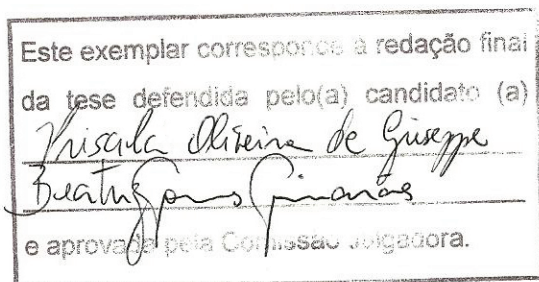


UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

PRISCILA OLIVEIRA DE GIUSEPPE

**“Estudos estruturais de proteínas de *Leptospira interrogans*
sorovar Copenhageni potencialmente localizadas no envelope
celular”**



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

Orientadora: Profa. Dra. Beatriz Gomes Guimarães

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

Campinas, 2010

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

G449e	<p>Giuseppe, Priscila Oliveira de Estudos estruturais de proteínas de <i>Leptospira interrogans</i> sorovar Copenhageni potencialmente localizadas no envelope celular / Priscila Oliveira de Giuseppe. – Campinas, SP: [s.n.], 2010.</p> <p>Orientadores: Beatriz Gomes Guimarães, Nilson Ivo Tonin Zanchin. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.</p> <p>1. <i>Leptospira interrogans</i>. 2. Proteínas - Estrutura. 3. Cristalografia. 4. Parede celular. I. Guimarães, Beatriz Gomes. II. Zanchin, Nilson Ivo Tonin. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.</p> <p style="text-align: right;">(rcdt/ib)</p>
--------------	--

Título em inglês: Structural studies of proteins from *Leptospira interrogans* serovar Copenhageni potentially located at the cell envelope.

Palavras-chave em inglês: *Leptospira interrogans*; Proteins - Structure; Crystallography; Cell wall.

Área de concentração: Genética de Microorganismos.

Titulação: Doutora em Genética e Biologia Molecular.

Banca examinadora: Beatriz Gomes Guimarães, Richard Charles Garratt, Marcos Roberto de Mattos Fontes, Anete Pereira de Souza, Celso Eduardo Benedetti.

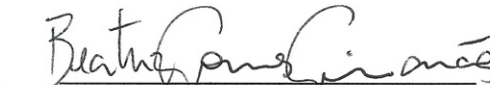
Data da defesa: 06/08/2010.

Programa de Pós-Graduação: Genética e Biologia Molecular.

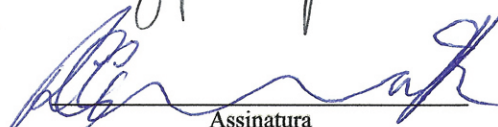
Campinas, 06 de Agosto de 2010.

BANCA EXAMINADORA

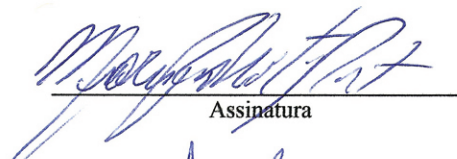
Profª. Dra. Beatriz Gomes Guimarães (Orientadora)


Assinatura

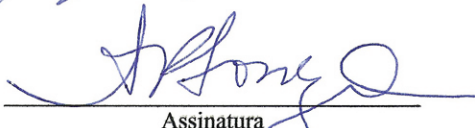
Prof. Dr. Richard Charles Garratt


Assinatura

Prof. Dr. Marcos Roberto de Mattos Fontes


Assinatura

Profª. Dra. Anete Pereira de Souza


Assinatura

Prof. Dr. Celso Eduardo Benedetti


Assinatura

Prof. Dr. Ronaldo Alves Pinto Nagem

Assinatura

Prof. Dr. Ricardo Aparicio

Assinatura

Prof. Dr. Mário Tyago Murakami

Assinatura

AGRADECIMENTOS

À Beatriz Guimarães pelos valiosos ensinamentos, pelo companheirismo, pelo comprometimento, pela dedicação e pela confiança ao longo de todos esses anos.

Ao Nilson Zanchin, pela acolhida no seu grupo de pesquisa, pelo que muito me ensinou e por todo seu apoio e auxílio.

À Profa. Dra. Ana Lúcia T.O. do Nascimento e às Dras. Fernanda O. Neves e Marina Von A. dos Reis pelos clones cedidos e pela colaboração na concepção desse projeto.

Ao Dr. Mathieu Picardeau e à Dra. Leyla Slamti pela oportunidade ímpar de aprender a manipular geneticamente leptospiras.

À Profa. Dra. Anete Pereira de Souza e aos Profs. Drs. Richard Charles Garratt, Marcos Roberto de Mattos Fontes e Celso Eduardo Benedetti pela participação na banca examinadora desse trabalho.

Aos pesquisadores, funcionários e alunos do LNBio pela troca de experiências, por todo auxílio e pelo agradável ambiente de trabalho.

À Andréia, Celisa, Adriana, Elaine, Tereza, Givanil, Eugênia, Veruska e Renata pelo apoio técnico e pela prestatividade.

À Tereza também pelo auxílio na clonagem de variantes de um dos alvos desse trabalho.

Ao Amadeu Iglesias, ao Eduardo Pilau e à Margareth Navarro pelos experimentos de espectrometria de massas.

Ao Lucas, Fernando e Walan pelo suporte técnico nas linhas de luz MX1 e MX2.

A toda equipe do Dr. Mathieu Picardeau pela receptividade e pelo agradável convívio.

Às equipes das linhas de luz PROXIMA 1 e 2 e a todos os colegas do Synchrotron SOLEIL, em especial à Blandine Pineau, pelo apoio e convivialidade.

Ao LNBio, ao LNLS, ao Instituto Pasteur e ao Synchrotron SOLEIL pelo fornecimento de uma excelente infra-estrutura para a pesquisa.

Ao Programa de Pós-Graduação em Genética e Biologia Molecular da UNICAMP.

À FAPESP e à CAPES pelo apoio financeiro.

A meus pais, Iveti e José Luís por tudo o que me ensinaram e continuam me ensinando, pelos exemplos de caráter, bondade e honestidade e por todo amor, carinho e incentivo, fundamentais para a superação dos momentos difíceis.

Ao meu marido, Daniel, por todo apoio, amor, paciência, dedicação e companheirismo que me auxiliaram muito a superar os obstáculos e seguir em frente.

Às minhas queridas irmãs, Gisele e Larissa, por todo apoio, amor e incentivo e por serem essas pessoas maravilhosas das quais eu tanto me orgulho.

Às minhas queridas amigas Daniela Thomazella, Karine Santos, Kívia Oliveira, Camila Santos, Juliana Oliveira e Tatiana Brasil pelo companheirismo, pelos exemplos de determinação, dedicação e perseverança e por todo apoio e incentivo.

Às minhas queridas amigas de república Carolina, Melina, Lívia, Bianca e Priscila, que acompanharam de perto as alegrias e as dificuldades por mim experimentadas ao longo de todos esses anos de trabalho. Sou grata pelo agradável convívio, por todo carinho, atenção e apoio.

A todos os meus parentes e amigos, sem o apoio e o carinho dos quais eu não teria chegado tão longe.

SUMÁRIO

LISTA DE FIGURAS.....	viii
LISTA DE ABREVIACÕES	ix
PREFÁCIO	xi
RESUMO.....	xii
ABSTRACT	xiii
1. INTRODUÇÃO	1
1.1. A LEPTOSPIROSE E SEUS AGENTES CAUSAIS.....	1
1.2. SEQUENCIAMENTO GENÔMICO DE <i>LEPTOSPIRA</i> SPP.	5
1.3. PREDIÇÃO DE FUNÇÃO DE PROTEÍNAS	7
1.4. ALVOS DO PROJETO	8
2. OBJETIVOS	10
3. RESULTADOS	11
3.1. ARTIGO I	12
3.2. ARTIGO II.....	21
3.3. PRODUÇÃO E CARACTERIZAÇÃO DE LIC10494	56
4. DISCUSSÃO.....	80
4.1. ESTUDOS ESTRUTURAIS DA PROTEÍNA LP49	80
4.1.1. Cristalização de Lp49.....	80
4.1.2. A estrutura tridimensional do antígeno leptospiral Lp49 sugere uma provável função de ligação a proteínas	82
4.2. ESTUDOS ESTRUTURAIS DA PROTEÍNA LIC12922	86
4.2.1. Cristalização de LIC12922	86
4.2.2. A estrutura tridimensional da proteína hipotética conservada LIC12922 revela homologia com a chaperona periplasmática SurA.....	89
4.3. PRODUÇÃO E CARACTERIZAÇÃO DE LIC10494.....	93

5. CONCLUSÕES	95
6. REFERÊNCIAS BIBLIOGRÁFICAS	98
ANEXO	110
A. ESTUDOS DE BIOLOGIA MOLECULAR E GENÉTICA EM <i>LEPTOSPIRA</i> SPP.	111
A.1. Introdução	111
A.2. Geração da biblioteca de mutantes de <i>L. biflexa</i>	112
A.3. Marcadores de resistência a antibiótico para manipulação de <i>Leptospira</i> spp.	115
<i>A.3.1. Artigo III</i>	116
A.4. Referências bibliográficas	121

LISTA DE FIGURAS

FIGURA 1.1. Esquema representativo do envelope celular de <i>Leptospira</i> spp.	2
TABELA 1.1. Características das proteínas alvo deste projeto.	8
FIGURA 4.1. Cristais de LIC10793 observados em condições dos ensaios iniciais.	80
FIGURA 4.2. Cristais de LIC10793 crescidos na condição 17,5 % PEG3000 0,175 M NaCl 0,09 M Tris HCl pH 8.5, aplicando-se a técnica de “ <i>microseeding</i> ”.	81
FIGURA 4.3. Cristais de Lp49 marcada com SeMet.	82
FIGURA 4.4. Comparação estrutural entre a cavidade eletropositiva da face superior do β - <i>propeller</i> de Lp49 e o sítio ativo da enzima NcCMLE.	86
FIGURA 4.5. Exemplos de cristais de LIC12922 observados nos refinamentos em que se aliaram a técnica de <i>microseeding</i> e a cristalização em gel.	88
FIGURA A1. Exemplo de placa EMJH 0,3 % ágar noble contendo colônias mutantes de <i>L. biflexa</i>	112
FIGURA A2. Fotos de microscopia eletrônica do mutante M41 (painel A) e de <i>L. biflexa</i> selvagem (painel B).	114
FIGURA A3. Curvas de crescimento do mutante M41 e de <i>L. biflexa</i> selvagem.	114

LISTA DE ABREVIACÕES

aa.	aminoácido
CD	dicroísmo circular (<i>circular dichroism</i>)
C-terminal	carboxy-terminal
CV	volume de coluna (<i>column volume</i>)
DLS	espalhamento de luz dinâmico
FPLC	cromatografia líquida rápida de proteínas (<i>Fast protein liquid chromatography</i>)
Int	intensidade
IPTG	isopropil- β -D-galactopiranosídeo
IZIT	corante azul de metileno tri-hidratado $C_{16}H_{18}ClN_3S \cdot 3H_2O$ (Hampton Research)
LPS	lipopolissacarídeo
MW	massa molecular (<i>molecular weight</i>)
N-terminal	amino-terminal
OD₆₀₀	densidade ótica a 600 nm
Pd	índice de polidispersividade
PDB	banco de dados de estrutura de proteínas (<i>Protein Data Bank</i>)
PEG	polietilenoglicol
PI	ponto isoelétrico
PMSF	fluoreto de fenilmetilsulfonil (<i>phenylmethanesulfonylfluoride</i>)
PPIase	peptidil prolil isomerase

PPI-Pred	método de predição de interface proteína-proteína (<i>Protein-Protein Interface Prediction Prediction Server</i>)
ProFunc	servidor para predição de função de proteínas a partir de sua estrutura tridimensional
PS	peptídeo sinal
PSIPRED	servidor de predição de estrutura secundária de proteínas (<i>Protein Structure Prediction Server</i>)
R_h	raio hidrodinâmico
SDS-PAGE	Eletroforese em gel de poliacrilamida na presença de duodecil-sulfato de sódio (<i>sodium duodecyl-sulphate polyacrilamide gel electrophoresis</i>)

PREFÁCIO

O conteúdo deste trabalho foi dividido em cinco capítulos. O primeiro deles contém uma introdução sobre as características da leptospirose e de *Leptospira interrogans*, as informações reveladas pelo sequenciamento genômico de *Leptospira* spp., uma breve justificativa para a escolha da cristalografia de proteínas como abordagem experimental e, por fim, a apresentação dos alvos do projeto. O segundo capítulo apresenta o objetivo central do trabalho bem como as etapas experimentais necessárias para o seu cumprimento. O terceiro capítulo é dedicado aos resultados e está dividido em três seções. A seção 3.1 consiste em um artigo, publicado na revista *Journal of Structural Biology*, que contém a determinação e a análise da estrutura cristalográfica do alvo LIC10793 (Lp49). A seção 3.2 refere-se a um manuscrito submetido para publicação no qual são apresentadas a determinação e a análise da estrutura do alvo LIC12922. A seção 3.3 descreve a expressão, purificação, cristalização e caracterização do alvo LIC10494. No capítulo quatro são discutidos, separadamente, os resultados obtidos para cada alvo. Nas subseções 4.1.1 e 4.2.1, destacam-se as dificuldades experimentais e as soluções encontradas para a produção de cristais de Lp49 e LIC12922 adequados para o experimento de difração. O quinto capítulo contém as conclusões do trabalho. Esta tese apresenta ainda um anexo no qual são relatadas as atividades desenvolvidas durante um estágio realizado na Unidade de Biologia de Espiroquetas do Instituto Pasteur (Paris, FR), no contexto da participação no Programa de Doutorado no País com Estágio no Exterior (PDPEE) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

RESUMO

Leptospira interrogans é uma bactéria espiroqueta que causa a leptospirose, uma zoonose de distribuição mundial que afeta mais de 500.000 pessoas anualmente. Pouco se sabe sobre a biologia de leptospiros, o que dificulta a elaboração de novas estratégias de prevenção e de tratamento contra a doença. Cerca de 60 % dos genes de *L. interrogans* codifica proteínas que não apresentam similaridade de sequência significativa com proteínas de função conhecida. Como a estrutura cristalográfica de uma proteína pode revelar vários indícios funcionais, este trabalho visou à determinação da estrutura cristalográfica das proteínas LIC10793, LIC12922 e LIC10494 de *L. interrogans*, que são potencialmente localizadas no envelope celular e não são funcionalmente caracterizadas. A estrutura do antígeno LIC10793 (Lp49) foi resolvida a 2 Å de resolução e revelou que essa provável lipoproteína apresenta dois domínios. O domínio N-terminal de Lp49 possui um enovelamento do tipo Imunoglobulina e sua topologia diverge das formas padrão do motivo estrutural “chave grega” (*Greek key motif*). O domínio C-terminal consiste em um β -propeller formado por sete folhas β . Comparações locais não identificaram nenhum sítio catalítico conhecido em Lp49, mas análises de sua superfície revelaram a presença de prováveis sítios de ligação a proteínas. Com base nesses indícios, na provável localização de Lp49 na membrana externa e em sua antigenicidade, postula-se que Lp49 tenha uma função de interação com outras proteínas podendo desempenhar um papel na interação entre leptospiros e seus hospedeiros. A estrutura cristalográfica de LIC12922, determinada a 3,1 Å de resolução, revelou a presença de dois domínios. O domínio NC é estruturalmente relacionado a domínios que apresentam atividade chaperona, encontrados nas proteínas SurA e *trigger factor* de *E. coli*. O domínio parvulina de LIC12922 não apresenta atividade de peptidil prolil isomerase, mas possui um provável sítio de interação a proteína que inclui o sítio de reconhecimento ao substrato proposto para o domínio parvulina P1 de SurA. Análises filogenéticas sugerem que LIC12922 e as chaperonas extracitoplasmáticas SurA, PpiD e PrsA apresentam um ancestral comum. Com base nesses indícios e na provável localização de LIC12922 no periplasma, propõe-se que LIC12922 seja uma chaperona periplasmática envolvida na biogênese de proteínas da membrana externa. LIC10494 foi expressa, purificada e cristalizada. Refinamentos das condições de cristalização não foram suficientes para se obter cristais adequados ao experimento de difração. Análises da sua sequência evidenciaram que LIC10494 apresenta uma extensa região central intrinsecamente desordenada rica em resíduos de treonina. Assim como proteínas que possuem domínios intrinsecamente desestruturados, LIC10494 apresenta mobilidade mais lenta do que o esperado em SDS-PAGE, um volume de eluição menor do que o esperado em ensaios de gel filtração e uma considerável contribuição de configurações randômicas em seu espectro de dicroísmo circular.

ABSTRACT

Leptospira interrogans is a spirochaetal bacterium which causes leptospirosis, a worldwide spread zoonosis that affects more than 500,000 people annually. Little is known about the biology of leptospire, which difficults the development of new preventive and treatment strategies for the disease. About 60 % of the genes from *L. interrogans* encode for proteins that did not show significative sequence similarity with proteins of known function. Since the tridimensional structure of a protein can contribute to the understanding of its function, this work aimed at the crystallographic structure determination of the proteins LIC10793, LIC12922 and LIC10494 from *L. interrogans*, which are potentially situated at the cell envelope and are not functionally characterized. The structure of the antigen LIC10793 (Lp49) was determined at 2 Å resolution and revealed that this probable lipoprotein possesses two domains. The Lp49 N-terminal domain presents an Immunoglobulin-like fold and its topology diverges from the standard patterns of the Greek key motif. The C-terminal domain is a 7-bladed β-propeller. Local structural comparisons did not identify known catalytic sites at Lp49, but surface analyses evidenced potential protein binding sites. Based on these results, the putative localization of Lp49 at the outer membrane and its role as an antigen, we postulate that Lp49 has a protein binding function involved in *Leptospira*-host interaction. The LIC12922 crystal structure, determined at 3.1 Å resolution, revealed two domains. The NC domain is structurally related to the chaperone domains of *E. coli* SurA and trigger factor proteins. The LIC12922 parvulin domain is devoid of peptidyl prolyl isomerase activity, but presents a putative protein binding site which includes the substrate recognition site proposed to the first parvulin domain of SurA. Phylogenetic analyses suggest that LIC12922 and the extracytoplasmic chaperones SurA, PpiD and PrsA have a common ancestor. Based on the structural and phylogenetic analyses and taking into account its probable periplasmic localization we postulate that LIC12922 is a periplasmic chaperone involved at the biogenesis of outer membrane proteins. The protein LIC10494 was expressed, purified and crystallized. In spite of extensive refinement of crystallization conditions the crystals were not adequate for diffraction experiments. Sequence analyses evidenced that LIC10494 has an extensive central region intrinsically disordered which is rich in threonine residues. Similarly to proteins that possess intrinsically disordered domains, LIC10494 presents mobility slower than expected at SDS-PAGE, elution volume smaller than expected in gel filtration assays and a considerable contribution of random coil structures in circular dichroism spectrum.

1. INTRODUÇÃO

1.1. A LEPTOSPIROSE E SEUS AGENTES CAUSAIS

A leptospirose é uma zoonose de distribuição mundial que, na última década, emergiu como um sério problema de saúde pública, dado o aumento repentino do número de casos e epidemias registrados (Sambsiava *et al.*, 2003). No estado de São Paulo, uma taxa de mortalidade de cerca de 10 % foi estimada num período de 23 anos entre pacientes hospitalizados (Lomar *et al.*, 1996). Grandes surtos ocorreram na América Central e na América do Sul após enchentes resultantes do excesso de chuva relacionado ao El Niño aumentando assim o ressurgente interesse internacional pela doença (Levett, 2001).

Apesar de ser considerada uma importante causa de morbidade e mortalidade entre populações pobres, principalmente em regiões tropicais e em países em desenvolvimento, a leptospirose é uma doença negligenciada (Ko *et al.*, 2009). Mais de 500.000 casos de leptospirose severa são registrados a cada ano (WHO, 1999).

Doença endêmica de trabalhadores da zona rural, a leptospirose tem se tornado um problema grave nas favelas dos grandes centros urbanos, onde a falta de saneamento básico gera condições para a transmissão via ratos (Ko *et al.*, 1999, Riley *et al.*, 2007, Reis, *et al.*, 2008). Além disso, a leptospirose também tem impactos consideráveis na pecuária por causar transtornos reprodutivos, abortos e redução na produção de leite em diversos animais domésticos (Ellis, 1994; Grooms, 2006; Lilenbaum *et al.*, 2007).

Bactérias patogênicas do gênero *Leptospira* são os agentes causadores da leptospirose. Esse gênero compreende bactérias espiroquetas altamente móveis, aeróbicas obrigatórias, que possuem, como outras espiroquetas, características tanto de bactérias Gram positivas quanto de Gram negativas (Haake, 2000). Como em bactérias Gram positivas, sua membrana citoplasmática é fortemente associada à parede celular de peptidoglicano. Como em bactérias Gram negativas, apresentam uma membrana externa que encerra o espaço periplasmático (Haake, 2000) (FIGURA 1.1).

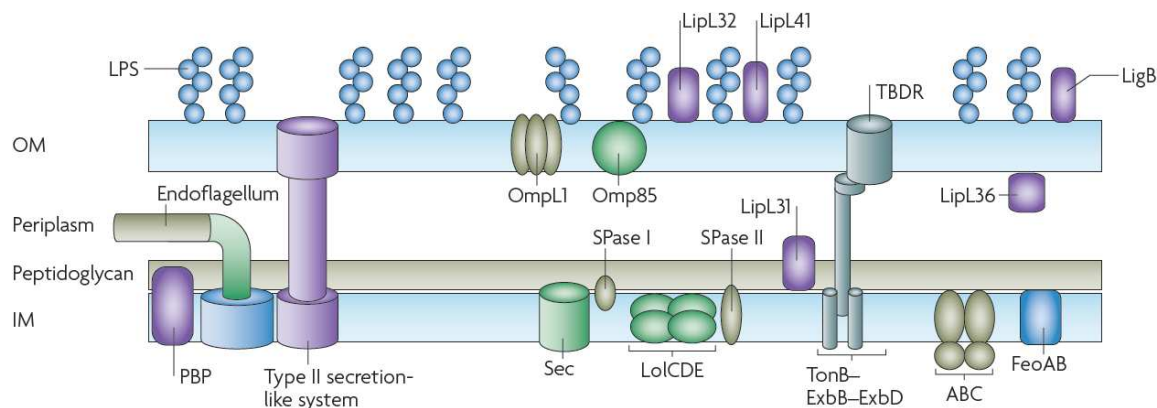


FIGURA 1.1. Esquema representativo do envelope celular de *Leptospira* spp. Na membrana externa (OM) de leptospiros observamos lipopolissacarídeos (LPS), lipoproteínas (ex. LipL32, LipL41, LigB e LipL36) e proteínas transmembrana (OMPs) tais como OmpL1, Omp85 e TBDR (receptor dependente de TonB). Como em outras espiroquetas, os flagelos de leptospiros localizam-se no periplasma e a parede de peptídeo glicano associa-se à membrana interna (IM). Nesta, encontram-se lipoproteínas (ex. LipL31) e proteínas transmembrana tais como proteínas ligadoras de penicilina (PBP), proteínas transportadoras de ferro (FeoAB), componentes do sistema de transporte ABC, proteínas do transloco Sec, através do qual são transportadas OMPs e lipoproteínas recém sintetizadas no citoplasma, proteínas do sistema Lol, que participam do endereçamento de lipoproteínas para a membrana externa, e as peptidases SPase I e II, responsáveis pela clivagem do peptídeo sinal de proteínas translocadas para o envelope celular. Nessa representação destacam-se ainda dois sistemas de transporte que apresentam componentes na membrana interna, no periplasma e na membrana externa (sistema de secreção tipo II e o complexo TonB-ExbB-ExbD-TBDR). Figura extraída de Ko *et al.* 2009.

O gênero *Leptospira* foi originalmente dividido em duas espécies, *L. biflexa*, compreendendo as linhagens saprófitas e *L. interrogans* abrangendo todas as linhagens patogênicas (Everard, 1996; Johnson & Faine, 1984). Porém, com o advento da classificação genotípica, o gênero sofreu uma re-classificação e apresenta, atualmente, 20 espécies (Ko *et al.*, 2009).

L. interrogans e *L. borgpetersenii* são as espécies patogênicas responsáveis pela maioria dos casos de leptospirose e compreendem quase a metade das 230 sorovares conhecidas (Brenner *et al.*, 1999, Bulach *et al.*, 2006). A classificação em sorovares é baseada na heterogeneidade estrutural do carboidrato que compõe o LPS, presente na membrana externa dessas bactérias, e é importante para estudos clínicos e epidemiológicos (Ko *et al.*, 2009).

Mamíferos selvagens e domésticos, bem como répteis e anfíbios servem como reservatórios das mais de 200 linhagens patogênicas conhecidas (Plank & Dean, 2000).

Ratos e outros roedores são a fonte mais importante para a infecção humana (Plank & Dean, 2000). Em ratos, *L. interrogans* coloniza cronicamente os túbulos renais proximais e são excretadas na urina, o que facilita a transmissão para outros hospedeiros (Faine *et al.*, 1999). Esta se dá principalmente pelo contato com a urina ou tecido de animais infectados e com solo ou água contaminados, através de cortes ou abrasões cutâneas ou via a conjuntiva (Lecour *et al.*, 1989; Everard *et al.*, 1995; Levett, 2001). Após entrar no hospedeiro, *L. interrogans* chega rapidamente à circulação sanguínea, através da qual é disseminada para múltiplos órgãos (Barochi *et al.*, 2002).

A leptospirose produz um amplo espectro de sintomas que vão desde febres de início repentino à icterícia, falência renal e hemorragia pulmonar potencialmente letal (Bharti *et al.*, 2003). Sua histopatologia é caracterizada por danos endoteliais, vasculites e infiltrados inflamatórios observados geralmente no fígado, rins, coração e pulmões (Levett, 2001).

O uso de antibióticos como penicilina G, ampicilina e doxiciclina tem sido a terapia adotada para combater o agente causador (Bharti *et al.*, 2003). Seus benefícios são maiores quando empregues na fase inicial da doença (Faine *et al.*, 1999, World Health Organization, 2003). Porém, o diagnóstico correto e rápido da leptospirose é dificultado pelos sintomas comuns a outras doenças febris, tais como a dengue, principalmente em países em desenvolvimento (Mcbride, *et al.*, 2005). Atualmente, dispõe-se de alguns testes diagnósticos para a leptospirose, porém eles têm sido associados a um alto custo e a uma baixa sensibilidade durante a fase aguda da doença (Suputtamongkol *et al.*, 2010).

Vacinas baseadas em bactérias mortas são disponíveis em alguns países para uso em humanos e têm sido usadas há anos no campo veterinário (Bharti *et al.*, 2003, McBride *et al.*, 2005). Porém, tais vacinas estão geralmente associadas a efeitos colaterais inaceitáveis e à eficácia altamente limitada e de curta duração (Bharti *et al.*, 2003).

Há alguns anos, Cuba e China desenvolveram, respectivamente, vacinas bacterianas e vacinas baseadas em membrana externa para uso humano (Yan *et al.*, 2003; Martinez *et al.*, 2004). Durante a triagem, não foram registradas reações adversas graves, geralmente associadas a vacinas bacterianas. Porém, elas não oferecem imunidade protetora cruzada para os sorogrupos que não estão representados em sua formulação (McBride *et al.*, 2005). Dessa forma, a busca por alvos para o desenvolvimento de vacinas de subunidades, ou seja, baseadas em antígenos protéicos, tem crescido nos últimos anos e pesquisas nesse sentido

estão sendo desenvolvidas (Haake *et al.*, 1999; Branger *et al.* 2001; Haake *et al.* 2004, Ko *et al.*, 2009).

Embora leptospiros tenham sido identificadas como o agente causal da leptospirose há cerca de 100 anos atrás (Inada *et al.*, 1916), pouco se sabe sobre a biologia e a patogenicidade dessas bactérias, o que dificulta a elaboração de novas estratégias de prevenção e de tratamento contra a leptospirose (Ko *et al.*, 2009).

Estudos *in vitro* mostram que *L. interrogans* é capaz de invadir células epiteliais e macrófagos e que sua entrada nessas células depende de receptores mediadores de endocitose (Merien, *et al.*, 1997). *L. interrogans* também induz apoptose em macrófagos através de vias dependentes de caspases 8 e 3 (Merien, *et al.*, 1997, Jin *et al.*, 2009).

Entretanto, diferentemente de *Salmonella* e *Shigella*, *L. interrogans* não parece ser um patógeno intracelular facultativo (Barochi *et al.*, 2002). Propõe-se que a translocação rápida de *L. interrogans* pelas células hospedeiras seja um mecanismo que permite à bactéria se evadir da morte pelas células hospedeiras e chegar rapidamente ao sistema sanguíneo para a disseminação para múltiplos órgãos (Barochi *et al.*, 2002).

Apesar das diversas evidências de que *L. interrogans* é uma bactéria invasiva, capaz de se aderir e de invadir células epiteliais e de induzir apoptose em macrófagos, os componentes leptospirais envolvidos nesses processos permanecem praticamente desconhecidos (Adler & de la Peña Moctezuma, 2010). A falta de conhecimento das bases moleculares da virulência de leptospiros é devida, em parte, à limitação imposta pela ausência de ferramentas genéticas disponíveis para o estudo dessas bactérias até há pouco tempo.

Os primeiros trabalhos utilizando-se genética molecular em leptospiros patogênicos começaram a ser publicados há cinco anos com a geração dos primeiros mutantes em *L. interrogans* (Bourhy *et al.*, 2005). O primeiro fator de virulência definido geneticamente em leptospiros foi a lipoproteína de superfície Loa22, cuja função molecular permanece desconhecida (Ristow *et al.*, 2007).

A partir de então, importantes avanços na compreensão da biologia e da patogenicidade leptospiral tem sido alcançados, auxiliados pelo recente seqüenciamento genômico de *Leptospira* spp. (revisão em Ko *et al.*, 2009). Nos últimos anos, foram identificadas proteínas de *Leptospira interrogans* que se ligam a componentes da matriz

extracelular, ao fator H regulador do complemento e ao plasminogênio humano (Barbosa *et al.*, 2006; Atzingen *et al.*, 2008; Longhi *et al.*, 2009; Vieira *et al.*, 2010; Verma *et al.*, 2006; Stevenson *et al.*, 2007; Verma *et al.*, 2010). Cadeias de glicosaminoglicanos presentes em proteoglicanos expostos na superfície de células humanas foram mostradas como sendo receptores para *L. interrogans* (Breiner *et al.*, 2009).

Diversas hemolisinas, para as quais se sugere um papel citotóxico, foram identificadas a partir do genoma de *L. interrogans* (Nascimento *et al.*, 2004b). A atividade citotóxica e o provável efeito apoptótico de uma delas, Sph2, foram recentemente descritos (Zhang *et al.*, 2008). Interessantemente, as hemolisinas já eram sugeridas como fatores de virulência muito antes do seqüenciamento genômico (Kasarov, 1970; Thompson & Manktelow, 1986). Em 2002, mostrou-se que o mecanismo citotóxico da hemolisina SphH envolve a formação de poros nas células de mamíferos (Lee, *et al.*, 2002).

Como referido acima, diversas proteínas envolvidas na adesão e na sobrevivência *in vivo* de leptospiros tem sido identificadas. Porém, o conhecimento que se tem atualmente sobre a biologia e a patogenicidade de leptospiros é exíguo se comparado ao conhecimento de patógenos bacterianos para os quais ferramentas de manipulação genética são disponíveis há mais de 25 anos (Adler & de la Peña Moctezuma, 2010).

1.2. SEQUENCIAMENTO GENÔMICO DE *Leptospira* spp.

Em humanos, a forma severa da leptospirose é frequentemente causada por linhagens do sorogrupo Icterohaemorrhagiae. As linhagens responsáveis pela doença dependem amplamente da localização geográfica e da ecologia dos hospedeiros reservatórios locais. Assim, na Europa, as sorovares Copenhageni e Icterohaemorrhagiae, carregadas por ratos, são predominantes enquanto que no sudeste asiático a sorovar Lai é a mais frequente (Levett, 2001).

No início dos anos 2000, os genomas das sorovares Lai e Copenhageni foram seqüenciados (Ren *et al.*, 2003; Nascimento *et al.*, 2004b). O genoma de *Leptospira interrogans* sorovar Lai (4691184 pb) é muito maior que o de outras duas espiroquetas patogênicas – *Treponema pallidum* (1138006 pb), responsável pela sífilis (Fraser *et al.*, 1998) e *Borrelia burgdorferi* (1519857 pb), responsável pela doença de Lyme (Fraser *et al.*,

1997). Ele é constituído de dois cromossomos circulares, um grande de 4332241 pb (CI) e um pequeno de 358943 pb (CII) (Ren *et al.*, 2003). O genoma da sorovar Copenhageni apresenta-se na mesma configuração, possuindo apenas 63818 pb a menos que o genoma da sorovar Lai (Nascimento *et al.*, 2004b; Ren *et al.*, 2003). A porcentagem média de identidade de nucleotídeos entre os dois genomas é de 95 % e a identidade média de nucleotídeos entre pares de genes ortólogos preditos, codificantes de proteínas, é de 99 % (Nascimento *et al.*, 2004a).

Recentemente, o genoma da espécie patogênica *Leptospira borgpetersenii* e da espécie saprófita *Leptospira biflexa* também foram sequenciados. O genoma de *L. borgpetersenii* é 700 Kb menor que o genoma de *L. interrogans*. Acredita-se que essa redução esteja associada a um processo evolutivo em direção a um ciclo de vida restrito aos hospedeiros (Bulach *et al.*, 2006). No genoma de *L. biflexa*, cerca de um terço dos genes não apresentam ortólogos nos genomas das espécies patogênicas (Picardeau *et al.*, 2008).

O seqüenciamento genômico de *Leptospira* spp. abriu caminho para a elucidação de novos aspectos da fisiologia dessas bactérias relacionados ao metabolismo energético, adaptação, sobrevivência em diferentes ambientes, colonização e mecanismos de patogenicidade. Além disso, ele possibilitou a identificação de proteínas potencialmente localizadas na superfície celular que podem servir como alvos para o desenvolvimento de vacinas de subunidades, além de poderem ser úteis para o diagnóstico sorológico da doença (Gamberini *et al.*, 2005).

Analisando-se a lista de produtos gênicos preditos pelo projeto genoma de *L. interrogans* sorovar Copenhageni no banco de dados do projeto (aeg.lbi.ic.unicamp.br/world/lic/), observa-se que grande parte daqueles que são potencialmente exportados para o envelope celular, incluindo lipoproteínas e proteínas transmembrânicas, não apresentam função molecular conhecida (Nascimento *et al.*, 2004a).

De uma maneira geral, proteínas transmembrânicas ou lipoproteínas expostas na superfície celular, podem estar relacionadas a processos de adesão, colonização e adaptação do patógeno (Haake, 2000). As lipoproteínas, em especial, são reconhecidas como as proteínas mais abundantes em várias espiroquetas, o que indica sua importância para essas bactérias (Haake, 2000). Além disso, proteínas expostas na superfície também podem servir como alvo do sistema imune do hospedeiro, sendo, por isso, alvos potenciais para o

desenvolvimento de vacinas (Pizza *et al.*, 2000, Gamberini *et al.*, 2005). Desse modo, a caracterização estrutural e funcional dessas proteínas deve contribuir para a ampliação do conhecimento dos mecanismos moleculares que regem a fisiologia de *Leptospira interrogans*, podendo ter implicações na elucidação dos mecanismos moleculares envolvidos na interação entre este patógeno e seus hospedeiros.

1.3. PREDIÇÃO DE FUNÇÃO DE PROTEÍNAS

Na ausência de dados experimentais, a função de uma proteína é geralmente inferida com base em sua similaridade de seqüência com uma proteína de função conhecida (Watson *et al.*, 2005). Entretanto, em muitos genomas seqüenciados, cerca de 30 % a 50 % dos produtos gênicos preditos possuem uma baixa identidade de seqüência (< 30 %) com proteínas funcionalmente caracterizadas (Yakunin *et al.*, 2004).

A estrutura tridimensional de uma proteína pode revelar vários indícios funcionais, que são obtidos, basicamente, por similaridades estruturais ou pela presença de ligantes na estrutura (Zhang & Kim, 2003). Zarembinski e colaboradores foram os primeiros a propor a função molecular de uma proteína hipotética a partir de sua estrutura cristalográfica (Zarembinski *et al.*, 1998). Desde então, vários trabalhos similares tem sido realizados (Hwang *et al.*, 1999; Boggon *et al.*, 1999, Roberts *et al.*, 2005, Sieminska *et al.*, 2007)

Sendo assim, nos últimos anos, muito tem se investido no desenvolvimento de métodos para se extrair indícios funcionais diretamente da estrutura tridimensional de uma proteína. Esses métodos vão desde a busca por estruturas de enovelamento similar até a identificação de agrupamentos tridimensionais de resíduos funcionais (Watson *et al.*, 2005), sendo crescente o número de proteínas cujas funções têm sido propostas a partir de informações estruturais (Yakunin *et al.*, 2004).

Além de prover importantes indícios funcionais, que podem ser experimentalmente testados, a determinação da estrutura tridimensional de proteínas também tem importantes implicações no estabelecimento de relações evolutivas e na determinação de mecanismos de interações entre domínios protéicos, sendo, portanto, de grande importância para o entendimento de como os vários atributos moleculares de uma célula se combinam para produzir suas propriedades anatômicas e fisiológicas particulares (Teichmann *et al.*, 2001).

1.4. ALVOS DO PROJETO

Dentro do contexto exposto acima, o objetivo desse projeto foi a obtenção de indícios funcionais a partir da resolução da estrutura cristalográfica de três proteínas de *L. interrogans* sorovar Copenhageni potencialmente localizadas no envelope celular (TABELA 1.1).

TABELA 1.1. Características das proteínas alvo deste projeto. código: identificação da proteína no banco de dados de *L. interrogans* sorovar Copenhageni (aeg.lbi.ic.unicamp.br/world/lic/). n° de aa.: número de resíduos de aminoácidos. M: massa molecular calculada pelo programa ProtParam (Gasteiger *et al.*, 2005). pI: ponto isoelétrico teórico calculado pelo programa ProtParam (Gasteiger *et al.*, 2005). descrição: anotação das proteínas alvo pelo projeto genoma. ps: presença de peptídeo sinal predito pelo programa signalP-HMM (Nielsen *et al.*, 1997; Nielsen & Krogh, 1998). localização celular: localização celular predita pelo programa PSORT (Nakai & Horton, 1999) ou PSORTB v.2.0.4[§] (Gardy *et al.*, 2005).

código	n° de aa.	M (kDa)	pI	descrição	ps	localização celular
LIC10793	456	51.8	8.9	antígeno de superfície	✓	membrana externa [§]
LIC12922	362	42.3	6.3	hipotética conservada	✓	periplasma
LIC10494	231	25.8	8.5	provável lipoproteína	✓	membrana interna/externa

Tais proteínas foram selecionadas a partir de um grupo de interesse do Centro de Biotecnologia do Instituto Butantan, São Paulo. Os critérios de seleção adotados foram a expressão solúvel em *E. coli*, baseada em experimentos feitos anteriormente, e a ausência de homólogos com estrutura e função conhecidas, de acordo com buscas realizadas pelo programa BLAST (Altschul *et al.*, 1997).

