultados Alcançados

### 2.1 – Elucidação da Estrutura Tri-dimensional da Bothropasina

Diferentes abordagens foram realizadas no intuito de se obter a estrutura cristalográfica completa da bothropasina. Dentre as abordagens, a mais produtiva foi a utilização do programa *Phaser* (McCoy *et al.*, 2004; 2005) somada às modificações das densidades eletrônicas e achatamento do solvente (CCP4, 1994; Cowtan, 1994), resultando em mapas de densidade eletrônica mais bem interpretáveis.

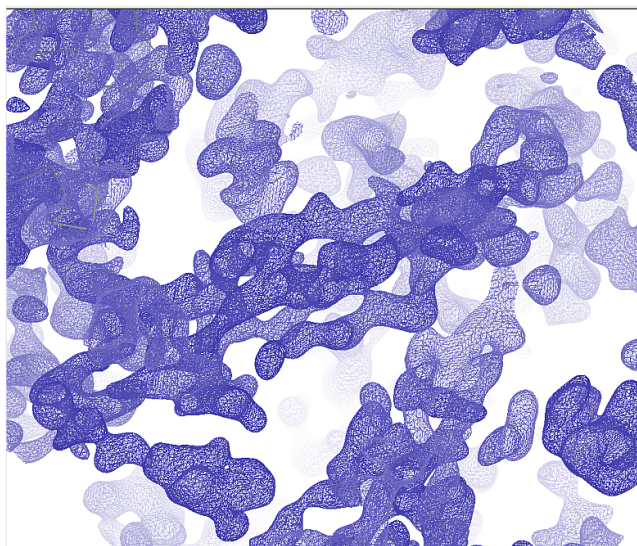
A versão do programa *Phaser* utilizada (versão 2.0) no presente trabalho ainda está em fase de testes e/ou aprimoramentos, mas foi gentilmente empregada na resolução do problema das fases pelo seu próprio criador, o Prof. Dr. Randy Read –

---

<sup>5</sup> Resultados ainda não publicados

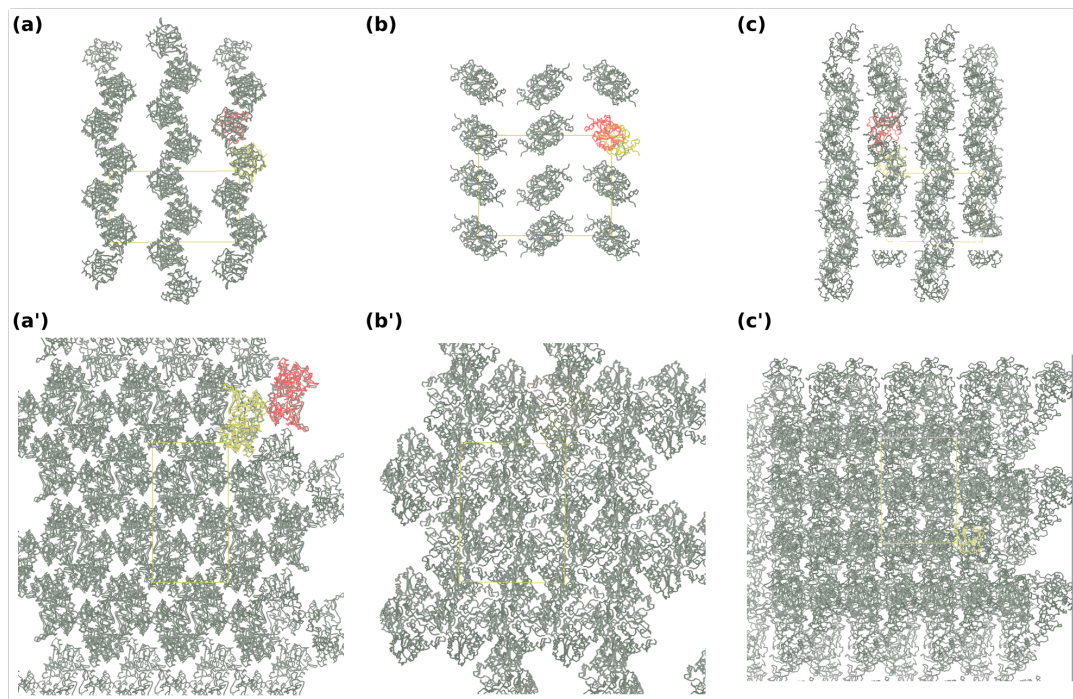
Professor do *Medical Research Council* (MRC) da Universidade de Cambridge. Dentre as características desse programa, pode-se citar a capacidade de faseamento utilizando-se de informações de espalhadores anômalos como átomos pesados (provenientes de um experimento de SAD, MAD, SIRAS, etc por exemplo) além da possibilidade de faseamento por Substituição Molecular.

De posse de mapas melhores (mais interpretáveis) e de estruturas homólogas à bothropasina a construção de todos os três domínios tornou-se possível (figura AI.1). Dados comparativos do tipo “antes e depois” podem ser visualizados na figura AI.2.



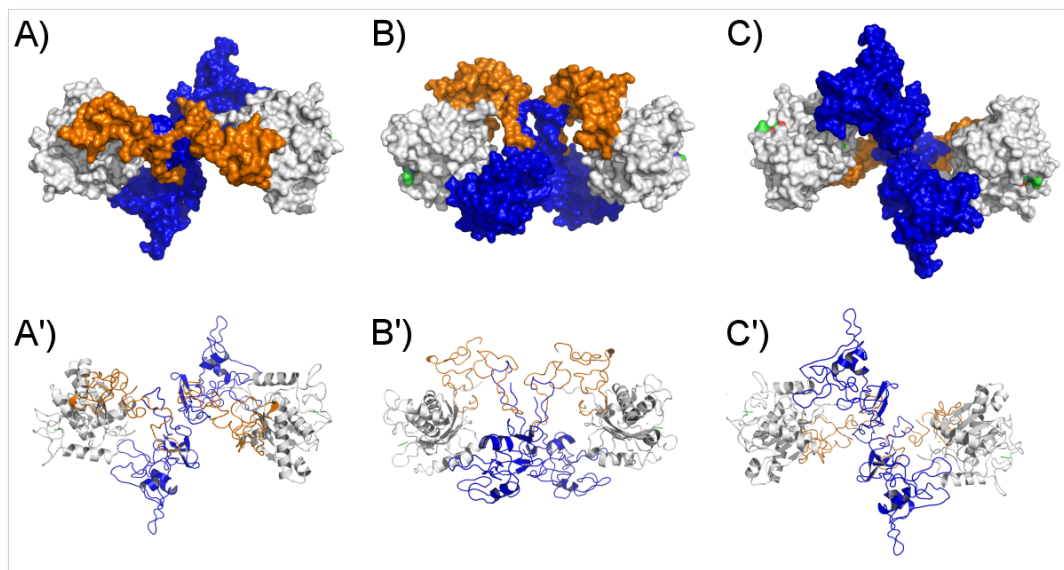
**Figura AI.1: Mapa de densidade eletrônica das regiões dos domínios disintegrina e rico em cisteína.** Com mapas mais bem interpretáveis foi possível traçar os átomos das cadeias principais e laterais para os dois domínios da bothropasina.





**Figura AI.2: Empacotamento cristalino ao longo dos eixos a, b e c da célula unitária.** Uma vez traçados todos os domínios da bothropasina, pode ser observado o contato cristalino entre as moléculas em seu respectivo empacotamento. As figuras (a), (b) e (c) mostram o empacotamento cristalino apenas para o domínio metaloprotease da bothropasina ao longo dos eixos a, b e c enquanto que as figuras (a'), (b') e (c') mostram o empacotamento cristalino para os três domínios da bothropasina e suas interações intermoleculares.

Uma vez traçada toda a cadeia principal, a etapa subsequente foi o posicionamento correto das cadeias laterais, otimização da geometria dos resíduos e refinamento de toda a estrutura. Esses processos foram realizados com o emprego dos programas CNS (Brunger *et al.*, 1998), Refmac5 e (Murshudov *et al.*, 1997). A figura AI.3 ilustra uma representação da bothropasina em sua unidade assimétrica (duas moléculas) destacando seus três domínios através de diferentes representações.



**Figura AI.3: Diferentes representações para as duas moléculas da bothropasina na unidade assimétrica.** A figura mostra a bothropasina em superfície e “cartoon” coloridas para os três domínios facilitando assim a identificação dos domínios metaloprotease, desintegrina e rico em cisteína (cinza, laranja e azul, respectivamente). A, B, C, A', B' e C' correspondem a diferentes orientações: A e A', visão superior; B e B', visão lateral; C e C', visão inferior, respectivamente.

A atual etapa de desenvolvimento do projeto de doutorado caracteriza-se pelo refinamento da estrutura da bothropasina com a utilização dos programas CNS e Refmac5. Dados do refinamento da estrutura da bothropasina são mostrados na Tabela AI.I. A estrutura será depositada em banco de estrutura protéica e um artigo descrevendo a estrutura da bothropasina e sua relação com suas diversas atividades biológicas está sendo redigido.

Portanto, será possível o cumprimento de todos os planos propostos no projeto de doutoramento dentro do prazo inicialmente estabelecido de quatro anos.

**Tabela AI.I** – Estatísticas da coleta de dados e refinamento. O conjunto coletado é proveniente de um “soaking” em NaI 0,5 M por aproximadamente 2 minutos.

Bothropasina – NaI	
Fonte de raios X	MX1, LNLS-Brasil
Comprimento de onda ( $\lambda$ )	1,4270
Ângulo de oscilação (°)	1,00
Número de imagens	344
Grupo espacial	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Parâmetros da célula unitária	
a (Å)	72,5
b (Å)	101,2
c (Å)	133,6
Limites na resolução	58,93 – 2,80 (2,9 – 2,80)
Mosaicidade	0,75
Reflexões totais	880327
Reflexões únicas	26293
Redundância	13,4 (13,3)
Completeza (%)	99,9 (100,0)
$\langle I/\sigma(I) \rangle$	25,0 (5,1)
R <sub>sym</sub>	8,3 (48,5)
R <sub>factor</sub>	0,293
R <sub>free</sub>	0,376

## 2.2 – Modelagem Molecular Comparativa

O processo de desenvolvimento e aprendizado da modelagem molecular comparativa foi cumprido de maneira mais que satisfatória. Além de poder contar com a colaboração direta dos pesquisadores e desenvolvedores das ferramentas de bioinformática, uma grande oportunidade me foi dada pelo próprio Prof. Blundell sob forma de convite a participar do sétimo encontro CASP<sup>6</sup> (“*Critical Assessment of Techniques for Protein Structure Prediction*” 7 – 2006 “<http://predictioncenter.org/casp7/Casp7.html>”) como um dos integrantes de seu grupo representando a Universidade de Cambridge.

A CASP tem como principais objetivos estabelecer as capacidades e limitações dos atuais métodos de modelagem de estrutura protéica a partir de sua seqüência de aminoácidos, visando determinar os progressos que vêm sendo feitos na área, assim como localizar os maiores “gargalos” nas predições estruturais (Moult *et al.*, 2003). Os projetos na CASP são divididos basicamente em três categorias: (1)

6 A CASP é uma reunião mundial bienal entre pesquisadores especializados na área de bioinformática.

7 Acesso em 08/11/2006

modelagem molecular comparativa; (2) reconhecimento de enovelamentos (padrões estruturais, domínios, motivos, etc.) e (3) métodos de obtenção de novos enovelamentos. Para isso, os mais diversos bancos de dados são selecionados e divididos em grupos com diferentes tipos de estruturas.

Um total de 23 proteínas (Figura AI.4) foram preditas pelo grupo utilizando-se diferentes abordagens como, principalmente, os programas MODELLER, RAPPER, CHORAL, CODA e FUGUE – todos desenvolvidos (com exceção do MODELLER, que tem o Prof. Blundell como um dos autores na primeira versão programa) no próprio Departamento de Bioquímica – Grupo de Cristalografia, Biologia Estrutural e Bioinformática da Universidade de Cambridge.

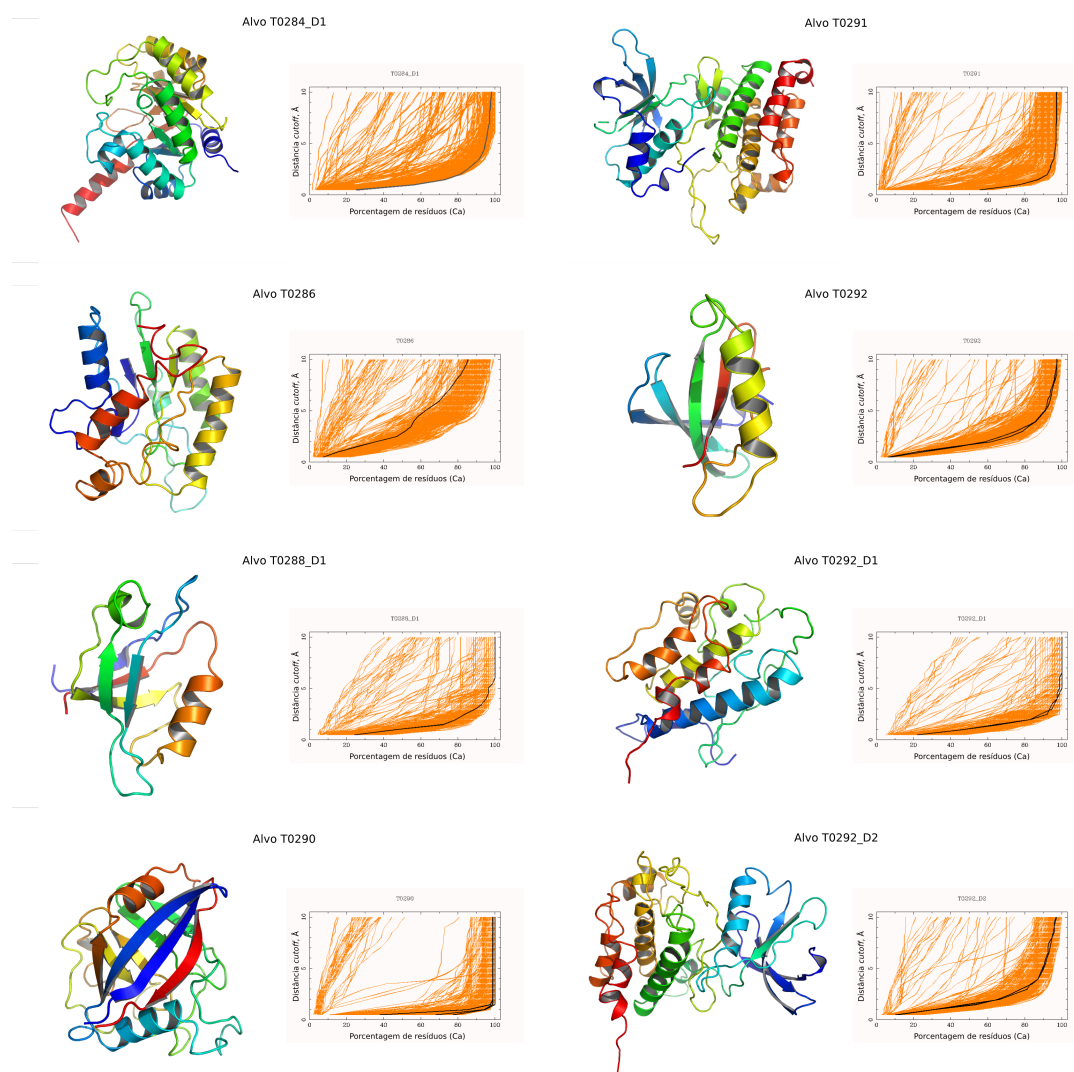
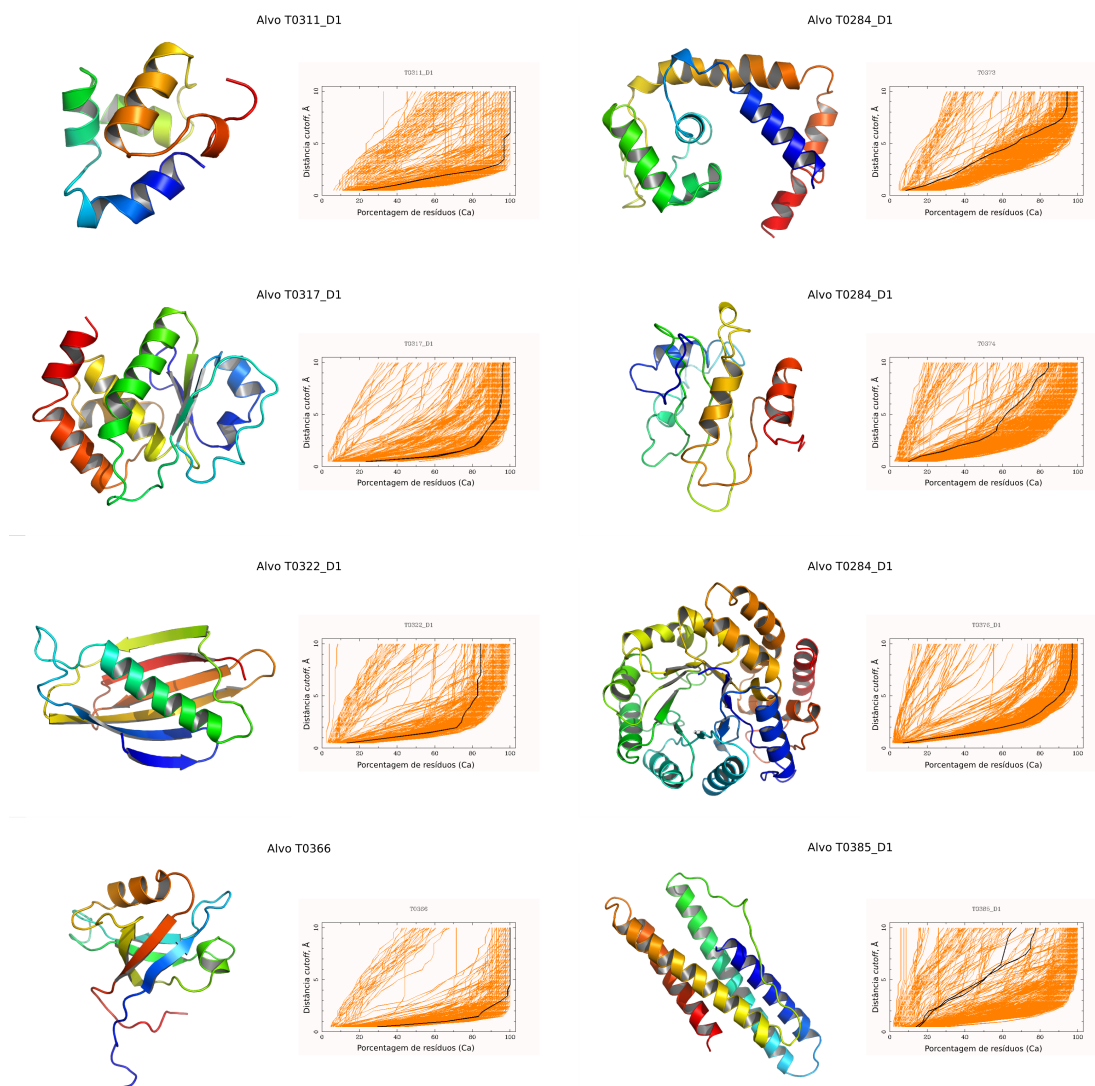


Figura AI.4 – continuação e legenda na próxima página.