Ensaio de *Western blot* realizados no Centro de Biotecnologia do Instituto Butantan com várias proteínas, inclusive as três selecionadas nesse projeto, indicaram que estas são reconhecidas por anticorpos presentes em soro de pacientes diagnosticados com a doença (Ana Lúcia T.O. Nascimento, *comunicação pessoal*).

Após o início deste trabalho, confirmou-se que a proteína LIC10793, renomeada como Lp49, é antigênica e propôs-se sua aplicação no desenvolvimento de testes sorodiagnostics (Neves *et al.*, 2007). Nesse mesmo trabalho, evidenciou-se a provável

localização de Lp49 na membrana externa e mostrou-se que o gene LIC10793 está presente em várias sorovares patogênicas, mas é ausente em *L. biflexa* (Neves *et al.*, 2007).

A busca por indícios funcionais a partir da resolução da estrutura cristalina das três proteínas alvo torna-se interessante pelo fato de que estas não apresentam similaridades significativas com famílias funcionalmente caracterizadas ou com proteínas de estrutura conhecida. Dessa forma, a determinação de suas estruturas tridimensionais, bem como os indícios funcionais obtidos a partir delas, pode significar a caracterização de novas famílias, contribuindo para a elucidação da estrutura e função de proteínas homólogas.

2. OBJETIVOS

O presente trabalho visou à obtenção de indícios funcionais das proteínas LIC10793, LIC12922 e LIC10494 de *Leptospira interrogans* sorovar Copenhageni a partir de suas estruturas tridimensionais. Para tanto, o plano de trabalho adotado incluiu a expressão, purificação, cristalização e resolução da estrutura cristalográfica por difração de raios X das três proteínas alvo.

3. RESULTADOS

Os resultados deste trabalho estão apresentados em três partes. A primeira delas consiste em um artigo já publicado, contendo a determinação e análise da estrutura cristalográfica de LIC10793, recentemente renomeada como Lp49. O segundo capítulo corresponde a um manuscrito, submetido para publicação, onde se analisa a estrutura cristalográfica de LIC12922. A terceira parte descreve a expressão, purificação, cristalização e caracterização de LIC10494. No que se refere ao artigo publicado e ao artigo submetido para a publicação, minha participação nos trabalhos inclui a redação dos manuscritos e todas as etapas experimentais com exceção da etapa de clonagem dos alvos nos vetores pDEST ou pAE (Ramos *et al.*, 2004), efetuada pelo grupo da Dra. Ana Lúcia T. O. Nascimento do Centro de Biotecnologia do Instituto Butantan (SP), e da construção de alguns dos variantes do gene LIC12922 em vetor pET-TEV, realizada por Tereza Cristina Lima Silva.

3.1. ARTIGO I

**The leptospiral antigen Lp49 is a two-domain protein with
putative protein binding function**

Priscila O. Giuseppe, Fernanda O. Neves, Ana Lúcia T.O. Nascimento and Beatriz G.
Guimarães



The leptospiral antigen Lp49 is a two-domain protein with putative protein binding function [☆]

Priscila Oliveira Giuseppe ^a, Fernanda Oliveira Neves ^{b,c}, Ana Lúcia T.O. Nascimento ^{b,c},
Beatriz Gomes Guimarães ^{a,*}

^a Centro de Biologia Molecular Estrutural, Laboratório Nacional de Luz Síncrotron (LNLS), Rua Giuseppe Máximo Scolfaro 10000, PO Box 6192, Campinas 13083-970, SP, Brazil

^b Centro de Biotecnologia, Instituto Butantan, São Paulo, SP, Brazil

^c Interunidades em Biotecnologia, Instituto de Ciências Biomédicas, USP, São Paulo, Brazil

ARTICLE INFO

Article history:

Received 23 January 2008
Received in revised form 7 April 2008
Accepted 8 April 2008
Available online 14 April 2008

Keywords:

Leptospira interrogans
Leptospirosis
Leptospiral antigen
Lp49
Crystal structure
LIC10793
Immunoglobulin-like domain
7-Bladed beta-propeller domain

ABSTRACT

Pathogenic *Leptospira* is the etiological agent of leptospirosis, a life-threatening disease that affects populations worldwide. Currently available vaccines have limited effectiveness and therapeutic interventions are complicated by the difficulty in making an early diagnosis of leptospirosis. The genome of *Leptospira interrogans* was recently sequenced and comparative genomic analysis contributed to the identification of surface antigens, potential candidates for development of new vaccines and serodiagnosis. Lp49 is a membrane-associated protein recognized by antibodies present in sera from early and convalescent phases of leptospirosis patients. Its crystal structure was determined by single-wavelength anomalous diffraction using selenomethionine-labelled crystals and refined at 2.0 Å resolution. Lp49 is composed of two domains and belongs to the all-beta-proteins class. The N-terminal domain folds in an immunoglobulin-like beta-sandwich structure, whereas the C-terminal domain presents a seven-bladed beta-propeller fold. Structural analysis of Lp49 indicates putative protein–protein binding sites, suggesting a role in *Leptospira*–host interaction. This is the first crystal structure of a leptospiral antigen described to date.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Leptospirosis is a re-emerging infectious disease caused by spirochaetes of the genus *Leptospira*. The disease is more prevalent in tropical and subtropical regions and, in developing countries, it produces large urban epidemics with mortality mainly during the rainy season (Bharti et al., 2003). The incidence of leptospirosis remains underestimated due to the broad spectrum of symptoms and similarities with other febrile illnesses. Five to 15% of the cases evolve to renal and hepatic failure with mortality rate of 4–40% (Faine et al., 1999; Plank and Dean, 2000). In addition, leptospirosis has a major impact in the agricultural industry since it affects the livestock causing abortions, stillbirths, infertility, reduced milk production and death (Faine et al., 1999).

The large number of pathogenic *Leptospira* serovars (>230) (Bharti et al., 2003) imposes a major limitation to the development of a multi-serovar vaccine based on inactivated whole cells or membrane preparations of pathogenic leptospires. Currently, the

available vaccines neither induce long-term protection against infection nor provide cross-protective immunity against leptospiral serovars not included in the vaccine preparation. The complete genome sequencing of *Leptospira interrogans* serovar Lai (Ren et al., 2003) and serovar Copenhageni (Nascimento et al., 2004a) contributes to the understanding of the biology of this versatile pathogen and to the development of new prophylactic interventions. Comparative genomics of the two serovars allowed the identification of a large number of exported lipoproteins and other membrane-associated proteins which may be involved in leptospiral protective immunity (Nascimento et al., 2004b). Outer membrane proteins that are conserved among pathogenic serovars might be useful for leptospirosis diagnosis as well as in the production of a recombinant vaccine (Gamberini et al., 2005).

LIC10793 encodes for a 49 kDa membrane-associated protein (Lp49) from *L. interrogans* serovar Copenhageni. Lp49 is predicted to be a lipoprotein by LipoP web server (Juncker et al., 2003) and SpLip algorithm (Setúbal et al., 2006) (Fig. 1A) but no other functional domain was identified based on its primary structure. Recent studies showed that Lp49 is recognized by antibodies present in early and convalescent sera of leptospirosis patients (Neves et al., 2007). Lp49 encoding sequence is conserved among several pathogenic serovars of *L. interrogans*, suggesting its use as antigenic mar-

[☆] Data deposition: The atomic coordinates of Lp49 have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 3BWS).

* Corresponding author. Fax: +55 19 3512 1006.

E-mail address: beatriz@lnls.br (B.G. Guimarães).

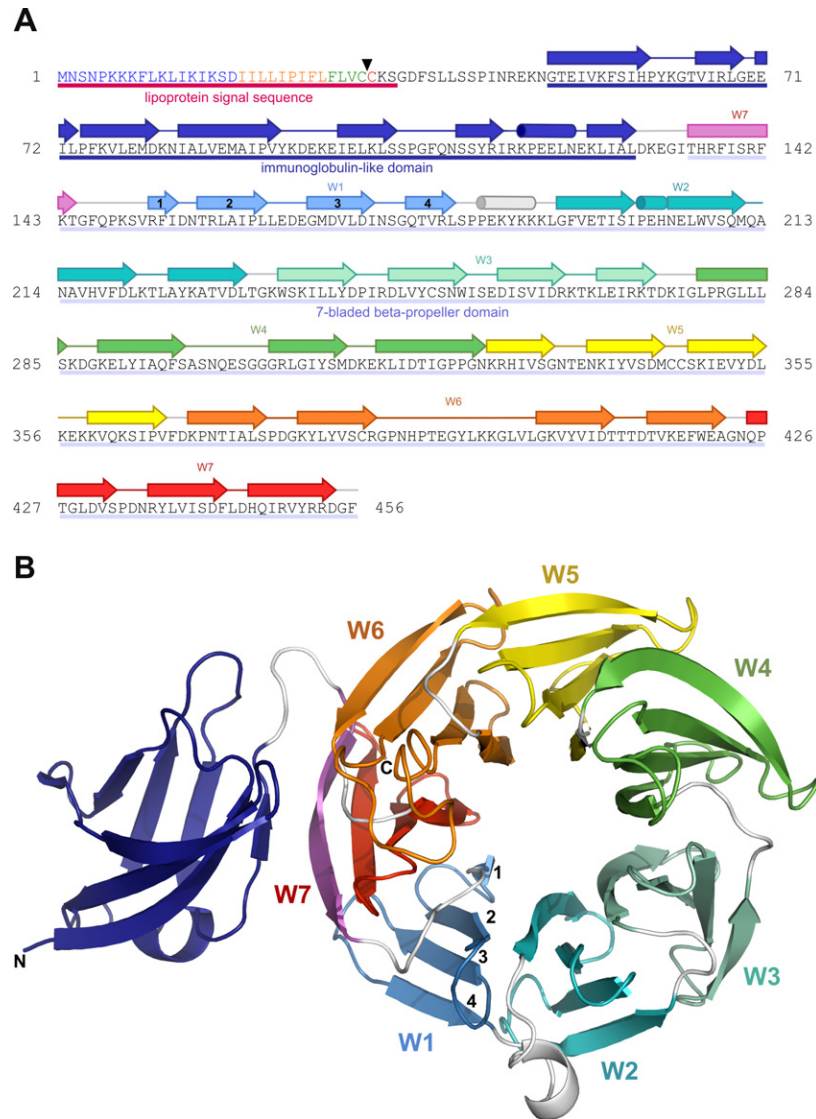


Fig. 1. Overall structure of *L. interrogans* Lp49. (A) Primary structure of Lp49. The three domain boundaries are indicated below the amino acids sequence. The predicted signal peptide from amino acid 1–34 (Schultz et al., 1998; Letunic et al., 2006) shows the typical features of spirochaetal lipoprotein signal sequences (Setúbal et al., 2006), which are composed of an amino-terminal region with a net positive charge (blue), a hydrophobic region (orange) and the characteristic carboxy-terminal region (green) consisting of I, L, F, or V at either position –3 or –4 relative to the predicted cleavage site (black arrow). The predicted lipidated amino-terminal cysteine responsible for anchoring the mature protein to a lipid bilayer is indicated in red. Secondary structures obtained from the crystallographic data are represented at the top of the sequence and colored according to (B). (B) Cartoon representation of Lp49 structure. Molecule orientation shows the β -propeller top face. The β -propeller blades are shown in different colors and W1 strands are numbered. N- and C- termini are indicated.

ker protein for detecting a broad spectrum of *Leptospira* infection. In addition, its absence from non pathogenic strain suggests a possible role of Lp49 in leptospiral pathogenicity (Neves et al., 2007). In this work we present the crystal structure of Lp49 solved at 2.0 Å resolution. Lp49 structure reveals a two-domain protein with putative protein binding function. To our knowledge this is the first three-dimensional structure of a leptospiral antigen protein reported.

2. Materials and methods

2.1. Protein expression and purification

Cloning of truncated LIC10793 (100–1371 bp) gene encoding for Lp49 protein without the signal peptide sequence was described previously (Neves et al., 2007). Preparation of selenomethi-

onine (SeMet) labeled protein was performed based on the metabolic inhibition method proposed by Van Duyne et al. (1993). *Escherichia coli* C41(DE3) competent cells were transformed with the recombinant expression vector. A freshly transformed single colony was grown for 16 h (310 K, 200 rpm) in 30 mL Luria Bertani medium supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin. Cells were harvested by centrifugation (300g, 2 min, 293 K) and resuspended in 30 mL M9 minimal medium (6.8 mg mL⁻¹ Na₂HPO₄, 3 mg mL⁻¹ KH₂PO₄, 0.5 mg mL⁻¹ NaCl, 2 mM MgSO₄, 2 mg mL⁻¹ NH₄Cl, 0.2 mM CaCl₂, 1% glucose) supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin, 10 $\mu\text{g mL}^{-1}$ thiamine and 50 μM FeCl₃ (M9^{supl}). The suspended cells were inoculated into 1 L M9^{supl} medium and grown at 310 K, 200 rpm until the optical density of the culture reached 0.6 at 600 nm. The culture was then supplemented with 10 mL of a sterile mixture containing lysine, phenylalanine and threonine (10 mg mL⁻¹), isoleucine, leucine and valine (5 mg mL⁻¹) and selenomethionine (6 mg mL⁻¹) diluted

into M9 medium without glucose. After 15 min of incubation (310 K, 200 rpm), expression was induced by addition of isopropyl-1-thio- β -galactopyranoside (IPTG) to a final concentration of 0.4 mM. Cells were grown for a further 4 h at 303 K, harvested by centrifugation at 3000g for 30 min and then frozen at 253 K.

For His₆-LIC10793(SeI₃₄-Phe₄₅₆) fusion protein purification, the frozen pellet was resuspended in 40 mL of buffer A (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 5 mM Imidazole) with 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were disrupted by sonication and the lysate was clarified by centrifugation (20,000g, 30 min, 277 K). For removal of nucleic acids, the lysate was incubated with 2% w v⁻¹ streptomycin sulfate while shaking (45 min, 277 K) and centrifuged (20,000g, 30 min, 277 K). The supernatant was then injected in a 5 mL HiTrap™ Chelating HP column (GE Healthcare Life Sciences) pre-charged with Ni₂SO₄ using an ÄKTA-FPLC system set at a flow rate of 2 mL min⁻¹. After washing the resin with 20 column volumes (CV) of buffer A, bound proteins were eluted using a 5 mM to 1 M imidazole gradient divided in 3 parts: a 1 CV linear gradient from 0% to 5% buffer B (buffer A + 1 M imidazole), a 2 CV step of 5% buffer B and a 20 CV linear gradient from 5% to 100% buffer B. Fractions were subjected to SDS-PAGE and those containing the target protein were pooled and dialyzed against buffer C (20 mM sodium citrate pH 6, 20 mM NaCl). The sample was then centrifuged (18,000g, 10 min, 277 K) and loaded onto a 5 mL HiTrap™ SP HP cation exchange column (GE Healthcare Life Sciences) via an ÄKTA-FPLC system at 1.5 mL min⁻¹. After washing with 4 CV of buffer C, the target protein was eluted from the resin using a 20 CV linear gradient from 20 mM to 1 M NaCl. The purified protein was dialyzed against buffer C and concentrated using an Amicon Ultra-15 10,000 MW filter (Millipore). Protein concentration was estimated from direct absorbance at 280 nm considering a predicted extinction molar coefficient of 47,580 M⁻¹ cm⁻¹.

2.2. Protein crystallization

SeMet-labeled protein was crystallized by vapor diffusion in hanging drops. Initial screening was performed by equilibrating a mixture of 1 μ L of protein solution (6 mg mL⁻¹) and 1 μ L of reservoir solution (Hampton Research Crystal Screen I) over 500 μ L of reservoir solution at 291 K. Small and clustered crystals were observed in the condition containing 0.2 M LiSO₄, 0.1 M Tris-HCl pH 8.5, 30% polyethylene glycol (PEG) 4000 after 24 h. Attempting to obtain suitable crystals for diffraction experiments, the microseeding technique (Stura and Wilson, 1990) was then applied. For this purpose, clustered crystals were transferred to an Eppendorf tube containing seed bead (Hampton Research) and 50 μ L of the crystallization condition. The tube was vortexed for 3 min to generate a seed stock. This stock was diluted 1000-fold and 0.3 μ L of the dilution was added to drops composed by 1.5 μ L of protein solution (6 mg mL⁻¹) and 1.5 μ L of reservoir solution pre-equilibrated over 500 μ L of reservoir solution (0.2 M LiSO₄, 0.1 M Tris-HCl pH 8.5 and 30, 25, 20 or 15% PEG 4000) for 1 h at 291 K. High-quality single crystal plates, with two dimensions higher than 100 μ m, were obtained at the condition containing 15% PEG 4000.

2.3. Data collection and processing

SeMet labeled crystals cryo-protected by soaking in reservoir solution plus 20% PEG400 were flash-cooled in a 100 K nitrogen-gas stream. X-ray diffraction data were collected at the protein crystallography W01B-MX2 beam line of the Brazilian Synchrotron Light Laboratory (LNLS), Campinas, Brazil. Anomalous scattering factors were derived from the emission fluorescence spectrum of the crystal using the program CHOOCH (Evans and Pettifer,

2001). Then, diffraction data were collected at a single wavelength (0.9795 Å) corresponding to the maximum of f'' . The crystal was exposed for 1 min per 0.7° rotation of ϕ with the distance to detector set to 170 mm. Data were recorded on a Marmosaic-225 CCD detector (MarUSA) and indexed using the program MOSFLM (Leslie, 1992). Scaling was performed by SCALA (Kabsch, 1988) from the CCP4 package (COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4, 1994). Crystal belongs to space group $P2_12_12_1$. Matthews (1968) coefficient calculation indicated the presence of two monomers in the asymmetric unit ($V_m = 2.23 \text{ \AA}^3 \text{ Da}^{-1}$) and a solvent content of 45% Table 1 summarizes data collection statistics.

2.4. Structure solution and refinement

The crystal structure of Lp49 was solved by the Single-wavelength Anomalous Diffraction (SAD) method. All 14 expected selenium were located by the Hybrid Sub-structure Search method (Grosse-Kunstleve and Adams, 2003) and initial phasing was calculated with Phaser (McCoy et al., 2007). Density modification was carried out with RESOLVE (Terwilliger, 2002) which was also used to build and refine a preliminary model. Programs mentioned above were used through the AutoSol Wizard of the PHENIX package (Adams et al., 2007). Further automatic model building was carried out with ARP/wARP (Lamzin et al., 2001). Model refinement was performed alternating cycles of REFMAC5 (Murshudov et al., 1997) with visual inspection of the electron density maps and manual rebuilding with COOT (Emsley and Cowtan, 2004). A total of 814 residues (out of the 864 expected for 2 monomers) were modelled, comprising Lp49 residues from Gly50 to Phe456. Water molecules were added using COOT (Emsley and Cowtan, 2004) during the last refinement cycles. Three sulfate ions were modelled based on the difference Fourier map. The final model was estab-

Table 1
Data Collection and Refinement Statistics

<i>Data statistics^a</i>	
Space group	$P2_12_12_1$
Unit cell	$a = 78.95 \text{ \AA}$, $b = 86.87 \text{ \AA}$, $c = 132.42 \text{ \AA}$
Wavelength (Å)	0.9795
Resolution range (Å)	53.45–1.99
No. of unique reflections	60735 (8497)
Redundancy ^b	6.1 (5.4)
Completeness ^b (%)	95.9 (94.0)
$\langle I/\sigma(I) \rangle$	19.3 (5.3)
$R_{\text{r.i.m.}}$ ^{b,c} (%)	0.126 (0.421)
<i>Refinement statistics</i>	
Resolution range (Å)	20.0–1.99
No. of reflections	57517
R_{work} ^d	0.185
R_{free} ^d	0.237
R.m.s.d. bond length (Å)	0.010
R.m.s.d. bond angle (°)	1.256
Mean B value (Å ²)	19.84
Mean B value (solvent) (Å ²)	27.74
Protein atoms	6438
Water molecules	602
SO ₄ molecules	3
<i>Residues in Ramachandran plot regions^e (%)</i>	
Most favored	86.4
Additional allowed	12.3
Generously allowed	1.0
Disallowed	0.3

^a Numbers in parentheses are for the highest resolution bin, 2.10–1.99 Å.

^b Considering bijvoet pairs to be independent reflections.

^c $R_{\text{r.i.m.}} = \sum_h \left(\frac{n_h}{n_h - 1} \right)^{1/2} \sum_i |I_{hi} - \bar{I}_h| / \sum_h \sum_i \bar{I}_h$.

^d $R_{\text{work}} = \sum_h |F_o - F_c| / \sum_h |F_o|$, R_{free} (test set = 5%).

^e Procheck (Laskowski et al., 1993).

lished after the convergence of R_{factor} and R_{free} values to 0.19 and 0.24, respectively. Stereochemistry of the model was analyzed with PROCHECK (Laskowski et al., 1993) and Molprobity (Davis et al., 2007). Residue Lys82 of both monomers falls in the disallowed region of the PROCHECK Ramachandran plot. However, electron density maps unambiguously support the model. Lys82 is located at a type II' beta-turn (Hutchinson and Thornton, 1994) and its side chain is hydrogen-bound to the Asp81 carboxyl group. No outlier was observed in the Molprobity Ramachandran plot and 95.6% of the residues are in its favored region. Additional refinement details are summarized in Table 1. Figures were produced using the program Pymol (DeLano, 2002) and TopDraw (Bond, 2003).

3. Results and discussion

3.1. Overall structure

SeMet-labeled protein Lp49 (Ser34–Phe456) was produced in *E. coli*, purified in two steps and crystallized. The crystal structure was solved by single-wavelength anomalous diffraction and refined at 2.0 Å resolution to a final R_{factor} of 0.19 ($R_{\text{free}} = 0.24$). The final atomic model includes two symmetrically independent monomers (referred as A and B) and quality of the electron density maps allowed modelling of amino acid residues 50–456 in both monomers. The side chain of a few solvent exposed glutamic acids and lysines were poorly defined in the electron density maps and were modelled as alanines. The asymmetric unit dimer interface buries approximately 590 Å² of solvent-accessible area per molecule (5% of its total surface) and analyses of monomer–monomer interfaces using the PISA web server (Krissinel and Henrick, 2007) did not indicate any biologically relevant dimer in the crystal. Superposition of the two crystallographically independent monomers resulted in an overall r.m.s. deviation of 0.28 Å for all C α atoms. The following analysis is performed using the monomer A.

Lp49 structure is composed of two domains, both belonging to the all β -proteins SCOP class (Lo Conte et al., 2000) (Fig. 1B). The N-terminal domain (residues 50–129) is a β -sandwich formed by two four-stranded β -sheets arranged face-to-face. It is 58 Å long, with an approximate diameter of 30 Å and is tightly attached to the C-terminal domain by a type II β -turn motif (Hutchinson and Thornton, 1994). The C-terminal domain (residues 135–456) is a β -propeller composed of 7 blades packed radially around a central pseudo-7-fold symmetry axis. Each blade consists of a β -sheet formed by four-twisted antiparallel β -strands. In the last blade (W7), ring closure is achieved by a typical molecular Velcro formed by a combination of three β -strands from the C-terminal (strands 1, 2 and 3) and one β -strand from the N-terminal of the propeller (strand 4) (Fig. 1). Structural superposition of the blades W2–W7 on blade W1 results in r.m.s. deviations of 1.75, 1.70, 1.43, 1.33, 1.55 and 1.79 Å, respectively. The β -propeller bottom face, defined by the loops connecting β -strands 1–2 and 3–4, has a wider diameter (75 Å) than the top face (42 Å), defined by the loops connecting β -strands 4–1 and 2–3. The first strand of each blade lines a central cavity which is occluded near the top surface entrance by residues mainly polar or positively charged (discussion below).

3.2. The atypical N-terminal domain topology

The Lp49 N-terminal domain has an immunoglobulin-like fold. Its topology shows a five-stranded structural unit ($\beta_1, 2, 4, 5$ and 6) corresponding to a “complete” Greek key motif (Ptitsyn and Finkelstein, 1980) (Fig. 2A). Two forms of the standard Greek key motif have been described (Zhang and Kim, 2000). In both forms (called A and B), β -strands named a_0, b_0 and c_0, d_0 define, respectively, two

β -arches (Hazes and Hol, 1992) (Fig. 2B). The difference between the two forms relies on the position of the middle strand, named b_1 when it is hydrogen-bound to b_0 (form A) or c_1 , when it is hydrogen-bound to c_0 (form B). The Lp49N-terminal domain shows an atypical variation of the Greek key motif, with the middle strand apparently splitting in two, b_1 and c_1 (Fig. 2B). Structural comparisons performed with SSM (Krissinel and Henrick, 2004) service indicated that the Lp49 N-terminal domain is most similar to structures which present the form B of the Greek key motif. It suggests that, in the Lp49 N-terminal domain, the original five-stranded unit $a_0b_0c_1c_0d_0$ was altered by insertion of the strand b_1 . Besides this variation, the Lp49 Greek key motif presents a C-terminal extension, comprising the strand β_7 , a 3_{10} -helix (h) and the strand β_8 . β_7 and β_8 are labeled d_1 and d_2 , respectively, according to their proximity to the strand d_0 in the polypeptide chain (Zhang and Kim, 2000). Thus, the Lp49 Ig-like β -sandwich domain topology can be defined as $d_2a_0c_0c_1b_1b_0d_0d_1$ (Zhang and Kim, 2000).

Structural comparisons of the Lp49 N-terminal atomic coordinates with the contents of the Protein Data Bank using the SSM (Krissinel and Henrick, 2004) and Dali (Holm and Sander, 1995) services identified distantly related structures. The most relevant hits are the 20th domain of the human filamin-b (pdb code 2e9i; r.m.s.d. of 2.3 Å for 62 C α aligned) and the mouse β_2 -microglobulin (Zhang et al., 1992) (pdb code 2mha; r.m.s.d. of 3.0 Å for 62 C α aligned). Following the SCOP classification (Lo Conte et al., 2000), β_2 -microglobulin belongs to the immunoglobulin superfamily whereas the filamin-b domain is suggested to belong to the E-set domains superfamily by the program fastSCOP (Tung and Yang, 2007). The E-set domains are “early” Ig-like fold families possibly related to the immunoglobulin superfamily. Comparison with the current content of protein sequence databases shows that Lp49 N-terminal domain does not share significant sequence similarity with other proteins, preventing the inference of homology relationships with members of the above mentioned superfamilies. Since structural comparisons also did not allow its unambiguous classification into the currently defined superfamilies, we suggest that the Lp49 β -sandwich domain may be the founder of a new protein superfamily.

3.3. The β -propeller domain

Lp49 C-terminal domain folds in a 7-bladed β -propeller structure and atomic coordinates comparison using the SSM (Krissinel and Henrick, 2004) service identified significant structural similarity to the YVTN β -propeller domain of an archaeal surface layer protein (SLP) (Jing et al., 2002) (pdb code 110q) and to members of the WD40 repeat-like superfamily. Lp49 C-terminal domain superimposes on the archaeal YVTN β -propeller with a r.m.s.d. of 1.78 Å for 267 C α atoms aligned and a sequence identity of 25%. The second hit in SSM score showed that Lp49 β -propeller superimposes on the WD40 region of the human GROUCHO/TL1 protein (Pickles et al., 2002) (pdb code 1gxr) with a r.m.s.d. of 1.90 Å for 271 aligned C α atoms and sequence identity of 15%.

Structural and sequence similarities indicate that Lp49 C-terminal domain is a member of the YVTN repeat-like/Quinoprotein amine dehydrogenase superfamily. Despite the overall similarity (Fig. 3A), Lp49 β -propeller is less regular than the structurally related YVTN β -propeller. In Lp49, β -sheets superimposes to each other with a r.m.s.d. of 1.85 ± 0.07 Å and a sequence identity of $16.20 \pm 1.52\%$, whereas the YVTN β -propeller modules superimpose on the six pseudo-symmetrical positions with an r.m.s.d. of 1.08 ± 0.08 Å and a sequence identity of $46.6 \pm 7.6\%$. Furthermore, Lp49 β -propeller blades do not share highly conserved motifs, although sequence consensus is observed at some positions, mainly in the second and third β -strands (Fig. 3B). Most of the con-

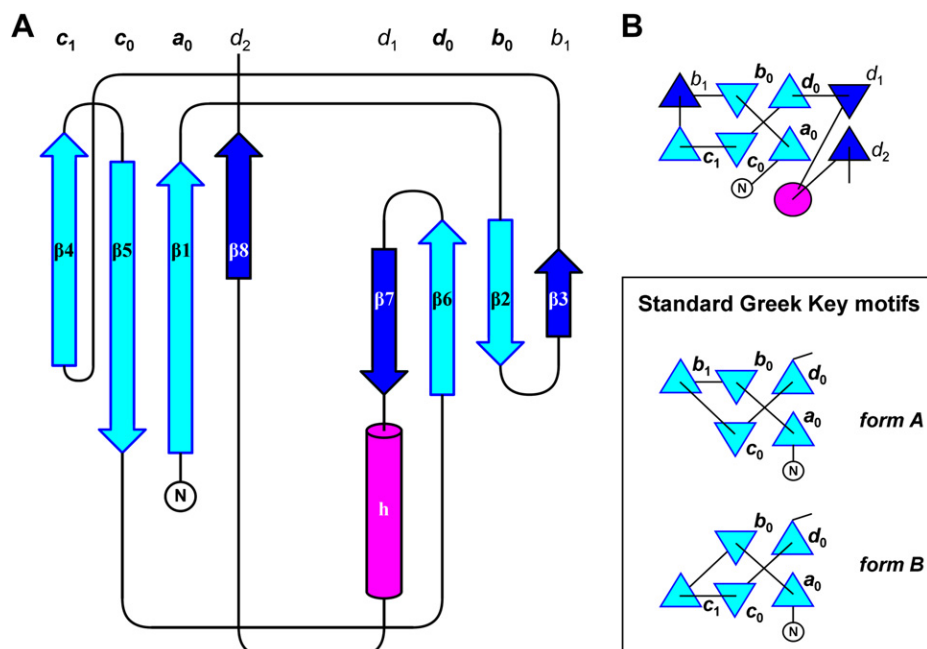


Fig. 2. The Lp49 N-terminal domain. (A) Domain topology and (B) compares Lp49 Greek key motif (colored in light blue) with forms A and B of the standard Greek key motif (Zhang and Kim, 2000). Lp49 strands $\beta 1$ to $\beta 8$ are named a_0 , b_0 , b_1 , c_1 , c_0 , d_0 , d_1 , d_2 , respectively, following the Zhang and Kim (2000) notation. Secondary structure elements not belonging to the Lp49 Greek key motif are colored in dark blue (β strands) and magenta (3_{10} helix).

served residues participate in inter β -sheets hydrophobic interactions, which stabilize the blades and their relative orientation. Besides, a conserved aspartic acid residue is located at the end of the third β -strands, corresponding to the aspartic acid of the SVIDT motif present in the YVTN β -propeller, and to the Asp residue of the Groucho/TL1 β -propeller WD repeat (Fig. 3B). As previously described, the role of this Asp residue is to stabilize a type I turn by hydrogen binding to the amide group of residue $n + 2$ (Jing et al., 2002; Lambright et al., 1996). Finally, Lp49 C-terminal domain shows long insertions between strands 2 and 3 of blades W4 and W6, respectively (Fig. 3). These insertions form prominent loops at the top side of the propeller, not observed in the Lp49 structural counterparts.

3.4. Putative protein binding function

Lp49 amino acid sequence shares no significant similarity with proteins of known function and aiming at detecting functional evidences from the three-dimensional structure, comparative structural analyses were performed. Analyses using the ProFunc service (Laskowski et al., 2005a) were carried out and the n -residue templates searches (Laskowski et al., 2005b) did not indicate the presence of known enzyme active sites, DNA binding sites or ligand binding sites. In the case of the β -sandwich domain, reverse template analyses (Laskowski et al., 2005b), which scan templates generated from the target structure against a representative set of structures in the PDB, returned no reliable matches. This result was to some extent expected since, as mentioned above, the Lp49 N-terminal domain is unrelated to other proteins present in sequence databases.

Although the searches for local similarities did not provide any hint about the Lp49 β -sandwich function, its fold suggests a role in protein binding. Ig-like domains are recurrently observed as a scaffold for interactions with proteins, for instance in the vertebrate immune system (such as antigen and cytokine receptors) (Barclay, 2003) and in cell surface adhesion molecules (such as VCAM-1, cadherin and polycystin-1) (Humphries and Newham, 1998; Ibra-

ghimov-Beskrovnyaya et al., 2000). In bacteria, Ig-like domains have already been found in virulence factors that mediate mammalian host cell adherence and invasion, such as *E. coli* intimin (Luo et al., 2000) and *Yersinia pseudotuberculosis* invasin (Hamburger et al., 1999).

Analyses of the Lp49 C-terminal domain by 3D reverse template searches (Laskowski et al., 2005b) provided corroborating evidence of its homology with the archaeal SLP YVTN β -propeller (Jing et al., 2002) and indicated local similarities with the 7-bladed WD40 (Pickles et al., 2002; Sprague et al., 2000; Tarricone et al., 2004) and the 6-bladed NHL (Edwards et al., 2003) and YWTD (Takagi et al., 2003) β -propeller domains. Sequence alignment driven by the template matches showed that most of the conserved residues shared by Lp49 β -propeller and the analogues cited above are those already mentioned as important for propeller folding and stability (Fig. 3). However, no significant similarity was detected with the characterized functional sites of the related structures, preventing a straightforward indication of the biochemical function of the Lp49 C-terminal domain.

Predicting function for a β -propeller domain is a challenging task since this fold is characterized by a remarkable functional and sequence diversity despite its structural rigidity (Fülöp and Jones, 1999). The Lp49 homologue YVTN β -propeller belongs to an archaeal surface layer protein (SLP) that has been located at cell-cell interfaces and might be involved in cell adhesion process of *Methanosarcina mazei* (Jing et al., 2002; Macario and Conway De Macario, 2001). Interestingly, the archaeal SLP is also a two-domain protein composed of a 7-bladed β -propeller and an Ig-like β -sandwich domain. Nevertheless, structural superposition of the Lp49 N-terminal domain with the SLP β -sandwich domain does not show a close match. Although a surface layer has not been observed in *L. interrogans*, the fact that at least two proteins with SLP homology (Lp49 and LIC12952, Nascimento et al., 2004a) were identified might suggest the formation of S-layer in leptospire.

The Lp49 β -propeller analogues that belong to the WD40 repeat superfamily are part of the human GROUCHO/TLE1 (Pickles et al., 2002) and yeast TUP1 (Sprague et al., 2000) transcriptional co-

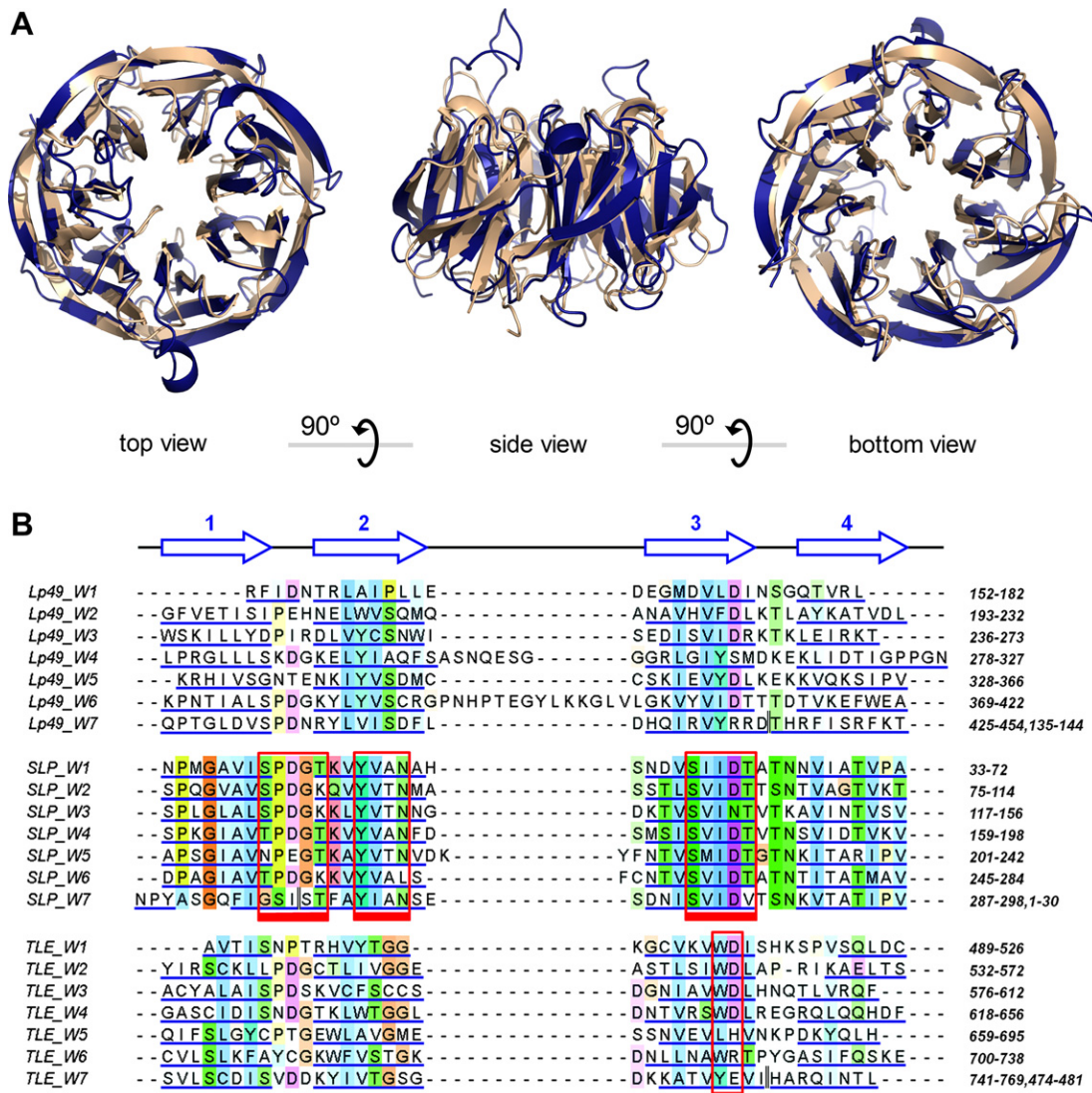


Fig. 3. Comparison of the Lp49 β -propeller domain and its structural counterparts. (A) Superposition of β -propeller structures of Lp49 (blue) and archaeal SLP (Jing et al., 2002, pdb code 110q) (light brown). (B) Alignment of the seven blades sequences of the Lp49 β -propeller domain (Lp49), the archaeal SLP YVTN β -propeller domain (SLP) and the WD40 domain of the human Groucho/TLE1 co-repressor (Pickles et al., 2002) (TLE). β -Strands residues are underlined and arrows on the top of the figure indicate the average length and position of the strands 1–4. On the right side of the figure, the number of the first and the last residue of each blade is indicated. In the last blade of each propeller (W7), adjacent N- and C-termini responsible for the “Velcro” closure are separated by two vertical lines. Residues are colored according to the conservation pattern derived from the alignment performed by Clustal W (Thompson et al., 1994). Conserved sequence motifs described for SLP (Jing et al., 2002) and TLE (Pickles et al., 2002) proteins are highlighted with red boxes.

repressors and of the human LIS1 (Tarricone et al., 2004). In all these proteins, the WD40 domains are involved in protein–protein interactions. Protein binding function has also been described for the Lp49 analogue NHL repeat domain of the protein Brat (Edwards et al., 2003), a translation regulator associated with brain tumor and for the YWTD domain of the extracellular matrix component nidogen (Takagi et al., 2003).

Lp49 structure was then analyzed using the PPI-Pred service (Bradford and Westhead, 2005) in order to verify its possible role in protein–protein interaction. PPI-Pred predicts protein–protein interaction sites based on surface properties likely to distinguish protein binding sites such as hydrophobicity, residue interface propensity, electrostatic potential, solvent accessible area and surface shape (Bradford and Westhead, 2005). Three putative protein binding sites were identified in Lp49 structure, the most probable located around the central cavity of the β -propeller top surface (Fig. 4A). Interestingly, this result agrees with the protein binding

sites described for the analogues of the Lp49 β -propeller domain which are also located at the top side of the propeller (Pickles et al., 2002; Sprague et al., 2000; Tarricone et al., 2004; Edwards et al., 2003; Takagi et al., 2003). In addition, Lp49 electrostatic surface shows that the bottom side of the β -propeller domain is mostly electronegative in contrast to the top side which reveals a central electropositive patch (Fig. 4B). This positively charged patch partially coincides with the most probable protein–protein interaction site (Fig. 4) and might have a role in protein binding. It has been found that charged amino acids can significantly enhance association rates of protein–protein complexes, even though the rate-limiting step involves the partial desolvation of charged and polar groups (Sheinerman et al., 2000).

In order to verify if the Lp49 C-terminal domain and its homologue YVTN beta-propeller (pdb code 110q) share conserved residues at the most probable protein–protein interaction site, the position of the conserved residues was analyzed. This analysis re-

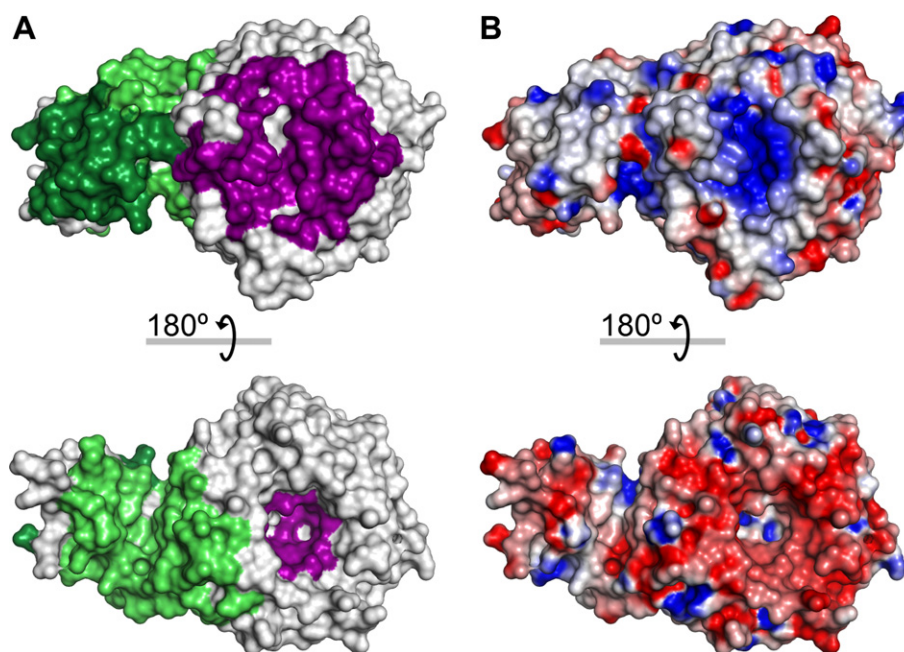


Fig. 4. Analyses of Lp49 surface. (A) Colored surface areas are predicted to be involved in protein binding according to the PPI-PRED algorithm (Bradford and Westhead, 2005). Probability of being a protein–protein interaction site increases from light to dark green to purple. (B) Electrostatic surface of Lp49 colored by charge, from red (–10 kV) to blue (+10 kV). Electrostatic potential was calculated using GRASP (Nicholls et al., 1991). On the top of the figure, molecule orientations are the same of Fig. 1, showing the top side of the β -propeller. On the bottom, the surface is rotated by 180° along the horizontal axis. The large purple area located at the β -propeller top side predicted to be involved in protein–protein interaction (top-left) includes and surrounds the positively charged central cavity (top-right).

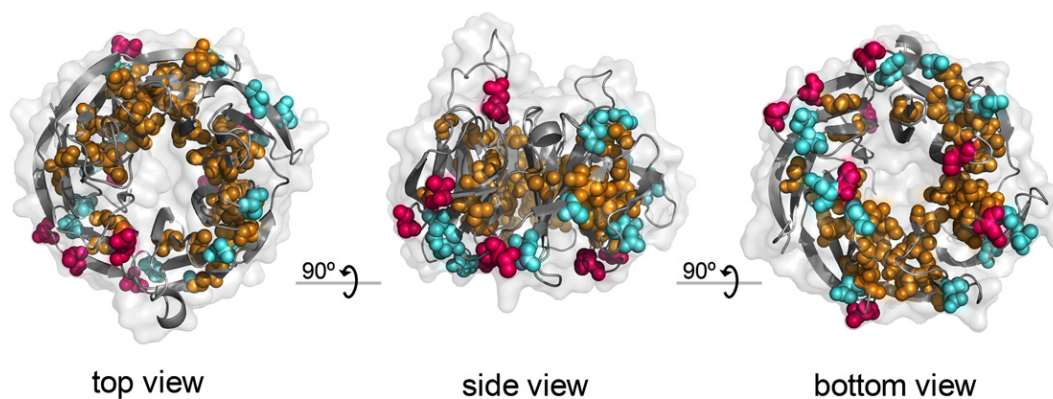


Fig. 5. Localization of the conserved residues in the Lp49 β -propeller domain. Residues conserved in both Lp49 and archaeal SLP β -propellers are represented in spheres and colored according to their exposed area: yellow indicates buried residues; cyan indicates partially exposed residues and magenta represents solvent exposed residues, according to GETAREA (Fraczkiewicz and Braun, 1998). Lp49 surface is shown in gray.

vealed that 67% of the 66 conserved residues in the β -propeller domains are buried, 21% are partially exposed and 12% are exposed to the solvent, according to the algorithm implemented in the GETAREA software (Fraczkiewicz and Braun, 1998). Conserved solvent-accessible residues are sparsely distributed in Lp49 surface (Fig. 5), giving no additional evidence of a common role of these residues in leptospiral and archaeal homologues.

4. Conclusion

In this paper, we present the first crystal structure of a leptospiral antigen. An Ig-like β -sandwich and a 7-bladed β -propeller domain accounts for the architecture of the Lp49 protein. The lack of reliable sequence and structural similarities to infer homology relationships to the Lp49 Ig-like domain suggests it might be the founder of a new protein superfamily. Domains structurally related to the Lp49 β -propeller were found in archaeal surface layer pro-

teins and in metazoan proteins, all of them involved in protein–protein interaction, despite participating in diverse biological processes such as transcription regulation and basement membrane assembly. *In silico* analysis of Lp49 structure also indicated the presence of putative protein–protein interaction sites. This result, together with the Lp49 localization at the leptospiral membrane and its role as antigen (Neves et al., 2007), suggests that Lp49 may have a protein binding function involved in *Leptospira*–host interaction. Biochemical assays aiming at the identification of possible Lp49 host cell receptors are in progress.

Acknowledgments

We would like to thank Dr. Robert Sweet, Dr. Alexei Soares and all staff of RapiData2007 course at NSLS for their assistance in preliminary data collection and processing. We also thank Celisa C. Tonoli and Andreia N. Meza for technical support. This work was

supported by FAPESP Grants CEPID/CBME 98/14138-2 and SMOI-B-Net 00/10266-8. P.O.G. is recipient of a FAPESP Ph.D. fellowship (05/04106-1).

References

- Adams, P.D., Afonine, P.V., Grosse-Kunstleve, R.W., Moriarty, N.W., Sauter, N.K., Zwart, P.H., Gopal, K., Ioerger, T.R., Kanbi, L., McKee, E., Pai, R.K., Hung, L.W., Radhakannan, T., McCoy, A.J., Read, R.J., Storoni, L.C., Romo, T.D., Sacchettini, J.C., Terwilliger, T.C., 2007. Automated structure determination with Phenix. In: Read, R.J., Sussman, J.L. (Eds.), *Evolving Methods for Macromolecular Crystallography*. Springer, Dordrecht, pp. 101–109.
- Barclay, A.N., 2003. Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecules. *Semin. Immunol.* 15, 215–223.
- Bharti, A.R., Nally, J.E., Ricaldi, J.N., Matthias, M.A., Diaz, M.M., Lovett, M.A., Levett, P.N., Gilman, R.H., Willig, M.R., Gotuzzo, E., Vinetz, J.M., Peru-United States Leptospirosis Consortium., 2003. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* 3, 757–771.
- Bond, C.S., 2003. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19, 311–312.
- Bradford, J.R., Westhead, D.R., 2005. Improved prediction of protein–protein binding sites using a support vector machines approach. *Bioinformatics* 21, 1487–1494.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35 (Web Server Issue), W375–W383.
- DeLano, W.L., 2002. The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>.
- Edwards, T.A., Wilkinson, B.D., Wharton, R.P., Aggarwal, A.K., 2003. Model of the brain tumor-Pumilio translation repressor complex. *Genes Dev.* 17, 2508–2513.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Cryst. D* 60, 2126–2132.
- Evans, G., Pettifer, R.F., 2001. *CHOOCH*: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.* 34, 82–86.
- Faine, S., Adler, B., Bolin, C., Perolat, P., 1999. *Leptospira* and leptospirosis, second ed. MediSci, Melbourne.
- Fraczkiewicz, R., Braun, W., 1998. Exact and efficient analytical calculation of the accessible surface area and their gradient for macromolecules. *J. Comput. Chem.* 19, 319–333.
- Fülöp, V., Jones, D.T., 1999. Beta propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* 9, 715–721.
- Gamberini, M., Gomez, R.M., Atzingen, M.V., et al., 2005. Whole-genome analysis of *Leptospira interrogans* to identify potential vaccine candidates against leptospirosis. *FEMS Microbiol. Lett.* 244, 305–313.
- Grosse-Kunstleve, R.W., Adams, P.D., 2003. Substructure search procedures for macromolecular structures. *Acta Cryst. D* 59, 1966–1973.
- Hamburger, Z.A., Brown, M.S., Isberg, R.R., Bjorkman, P.J., 1999. Crystal structure of invasins: a bacterial integrin binding protein. *Science* 286, 291–295.
- Hazes, B., Hol, W.G.J., 1992. Comparison of the hemocyanin β -barrel with other Greek key β -barrels: possible importance of the β -zipper in protein structure and folding. *Proteins* 12, 278–293.
- Holm, L., Sander, C., 1995. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478–480.
- Humphries, M.J., Newham, P., 1998. The structure of cell-adhesion molecules. *Trends Cell. Biol.* 8, 78–83.
- Hutchinson, E.G., Thornton, J.M., 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* 3, 2207–2216.
- Ibraghimov-Beskrovnaya, O., Bukanov, N.O., Donohue, L.C., Dackowski, W.R., Klinger, K.W., Landes, G.M., 2000. Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. *Hum. Mol. Genet.* 9, 1641–1649.
- Jing, H., Takagi, J., Liu, J.H., Lindgren, S., Zhang, R.G., Joachimiak, A., Wang, J.H., Springer, T.A., 2002. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure* 10, 1453–1464.
- Juncker, A.S., Willenbrock, H., von Heijne, G., Nielsen, H., Brunak, S., Krogh, A., 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662.
- Kabsch, W., 1988. Evaluation of single-crystal X-ray diffraction data from a position sensitive detector. *J. Appl. Cryst.* 21, 916–924.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst. D* 60, 2256–2268.
- Lambrigt, D.G., Sondck, J., Bohm, A., Skiba, N.P., Hamm, H.E., Sigler, P.B., 1996. The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* 379, 311–319.
- Lamzin, V.S., Perrakis, A., Wilson, K.S., 2001. The ARP/WARP suite for automated construction and refinement of protein models. In: Rossmann, M.G., Arnold, E. (Eds.), *International Tables for Crystallography: Crystallography of Biological Macromolecules*, vol. F. Kluwer Academic Publishers, Dordrecht, pp. 720–722.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
- Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005a. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* 33 (Web Server issue), W89–W93.
- Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005b. Protein function prediction using local 3D templates. *J. Mol. Biol.* 351, 614–626.
- Leslie, A.G.W., 1992. Joint CCP4 + ESF-EAMCB. Newsletter on Protein Crystallography. n° 26.
- Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., Bork, P., 2006. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* 34, D257–D260.
- Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G., Chothia, C., 2000. SCOP: a structural classification of proteins database. *Nucleic Acids Res.* 28, 257–259.
- Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., et al., 2000. Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex. *Nature* 405, 1073–1077.
- Macario, A.J., Conway De Macario, E., 2001. The molecular chaperone system and other anti-stress mechanisms in archaea. *Front. Biosci.* 6, D262–D283.
- Matthews, B.W., 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33, 491–497.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D* 53, 240–255.
- Nascimento, A.L., Verjovski-Almeida, S., Van Sluys, M.A., et al., 2004a. Genome features of *Leptospira interrogans* serovar Copenhageni. *Braz. J. Med. Biol. Res.* 37, 459–477.
- Nascimento, A.L., Ko, A.I., Martins, E.A., et al., 2004b. Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J. Bacteriol.* 186, 2164–2172.
- Neves, F.O., Abreu, P.A., Vasconcelos, S.A., de Morais, Z.M., Romero, E.C., Nascimento, A.L., 2007. Identification of a novel potential antigen for early-phase serodiagnosis of leptospirosis. *Arch. Microbiol.* 188, 523–532.
- Nicholls, A., Sharp, K.A., Honig, B., 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281–296.
- Pickles, L.M., Roe, S.M., Hemingway, E.J., Stifani, S., Pearl, L.H., 2002. Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. *Structure* 10, 751–761.
- Plank, R., Dean, D., 2000. Overview of the epidemiology, microbiology, and pathogenesis of *Leptospira spp* in humans. *Microbes Infect.* 2, 1265–1276.
- Pitsyn, O.B., Finkelstein, A.V., 1980. Similarities of protein topologies: evolutionary divergence, functional convergence or principles of folding? *Q. Rev. Biophys.* 13, 339–386.
- Ren, S.X., Fu, G., Jiang, X.G., Zeng, R., et al., 2003. Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 422, 888–893.
- Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* 95, 5857–5864.
- Setubal, J.C., Reis, M., Matsunaga, J., Haake, D.A., 2006. Lipoprotein computational prediction in spirochaetal genomes. *Microbiology* 152, 113–121.
- Sheinerman, F.B., Norel, R., Honig, B., 2000. Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* 10, 153–159.
- Sprague, E.R., Redd, M.J., Johnson, A.D., Wolberger, C., 2000. Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. *EMBO J.* 19, 3016–3027.
- Stura, E.A., Wilson, I.A., 1990. *Methods: A Companion to Methods in Enzymology*, vol. 1. Academic Press, New York, pp. 38–49.
- Takagi, J., Yang, Y., Liu, J.H., Wang, J.H., Springer, T.A., 2003. Complex between nidogen and laminin fragments reveals a paradigmatic beta-propeller interface. *Nature* 424, 969–974.
- Tarricone, C., Perrina, F., Manzani, S., Massimiliano, L., Kim, M.H., Derewenda, Z.S., Knapp, S., Tsai, L.H., Musacchio, A., 2004. Coupling PAF signaling to dynein regulation: structure of LIS1 in complex with PAF-acetylhydrolase. *Neuron* 44, 809–821.
- Terwilliger, T.C., 2002. Automated structure solution, density modification and model building. *Acta Cryst. D* 58, 1937–1940.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tung, C.H., Yang, J.M., 2007. fastSCOP: a fast web server for recognizing protein structural domains and SCOP superfamilies. *Nucleic Acids Res.* 35 (Web Server issue), W438–W443.
- Van Duyn, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J., 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* 229, 105–124.
- Zhang, C., Kim, S.H., 2000. A comprehensive analysis of the Greek Key motifs in protein β -barrels and β -Sandwiches. *Proteins* 40, 409–419.
- Zhang, W., Young, A.C., Imarai, M., Nathenson, S.G., Sacchettini, J.C., 1992. Crystal structure of the major histocompatibility complex class I H-2Kb molecule containing a single viral peptide: implications for peptide binding and T-cell receptor recognition. *Proc. Natl. Acad. Sci. USA* 89, 8403–8407.

3.2. ARTIGO II

MANUSCRITO SUBMETIDO À REVISTA *JOURNAL OF STRUCTURAL BIOLOGY*

**The crystal structure of the leptospiral hypothetical protein
LIC12922 reveals homology with the periplasmic chaperone
SurA.**

Priscila O. Giuseppe, Marina V. Atzingen, Ana Lúcia T.O. Nascimento, Nilson I. T.
Zanchin and Beatriz G. Guimarães.

TITLE

The crystal structure of the leptospiral hypothetical protein LIC12922 reveals homology with the periplasmic chaperone SurA.

AUTHORS

Priscila O. Giuseppe¹, Marina Von Atzingen², Ana Lúcia T.O. Nascimento^{2,3}, Nilson I.T. Zanchin⁴, Beatriz G. Guimarães^{5,*}

AFFILIATIONS

¹ Brazilian Biosciences National Laboratory (LNBio), Brazilian Center of Research in Energy and Materials (CNPEM), Giuseppe Máximo Scolfaro 10000, 13083-970, Campinas SP, Brazil.

² Centro de Biotecnologia, Instituto Butantan, Avenida Vital Brasil 1500, 05503-900, São Paulo, SP, Brazil.

³ Interunidades em Biotecnologia, Instituto de Ciências Biomédicas, USP, Avenida Prof. Lineu Prestes 1730, 05508-900, São Paulo, SP, Brazil

⁴ Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Avenida Cândido Rondon 400, 13083-875, Campinas SP, Brazil.

⁵ Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin - BP 48 91192 GIF-sur-YVETTE CEDEX, France.

*Corresponding author: Beatriz Gomes Guimarães, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin - BP 48 91192 GIF-sur-YVETTE CEDEX, France. E-mail address: beatriz.guimaraes@synchrotron-soleil.fr. Phone: +33(0)1-69 35 97 94. Fax : +33-(0)1-69 35 94 56.

ABSTRACT

Leptospirosis is a world spread zoonosis caused by members of the genus *Leptospira*. Although leptospires were identified as the causal agent of leptospirosis almost 100 years ago, little is known about their biology, which hinders the development of new treatment and prevention strategies. One of the several aspects of the leptospiral biology not yet elucidated is the process by which outer membrane proteins (OMPs) traverse the periplasm and are inserted into the outer membrane. The crystal structure determination of the conserved hypothetical protein LIC12922 from *Leptospira interrogans* revealed a two domain protein homologous to the *Escherichia coli* periplasmic chaperone SurA. The LIC12922 NC-domain is structurally related to the chaperone modules of *E. coli* SurA and trigger factor, whereas the parvulin domain is devoid of peptidyl prolyl cis-trans isomerase activity. Phylogenetic analyses suggests a relationship between LIC12922 and the chaperones PrsA, PpiD and SurA. Based on our structural and evolutionary analyses, we postulate that LIC12922 is a periplasmic chaperone involved in OMPs biogenesis in *Leptospira spp.* Since LIC12922 homologs were identified in all spirochetal genomes sequenced to date, this assumption may have implications for the OMPs biogenesis studies not only in leptospires but in the entire Phylum.

KEYWORDS

LIC12922, crystal structure, parvulin domain, SurA homology, chaperone

ABBREVIATIONS¹

¹OMP, outer membrane protein;
PPIase, peptidyl prolyl cis-trans isomerase;
SeMet, selenomethionine;
SAD, single-wavelength anomalous diffraction;
PDB, protein data bank.

INTRODUCTION

Leptospirosis is a zoonosis caused by spirochetes bacteria of the genus *Leptospira* (Levett, 2001). In the past decade, international attention on leptospirosis has resurged due to occurrence of important outbreaks (Levett, 2001). This re-emerging infectious disease has been identified as an important public health problem in Asia and Latin America (McBride et al., 2005). Traditionally encountered at rural settings, it has emerged as the cause of epidemics in poor urban slum communities of developing countries (McBride et al., 2005). Despite its prevalence in the tropics, leptospirosis occurs worldwide (Levett, 2001). It affects not only humans but also several animal species, having remarkable impacts on livestock production (Faine et al., 1999).

The lack of effective treatment and prevention procedures as well as of adequate diagnostic methods has limited public health responses to leptospirosis (McBride et al., 2005). Although leptospires were identified as the causal agent of leptospirosis almost 100 years ago (Inada et al., 1916), little is known about their biology, which hinders the development of new treatment and prevention strategies. For several decades, the study of *Leptospira* spp. was hampered by the absence of genetic tools for use in leptospires (Ko et al., 2009). However, the recent publication of the genome sequence of three *Leptospira* species and the development of methods for genetic transformation of *Leptospira* spp. have allowed important advances in our understanding of leptospires biology and pathogenesis (Ko et al., 2009).

One of the several aspects of the biology of spirochetes not yet elucidated is the process by which outer membrane proteins (OMPs) traverse the periplasm and are inserted into the outer membrane (Xue et al., 2009; Cullen et al., 2004). The main source of information in this field comes from comparative genome analyses. Homologs of genes involved in protein secretion such as the *E. coli* *sec* (Valent et al., 1998) and signal recognition particle (Valent et al., 1995) (SRP)-related genes are found in *Leptospira* spp. and other spirochetes (Cullen et al., 2004). However, homologs of periplasmic chaperones such as *E. coli* PpiD, DegP and Skp, implicated in the folding of OMPs (Sklar et al., 2007; Müller et al., 2001), have not been identified in spirochetal genomes (Cullen et al., 2004). Another important chaperone involved in OMPs biogenesis in *E. coli* is the protein SurA

(Lazar et al., 1996). BLAST (Altschul et al., 1997) analysis using this chaperone as bait indicates distant homologs in some spirochetes, but not in *Leptospira* spp.

LIC12922 gene encodes a protein of unknown function in *Leptospira interrogans* serovar Copenhageni (Nascimento et al., 2004). Amino acid sequence analyses using the programs SignalP (Emanuelsson et al., 2007) and PSORT (Nakai and Horton, 1999) indicate that LIC12922 possesses a signal peptide and is probably addressed to the periplasm. Microarray data showed that LA0675 gene (the equivalent to LIC12922 in *L. interrogans* sv. Lai) is expressed at 28 °C and 37 °C in similar levels (Qin et al., 2006). The same was observed by Lo and colleagues by using a proteomic approach (Lo et al., 2009).

Moreover, LIC12922 orthologs are found in all leptospiral genomes sequenced to date, suggesting its importance for leptospire survival (Nascimento et al., 2004; Picardeau et al., 2008; Bulach et al., 2006). Recently, it has been shown that rLIC12922 do not react with plasminogen to generate active plasmin (Vieira et al., 2010). In addition, the prevalence of antibodies directed against the recombinant protein in convalescent human and hamsters sera is low (Atzingen et al., 2010).

The crystal structure of the conserved protein LIC12922 presented in this work adds a piece to the puzzle of OMPs biogenesis in *Leptospira* spp. It reveals the homology of LIC12922 to the *E. coli* periplasmic chaperone SurA (Lazar et al., 1996), and constitutes to our knowledge the first structure of a spirochetal homologous to SurA to be elucidated. Based on structural and phylogenetic analyses, we postulate that LIC12922 is a periplasmic chaperone involved in OMPs biogenesis in *Leptospira* spp. This assumption may have implications for the OMPs biogenesis studies not only in leptospire but in the entire Phylum Spirochaetes.

MATERIALS AND METHODS

Expression and purification

The gene LIC12922 (1089 pb) was cloned in pDESTTM17 (Invitrogen) without the sequence coding for the signal peptide (141 pb) defined by SignalP (Emanuelsson et al., 2007) as described in Vieira *et al.* (Vieira et al., 2010) and Atzingen *et al.* (Atzingen et al.,

2010). Recombinant His₆-LIC12922 (Glu₄₈-Asp₃₆₂) protein was produced in *Escherichia coli* C41(DE3) cells (Avidis) transformed with the recombinant expression vector and grown in LB medium supplemented with 100 µg/ml ampicillin.

After the culture reached OD₆₀₀ ~ 0.6, expression was induced for 4 h at 30 °C, 200 rpm, with 0.4 mM IPTG (isopropyl-1-thio-β-galactopyranoside). Cells were harvested by centrifugation at 3000 g for 30 min. Preparation of selenomethionine-substituted (SeMet) protein was performed based on a metabolic inhibition method (Van Duyne et al., 1993).

Briefly, *E. coli* C41(DE3) cells (Avidis) transformed with the recombinant expression vector were grown at 37 °C, 200 rpm in M9 minimal medium supplemented with 100 µg/ml ampicillin. When the culture reached OD₆₀₀ ~ 0.6, it was supplemented with 10 ml of a sterile mixture containing lysine, phenylalanine and threonine (10 mg/ml), isoleucine, leucine and valine (5 mg/ml) and selenomethionine (6 mg/ml) diluted into M9 medium without glucose and thiamine. After 15 min of incubation (30 °C, 200 rpm), expression was induced by addition of 0.4 mM IPTG. Cells were grown for a further 4 h at 30 °C and harvested by centrifugation at 3000 g for 30 min.

LIC12922 and LIC12922-SeMet were purified using the same protocol, except that 5 mM β mercaptoethanol was added to the buffers used for LIC12922-SeMet purification. The pellets from 1 L cell culture were resuspended in 40 ml of buffer A (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 5 mM imidazole) containing 1 mM PMSF (phenylmethylsulfonyl fluoride). Cells were disrupted by sonication and the lysate was clarified by centrifugation (30 min at 20000 g). For removal of nucleic acids, the lysate was incubated with 2 % w/v streptomycin sulfate under agitation (45 min, 4°C), followed by centrifugation (30 min at 20000 g). Subsequently the supernatant was loaded onto a 5 ml HiTrapTM Chelating HP column (GE Healthcare) pre-charged with Ni₂SO₄ using an ÄKTA-FPLC system set at a flow rate of 1.5 ml/min.

After washing the column with 10 CV (column volume) of buffer A, bound proteins were eluted using a 5 mM to 1 M imidazole gradient divided in 3 parts: a 4 CV gradient from 0 % to 10 % buffer B (buffer A + 1 M imidazole), a 2 CV step of 10 % buffer B and a 18 CV linear gradient from 10 % to 100 % buffer B. Fractions containing the target protein were pooled and dialyzed against 20 mM sodium acetate pH 5.0, 20 mM NaCl. Protein samples were concentrated using Amicon Ultra-15 10000 MW filters (Millipore). Protein

concentration was estimated from direct absorbance at 280 nm considering a predicted extinction molar coefficient of $40680 \text{ M}^{-1}\text{cm}^{-1}$.

For removal of N-His tag, protein sample at 1 mg/ml was submitted to limited proteolysis by trypsin (0.01 mg/ml) for 2 h in 20 mM sodium acetate pH 6.2, 20 mM NaCl at room temperature. The reaction was stopped with 0.5 mM PMSF and concentrated to 1 ml. The protein was further purified on a HiLoad™ 16/60 Superdex™ 75 PG column (GE Healthcare) equilibrated with 20 mM sodium acetate pH 5.0, 150 mM NaCl using an ÄKTA-FPLC system set at a flow rate of 1 ml/min.

Crystallization

Native and SeMet-labeled proteins were crystallized by vapor diffusion in hanging drops at 20 °C. The best crystals were obtained by combining the techniques of microseeding (Stura and Wilson, 1992) and crystallization in gel (Biertümpfel et al., 2002). For the native protein, crystals grown in 0.1 M sodium citrate pH 5.5, 1.2 M ammonium phosphate, 0.2 M sodium chloride were transferred to an Eppendorf tube containing SEED BEAD (Hampton Research) and 50 µl of stabilization buffer (mother liquor). The tube was vortexed to produce microseeds. For the SeMet labeled protein, microseeds were produced using crystals grown in 0.1 M sodium citrate pH 5.5, 1 M ammonium sulfate, 0.2 M sodium chloride. The seed stocks were serially diluted (10^2 - 10^6 fold) in the respective stabilization buffers.

Drops containing 2.5 µl of native protein solution (2 mg/ml) and 2.5 µl of reservoir solution containing 0.2 % low melting point agarose (Sigma-Aldrich) were pre-equilibrated over 200 µl of reservoir solution (0.1 M sodium citrate pH 5.5, 1 M ammonium phosphate, 0.2 M sodium chloride) for 1 h at 20 °C. Microseeds were then introduced by adding 0.5 µl of each dilution to the drops. SeMet labeled crystals were produced in drops containing 3 µl of protein solution (1.2 mg/ml) and 1 µl of reservoir solution containing 0.2 % low melting point agarose (Sigma-Aldrich). Microseeds dilutions (0.4 µl) were added to the drops after pre-equilibration over 200 µl of reservoir solution (0.1 M sodium citrate pH 5.5, 0.8 M ammonium sulfate, 0.2 M sodium chloride) for 1 h at 20 °C. Average dimensions of native and SeMet crystals used for data collection were 300 µm and 200 µm, respectively.

Data collection and processing

SeMet-labeled crystals were cryo-protected by a 2 min soaking in the reservoir solution supplemented with 10 % ethylene glycol and 10 % glycerol prior to flash-cooling in a 100 K nitrogen-gas stream. X-ray diffraction data were collected on the PROXIMA 1 beamline at Synchrotron SOLEIL, France. Anomalous scattering factors were derived from the emission fluorescence spectrum of the crystal using the program CHOOCH (Evans and Pettifer, 2001). Diffraction data were collected at a single energy (12662 eV) corresponding to the maximum of f'' . The crystal was exposed for 0.5 s per 0.25 ° rotation of φ with the distance to detector set to 496 mm.

Native crystals cryo-protected by soaking in reservoir solution plus 15 % ethylene glycol and 10 % glycerol were flash-cooled in an 85 K nitrogen-gas stream. X-ray diffraction data were collected on PROXIMA 1 beamline at 12650 eV in two steps. Firstly, the crystal was exposed for 0.4 s per 0.5 ° rotation of φ with the distance to detector set to 446 mm and the beam attenuated to 40 %. Then, the crystal was exposed for 1 s per 0.5 ° rotation of φ with full intensity of the beam in order to get the maximum resolution. Data were recorded on an ADSC Quantum 315r detector and processed and scaled using the program XDS and XSCALE (Kabsch, 1993). Table I summarizes data collection statistics.

Table I. Data Collection and Refinement Statistics

Data Statistics^a

	Native	SeMet
Space group	<i>P</i> 23	<i>P</i> 23
Unit cell (Å)	a= b=c= 137.12	a= b=c= 138.07
Wavelength (Å)	0.980	0.979
Resolution range (Å)	35.00 - 3.10 (3.18 – 3.10)	35.00 - 3.48 (3.68 – 3.48)
N° of unique reflections	15872 (1177)	21868 (3508) ^b
Redundancy	10.2 (7.7)	7.8 (7.6)
Completeness (%)	99.8 (100.0)	99.7 (99.3)
I/σ(I)	21.5 (2.7)	15.4 (2.7)
R _{meas} (%) ^c	8.8 (88.1)	8.5 (71.3)

Refinement Statistics

Resolution range (Å)	33.25 - 3.10
N° of reflections	15869
R _{factor}	0.20
R _{free} ^d	0.22
R.m.s.d. bond length (Å)	0.01
R.m.s.d. bond angle (°)	1.2
Mean B value (Å ²)	141.07
Protein atoms	2438
Residues in Molprobity Ramachandran plot regions ^e (%)	
favoured	94.5
disallowed	0.3
Residues in PROCHECK Ramachandran plot regions ^f (%)	
most favoured	91.5
additional allowed	7.8
generously allowed	0.7
disallowed	0.0

^a Numbers in parentheses are for the highest resolution bin

^b considering bijvoet pairs to be independent reflections

$$^c R_{meas} = \sum_h \left(\frac{n_h}{n_h - 1} \right)^{1/2} \frac{\sum_i | \hat{I}_h - I_{h,i} |}{\sum_h \sum_i I_{h,i}} \quad \text{with} \quad \hat{I}_h = \frac{1}{n_h} \sum_i I_{h,i}$$

^d R_{free} (test set = 5 %)

^e Molprobity (Davis *et al.*, 2007)

^f PROCHECK (Laskowski *et al.*, 1993)

Structure determination and refinement

The crystal structure of LIC12922 was initially determined by single-wavelength anomalous diffraction method using the diffraction data from SeMet-labeled crystals. Six selenium sites were found using the program SHELXD (Sheldrick, 2008) and initial phasing calculation was performed with SHARP (Bricogne et al., 2003). Density modification was carried out with SOLOMON (Abraham and Leslie, 1996). A preliminary model was obtained by superposing a fragment of the model of *E. coli* SurA (PDB ID: **2PV3**) to the experimental map using the program MOLREP (Vagin and Isupov, 2001).

Alternate cycles of manual model building using COOT (Emsley and Cowtan, 2004) and refinement with REFMAC5 (Murshudov et al., 1997) allowed the building of a partial model containing 234 out of the 325 residues expected in the asymmetric unit (1 monomer). Native dataset at 3.1 Å resolution was later obtained and combined with the SeMet data set for phase extension using SHARP/SOLOMON (Bricogne et al., 2003; Abraham and Leslie, 1996). The partial model was then positioned in the new experimental map using MOLREP (Vagin and Isupov, 2001).

Cycles of refinement with BUSTER 2.9.1 (Bricogne et al., 2010) were performed, alternating with visual inspection of the electron density maps and manual rebuilding with COOT (Emsley and Cowtan, 2004). Due to the limited resolution of the data, the torsion, planar and Ramachandran restraints were applied for real space refinement using COOT (Emsley and Cowtan, 2004). The map sharpening tool available in COOT (Emsley and Cowtan, 2004) (B factor set to -30) was employed in the last cycles of model building.

A total of 312 residues (out of the 325 expected for 1 monomer) were modeled, comprising LIC12922 residues from Leu₅₀ to Leu₃₆₁. The final model was established after the convergence of R_{factor} and R_{free} values to 0.20 and 0.22 respectively. Stereochemistry of the model was analyzed with PROCHECK (Laskowski et al., 1993) and Molprobit (Davis et al., 2007). In the Molprobit Ramachandran plot, 94.5 % of the residues are in the favored region and only one residue (Trp257) falls in the disallowed region. Additional refinement details are summarized in table I. Figures were produced using the program Pymol (DeLano, 2002).

Protease-coupled PPIase assay

A protease-coupled assay based on Kofron and colleagues (Kofron et al., 1991) was used to verify if LIC12922 parvulin domain presents PPIase activity. The substrate *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (Sigma-Aldrich) was solubilized in TFE (2,2,2-trifluoroethanol) containing 470 μ M LiCl.

The reaction was carried out in 50 mM HEPES [(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)] pH 8, 100 mM NaCl with 1 mg/ml α -chymotrypsin at 10 °C. LIC12922 and substrate concentrations were 3 μ M and 50 μ M, respectively. The absorption at 390 nm was measured in intervals of 0.5 s during 10 min using an Analytik Jena spectrophotometer SPECORD 210. Human cyclophilin A (Sigma-Aldrich) at 0.1 μ M was employed as positive control.

Phylogenetic analyses

BLASTP (Altschul et al., 1997) searches (BLOSUM45) against the non-redundant protein database of NCBI were performed using SurA (REFSEQ ID: **NP_752018.1**), PpiD (EMBL ID: **CAQ30914.1**) and PrsA (REFSEQ ID: **NP_388876.1**) protein sequences as baits. Alignment of the LIC12922 sequence (REFSEQ ID: **YP_002837.1**) was performed using the PSI-BLASTP algorithm (Altschul et al., 1997) (3-iterations) and restricting the search set to the phylum Spirochaetes. Nine homologous sequences were selected for each bait. Since PpiD and SurA parvulin domains are paralogs, the respective homologous sequences were selected from the same species.

Multiple sequence alignment was generated by M-Coffee (Moretti et al., 2007) using default parameters. Multiple structural alignment of the parvulin domains from LIC12922 (PDB ID **3NRK**, G194-T304), SurA (PDB ID: **2PV2**, T172-R274), PpiD (PDB ID: **2KGJ**) and PrsA (PDB ID: **1ZK6**), was performed by using SSM (Krissinel and Henrick, 2004).

The region corresponding to the parvulin domain in the multiple sequence alignment was manually edited using the structural alignment as a guide. A phylogenetic tree for the parvulin domain was built by using the PHYLIP 3.69 package (Felsenstein, 2009). For bootstrap analyses, 100 data sets were created by the SEQBOOT program

(Felsenstein, 2009). The parsimony trees were built with the PROTPARS (Felsenstein, 2009) program. The consensus bootstrap tree was obtained with the CONSENSE program (Felsenstein, 2009) and drawn by using the TreeView program (Page, 1996). PrsA clade was set as outgroup since Firmicutes are considered the earliest branching bacterial phylum in relation to Spirochaetes and Proteobacteria (Ciccarelli et al., 2006).

The distribution of SurA (Bitto and McKay, 2002), LIC12922/PrsA (Vitikainen et al., 2004) and PpiD (Weininger et al., 2010) similar domain architectures within the kingdom Bacteria was generated based on searches of the domain combination “Trigger Factor/SurA peptide binding domain-like” and “FKBP-like” using the SUPERFAMILY database (Gough, 2002). Proteins presenting similar architectures and belonging to distinct genera were grouped in the respective phylum or class, according to the classification available at the NCBI taxonomy database (Sayers et al., 2009).

RESULTS AND DISCUSSION

Overall Structure

SeMet-labeled and native protein LIC12922 (Glu₄₈-Asp₃₆₂) were produced in *E. coli* fused to a N-terminal His-tag, purified, submitted to limited proteolysis followed by size exclusion chromatography and crystallized. The crystal structure was initially solved by SAD methods using diffraction data from SeMet-labeled crystals and subsequently a native data set at 3.1 Å resolution was used for phase extension and refinement to a final R_{factor} of 0.20 ($R_{\text{free}} = 0.22$).

The final atomic model consists of one monomer and the quality of the electron density maps allowed modelling of amino acid residues 50 to 361. The side chains of about 10 % of the residues could not be modeled. The backbone conformation of a single residue, Trp₂₅₇, lies in the disallowed region of the Molprobit Ramachandran plot (Davis et al., 2007).

Analyses of assemblies using the PISA web server (Krissinel and Henrick, 2007) identified a trimer as a possible stable quaternary structure for LIC12922 in solution. However, size exclusion chromatography (Supplemental Fig. S1) and Dynamic Light Scattering (DLS) experiments at low salt concentrations indicated the presence of

monomers in solution. According to DLS data, recombinant LIC12922 (without part of its His-tag) presents an average hydrodynamic radius of 3,2 nm (polydispersity index 12,6 %). This value corresponds to a mass of 50 kDa, which is closer to the mass of LIC12922 monomers (38 kDa) and is in agreement with the size exclusion chromatography results (Supplemental Fig. S1). Thus, the multimeric state suggested by PISA analysis may have been induced by the conditions of high protein and salt concentrations present in the crystallization condition.