Figura AI.4 - continuação



**Figura AI.4: Resultados obtidos pelo grupo na sétima reunião da CASP.** O código de uma letra e quatro dígitos numéricos, refere-se ao nome dado ao alvo a ser modelado. À esquerda, uma representação sob forma de “cartoon” do modelo predito a partir da seqüência primária da proteína e a direita, uma comparação entre o modelo predito pelo grupo (linha destacada na cor cinza escura), em relação a todos os modelos submetidos à CASP (linhas alaranjadas). No eixo das abscissas temos a porcentagem do número de carbonos alfa (CA) corretamente alinhados em relação a estrutura original e nas ordenadas, o desvio entre o modelo final e a estrutura cristalográfica em Å.

Como observado na Figura AI.4, a grande maioria dos modelos preditos foi muito próxima às estruturas cristalográficas, como por exemplo os modelos T0284, T0288, T0290, T091, T092, T0311, T0366 e T0376 que obtiveram, quando

comparadas as estruturas, um desvio máximo de 10 Å e mais de 95% de resíduos alinhados.

No entanto, alguns modelos ficaram aquém da expectativa, como foi o caso do T0373, T0374 e T0377 com altos desvios em cerca de 60% a 80% de resíduos que representariam parte da estrutura cristalográfica. Nesses casos, a baixa identidade sequencial (em certos casos menores que 5% de identidade), foi o fator determinante no insucesso das predições, um dos motivos que justificam a necessidade de reuniões como essas.

### **3 – Relevância na Formação Intelecto-Cultural**

Indubitavelmente, a possibilidade que me foi dada pela CAPES sob forma de auxílio do tipo bolsa de estudos para desenvolvimento de estágio no exterior, contribuiu de maneira veemente em minha formação intelecto-cultural por vários motivos. O contato direto com a língua inglesa foi muito importante para o aperfeiçoamento e aprendizagem da fala, escuta, leitura e escrita do idioma. De certa maneira, dada a exposição à língua, você se sente obrigado a se esforçar de maneira saudável para entender e compreender o idioma anglo-saxônico falado ao seu redor, seja através de contatos com outras pessoas, seja através da mídia como TV, cinema, rádio, jornais, internet e revistas. Todos esses fatores contribuem de maneira importante no processo de aperfeiçoamento e treinamento do idioma.

Minha adaptação foi bastante tranquila, calma e naturalmente. Não encontrei grandes dificuldades em me adaptar à língua e começar a me comunicar de maneira fluente com as pessoas do grupo onde trabalhei, com companheiros de alojamento e colegas que tive.

Uma vez melhor adaptado, passei a desfrutar melhor da infra-estrutura ao meu redor. O laboratório tinha todas as condições necessárias para o desenvolvimento do meu trabalho, como mesa, computador de uso exclusivo, acesso a rede de periódicos, livros, revistas, etc... além de contar com o total apoio intelectual do meu co-orientador estrangeiro, Prof. Dr. Tom Blundell, do Prof. Dr. Randy Read, Dr. Dima Chirgadze, Dr. Rinaldo Montalvão, Dr. Nicholas Furnham, Dra. Eva Maria Priego, Tammy Cheng, dentre tantos outros que de maneira direta ou indireta me ajudaram nas discussões, implementações e apoio ao meu projeto me deixando ainda mais a vontade para desenvolvê-lo.



## 4 – Agradecimentos

Gostaria de agradecer mais uma vez ao apoio da minha orientadora Profa. Dra. Dulce Helena Ferreira de Souza por toda orientação, apoio e incentivo, ao Prof. Dr. Glaucius Oliva e Prof. Dr. Richard Charles Garratt por todos os contatos, apoios e incentivos, assim como Cristiane Gomes Lazarini Estella e Wladerez Aparecida Gounella Caiado do Serviço de Pós-Graduação do Instituto de Física de São Carlos, por toda a atenção, coleguismo, apoio, incentivo e principalmente a CAPES por todo o suporte que me foi dado para a realização desse Programa de Desenvolvimento de Estágio no Exterior. Muito obrigado.

## 5 – Referências Bibliográficas

- Baker, D. and Sali, A. (2001) *Science*, **294**, 93-96.
- Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Cryst.*, **D54**, 905-921.
- CCP4 (1994), *Acta Cryst.*, **D50**, 760-763.
- Cowtan, K. (1994) *ESF-EACBM Newsletter on Protein Crystallography*, **31**, p34-38.
- Furnham, N., Doré, A., Chirgadze, D., de Bakker, P., DePristo, M. and Blundell, T. (2006) *Structure*, **14(8)**, 1313-1320.
- Gomis-Ruth, F.X., Meyer, E.F., Kress, L.F. and Politi, V. (1998) *Protein Sci.*, **7**, 283-292.
- Harrison, A., Pearl, F., Mott, R., Thornton, J. and Orengo, C. (2002) *J. Mol. Biol.*, **323**, 909-926.
- Kamiguti, A.S., Hay, C.R.M. and Zuzel, M. (1996) *Biochem J.*, **320**, 635-641.
- Kamiguti, A.S., Moura-da-Silva, A.M., Laing, G.D., Knapp, T., Zuzel, M., Crampton, J.M. and Theakston, R.D.G. (1997) *Biochim. Biophys. Acta*, **1335**, 209-217.
- Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291-325.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. (2005) *Acta Cryst.*, **D61**, 458-464.
- McCoy, A.J., Storoni, L.C. and Read, R.J. (2004) *Acta Cryst.*, **D60**, 1220-1228.
- Moura-da-Silva, A.M., Marcinkiewicz, C., Marcinkiewicz, M., Niewiarowski, S. (2001) *Thrombosis Research*, **2**, 153-159.
- Murshudov, G.N., Vagin, A.A. and E.J.Dodson (1997), *Acta Cryst.*, **D53**, 240-255.
- Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) *Structure*, **5**, 1093-1108.
- Pazos, F. and Sternberg, M.J.E. (2004) *Proc. Natl. Acad. Sci.*, **101**, 14754-14759.
- Skolnick, J., Fetrow, J.S. and Kolinski, A. (2000) *Nature*, **18**, 283-287.
- Takeda, S., Igarashi, T., Mori, H. and Araki, S. (2006) *Embo J.*, **25**, 2388-2396.
- Yang, A.S. and Honig, B. (2000) *J. Mol. Biol.*, **301**, 665-678.

**Two eIF4A helicases in *Trypanosoma brucei* are functionally distinct**

Julia, Nina B. <sup>1</sup>, Christian R. S. Reis, Rodolfo Katz, João R. C. Muniz<sup>2</sup>, Nancy Standart<sup>1</sup>, and Osvaldo P. de Melo Neto<sup>1\*</sup>

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

**Analysis of eIF4F subunit homologues in *Xylella fastidiosa*: correlations between <sup>13</sup>C NMR data, and molecular models**

João Renato C. Muniz<sup>2</sup>, Helvécio Della Coleta Filho<sup>4</sup>, João Alves de Sousa<sup>4,5</sup>, Marcos Antonio Machado<sup>6</sup>, Richard C. Garratt<sup>6</sup>, Luiz Alberto Colnago<sup>7\*</sup>

# Apêndice

## ARTIGOS PUBLICADOS EM REVISTAS INDEXADAS





ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Biochemical and Biophysical Research Communications 323 (2004) 987–995

BBRC

[www.elsevier.com/locate/ybbrc](http://www.elsevier.com/locate/ybbrc)

## Fatty acid synthesis in *Xylella fastidiosa*: correlations between genome studies, $^{13}\text{C}$ NMR data, and molecular models

Denise Osiro<sup>a,b</sup>, João Renato C. Muniz<sup>c</sup>, Helvécio Della Coleta Filho<sup>d</sup>,  
Alessandra Alves de Sousa<sup>d,e</sup>, Marcos Antonio Machado<sup>d</sup>,  
Richard C. Garratt<sup>c</sup>, Luiz Alberto Colnago<sup>a,\*</sup>

<sup>a</sup> Embrapa Instrumentação Agropecuária, Rua XV de Novembro 1452, CEP 13560-970 São Carlos, SP, Brazil

<sup>b</sup> Instituto de Química de São Carlos, Universidade de São Paulo, CEP 13560-970 São Carlos, SP, Brazil

<sup>c</sup> Instituto de Física de São Carlos, Universidade de São Paulo, CEP 13560-970 São Carlos, SP, Brazil

<sup>d</sup> Centro APTA Citros 'Sylvio Moreira' Instituto Agronômico—CP 04, 13490-970 Cordeirópolis SP, Brazil

<sup>e</sup> Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica W5 Norte Final, 70770-900 Brasília, DF, Brazil

Received 19 August 2004

Available online 11 September 2004

### Abstract

*Xylella fastidiosa* was the first plant pathogen to have its complete genome sequence elucidated. Routine database analyses suggested that two enzymes essential for fatty acid synthesis were missing, one of these is the holo-acyl-carrier-protein synthase. However, here we demonstrate, using  $^{13}\text{C}$  NMR spectroscopy, that *X. fastidiosa* is indeed able to synthesize fatty acids from acetate via an apparently conventional metabolic pathway. We further identify a gene product HetI, an alternative phosphopantetheinyl transferase, which we propose to fill the missing link. Homology modeling of HetI shows conservation of the Coenzyme A binding site suggesting it to be an active enzyme and reveals several interesting structural features when compared with the surfactin synthase-activating enzyme, on which the model was built. These include a simplified topology due to N- and C-terminal deletions and the observation of a novel serine ladder.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** *Xylella fastidiosa*; NMR; Fatty acid biosynthesis; Molecular modeling; Holo-acyl-carrier-protein synthase; HetI

*Xylella fastidiosa* is a gram-negative, xylem-limited bacterium, pathogenic to several plant species, and is the causal agent of citrus variegated chlorosis (CVC) in several sweet orange varieties in Brazil, as well as Pierce's disease in grapevines in the United States [1,2]. As a consequence of its significant impact on commercial agriculture it was the first plant pathogen to have its complete genome sequence published [1].

Almost all of our current knowledge concerning the composition and metabolic pathways of *X. fastidiosa* is derived from the comparison of its open reading frames

with the sequences deposited in public databases. These comparisons have allowed the assignment of 47% of the ORFs to a putative biological function [1]. However, several genes related to important metabolic pathways were not identified by routine database searching algorithms for one of a number of possible reasons. Either they (a) are genuinely lacking in the genome, (b) are too distantly related to homologs within the database to be readily detectable on the basis of sequence alone or (c) have been replaced by a non-homologous counterpart.

One of the metabolic gaps reported in the original study is related to the ability of the bacterium to synthesize fatty acids. On the one hand, the genome analysis

\* Corresponding author. Fax: +55 16 33725958.

E-mail address: [colnago@cnpdia.embrapa.br](mailto:colnago@cnpdia.embrapa.br) (L.A. Colnago).

showed the presence of most of the enzymes necessary for synthesizing and elongating fatty acids from acetate. On the other hand, two *Escherichia coli* enzymes, holoacyl-carrier-protein synthase (ACPs), which converts inactive apoACP to its active holo form, and Enoyl-ACP-reductase, which reduces the double bonds, were not identified in the *X. fastidiosa* genome [1]. This is a surprising observation given that fatty acids are ubiquitous in living organisms, playing essential roles in the makeup of biological membranes, for example, a vital component of every cell [3]. Certainly an intact biosynthetic pathway for fatty acids would be anticipated in *X. fastidiosa*.

Functional genomic studies of *X. fastidiosa* strains have thus far focused mainly on the transcriptome [4,5] and proteome [6–8] but there has been little work so far reporting studies of the set of synthetic metabolites which constitute the bacterium's metabolome [9]. This is the focus of the current report which concerns the use of nuclear magnetic resonance (NMR) spectroscopy to identify the presence of fatty acids in *X. fastidiosa* and to demonstrate their *in vivo* synthesis. NMR spectroscopy is a recognized technique used in the determination of the chemical structure and composition of purified and raw biological materials in both solution and the solid state [10–13] as well as for following metabolic pathways *in vivo* [10,14]. Here, its application to the demonstration of an intact metabolic pathway for the synthesis of fatty acids in *X. fastidiosa* begs the question of how the apparent gaps identified in the original genome study are effectively overcome. An answer is given for the case of the ACP synthase, where the identification and homology modeling of an alternative 4'-phosphopantetheinyl transferase is suggested to assume this role.

The information generated by such metabolomic approaches will provide a complementary dataset, which, together with genomic data, should help in the elucidation of the mechanism of fatty acid biosynthesis in *X. fastidiosa* as well as in the identification of missing links in other apparently incomplete metabolic pathways.

## Materials and methods

**Nuclear magnetic resonance.** Cells of *X. fastidiosa* 9a5c strain were grown in PW broth [15]. The cultures were agitated at 128 rpm at 28 °C for 8 days and centrifuged at 4000g for 20 min. The pellet was suspended and centrifuged twice. The final pellet was lyophilized for use in solid state NMR analysis. The dried sample was also used for the extraction with chloroform to evaluate the presence of free fatty acid [12].

For the acetate incorporation experiment, the pellet obtained in the first centrifugation was agitated in 200 mL of minimal mineral media [16], using 100 mg of sodium acetate 99% [ $^{13}\text{C}$ ] from Cambridge Isotopes laboratory, as a carbon source, for four days. The culture was then centrifuged and re-suspended in 0.5 mL of deuterated water.

The  $^{13}\text{C}$  NMR experiments were performed on a Varian Inova 400 (9.4 T) spectrometer. The solid state spectrum was acquired using a single pulse sequence with magic angle sample spinning [12,13]. The

sample (about 200 mg) was packed into a 5 mm zirconium rotor and spun at the magic angle at a frequency of 9 kHz. The spectra were averaged 10,000 times with a flip angle of  $\pi/2$  ( $\text{pw} = 7.2 \mu\text{s}$ ), acquisition time,  $a_1 = 100 \text{ ms}$ ; spectral width,  $s_w = 40 \text{ kHz}$ ; and recycle time,  $r_1 = 3 \text{ s}$ . The  $^{13}\text{C}$  NMR spectrum in suspension in deuterated water was acquired using a 5 mm solution probe,  $\text{pw} = 19 \mu\text{s}$  ( $\pi/2$ ),  $a_1 = 1 \text{ s}$ ,  $s_w = 40 \text{ kHz}$ , and  $r_1 = 3 \text{ s}$ .

**Database searches.** The recent description of a second 4'-phosphopantetheinyl transferase gene in *E. coli*, coding for a product named YhhU, which is capable of suppressing an ACPs conditional lethal phenotype of the bacterium, suggested the presence of an alternative pathway for the generation of holoACP from apoACP in *E. coli* [17]. This observation together with the absence of a homolog to the ACPs itself in *X. fastidiosa* stimulated a specific re-examination of the genome for a possible YhhU like molecule. Database searches of the *X. fastidiosa* genome [18] (<http://aeg.lbi.ic.unicamp.br/xf/>) using the YhhU sequence as probe were performed with the BLASTp program [19] employing default parameters.

Subsequent searches of GenBank [20] (<http://www.ncbi.nlm.nih.gov/>) were similarly performed with BLASTp using the *X. fastidiosa* HetI sequence (XF1934) for the identification of 4'-phosphopantetheinyl transferase homologs. Fold recognition searches were made using GenThreader [21] and PSI-BLAST [22] using default parameters against the Protein Database [23] in order to improve the alignments to be used for subsequent homology modeling. The final alignment was manually adjusted after visual inspection of the structural homolog Sfp (surfactin synthetase-activating enzyme), a 4'-phosphopantetheinyl transferase of known structure from *Bacillus subtilis* [24].

**Molecular modeling.** Molecular models for HetI were built using restraint based homology modeling as implemented in the program MODELLER 6a [25]. The final alignment of the *X. fastidiosa* HetI protein against the Sfp sequence as described above was initially used as input to the MODELLER program together with the atomic coordinates of the latter (PDB file 1QR0). The low sequence identity between HetI and Sfp inevitably led to alignment ambiguities requiring the construction of several alternative models based on slightly different alignments (over 400 in total). This process included the deliberate misalignment of residues adjacent to deletions in order to facilitate gap closure.

Models were generated by coordinate copying for all atoms in common between the two sequences and completed using internal coordinates from the CHARM topology library [26]. A 4 Å coordinate randomization was applied to each structure prior to optimization using up to 300 cycles of the variable target function method employing conjugate gradients. Refinement was subsequently performed in Cartesian space using the standard full molecular dynamics algorithm with simulated annealing as implemented in MODELLER 6a. Default limits were applied to the distance restraints and the Coenzyme A molecule of the Sfp structure was included in the model calculation.

Besides the MODELLER pseudo-energy term, the quality of the models generated was also independently evaluated by the programs PROCHECK [27], VERIFY 3D [28], and the QUALTY option [29] of WHATIF [30] and used to select a representative model for structure analysis.

## Results

### Nuclear magnetic resonance

*Xylella fastidiosa*, as its name suggests, is a very slow growing bacterium [31] and the  $^{13}\text{C}$  NMR procedures normally used for the analysis of fast growing bacteria such as *E. coli* [14] are not efficient. To verify the pres-

ence of fatty acids as part of the makeup of the bacterium we analyzed a dried sample of *X. fastidiosa* with  $^{13}\text{C}$  NMR spectroscopy using a single pulse sequence combined with magic angle sample spinning (Fig. 1A). This technique has been used to enhance spectral resolution and for the detection of mobile molecules, like fatty acid derivatives [12,13]. The *X. fastidiosa* spectrum is typical of a mono-unsaturated fatty acid chain, having strong signals at  $\delta 14$ , assigned to the terminal methyl group,  $\delta 23$ – $\delta 32$  due to the methylene groups, and weak peaks at  $\delta 130$  and  $\delta 173$  assigned to olefinic and carboxyl carbons, respectively [10,12,13]. The signal often seen at  $\delta 34$  and assigned to  $\text{C}_2$  is absent in the spectrum from *X. fastidiosa*. This signal is not observed in experiments using low power decoupling [12] as used here. The strong dipolar interaction with bonded hydrogens has not been eliminated, leading to broadening of the signal. The peak due to the carboxyl carbon ( $\text{C}_1$ ) is, however, observed, albeit broadened, as a consequence of its weaker dipolar interaction. The signals expected at  $\delta 62$  and  $\delta 70$  relative to the glycerol moiety of triglycerides and phospholipids were not observed in the present spectrum.

To confirm the ability of *X. fastidiosa* to synthesize fatty acids from acetate we analyzed a *X. fastidiosa* sample incubated for 4 days in the presence of  $[1-^{13}\text{C}]$ acetate as the only source of carbon. The  $^{13}\text{C}$  NMR spectrum (using a solution probe) of suspended bacteria in deuterated water is given in Fig. 1B, showing signals due to molecules soluble in water or with high inherent mobility. In this spectrum, besides several other compounds such as bicarbonate ( $\delta 160.5$ ), the characteristic peaks of *cis* mono-unsaturated fatty acids, like oleic, palmitoleic or vaccenic acids, with signals at  $\delta 23$ ,  $\delta 25$ ,  $\delta 27$ ,  $\delta 30$ ,  $\delta 130$ ,

and  $\delta 172$ , are clearly observable [10,12,13]. It also shows peaks from  $\delta 60$  to  $\delta 85$ , related to glycerol and saccharide carbons.

#### Database searches and sequence alignment

A BLAST search against the *X. fastidiosa* genome using the *E. coli* YhhU gene product as probe identified the HetI protein (XF1934) to be a distant homolog (32% identity over 126 alignment positions) and therefore a potential 4'-phosphopantetheinyl transferase. No crystal structure is available for YhhU, so the HetI sequence was subsequently used as probe against the PDB in the search for 4'-phosphopantetheinyl transferases of known crystal structure. The GenThreader and PsiBLAST algorithms aided in identifying and aligning HetI to the surfactin synthetase-activating enzyme, Sfp, from *B. subtilis*, the only homolog of known three-dimensional structure. Fig. 2 shows the final alignment between HetI and Sfp in which the secondary structures derived from the crystal structure are identified. Here, we use a nomenclature for the secondary structures derived from the assignments in the header of the PDB file (1QR0), which differ slightly from those used by [24] in their original description of the structure. Only 40 of the 197 aligned residues are conserved, corresponding to an overall sequence identity of 20%. These residues are not homogeneously distributed along the sequence but concentrate in the major elements of secondary structure. All residues identified as important for the binding of the pyrophosphate and adenine base of Coenzyme A are conserved or conservatively substituted in HetI (Ser67, His68, Lys136, Gly51, Arg52, and Pro53) as are the acidic residues (Asp86, Glu88, and Glu132) responsible for the concomitant binding of  $\text{Mg}^{2+}$  (Fig. 2). Two major deletions are observed in HetI, corresponding to the loss of 23 residues at the N-terminus and 15 at the C-terminus as compared to Sfp.

#### Quality of the model

The HetI model was built on the basis of the only available parent structure, Sfp, using the coordinate file of its complex with CoA (1QR0) as described above. Stereochemical analysis of the final model using PROCHECK showed that 95.4% of the residues lie in the most favorable regions of the Ramachandran plot, with 3.6% in the additionally allowed regions and only 1.0% of outliers. Model validation using chemical environments as defined by VERIFY 3D yielded a total score of 53.6 compared with a mean expected score of 89.6 and a minimal acceptable score of 40.3, for a protein of 197 residues. This suggests that the model has overall self-consistency in terms of sequence-structure compatibility. However, one region of the structure (His147–Val161) was detected to be problematic by this method. This part of the structure corre-

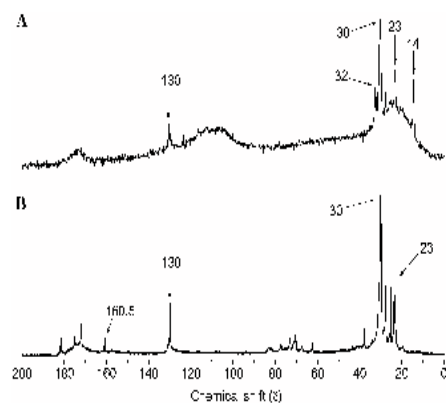


Fig. 1. (A)  $^{13}\text{C}$  NMR spectrum of the lyophilized *X. fastidiosa* sample using a single pulse sequence combined with magic angle spinning. (B)  $^{13}\text{C}$  NMR spectrum of *X. fastidiosa* in deuterated water suspension, after four days culture in enriched  $[1-^{13}\text{C}]$ acetate as a sole carbon source.



990

D. Osiro et al. / Biochemical and Biophysical Research Communications 323 (2004) 987–995

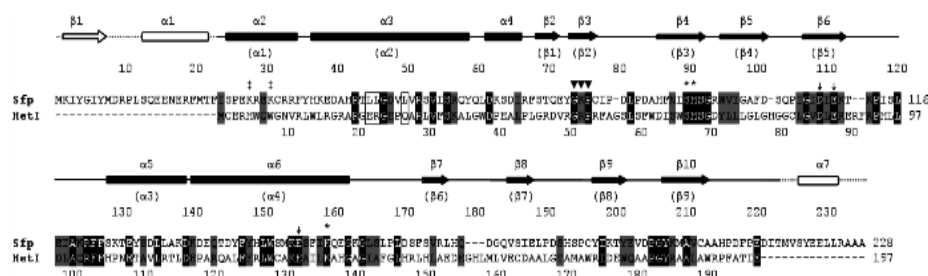


Fig. 2. The final sequence alignment for Sfp and HetI with the nomenclature for the secondary structural elements indicated (HetI in parentheses). Residue identities and conservative substitutions are shown with black and grey backgrounds, respectively. The CoA/Mg<sup>2+</sup> cofactor binding site is indicated in the following manner: conserved residues involved in binding the adenine base (▼), the pyrophosphate group (\*), and the Mg<sup>2+</sup> (‡). Lys28 and Lys31 of Sfp, which interact with the 3' phosphate, are also indicated (‡). The numbers which appear above the sequences refer to alignment positions and those below refer specifically to the HetI sequence itself. The open boxed residues which appear in helix  $\alpha 3$  of Sfp are hydrophobic residues which participate in the occluded face of the helix but which are solvent-exposed hydrophilic residues in HetI.

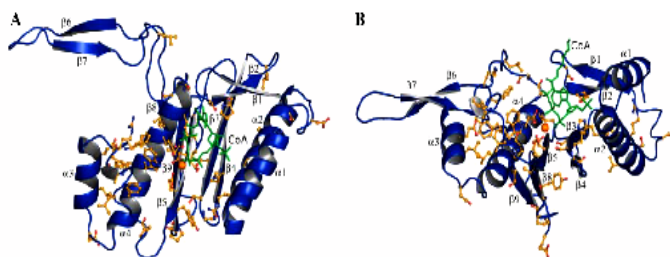


Fig. 3. Ribbon representation of HetI in two different views. The conserved residue side chains (orange), CoA (green), and the Mg<sup>2+</sup> ion (orange sphere) are shown explicitly. Structural figures were generated with PyMOL [32]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sponds to a protruding wing in the structure of Sfp, including strands  $\beta 6$  and  $\beta 7$  (Fig. 3). This unusual structure may be the result of crystal packing in the case of the parent structure Sfp, where these two  $\beta$ -strands form part of a four-stranded sheet with a symmetry-related molecule. The poor evaluation of this part of the HetI model in terms of residue environments is probably a consequence of this artifact and suggests that in solution this protrusion may not exist.

Atomic contact analysis of the model using the QUALTY option of the program WHATIF gave an overall mean value of  $-1.637$ , which is acceptable given the reference values. Some regions of the structure show poor atomic contacts but these are generally loop regions and not important for Coenzyme A binding. Given the low sequence identity between parent structure and target, the representative model chosen for analysis appears to be of good overall quality.

*Overall description of the structure*

By definition the HetI structure is similar to that of Sfp but with some notable differences. It is divided into two similarly folded domains of approximately equal size. The first domain corresponds to the N-terminal

half of the structure and extends from Met1 to Asp86, while the C-terminal domain runs from Leu87 to Glu197. Each domain has a similar mixed  $\alpha/\beta$  topology based on that originally described for Sfp, but modified as a consequence of two large deletions at the N- and C-termini in HetI, as described above. The main  $\beta$ -sheet in the N-terminal domain is composed of the anti-parallel strands  $\beta 3$  and  $\beta 4$  (Figs. 2 and 3) which are packed against the corresponding sheet from the C-terminal domain (strands  $\beta 5$ ,  $\beta 8$ , and  $\beta 9$ ) forming an unusual wedge-shaped  $\beta$ -sandwich at the core of structure. Each domain has two large  $\alpha$ -helices packed against the external sides of this  $\beta$ -sandwich and two additional  $\beta$ -strands ( $\beta 1/\beta 2$  and  $\beta 6/\beta 7$ , respectively) forming hairpins. In the N-terminal domain this hairpin is folded back onto the body of the structure whilst in the C-terminal domain it protrudes into the solvent and corresponds to the region described above (147–161) which is poorly evaluated in terms of residue environments.

*CoA/Mg<sup>2+</sup> binding site*

Fig. 4 shows a schematic diagram of the HetI model including CoA/Mg<sup>2+</sup> bound to the structure. The

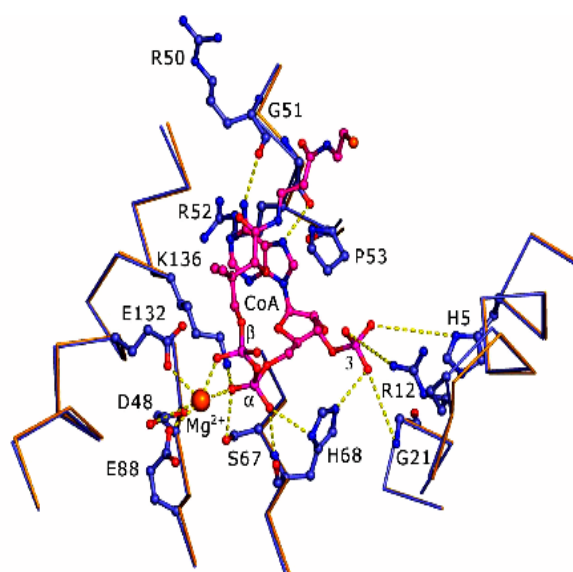


Fig. 4. The CoA (purple)/Mg<sup>2+</sup> (orange sphere) binding pocket of HetI. The C $\alpha$  trace of selected regions of HetI is shown in blue superposed on Sfp in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cofactor binding site is located at the exposed edge of the wedge-shaped  $\beta$ -sandwich (composed of strands  $\beta$ 3 and  $\beta$ 5), where the two sheets come closest together. Most of the intermolecular interactions observed in the Sfp structure are conserved in the complex HetI–CoA/Mg<sup>2+</sup>, indicating that the cofactor would be expected to bind with similar affinity. Residues Gly51–Pro53 of HetI correspond to Gly74–Pro76 in Sfp in which the proline ring nestles between the ribose sugar and the adenine base of the CoA and Arg52 stacks across the latter in a manner similar to that of Lys75 in Sfp. Ser67, His68, and Lys136 (HetI numbering) are all conserved and play important roles in binding the CoA pyrophosphate and 3'-phosphate (P<sub>3</sub>). Particularly important is His68 whose N<sub>62</sub> forms a salt bridge with P<sub>3</sub> and whose N<sub>61</sub> and amide nitrogen bind the  $\alpha$ -phosphate. His68 is part of the largest single conserved stretch of residues (SHSG) whose serines appear to play an unusual structural role (see Discussion). CoA binding is further stabilized by the conserved acidic side chains of Asp86, Glu88, and Glu132 which directly coordinate the metal ion (supposedly Mg<sup>2+</sup>) which bridges the protein to the CoA pyrophosphate group.

The only interactions important for CoA binding in Sfp which are not conserved in HetI are two lysine residues involved in salt bridges to the 3'-phosphate. The model suggests that this role may be assumed by basic residues such as His5 and Arg12 of HetI which lie approximately 5.5 Å from the 3'-phosphate in the repre-

sentative model but which could potentially interact directly via hydrogen bonds. Fig. 4 shows the predicted cofactor binding site.

## Discussion

With the complete genome sequences of several strains of *X. fastidiosa* as well as related bacteria already published [1,2], future studies have now to focus on gene expression/regulation and in the identification of metabolites (metabolomics) and the enzymes necessary for their synthesis (proteomics). Here, we have concentrated specifically on fatty acid biosynthesis.

The <sup>13</sup>C NMR spectrum (Fig. 1A) shows the presence of the expected resonances for mono-unsaturated fatty acids in a dry sample of *X. fastidiosa*. Furthermore, the signals which remain after growing the bacteria in the presence of [1-<sup>13</sup>C]acetate as a sole source of carbon (Fig. 1B) are consistent with classical fatty acid synthesis from acetate. For example, as the fatty acid is elongated by adding [1-<sup>13</sup>C]acetate, it will not label the terminal methyl group (explaining the absence of a signal at  $\delta$ 14) or the methylene carbons at  $\delta$ 32 and  $\delta$ 34, assigned to <sup>13</sup>CH<sub>2</sub>–CH<sub>2</sub>–CH<sub>3</sub>, and <sup>13</sup>C<sub>2</sub>, respectively.

However, the absence in the spectra of Fig. 1A of the expected resonances associated with glycerol raises the question of the chemical nature of the fatty acids present. The lack of glycerol signals could be explained

either by its lower mobility and stronger dipolar interaction or alternatively because the fatty acids are either free or present as part of structures which do not contain glycerol. The free fatty acid hypothesis was eliminated by extraction with chloroform [12]. Furthermore, the fatty acids do not seem to be part of lipopolysaccharides, normally present in the cell wall of gram-negative bacteria like *X. fastidiosa*, since the fatty acids of lipid A are normally  $\beta$  hydroxyl-myristic acids [33], which have very different chemical shift to those observed here.

On the whole, these data indicate that *X. fastidiosa*, as expected, is able to synthesize fatty acid, presumably present mainly in the form of membrane phospholipids, a vital component of almost all living organisms. The large amount of mono-unsaturated fatty acids in the membrane may be related to the lower growth temperature (28 °C) preferred by the bacterium [34,35].

Normal fatty acid synthesis involves the initial activation of the acyl carrier protein (ACP) of the fatty acid synthase by the holo-acyl-carrier-protein synthase (ACPs) which transfers the phosphopantetheinyl arm derived from Coenzyme A onto a specific serine residue of the ACP. However, despite the fact that the *X. fastidiosa* genome encodes an acyl carrier protein, and is capable of fatty acid synthesis, it was originally reported to lack ACPs.

Classical ACPs molecules are symmetric homotrimers made of small (~100 residue) subunits [36,37] in which the active sites lie at the intersubunit interfaces, where CoA binds. However, a second family of phosphopantetheinyl transferase is described to be involved in processes such as the activation of peptide carrier proteins (PCP) of peptide synthetases [24], lysine biosynthesis in fungi [38], and cyanobacterial heterocyst differentiation [39]. Members of this second family are approximately twice the size of a normal ACPs but are monomeric enzymes, composed of two domains (structurally equivalent to two of the three subunits of ACPs) with an active site at their interface.

*Escherichia coli* possesses at least three enzymes capable of phosphopantetheinyl transfer, including one member of each of these two families. The YhhU gene product belongs to the second family and is less active than ACPs in activating ACP but can maintain an *E. coli* ACPs conditional lethal phenotype viable when expressed in higher concentration.

We propose here that HetI, the *X. fastidiosa* homolog of YhhU, may therefore represent one of the supposed missing links for fatty acid biosynthesis. However, the identification of a homolog is not sufficient to guarantee its functionality and this is particularly critical in the case of low sequence identity, as observed here. First the gene product must be structurally viable and second it must be catalytically active. The homology model of HetI helps to address these questions, at least in part.

All validation methods applied to the selected representative model showed it to be acceptable. Furthermore, despite presenting only 20% sequence identity compared with the parent structure, Sfp, the pattern of sequence conservation is coherent. Besides those residues directly involved in CoA binding, the majority of conserved residues are either hydrophobic and participate in the core of the structure or appear to play specific structural roles, such as glycines with positive  $\phi$  angles or charged residues involved in salt bridges. Fig. 3 shows all such residues explicitly. Furthermore, the model suggests that His5 and Arg12 may replace the only non-conserved CoA binding residues of Sfp, Lys28, and Lys31, in interacting with P<sub>3</sub> (Fig. 4).