The LIC12922 structure reveals two domains (Fig. 1). The sequentially discontinuous NC-domain is composed of segments from the N- (Leu50-Val193) and C- (Met306-Leu361) termini of the polypeptide chain. It is mainly constituted of α -helices, except for the antiparallel β -sheet formed by two strands of the N-segment and one strand of the C-segment. In addition, a short α -helix (α 8) followed by an unstructured region form a 21-residue C-terminal extension (Fig. 1B).

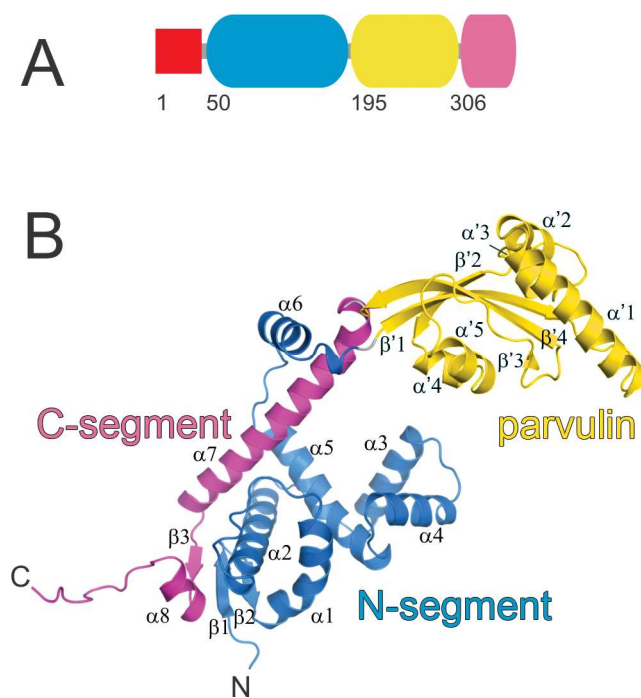


Fig. 1. Overall structure of LIC12922.

A. LIC12922 primary structure representation showing the predicted signal peptide (*red*), the parvulin domain (*yellow*) and the N- (*blue*) and C- (*pink*) segments which form the NC-domain. Numbers indicate the first residue of each segment. **B.** Cartoon representation of LIC12922. Colors are the same used in A and N- and C- termini are indicated. Secondary structure elements are numbered from β 1 to α 8 (NC domain) and from β '1 to β '4 (parvulin domain).

The second LIC12922 domain (Phe195-Thr304) is connected to the N- and C-segments through short linkers and displays a parvulin fold (Fig. 1B). The parvulin fold was first described for the human enzymes Pin1 and Par14 that catalyze the *cis/trans*

isomerization of peptidyl-prolyl bonds (Ranganathan et al.; Terada et al., 2001). It consists of a four-stranded antiparallel β -sheet core surrounded by four α -helices ($\beta\alpha_3\beta\alpha_2$ topology). An additional α -helix is present in the LIC12922 parvulin domain, generating a $\beta\alpha_3\beta\alpha_2\beta_2$ topology (Fig. 1B).

The LIC12922 NC domain is structurally related to chaperone domains and is intrinsically flexible

Comparative analyses evidence that the NC-domains of LIC12922 (LIC12922_NC) and *E. coli* periplasmic chaperone SurA (Bitto and McKay, 2002) (SurA_NC) share the same folding and topology, except for the C-terminal extension present in LIC12922.

Structural superposition of LIC12922_NC and SurA_NC using SSM (Krissinel and Henrick, 2004) results in an overall r.m.s deviation of 2.39 Å for 116 C α atoms aligned. Moreover, LIC12922 and SurA NC-domains are also structurally related to the C-terminal domain of the *E. coli* cytoplasmic chaperone trigger factor (TF_C) (Merz et al., 2006) and to the *Mycoplasma pneumoniae* hypothetical protein MPN555 (Schulze-Gahmen et al., 2005). Interestingly, the “arms” and “body” structures described in TF_C (Merz et al., 2006) are also present in LIC12922_NC. We propose here a bisection of the “body” module in “head” and “spine” to describe the LIC12922 NC-domain by the four elements: “head” (α -helix 6), “spine” (α -helix 7) and “arms” 1 (α -helices 1-2) and 2 (α -helices 3-5) (Fig. 2A). In addition to TF_C, these structural modules can also be observed in SurA_NC and MPN555 (Fig. 2B).

Comparison of these four structures reveals distinct arrangements of the secondary structure elements. Notably, the orientation of the helices from the “head” and “arms” generates a compact structure in the case of MPN555 whereas TF_C and SurA_NC adopt an open conformation. LIC12922_NC assumes an intermediate arrangement (Fig. 2B).

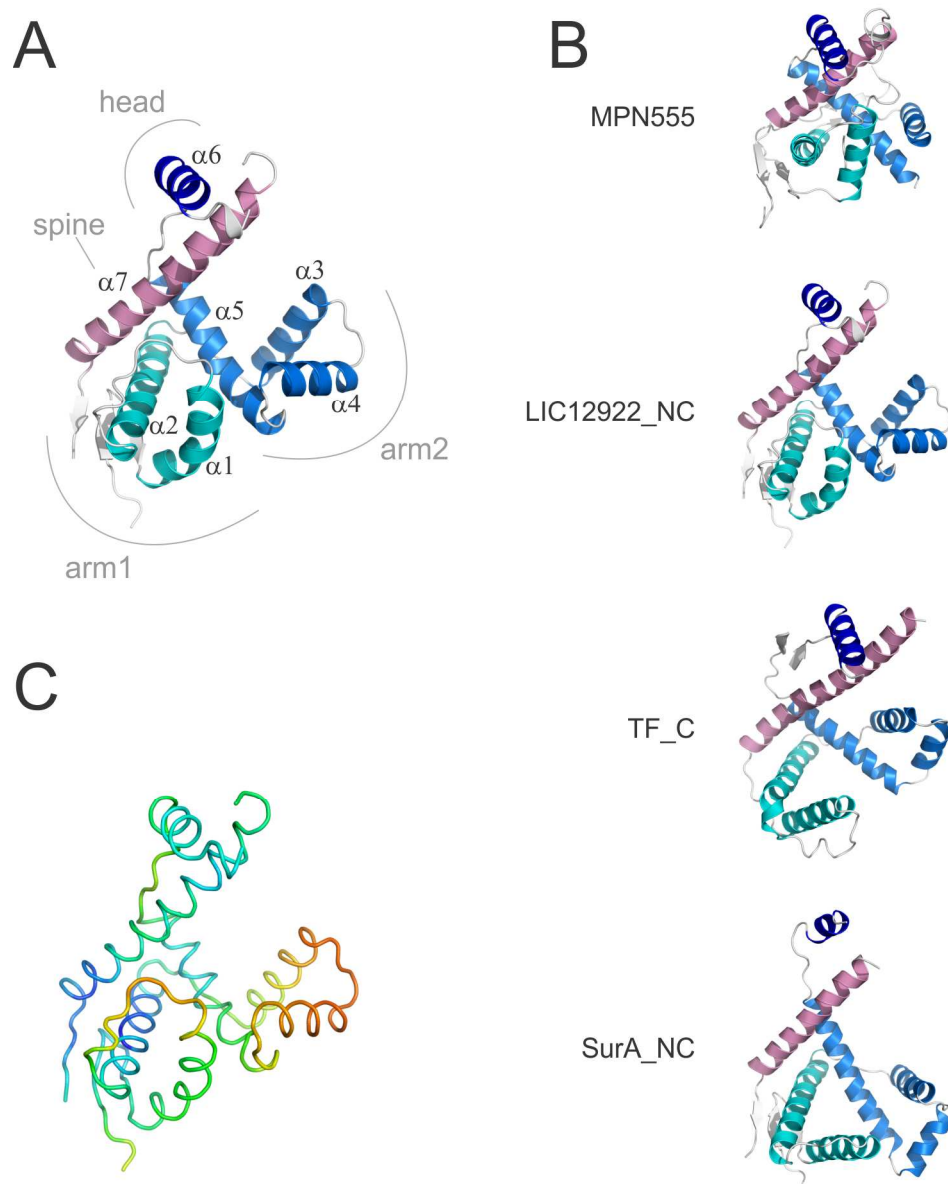


Fig. 2. **The LIC12922 NC-domain.** A. Cartoon representation of the LIC12922 NC-domain evidencing the “head”, “spine” and “arms” structures. N-segment α -helices are highlighted in different shades of blue to distinguish “arm1” ($\alpha 1$ and $\alpha 2$), “arm2” ($\alpha 3$, $\alpha 4$ and $\alpha 5$) and the “head” ($\alpha 6$). The “spine” ($\alpha 7$) belongs to the C-segment (*pink*). B. Structural comparison between the LIC12922 NC-domain, the putative chaperone MPN555 from *M. pneumoniae*, the C-terminal domain of the *E. coli* trigger factor (TF_C) and the NC-domain of the *E. coli* chaperone SurA (SurA_NC). The correspondent helices are displayed using the same color code of the panel A. C. Ribbon representation of LIC12922 NC-domain colored accordingly to the B-factor values, from dark blue (low) to red (high). For clarity purposes, the LIC12922 C-terminal extension was omitted.

Analysis of the B-factor distribution in the LIC12922 crystal structure shows that the helices $\alpha 1$, $\alpha 3$ and $\alpha 4$ and loops connecting $\alpha 1$ - $\alpha 2$, $\alpha 3$ - $\alpha 4$ and $\alpha 4$ - $\alpha 5$ present the highest B-factor values (Fig. 2C). Calculation of the mean B-factor value (main-chain atoms) results in 167.6 \AA^2 for helices $\alpha 1$, $\alpha 3$ and $\alpha 4$ against 119.8 \AA^2 for the remaining helices. Crystal contacts can only partially explain these differences. Together with the comparative analyses with the structurally related proteins (Fig. 2B), this observation suggests that the NC-domain is intrinsically flexible and the structural rearrangement of the “arms” could be favored by the mobility of the loops connecting the helices.

The LIC12922 parvulin domain is devoid of PPIase activity

A protease-coupled assay (Kofron et al., 1991) using *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide as substrate revealed that LIC12922 does not present PPIase activity (Fig. 3). Catalytically inactive parvulin domains have already been described, for instance in the chaperones SurA (P1 domain) (Rouvière and Gross, 1996) and PpiD (Weininger et al., 2010).

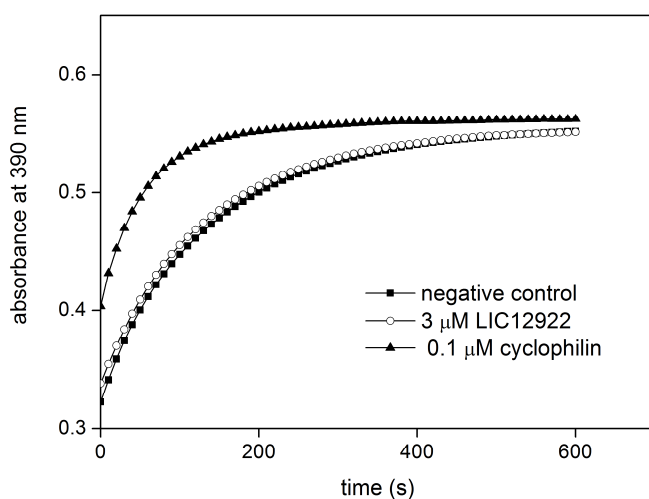


Fig. 3. **Protease-coupled PPIase assay.** Cyclophilin was used as positive control in concentration 30-fold lower than LIC12922.

The catalytic mechanism of parvulin PPIases is not completely elucidated (Heikkinen et al., 2009). Two main hypothesis have been proposed so far. The first is based on the crystal structure of the human Pin1 in complex with the *cis* dipeptide AlaPro (Ranganathan et al., 1997). The crystallographic model together with biochemical assays

suggested a covalent catalytic mechanism involving the Pin1 residues H59, S154, H157 and the proposed nucleophile C113 (Ranganathan et al., 1997).

The second hypothesis, that has emerged ten years later, contradicts the nucleophilic model (Behrsin et al., 2007). By analysis of Pin1 variants, Behrsin and colleagues observed that H157 is not essential for PPIase activity and proposed a mechanism according to which C113 is maintained ionized by H59 and presents a partial or full negative charge to the carbonyl oxygen atom of the *cis* substrate. Interaction with C113 would weaken the N–C double bond character of the substrate promoting rotation (Behrsin et al., 2007).

LIC12922 lack of PPIase activity can be explained by means of comparative structural analysis (Fig. 4). The Pin1 cysteine residue C113 is substituted by an aspartic acid in LIC12922 (D245). However, the same substitution occurs in other parvulins (Fig. 4A), including the PPIase PrsA from *Bacillus subtilis* (Vitikainen et al., 2004), and thus should not have a deleterious effect on the activity.

In contrast, replacement of the Pin1 active site H59 by the hydrophobic isoleucine I201 in LIC12922 (Fig. 4A) may have a preponderant role in the loss of PPIase activity. In addition, although the residues forming the proline binding pocket in Pin1 are partially conserved in LIC12922, the residues R290, R292 and Y266 occlude the putative substrate binding site (Fig. 4B). Tyrosine Y266 is part of an insertion in the LIC12922 polypeptide chain which is not observed in the related parvulin domains and R292 substitutes a highly conserved glycine (Fig. 4A). Moreover, R292 makes a hydrogen bond with D245, which in turn is not in position to interact with putative substrates (Fig. 4).

The exact relationship between PPIase activity and *in vivo* function of parvulins remains unclear (Behrsin et al., 2007). Examples have been described where severe loss of PPIase activity did not result in loss of cell viability, as it might be expected (Wu et al., 2000; Gemmill et al., 2005). A role involving peptide binding is generally proposed for the catalytically inactive parvulin domains. For instance, the P1 domain of chaperone SurA is suggested to be responsible for substrate recognition (Xu et al., 2007). Despite the lack of PPIase activity, it maintains the ability to bind peptides in a region which overlaps with the substrate binding site of the active parvulin Pin1 (Ranganathan et al., 1997). The same feature is observed for the parvulin domain of the chaperone PpiD (Weininger et al., 2010).

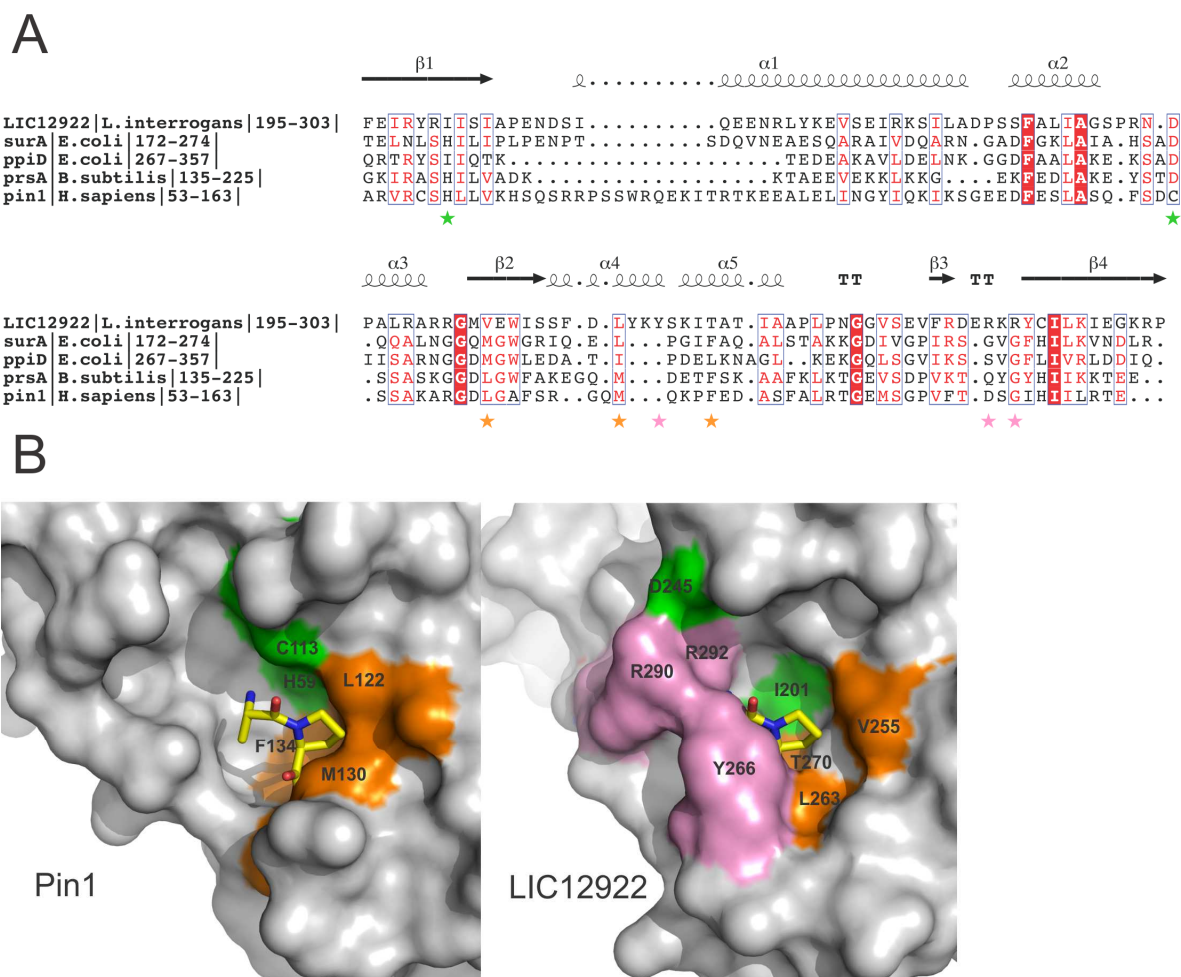


Fig. 4. Comparative structural analysis of parvulin domains. A. Structure based sequence alignment of the parvulin domains from LIC12922, *E. coli* SurA and PpiD, *B. subtilis* PrsA and human Pin1. Structural alignment was performed by using SSM (Krissinel and Henrick, 2004). Sequence alignment was formatted with ESPrIPT 2.2 (Gouet et al., 1999). The protein identification and the numbers of the first and last aligned residues are indicated on the left. The secondary structure elements of LIC12922 parvulin domain are displayed at the top of the alignment (arrows, helices and "T" indicate β -strands, α -helices and β -turns, respectively). Blue boxes highlight the conserved regions with similar residues represented by red letters and identical residues shown as white letters. The green and orange stars indicate, respectively, the Pin1 residues involved in catalysis and those forming the proline binding pocket (Ranganathan et al., 1997; Behrsin et al., 2007). The pink stars indicate the LIC12922 residues which occlude the potential substrate binding site. B. Representation of the parvulin active site of Pin1 (Ranganathan et al., 1997; Behrsin et al., 2007) in complex with a dipeptide and the correspondent region in LIC12922. Residues occupying equivalent positions are represented in the same color, according to the stars code used in panel A. The dipeptide AlaPro (stick model) was modeled in the LIC12922 putative binding site by superposing LIC12922 and Pin1 parvulin domains structures using SSM (Krissinel and Henrick, 2004).

Putative peptide interaction sites

The LIC12922 crystallographic model was analyzed using the Protein-Protein Interface Prediction (PPI-Pred) algorithm (Bradford and Westhead, 2005). LIC12922 shows an extensive surface potentially involved in protein interaction, which includes both the NC- and parvulin domains (Fig. 5).

The first most probable protein binding site is located at the parvulin domain (Fig. 5A). It includes the region corresponding to the active site of Pin1 (Ranganathan et al., 1997) and to the peptide binding site of the SurA P1 domain (Xu et al., 2007). The second and third most probable protein binding sites according to the PPI-Pred score are contiguous and located at the NC-domain. They include a narrow channel formed by the “arms” modules and extend to a cavity present at the bottom face of NC-domain (Fig. 5A).

The putative protein interaction surface predicted for the NC-domain agrees with the substrate binding site proposed to the correspondent *E. coli* SurA domain and the trigger factor C-domain (Bitto and McKay, 2002; Merz et al., 2006; Behrens, 2002; Lakshmipathy et al., 2007). In these structures, a crevice between the two “arms” modules gives access to a cleft formed by the “arms” and the “spine” elements (Bitto and McKay, 2002; Ferbitz et al., 2004). Biochemical studies on the *E. coli* trigger factor indicated that polypeptide chains traverse this crevice engaging the entire C-domain to progress towards the PPIase domain (Merz et al., 2008). In LIC12922, this crevice is closed and the central cleft is reduced to a narrow channel due to a different arrangement of the helices $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$ and $\alpha 5$ (Fig. 6).

Interestingly, the crystal structures of apo-forms of *E. coli* SurA (Bitto and McKay, 2002) and trigger factor (Merz et al., 2006) are not strictly in the unbound state, since portions of symmetry-related molecules are observed interacting with the crevice or the central cleft. In contrast, there is a minimal involvement of the corresponding “arms” in LIC12922 crystal contacts and we speculate that the conformation of LIC12922 NC-domain might correspond to an unbound state that could undergo a structural rearrangement to enable the binding of the substrate.

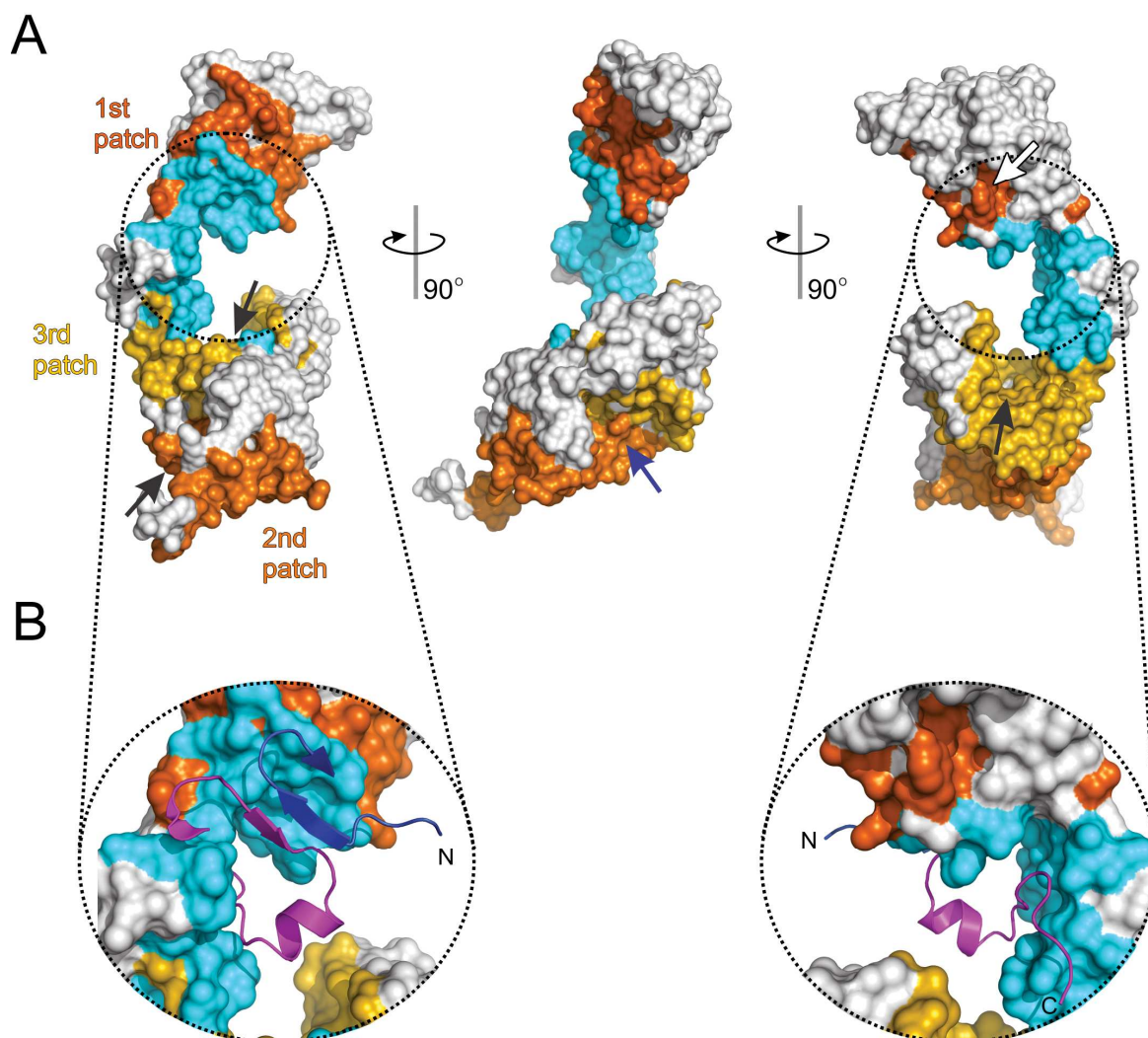


Fig. 5. **Putative peptide binding sites of LIC12922.** A. Molecular surface representation showing the potential protein-protein interaction sites predicted by PPI-Pred analyses (Bradford and Westhead, 2005). The gradient *dark orange* to *yellow* indicates a decreasing PPI-Pred score. The region defined by PISA (Krissinel and Henrick, 2007) analyses as forming a stable interaction with a crystal neighbor molecule is colored in *cyan*. *Black* arrows (right panel) indicate the entrances of the NC-domain channel. The *blue* arrow (central panel) indicates the region which corresponds to the SurA and Trigger Factor crevices (Bitto and McKay, 2002; Ferbitz et al., 2004) and the *white* arrow (right panel) indicates the region corresponding to the Pin1 active site (Ranganathan et al., 1997). B. Closer view of the LIC12922 assembly interface with a crystal neighbor (cartoon). The cartoon representation is colored according to the fig. 1 and N- and C- termini are indicated.

Analysis of the stable oligomeric assembly of LIC12922 proposed by the PISA algorithm (Krissinel and Henrick, 2007) shows that the short antiparallel β -sheet formed by the N- and C-segment extremities and the C-terminal extension tightly interact with the crystal neighbor molecule at the interface of the NC- and parvulin domains (Fig. 5B). Although the trimer suggested by PISA could not be identified in solution, we speculate whether this oligomer indicates a propensity for protein-protein complex formation, evidenced by the conditions of high protein and salt concentrations present in the crystal. This crystal packing interface connects the three putative protein binding sites indicated by PPI-Pred, suggesting that, similarly to the trigger factor mechanism (Merz et al., 2008), substrate binding may encompass the entire concave surface of LIC12922.

Structural superposition of LIC12922 and SurA crystallographic models reveals that the orientation of the parvulin domains relative to the NC-domain is remarkably variable. The crystal structure of full length SurA shows the second parvulin domain (P2) as a satellite module whereas the NC-domain and the first parvulin domain (P1) form a core module (Bitto and McKay, 2002). Interestingly, the position of the LIC12922 parvulin domain does not match any of the SurA parvulin domains, although it is closer to the P2 orientation (Fig. 6). Moreover, the structure of a truncated form of SurA (lacking the P2 domain) in complex with a dodecapeptide revealed a rearrangement of the P1 domain (Xu et al., 2007) (Fig. 6).

Thus, the analysis of LIC12922 and structure comparison with the homologous proteins shows that in addition to the “arms” flexibility, the orientation between the NC- and the parvulin domain seems to be quite variable, indicating that structural flexibility should play a role in substrate binding mechanism.

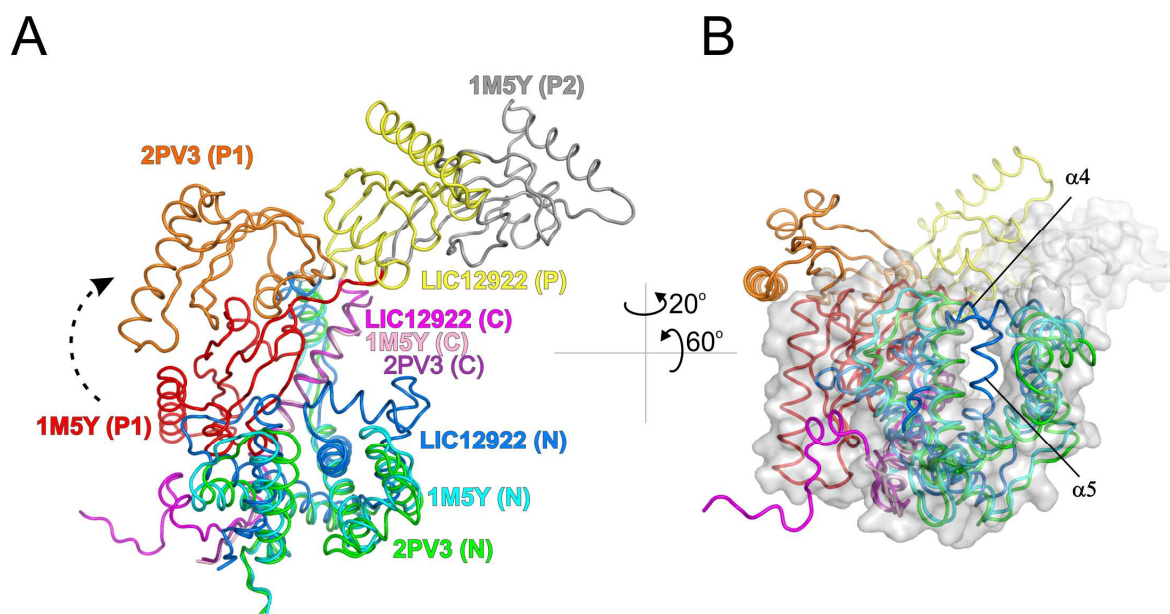


Fig. 6. **Structural alignment of LIC12922 and *E. coli* SurA.** A. The SurA models are identified by their PDB codes. The model 1M5Y includes the complete SurA structure, whereas 2PV3 corresponds to a truncate form, which lacks the second parvulin domain, in complex with a peptide (not shown in the figure). Parvulin domains, N- and C- segments are labeled by P, N and C respectively. For clarity, code colors are indicated in the figure. The dotted arrow highlights the rearrangement of SurA P1 domain. B. Bottom view of the aligned structures evidencing the LIC12922 helices $\alpha 4$ and $\alpha 5$ which occlude the crevice observed in the SurA structure. Surface representation corresponds to the model 1M5Y. Structures superposition was performed with the LSQKAB program (Kabsch, 1976) using the conserved helix $\alpha 7$ (pink) as reference.

Evolutionary relationship between LIC12922 and extracytoplasmic chaperones

To further investigate a possible chaperone function of LIC12922, we analysed the evolutionary relationship between LIC12922 and related extracytoplasmic chaperones. Since different domains may evolve independently (Buljan and Bateman, 2009), the most conserved LIC12922 domain, the parvulin domain, was selected to be used in the multiple alignments. From these alignments (Supplemental Fig. S2), a maximum parsimony tree was inferred, revealing the probable relationship between PrsA (Vitikainen et al., 2004), PpiD (Weininger et al., 2010), SurA (Bitto and McKay, 2002) and LIC12922 (Fig. 7).

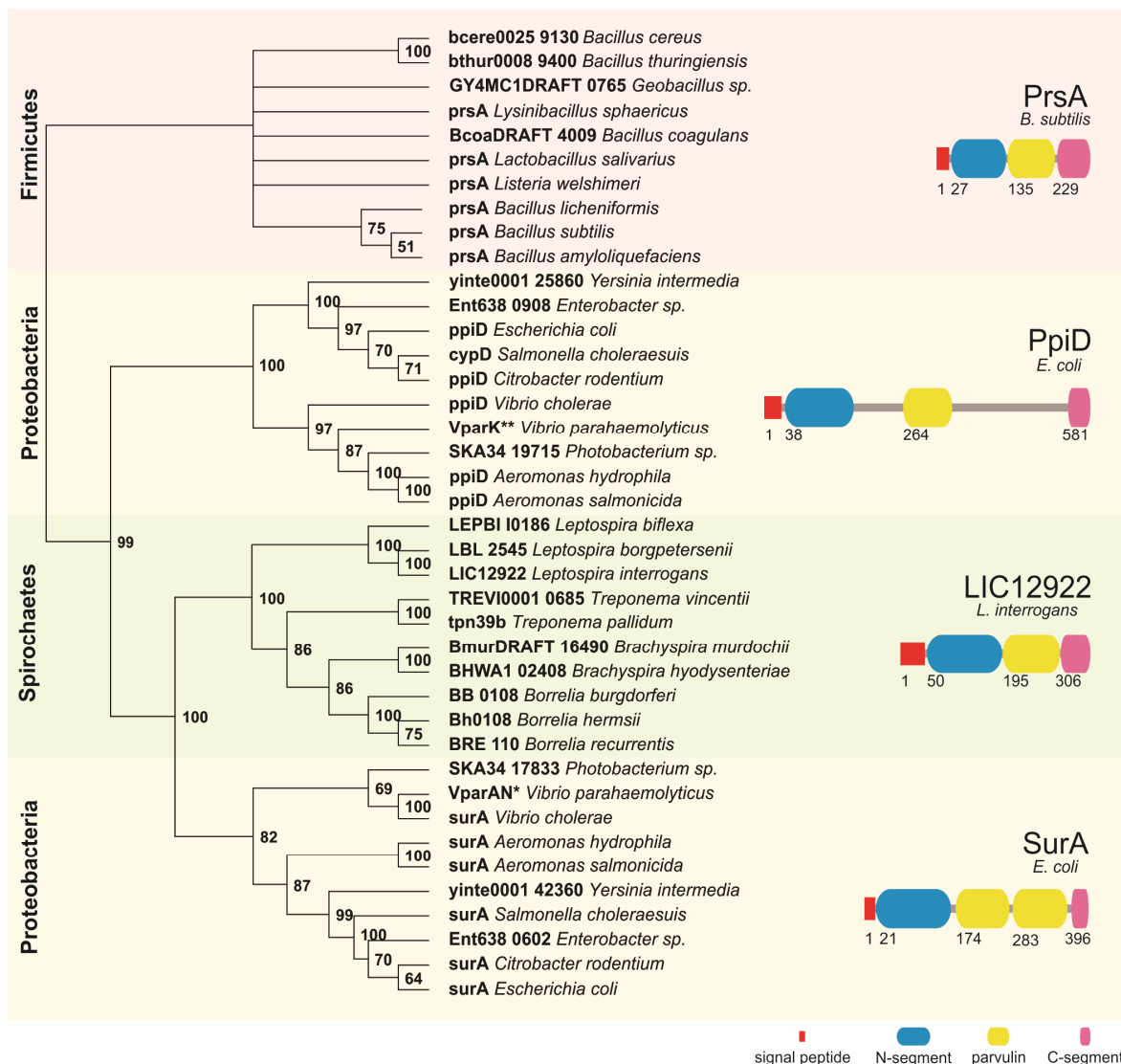


Fig. 7. Evolutionary tree evidencing a relationship between LIC12922 and extracytoplasmic chaperones. Maximum parsimony bootstrap tree was based on multiple amino acid sequence alignment of parvulin domains of LIC12922, the spirochaetal homologs and the extracytoplasmic chaperones PrsA, PpiD and SurA. The numbers at branch points indicate the percentage of the 100 bootstrapped data sets that supported the specific internal branches of the cladogram. Branches with less than 50% bootstrap support were collapsed. The taxa are indicated by the gene name and species. PrsA taxa (Firmicutes) were set as outgroup. PpiD and SurA genes belong to Proteobacteria while LIC12922 clade is composed by genes from Spirochaetes. For each cluster, the conserved segments or domains of a representative member are shown on the right. Signal peptide sequences (red) were predicted using SignalP (Emanuelsson et al., 2007). N-segments (blue) and C-segments (pink) of PrsA and PpiD were assigned based on sequence analyses using SUPERFAMILY (Gough et al., 2001) and PSIPRED (Bryson et al., 2005). The other domains were defined according to structural data: SurA (PDB ID [1M5Y](#)), PpiD (PDB ID [2KGJ](#)) and PrsA (PDB ID [1ZK6](#)) Numbers indicate the first residue of each segment. Gene names identified with asterisks correspond to *VparA_N010200009300 and **VparK_010100022392.

The resulting cladogram indicates that SurA P1 domain is the closest homolog of LIC12922 parvulin domain, which is in agreement with their structural similarity (r.m.s. deviation of 1.43 Å for 100 C- α atoms aligned). The evolutionary tree also suggests that the gene duplication event responsible for the divergence of SurA and PpiD parvulin domains occurred early in evolution, probably before the speciation event that separated SurA from LIC12922 parvulin ancestors (Fig. 7).

In addition to the evidences of evolutionary relationship between the parvulin domains, comparison of the architecture of LIC12922 and the extracytoplasmic chaperones PrsA (Vitikainen et al., 2004), PpiD (Weininger et al., 2010) and SurA (Bitto and McKay, 2002) reveals a similar organisation of domains. The primary and tertiary structure analyses indicate that these four related proteins present the N-segment, one or two parvulin domains and the C-segment arranged sequentially (Fig. 7).

In order to analyse the distribution of proteins with similar domain arrangements within the Kingdom Bacteria, we verified the occurrence of the combination of the segment N (“Trigger Factor/SurA peptide binding domain-like” in the SCOP classification (Hubbard et al., 1997)) and the parvulin domain (“FKBP-like” in the SCOP classification (Hubbard et al., 1997)) in the bacterial genomes sequenced to date, using the SUPERFAMILY algorithm (Gough et al., 2001) (Fig. 8).

The domain architecture similar to that of PrsA and LIC12922 structures is the most ubiquitous, being identified in all bacterial groups analysed, including Cyanobacteria (Fig. 8). Interestingly, multiple genes encoding for proteins with this domain arrangement were identified in several Firmicutes and Proteobacteria species, indicating that, for this gene, duplication is a common event. Thus, it is plausible to suggest that SurA and PpiD parvulin domains evolved from distinct gene copies, present in the ancestor of Proteobacteria and Spirochaetes as proposed by the phylogenetic analysis (Fig. 6).

The SurA and PpiD-like domain architectures are not observed at Firmicutes, Actinobacteria and Spirochaetes and show a complex distribution. In spite of being observed at the deep branching phylum *Thermotoga* (Ciccarelli et al., 2006), PpiD-like architectures were not identified at late branch groups, such as ϵ -proteobacteria (Fig. 8). A similar observation is valid for SurA-like domain arrangements. However, PpiD- and SurA-like proteins are frequently observed at β - and γ -proteobacteria. The restricted distribution

of the SurA domain architecture in comparison with PrsA-like arrangements suggests that the presence of an additional parvulin domain may not be critical for chaperone function. Indeed, functional experiments have shown that the lack of the second parvulin domain of SurA, which presents PPIase activity, does not affect SurA function *in vivo* (Behrens et al., 2001).