A further conserved structural feature is a series of serine sidechains which emanate from the inward pointing face of  $\beta$ -strand  $\beta$ 3 so as to form hydrogen bonds with the exposed amides of the adjacent edge strand  $\beta$ 5, which is part of the second  $\beta$ -sheet of the central sandwich. A BLASTp search of GenBank with HetI showed this SXSXS motif to be generally conserved and characteristic of most such phosphopantetheinyl transferase sequences, including ACPs. In some sequences, such as Sfp, one serine is replaced by an alternative amino acid with similar hydrogen-bonding potential (Asn, Asp, and Thr). However, a search of the PDB using the PATTINPROT software [40] shows this motif, which we term a serine ladder, to be unique to such molecules and is probably correlated to the wedge-like arrangement of the  $\beta$ -sandwich which favors the correct orientation of the serine side chains with respect to the leftover amides of the adjacent edge strand. No similar mechanism has thus far been described for satisfying or hiding edge-strand hydrogen-bonding potential [41] (Fig. 5).

Nevertheless, one major difference was noted between HetI and Sfp; the fact that the former is approximately 40 residues shorter due to deletions at both the N- and C-termini. The structural consequences of these deletions (Figs. 2 and 6) are the loss of a  $\beta$ -strand, two  $\alpha$ -helices, and connecting loops at each end of the molecule. These regions are spatially proximal and in Sfp cover one side of the large buried helix  $\alpha$ 3. Their deletion leads to the solvent exposure of the equivalent helix in HetI ( $\alpha$ 2) which has suffered a series of compensatory mutations. Leu45, Leu46, and Leu50 in Sfp have been replaced by Glu22, Arg23, and Gln27 on the exposed side of  $\alpha$ 2 in HetI, consistent with the N- and C-terminal deletions.

Overall, these observations appear to suggest that the HetI gene product should be both structurally viable and catalytically active. However, it would appear to present a somewhat simplified fold compared to Sfp. At this stage it is difficult to know what might be the metabolic consequences of this, if any. However, the observation that the YhhU protein in *E. coli* is less

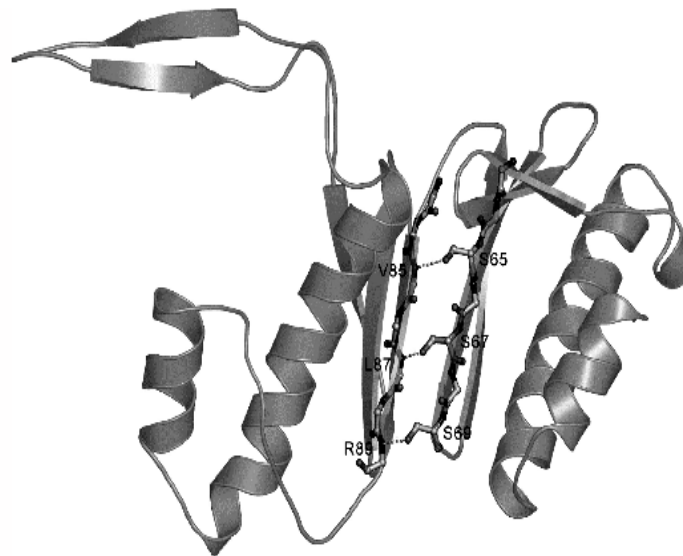


Fig. 5. The serine ladder. The side chains of serines 65, 67, and 69 in  $\beta_3$  in one  $\beta$ -sheet form hydrogen bonds with the free main chain amides from edge strand  $\beta_5$  of the other sheet of the wedge-like sandwich.

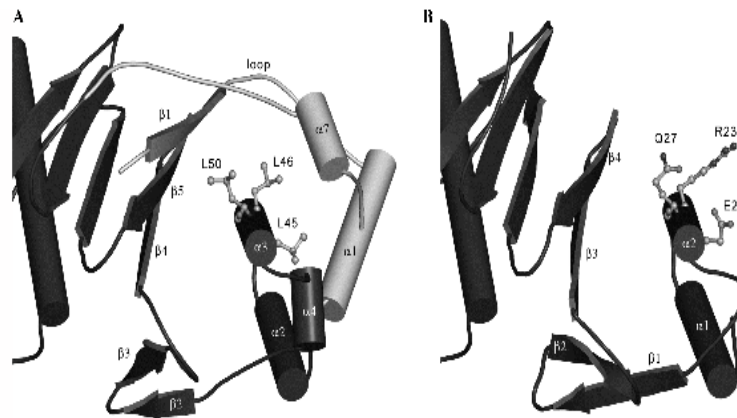


Fig. 6. Ribbons figure of part of the Sfp structure centered on the buried  $\alpha_3$  helix (A) compared with the exposed homologous helix,  $\alpha_2$  in HetI (B). The elements of structure shown in light grey in Sfp are absent from HetI due to large N- and C-terminal deletions. Selected residues which are different in the two molecules are highlighted in ball-and-stick.

efficient in modifying ACP compared with ACPs suggests that a less than optimal production of holo-ACP in *X. fastidiosa*, and consequently a reduced level of fatty acid, phospholipid, and lipopolysaccharide synthesis, may be one of the causes for the slow growth rate of the bacterium in vivo and in vitro.

Clearly the HetI gene product now needs to be expressed and biochemically characterized in order to elucidate its full role in *X. fastidiosa* metabolism. It is even possible that it serves as a generic phosphopantetheinyl transferase for more than one substrate given

that it is the only such molecule thus far identified in *X. fastidiosa*. Recently, interest in novel phosphopantetheinyl transferases capable of activating the PCP domains of non-ribosomal peptide synthetases is growing due to the search for broad specificity enzymes which have potential applications in the in vivo synthesis of novel biotechnologically interesting oligopeptide metabolites. The apparently simplified version of a phosphopantetheinyl transferase, represented by HetI, means, at the very least, that it merits further investigation.



## Acknowledgments

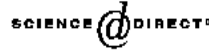
We thank FAPESP, CAPES, and CNPq for financial support.

## References

- [1] A.J.G. Simpson, F.C. Reinach, P. Arruda, F.A. Abreu, M. Acencio, R. Alvarenga, L.M.C. Alves, J.E. Araya, G.S. Baia, C.S. Baptista, M.H. Barros, E.D. Bonaccorsi, S. Bordin, J.M. Bove, M.R.S. Briones, M.R.P. Bueno, A.A. Camargo, L.E.A. Camargo, D.M. Carraro, H. Carrer, N.B. Colauto, C. Colombo, F.F. Costa, M.C.R. Costa, C.M. Costa-Neto, L.L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El Dorry, A.P. Facincani, A.J.S. Ferreira, V.C.A. Ferreira, J.A. Ferro, J.S. Fraga, S.C. Franca, M.C. Franco, M. Frohne, L.R. Furlan, M. Garnier, G.H. Goldman, M.H.S. Goldman, S.L. Gomes, A. Gruber, P.L. Ho, J.D. Hoheisel, M.L. Junqueira, E.L. Kemper, J.P. Kitajima, J.E. Krieger, E.E. Kuramae, F. Laigret, M.R. Lambais, L.C.C. Leite, E.G.M. Lemos, M.V.F. Lemos, S.A. Lopes, C.R. Lopes, J.A. Machado, M.A. Machado, A.M.B.N. Madeira, H.M.F. Madeira, C.L. Marino, M.V. Marques, E.A.L. Martins, E.M.F. Martins, A.Y. Matsukuma, C.F.M. Menck, E.C. Miracca, C.Y. Miyaki, C.B. Monteiro-Vitorello, D.H. Moon, M.A. Nagai, A.L.T.O. Nascimento, L.E.S. Netto, J. Nhani, F.G. Nobrega, L.R. Nunes, M.A. Oliveira, M.C. de Oliveira, R.C. de Oliveira, D.A. Palmieri, A. Paris, B.R. Peixoto, G.A.G. Pereira, J. Pereira, J.B. Pesquero, R.B. Quaggio, P.G. Roberto, V. Rodrigues, A.J.D. Rosa, V.E. de Rosa, R.G. de Sa, R.V. Santelli, H.E. Sawasaki, A.C.R. da Silva, F.R. da Silva, A.M. da Silva, J. Silva, J.F. da Silveira, M.L.Z. Silvestri, W.J. Siqueira, A.A. de Souza, A.P. de Souza, M.F. Terenzi, D. Truffi, S.M. Tsai, M.H. Tshako, H. Vallada, M.A. Van Sluys, S. Verjovski-Almeida, A.L. Vettore, M.A. Zago, M. Zatz, J. Meidanis, J.C. Setubal, The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (2000) 151–157.
- [2] M.A. Van Sluys, M.C. de Oliveira, C.B. Monteiro-Vitorello, C.Y. Miyaki, L.R. Furlan, L.E. Camargo, A.C. da Silva, D.H. Moon, M.A. Takita, E.G. Lemos, M.A. Machado, M.I. Ferro, F.R. da Silva, M.H. Goldman, G.H. Goldman, M.V. Lemos, H. El-Dorri, S.M. Tsai, H. Carrer, D.M. Carraro, R.C. de Oliveira, L.R. Nunes, W.J. Siqueira, L.L. Coutinho, E.T. Kimura, E.S. Ferro, R. Harakava, E.E. Kuramae, C.L. Marino, E. Giglioti, I.L. Abreu, L.M. Alves, A.M. do Amaral, G.S. Baia, S.R. Blanco, M.S. Brito, F.S. Cannavan, A.V. Celestino, A.F. da Cunha, R.C. Fenille, J.A. Ferro, E.F. Formighieri, L.T. Kishi, S.G. Leoni, A.R. Oliveira, V.E. Rosa Jr., F.T. Sasaki, J.A. Sena, A.A. de Souza, D. Truffi, F. Tsukumo, G.M. Yanai, L.G. Zarus, E.L. Civerolo, A.J. Simpson, N.F. Almeida Jr., J.C. Setubal, J.P. Kitajima, Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*, *J. Bacteriol.* 185 (2003) 1018–1026.
- [3] C.O. Rock, S. Jackowski, Forty years of bacterial fatty acid synthesis, *Biochem. Biophys. Res. Commun.* 292(2002) 1155–1166.
- [4] A. Bhattacharyya, S. Stilwagen, G. Reznik, H. Feil, W.S. Feil, I. Anderson, A. Bernal, M. D'Souza, N. Ivanova, V. Kapatral, N. Larsen, T. Los, A. Lykidis, E. Selkov Jr., T.L. Walunas, A. Purcell, R.A. Edwards, T. Hawkins, R. Haselkorn, R. Overbeek, N.C. Kyrpides, P.F. Predki, Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights, *Genome Res.* 12 (2002) 1556–1563.
- [5] E.G.M. Lemos, L.M.C. Alves, J.C. Campanharo, Genomics based design of defined growth media for the plant pathogen *Xylella fastidiosa*, *FEMS Microbiol. Lett.* 219 (2003) 39–45.
- [6] M.B. Smolka, D. Martins, F.V. Winck, C.E. Santoro, R.R. Castellari, F. Ferrari, I.J. Brum, E. Galembeck, H.D. Coletta-Filho, M.A. Machado, S. Marangoni, J.C. Novello, Proteome analysis of the plant pathogen *Xylella fastidiosa* reveals major cellular and extracellular proteins and a peculiar codon bias distribution, *Proteomics* 3 (2003) 224–237.
- [7] A.P. Facincani, J.A. Ferro, J.M. Pizauro Jr., H.A. Perreira Jr., E.G.M. Lemos, A.L. do Prado, M.I.T. Ferro, Carbohydrate metabolism of *Xylella fastidiosa*: detection of glycolytic and pentose phosphate pathway enzymes and cloning and expression of the enolase gene, *Genet. Mol. Biol.* 26 (2003) 203–211.
- [8] C. de Peri, L.M. Beltrami, H.S. Selistre-de-Araújo, A.L. Vettore, F.R. da Silva, P. Arruda, G. Oliva, D.H.F. de Souza, Overexpression, purification, and biochemical characterization of GumC, an enzyme involved in the biosynthesis of exopolysaccharide by *Xylella fastidiosa*, *Protein Expr. Purif.* 34 (2004) 223–228.
- [9] D. Osiro, L.A. Colnago, A.M.M.B. Otoboni, E.G.M. Lemos, A.A. Souza, H.D. Coletta Filho, M.A. Machado, A kinetic model for *Xylella fastidiosa* adhesion, biofilm formation, and virulence, *FEMS Microbiol. Lett.* 236 (2004) 313–318.
- [10] L.A. Colnago, P.R. Seidl, Application of carbon-13 nuclear magnetic resonance to the germination of soybean seeds in vivo, *J. Agric. Food Chem.* 31 (1983) 459–461.
- [11] K. Shon, Y. Kim, L.A. Colnago, S.J. Opella, NMR studies of the structure and dynamics of membrane bound bacteriophage Pfl coat protein, *Science* 252 (1991) 1303–1305.
- [12] L.A. Forato, L.A. Colnago, R.C. Garratt, M.A. Lopes, Identification of free fatty acids in maize protein bodies and purified alpha zeins by  $^{13}\text{C}$  and  $^1\text{H}$  nuclear magnetic resonance, *Biochim. Biophys. Acta* 1543 (2000) 106–114.
- [13] L.A. Forato, V. Yushmanov, L.A. Colnago, Interaction of two prolamins with [ $^{13}\text{C}$ ]oleic acid by  $^{13}\text{C}$  NMR, *Biochemistry* 43 (2004) 7121–7126.
- [14] D.G. Gadian, NMR and its Applications to Living Systems, second ed., Oxford University Press, New York, 1995.
- [15] M.J. Davis, W.J. French, N.W. Schaad, Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald, *Curr. Microbiol.* 6 (1981) 309–314.
- [16] L.A. Colnago, K.G. Valentine, S.J. Opella, Dynamics of the fd coat protein in the bacteriophage, *Biochemistry* 26 (1987) 847–853.
- [17] R.S. Flugel, Y. Hwangbo, R.H. Lambalot, J.E. Cronan, C.T. Walsh, Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in *Escherichia coli*, *J. Biol. Chem.* 275 (2000) 959–968.
- [18] Available from: <<http://aeg.lbi.ic.unicamp.br/xf/>>.
- [19] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410.
- [20] Available from: <<http://www.ncbi.nlm.nih.gov/>>.
- [21] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292 (1999) 195–202.
- [22] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402.
- [23] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [24] K. Reuter, M.R. Mofid, M.A. Marahiel, R. Ficner, Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily, *EMBO J.* 18 (1999) 6823–6831.
- [25] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.



- [26] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comp. Chem.* 4 (1983) 187–217.
- [27] R.A. Laskowski, M.W. MacArthur, J.M. Thornton, Validation of protein models derived from experiment, *Curr. Opin. Struct. Biol.* 8 (1998) 631–639.
- [28] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature* 356 (1992) 83–85.
- [29] G. Vriend, C. Sander, Quality control of protein models: directional atomic contact analysis, *J. Appl. Crystallogr.* 6 (1993) 47–60.
- [30] G. Vriend, What if: a molecular modeling and drug design program, *J. Mol. Graph.* 8 (1990) 52–56.
- [31] A.H. Purcell, D.L. Hopkins, Fastidious xylem-limited bacterial plant pathogens, *Annu. Rev. Phytopathol.* 34 (1996) 131–151.
- [32] W.L. DeLano, The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. Available from: <<http://www.pymol.org>>.
- [33] A. Bhattacharyya, S. Stilwagen, N. Ivanova, M. D'Souza, A. Bernal, A. Lykidis, V. Kapatral, I. Anderson, N. Larsen, T. Los, G. Reznik, E. Selkov Jr., T.L. Walunas, H. Feil, W.S. Feil, A. Purcell, J.-L. Lassez, T.L. Hawkins, R. Haselkorn, R. Overbeck, P.F. Predki, N.C. Kyrpides, Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains, *Proc. Natl. Acad. Sci. USA* 99 (2002) 12403–12408.
- [34] D.A. Los, N. Murata, Structure and expression of fatty acid desaturases, *Biochim. Biophys. Acta* 1394 (1998) 3–15.
- [35] N. Murata, D.A. Los, Membrane fluidity and temperature perception, *Plant Physiol.* 115 (1997) 875–879.
- [36] K.D. Parris, L. Lin, A. Tam, R. Mathew, J. Hixon, M. Stahl, C.C. Fritz, J. Seehra, W.S. Somers, Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites, *Structure* 8 (2000) 883–895.
- [37] N.Y. Chirgadze, S.L. Briggs, K.A. McAllister, A.S. Fischl, G. Zhao, Crystal structure of *Streptococcus pneumoniae* acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis, *EMBO J.* 19 (2000) 5281–5287.
- [38] D.E. Ehmann, A.M. Gehring, C.T. Walsh, Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of  $\alpha$ -amino adipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5, *Biochemistry* 38 (1999) 6171–6177.
- [39] T.A. Black, C.P. Wolk, Analysis of a Het<sup>-</sup> mutation in *Anabaena* sp. strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing, *J. Bacteriol.* 176 (1994) 2282–2292.
- [40] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, Network protein sequence analysis, *TIBS* 25 (2000) 147–150.
- [41] J.S. Richardson, D.C. Richardson, Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation, *Proc. Natl. Acad. Sci. USA* 99 (2002) 2754–2759.

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Molecular &amp; Biochemical Parasitology 140 (2005) 23–41

---



---

MOLECULAR  
& BIOCHEMICAL  
PARASITOLOGY

---



---

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

Rafael Dhalia<sup>a,c</sup>, Christian R.S. Reis<sup>b,c</sup>, Eden R. Freire<sup>b,c</sup>, Pollyanna O. Rocha<sup>b,c</sup>,  
Rodolfo Katz<sup>b,c</sup>, João R.C. Muniz<sup>c</sup>, Nancy Standart<sup>d</sup>, Osvaldo P. de Melo Neto<sup>c,\*</sup>

<sup>a</sup> Departamento de Biologia Celular, Universidade de Brasília, Brasília 70910-900, D.F., Brazil<sup>b</sup> Departamento de Genética, Universidade Federal de Pernambuco, Avenida Professor Moraes Rego s/n, Cidade Universitária, Recife 50732-970, PE, Brazil<sup>c</sup> Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos 13560-970, SP, Brazil<sup>d</sup> Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB21 9A, UK<sup>e</sup> 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 (*LmEIF4E1-4*), two eIF4As (*LmEIF4A1-2*) and five eIF4Gs (*LmEIF4G1-5*). We report the initial characterisation of *LmEIF4E1-3*, *LmEIF4A1-2* and *LmEIF4G3*. First, the expression of these proteins in *L. major* promastigotes was quantitated by Western blotting using isoform specific antibodies. *LmEIF4A1* and *LmEIF4E3* are very abundant, *LmEIF4G3* is moderately abundant and *LmEIF4E1/LmEIF4E2/LmEIF4A2* are rare or not detected. In cap-binding assays, only *LmEIF4E1* bound to the 7-methyl-GTP-Sepharose resin. Molecular modelling confirmed that *LmEIF4E1* has all the structural features of a cap-binding protein. Finally, pull-down assays were used to investigate the potential interaction between the eIF4A (*LmEIF4A1/LmEIF4A2*) and eIF4G (*LmEIF4G1-3*) homologues. Only *LmEIF4G3*, via the HEAT domain, bound specifically both to *LmEIF4A1* 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.

© 2004 Elsevier B.V. All rights reserved.

**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, Sleep-

ing 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](mailto:opmn@cpqam.fiocruz.br) (O.P. de Melo Neto).

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. tarantolae* [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 cerevisiae* 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 *LmEIF4E1* and the *LmEIF4G3* 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 *KpnI/XbaI* (*LmEIF4E2*) or *BamHI/XbaI* (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 *AflIII/NorI* (*LmEIF4E1*), *NcoI/NorI* (*LmEIF4E2* and *LmEIF4E3*)—in the latter a N to D mutation was introduced in the second codon to accommodate the *NcoI* site) or *NcoI/XhoI* (*LmEIF4A1*) and cloned into the *NcoI/NorI* or *NcoI/XhoI* 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 *LmEIF4E2b* were recovered from the pGEM derived plasmids and subcloned into the *BamHI/SalI* sites of the vector pGEX4T3 (Amersham Biosciences). For the same reason *LmEIF4E3* and *LmEIF4A1* were reamplified flanked by sites for *BamHI/NorI* and *BamHI/XhoI*, 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 *BamHI/XhoI* and cloned directly into the same sites of the vectors pRSETA and pGEX4T3.

Fragments coding for the HEAT repeats domain from the *LmEIF4G1*-3 homologues, *LmEIF4G1*<sub>128–475</sub> (the numbers indicate the amino acids remaining from the wild type protein), *LmEIF4G2*<sub>387–705</sub> (missing the last 22 amino acids from the HEAT domain in Fig. 7) and *LmEIF4G3*<sub>26–310</sub>, flanked by sites for *AflIII/NorI* (*LmEIF4G1*<sub>128–475</sub>) or *NcoI/NorI* (the others), were amplified and cloned into the *NcoI/NorI* sites of the pET21D plasmid for the expression of C-terminal His-tagged proteins. For both *LmEIF4G2*<sub>387–705</sub> and *LmEIF4G3*<sub>26–310</sub>, an AUG start codon was introduced in

the constructs with the *NcoI* site. The full length *LmEIF4G3* (*LmEIF4G3*<sub>1–636</sub>), as well as its first half (*LmEIF4G3*<sub>1–343</sub>) were also amplified and cloned into the same sites of pET21D. The latter two fragments were then reamplified and cloned into the *BamHI/NorI* 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

<sup>35</sup>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 *HinDIII/BglII* and subcloned into the *HinDIII/BamHI* 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<sup>7</sup>G(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 2M 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 *LmEIF4A1*-2, *LmEIF4E1*-2 and *LmEIF4G3*<sub>26–310</sub> by immunising adult New Zealand White rabbits with the His-tagged recombi-



nant forms. For *LmEIF4E3*, 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<sub>2</sub>, 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 <sup>35</sup>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 *NotI* for pET-*LmEIF4G1*<sub>128–475</sub> and the *LmEIF4G3* variants and with *XhoI* for pET-*LmEIF4G2*<sub>387–705</sub> 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 <sup>35</sup>S-methionine. To obtain the labelled human eIF4A, its cDNA in plasmid pET(His<sub>6</sub>-eIF4A) [48] was first recovered by digestion with *NdeI/NotI* and subcloned into pET21A. The resulting plasmid was linearized with *NotI* 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 *LmEIF4E1* (GeneDB ID—LmjF27.1620), *LmEIF4E2* (LmjF19.1500/LmjF19.1480), *LmEIF4E3* (LmjF28.2500) and *LmEIF4E4* (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 *LmEIF4E1-4*, 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

<i>L. major</i> sequence	GeneDB ID	Predicted molecular weight (kDa)	Chromosome localization	Identity (similarity) to Human homologue <sup>a</sup>	Identity (similarity) to the <i>T. brucei</i> orthologue
<i>LmEIF4E1</i> <sup>b</sup>	LmjF27.1620	24	27	22% (42)	48% (61)
<i>LmEIF4E2</i> <sup>b</sup>	LmjF19.1500/LmjF19.1480	31.5	19 (2 genes)	27% (41)	45% (57)
<i>LmEIF4E3</i>	LmjF28.2500	38	28	27% (43)	49% (61)
<i>LmEIF4E4</i>	LmjF30.0450 <sup>c</sup>	33.8	30	28% (45)	37% (50)
<i>LmEIF4A1</i> <sup>b</sup>	LmjF01.0780/LmjF01.0770	45.3	1 (2 genes)	56% (74)	85% (91)
<i>LmEIF4A2</i>	LmjF28.1530	43.9	28	50% (71)	79% (89)
<i>LmDHH1</i>	LmjF35.0370	46.4	35	30% (50)	86% (92)
<i>LmEIF4G1</i>	LmjF15.0060	114	15	25% (43) <sup>d</sup>	31% (47)
<i>LmEIF4G2</i>	LmjF15.1320	145.9	15	21% (37) <sup>d</sup>	33% (49)
<i>LmEIF4G3</i>	LmjF16.1600	71.2	16	26% (39) <sup>d</sup>	38% (55)
<i>LmEIF4G4</i>	LmjF36.6060	84.6	36	22% (38) <sup>d</sup>	26% (45)
<i>LmEIF4G5</i>	LmjF10.1080	88.8	10	21% (37) <sup>d</sup>	43% (60)

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

<sup>a</sup> Human eIF4G1. GenBank accession: Q04637.

<sup>b</sup> GenBank accessions available: *LmEIF4E1*—CAB77676; *LmEIF4E2*—CAB94109 and CAB94111; *LmEIF4A1*—NP\_047099.

<sup>c</sup> 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.

<sup>d</sup> These sequences show similarity to the human sequence only at the level of the central HEAT domain.

*LmEIF4E1*, 3 and 4, respectively. In contrast, *LmEIF4E2* 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]. *LmEIF4E2* contains all eight of these tryptophan residues, *LmEIF4E1* contains seven, *LmEIF4E4* five and *LmEIF4E3* 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 *LmEIF4E1* and *LmEIF4E2*, while in *LmEIF4E3-4* 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, *LmEIF4E1* and 2 contain K112, R157 and R/K162, respectively, while both *LmEIF4E3-4* contain only R157, and lack basic residues at 112 and 162. On this basis, *LmEIF4E1* and 2 possess all the features expected of a 7-methyl-GDP-binding protein, while *LmEIF4E3-4* have some, but not all, the predicted features.

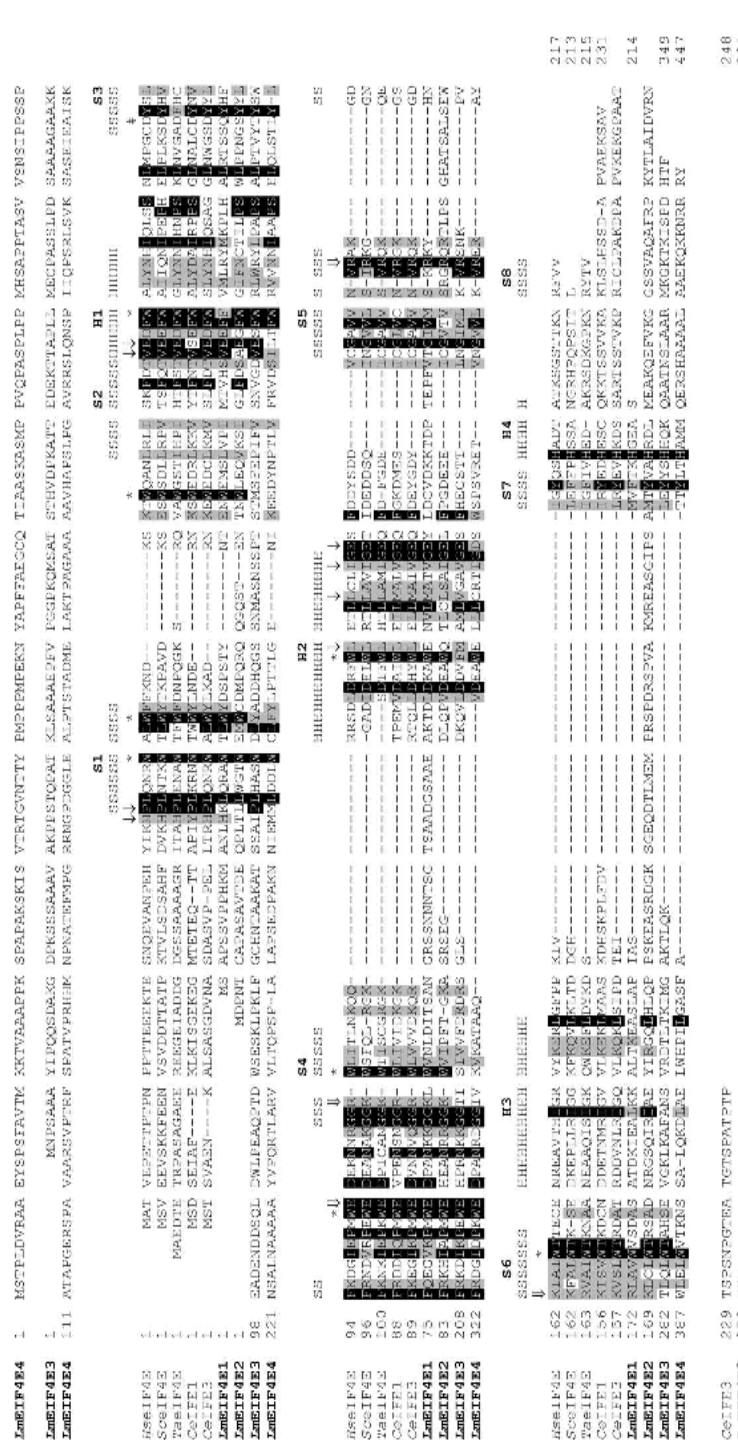
eIF4G binds the convex dorsal side of eIF4E primarily due to the interaction between W73 of eIF4E and three amino acid side chains of the eIF4E binding peptide, YXXXXLphi (where X can be any amino acid and phi is usually a hydrophobic amino acid such as L, M or F [23]), also present in eIF4E-binding 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]. *LmEIF4E2*, 3 and 4 contain W73, while *LmEIF4E1* 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 *LmEIF4E3-4* N-terminal extensions of about 80 and 190 amino acids, respectively, a shorter C-terminus for *LmEIF4E1* and the significant insertions of variable number of amino acids at specific positions in the first three *Leishmania* proteins. *LmEIF4E1* is distinguished by the presence of a 21 amino acids long insertion not observed in other eIF4E sequences. *LmEIF4E2* has two short insertions, one of them also present in *LmEIF4E3*, 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

*LmEIF4E1-3* (*LmEIF4E4* 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 *LmEIF4E1-3* in extracts of exponentially growing *L. major* promastigotes (Fig. 2). All three proteins were detected in these extracts although their levels varied significantly, with *LmEIF4E3* being very abundant ( $>5 \times 10^4$  molecules/cell) and both *LmEIF4E1* 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 *LmEIF4E3* 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 *LmEIF4E1* or *LmEIF4E2* 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* cIF4E homologues with the human, yeast, plant and nematode sequences. Chiral W alignment of different cIF4E homologues from selected organisms. Amino acids identical in more than 60% of the sequences are highlighted in dark grey, while amino acids defined as similar, based on the BLOSUM62 Matrix, on more than 60% of the sequences, are shown in pale grey. 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 cIF4G binding [53]. # indicates the amino acid D90, replaced by a Q in *LmEIF4E1* (Q71) and shown in Fig. 4. Relevant GenBank accession numbers: human (Hf)—P06730; yeast (*S. cerevisiae*, Sc)—P07260; plant (*Thiitewm acerrivum*, Ta) P29557; nematode (*C. elegans*, Ce) IF4E1 and IF4E3—NP\_499751 and NP\_503124, respectively.

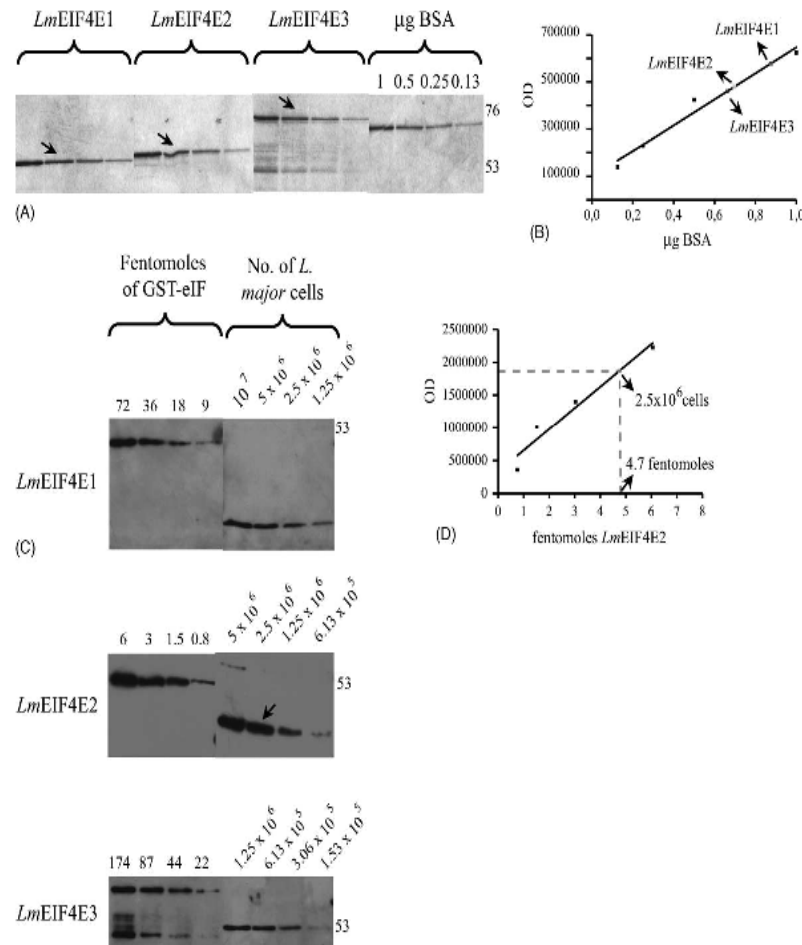


Fig. 2. Expression analysis and quantitation of *LmEIF4E1-3* in exponentially grown *L. major* promastigotes. (A) Serial dilutions of recombinant GST-tagged *LmEIF4E1-3* compared in Coomassie Blue stained gels with known concentrations of BSA. (B) Quantitation of the recombinant proteins (a representative 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 *LmEIF4E1-3*, 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 *LmEIF4E1-3* 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 analyzed by western blotting with the isoform specific sera. (D) Procedure for the quantitation of the endogenous levels of the various factors using *LmEIF4E2* as an example. The Western blot result shown in C for GST-*LmEIF4E2* was quantitated by densitometry and plotted as a function of fentomoles of recombinant protein. The optical density obtained for the endogenous factor at  $2.5 \times 10^6$  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 *LmEIF4E1* and *LmEIF4E3*. 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*

*LmEIF4E1-3* were then labelled with  $^{35}\text{S}$ -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 *LmEIF4E1* 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

<i>L. major</i> sequence	Femtomoles/10 <sup>6</sup> cell	No. of molecules/cell	No. of molecules/cell in yeast <sup>a</sup>
<i>LmEIF4E1</i>	5.2 ± 2.5	3.2 ± 1.5 × 10 <sup>3</sup>	(3.2–3.6) × 10 <sup>5</sup>
<i>LmEIF4E2</i>	1.8 ± 0.5	1 ± 0.3 × 10 <sup>3</sup>	
<i>LmEIF4E3</i>	117 ± 56	7.1 ± 3.4 × 10 <sup>4</sup>	
<i>LmEIF4A1</i>	594 ± 218	3.6 ± 1.3 × 10 <sup>5</sup>	(7–9) × 10 <sup>5</sup>
<i>LmEIF4A2</i>	ND	ND	
<i>LmEIF4G3</i>	10.6 ± 4.1	6.4 ± 2.5 × 10 <sup>3</sup>	(1.5–2) × 10 <sup>4</sup>

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

ing was observed (not shown). *LmEIF4E2* 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 (*LmEIF4E1*) 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 (*LmEIF4E1–3*), we attempted to model their structure based on the mammalian eIF4E structure [41]. However, adequate modelling of both *LmEIF4E2* and *LmEIF4E3* 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 *LmEIF4E1*, 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 *LmEIF4E1* 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 *LmEIF4E1* 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 *LmEIF4E1* model confirms that, despite the minor variations in sequence, not only *LmEIF4E1* 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 *LmEIF4E1* model, using the program GRASP [57], does not show any significant differences in overall charges

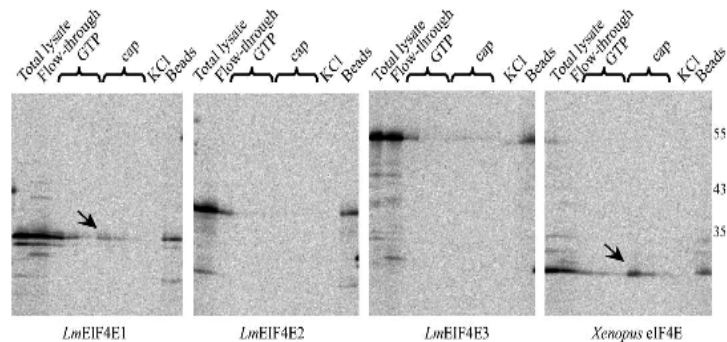


Fig. 3. *LmEIF4E1*, but not *LmEIF4E2* or 3, binds specifically to the eukaryotic cap analogue. The three *L. major* eIF4E homologues (*LmEIF4E1–3*), as well as the *X. laevis* eIF4E (*Xenopus* in the figure) used as positive control, were labelled with <sup>35</sup>S-methionine by in vitro translation 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 run 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 2M 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.