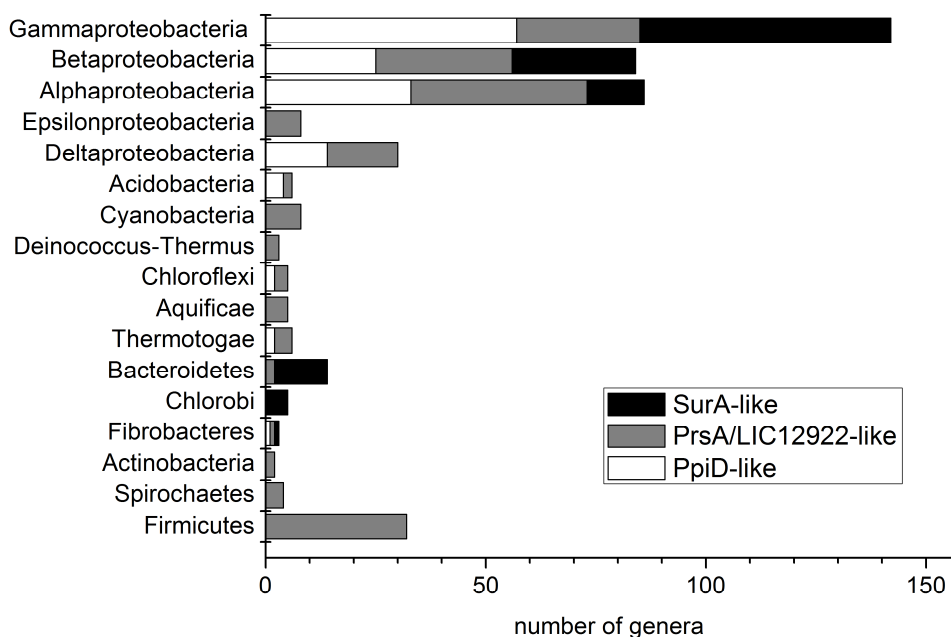


Fig. 8. Distribution of the domain architectures found in SurA, PrsA/LIC12922 and PpiD within the Kingdom Bacteria. The number of genera in function of phylum or class was generated based on searches of the domain combination “Trigger Factor/SurA peptide binding domain-like” and “FKBP-like” using the SUPERFAMILY database (Gough, 2002).

The evidences of a relationship between LIC12922 and the chaperones PrsA (Vitikainen et al., 2004), PpiD (Weininger et al., 2010) and SurA (Bitto and McKay, 2002), which perform similar functions at extracytoplasmic environments, strongly supports the hypothesis that LIC12922 is also a chaperone. Furthermore, the homology between LIC12922 and SurA together with the putative location of LIC12922 at the periplasm lead us to postulate that LIC12922 is also involved in the biogenesis of outer membrane proteins.

Finally, although LIC12922 does not share a high sequence identity with the spirochetal orthologs (15 % to 20 %), analyses using the SUPERFAMILY program (Gough et al., 2001) indicated the presence of the N-segment and the parvulin domain in all related

sequences. This result suggests that homologs of LIC12922 are indeed present in other spirochetes and reinforce the importance of the gene for these bacteria.

To conclude, the inference of a chaperone function for LIC12922 may have implications for OMPs maturation studies on the entire Phylum Spirochaetes. Moreover, it was recently demonstrated that SurA is essential to the virulence of uropathogenic *E. coli*, *Salmonella*, and *Shigella* spp. and it was proposed as a potential drug target against Gram-negative pathogens (Behrens-Kneip et al., 2010). The identification of a SurA homolog in leptospire and other spirochetes may contribute to the development of new treatments against diseases caused by members of the phylum.

ACCESSION NUMBER

Coordinates and structure factors have been deposited in the Protein Data Bank with the accession number 3NRK.

ACKNOWLEDGEMENTS

We thank Celisa C. Tonoli, Andréia N. Meza, Elaine C. Teixeira, Adriana C. A. Pinto, Tereza C. L. Silva and Blandine Pineau for technical support. This work was supported by FAPESP grants CEPID/CBME 98/14138-2, SMoIBNet 00/10266-8 and SOLEIL Synchrotron. POG was recipient of FAPESP and PDEE/CAPES fellowships.

SUPPLEMENTARY INFORMATION

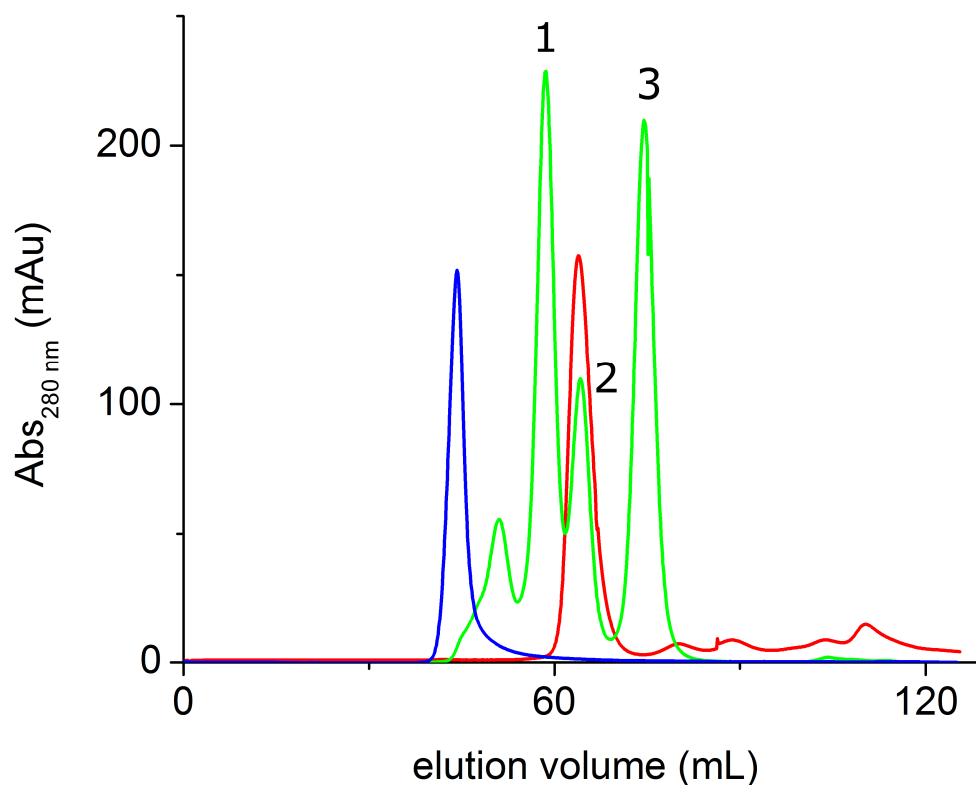


Fig. S1. Analytical gel filtration performed on a HiLoad™ 16/60 Superdex™ 75 PG column at a flow rate of 1 ml/min. The column was pre-equilibrated with 20 mM sodium acetate pH 5, 150 mM NaCl. Elution pattern of LIC12922 after limited proteolysis is shown in *red*, standard proteins in *green* and blue dextran 2000 in *blue*. Numbers indicate the standard proteins elution peaks: 1 - albumin (67 kDa), 2 - ovalbumin (43 kDa), 3 - chymotrypsinogen (25 kDa). The molecular mass estimated for LIC12922 based on this experiment is 46 kDa which is compatible with a monomeric state (theoretical mass for a monomer is approximately 38 kDa).

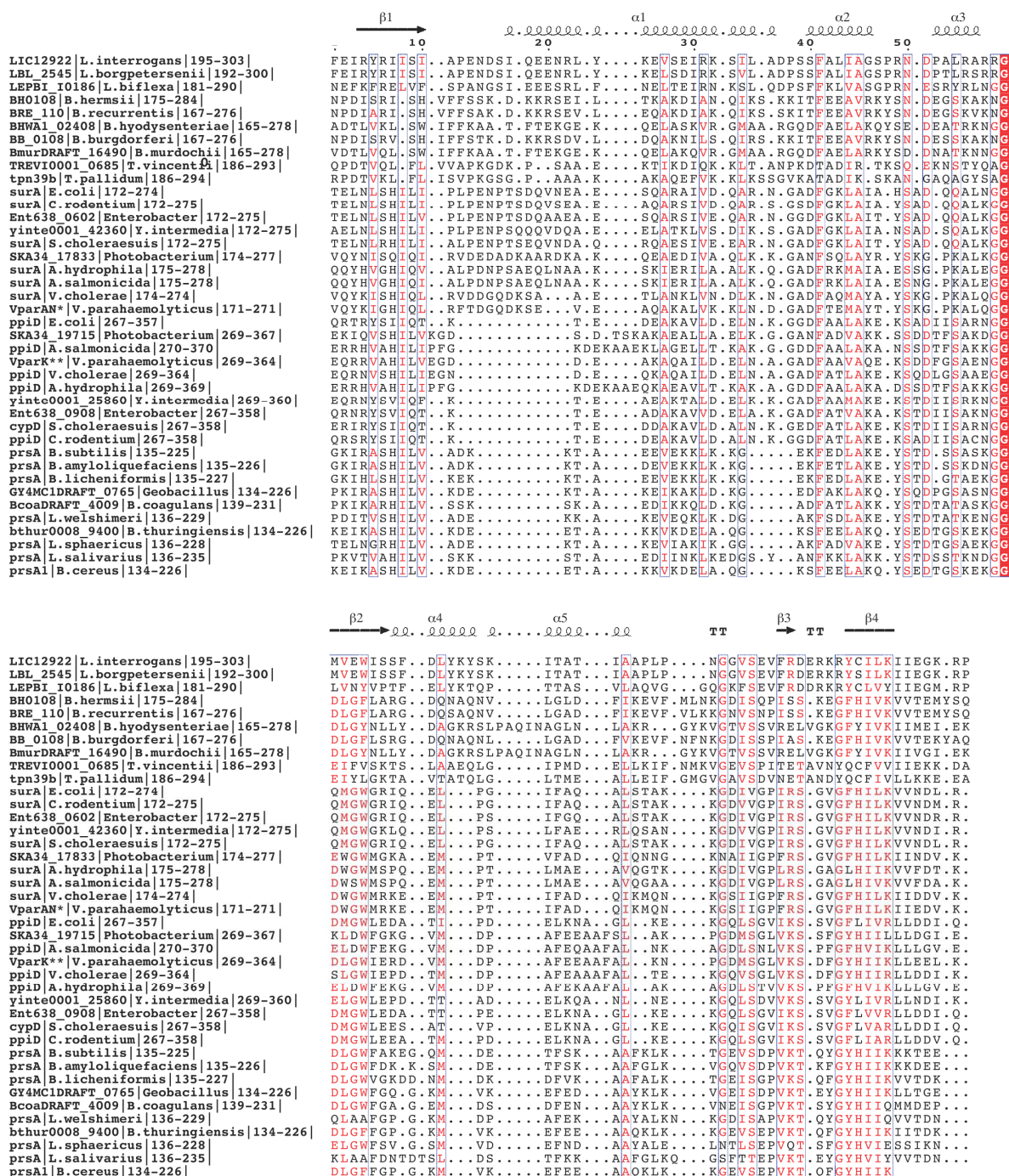


Fig. S2. Multiple sequence alignment of parvulin domains used for the construction of the evolutionary tree presented in figure 7. Secondary structure elements of LIC12922 parvulin domain are displayed at the top of the alignment (arrows, helices and “T” indicate β -strands α -helices and β -turns, respectively). On the left of the gene name, the organism and the numbers of the first and last residue aligned are indicated. Conserved regions are highlighted by blue boxes. Conserved residues are displayed in red and identical residues are displayed as white letters in a red background. Gene names identified with asterisks correspond to *VparA_N010200009300 and **VparK_010100022392. Sequence alignment was formatted using ESPrnt 2.2 (Gouet et al., 1999).

REFERENCES

- Abrahams, J.P., Leslie, A.G., 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* 52, 30-42.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- Atzingen, M., Goncales, A., Morais, Z., Araujo, E., De Brito, T., Vasconcellos, S., Nascimento, A.L., 2010. Characterization of leptospiral proteins that afforded partial protection in hamsters against lethal challenge with *Leptospira interrogans*. *J. Med. Microbiol.*
- Behrens, S., 2002. Periplasmic chaperones--new structural and functional insights. *Structure.* 10, 1469-1471.
- Behrens, S., Maier, R., de Cock, H., Schmid, F.X., Gross, C.A., 2001. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J.* 20, 285-294.
- Behrens-Kneip, S., 2010. The role of SurA factor in outer membrane protein transport and virulence. *Int. J. Med. Microbiol.*
- Behrsin, C.D., Bailey, M.L., Bateman, K.S., Hamilton, K.S., Wahl, L.M., Brandl, C.J., Shilton, B.H., Litchfield, D.W., 2007. Functionally important residues in the peptidyl-prolyl isomerase Pin1 revealed by unigenic evolution. *J. Mol. Biol.* 365, 1143-1162.
- Biertumpfel, C., Basquin, J., Suck, D., Sauter, C., 2002. Crystallization of biological macromolecules using agarose gel. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1657-1659.
- Bitto, E., McKay, D.B., 2002. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure.* 10, 1489-1498.
- Bradford, J.R., Westhead, D.R., 2005. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics.* 21, 1487-1494.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W., 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2023-2030.

- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Smart, O.S., Vonrhein, C., Womack, T.O. 2010. BUSTER, version 2.9.1. Global Phasing Ltd., Cambridge, United Kingdom.
- Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T., 2005. Protein structure prediction servers at University College London. *Nucleic Acids Res.* 33, W36-38.
- Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E., Alt, D.P., Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K., Adler, B., 2006. Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14560-14565.
- Buljan, M., Bateman, A., 2009. The evolution of protein domain families. *Biochem. Soc. Trans.* 37, 751-755.
- Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science.* 311, 1283-1287.
- Cullen, P.A., Haake, D.A., Adler, B., 2004. Outer membrane proteins of pathogenic spirochetes. *FEMS Microbiol. Rev.* 28, 291-318.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375-383.
- DeLano, W.L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>.
- Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953-971.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126-2132.
- Evans, G., Pettifer, R.F., 2001. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.* 34, 82-86.
- Faine, S., Adler, B., Bolin, C., Perolat, P., 1999. *Leptospira and leptospirosis*. second ed. MediSci, Melbourne.

- Felsenstein, J. 2009. PHYLIP (Phylogeny Inference Package), version 3.69, *Distributed by the author*. Department of Genome Sciences, University of Washington. Seattle, WA, USA, 2009. Available online: <http://evolution.genetics.washington.edu/phylip.html> (accessed on 18 March 2010).
- Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., Ban, N., 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*. 431, 590-596.
- Gemmill, T.R., Wu, X., Hanes, S.D., 2005. Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 280, 15510-15517.
- Gouet, P., Courcelle, E., Stuart, D.I. and Metz, F., 1999. ESPript: multiple sequence alignments in PostScript. *Bioinformatics*. 15, 305-308.
- Gough, J., 2002. The SUPERFAMILY database in structural genomics. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1897-1900.
- Gough, J., Karplus, K., Hughey, R., Chothia, C., 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903-919.
- Heikkinen, O., Seppala, R., Tossavainen, H., Heikkinen, S., Koskela, H., Permi, P., Kilpelainen, I., 2009. Solution structure of the parvulin-type PPIase domain of *Staphylococcus aureus* PrsA--implications for the catalytic mechanism of parvulins. *BMC Struct. Biol.* 9, 17.
- Hubbard, T.J., Murzin, A.G., Brenner, S.E., Chothia, C., 1997. SCOP: a structural classification of proteins database. *Nucleic Acids Res.* 25, 236-239.
- Inada, R., Ido, Y., Hoki, R., Kaneko, R., Ito, H., 1916. The Etiology, Mode of Infection, and Specific Therapy of Weil's Disease (*Spirochaetosis Icterohaemorrhagica*). *J. Exp. Med.* 23, 377-402.
- Kabsch, W., 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A*. 32, 922-923.
- Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26, 795-800.
- Ko, A.I., Goarant, C., Picardeau, M., 2009. *Leptospira*: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat. Rev. Microbiol.* 7, 736-747.

- Kofron, J.L., Kuzmic, P., Kishore, V., Colon-Bonilla, E., Rich, D.H., 1991. Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry*. 30, 6127-6134.
- Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256-2268.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774-797.
- Lakshmipathy, S.K., Tomic, S., Kaiser, C.M., Chang, H.C., Genevoux, P., Georgopoulos, C., Barral, J.M., Johnson, A.E., Hartl, F.U., Etchells, S.A., 2007. Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* 282, 12186-12193.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26 283-291.
- Lazar, S.W., Kolter, R., 1996. SurA assists the folding of Escherichia coli outer membrane proteins. *J. Bacteriol.* 178, 1770-1773.
- Levett, P.N., 2001. Leptospirosis. *Clin. Microbiol. Rev.* 14, 296-326.
- Lo, M., Cordwell, S.J., Bulach, D.M., Adler, B., 2009. Comparative transcriptional and translational analysis of leptospiral outer membrane protein expression in response to temperature. *PLoS Negl. Trop. Dis.* 3, e560.
- McBride, A.J., Athanazio, D.A., Reis, M.G., Ko, A.I., 2005. Leptospirosis. *Curr. Opin. Infect. Dis.* 18, 376-386.
- Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., Deuerling, E., 2006. The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. *J. Biol. Chem.* 281, 31963-31971.
- Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B., Deuerling, E., 2008. Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *EMBO J.* 27, 1622-1632.
- Moretti, S., Armougom, F., Wallace, I.M., Higgins, D.G., Jongeneel, C.V., Notredame, C., 2007. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res.* 35, W645-648.

- Muller, M., Koch, H.G., Beck, K., Schafer, U., 2001. Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane. *Prog. Nucleic Acid Res. Mol. Biol.* 66, 107-157.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240-255.
- Nakai, K., Horton, P., 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34-36.
- Nascimento, A.L., Verjovski-Almeida, S., Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Digiampietri, L.A., Harstkeerl, R.A., Ho, P.L., Marques, M.V., Oliveira, M.C., Setubal, J.C., Haake, D.A., Martins, E.A., 2004. Genome features of *Leptospira interrogans* serovar Copenhageni. *Braz. J. Med. Biol. Res.* 37, 459-477.
- Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12, 357-358.
- Picardeau, M., Bulach, D.M., Bouchier, C., Zuerner, R.L., Zidane, N., Wilson, P.J., Creno, S., Kuczek, E.S., Bommezzadri, S., Davis, J.C., McGrath, A., Johnson, M.J., Boursaux-Eude, C., Seemann, T., Rouy, Z., Coppel, R.L., Rood, J.I., Lajus, A., Davies, J.K., Medigue, C., Adler, B., 2008. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS One.* 3, e1607.
- Qin, J.H., Sheng, Y.Y., Zhang, Z.M., Shi, Y.Z., He, P., Hu, B.Y., Yang, Y., Liu, S.G., Zhao, G.P., Guo, X.K., 2006. Genome-wide transcriptional analysis of temperature shift in *L. interrogans* serovar lai strain 56601. *BMC Microbiol.* 6, 51.
- Ranganathan, R., Lu, K.P., Hunter, T., Noel, J.P., 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell.* 89, 875-886.
- Rouviere, P.E., Gross, C.A., 1996. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev.* 10, 3170-3182.
- Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Mizrahi, I., Ostell, J., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner,

- L., Yaschenko, E., Ye, J., 2009. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 37, D5-15.
- Schulze-Gahmen, U., Aono, S., Chen, S., Yokota, H., Kim, R., Kim, S.H., 2005. Structure of the hypothetical Mycoplasma protein MPN555 suggests a chaperone function. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1343-1347.
- Sheldrick, G.M., 2008. A short history of SHELX. *Acta Crystallogr. A.* 64, 112-122.
- Sklar, J.G., Wu, T., Kahne, D., Silhavy, T.J., 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev.* 21, 2473-2484.
- Stura, E.A., Wilson, I.A., 1992. *Seeding Techniques, Crystallization of Nucleic Acids and Proteins: A Practical Approach*, Oxford University Press, Oxford.
- Terada, T., Shirouzu, M., Fukumori, Y., Fujimori, F., Ito, Y., Kigawa, T., Yokoyama, S., Uchida, T., 2001. Solution structure of the human parvulin-like peptidyl prolyl cis/trans isomerase, hPar14. *J. Mol. Biol.* 305, 917-926.
- Vagin, A.A., Isupov, M.N., 2001. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1451-1456.
- Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., Luirink, J., 1995. Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *EMBO J.* 14, 5494-5505.
- Valent, Q.A., Scotti, P.A., High, S., de Gier, J.W., von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B., Luirink, J., 1998. The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J.* 17, 2504-2512.
- Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J., 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* 229, 105-124.
- Vieira, M.L., Atzingen, M.V., Oliveira, T.R., Oliveira, R., Andrade, D.M., Vasconcellos, S.A., Nascimento, A.L., 2010. In vitro identification of novel plasminogen-binding receptors of the pathogen *Leptospira interrogans*. *PLoS One.* 5, e11259.
- Vitikainen, M., Lappalainen, I., Seppala, R., Antelmann, H., Boer, H., Taira, S., Savilahti, H., Hecker, M., Vihinen, M., Sarvas, M., Kontinen, V.P., 2004. Structure-function analysis of PrsA reveals roles for the parvulin-like and flanking N- and C-terminal domains in protein folding and secretion in *Bacillus subtilis*. *J. Biol. Chem.* 279, 19302-19314.

- Weininger, U., Jakob, R.P., Kovermann, M., Balbach, J., Schmid, F.X., 2010. The prolyl isomerase domain of PpiD from *Escherichia coli* shows a parvulin fold but is devoid of catalytic activity. *Protein Sci.* 19, 6-18.
- Wu, X., Wilcox, C.B., Devasahayam, G., Hackett, R.L., Arevalo-Rodriguez, M., Cardenas, M.E., Heitman, J., Hanes, S.D., 2000. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.* 19, 3727-3738.
- Xu, X., Wang, S., Hu, Y.X., McKay, D.B., 2007. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J. Mol. Biol.* 373, 367-381.
- Xue, F., Yan, J., Picardeau, M., 2009. Evolution and pathogenesis of *Leptospira* spp.: lessons learned from the genomes. *Microbes Infect.* 11, 328-333.

3.3. PRODUÇÃO E CARACTERIZAÇÃO DE LIC10494

Expressão, purificação, cristalização e caracterização de LIC10494

RESUMO

LIC10494, identificada como provável lipoproteína, foi expressa em *Escherichia coli*, purificada e cristalizada em algumas condições das triagens iniciais na forma de agulhas ou esferulitas. Três condições foram refinadas, mas não renderam cristais adequados para o experimento de difração. Experimentos de gel filtração e DLS revelaram a tendência de agregação de LIC10494, a qual foi eficientemente suprimida com a utilização do aminoácido arginina como aditivo.

Estudos de gel filtração analítica sugerem que a fração não agregada de LIC10494 apresenta-se no estado monomérico e possui uma forma não globular. Análises de dicroísmo circular associadas a análises de sequência revelaram que LIC10494 é uma proteína parcialmente desenovelada que apresenta uma região rica em resíduos de treonina.

INTRODUÇÃO

O gene LIC10494 codifica para uma proteína de 25,8 kDa. Ele está localizado no cromossomo I de *L. interrogans* sorovar Copenhageni entre as ORFs LIC10493 e LIC10495 que codificam, respectivamente, para uma proteína hipotética e para uma proteína da membrana citoplasmática. Buscas em bancos de dados, utilizando-se o algoritmo PSI_BLAST (Altschul *et al.*, 1997), não encontram proteínas ortólogas a LIC10494. A presença de um lipobox característico em seu peptídeo sinal, seguido de um resíduo de cisteína, é uma evidência sugestiva de que ela seja uma lipoproteína (Pugsley, 1993; Haake, 2000) (FIGURA 1).

MNQLFYKIIILFFTSIILNGC

FIGURA 1. Peptídeo sinal de LIC10494.

Lipoproteínas de bactérias espiroquetas possuem em seu peptídeo sinal uma seqüência (lipobox) caracterizada por apresentar: uma região hidrofóbica com mais de 6 resíduos (*azul*); uma região carboxiterminal (*cinza*) com I, L, F ou V na posição -3 e -4 e G, A, S, ou N na posição -1 em relação ao sítio de clivagem predito (*vermelho*) (Pugsley, 1993; Haake, 2000).

Análises da seqüência de LIC10494 com os programas PSIPRED (Jones, 1999, Bryson *et al.*, 2005), PONDR (Li *et al.*, 1999) e FoldIndex (Prilusky *et al.*, 2005) indicam que LIC10494 apresenta uma região central intrinsecamente desestruturada de aproximadamente 70 resíduos (FIGURA 2). O programa FoldIndex prediz se uma dada seqüência polipeptídica é intrinsecamente desestruturada com base na hidrofobicidade residual média e na carga total da mesma (Prilusky *et al.*, 2005). O programa PONDR aplica redes neurais para a predição de ordem ou desordem em uma seqüência polipeptídica e se baseia em sua composição de aminoácidos e em propriedades das cadeias laterais tais como flexibilidade, hidropatia e número de coordenação (Li *et al.*, 1999). O número de coordenação é o número médio de cadeias laterais vizinhas que estão em contato com uma dada cadeia lateral quando ela está totalmente “enterrada” na estrutura. Este número foi calculado a partir de um conjunto de 33 proteínas não homólogas (Galaktionov & Marshall, 1996).

Na região de LIC10494 predita ser desestruturada, observa-se uma grande quantidade de resíduos de treonina (FIGURA 2A). O programa PSIPRED prediz algumas fitas beta

MATERIAIS E MÉTODOS

Expressão e purificação

A expressão do gene que codifica para a proteína LIC10494 foi induzida em células da linhagem C41(DE3) de *E. coli* (Avidis) transformadas por choque térmico com o vetor pDEST (Gateway) contendo o gene LIC10494 (84-696 pb) sem a região codificante para o peptídeo sinal. A indução foi realizada por 4h, a 30 °C, adicionando-se 0,4 mM IPTG em 1 L de cultura crescida em LB/ampicilina, sob agitação de 200 rpm, após esta ter atingido uma OD₆₀₀ entre 0,4-0,6. Subseqüentemente, a cultura foi centrifugada a 3000 g durante 30 min a 4 °C.

Para o preparo do extrato, as células foram ressuspensas em 40 mL de tampão de lise (50 mM Tris HCl pH 8, 300 mM NaCl, 1mM PMSF e 20 µg/mL DNase). A suspensão foi sonicada com 20 pulsos de 15 s, em intervalos de 30 s, em banho de gelo. O produto dessa sonicação foi centrifugado a 20000 g por 30 min a 4 °C e o sobrenadante incubado por 2 h, sob agitação a 4 °C com 1 mL de resina Ni-NTA (Qiagen) pré-equilibrada em tampão A (50 mM Tris HCl pH 8, 300 mM NaCl).

Em seguida, a mistura de extrato e resina foi transferida para uma coluna mantida à temperatura ambiente. Após a passagem do extrato, a resina foi lavada com 10 mL de tampão A. Para a eluição de LIC10494, adotou-se um gradiente de imidazol dividido em 6 degraus nas concentrações 10 mM (8 CV), 25 mM (8 CV), 50 mM (6 CV), 100 mM (4 CV), 150 mM (6 CV) e 500 mM (6 CV) de imidazol.

A eficiência e o rendimento dessa purificação foram analisados por SDS-PAGE (Laemmli, 1970). Determinadas as frações contendo a proteína alvo, estas foram agrupadas e dialisadas três vezes contra 0,5 L de tampão 50 mM Tris HCl pH 8, 150 mM NaCl.

Subseqüentemente, LIC10494 foi submetida a um segundo passo de purificação pelo método de cromatografia de exclusão molecular em colunas Superdex™ 75 10/30 (GE Healthcare) ou Superdex™ 75 16/60 (GE Healthcare), equilibradas em tampão 50 mM Tris HCl pH 8, 150 mM NaCl.

Na cromatografia realizada em coluna Superdex™ 75 10/30 (GE Healthcare), utilizou-se um fluxo de 0,8 mL/min e coletaram-se frações de 0,5 mL. Na realizada em coluna Superdex™ 75 16/60 (GE Healthcare), utilizou-se um fluxo de 1 mL/min e coletaram-se frações de 1 mL. As frações contendo a proteína alvo foram agrupadas,

concentradas em filtro Amicon Ultra-15 5000 MW (Millipore) e submetidas a ensaios de espalhamento de luz dinâmico, dicroísmo circular e cristalização.

Após cada etapa de purificação e de concentração, LIC10494 foi quantificada pela medida da absorção da amostra em 280 nm (Edelhoch, 1967) utilizando-se o coeficiente de extinção molar teórico $10240 \text{ M}^{-1} \text{ cm}^{-1}$ calculado pelo programa ProtParam (Gasteiger *et al.*, 2005).

Gel filtração analítica

Para os estudos de oligomerização de LIC10494 utilizou-se a coluna Superdex™ 75 16/60 (GE Healthcare) equilibrada em tampão 50 mM Tris HCl pH 8, 150 mM NaCl. A calibração da mesma foi realizada com uma mistura contendo os seguintes padrões: albumina (3,55 nm), ovalbumina (3,05 nm) e quimiotripsinogênio (2,09 nm) diluídos em tampão 50 mM Tris HCl pH 8, 150 mM NaCl. As eluições de LIC10494 e da amostra padrão foram realizadas separadamente, injetando-se na coluna um volume de 1 mL das respectivas amostras a um fluxo de 1 mL/min. Para se estimar o volume morto da coluna (V_0), utilizou-se *blue dextran* 2000. A partir do volume de eluição dos padrões (V_e) calculou-se a constante K_{av} dada por $\frac{V_e - V_0}{V_t - V_0}$, onde V_t é o volume total da coluna (124 mL).

A partir da curva obtida por regressão linear dos pontos $\sqrt{-\log(K_{av})}$ vs R_h , determinados para as proteínas padrão, estimou-se o R_h de LIC10494.

Espalhamento de luz dinâmico

Para se avaliar o estado oligomérico e a polidispersividade de LIC10494 após cada etapa cromatográfica e após ser concentrada para os ensaios de cristalização, mediu-se o espalhamento de luz dinâmico de amostras de LIC10494 a 18 °C em cubetas de quartzo utilizando-se um equipamento *Protein Solutions™ DynaPro™* (Wyatt Technology Corporation). Os dados foram analisados pelo programa *Dynamics™ V6* (Wyatt Technology Corporation).

Dicroísmo circular

O espectro de dicroísmo circular de amostras de LIC10494 foi medido em um espectropolarímetro Jasco J-810 utilizando-se uma cubeta de quartzo de 0,1 cm de caminho óptico. As medidas foram realizadas em quadruplicata, no intervalo de 195 nm a 260 nm, a cada 0,5 nm, a uma velocidade de 100 nm/min com um tempo de resposta de 1 s. Após serem corrigidas pela subtração do espectro de CD do tampão, as medidas foram convertidas de elipticidade para elipticidade molar residual média pela fórmula:

$$[\theta]_{\lambda} = \frac{\theta_{\lambda}^{obs} \times MRM}{10 \times d \times c}, \text{ onde } \theta_{\lambda}^{obs} \text{ (graus) é a elipticidade medida no comprimento de}$$

onda λ ; MRM é a massa molecular residual média (Da), obtida dividindo-se a massa molecular da proteína pelo seu número de resíduos de aminoácidos; d é o caminho óptico da cubeta (cm) e c é a concentração de proteína (g/mL).

Ensaio para análise de agregação

Com o intuito de se encontrar um tampão capaz de deslocar o equilíbrio agregado \leftrightarrow monômero para o estado monomérico de LIC10494, realizou-se o seguinte experimento, baseado em Bondos & Bicknell (2003). Incubaram-se por 1h, à temperatura ambiente, alíquotas compostas por 60 μ L de LIC10494 (5 mg/mL) em 20 mM Tris HCl pH 8, 20 mM NaCl e 60 μ L do mesmo tampão contendo os seguintes aditivos:

1. sem aditivo (controle)
2. 50 mM sulfato de amônio
3. 50 mM cloreto de magnésio
4. 50 mM L-arginina
5. 5 % glicerol
6. 0,1% tween 20

Além dessas condições, o tampão 20 mM HEPES de sódio pH 7, 20 mM NaCl também foi testado.

A separação das frações “agregada” (> 100 kDa) e “não agregada” (< 100 kDa) foi realizada por ultracentrifugação 1000 g (10 min) em membranas de exclusão molecular de 100 kDa, utilizando-se o sistema Centricon (Millipore). A fração que passou pelo filtro

(proteínas < 100 kDa) foi transferida para tubos novos. Os agregados retidos na membrana foram ressuspensos em 60 μL de água e coletados centrifugando-se, por 15 min, a 1000 g, o filtro invertido em um novo tubo. A proporção entre as frações agregadas e não agregadas foi analisada qualitativamente em SDS-PAGE.

Cromatografia de exclusão molecular em presença de arginina

Para se avaliar a influência da arginina no equilíbrio agregado \leftrightarrow monômero de LIC10494 em solução, realizaram-se, paralelamente, cromatografias de exclusão molecular de amostras de LIC10494 na ausência e na presença de 50 mM de arginina. Para tanto, uma amostra de LIC10494, purificada por cromatografia de afinidade, foi dividida em duas partes iguais, as quais foram dialisadas contra 50 mM Tris HCl pH 8, 150 mM NaCl (amostra I) e contra 50 mM Tris HCl pH 8, 150 mM NaCl contendo 50 mM arginina (amostra II). Após a diálise, as amostras foram concentradas em filtros Amicon Ultra-15 5000 MW para um volume aproximado de 1 mL e aplicadas em coluna Superdex™ 75 16/60 equilibradas no mesmo tampão das respectivas amostras.

As cromatografias foram realizadas em um sistema FPLC, com fluxo de 1 mL/min. As amostras I e II foram quantificadas antes e depois da cromatografia de exclusão molecular pela medida da absorção da amostra em 280 nm (Edelhoch, 1967), usando-se o coeficiente de extinção molar teórico $10240 \text{ M}^{-1} \text{ cm}^{-1}$, calculado pelo programa ProtParam (Gasteiger *et al.*, 2005). As frações correspondentes à amostra não agregada foram juntadas, concentradas em filtros Amicon Ultra (Millipore) e submetidas a ensaios de cristalização. Além disso, as amostras foram analisadas por SDS-PAGE e DLS.

Cristalização

Triagens iniciais

Ensaio de cristalização de LIC10494 foram realizados em microplacas de 96 poços crystal EX™ (Corning). O método empregado foi o da difusão de vapor em gota sentada. Gotas contendo 0,5 μL de solução de proteína (6,3 mg/mL) e 0,5 μL de solução do reservatório foram equilibradas contra 80 μL de solução do reservatório.

As soluções foram automaticamente dispensadas nas placas utilizando-se o robô de cristalização Honeybee™ (Genomic Solutions). Foram testados 6 kits [JCSG (Nextal/Qiagen), PACT (Nextal/Qiagen), Precipitant Synergy (Emerald BioSystems), SaltRX (Hampton Research), Crystal Screen I e II (Hampton Research), Wizard Screen I e II (Emerald BioSystems)] confeccionados usando-se o robô Matrix Maker™ (Emerald BioSystems). Os ensaios foram realizados a 18 °C.

Refinamento I

A partir das triagens iniciais, selecionaram-se três condições para refinamento: A (0,1 M Na₂HPO₄/ácido cítrico pH 4,2, 0,4 M K₂HPO₄, 1,6 M NaH₂PO₄); B (20 % PEG3350, 0,2 M Na₂HPO₄/KH₂PO₄) e C (0,1 M citrato de sódio pH 5,5, 2 M sulfato de amônio) (FIGURA 3). Os ensaios foram realizados em placas de 96 poços, equilibrando-se gotas de 2 μL (1 proteína :1 poço) contra 80 μL de solução do poço.

As soluções foram automaticamente dispensadas nas placas utilizando-se o robô de cristalização Honeybee™ (Genomic Solutions). Nestes refinamentos, fixou-se a concentração de proteína em 6 mg/mL e variou-se o pH vs concentração dos agentes precipitantes (FIGURA 3). Em uma das condições do refinamento de A, coraram-se os cristais com IZIT (Hampton Research) para se descartar a hipótese de que estes fossem de sal.

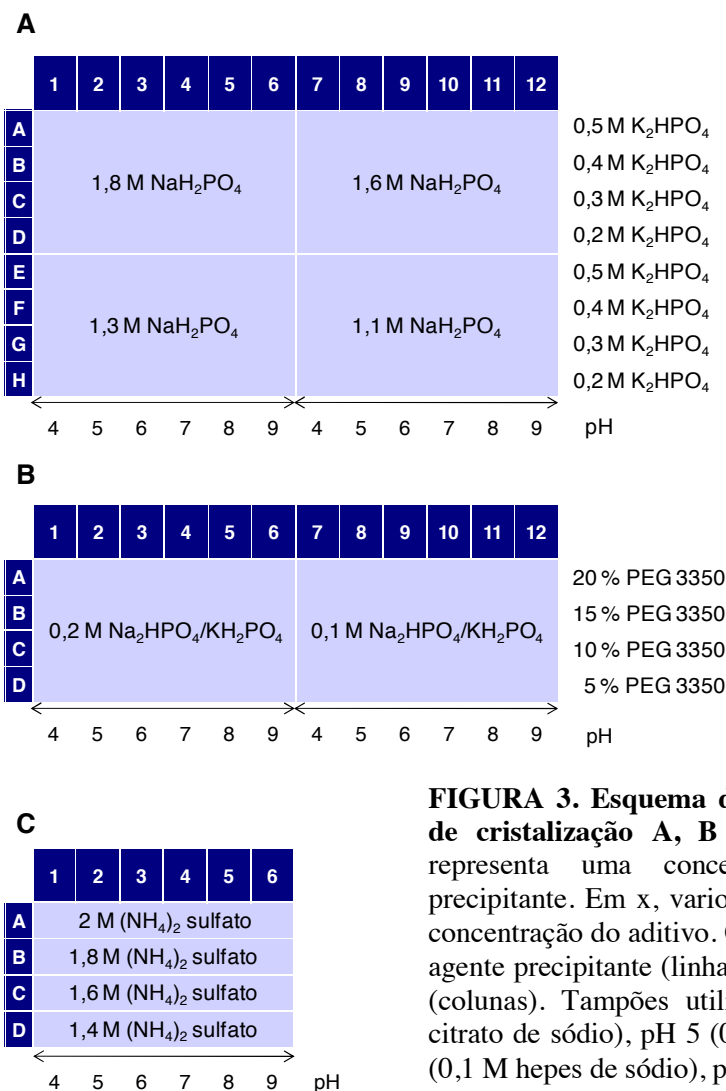


FIGURA 3. Esquema dos refinamentos das condições de cristalização A, B e C. A e B. Cada quadrante representa uma concentração diferente do agente precipitante. Em x, variou-se o pH da condição e em y a concentração do aditivo. C. Concentrações decrescentes do agente precipitante (linhas) são testadas em diferentes pHs (colunas). Tampões utilizados: pH 4 (0,1 M fosfato e citrato de sódio), pH 5 (0,1 M acetato de sódio), pH 6 - 8 (0,1 M hepes de sódio), pH 9 (0,1 M bicina).

Refinamento II

O segundo ciclo de refinamento foi realizado em presença e em ausência de 50 mM de arginina, indicado como supressor de agregação de LIC10494 em ensaios para análise de agregação. As amostras I (controle) (6 mg/mL) e II (+ arginina) (5 mg/mL) de LIC10494 foram submetidas a ensaios de cristalização nas condições A-D vs 1-12 do refinamento da condição A (FIGURA 3). Esses ensaios foram realizados a 18 °C em placas de 48 poços usando-se o método de gota pendurada.