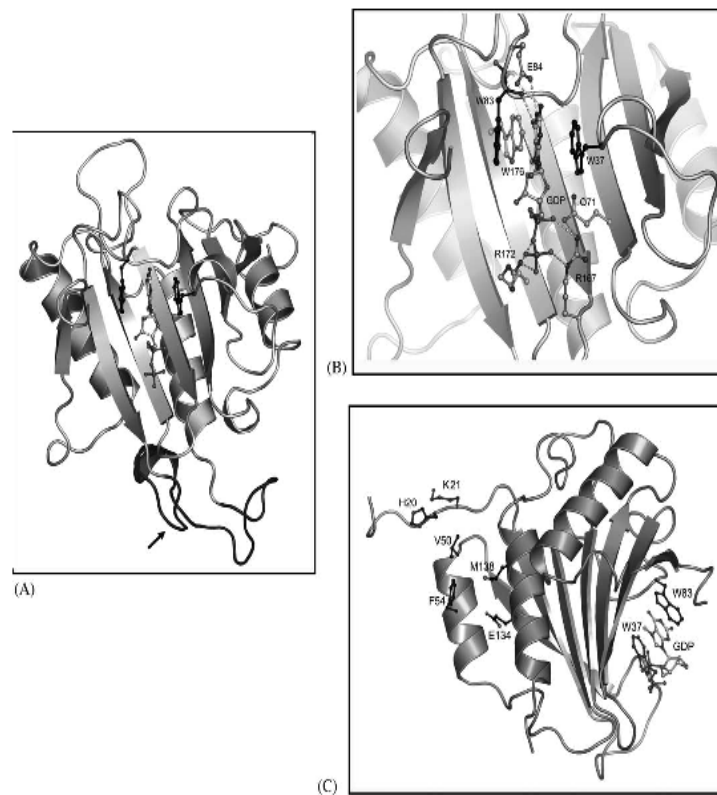


Fig. 4. Model of the predicted *LmEIF4E1* structure bound to 7-methylGDP. Ribbon diagrams of the overall predicted *LmEIF4E1* structure and detail of the cap binding slot created using the program PyMol (<http://www.pymol.org>). (A) Model of the structure of *LmEIF4E1* bound to 7-methylGDP, based on 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 *LmEIF4E1* 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 E93 interaction (corresponding to R112 in mammals) is omitted. (C) Side view of the *LmEIF4E1* 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 *LmEIF4E1* 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 *LmEIF4A1*, and the human eIF4A1 homologue is 74% (Table 1). *LmEIF4A1* (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

*LmEIF4A1* as well as a possible second eIF4A homologue from *L. major* (*LmEIF4A2*—*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 (*EcW2*) which may also be involved in translation initiation [59], an Archae DEAD-box protein (*MjDEAD*) whose structure has been solved [60], as well as two unrelated eukaryotic RNA helicases, *LmDHH1* and *Spisula p47* (see below). eIF4A and other related DEAD-box RNA helicases are characterised by nine motifs (motifs I,

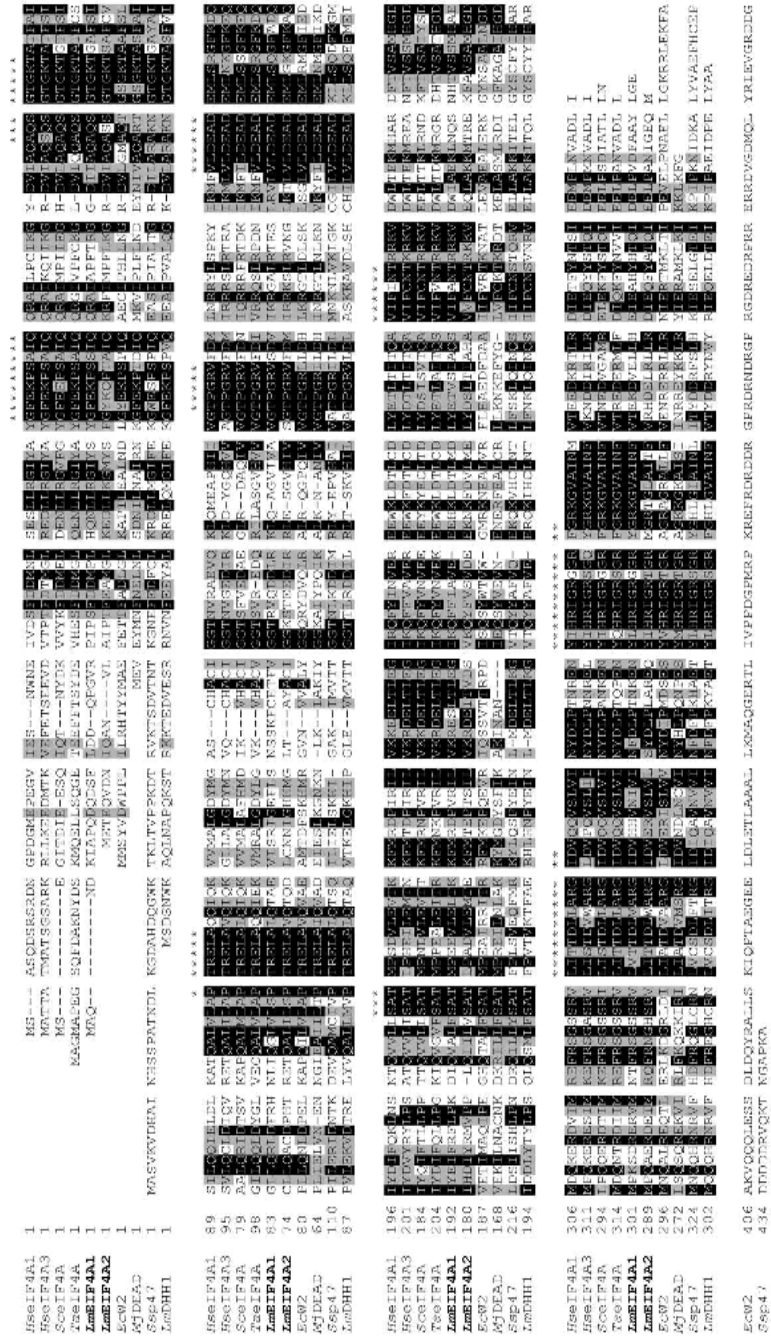


Fig. 5. Sequence comparison of the *L. major* eIF4A homologs with various related protein sequences. Clustal W alignment performed as described in Fig. 1 comparing the sequences of the *L. major* eIF4A homologs with selected sequences from several different organisms. \* indicates the nine motifs typical of DEAD box RNA helicases [20,21]. Relevant GenBank accession numbers: human (Hs) eIF4A1—P04765; human eIF4AIII (Hs6IF4A3)—P38919; yeast (*S. cerevisiae*, 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* MjDEAD 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 *LmEIF4A1* and 2 as possible eIF4A homologues, we searched the *L. major*

sequence databases for any related RNA helicases. The sequence with the nearest match (*LmDHH1* 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 *LmDHH1* and bacterial proteins, and also highlights the close

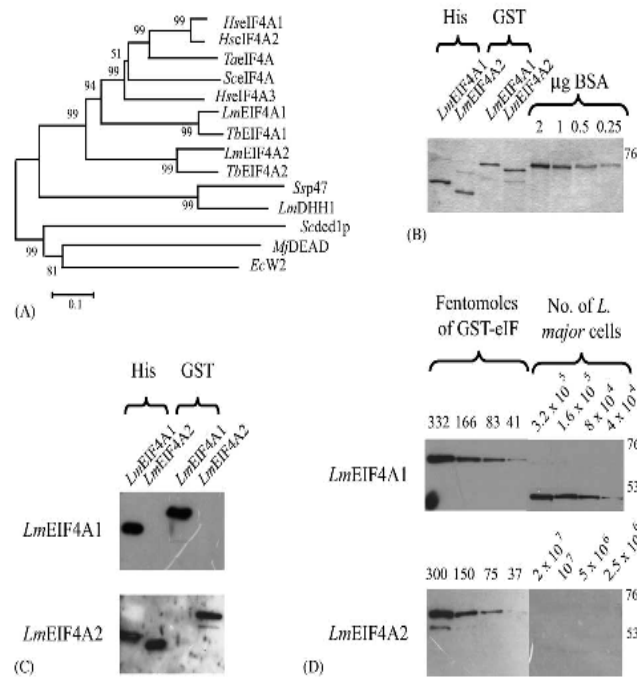


Fig. 6. *LmEIF4A1*, but not *LmEIF4A2*, is expressed as an abundant protein in *L. major* promastigotes. (A) Evolutionary relationship of the *L. major* eIF4A sequences. Neighbour-joining tree based on the alignment of various eIF4A sequences including *LmEIF4A1-2*. Bootstrap values are shown next to the respective branches (10,000 replicates). When compared with Fig. 5 we have included for this tree the sequences from: human eIF4A2 (GeneBank accession AAH12547); *S. cerevisiae* 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 *LmEIF4A1-2* as well as isoform specific purified antibodies. (B) Recombinant proteins used for antibody production (His) and quantitation assays (GST). Approximately 0.5 (*His-LmEIF4A2*) and both GST proteins) or 1  $\mu$ g (*His-LmEIF4A1*) 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 *LmEIF4A1-2* were tested with the two antisera. For the *LmEIF4A1* assay, 12.5 ng of each recombinant protein were loaded on the gel whilst for *LmEIF4A2* 6.25 ng of the His and 25 ng of the GST fusions were used. Note that the antibody produced against *His-LmEIF4A2* cross-reacts with *His-LmEIF4A1* 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-*LmEIF4A1* antibody. (D) Analysis of the expression of both *LmEIF4A1* and *LmEIF4A2* in total *L. major* extracts. Quantitation of *LmEIF4A1* in promastigotes was performed as described for the *LmEIF4E* homologues in Fig. 2.

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

### 3.6. *LmEIF4As* quantitation and expression analysis in *L. major*

Both the *LmEIF4A1* and the *LmEIF4A2* 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 *LmEIF4A1* 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 *LmEIF4Es* (Fig. 6C and D). *LmEIF4A1*, 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 *LmEIF4A1/LmEIF4E3*. 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 *LmEIF4A2*, even in  $2 \times 10^7$  promastigote cells (Fig. 6D). We estimate from the sensitivity of the *LmEIF4A2* antibody that levels above  $3 \times 10^3$  molecules/cell would have been detected. However in this case, its abundance would be at least 60–150-fold lower than that of *LmEIF4A1*. We conclude then that *LmEIF4A1* is the functional homologue of eIF4A in *L. major* promastigotes and that *LmEIF4A2* 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 (*LmEIF4G1*, GeneDB ID—LmjF15.0060;

*LmEIF4G2*—LmjF15.1320; *LmEIF4G3*—LmjF16.1600; *LmEIF4G4*—LmjF36.6060; *LmEIF4G5*—LmjF10.1080).

The five proteins vary significantly in size (ranging from 1016 to 1425, 635, 765 and 782 amino acids for *LmEIF4G1*–*5*, respectively), overall charge (positive in *LmEIF4G1* and negative in the other four) and in the relative location of the conserved eIF4G domain (central in *LmEIF4G1*, 2, 5 and N-terminal in *LmEIF4G3*–*4*). Both the *LmEIF4G1* and *LmEIF4G2* genes are located on chromosome 15 while the *LmEIF4G3*–*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 *LmEIF4G3* 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. cerevisiae* Tif4631p, and wheat eIFiso4G as well as the translation regulator PAIP1. Overall, the domain from *LmEIF4G1* is the most similar to human eIF4GI whilst the ones from *LmEIF4G2* 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 *LmEIF4G1*, 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

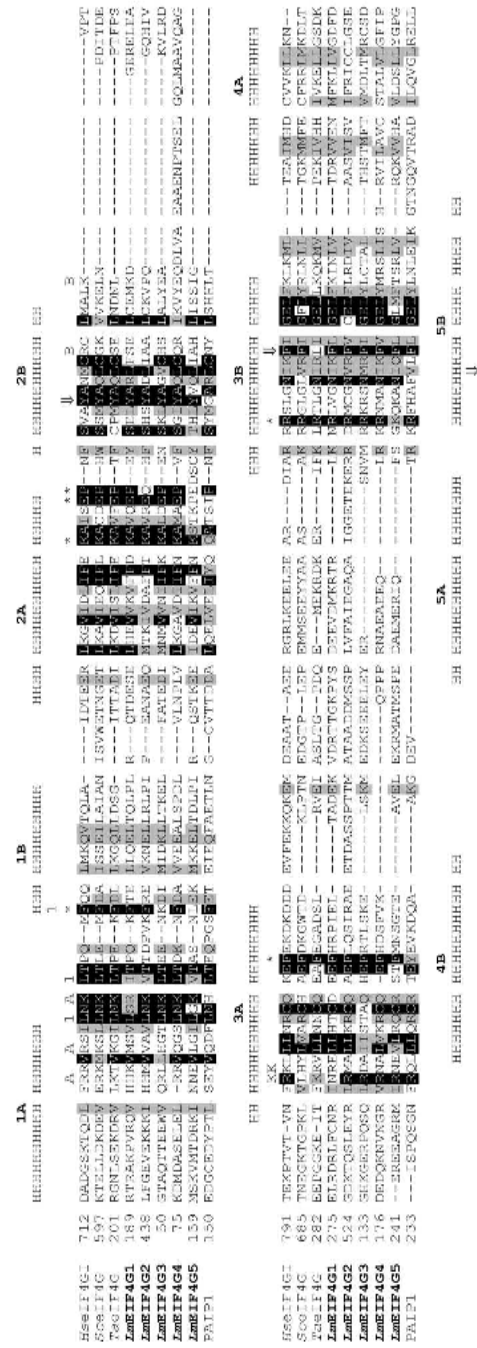


Fig. 7. Sequence comparison of the HEAT domain of the putative *L. major* eIF4G homologues with selected sequences from various organisms. Clustal W alignment performed as described in Fig. 1 comparing the sequences of the HEAT domain from the putative *L. major* eIF4G homologues (LmEIF4G1–5) with the equivalent domain from human, yeast and plant eIF4G homologues as well as the human PAIP1 protein. The predicted five antiparallel alpha 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 P938A [29]. The R801D/K803D mutant (6; in the figure) does not prevent eIF4A binding although it does abolish IRES binding [30]. 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 [30,36]. Relevant GenBank accession numbers: human eIF4G (*Hs*)—Q04637; yeast TIR463.1p (*S. cerevisiae*, *Sc*)—NP\_011678; plant eIF5a-4G (*T. aestivum*, *Ta*)—Q03387; human PAIP1—NP\_006442.

recombinant *LmEIF4A1-2* and human eIF4A (His<sub>6</sub>-eIF4A [48]) on Ni-NTA beads (Fig. 8A—left panel), and incubated the beads with <sup>35</sup>S-labelled human eIF4G [70]. Human eIF4A efficiently bound to human eIF4G, however neither *LmEIFA1* nor *LmEIFA2* was capable of binding to the human protein (Fig. 8A—right panel). Next, fragments cod-

ing for the HEAT domain from three of the *L. major* candidate eIF4G proteins (*LmEIF4G1*<sub>128–475</sub>, *LmEIF4G2*<sub>387–705</sub>, *LmEIF4G3*<sub>26–310</sub>), were synthesised in the presence of <sup>35</sup>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

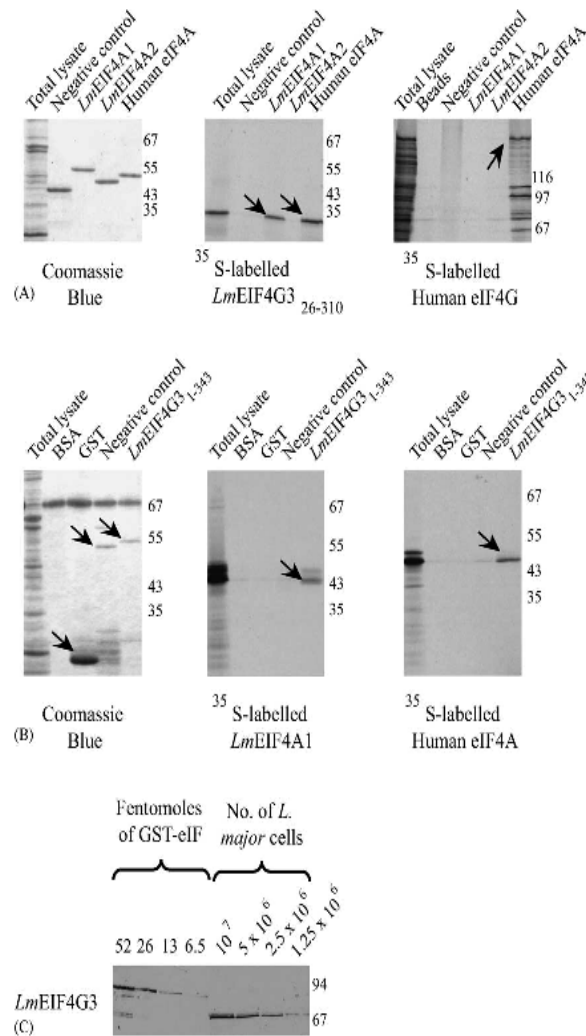


Fig. 8. Analysis of the interaction between human and *L. major* eIF4A homologues with the *LmEIF4G3* protein. (A) Pull-down assay using recombinant His-tagged *LmEIF4A1-2* and human eIF4A incubated with either <sup>35</sup>S-labelled human eIF4G or *LmEIF4G3*<sub>26–310</sub>. Proteins bound to the Ni-NTA beads were analysed through 15% (*LmEIF4G3*<sub>26–310</sub>) 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 *LmEIF4A1*-His/labelled *LmEIF4G3*<sub>26–310</sub> and human eIF4A-His/labelled *LmEIF4G3*<sub>26–310</sub>. 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 *LmEIF4G3*<sub>1–343</sub> and <sup>35</sup>S-labelled *LmEIF4A1*/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 *LmEIF4G3*<sub>1–343</sub>-GST/labelled *LmEIF4A1* and *LmEIF4G3*<sub>1–343</sub>-GST/labelled human eIF4A. (C) Quantitation of *LmEIF4G3* in *L. major* promastigotes. The quantitation of the endogenous protein levels was performed as described in Fig. 2 with recombinant GST-*LmEIF4G3*<sub>1–636</sub> and isoform specific polyclonal sera against *LmEIF4G3*.



complex formation on the EMCV IRES [28]. Only labelled *LmEIF4G3*<sub>26–310</sub> was able to bind efficiently to *LmEIF4A1* and to human eIF4A as shown in Fig. 8A (middle panel). No specific binding by the *LmEIF4G1–2* constructs to the various recombinant proteins was observed (data not shown). Moreover, even though *LmEIF4A1* and 2 share a similarity of 71%, no significant binding by any of the three *LmEIF4G* proteins to *LmEIF4A2* was observed (Fig. 8A and data not shown).