Além disso, aplicou-se a técnica de *microseeding* (Bergfors, 2003) nos ensaios descritos na figura 4. As sementes foram produzidas a partir de cristais de LIC10494

utilizando-se o aparato SEED BEAD (Hampton Research) de acordo com as instruções fornecidas pelo fabricante. Gotas de 4 μL (1 proteína:1 poço) foram semeadas com 0,4 μL das diluições seriadas das sementes após 30 min ou 90 min de equilíbrio contra 200 μL de solução do poço.

		diluições das sementes								
		10 ⁻¹	10 ⁻²	10 ⁻³	10 ⁻⁴	10 ⁻⁵	10 ⁻⁶	10 ⁻⁷	-	
		1	2	3	4	5	6	7	8	Amostra (concentração)
A										I (3 mg/mL)
B										I (1,5 mg/mL)
C										II (5 mg/mL)
D										II (5 mg/mL)
E										II (5 mg/mL)

soluções

- A)** 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 , 0,1 M hepes de sódio pH 8
B) 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 , 0,1 M hepes de sódio pH 8
C) 1,3 M NaH_2PO_4 , 0,3 M K_2HPO_4 , 0,1 M hepes de sódio pH 8
D) 1,3 M NaH_2PO_4 , 0,2 M K_2HPO_4 , 0,1 M hepes de sódio pH 8
E) 1,0 M NaH_2PO_4 , 0,3 M K_2HPO_4 , 0,1 M hepes de sódio pH 8

		diluições das sementes								
		10 ⁻¹	10 ⁻²	10 ⁻³	10 ⁻⁴	10 ⁻⁵	10 ⁻⁶	10 ⁻⁷	-	
		1	2	3	4	5	6	7	8	Amostra (concentração)
A										II (5 mg/mL)
B										II (4 mg/mL)
C										II (3 mg/mL)
D										II (2 mg/mL)
E										II (1 mg/mL)

solução

- A-E)** 1,6 M NaH_2PO_4 , 0,2 M K_2HPO_4 , 0,1 M hepes de sódio pH 8

FIGURA 4. Esquema dos ensaios de cristalização das amostras I (- arginina) e II (+ arginina) de LIC10494 aplicando-se a técnica de *microseeding*.

Uma terceira estratégia de refinamento consistiu na triagem de aditivos dos kits *Additive Screen* I e II (Hampton Research). Para tanto, gotas compostas por 2 μL de amostra II (3 mg/mL), 2 μL de solução do poço (1,6 M NaH_2PO_4 , 0,2 M K_2HPO_4 , 0,1 M

HEPES de sódio pH 8) e 0,4 μL de aditivo foram equilibradas contra 200 μL de solução do poço a 18 °C. O método empregado foi o da difusão de vapor em gota pendurada.

Coleta de dados teste

A difração de cristais (25 μm – 50 μm) de LIC10494 foi testada na linha W01B-MX2 do LNLS. Imagens de difração foram coletadas em um detector do tipo CCD (Mar Research) fixado a 100 mm do cristal. Para cada imagem, o cristal foi rotacionado em 1° ou 10° durante 60 s ou 240 s. Como solução crioprotetora utilizou-se a solução do poço acrescida de 20 % de etileno glicol.

RESULTADOS

Expressão e Purificação

A proteína LIC10494 foi satisfatoriamente produzida solúvel em células da linhagem C41(DE3) de *E. coli*. Em SDS-PAGE de amostras de LIC10494 recombinante, observa-se uma massa aparente significativamente maior do que sua massa teórica (25,9 kDa). Esse tipo de mobilidade anormal em SDS-PAGE é frequentemente observado para proteínas intrinsecamente desestruturadas que, por apresentarem composições de aminoácidos únicas, se ligam a menos moléculas de SDS e migram, portanto, mais lentamente do que proteínas globulares (Receveur-Bréchet *et al.*, 2006). Tal observação apóia a hipótese de que LIC10494 apresenta uma grande região intrinsecamente desestruturada.

Na primeira etapa de purificação, obteve-se um rendimento de 16 mg de proteína por litro de cultura. As frações mais puras e concentradas de LIC10494 foram agrupadas e dialisadas em tampão 50 mM Tris-HCl pH 8, 150 mM NaCl. Ensaio de DLS utilizando-se essa amostra mostraram que LIC10494 apresentava-se agregada (FIGURA 5). A alta polidispersividade da amostra (39 %) sugeriu a presença de vários estados oligoméricos de LIC10494. O raio hidrodinâmico médio calculado corresponde a uma massa molecular cerca de 15 vezes a massa de LIC10494.

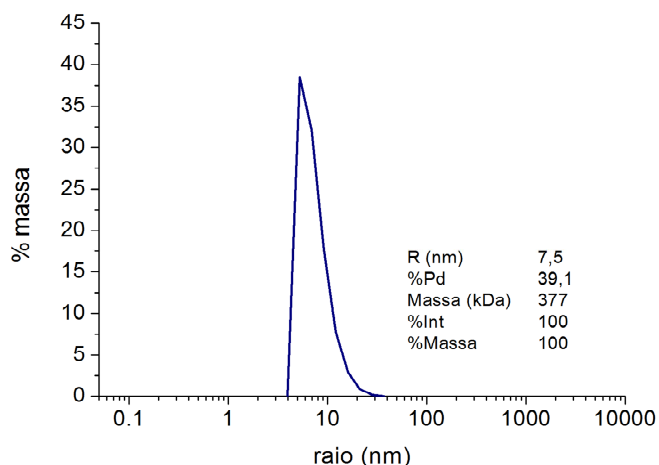


FIGURA 5. DLS de LIC10494 após a primeira etapa de purificação. Amostra a 4 mg/mL em 50 mM Tris HCl pH 8, 150 mM NaCl.

Assim, uma segunda etapa de purificação por exclusão molecular foi adotada. Esta foi suficiente para separar a fração não agregada de LIC10494 tanto de proteínas contaminantes quanto da fração agregada (FIGURA 6).

O rendimento final, considerando-se apenas a fração não agregada de LIC10494, (pico 2 da cromatografia de exclusão molecular), foi estimado em 2 mg de proteína por litro de cultura, cerca de 13 % do rendimento obtido após a primeira etapa de purificação. Constata-se, portanto, que a maior parte das moléculas de LIC10494 forma agregados nas condições empregadas para sua purificação.

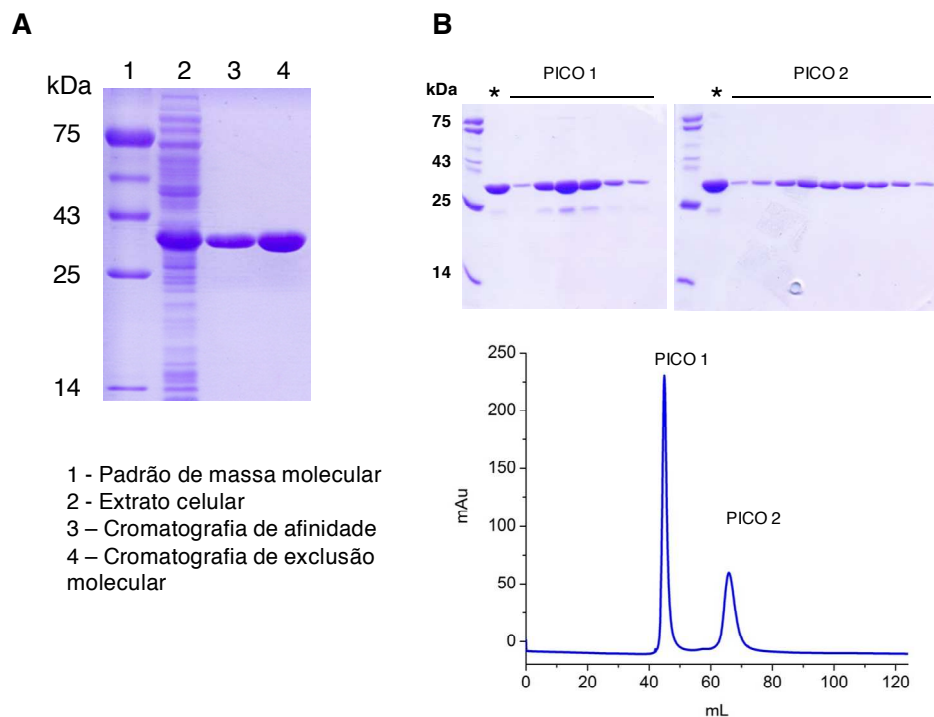
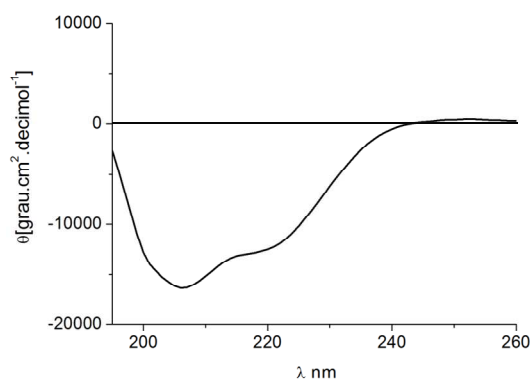


FIGURA 6. Purificação de LIC10494. A. SDS-PAGE 12 % de amostras coletadas a cada etapa de purificação de LIC10494 recombinante (25,9 kDa). B. SDS-PAGE 12 % de frações coletadas durante a eluição de LIC10494 em coluna Superdex™ 75 16/60 (GE Healthcare) e o respectivo cromatograma. * indica a amostra injetada na coluna.

Estrutura secundária de LIC10494

Analisando-se o espectro de dicroísmo circular de LIC10494, observou-se a presença de bandas negativas em 208 nm e 222 nm, evidenciando a presença de estruturas do tipo α -hélice (FIGURA 7). Além disso, observou-se que a banda em 208 nm é maior (em módulo) do que a banda em 222 nm, o que sugere uma considerável contribuição de estruturas do tipo *random coil* (banda negativa em 195 nm) no espectro de CD de LIC10494. Isso é evidenciado pela estimativa da quantidade relativa de estruturas secundárias de LIC10494 a partir da deconvolução do seu espectro de CD. De acordo com essa estimativa, aproximadamente 45 % de LIC10494 corresponde a regiões desestruturadas (FIGURA 7).



α -hélice	14 %	folhas β	23 %
volta β	18 %	outras	45 %

FIGURA 7. Dicroísmo Circular de LIC10494. Espectro de CD de LIC10494 (painel superior) e quantidade relativa de estrutura secundária de LIC10494 estimada pela deconvolução do espectro pelo programa CDNN (Bohm *et al.*, 1992) no intervalo de 195 nm a 260 nm (painel inferior).

A abundância de estruturas *random coil* em LIC10494 condiz com o modelo teórico de sua estrutura secundária, predito pelo programa PSIPRED (Jones, 1999, Bryson *et al.*, 2005), e com as previsões dos programas PONDR e FoldIndex, que indicam a presença de uma extensa região central intrinsecamente desestruturada em sua cadeia polipeptídica (FIGURA 2).

Comportamento hidrodinâmico de LIC10494

Como forma de se estimar o estado oligomérico da fração não agregada de LIC10494, realizou-se uma cromatografia de exclusão molecular analítica em coluna Superdex™ 75 16/60. Como mostrado na seção anterior, a amostra de LIC10494 se apresenta, basicamente, como uma mistura composta por agregados, que eluem no volume morto da coluna, e por uma fração não agregada (FIGURA 8).

A partir do K_{av} calculado para o segundo pico de LIC10494, estimou-se um Rh de 2,84 nm e 2,65 nm em experimentos realizados a 426 μ M e 155 μ M de LIC10494, respectivamente. Tais valores são maiores do que o Rh esperado para um monômero (2,19 nm) e menores do que o Rh esperado para um dímero (3,19 nm) de uma proteína globular de 25,9 kDa.

Proteínas intrinsecamente desordenadas apresentam uma mobilidade mais rápida em cromatografia de exclusão molecular em relação a proteínas globulares, visto que conformações estendidas resultam em dimensões hidrodinâmicas maiores (Receveur-Bréchet *et al.*, 2006). Assim, os Rhs estimados para LIC10494 sugerem que a fração não agregada corresponde a monômeros cujas formas desviam da globular.

Amostras da cromatografia realizada em coluna Superdex™ 75 10/30 foram monitoradas por DLS (FIGURA 9). Frações

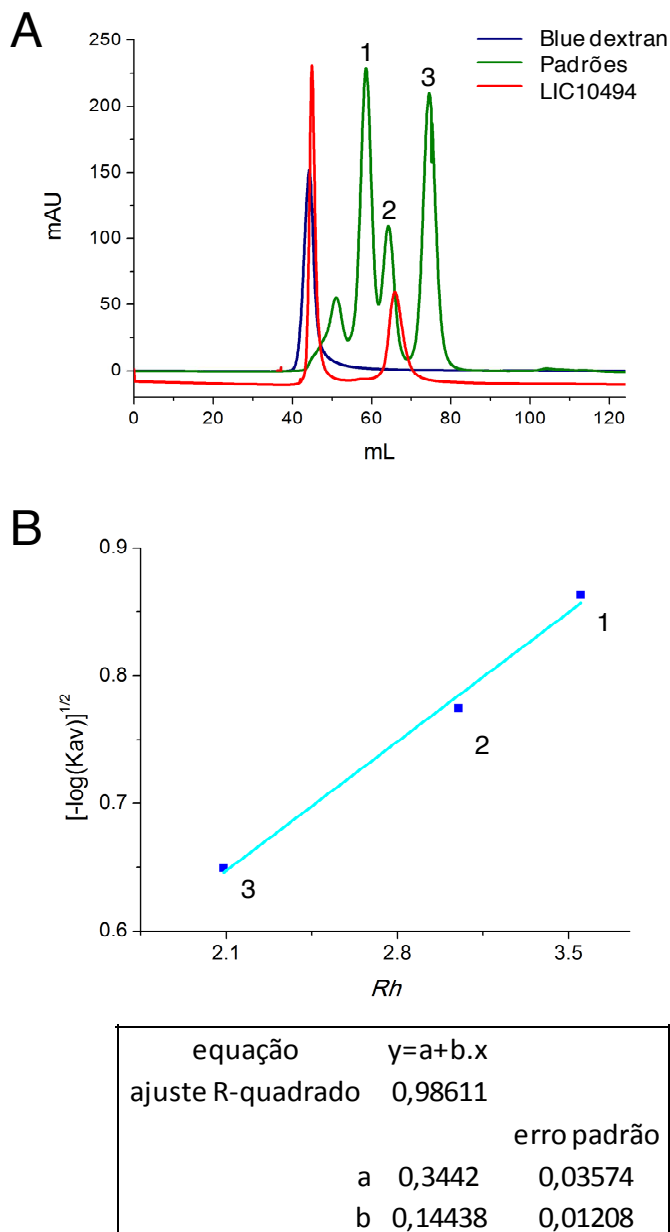


FIGURA 8. Cromatografia de exclusão molecular analítica de LIC10494 em coluna Superdex™ 75 16/60 para estimativa de seu estado oligomérico. A. Cromatograma da eluição de LIC10494 (426 μ M) (vermelho), Blue Dextran (azul) e das amostras padrão (verde): [1] albumina (3,55 nm), [2] ovalbumina (3,05 nm) e [3] quimiotripsinogênio (2,09 nm). B. Regressão linear dos dados experimentais obtidos a partir do cromatograma apresentado em A.

eluídas no volume morto da coluna (pico 1) apresentaram valores estimados para o Rh médio entre 7,4 nm e 10,3 nm, equivalentes a agregados de massas moleculares de 369 kDa a 792 kDa. Frações correspondentes à transição entre o primeiro e o segundo pico apresentaram alto índice de polidispersividade (30 % - 60 %), indicando sua grande heterogeneidade (FIGURA 9).

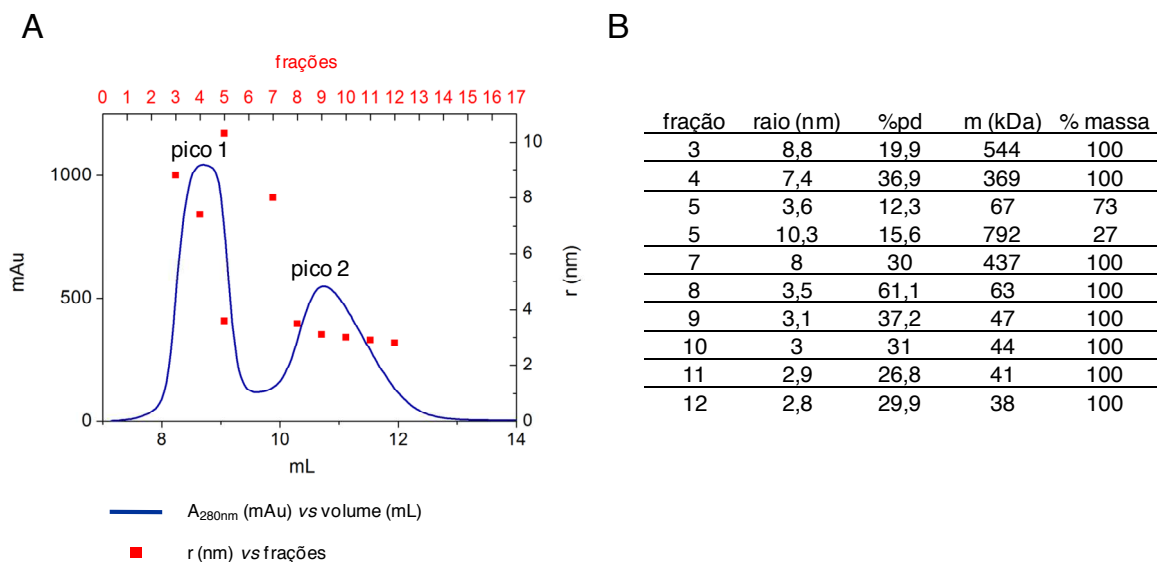


FIGURA 9. DLS das frações de LIC10494 eluídas da coluna Superdex™ 75 10/30. A. Cromatograma da eluição de LIC10494 da coluna Superdex™ 75 10/30 (azul). Pontos vermelhos indicam o raio hidrodinâmico médio r (nm) estimado para determinadas frações a partir de medidas de DLS. **B.** Dados obtidos a partir das medidas de DLS das frações mostradas em A. raio - raio hidrodinâmico (Rh) médio, %pd - polidispersividade da distribuição de tamanhos em porcentagem, m - massa estimada a partir do raio hidrodinâmico médio, % massa - fração em massa do respectivo componente em relação à distribuição total de tamanhos.

Para as frações eluídas no segundo pico do cromatograma, observaram-se Rhs entre 2,8 – 3,1 nm, equivalentes a massas moleculares entre 38 kDa – 47 kDa e comparáveis aos Rhs estimados por gel filtração analítica. Tais frações apresentaram um alto índice de polidispersividade (> 20 %), sugestivo de agregação de parte da amostra após a cromatografia de exclusão molecular (FIGURA 9B).

Após se agrupar as frações 9 a 12 do pico 2, mediu-se o DLS da amostra resultante contendo LIC10494 a 2 mg/mL. Esta apresentou um raio hidrodinâmico médio de 2,6 nm (31 kDa) e um índice de 46 % de polidispersividade. Após ser concentrada para 6 mg/mL, LIC10494 apresentou dois picos no DLS. Um pico monodisperso (12,6 % Pd) de raio

hidrodinâmico médio de 3,2 nm (50 kDa) e um pico de raio hidrodinâmico médio de 15 nm (1959 kDa) correspondendo à apenas 9 % da massa total. Com essa amostra, realizaram-se os primeiros ensaios de cristalização para LIC10494.

Ensaio para análise de agregação

Dada a grande proporção de proteínas agregadas em amostras de LIC10494 em tampão 50 mM Tris HCl pH 8, 150 mM NaCl, decidiu-se testar alguns aditivos descritos na literatura como sendo capazes de minimizar a agregação de proteínas (Bondos & Bicknell, 2003). Isso para melhorar tanto o rendimento da cromatografia de exclusão molecular quanto a polidispersividade das amostras submetidas à cristalização.

Como mostra a figura 10, dos 5 aditivos testados, apenas a arginina mostrou claramente ser capaz de aumentar a quantidade relativa de proteínas não agregadas. A redução do pH em 1 ponto (amostra 7) aparentemente não influenciou no equilíbrio agregado ↔ monômero (FIGURA 10).

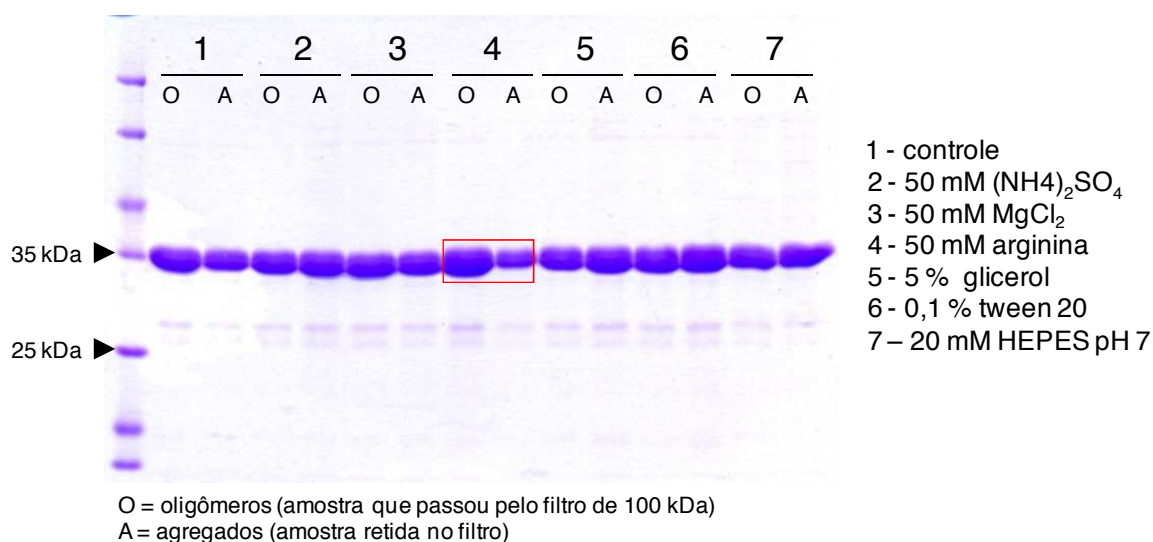


FIGURA 10. SDS-PAGE das amostras do ensaio para análise de agregação de LIC10494 em presença de aditivos. Em vermelho destaca-se o melhor resultado, observado no teste com arginina.

Embora a arginina seja amplamente empregada como um supressor de agregação de proteínas, especialmente em processos de renaturação, seu mecanismo de ação ainda não é

bem conhecido (Baynes *et al.*, 2005; Das *et al.*, 2007). Estudos sugerem que, em solução aquosa, a arginina forme agrupamentos que apresentam uma superfície hidrofóbica devido ao alinhamento de seus 3 grupos metileno. De acordo com esse modelo, a inibição da agregação de proteínas se dá pela interação da superfície hidrofóbica das mesmas com a superfície hidrofóbica dos agrupamentos de arginina (Das *et al.*, 2007).

Outro modelo proposto para explicar o efeito da arginina na supressão da agregação de proteínas baseia-se em duas observações (Arakawa, *et al.*, 2007). A primeira delas é a de que a arginina, similarmente ao hidrocloreto de guanidina (GdnHCl), interage favoravelmente com a maioria das cadeias laterais de aminoácidos. A segunda é a de que a ligação de arginina na superfície de proteínas é limitada, diferentemente da ligação de GdnHCl. Assim, postula-se que a limitada interação de arginina com a superfície de proteínas tenha um papel preponderante em sua habilidade de suprimir a agregação de proteínas sem desestabilizá-las (Arakawa, *et al.*, 2007).

Cromatografia de exclusão molecular em presença de arginina

Para se analisar quantitativamente o efeito da adição de 50 mM de arginina no equilíbrio agregado \leftrightarrow monômero em amostras de LIC10494, realizaram-se ensaios de cromatografia de exclusão molecular na presença e na ausência de arginina (FIGURA 11).

Como esperado, em amostras dialisadas e purificadas em tampão 50 mM Tris HCl pH 8, 150 mM NaCl, 50 mM arginina, a proporção da fração não agregada aumentou substancialmente (FIGURA 11). No cromatograma da amostra controle, observou-se que o pico da fração não agregada correspondeu a 48 % da área total sob a curva, ou seja, cerca de 50 % da amostra apresentou-se agregada em tampão 50 mM Tris HCl pH 8, 150 mM NaCl (FIGURA 11). Em presença de 50 mM de arginina, a proporção do segundo pico foi para 90 % da área total sob a curva, o que mostra que a adição de arginina inibiu drasticamente a agregação de LIC10494 (FIGURA 11).

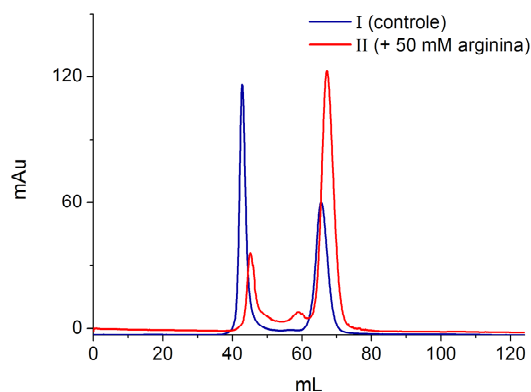


FIGURA 11. Inibição da agregação de LIC10494 com a adição de 50 mM de arginina no tampão de purificação.

As amostras I e II foram analisadas por DLS antes e depois da cromatografia de exclusão molecular (FIGURA 12). Em relação à amostra I, esta se apresentou polidispersa (39 % Pd) e com um raio hidrodinâmico médio estimado em 7,5 nm antes de ser purificada por cromatografia de exclusão molecular (FIGURA 12A). Com a adição de 50 mM de arginina (amostra II), o raio hidrodinâmico médio caiu para 3,7 nm e a polidispersividade aumentou para 69 %, indicando um aumento na proporção de moléculas não agregadas (FIGURA 12A).

Após ser purificada por cromatografia de exclusão molecular, a fração monomérica da amostra I foi concentrada para 6 mg/mL. A análise de DLS da amostra concentrada apresentou dois picos monodispersos, um com raio hidrodinâmico médio de 2,8 nm (36 kDa), correspondendo a 81 % da massa total de proteínas espalhadoras e um pico com raio hidrodinâmico médio de 9,4 nm (636 kDa) (FIGURA 12B). Na amostra contendo arginina, observou-se uma população monodispersa de moléculas de 3,3 nm (57 kDa) de raio hidrodinâmico médio correspondendo a praticamente 100 % em massa (FIGURA 12B).

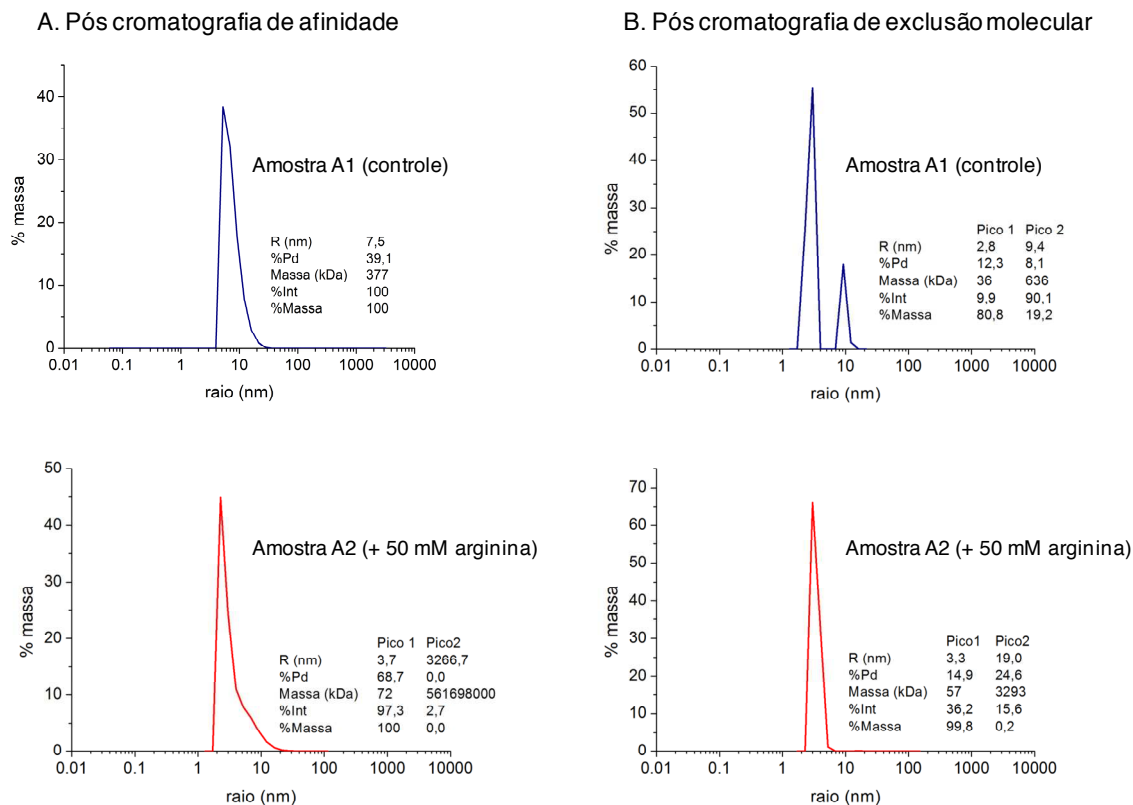


FIGURA 12. DLS das amostras I (controle) e II (+ arginina) de LIC10494 antes e depois da cromatografia de exclusão molecular.

Cristalização

Ensaio iniciais

Nos ensaios iniciais, foram testadas 576 condições de cristalização, utilizando-se amostras de LIC10494 a 6 mg/mL em tampão 50 mM Tris HCl pH 8, 150 mM NaCl. Na maioria das condições, observaram-se gotas claras. Cristais de tamanho $\leq 50 \mu\text{m}$ foram observados em três condições as quais foram selecionadas para serem refinadas. Em duas delas os cristais apresentaram-se no formato de agulhas. Tentativas de se testar a difração desses cristais na linha W01B-MX2 do LNLS foram limitadas pelas pequenas dimensões e pela fragilidade dos cristais observados.

Refinamento I

A primeira estratégia adotada para se refinar as condições que apresentaram resultados positivos nas triagens iniciais foi a variação da concentração dos precipitantes em função do pH. No refinamento da condição A (0,1 M Na_2HPO_4 /ácido cítrico pH 4,2, 0,4 M K_2HPO_4 e 1,6 M NaH_2PO_4), esferulitas ou agulhas foram observadas, principalmente, nas concentrações 1,8 M e 1,6 M NaH_2PO_4 (FIGURA 13). Cristais desse refinamento foram corados com IZIT, o que sugere que os mesmos são compostos de proteínas (FIGURA 13).

Tanto a quantidade de nucleação quanto o tamanho e a forma dos cristais aparentemente não variaram em função do pH e da diminuição da concentração de agentes precipitantes. Em todas as gotas que apresentaram cristais, observaram-se alta nucleação e, conseqüentemente, cristais muito pequenos ($< 25 \mu\text{m}$).

Ao longo do gradiente de concentração dos agentes precipitantes testado nesse ensaio, a transição entre condições que apresentaram cristais para gotas claras ocorreu abruptamente, sem serem observadas condições intermediárias que apresentassem, gradativamente, menor nucleação e cristais maiores. Isso sugere que, nas condições desse ensaio, o diagrama de fases de LIC10494 apresenta uma zona de nucleação muito estreita. No refinamento da condição B (20 % PEG3350, 0,2 M $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$), observou-se precipitado protéico gelatinoso na condição 20 % PEG3350, 0,2 M $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ na maioria dos pHs testados. Em concentrações de PEG3350 abaixo de 20 % só foram observadas gotas claras, salvo raras exceções que apresentaram precipitação leve. Agulhas foram observadas apenas na condição 20 % PEG3350, 0,2 M $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ em pH 5. Esferulitas foram observadas na condição 20 % PEG3350, 0,1 M $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ pH (7-9).

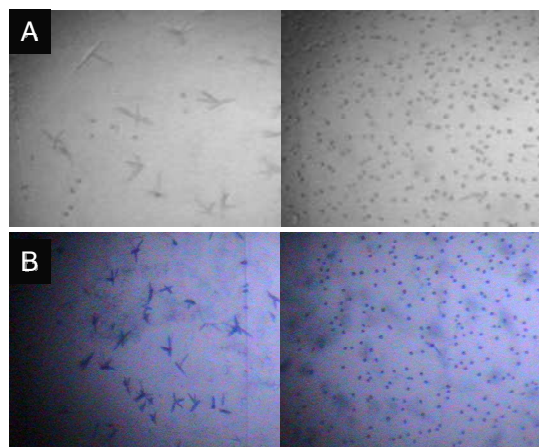


FIGURA 13. Cristais de LIC10494 observados nos primeiros ensaios de refinamento da condição A. No painel B observam-se os cristais do painel A corados com IZIT.

Em relação ao refinamento da condição C (0,1 M citrato de sódio pH 5,5, 2 M sulfato de amônio), observou-se precipitado protéico gelatinoso nos pHs 4 e 5 e esferulitas em pH ≥ 7 nas concentrações de sulfato de amônio 2 M e 1,8 M. No restante das condições desse refinamento, foram observadas gotas claras.

Refinamento II

A segunda estratégia de refinamento consistiu no uso de uma amostra de LIC10494 menos polidispersa, obtida pela adição de 50 mM de arginina no tampão da cromatografia de exclusão molecular. As amostras I (controle) e II (+ arginina) de LIC10494 foram submetidas a ensaios de cristalização nas condições A-D vs 1-12 do refinamento da condição A (0,1 M Na_2HPO_4 /ácido cítrico pH 4,2, 0,4 M K_2HPO_4 e 1,6 M NaH_2PO_4) (FIGURA 3).

Tanto no refinamento da amostra controle (I), quanto no refinamento da amostra

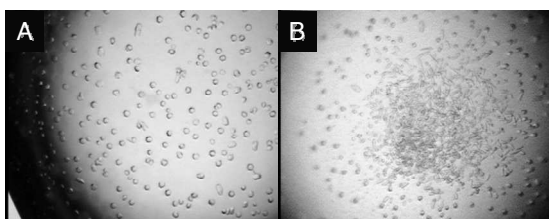


FIGURA 14. Cristais de LIC10494 observados no refinamento II. A. Amostra I. Condição - 1,8 M NaH_2PO_4 , 0,2 M K_2HPO_4 pH 8. B. Amostra II. Condição - 1,8 M NaH_2PO_4 , 0,3 M K_2HPO_4 pH 8.

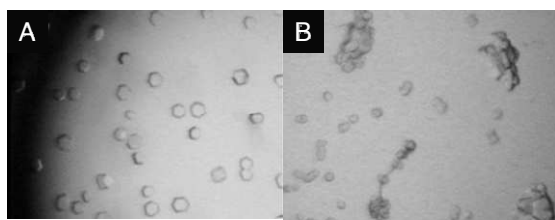


FIGURA 15. Cristais de LIC10494. A. Cristais da amostra I (25 μm). Condição 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 pH 7. B. Aglomerados de cristais da amostra II (50 μm). Condição - 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 pH 8.

contendo arginina (II), observou-se alto nível de nucleação e cristais muito pequenos ($\leq 25 \mu\text{m}$) na forma de agulhas/bastões ou esferas/hexágonos em quase todas as condições testadas em presença de 1,8 M NaH_2PO_4 (FIGURA 14). Em 1,6 M NaH_2PO_4 , observou-se uma leve diminuição na nucleação em função da diminuição da concentração de K_2HPO_4 na maioria dos pHs testados.

Cristais com a melhor morfologia foram observados no ensaio realizado com a amostra controle na condição 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 pH 7 (FIGURA 15A). As maiores formações cristalinas observadas foram aglomerados de 50 μm

crescidos na condição 1,6 M NaH_2PO_4 , 0,3 M K_2HPO_4 pH 8, do ensaio com a amostra II (FIGURA 15B).

Ensaio de cristalização realizados com a amostra II aplicando-se a técnica de *microseeding* renderam apenas gotas claras. Provavelmente, nas condições testadas, as gotas estavam num nível de saturação abaixo da curva de solubilidade de LIC10494. No ensaio realizado com a amostra I, observaram-se apenas esferulitas. Na triagem de aditivos dos kits *Additive Screen* I e II (Hampton Research), observou-se precipitado cristalino na maioria das condições. O resultado mais promissor foi obtido com o aditivo cloreto de zinco (10 mM). Nessa condição foram observados aglomerados de cristais ($50 \mu\text{m}$).

Coleta de dados teste

A difração de cristais de LIC10494 foi testada na linha W01B-MX2 do LNLS. Todos apresentavam tamanhos $\leq 50 \mu\text{m}$ e não difrataram, mesmo quando expostos à radiação por um longo período de tempo. Como o brilho da linha de luz W01B-MX2 não é favorável para a coleta de pequenos cristais, o tamanho dos cristais de LIC10494 torna-se um fator limitante. Além da limitação de tamanho, os cristais de LIC10494 provavelmente não possuem uma rede cristalina bem ordenada. A presença de uma grande região intrinsecamente desestruturada em LIC10494 deve ser o principal fator limitante para a formação de cristais adequados para o experimento de difração.

4. DISCUSSÃO

4.1. ESTUDOS ESTRUTURAIS DA PROTEÍNA LP49

4.1.1. Cristalização de Lp49

Como o artigo referente à Lp49 (seção 3.1) não aborda as dificuldades encontradas na etapa de cristalização dessa proteína, convém apresentar e discutir brevemente alguns resultados dessa etapa bem como as estratégias escolhidas na tentativa de se obter cristais adequados para o experimento de difração.

Nos ensaios iniciais de cristalização de Lp49, diversas condições apresentaram cristais, os quais se caracterizaram como placas, geralmente finas e de morfologia irregular, crescidas conectadas umas às outras (FIGURA 4.1).

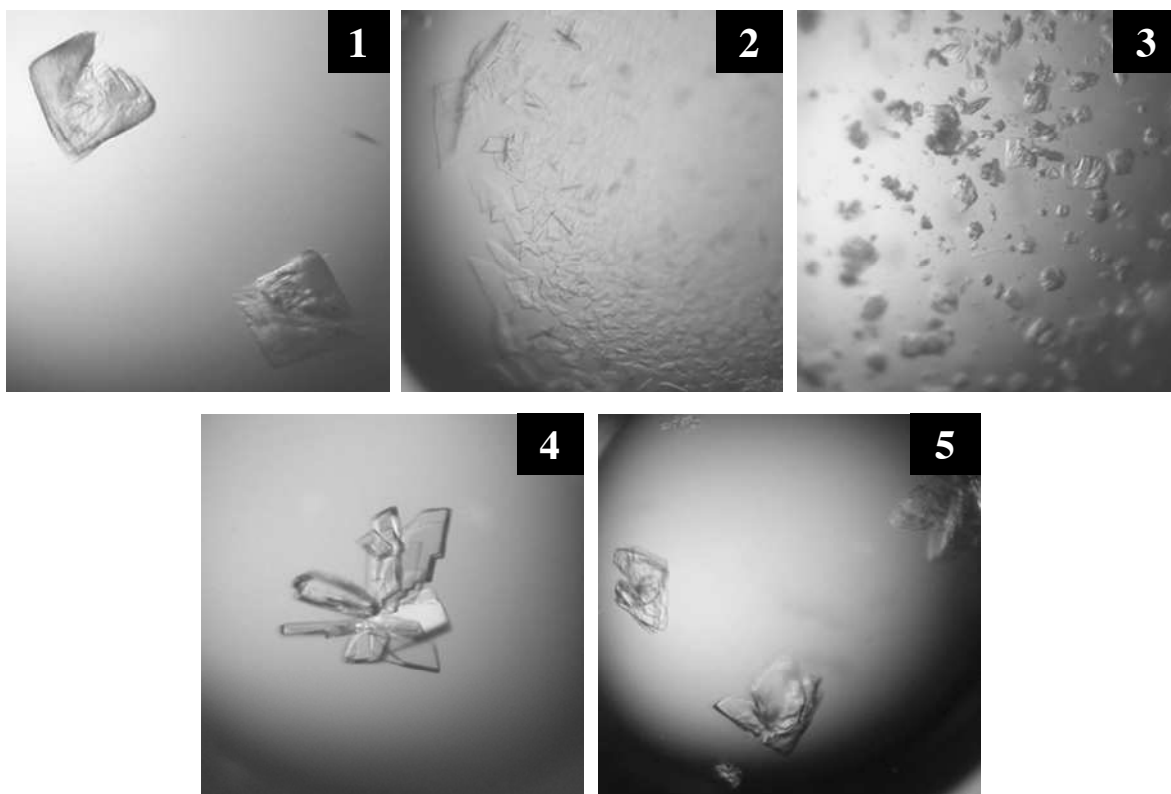


FIGURA 4.1. Cristais de Lp49 observados em condições dos ensaios iniciais. 1) 0,1 M HEPES de sódio pH 7,5, 1,5 M Li_2SO_4 ; 2) 0,1 M Tris HCl pH 8,5, 30 % PEG4000, 0,2 M Li_2SO_4 ; 3) 0,1 M HEPES de sódio pH 7,5, 2 % PEG400, 2 M $(\text{NH}_4)_2\text{SO}_4$; 4) 0,1 M HEPES de sódio pH 7,5, 20 % PEG3000, 0,2 M NaCl; 5) 0,1 M Tris HCl pH 8,5, 20 % PEG2000 MME, 0,01 M NiCl_2 .

O refinamento de algumas dessas condições foi realizado variando-se a concentração de agente precipitante em função da variação do pH. Embora ele tenha sido eficiente na busca por condições que propiciassem o maior crescimento dos cristais, não foi suficiente para se determinar condições que favorecessem o crescimento de monocristais isolados e de boa qualidade de difração.

Sendo assim, realizou-se um segundo ciclo de refinamento, baseado em novas estratégias de otimização. Estas incluíram o uso de aditivos (Hampton Research), o controle da taxa de difusão de vapor por meio de aplicação de óleo no poço (Chayen, 1997) e o emprego da técnica de *microseeding* (Bergfors, 2003). Destas, a técnica de *microseeding* apresentou-se como a melhor estratégia para a obtenção de cristais isolados de boa qualidade de difração (FIGURA 4.2).

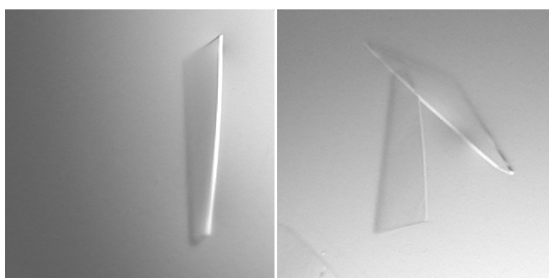


FIGURA 4.2. Cristais de Lp49 crescidos na condição 17,5 % PEG3000, 0,175 M NaCl, 0,09 M Tris HCl pH 8,5, aplicando-se a técnica de “microseeding”. Cristal da esquerda (500 μ m x 100 μ m). Cristais da direita (250 μ m x 100 μ m).

A taxa de crescimento de um cristal, assim como o processo de nucleação, está diretamente relacionada com o nível de supersaturação em que a proteína se encontra, de forma que, quanto maior a supersaturação, maiores são as taxas de nucleação e de crescimento de cristais e, conseqüentemente, maior é a probabilidade do surgimento de imperfeições na superfície cristalina que podem tanto limitar o crescimento quanto favorecer a ocorrência de cristais múltiplos (Blow *et al*, 2004).

Através da técnica de *microseeding* é possível desvincular o processo de nucleação, que ocorre em condições de alta supersaturação, do processo de crescimento, que ocorre mais lentamente em condições de baixa supersaturação (Bergfors, 2003). A importância de se desvincular esses dois eventos consiste no fato de que, em condições de baixa supersaturação, que compreendem a zona metaestável num diagrama de solubilidade, o processo de nucleação é inibido, porém, o processo de crescimento de núcleos já formados

ainda ocorre em taxas lentas, o que favorece a formação de cristais bem ordenados (Blow *et al.*, 2004). Dessa forma, estabeleceu-se como protocolo padrão para a cristalização de Lp49, a aplicação da técnica de “*microseeding*” na condição 17,5 % PEG3000 0,175 M NaCl 0,09 M Tris HCl pH 8,5.

Ensaio realizado com a proteína marcada com SeMet a 6 mg/mL na condição de cristalização da proteína nativa, usando-se a técnica de “*microseeding*”, não resultaram em cristais. Sendo assim, optou-se por buscar uma nova condição de cristalização, através de ensaios com o kit “*Crystal Screen I*” (Hampton). Em 4 condições dessa triagem, foram observados pequenos cristais e uma alta taxa de nucleação, 1 dia após a confecção da placa.

Duas dessas condições (0,1 M Tris HCl pH 8,5, 0,2 M LiSO₄, 30 % PEG4000 e 0,1 M cacodilato de sódio pH 6,5, 0,2 M Mg₂(C₂H₃O₂), 20 % PEG8000) foram refinadas combinando-se a técnica de “*microseeding*” com a diminuição da concentração de PEG. Dessa forma, obtiveram-se cristais semelhantes aos nativos na condição 0,1M Tris HCl pH 8,5 0,2M LiSO₄ 15 % PEG4000 semeada com o estoque de sementes diluído a 10⁻³ (FIGURA 4.3).

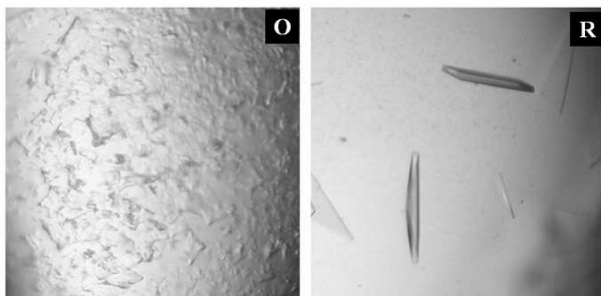


FIGURA 4.3. Cristais de Lp49 marcada com SeMet. Legenda: O: cristais crescidos em 0,1 M Tris HCl pH 8,5, 0,2 M LiSO₄, 30 % PEG4000. R: cristais crescidos em 0,1 M Tris HCl pH 8,5, 0,2 M LiSO₄, 15 % PEG4000, aplicando-se “*microseeding*”.

4.1.2. A estrutura tridimensional do antígeno leptospiral Lp49 sugere uma provável função de ligação a proteínas

A estrutura cristalográfica de Lp49 revelou a presença de dois domínios, ambos pertencentes à classe “*all β-proteins*” do SCOP (Lo Conte *et al.*, 2000). O domínio N-terminal de Lp49 possui um enovelamento do tipo Imunoglobulina (*Ig-like fold*) e uma topologia provavelmente derivada da forma B do arranjo *Greek key* (Zhang and Kim, 2000). Sua sequência é altamente divergente (identidade de sequência em torno de 10 %

em alinhamentos estruturais) e sua estrutura é remotamente relacionada a membros de superfamílias distintas, o que sugere que ele pode ter evoluído de um ancestral independente, sendo, portanto, membro de uma nova superfamília.

Na classificação SCOP, proteínas são agrupadas na mesma superfamília quando a similaridade estrutural e, em muitos casos, a similaridade funcional, sugerem uma provável origem evolutiva comum (Lo Conte et al., 2000). É interessante observar que domínios que apresentam o enovelamento do tipo Imunoglobulina são divididos, atualmente, em 28 superfamílias distintas, o que evidencia a dificuldade de se estabelecer relações evolutivas entre os mesmos.

Embora buscas por similaridades locais não tenham fornecido nenhum indício sobre a função do domínio N-terminal de Lp49, seu enovelamento sugere um papel na interação com outras proteínas. Domínios *Ig-like* são recorrentemente observados como uma plataforma para interações com proteínas, por exemplo, no sistema imune de vertebrados (como os receptores de antígeno e de citocinas) (Barclay, 2003) e em moléculas de adesão expostas na superfície das células (como a proteína VCAM-1, caderina e policistina-1) (Humphries and Newham, 1998; Ibraghimov-Beskrovnaya *et al.*, 2000). Em bactérias, domínios *Ig-like* já foram encontrados em fatores de virulência que medeiam sua aderência a células do hospedeiro e fatores de invasão, como as proteínas intimina (Luo *et al.*, 2000) de *E. coli* e invasina de *Yersinia pseudotuberculosis* (Hamburger *et al.*, 1999).

O domínio C-terminal de Lp49 apresenta similaridade estrutural significativa com domínios encontrados em proteínas da camada S de arqueobactérias e de proteínas encontradas em metazoários. Similaridades estruturais e sequenciais indicam que ele deve ser um membro da superfamília *YVTN repeat-like/Quinoprotein amine dehydrogenase*.

Apesar da similaridade global, o β -*propeller* de Lp49 é menos regular do que seu homólogo YVTN β -*propeller*. As folhas β do β -*propeller* de Lp49 não compartilham motivos altamente conservados, apesar de se observar consensos em algumas posições da sequência, principalmente nas fitas 2 e 3 (Fig. 3, artigo I, p. 18). Além disso, o domínio C-terminal de Lp49 apresenta longas inserções entre as fitas 2 e 3 das pás W4 e W6, respectivamente (Fig. 3, artigo I, p. 18). Essas inserções formam *loops* proeminentes na face superior do *propeller*, não observados nos domínios YVTN β -*propeller* e WD40 de Groucho/TL1, e podem ter algum papel funcional em Lp49.

A maioria dos resíduos conservados no β -*propeller* de Lp49 e nos domínios a ele relacionados apresenta um papel na estabilidade estrutural dos mesmos. Resíduos conservados e expostos ao solvente apresentam-se dispersos, o que dificulta a identificação de resíduos potencialmente funcionais (Fig. 5, artigo I, p. 19).

Predizer a função de um domínio do tipo β -*propeller* é um grande desafio, visto que esse *fold* é caracterizado por uma diversidade funcional e sequencial marcante, apesar de sua rigidez estrutural (Fülöp and Jones, 1999). O domínio homólogo ao domínio C-terminal de Lp49, o YVTN β -*propeller*, pertence a uma proteína de camada S (SLP) que foi localizada em interfaces intercelulares e pode estar envolvida em processos de adesão em *Methanosarcina mazei* (Jing *et al.*, 2002; Macario and Conway De Macario, 2001). É interessante observar que essa SLP de archaea é também composta por um domínio 7-bladed β -*propeller* e um domínio *Ig-like*. Contudo, o domínio N-terminal de Lp49 não apresenta uma similaridade estrutural significativa ao domínio *Ig-like* desta SLP. Apesar de não se ter observado ainda uma camada S em *L. interrogans*, o fato de pelo menos três proteínas com homologia a SLPs (Lp49, LIC10868 e LIC12952 (Nascimento *et al.*, 2004b)) terem sido identificadas pode sugerir a formação de camada S em leptospiros.

Análogos do domínio β -*propeller* que pertencem à superfamília WD40 *repeat* fazem parte dos co-repressores de transcrição Groucho/TL1 (humano) (Pickles *et al.*, 2002) e TUP1 (fungo) (Sprague *et al.*, 2000) e da proteína humana LIS1 (Tarricone *et al.*, 2004). Em todas essas proteínas, os domínios WD40 estão envolvidos em interações proteína-proteína. Uma função de ligação a proteínas também já foi descrita para o domínio análogo presente na proteína Brat (Edwards *et al.*, 2003), membro da superfamília NHL *repeat*, e no domínio análogo YWTD, presente no componente de matriz extracelular chamado nidogênio (Takagi *et al.*, 2003).

Três possíveis sítios de ligação a proteínas foram identificados na estrutura de Lp49, segundo análises do programa PPI-Pred (Bradford and Westhead, 2005). O mais provável localiza-se ao redor da cavidade central da superfície superior do β -*propeller* (Fig. 4A, artigo I, p. 19). Interessantemente, esse resultado concorda com a localização dos sítios de ligação a proteínas descritos para os análogos do domínio C-terminal de Lp49, que também se encontram na face superior do β -*propeller* (Pickles *et al.*, 2002; Sprague *et al.*, 2000; Tarricone *et al.*, 2004; Edwards *et al.*, 2003; Takagi *et al.*, 2003).

A superfície eletrostática de Lp49 mostra que a face inferior do β -*propeller* é predominantemente eletronegativa em contraste com a face superior que revela uma região central eletropositiva (Fig. 4B, artigo I, p. 19). Essa região eletropositiva coincide parcialmente com o sítio mais provável de interação proteína-proteína, podendo ter implicações na formação de complexos protéicos, já que aminoácidos carregados podem aumentar significativamente as taxas de associação de complexos proteína-proteína, mesmo que a etapa limitante envolva a desolvatação de grupos polares e carregados (Sheinerman *et al.*, 2000).

É importante mencionar que a região central eletropositiva, presente na face superior do β -*propeller* de Lp49, compreende uma cavidade que apresenta similaridades com o sítio ativo da enzima *3-Carboxy-cis,cis-Muconate Lactonizing* de *Neurospora crassa* (NcCMLE) (Kajander *et al.*, 2002). A proteína NcCMLE também possui o enovelamento do tipo *7-bladed β -propeller* e seu sítio ativo está localizado na face superior do mesmo. Ela compõe a via β -keto adipato, responsável pela degradação de compostos aromáticos.

Em NcCMLE, a conversão de *cis,cis*-muconatos em muconolactonas é realizada pelos resíduos catalíticos His148 e Glu212. Os mesmos são substituídos pelos resíduos Arg280 e Met345 em Lp49 (FIGURA 4.4). Apesar de não conservar os resíduos catalíticos, Lp49 conserva os resíduos R196 e R274 que participam da ligação ao substrato em NcCMLE (FIGURA 4.4). Similarmente a NcCMLE, a superfície eletrostática positiva da cavidade presente na face superior do β -*propeller* de Lp49 pode ser capaz de atrair e ligar moléculas negativamente carregadas semelhantes aos muconatos ou às muconolactonas. A presença de um sulfato na cavidade de Lp49 apóia essa hipótese.

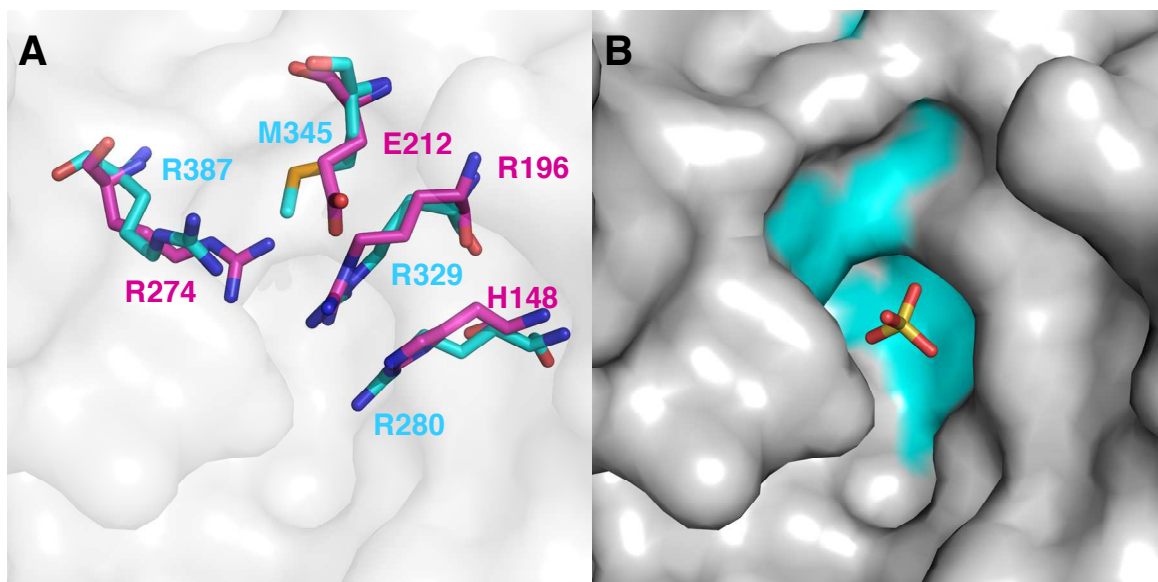


FIGURA 4.4. Comparação estrutural entre a cavidade eletropositiva da face superior do β -propeller de Lp49 e o sítio ativo da enzima NcCMLE. A. Representação dos resíduos do sítio ativo de NcCMLE (rosa) e seus correspondentes em Lp49 (azul). A superfície de Lp49 é mostrada em branco. B. Representação da superfície de Lp49 destacando-se o íon sulfato (*stick*) observado na estrutura cristalográfica de Lp49 ligado à cavidade da face superior do β -propeller. Em azul estão destacados os resíduos de Lp49 apresentados no painel A. As figuras A e B apresentam-se na mesma orientação.

Com base nos indícios funcionais obtidos a partir das análises da estrutura de Lp49, na provável localização dessa proteína na membrana externa e no fato dela ser um antígeno ausente em *L. biflexa* (Neves *et al.*, 2007; Picardeau *et al.*, 2008), nós postulamos que Lp49 desempenha uma função que envolve a interação com outras proteínas e que pode ter um papel nas interações entre *L. interrogans* e seus hospedeiros.

4.2. ESTUDOS ESTRUTURAIS DA PROTEÍNA LIC12922

4.2.1. Cristalização de LIC12922

Como o manuscrito referente à LIC12922 (seção 3.2) não aborda as dificuldades encontradas na obtenção de cristais dessa proteína que difratassem à alta resolução, convém apresentar e discutir brevemente alguns resultados dessa etapa bem como as estratégias escolhidas na tentativa de se obter cristais adequados para o experimento de difração.

LIC12922 cristalizou em inúmeras condições das triagens iniciais de condições de cristalização. Diversas delas foram refinadas variando-se a concentração dos agentes precipitantes em função da variação do pH. Tal refinamento resultou em cristais maiores e aparentemente bem formados. Porém, ao se testar a difração dos mesmos, observou-se que a maioria difratava entre 7 e 10 Å de resolução. Uma das hipóteses levantadas para se explicar a baixa qualidade e a fragilidade dos cristais de LIC12922 foi a de que sua cauda poli-His N-terminal poderia estar atrapalhando o empacotamento cristalino, por ser, provavelmente, uma região muito flexível. Dada a ausência de um sítio protease-específico para a remoção desta cauda, optou-se por tentar removê-la por meio de proteólise limitada.

Parte da cauda de LIC12922 foi removida pela protease tripsina, como confirmado por experimentos de *western blot* anti-His₆ e espectrometria de massas (dados não mostrados). Porém, a remoção de parte da cauda de histidina não foi suficiente para se melhorar a resolução máxima de difração dos cristais de LIC12922. Assim, diversas técnicas tais como otimização da solução crioprotetora, desidratação, ligação cruzada com glutaraldeído e *annealing*, conhecidas por melhorarem a difração de cristais de proteínas, foram testadas (Heras & Martin, 2005). De todas elas, a que apresentou melhores resultados foi a incubação dos cristais por 2 min em soluções crioprotetoras contendo 10 % de etileno glicol. Nesse teste, cristais difrataram a 4 Å de resolução.

Paralelamente, variantes de LIC12922 foram subclonadas no vetor pET-TEV (Carneiro *et al.*, 2006). Estas incluíram construções onde a região codificante para a porção C-terminal desestruturada ou para todo o segmento C-terminal de LIC12922 foram deletadas. Além disso, a sequência que codifica para a proteína LIC12922 sem o peptídeo sinal também foi sub-clonada no vetor pET-TEV, permitindo, assim, que a cauda de histidina fosse totalmente removida pela protease TEV. Dessas variantes, apenas a proteína inteira (sem o peptídeo sinal) foi satisfatoriamente expressa em *E. coli*. Entretanto, a remoção total da cauda de histidinas pela protease TEV também não foi suficiente para se aumentar a resolução de difração dos cristais de LIC12922. Assim, continuou-se o trabalho com a proteína recombinante original submetida à proteólise limitada.

Outras estratégias de refinamento incluíram o uso de aditivos (Hampton Research), a aplicação da técnica de *microseeding* e a cristalização em gel (Bergfors, 2003; Biertümpfel *et al.*, 2002). Esta última técnica consiste na adição de agarose de baixo ponto de fusão

(Sigma-Aldrich) na condição de cristalização a ser adicionada na gota. No caso de LIC12922, as gotas de cristalização foram compostas pela condição de cristalização contendo 0,2 % de agarose de baixo ponto de fusão, previamente aquecida a 40 °C, e pela solução de proteína, misturadas na proporção de 1:1. A condição de cristalização suplementada com agarose era inicialmente depositada na lamínula para que voltasse à temperatura ambiente antes da adição da solução de proteína. O método de cristalização usado foi a difusão de vapor em gota pendurada.

A combinação da técnica de *microseeding* com a cristalização em gel rendeu os melhores resultados. A técnica de *microseeding* permite uma taxa de crescimento mais lenta, favorecendo a formação de cristais bem ordenados (Bergfors, 2003). A técnica de cristalização em gel previne a sedimentação dos cristais e reduz a convecção na gota de cristalização, permitindo um ambiente mais estável para o crescimento cristalino (Biertümpfel *et al.*, 2002). Aplicando-se as duas técnicas, obteve-se um alto rendimento de cristais com tamanho médio na faixa de 200 μm a 400 μm , tanto para a proteína nativa quanto para a proteína marcada com SeMet (FIGURA 4.5).

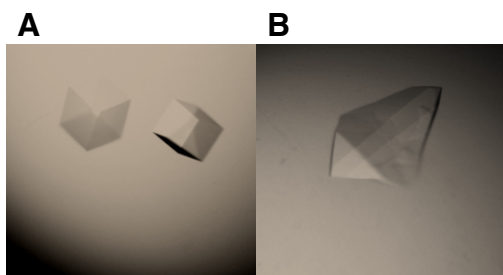


FIGURA 4.5. Exemplos de cristais de LIC12922 observados nos refinamentos em que se aliaram a técnica de *microseeding* e a cristalização em gel. A. cristais da proteína marcada com SeMet (200 μm). B. Cristal da proteína nativa (520 μm).

A cristalização em gel favoreceu também a manipulação dos cristais, por evitar a sua sedimentação. A quantidade de cristais danificados no momento da pesca reduziu drasticamente após a adoção desse método.

A combinação de ambas as técnicas contribuiu para uma melhoria na resolução máxima de difração dos cristais de LIC12922. Em relação à proteína marcada com SeMet, cristais crescidos em condições onde apenas a técnica de *microseeding* (Bergfors, 2003) foi aplicada apresentaram uma difração máxima que variou de 4,2 Å a 7,6 Å. Cristais crescidos aplicando-se a técnica de *microseeding* (Bergfors, 2003) aliada à cristalização em gel apresentaram uma difração máxima que variou entre 3,5 Å e 4,7 Å. Para os cristais nativos,

crescidos em condições onde apenas a técnica de *microseeding* foi aplicada, a difração máxima variou de 3,5 Å a 6,1 Å. Após o uso do método de cristalização em gel, a difração máxima passou a variar entre 3,1 Å e 3,6 Å. A otimização da condição crio protetora também teve papel importante nas melhorias observadas, bem como na reprodutibilidade dos resultados.

Centenas de cristais de LIC12922 foram testados a fim de se obterem dados com a maior resolução possível. Nesses testes, observou-se que a resolução máxima dos dados era diretamente proporcional ao tamanho dos cristais. Entretanto, um fator limitante para a coleta de dados de cristais maiores de 200 μm foi a formação de gelo durante o processo de criogenização. Assim, diferentes soluções crioprotetoras foram testadas.

Existem relatos na literatura de que alguns agentes crioprotetores ajudaram a melhorar a difração de cristais de proteína como, por exemplo, PEG 3350 (ou PEG 4000) e a mistura de etileno glicol com glicerol (Heras & Martin, 2005; Petock *et al.*, 2001; Hunsicker-Wang *et al.*, 2005). Por isso, essas condições também foram incluídas no teste. Os melhores resultados foram observados para a solução do poço adicionada de 15 % de etileno glicol e 10 % de glicerol.

4.2.2. A estrutura tridimensional da proteína hipotética conservada LIC12922 revela homologia com a chaperona periplasmática SurA.

A estrutura cristalográfica de LIC12922 revelou que a mesma apresenta uma forma alongada e é composta por 3 segmentos: um N-terminal (N), rico em alfa hélices, um domínio com enovelamento do tipo parvulina (P) e uma porção C-terminal (C) composta de uma longa alfa hélice seguida de duas fitas beta, uma alfa hélice e uma extensão desestruturada (Fig. 1, artigo II, p. 33). Os segmentos N e C formam o domínio NC. Nesse domínio, observa-se uma fenda com prevalência de resíduos hidrofóbicos e uma profundidade média de 16 Å (Fig. 1, artigo II, p. 33). O domínio parvulina (P) é composto por uma folha beta antiparalela de 4 fitas circundada por 5 alfa-hélices. Curtas extensões o conectam aos segmentos N e C. Domínios que possuem enovelamento do tipo parvulina geralmente são capazes de catalisar a isomerização *cis-trans* de ligações peptídil proilil, tendo um papel importante no processo de enovelamento das proteínas (Wedemeyer *et al.*, 2002).

Para se verificar se o domínio parvulina de LIC12922 possui atividade de peptidil prolil isomerase (PPIase), realizou-se um ensaio de PPIase acoplado à protease (Kofron *et al.*, 1991). A ligação peptidil prolil do substrato *N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide* é mais rapidamente convertida para a conformação *trans* na presença de uma PPIase. Essa conversão torna-o susceptível à clivagem pela quimotripsina, levando à formação do produto 4-nitroanilina, que absorve luz a 390 nm (Kofron *et al.*, 1991). Como mostrado na figura 3 (artigo II, p. 36), LIC12922 não apresentou atividade peptidil prolil isomerase, mesmo estando em grande excesso molar em relação à proteína usada como controle positivo.

Como podemos notar no alinhamento estrutural entre o domínio parvulina de LIC12922 e seus homólogos, LIC12922 não apresenta nenhum dos resíduos catalíticos descritos para a PPIase Pin1 (Behrsin *et al.*, 2007) (Fig. 4, artigo II. p. 38). O resíduo de cisteína C113 de Pin1 é substituído por um ácido aspártico em LIC12922 (D245), mas essa substituição ocorre em outras parvulinas sem causar perda de função. Entretanto, o suposto resíduo catalítico H59 de Pin1 é substituído pelo hidrofóbico resíduo de isoleucina I201 em LIC12922, o que pode ter contribuído para a perda de atividade PPIase da mesma. Além disso, apesar dos resíduos que formam o sítio de ligação à prolina identificados em Pin1 serem parcialmente conservados em LIC12922, os resíduos R290, R292 e Y266 ocluem o potencial sítio de ligação ao substrato. Tais indícios estruturais podem explicar porque o domínio parvulina de LIC12922 não apresenta a função PPIase.

A falta de atividade PPIase é fato recorrente em domínios parvulinas de chaperonas periplasmáticas. Exemplos disso são o domínio P1 da proteína SurA (Xu *et al.*, 2007) e o domínio parvulina da proteína PpiD (Weininger *et al.*, 2010), ambas de *E. coli*. Apesar de não apresentarem atividade catalítica, esses domínios ainda possuem a habilidade de se ligarem a peptídeos. Sugere-se para o domínio P1 de SurA um papel no reconhecimento do substrato (Xu *et al.*, 2007).

O domínio NC de LIC12922 apresenta similaridade estrutural com o domínio NC da chaperona periplasmática SurA (Bitto & MacKay, 2002) e com o domínio C-terminal da chaperona citoplasmática *trigger factor* de *E. coli* (Ferbitz *et al.*, 2004), o que apóia a hipótese da possível função chaperona de LIC12922. A posição e a orientação das hélices variam entre o domínio NC de LIC12922 e as estruturas a ele relacionadas, o que sugere

uma mobilidade estrutural intrínseca (Fig. 2, artigo II, p. 35). Análises do fator B de deslocamento atômico do modelo de LIC12922 apóiam a hipótese de flexibilidade estrutural, que pode ter implicações no mecanismo funcional da proteína.

Além disso, a superposição estrutural de LIC12922 nos modelos atualmente disponíveis para SurA (PDB 1M5Y, 2PV3) revelam que a orientação dos domínios parvulina em relação ao domínio NC é notavelmente variável. Embora estruturalmente e sequencialmente mais semelhante ao domínio P1 de SurA, o domínio parvulina de LIC12922 adota uma posição relativa ao domínio NC mais próxima da adotada pelo domínio P2 de SurA. Postula-se que a flexibilidade entre os domínios parvulina e NC também pode ter um papel no mecanismo de ligação ao substrato (Fig. 6, artigo II, p. 42).

Análises da estrutura de LIC12922 com o programa PPI-Pred (Bradford & Westhead, 2005) indicam a presença de uma extensa superfície potencialmente envolvida na ligação a proteínas que inclui os domínios NC e parvulina (Fig. 5, artigo II, p. 40). Uma das interfaces proteína-proteína observadas no empacotamento cristalino de LIC12922 conecta os prováveis sítios indicados pelo programa PPI-Pred, sugerindo que, similarmente ao mecanismo descrito para o *trigger factor* (Merz *et al.*, 2008), a ligação ao substrato deve envolver toda a superfície côncava de LIC12922.

Análises filogenéticas aliadas à comparação da arquitetura de domínios de LIC12922 e das chaperonas PrsA, PpiD e SurA sugerem que essas proteínas evoluíram de um ancestral comum. Em todas elas é possível identificar o segmento N, seguido de 1 ou 2 domínios parvulina e o segmento C (Fig. 7, artigo II, p. 43). A localização extracitoplasmática confirmada para PrsA, PpiD e SurA bem como a similaridade de suas funções apóiam essa hipótese. Com base nesses indícios e na provável localização de LIC12922 no periplasma, postulamos que LIC12922 seja uma chaperona periplasmática.

LIC12922 compartilha uma relação filogenética próxima com SurA, segundo o cladograma apresentada na figura 7 (artigo II, p. 43). Assim, postula-se que LIC12922, assim como sua ortóloga, esteja envolvida na biogênese de proteínas de membrana externa. A hipótese de que LIC12922 seja uma chaperona periplasmática envolvida na biogênese de OMPs em *L. interrogans* é apoiada pelo fato de que nenhum outro homólogo de SurA é encontrado no genoma leptospiral (Xue *et al.*, 2009). Como sabemos de estudos realizados em *E. coli*, a proteína SurA atua no enovelamento de diversas proteínas de membrana

externa, incluindo as porinas. Mutações de perda de função simultâneas de SurA e de PpiD, outra chaperona periplasmática, são letais, indicando que SurA é responsável por uma atividade essencial em *E. coli* codificada por genes redundantes (Bitto & MacKay, 2002).

Homólogos de LIC12922 foram identificados em todas as bactérias espiroquetas cujos genomas já foram seqüenciados (Fig, 7, artigo II, p. 43). A conservação desse gene sugere que ele desempenhe uma função essencial nessas bactérias. Confirmando-se que LIC12922 apresente uma função similar à de SurA, abre-se um amplo caminho para o estudo da biogênese de proteínas de membrana externa em todo o filo Espiroqueta.

Com o objetivo de testar a hipótese de que LIC12922 apresenta atividade chaperona, tentou-se realizar diversos ensaios *in vitro* utilizando-se a citrato sintase como substrato não nativo (Buchner *et al.*, 1998). Na primeira abordagem, optou-se por um experimento clássico no qual se observa a influência da suposta chaperona na agregação da citrato sintase durante sua desnaturação térmica (Buchner, *et al.*, 1998). Espera-se que, em presença de uma chaperona, as formas intermediárias de citrato sintase sejam protegidas da agregação, o que resulta na supressão do sinal de espalhamento a 500 nm. Entretanto, esse experimento apresenta uma limitação. A chaperona em questão precisa ser estável à 43°C, temperatura usada para a desnaturação da citrato sintase. Em ensaios preliminares, observou-se que LIC12922 precipita nessa temperatura, o que inviabiliza a demonstração de sua possível atividade chaperona através desse experimento.

Em uma segunda abordagem, testou-se se LIC12922 inibe a agregação da citrato sintase (CS) durante o processo de renaturação. Após transferência da CS do tampão desnaturante (contendo 6 M de guanidina HCl) ao tampão renaturante, formas intermediárias são formadas, das quais a maioria precipita nos primeiros 6 min após o início do processo de renaturação (Buchner, *et al.*, 1998). Sabe-se que chaperonas que interagem com formas intermediárias iniciais de CS são capazes de suprimir completamente sua agregação durante a reação de renovelamento. Nesse caso, observa-se a supressão do sinal de espalhamento de luz monitorado durante a reação (Buchner, *et al.*, 1998).

Testes foram realizados utilizando-se como tampão de desnaturação o próprio tampão de LIC12922, uma vez que essa proteína precipita no tampão 50 mM Tris HCl pH 8, 100 mM NaCl, normalmente empregado nesse tipo de ensaio. Entretanto, observou-se a

agregação de LIC12922 durante o ensaio, provavelmente devido a presença de guanidina HCl nos tampões usados para a coleta de dados, o que inviabilizou o experimento.

Em uma última tentativa, testou-se se LIC12922 é capaz de proteger a citrato sintase da inativação térmica irreversível. Como se sabia previamente que LIC12922 não é estável a 43°C, esse ensaio foi realizado em temperaturas menores (42, 40, 38 °C). Em um dos testes realizados, conseguiu-se observar um efeito de LIC12922 na supressão da inativação térmica de CS a 38 °C. Porém, o mesmo efeito foi observado em presença de BSA, usada como controle negativo.

Dessa forma, as abordagens *in vitro* utilizadas para se confirmar a provável atividade chaperona de LIC12922 mostraram-se inadequadas nesse caso. Idealmente, essa hipótese deveria ser testada em experimentos *in vivo*, utilizando-se mutantes de *L. interrogans* em que o gene LIC12922 esteja inativado. A detecção de eventuais alterações na quantidade de proteínas de membrana externa (OMPs) bem como o aumento da atividade do regulon sigma E nos mutantes seriam indicativos de que LIC12922 atua na via de biogênese das OMPs (Vertommen *et al.*, 2009; Missiakas *et al.*, 1996). Além desses ensaios, experimentos *in vitro* tais como *phage display* seriam interessantes para se tentar identificar sua especificidade (Bitto & McKay, 2003).

4.3. PRODUÇÃO E CARACTERIZAÇÃO DE LIC10494

A provável lipoproteína LIC10494 de *L. interrogans* sorovar Copenhageni foi produzida solúvel em *E. coli*, purificada e cristalizada. Estratégias de refinamento das condições de cristalização envolvendo variação de pH, de concentração de agente precipitante, uso de aditivos e aplicação da técnica de *microseeding* não resultaram em melhorias significativas no tamanho e na qualidade dos cristais. Os experimentos de difração de raios X não apresentaram resultados satisfatórios. Para nenhum dos cristais foi observada difração mesmo à baixa resolução. Possíveis causas da dificuldade em se melhorar os cristais de LIC10494 incluem uma zona de nucleação provavelmente muito estreita em seu diagrama de fases e a presença de regiões altamente flexíveis em sua estrutura.

Com base em análises de sequência e nos experimentos de CD, DLS, gel filtração analítica e SDS-PAGE, concluiu-se que LIC10494 apresenta uma região intrinsecamente desestruturada. Interessantemente, sua porção central desordenada, que corresponde a 30 % do tamanho de sua sequência, apresenta uma grande quantidade de resíduos de treonina, o que pode estar relacionado ao seu papel funcional.

LIC10494 apresenta uma forte tendência a formar agregados não covalentes. Em experimentos de gel filtração, observou-se a presença predominante de agregados e a presença minoritária de prováveis monômeros em amostras de LIC10494 em tampão Tris pH 8. Na busca por aditivos que pudessem diminuir essa agregação, observou-se o aminoácido arginina como um excelente supressor. Além de deslocar o equilíbrio agregado \leftrightarrow monômero para o estado monomérico, a adição de arginina diminuiu a polidispersividade de amostras submetidas à cristalização. Porém, o uso de amostras mais homogêneas não foi suficiente para se melhorar a qualidade dos cristais de LIC10494.

Frente às dificuldades de se obter cristais de LIC10494 adequados ao experimento de difração, acredita-se que a próxima estratégia a ser seguida seria a construção de variantes truncadas contendo apenas os segmentos enovelados adjacentes à porção central desestruturada de LIC10494.

Proteínas intrinsecamente desordenadas apresentam uma mobilidade mais rápida em cromatografia de exclusão molecular, visto que conformações estendidas resultam em dimensões hidrodinâmicas maiores (Receveur-Bréchet *et al.*, 2006). Assim, os *R_h* estimados para LIC10494 sugerem que sua fração não agregada corresponde a monômeros cujas formas desviam da globular. Porém para se validar essa hipótese, ensaios complementares precisam ser realizados tais como espectrometria de massas (Cohen, 1996), para se confirmar a massa teórica de LIC10494, e análises de ultracentrifugação analítica (Lebowitz *et al.*, 2002) para se confirmar seu estado oligomérico.

5. CONCLUSÕES

Os estudos estruturais apresentados nesse trabalho geraram hipóteses importantes acerca da função de três proteínas recentemente identificadas no projeto genoma de *L. interrogans* sorovar Copenhageni. Essas hipóteses formam a base para a elaboração de ensaios *in vitro* e *in vivo* que visem à caracterização funcional de Lp49, LIC12922 e LIC10494.

O antígeno Lp49 é composto por dois domínios cujos respectivos enovelamentos, *Ig-like* e *7-bladed β -propeller*, são recorrentemente observados como plataformas de interação a proteínas em domínios que desempenham as mais diversas funções tanto em procariotos como em eucariotos. Lp49 apresenta três prováveis sítios de interação proteína-proteína. Dois deles compreendem o domínio Imunoglobulina e o mais provável localiza-se na face superior do *β -propeller*, região que geralmente abriga o sítio de ligação a substrato em estruturas com esse tipo de enovelamento. Logo, propõe-se que a função de Lp49 envolva a interação com outras proteínas. Dada sua provável localização na membrana externa, sua antigenicidade e a ausência de homólogo na espécie saprófita *L. biflexa*, acredita-se que Lp49 possa ter um papel na patogenicidade de leptospirosas, participando da interação entre patógeno e hospedeiro.

A estrutura cristalográfica de LIC12922 revelou a presença de dois domínios: o domínio NC e o domínio parvulina. O domínio NC é estruturalmente relacionado aos domínios chaperona presentes na chaperona periplasmática SurA e na chaperona citoplasmática *trigger factor*, ambas de *E. coli*. A flexibilidade estrutural evidenciada para esse domínio, que envolve principalmente os módulos denominados “braços”, pode ter implicações em um mecanismo de interação com substratos. Demonstrou-se ainda neste trabalho que o domínio parvulina de LIC12922 não possui atividade de PPIase. Além disso, foram identificados potenciais sítios de interação proteína-proteína que podem ter um papel de reconhecimento ao substrato, como proposto para o domínio P1 de SurA, também desprovido dessa função catalítica (Xu *et al.*, 2007).