To confirm the specific interaction between *LmEIF4G3*<sub>26–310</sub> with *LmEIF4A1*, we expressed the N-terminal half of *LmEIF4G3* containing the HEAT domain (*LmEIF4G3*<sub>1–310</sub>) 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

<sup>35</sup>S-labelled *LmEIF4A1* 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-*LmEIF4G3*<sub>1–310</sub> protein. This reverse assay confirmed the interaction between the GST-*LmEIF4G3*<sub>1–310</sub> fusion protein and both labelled *LmEIF4A1* 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 *LmEIF4G3* interacts specifically with *LmEIF4A1* as well as with human eIF4A.

Thus, our results are consistent with roles for both *LmEIF4A1* and *LmEIF4G3* 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-*LmEIF4G3*<sub>26–310</sub>

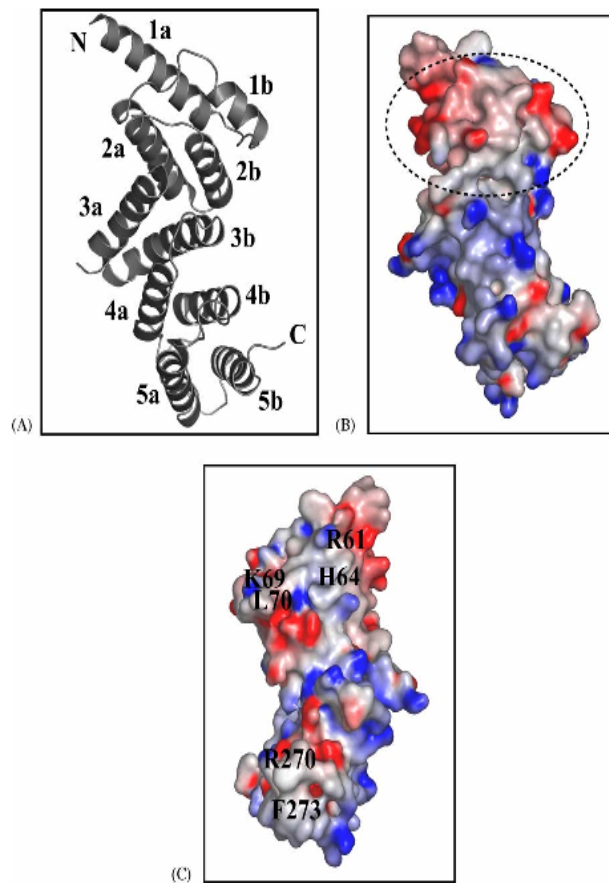


Fig. 9. Model of the structure of the *LmEIF4G3* HEAT domain. (A) Ribbon diagram of the structure of the predicted *LmEIF4G3* HEAT domain with the concave surface on the right and the convex surface on the left. The  $\alpha$  helices are labelled as described for human eIF4GII [26] and as shown in Fig. 7. (B)–(C) Grasp representation of the surface of the domain colored coded for electrostatic potential (red <math>-8kBT</math>; blue >math>8kBT</math>). 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 *LmEIF4G3* 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* LmEIF4G3 levels, as previously done for LmEIF4E1-3 and LmEIF4A1-2, using full length GST-LmEIF4G3<sub>1-636</sub> for comparison. This quantitation (Fig. 8C and Table 2) indicates that, at about  $(4-8) \times 10^3$  molecules per cell, LmEIF4G3 is present at about twice the level observed for LmEIF4E1 in *L. major* promastigotes, and is about 10- and 50-fold less abundant than LmEIF4E3 and LmEIF4A1, 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 LmEIF4G3 molecules in *L. major* is about one-third to one half of the levels in yeast, it is unclear whether the intracellular levels of LmEIF4G3 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 LmEIF4A 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 LmEIF4G3 in translation in *L. major*, we decided to attempt the molecular modelling of LmEIF4G3. 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 LmEIF4G3 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 LmEIF4G3 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 LmEIF4E1 model, the overall quality of the final model was confirmed by the programs PROCHECK, Verify 3D and WHATIF. The LmEIF4G3 HEAT domain folds into the same crescent-shaped molecule described for the human protein, with the five pairs of antiparallel  $\alpha$  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

EMCVIRES [26]. In human eIF4GII this region has an overall positive charge but in the LmEIF4G3 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 LmEIF4G3 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 LmEIF4E1 protein described here), to this cap4 structure investigated through fluorescence titration measurements [72,73]. LeishIF4E-1 bound to both the cap4 and m<sup>7</sup>GTP 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 LmEIF4E2 or 3, which do not bind cap-Sepharose, as well as LmEIF4E4, 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 LmEIF4E1 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 TbEIF4E3 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 *LmEIF4E1* 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 *LmEIF4E1* may be enhanced at the amastigote stage specially considering that, assuming a similar behaviour to the *T. brucei* orthologue, *LmEIF4E3* (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 *LmEIF4E2* 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 *LmEIF4A1* in translation in promastigotes. Since equivalent levels of expression have been reported for *LmEIF4A1* in both promastigote and amastigote stages of the *Leishmania* life cycle [37], it is unlikely that *LmEIF4A2* 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 *LmEIF4G3* (and possibly the related *LmEIF4G4*) 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 *LmEIF4G3-4* 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 *LmEIF4G3-4* 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, FGRDFMV (GenBank accession NP495729). Similar sequences FSLDEVV and FSLERVL are present in the short N-termini of *LmEIF4G3* and *LmEIF4G4* 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 *LmEIF4G3* and *LmEIF4G4* 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 *LmEIF4E1*, *LmEIF4A1* and *LmEIF4G3* 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<sub>6</sub>-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 NIAD. 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](https://doi.org/10.1016/j.molbiopara.2004.12.001).

#### References

- [1] Mubich 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, Haitan A, Uziel 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] Nilzen TW. *Trans*-splicing: an update. *Mol Biochem Parasitol* 1995;73:1–6.
- [8] Pascolo E, Blonzi 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] Mauney PA, Denker JA, Darzynkiewicz E, Laneve R, Nilzen 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] Henzley 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] Prevet D, Dailix 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 eIF4E 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] Moino 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] Pironnet 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, Neanzakis 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, Kneepfer E, Smith DF. Poly(A)-binding protein I of *Leishmania*: functional analysis and localization 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, 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. Cocystal 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] Pectova 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 jamareschii*. *Proc Natl Acad Sci USA* 2001;98:1465–70.
- [61] Tanner NK, Cordin O, Barroques 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] Collier JM, Tucker M, Sheth U, Valencia-Sanchez MA, Parker R. The DEAD box helicase, Dhhlp, 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, Deaudoiff 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;269:3048–55.
- [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 Fakhy Y, Sereno D, Tamar S, Papadopoulos B. A new developmentally regulated gene family in *Leishmania amazonensis* encoding a homolog of amastin surface proteins. *Mol Biochem Parasitol* 2000;110:345–57.
- [78] Bowcher 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.



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<sup>1</sup>, Christian R. S. Reis, Rodolfo Katz, João R. C. Muniz<sup>2</sup>, Nancy Standart<sup>1</sup>, Mark Carrington<sup>1</sup> 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, <sup>1</sup>Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK and <sup>2</sup>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 *TbEIF4AI* and *TbEIF4AIII*). 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. *TbEIF4AI* is a very abundant predominantly cytoplasmic protein (over  $1 \times 10^5$  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, *TbEIF4AIII* 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 *TbEIF4AI*, but not of *TbEIF4AIII*, induced a slow growth phenotype in transfected cells. Overall, our results suggest that only *TbEIF4AI* is involved in protein synthesis while the properties and sequence of *TbEIF4AIII* 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

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

2496 *Nucleic Acids Research*, 2006, Vol. 34, No. 9

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, *LmEIF4A1* (LeiF) and *LmEIF4A2*, 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. *LmEIF4A1* is very abundant with over  $10^5$  molecules/cell whilst *LmEIF4A2* is either absent or present at levels below  $10^4$  molecules/cell. Furthermore, only *LmEIF4A1* 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 *LmEIF4A1* (named as *TbEIF4AI*) is the functional homologue of the eIF4A present in eIF4F. As for the orthologue of *LmEIF4A2*, it seems to be the functional homologue of the nuclear eIF4AIII present in higher eukaryotes and has been named here as *TbEIF4AII*.

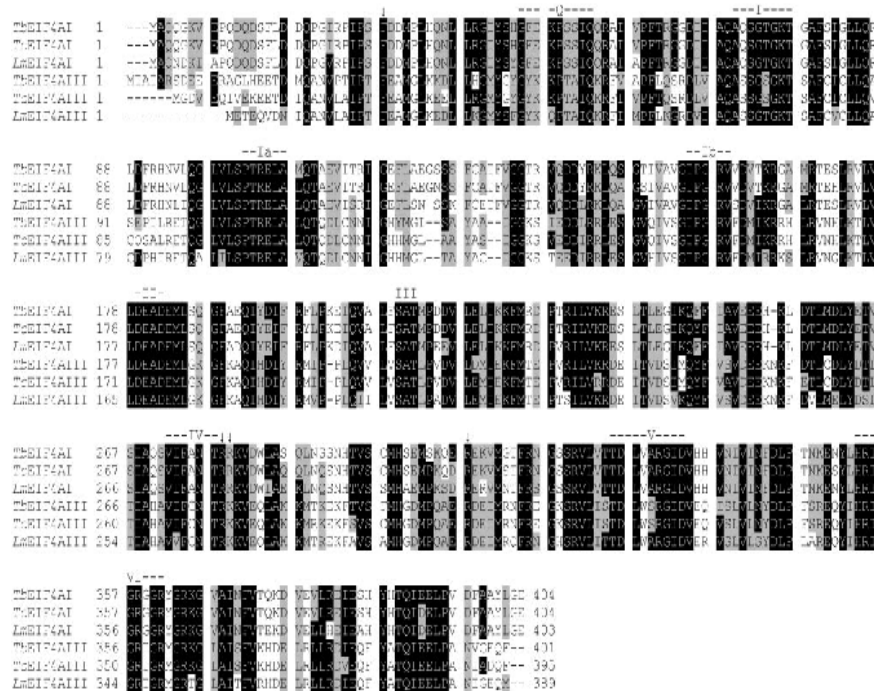
## MATERIALS AND METHODS

### 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](http://www.genedb.org)). Further sequence searches, Clustal W alignments and molecular modeling were done as described previously (33).

### PCR and cloning methods

The *TbEIF4AI* 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 *TbEIF4AIII* 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 *TbEIF4AI*-EYFP and *TbEIF4AIII*-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 ClustalW 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: *LmEIF4AI*, AAC24684/AAC24685; *LmEIF4AIII*, CAJ05468; *TbEIF4AI*, EAN76544; *TbEIF4AIII*, EAN79829; *TrEIF4AI*, EAN98527; *TrEIF4AIII*, EAN88971.

eIF4A-GSGSGPREQKLISEEDLPREQKLISEEDLPREQKLISEEDLPR.

**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% fetal 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<sup>6</sup>-10<sup>7</sup> 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<sub>2</sub>, supplemented with 10% FCS. Cultures grown to mid-log phase cultures (10<sup>5</sup>-10<sup>6</sup> 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 *TbEIF4As* and EP procyclin. A genomic repeat containing both  $\alpha$ - and  $\beta$ -tubulin genes was used to detect tubulin mRNA.

**Recombinant protein expression, antibody production and western blots**

His-tagged *TbEIF4AI* and *TbEIF4AIII* 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-*TbEIF4AI* antisera with *TbEIF4AIII* 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

2498 *Nucleic Acids Research*, 2006, Vol. 34, No. 9

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 ( $5 \times 10^6$ /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  $-20^\circ\text{C}/15$  min. Antibody detection of TbEIF4AI and III followed standard procedures. DNA was stained using Hoechst 33258. For the analysis of the cells expressing TbEIF4AI-EYFP and TbEIF4AIII-EYFP, aliquots of  $5 \times 10^6$  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, [ $^{35}\text{S}$ ]methionine (10  $\mu\text{Ci}/\text{ml}$ ) 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  $\mu\text{g}/\text{ml}$  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  $1 \times 10^7$  cells/ml in methionine-free RPMI 1640 containing 50  $\mu\text{Ci}/\text{ml}$  [ $^{35}\text{S}$ ]methionine and incubated for 1 h at  $28^\circ\text{C}$  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, TbEIF4AI and TbEIF4AIII, 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 LmEIF4A1 and 2, have also been renamed to LmEIF4AI and LmEIF4AIII 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 TbEIF4A mRNAs in *T. brucei* bloodstream and procyclic forms

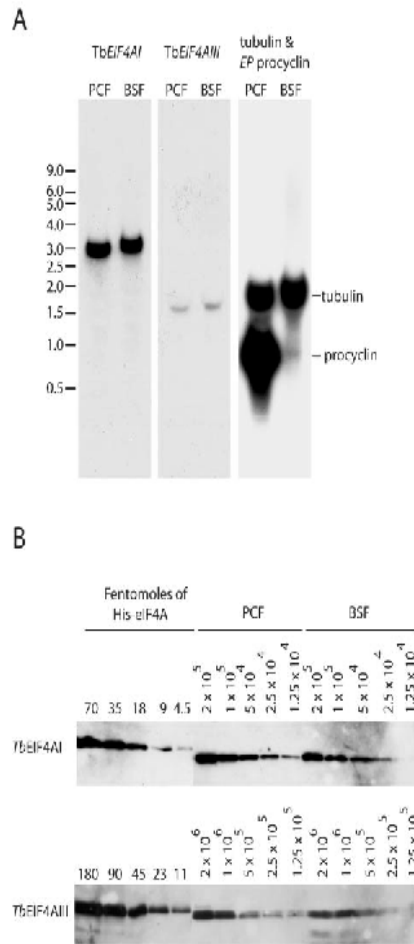
To begin the functional characterization of *T. brucei* eIF4A homologues and assay their expression at the mRNA level, the two genes were amplified, cloned and used as probes in northern blots of RNA from *T. brucei* procyclic and bloodstream forms (Figure 2A). The membranes were also probed for the 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 TbEIF4AI mRNA produces a much stronger signal than TbEIF4AIII. Since both probes used were of similar specific activity, and the exposures times for the films were similar as well, it seems that the TbEIF4AIII mRNA is far less abundant than TbEIF4AI. Remarkably, although the ORFs for both proteins are similar, 1215 versus 1206 bp for TbEIF4AI and TbEIF4AIII, respectively, their mRNAs differ significantly, with the TbEIF4AI message, at ~3 kb, being nearly twice the length of TbEIF4AIII (~1.6 kb), probably reflecting a considerable difference in the length of the 3'-UTR.

### TbEIF4AI is about 10-fold more abundant than TbEIF4AIII in both procyclic and bloodstream forms

Recombinant His-TbEIF4AI 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 *TbEIF4AI* 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 *TbEIF4AI* 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 *TbEIF4AI* and *TbEIF4AIII* in the procyclic and bloodstream forms of *T. brucei*. Recombinant His-tagged *TbEIF4AI* 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 ( $1.25 \times 10^5$ – $2 \times 10^5$  for *TbEIF4AI* and  $1.25 \times 10^5$ – $2 \times 10^6$  for *TbEIF4AIII*). 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 femtomoles/ $10^5$  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. *TbEIF4AI* 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, *TbEIF4AIII* levels were estimated at  $\sim 2 \times 10^4$  and  $1 \times 10^4$  molecules/cell in procyclic and bloodstream forms, respectively (Figure 2B). These data indicate that *TbEIF4AI* is present at levels at least 10-fold higher than *TbEIF4AIII*, 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), *TbEIF4AI* is in excess relative to mRNA, in contrast to *TbEIF4AIII*. Overall we conclude that both proteins are expressed constitutively and that only the *TbEIF4AI* levels are compatible with a role in translation.

**Subcellular localization of *TbEIF4AI* and III**

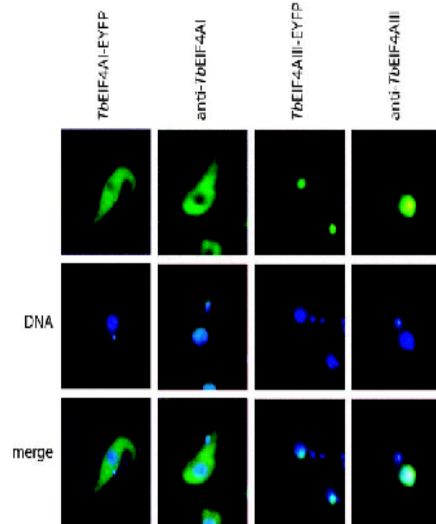
To determine the subcellular localization of the *T. brucei* eIF4A homologues we used two different experimental approaches. First the *TbEIF4AI* and *TbEIF4AIII* ORFs were cloned into the vector p2215 and the construct integrated into the non-transcribed spacer of a ribosomal RNA gene locus in the procyclic 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 *TbEIF4AI* 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. *TbEIF4AI*-EYFP is found predominantly in the cytoplasm, whilst *TbEIF4AIII*-EYFP is only found in the nucleus. These results were confirmed for the endogenous proteins through indirect immunofluorescence using isoform specific antibodies: again *TbEIF4AI* was mainly found in the cytoplasm whilst *TbEIF4AIII* was only detected in the nucleus (Figure 3).

**RNAi of *TbEIF4AI* and *TbEIF4AIII***

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 *TbEIF4AI* 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 *TbEIF4AI* 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 *TbEIF4AI* 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.

2500 *Nucleic Acids Research*, 2006, Vol. 34, No. 9

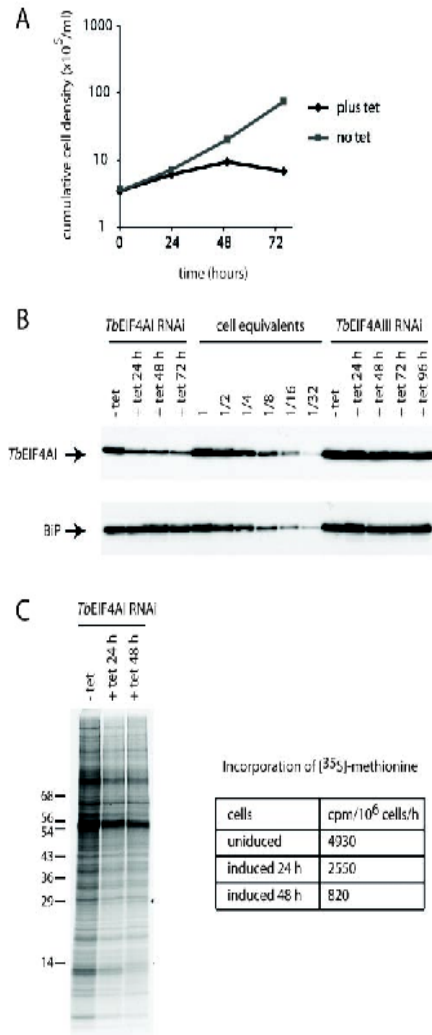


**Figure 3.** Subcellular localization of *TbEIF4AI* and *III* in *T. brucei* procyclic forms. Subcellular localization of the *TbEIF4AI* and *III* EYFP fusion proteins in transfected *T. brucei* cells was examined with a fluorescence microscope. The localization of native *TbEIF4AI* and *III* was also confirmed in wild-type procyclic cells (WT 427) by indirect immunofluorescence using the *TbEIF4AI* or *TbEIF4AIII* 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 *TbEIF4AIII* staining of the kinetoplast.

The phenotype of cells after RNAi ablation of *TbEIF4AIII* 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 *TbEIF4AIII* 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 *TbEIF4AIII* 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 *TbEIF4AIII* is not involved in the phenotype induced by the depletion of *TbEIF4AI* since no reduction in levels of *TbEIF4AIII* was observed in cells submitted to the *TbEIF4AI* RNAi procedure (Figure 5B). Likewise the *TbEIF4AIII* RNAi does not lead to any reduction in the levels of *TbEIF4AI* (Figure 4B). These results are compatible with *TbEIF4AIII* 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 *TbEIF4AI*, despite its abundance in wild-type cells, are much more sensitive to RNAi mediated depletion, consistent with a role in overall protein synthesis.

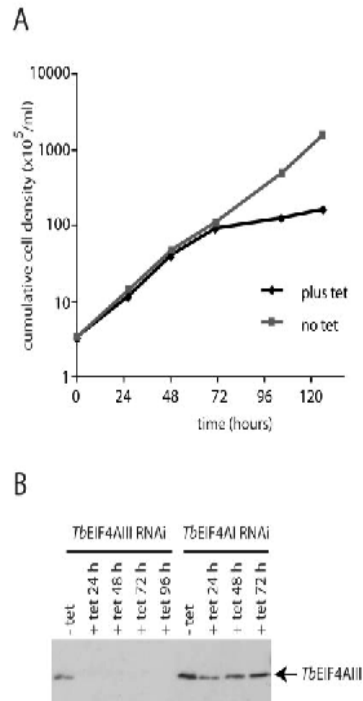
**Expression of dominant negative mutants of *TbEIF4AI* and *III* in transfected procyclic cells**

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



**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 phloeromycin and RNA interference induced after tetracycline addition. At regular intervals, cellular growth was monitored by counting the number of viable cells, expression of *TbEIF4AI* assayed and total protein synthesis investigated by [<sup>35</sup>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<sup>6</sup> cells and was used in the various RNAi lanes). *TbEIF4AI* 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/*TbEIF4AIII* construct (see also Figure 5) were also used in the blot to monitor for *TbEIF4AI* levels. (C) [<sup>35</sup>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 *TbEIF4AI* by incubating aliquots of the cells in the presence of [<sup>35</sup>S]methionine for 1 h followed by TCA precipitation, quantitation of the incorporated radioactivity or SDS-PAGE followed by autoradiography of the selected samples.

Incorporation of [ <sup>35</sup> S]-methionine	
cells	cpm/10 <sup>6</sup> cells/h
uninduced	4930
induced 24 h	2550
induced 48 h	820



**Figure 5.** RNAi of TbEIF4AIII. Procytic *T. brucei* cells were transfected with the p2T7-177/ TbEIF4AIII construct as described for Figure 4, monitored for cellular growth and assayed for expression of TbEIF4AIII. (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 TbEIF4AIII and TbEIF4AI RNAi experiments using the TbEIF4AIII antibodies. Samples from the same experiment assayed in Figure 4B were assayed for TbEIF4AIII 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), 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 TbEIF4AI transgenes was lower than expression from the endogenous gene whereas expression from the TbEIF4AIII transgenes was several fold higher than the endogenous protein. All the myc-tagged transgenes localized correctly (data not shown).

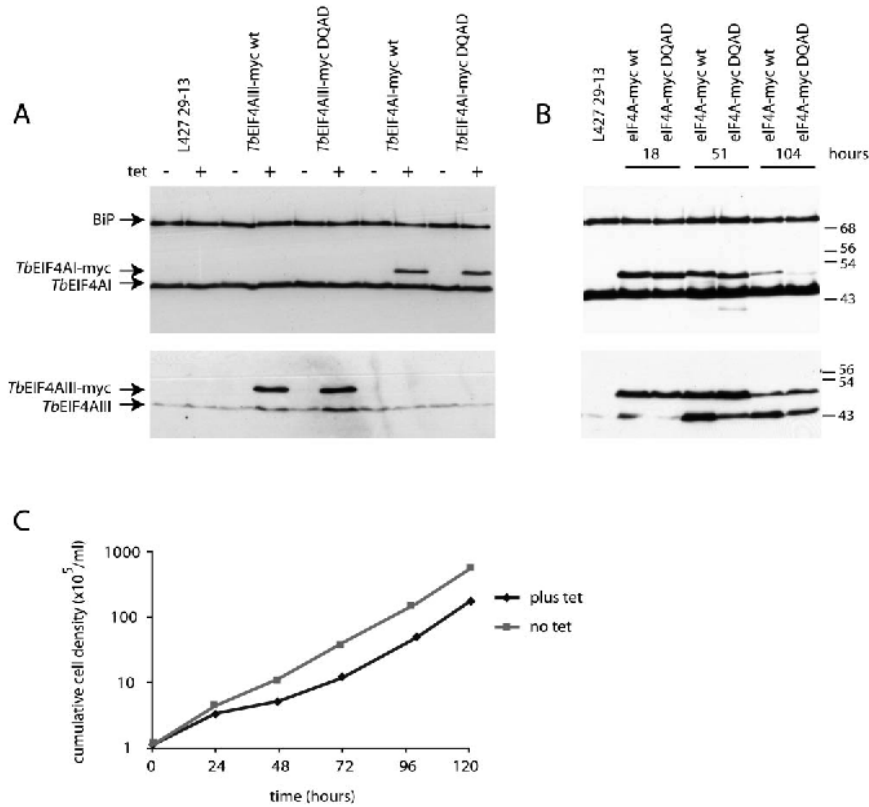
Cell growth and transgene expression was monitored over a time course (Figure 6B). Expression from the TbEIF4AI transgene reduced over the time course and was barely detectable by 104 h. We have observed this diminution of expression over time with other, but not all, transgenes expressed from vectors derived from pLEW100 and are unsure of the cause. The only transgene that had any effect on growth was the mutant form of TbEIF4AI (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 TbEIF4AI 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 TbEIF4AIII mutant transgene had no effect on growth of the culture. Interestingly, the expression of the TbEIF4AIII transgenes, but not the TbEIF4AI transgenes, resulted in increased levels of the endogenous protein. Overall, the results are compatible with the RNA helicase activity of TbEIF4AI being strictly required for growth. As for TbEIF4AIII, 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 TbEIF4AIII is not active in translation.

**Mapping of isoform specific amino acids**

The results described above for TbEIF4AIII 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 TbEIF4AI 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 cerevisiae*, Fal1p, a nucleolar protein shown to be required for 40S ribosomal subunit formation (45). Fal1p, however, does not seem to be involved in EJC

2502 *Nucleic Acids Research*, 2006, Vol. 34, No. 9



**Figure 6.** Expression of myc-tagged dominant negative mutants of *TbEIF4AI* and *III* in procyclic cells. (A) Western blot analysis of the expression of the various *TbEIF4AI* 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 *TbEIF4AI* western blot was simultaneously probed with anti-BIP as a loading control. (B) Time course expression of the different versions of *TbEIF4A-myc* after tetracycline addition to the culture. The *TbEIF4AI* western blot was simultaneously probed with anti-BIP as a loading control. (C) Effect of the expression of the dominant negative form of *TbEIF4AI-myc* on the growth of the transfected cells in culture.

formation since a search in *S.cerevisiae* 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 (*TbEIF4A 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 *TbEIF4AI* and *TbEIF4AIII*

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



2504 *Nucleic Acids Research*, 2006, Vol. 34, No. 9

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

Position: TbEIF4I	Substitution: 4AI > 4AIII	Secondary structure	Domain	Overall position in predicted tertiary structure
93	K/N/V > R	Loop	N-terminal	Near $\alpha 5$ /exposed
139	Q/R > E/G	$\alpha 5$	N-terminal	Exposed
146	Q/A > E/D	$\alpha 5$	N-terminal	Exposed
153	V > S/A <sup>a</sup>	$\beta 4$	N-terminal	Buried
197	F > Y	$\alpha 8$	N-terminal	Partially exposed
213	M > L	Loop	N-terminal	Interface/next to Motif III/partially buried
220	L/V/I > M	$\alpha 9$	N-terminal	Buried
226	R/N > T	Loop	C-terminal	Exposed
256	L > F <sup>a</sup>	$\alpha 10$	C-terminal	Partially buried
264	E > D <sup>a</sup>	$\alpha 10$	C-terminal	Exposed
303	S/D > P <sup>a</sup>	Loop	C-terminal	Exposed
328	V/L > W <sup>a</sup>	Loop	C-terminal	Interface/Motif V/Buried
388	H/N > A/S <sup>a</sup>	Loop	C-terminal	Exposed

<sup>a</sup>These substitutions but not the others are present in the *S.cerevisiae* 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 TbEIF4AI 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 TbEIF4AIII 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 TbEIF4AI is very limited in the number of interactions it can establish. It is possible then that the presence of the W residue in TbEIF4AIII, 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 TbEIF4AI 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, TbEIF4AIII does not seem to play an obvious role in protein synthesis. The nuclear localization of TbEIF4AIII, 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; *LmEIF4AI* binds strongly to at least two eIF4G homologues whereas *LmEIF4AIII* 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, TbEIF4AIII 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

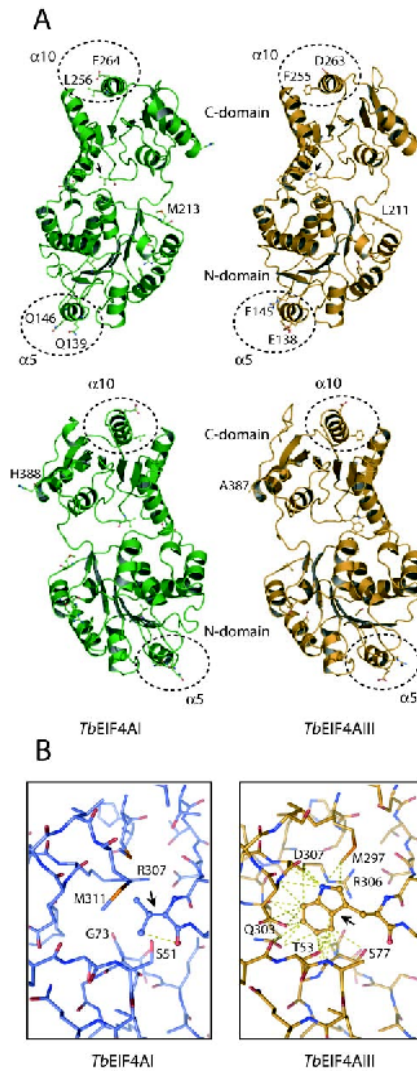


homologues but so far without success. However, both Y14 and Aly/REF are small RNA-binding proteins with single RRM, a category which includes many proteins with unassigned functions in those three genomes (48). It may

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 *TbEIF4AI* 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 *TbEIF4AI* 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,  $\alpha 5$  and  $\alpha 10$ . The H/N388A/S and M213L substitutions are also indicated (their numbering differ however from the eIF4AI/eIF4AIII sequences—for instance, H388 in *TbEIF4AI* is equivalent to A387 in *TbEIF4AIII* and so on). (B) Balls and sticks representation showing the neighbourhood of the L328W substitution in both *TbEIF4AI* 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.

2506 *Nucleic Acids Research*, 2006, Vol. 34, No. 9

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- $\alpha$  (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 *TbEIF4AIII* 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

- Campbell,D.A., Thomas,S. and Sturm,N.R. (2003) Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect.*, **5**, 1231–1240.
- Liang,X.H., Hanitan,A., Uliel,S. and Michaeli,S. (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell*, **2**, 830–840.
- 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.
- 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.
- 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.
- Preis,T. and Hentze,M.W. (2003) Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays*, **25**, 1201–1211.
- Sonenberg,N. and Dever,T.E. (2003) Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol.*, **13**, 56–63.
- 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.
- Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases-amino-acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.
- Linder,P. (2003) Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol. Cell*, **95**, 157–167.
- Rocak,S. and Linder,P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell Biol.*, **5**, 232–241.
- 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.
- Story,R.M., Li,H. and Abelson,J.N. (2001) Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jamastris*. *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
- Caruthers,J.M. and McKay,D.B. (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol.*, **12**, 123–133.
- 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.
- 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.
- 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 Pdc4d is essential for its binding to eukaryotic translation initiation factor 4A. *Mol. Cell. Biol.*, **24**, 3894–3906.
- Svitkin,Y.V., Pause,A., Haghghat,A., Pironnet,S., Witherell,G., Belsham,G.J. and Sonenberg,N. (2001) The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA*, **7**, 382–394.
- 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.
- 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.
- Yoder-Hill,J., Pause,A., Sonenberg,N. and Merrick,W.C. (1993) The p46 subunit of eukaryotic initiation factor (eIF)-4E exchanges with eIF-4A. *J. Biol. Chem.*, **268**, 5566–5573.
- 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.
- Holzmann,K., Gerner,C., Poll,A., Schafer,R., Obrist,P., Ensinger,C., Grimm,R. and Saueremann,G. (2000) A human common nuclear matrix protein homologous to eukaryotic translation initiation factor 4A. *Biochem. Biophys. Res. Commun.*, **267**, 339–344.
- 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.
- Cham,C.C., Dostie,J., Diem,M.D., Feng,W., Mann,M., Rappilber,J. and Dreyfuss,G. (2004) eIF4A3 is a novel component of the exon junction complex. *RNA*, **10**, 200–209.
- Palacios,J.M., Gatfield,D., Slj.D. and Izauralde,E. (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*, **427**, 753–757.
- 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.
- Ballut,L., Marchadier,B., Bague,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.
- 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.
- 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.
- 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. Dhaliya, 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., Method, 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., Doerre, M., Linder, P. and Barroques, 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., Barroques, J., Doerre, 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) Fall1p is an essential DEAD-box protein involved in 40S-ribosomal-subunit biogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **17**, 7283–7294.
46. Cheng, Z., Collier, 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 mRNA 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.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)