LIC12922 apresenta uma arquitetura de domínios semelhante à de três chaperonas extracitoplasmáticas: PrsA, que participa do enovelamento de proteínas secretadas em Firmicutes (Vitikainen *et al.*, 2004), PpiD e SurA, chaperonas periplasmáticas que atuam na via de biogênese de proteínas de membrana externa (OMPs) em Proteobactérias

(Dartigalongue & Raina, 1998; Lazar & Kolter, 1996). Comparações estruturais aliadas a estudos filogenéticos realizados no presente trabalho sugerem que LIC12922, PrsA, PpiD e SurA tenham evoluído de um ancestral comum, o que é apoiado pelas funções similares desempenhadas por PrsA, PpiD e SurA. Dada a maior similaridade entre LIC12922 e SurA e a possível localização periplasmática de LIC12922, postula-se que LIC12922 também seja uma chaperona periplasmática envolvida na biogênese de proteínas de membrana externa em leptospiros.

Homólogos de LIC12922 foram identificados em todas as bactérias espiroquetas cujos genomas já foram seqüenciados. A conservação desse gene sugere que ele desempenhe uma função essencial nessas bactérias. Confirmando-se que LIC12922 apresente uma função similar à SurA, abre-se um amplo caminho para o estudo da biogênese de proteínas de membrana externa em todo o filo Espiroqueta.

A provável lipoproteína LIC10494, aparentemente exclusiva de *L. interrogans*, apresenta uma extensa região central intrinsecamente desestruturada rica em resíduos de treonina. Essa observação traz uma pista importante para a elucidação de sua função. Isso porque regiões intrinsecamente desestruturadas geralmente têm o papel de se ligar a outras macromoléculas para formar complexos (Dyson & Wright, 2005). Diversos exemplos de proteínas que apresentam essa característica são encontrados em inúmeros processos biológicos tais como transcrição, tradução e transdução de sinal (Dyson & Wright, 2005). A ligação a diversos alvos com alta especificidade e baixa afinidade e a habilidade de superar restrições estéricas permitindo extensas superfícies de interação em proteínas relativamente pequenas, são algumas das vantagens propostas para regiões que não apresentam uma estrutura enovelada (Gunasekaran *et al.*, 2003, Wright and Dyson, 1999).

Coincidentemente, as funções das três proteínas alvo desse trabalho provavelmente envolvem a interação com outras proteínas. Portanto, uma etapa fundamental na determinação de suas funções será a identificação de seus supostos alvos, substratos ou parceiros. Em Lp49 e em LIC12922, prováveis sítios de ligação à proteína foram identificados, o que pode ser de grande valia para a elaboração de ensaios de mutagênese sítio dirigida que visem, posteriormente, a identificar resíduos chave para a interação entre essas proteínas e seus alvos. No caso do antígeno Lp49, sua estrutura tridimensional pode ainda auxiliar no desenho de ensaios que visem à identificação e caracterização de epítomos,

o que, conseqüentemente, pode ajudar na descoberta e no desenvolvimento de novas terapias, vacinas e diagnósticos (Westwood & Hay, 2001; Gershoni *et. al.*, 2007).

Em relação a LIC12922, a descoberta de sua homologia com a chaperona periplasmática SurA confere a ela o atributo de possível alvo para o desenvolvimento de drogas. Isso porque, recentemente, foi demonstrado que SurA, além de ter um papel importante na homeostase do envelope celular, também afeta a patogenicidade de diversas bactérias Gram negativas, sendo necessária para a completa virulência de bactérias uropatogênicas, incluindo *E. coli* (Behrens-Kneip, 2010).

Como pudemos observar nesse trabalho, a análise da estrutura tridimensional de proteínas de função desconhecida pode não fornecer, por si só, os indícios necessários para a completa elucidação da função dessas macromoléculas. Em um primeiro momento, essa constatação pode parecer contra-intuitiva, dado que a estrutura de uma proteína determina sua função. Entretanto, para podermos interpretar a estrutura de uma proteína, descobrir o que ela faz e como ela funciona, necessitamos de conhecimentos obtidos por meio de outras abordagens experimentais. Isso é evidente no caso de domínios cujos enovelamentos podem desempenhar as mais diferentes funções biológicas, como, por exemplo, os domínios *Immunoglobulin-like* e *7-bladed β -propeller*.

Por outro lado, a abordagem estrutural é de grande valia para se gerar hipóteses que direcionem os próximos passos a serem dados na busca pela caracterização funcional de proteínas. Além disso, uma vez desvendada a função das mesmas, suas estruturas tridimensionais se tornam fundamentais para a elucidação de seus mecanismos de ação. Dessa forma, espera-se que os resultados desse trabalho sejam um passo inicial na busca pelo conhecimento detalhado acerca da função biológica de Lp49, LIC12922 e LIC10494 e que, no futuro, as estruturas tridimensionais de Lp49 e de LIC12922 sejam úteis para a interpretação de resultados obtidos por meio de outras abordagens experimentais.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ADLER, B. and DE LA PEÑA MOCTEZUMA, A. *Leptospira* and leptospirosis. *Veterinary Microbiology*, v. 140, p. 287-296, 2010.

ALTSCHUL, S. F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, v. 25, p. 3389-3402, 1997.

ARAKAWA T., et al. Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. *Biophysical Chemistry*. v. 127, p. 1-8, 2007.

ATZINGEN, M. V., et al. Lsa21, a novel leptospiral protein binding adhesive matrix molecules and present during human infection. *BMC Microbiology*, v. 8, p. 70, 2008.

BARBOSA, A. S., et al. A newly identified leptospiral adhesin mediates attachment to laminin. *Infection and immunity*, v. 74, p. 6356-6364, 2006.

BARCLAY, A. N. Membrane proteins with immunoglobulin-like domains--a master superfamily of interaction molecules. *Seminars in immunology*, v. 15, p. 215-223, 2003.

BAROCCHI, M. A., et al. Rapid translocation of polarized MDCK cell monolayers by *Leptospira interrogans*, an invasive but nonintracellular pathogen. *Infection and immunity*, v. 70, p. 6926-6932, 2002.

BAYNES, B. M., WANG, D. I. and TROUT, B. L. Role of arginine in the stabilization of proteins against aggregation. *Biochemistry*, v. 44, p. 4919-4925, 2005.

BEHRENS-KNEIP, S. The role of SurA factor in outer membrane protein transport and virulence. *International Journal of Medical Microbiology*, v. p. 2010.

BEHRSIN, C. D., et al. Functionally important residues in the peptidyl-prolyl isomerase Pin1 revealed by unigenic evolution. *Journal of molecular biology*, v. 365, p. 1143-1162, 2007.

BERGFORS, T. Seeds to crystals. *Journal of structural biology*, v. 142, p. 66-76, 2003.

BHARTI, A. R., et al. Leptospirosis: a zoonotic disease of global importance. *The Lancet infectious diseases*, v. 3, p. 757-771, 2003.

BIERTUMPFEL, C., et al. Crystallization of biological macromolecules using agarose gel. *Acta crystallographica D*, v. 58, p. 1657-1659, 2002.

BITTO, E. and MCKAY, D. B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure*, v. 10, p. 1489-1498, 2002.

BITTO, E. and MCKAY, D. B. The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *The Journal of biological chemistry*, v. 278, p. 49316-49322, 2003.

BOGGON, T. J., et al. Implication of tubby proteins as transcription factors by structure-based functional analysis. *Science*, v. 286, p. 2119-2125, 1999.

BOHM, G., MUHR, R. and JAENICKE, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein engineering*, v. 5, p. 191-195, 1992.

BONDOS, S. E. and BICKNELL, A. Detection and prevention of protein aggregation before, during, and after purification. *Analytical biochemistry*, v. 316, p. 223-231, 2003.

BOURHY, P., et al. Random insertional mutagenesis of *Leptospira interrogans*, the agent of leptospirosis, using a mariner transposon. *Journal of bacteriology*, v. 187, p. 3255-3258, 2005.

BRADFORD, J. R. and WESTHEAD, D. R. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, v. 21, p. 1487-1494, 2005.

BRANGER, C., et al. Identification of the hemolysis-associated protein 1 as a cross-protective immunogen of *Leptospira interrogans* by adenovirus-mediated vaccination. *Infection and immunity*, v. 69, p. 6831-6838, 2001.

BREINER, D. D., et al. *Leptospira interrogans* binds to human cell surface receptors including proteoglycans. *Infection and immunity*, v. 77, p. 5528-5536, 2009.

BRENNER, D. J., et al. Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and four new *Leptospira* genomospecies. *International journal of systematic bacteriology*, v. 49, p. 839-858, 1999.

BRYSON K., et al. Protein structure prediction servers at University College London. *Nucleic Acids Research*, v. 33(Web Server issue), p. W36-38, 2005.

BUCHNER, J., GRALLERT, H. and JAKOB, U. Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods in enzymology*, v. 290, p. 323-338, 1998.

BULACH, D. M., et al. Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proceedings of the National Academy of Sciences of the United States of America*, v. 103, p. 14560-14565, 2006.

CHAYEN, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *Journal of Applied Crystallography*, v. 30, p. 198-202, 1997.

COHEN, S. L. Domain elucidation by mass spectrometry. *Structure*, v. 4, p. 1013-1016, 1996.

CSERZO, M., et al. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein engineering*, v. 10, p. 673-676, 1997.

DARTIGALONGUE, C. and RAINA, S. A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *The EMBO journal*, v. 17, p. 3968-3980, 1998.

DAS, U., et al. Inhibition of protein aggregation: supramolecular assemblies of arginine hold the key. *PloS one*, v. 2, p. e1176, 2007.

DYSON, H. J. and WRIGHT, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews*, v. 6, p. 197-208, 2005.

EDELHOCH, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, v. 6, p. 1948-1954, 1967.

EDWARDS, T. A., et al. Model of the brain tumor-Pumilio translation repressor complex. *Genes & development*, v. 17, p. 2508-2513, 2003.

ELLIS, W. A. Leptospirosis as a cause of reproductive failure. *The Veterinary clinics of North America*, v. 10, p. 463-478, 1994.

EVERARD, C. O., et al. A twelve-year study of leptospirosis on Barbados. *European journal of epidemiology*, v. 11, p. 311-320, 1995.

FERBITZ, L., et al. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, v. 431, p. 590-596, 2004.

FRASER, C. M., et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, v. 390, p. 580-586, 1997.

FRASER, C. M., et al. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science*, v. 281, p. 375-388, 1998.

FULOP, V. and JONES, D. T. Beta propellers: structural rigidity and functional diversity. *Current opinion in structural biology*, v. 9, p. 715-721, 1999.

GALAKTIONOV, S.G. and MARSHALL, G.R. Prediction of Protein Structure in Terms of Intraglobular Contacts: 1D to 2D to 3D, Technical Report, Center for Molecular Design, Washington University, St. Louis, 1996.

GAMBERINI, M., et al. Whole-genome analysis of *Leptospira interrogans* to identify potential vaccine candidates against leptospirosis. *FEMS microbiology letters*, v. 244, p. 305-313, 2005.

GARDY, J. L., et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, v. 21, p. 617-623, 2005.

GERSHONI, J. M., et al. Epitope mapping: the first step in developing epitope-based vaccines. *BioDrugs*, v. 21, p. 145-156, 2007.

GROOMS, D. L. Reproductive losses caused by bovine viral diarrhea virus and leptospirosis. *Theriogenology*, v. 66, p. 624-628, 2006.

GUNASEKARAN, K., et al. Extended disordered proteins: targeting function with less scaffold. *Trends in biochemical sciences*, v. 28, p. 81-85, 2003.

HAAKE, D. A. Spirochaetal lipoproteins and pathogenesis. *Microbiology*, v. 146 (Pt 7), p. 1491-1504, 2000.

HAAKE, D. A., et al. Leptospiral outer membrane proteins OmpL1 and LipL41 exhibit synergistic immunoprotection. *Infection and immunity*, v. 67, p. 6572-6582, 1999.

HAAKE, D. A., et al. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. *Journal of bacteriology*, v. 186, p. 2818-2828, 2004.

HAMBURGER, Z. A., et al. Crystal structure of invasin: a bacterial integrin-binding protein. *Science*, v. 286, p. 291-295, 1999.

HERAS, B. and MARTIN, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta crystallographica D*, v. 61, p. 1173-1180, 2005.

HOLM, L. and SANDER, C. Dali: a network tool for protein structure comparison. *Trends in biochemical sciences*, v. 20, p. 478-480, 1995.

HUMPHRIES, M. J. and NEWHAM, P. The structure of cell-adhesion molecules. *Trends in cell biology*, v. 8, p. 78-83, 1998.

HUTCHINSON, E. G. and THORNTON, J. M. A revised set of potentials for beta-turn formation in proteins. *Protein Science*, v. 3, p. 2207-2216, 1994.

HWANG, K. Y., et al. Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nature structural biology*, v. 6, p. 691-696, 1999.

IBRAGHIMOV-BESKROVNAYA, O., et al. Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. *Human molecular genetics*, v. 9, p. 1641-1649, 2000.

INADA, R., et al. The Etiology, Mode of Infection, and Specific Therapy of Weil's Disease (Spirochaetosis Icterohaemorrhagica). *The Journal of experimental medicine*, v. 23, p. 377-402, 1916.

JIN, D., et al. *Leptospira interrogans* induces apoptosis in macrophages via caspase-8- and caspase-3-dependent pathways. *Infection and immunity*, v. 77, p. 799-809, 2009.

JING, H., et al. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure*, v. 10, p. 1453-1464, 2002.

JONES, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, v. 292, p. 195-202, 1999.

KAJANDER, T., et al. The structure of *Neurospora crassa* 3-carboxy-cis,cis-muconate lactonizing enzyme, a beta propeller cycloisomerase. *Structure*, v. 10, p. 483-492, 2002.

KASAROV, L. B. Degradation of the erythrocyte phospholipids and haemolysis of the erythrocytes of different animal species by leptospirae. *Journal of Medical Microbiology*, v. 3, p. 29-37, 1970.

KO, A. I., et al. Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. *Lancet*, v. 354, p. 820-825, 1999.

KO, A. I., GOARANT, C. and PICARDEAU, M. *Leptospira*: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nature reviews*, v. 7, p. 736-747, 2009.

KOFRON, J. L., et al. Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry*, v. 30, p. 6127-6134, 1991.

KRISSINEL, E. and HENRICK, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica D*, v. 60, p. 2256-2268, 2004.

KRISSINEL, E. and HENRICK, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, v. 372, p. 774-797, 2007.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, v. 227, p. 680-685, 1970.

LAMBRIGHT, D. G., et al. The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature*, v. 379, p. 311-319, 1996.

LASKOWSKI, R. A., WATSON, J. D. and THORNTON, J. M. ProFunc: a server for predicting protein function from 3D structure. *Nucleic acids research*, v. 33, p. W89-93, 2005.

LASKOWSKI, R. A., WATSON, J. D. and THORNTON, J. M. Protein function prediction using local 3D templates. *Journal of molecular biology*, v. 351, p. 614-626, 2005.

LAZAR, S. W. and KOLTER, R. SurA assists the folding of *Escherichia coli* outer membrane proteins. *Journal of bacteriology*, v. 178, p. 1770-1773, 1996.

LEBOWITZ, J., LEWIS, M. S. and SCHUCK, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Science*, v. 11, p. 2067-2079, 2002.

LECOUR, H., et al. Human leptospirosis -a review of 50 cases. *Infection*, v. 17, p. 8-12, 1989.

LEE, G. J. Assaying proteins for molecular chaperone activity. *Methods in cell biology*, v. 50, p. 325-334, 1995.

LEE, S. H., et al. Cytotoxic activities of *Leptospira interrogans* hemolysin SphH as a pore-forming protein on mammalian cells. *Infection and immunity*, v. 70, p. 315-322, 2002.

LEVETT, P. N. Leptospirosis. *Clinical microbiology reviews*, v. 14, p. 296-326, 2001.

LI, X., et al. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics*, v. 10, p. 30-40, 1999.

LILENBAUM, W., et al. A serological study on *Brucella abortus*, caprine arthritis-encephalitis virus and *Leptospira* in dairy goats in Rio de Janeiro, Brazil. *Veterinary Journal*, v. 173, p. 408-412, 2007.

LO CONTE, L., et al. SCOP: a structural classification of proteins database. *Nucleic acids research*, v. 28, p. 257-259, 2000.

LONGHI, M. T., et al. A newly identified protein of *Leptospira interrogans* mediates binding to laminin. *Journal of Medical Microbiology*, v. 58, p. 1275-1282, 2009.

LUO, Y., et al. Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature*, v. 405, p. 1073-1077, 2000.

MACARIO, A. J. and CONWAY DE MACARIO, E. The molecular chaperone system and other anti-stress mechanisms in archaea. *Frontiers in Bioscience*, v. 6, p. D262-283, 2001.

MARTINEZ, R., et al. [Efficacy and safety of a vaccine against human leptospirosis in Cuba]. *Revista panamericana de salud publica = Pan American journal of public health*, v. 15, p. 249-255, 2004.

MCBRIDE, A. J., et al. Leptospirosis. *Current opinion in infectious diseases*, v. 18, p. 376-386, 2005.

MERIEN, F., BARANTON, G. and PEROLAT, P. Invasion of Vero cells and induction of apoptosis in macrophages by pathogenic *Leptospira interrogans* are correlated with virulence. *Infection and immunity*, v. 65, p. 729-738, 1997.

MERZ, F., et al. Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *The EMBO journal*, v. 27, p. 1622-1632, 2008.

MISSIAKAS, D., BETTON, J. M. and RAINA, S. New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Molecular microbiology*, v. 21, p. 871-884, 1996.

NAKAI, K. and HORTON, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in biochemical sciences*, v. 24, p. 34-36, 1999.

NASCIMENTO, A. L., et al. Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *Journal of bacteriology*, v. 186, p. 2164-2172, 2004a.

NASCIMENTO, A. L., et al. Genome features of *Leptospira interrogans* serovar Copenhageni. *Brazilian journal of medical and biological research*, v. 37, p. 459-477, 2004b.

NEVES, F. O., et al. Identification of a novel potential antigen for early-phase serodiagnosis of leptospirosis. *Archives of microbiology*, v. 188, p. 523-532, 2007.

NIELSEN, H., et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein engineering*, v. 10, p. 1-6, 1997.

NIELSEN, H. and KROGH, A. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology ; ISMB*, v. 6, p. 122-130, 1998.

PICARDEAU, M., et al. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PloS one*, v. 3, p. e1607, 2008.

PICKLES, L. M., et al. Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. *Structure*, v. 10, p. 751-761, 2002.

PIZZA, M., et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*, v. 287, p. 1816-1820, 2000.

PLANK, R. and DEAN, D. Overview of the epidemiology, microbiology, and pathogenesis of *Leptospira* spp. in humans. *Microbes and infection*, v. 2, p. 1265-1276, 2000.

PRILUSKY, J., et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, v. 21, p. 3435-3438, 2005.

PUGSLEY, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiological reviews*, v. 57, p. 50-108, 1993.

RAMOS C.R., et al. A high-copy T7 *Escherichia coli* expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Brazilian Journal of Medical and Biological Research*, v. 37, p. 1103-1109, 2004.

RANGANATHAN, R., et al. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell*, v. 89, p. 875-886, 1997.

RECEVEUR-BRECHOT, V., et al. Assessing protein disorder and induced folding. *Proteins*, v. 62, p. 24-45, 2006.

REIS, R. B., et al. Impact of environment and social gradient on *Leptospira* infection in urban slums. *PLoS neglected tropical diseases*, v. 2, p. e228, 2008.

REN, S. X., et al. Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature*, v. 422, p. 888-893, 2003.

RILEY, L. W., et al. Slum health: diseases of neglected populations. *BMC international health and human rights*, v. 7, p. 2, 2007.

ROBERTS, A., et al. YbiV from *Escherichia coli* K12 is a HAD phosphatase. *Proteins*, v. 58, p. 790-801, 2005.

SAMBSIAVA, R. R., et al. Leptospirosis in India and the rest of the world. *Brazilian Journal of Infectious Disease*, v. 7, p. 178-193, 2003.

SHEINERMAN, F. B., NOREL, R. and HONIG, B. Electrostatic aspects of protein-protein interactions. *Current opinion in structural biology*, v. 10, p. 153-159, 2000.

SIEMINSKA, E. A., et al. The X-ray crystal structure of PA1607 from *Pseudomonas aureginosa* at 1.9 Å resolution--a putative transcription factor. *Protein Science*, v. 16, p. 543-549, 2007.

SPRAGUE, E. R., et al. Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. *The EMBO journal*, v. 19, p. 3016-3027, 2000.

STEVENSON, B., et al. *Leptospira interrogans* endostatin-like outer membrane proteins bind host fibronectin, laminin and regulators of complement. *PloS one*, v. 2, p. e1188, 2007.

SUPUTTAMONGKOL, Y., et al. Strategies for diagnosis and treatment of suspected leptospirosis: a cost-benefit analysis. *PLoS neglected tropical diseases*, v. 4, p. e610, 2010.

TAKAGI, J., et al. Complex between nidogen and laminin fragments reveals a paradigmatic beta-propeller interface. *Nature*, v. 424, p. 969-974, 2003.

TARRICONE, C., et al. Coupling PAF signaling to dynein regulation: structure of LIS1 in complex with PAF-acetylhydrolase. *Neuron*, v. 44, p. 809-821, 2004.

TEICHMANN, S. A., MURZIN, A. G. and CHOTHIA, C. Determination of protein function, evolution and interactions by structural genomics. *Current opinion in structural biology*, v. 11, p. 354-363, 2001.

THOMPSON, J. C. and MANKTELOW, B. W. Pathogenesis and red blood cell destruction in haemoglobinaemic leptospirosis. *Journal of Comparative Pathol*, v. 96, p. 529-540, 1986.

TUNG, C. H. and YANG, J. M. fastSCOP: a fast web server for recognizing protein structural domains and SCOP superfamilies. *Nucleic acids research*, v. 35, p. W438-443, 2007.

VERMA, A., et al. Leptospiral endostatin-like protein A is a bacterial cell surface receptor for human plasminogen. *Infection and immunity*, v. 78, p. 2053-2059, 2010.

VERMA, A., et al. LfhA, a novel factor H-binding protein of *Leptospira interrogans*. *Infection and immunity*, v. 74, p. 2659-2666, 2006.

VERTOMMEN, D., et al. Characterization of the role of the *Escherichia coli* periplasmic chaperone SurA using differential proteomics. *Proteomics*, v. 9, p. 2432-2443, 2009.

VIEIRA, M. L., et al. Lsa63, a newly identified surface protein of *Leptospira interrogans* binds laminin and collagen IV. *Journal of Infection*, v. 60, p. 52-64, 2010.

VITIKAINEN, M., et al. Structure-function analysis of PrsA reveals roles for the parvulin-like and flanking N- and C-terminal domains in protein folding and secretion in *Bacillus subtilis*. *The Journal of biological chemistry*, v. 279, p. 19302-19314, 2004.

WATSON, J. D., LASKOWSKI, R. A. and THORNTON, J. M. Predicting protein function from sequence and structural data. *Current opinion in structural biology*, v. 15, p. 275-284, 2005.

WEDEMEYER, W. J., WELKER, E. and SCHERAGA, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry*, v. 41, p. 14637-14644, 2002.

WEININGER, U., et al. The prolyl isomerase domain of PpiD from *Escherichia coli* shows a parvulin fold but is devoid of catalytic activity. *Protein Science*, v. 19, p. 6-18, 2010.

WHO. Leptospirosis worldwide, 1999. *Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*, v. 74, p. 237-242, 1999.

WRIGHT, P. E. and DYSON, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, v. 293, p. 321-331, 1999.

XU, X., et al. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *Journal of molecular biology*, v. 373, p. 367-381, 2007.

XUE, F., YAN, J. and PICARDEAU, M. Evolution and pathogenesis of *Leptospira* spp.: lessons learned from the genomes. *Microbes and infection*, v. 11, p. 328-333, 2009.

YAKUNIN, A. F., et al. Structural proteomics: a tool for genome annotation. *Current opinion in chemical biology*, v. 8, p. 42-48, 2004.

YAN, Y., et al. An evaluation of the serological and epidemiological effects of the outer envelope vaccine to leptospira. *Journal of the Chinese Medical Association*, v. 66, p. 224-230, 2003.

ZAREMBINSKI, T. I., et al. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proceedings of the National Academy of Sciences of the United States of America*, v. 95, p. 15189-15193, 1998.

ZHANG, C. and KIM, S. H. A comprehensive analysis of the Greek key motifs in protein beta-barrels and beta-sandwiches. *Proteins*, v. 40, p. 409-419, 2000.

ZHANG, C. and KIM, S. H. Overview of structural genomics: from structure to function. *Current opinion in chemical biology*, v. 7, p. 28-32, 2003.

ZHANG, W., et al. Crystal structure of the major histocompatibility complex class I H-2Kb molecule containing a single viral peptide: implications for peptide binding and T-cell receptor recognition. *Proceedings of the National Academy of Sciences of the United States of America*, v. 89, p. 8403-8407, 1992.

ZHANG, Y. X., et al. Cytotoxic activity and probable apoptotic effect of Sph2, a sphigomyelinase hemolysin from *Leptospira interrogans* strain Lai. *BMB Reports*, v. 41, p. 119-125, 2008.

ANEXO

Estudos de biologia molecular e genética em *Leptospira* spp.

Projeto de pesquisa desenvolvido na Unidade de Biologia de Espiroquetas do Instituto Pasteur (Paris, FR) durante estágio realizado de setembro de 2009 a fevereiro de 2010.

orientador: Dr. Mathieu Picardeau

co-orientadora: Dra. Leyla Slamti

Estágio financiado pelo Programa de Doutorado no País com Estágio no Exterior (PDEE) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Processo n°. BEX 0689/09-2.

A. ESTUDOS DE BIOLOGIA MOLECULAR E GENÉTICA EM *LEPTOSPIRA* SPP.

A.1. Introdução

A Unidade de Biologia de Espiroquetas do Instituto Pasteur (Paris, França) é reconhecida pela sua liderança na área de genética de Espiroquetas e é pioneira no desenvolvimento de técnicas de manipulação genética de *Leptospira* spp. Nos últimos anos, o grupo desenvolveu uma série de ferramentas genéticas para manipulação tanto da espécie saprófita *L. biflexa* quanto da espécie patogênica *L. interrogans*. Dentre as ferramentas genéticas desenvolvidas para manipulação da espécie saprófita *L. biflexa* destacam-se a mutagênese randômica (Louvel *et al.*, 2005) e sítio dirigida (Louvel *et al.*, 2008). Métodos similares têm facilitado a manipulação de genes em linhagens patogênicas (Bourhy *et al.*, 2005, Picardeau, 2008). Esses progressos fornecem a oportunidade de se aplicar os mais recentes e avançados métodos para a identificação de funções de genes em *Leptospira* spp.

Uma das principais questões ainda não elucidadas em relação às leptospiros é o que determina sua forma espiralada, ou seja, quais são os genes responsáveis pela sua peculiar morfologia. Sabe-se que mutantes que formam flagelos periplasmáticos estendidos ou perdem os flagelos perdem as extremidades aduncas, características de leptospiros, mas ainda apresentam uma forma espiralada (Bromley, *et al.*, 1979, Picardeau *et al.*, 2001). A remoção da membrana externa de leptospiros também não resulta em perda de seu formato celular (Auran *et al.*, 1972).

Assim, com o objetivo de se determinar os genes envolvidos na morfologia espiralada de leptospiros decidiu-se criar uma biblioteca de mutantes de *L. biflexa* por mutagênese randômica como descrito em Louvel & Picardeau (2007). As atividades desenvolvidas nesse trabalho incluíram a elaboração da biblioteca de mutantes, a seleção de fenótipos condizentes com uma possível alteração morfológica e a identificação do sítio de inserção do transposon nos mutantes selecionados.

Paralelamente, outro projeto foi realizado com o intuito de se desenvolver novos marcadores de seleção para estudos genéticos de *Leptospira* spp. Os resultados desse trabalho estão reunidos em um manuscrito recentemente publicado pela revista *Applied and Environmental Microbiology* (AEM).

A.2. Geração da biblioteca de mutantes de *L. biflexa*

A biblioteca de mutantes de *L. biflexa* foi gerada por um sistema de mutagênese randômica baseado no transposon de origem eucariótica Himar1 (Chiang & Rubin, 2002). Plasmídeos derivados de pSC189 carregando a transposase hiperativa C9 e o transposon *Himar1* contendo um cassete de resistência a canamicina foram introduzidos em *L. biflexa* por conjugação, utilizando-se como linhagem doadora células *E. coli* β 2163, como descrito em Picardeau (2008).

As bactérias transformantes foram plaqueadas em meio EMJH sólido (Louvel & Picardeau, 2007) suplementado com 50 μ g/mL de canamicina. Após algumas semanas de incubação a 30 °C, as placas foram observadas e cerca de 10.000 colônias consideradas pequenas foram repicadas em meio EMJH sólido (1,2 % ágar noble) e semi-sólido (0,3 % ágar noble) para análise do fenótipo “tamanho de colônia” utilizando-se colônias selvagens como referência (FIGURA A1).

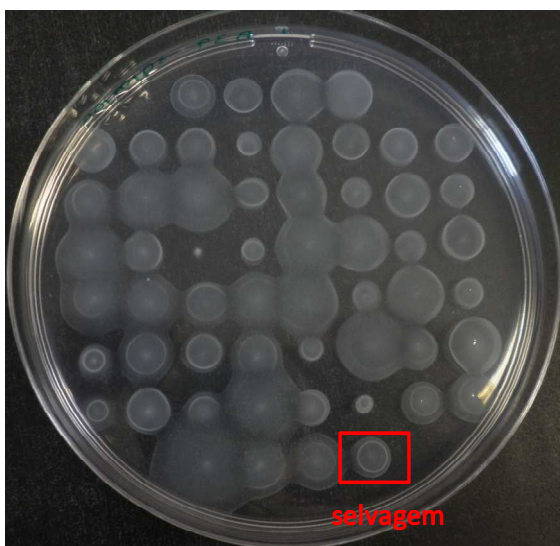


FIGURA A1. Exemplo de placa EMJH 0,3 % ágar noble contendo colônias mutantes de *L. biflexa*. Em todas as placas da biblioteca, repica-se uma cultura selvagem (quadrado vermelho) para servir como referência para a seleção dos mutantes.

O critério de seleção “colônias pequenas” parte do pressuposto de que bactérias que apresentam alterações morfológicas possuem menor mobilidade, formando colônias menores do que a selvagem principalmente em meio semi-sólido. Evidentemente, a formação de colônias menores do que a selvagem pode ter outras causas tais como crescimento lento, em função de mutações em vias metabólicas importantes, interrupção de

genes relacionados à quimiotaxia e mutações que afetem a formação ou o funcionamento dos flagelos. Mesmo assim, esse critério foi adotado pela sua simplicidade e rapidez.

O sítio de inserção do transposon foi identificado em aproximadamente 100 mutantes selecionados na triagem fenotípica descrita acima. O método empregado nessa identificação foi o PCR semi randômico (Beeman & Stauth, 1997). Esse método envolve duas rodadas consecutivas de PCR. Na primeira rodada, utilizam-se oligonucleotídeos degenerados contendo uma etiqueta 5' e oligonucleotídeos específicos para o gene de resistência à canamicina presente no transposon. Na segunda rodada, utilizam-se oligonucleotídeos desenhados para amplificação de um produto secundário dentro do produto originado na primeira rodada. Como molde para o PCR utilizou-se DNA genômico extraído dos mutantes utilizando-se o sistema Maxwell 16 (Promega). Os produtos da segunda rodada do PCR semi randômico foram seqüenciados pela empresa Cogenics Genome Express (Meylan, França). Grande parte dos sítios de inserção localizou-se em genes que codificam proteínas hipotéticas.

Em análises dos mutantes selecionados por microscopia de campo escuro, não se observaram alterações de mobilidade e de morfologia para a maioria deles. Porém, um dos mutantes analisados, denominado M41, apresentou-se menos móvel e aparentava ter extremidades retas quando comparado a bactérias selvagens. Análises de microscopia eletrônica mostraram que, de fato, esse mutante não apresentava extremidades aduncas e revelou também que ele mantinha a forma espiralada (FIGURA A2).

No mutante M41, o transposon está inserido no gene LEPBI_I1597 (NC_010602.1), que codifica uma proteína hipotética conservada portadora de um provável peptídeo sinal. O gene LEPBI_I1597 localiza-se próximo ao gene *fliN* (LEPBI_I1599) que codifica uma das proteínas do motor flagelar. Com já mencionado, mutantes que formam flagelos defeituosos perdem as extremidades encurvadas, características de leptospiros, mas ainda apresentam uma forma espiralada (Bromley, *et al.*, 1979, Picardeau *et al.*, 2001), semelhante ao mutante M41. Assim, sugere-se que o gene LEPBI_I1597 pode estar envolvido com a formação ou com o funcionamento dos flagelos.

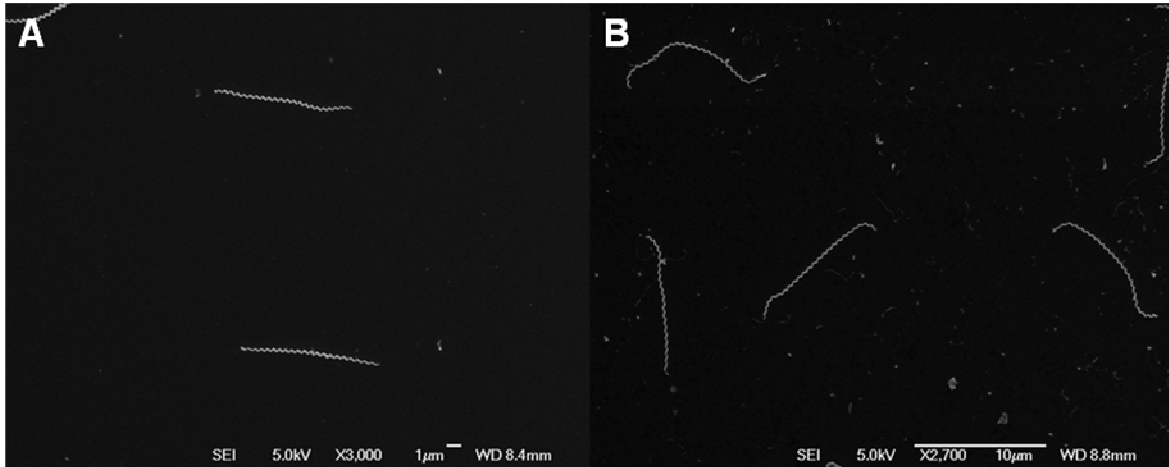


FIGURA A2. Fotos de microscopia eletrônica do mutante M41 (painel A) e de *L. biflexa* selvagem (painel B).

Para se testar se o fenótipo “colônia pequena” de M41 não foi resultante de uma taxa de crescimento mais lenta, mediu-se a curva de crescimento de M41 e de *L. biflexa* selvagem. A curva de crescimento do mutante M41 apresentou uma fase *lag* mais curta e uma fase *log* semelhante à da curva de *L. biflexa* selvagem (FIGURA A3).

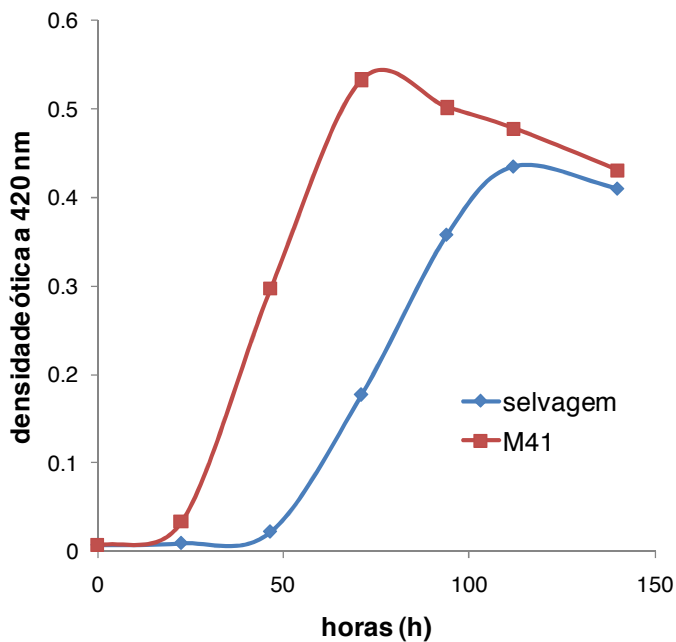


FIGURA A3. Curvas de crescimento do mutante M41 e de *L. biflexa* selvagem.

O prosseguimento deste trabalho incluirá estudos sobre a função do gene LEPBI_I1597 bem como a validação e a caracterização de outros mutantes gerados nessa biblioteca e será conduzido pela pós-doutoranda Leyla Slamti.

A.3. Marcadores de resistência a antibiótico para manipulação de *Leptospira* spp.

Os resultados desse trabalho estão reunidos em um manuscrito recentemente publicado pela revista *Applied and Environmental Microbiology* (AEM) e apresentado a seguir. As atividades desenvolvidas incluíram a inserção do cassete de resistência à gentamicina no vetor pMKL, a transformação de *L. interrogans* sorovar Lai com essa construção, a determinação de curvas de crescimento para linhagens selvagens e transformantes de *L. biflexa* (expressando o cassete de resistência a gentamicina) e a determinação da concentração inibitória mínima (CMI) dos antibióticos canamicina, gentamicina e espectinomicina para linhagens selvagens e mutantes (resistentes espontâneos ou transformados com o vetor pMKL) de *Leptospira* spp.

A.3.1. Artigo III

**Antibiotic-resistance markers for genetic manipulations of
Leptospira spp.**

Dante Poggi, Priscila O. Giuseppe and Mathieu Picardeau

Antibiotic Resistance Markers for Genetic Manipulations of *Leptospira* spp.[∇]

Dante Poggi, Priscila Oliveira De Giuseppe, and Mathieu Picardeau*

Unité de Biologie des Spirochètes, Institut Pasteur, Paris, France

Received 29 March 2010/Accepted 19 May 2010

We measured the frequency of appearance of spontaneous mutants resistant to gentamicin, kanamycin, streptomycin, and spectinomycin in saprophytic and pathogenic *Leptospira* strains. The mutations responsible for the spontaneous resistance to streptomycin and spectinomycin were identified in the *rpsL* and *rrs* genes, respectively. We also generated a gentamicin resistance cassette that allows the use of a third selectable marker in leptospires. These results may facilitate further advances in gene transfer systems in *Leptospira* spp.

AQ: A

AQ: B

Fn1/AQ:C Our understanding of leptospiral pathogenesis depends on reliable genetic tools for fully characterizing genes of interest. Significant advances in genetics of *Leptospira* spp. have been made over the last few years (8, 11). For generating antibiotic resistance genetic markers, our group focused on antibiotics other than those used therapeutically. We therefore excluded the use of β -lactams, as they are used to treat leptospirosis, which is an emerging disease with more 500,000 severe cases occurring annually (8). Plasmid DNA can be introduced into *Leptospira* by electroporation (2, 21) or conjugation (16). In 1990, Saint Girons et al. used the replication origin of the LE1 leptophage (22) to generate a plasmid that was able to replicate autonomously in both the saprophyte *Leptospira biflexa* and *Escherichia coli* (21). They used resistance to kanamycin (Kan), which was conferred by a gene from the Gram-positive bacterium *Enterococcus faecalis*, as a genetic marker to select for introduced DNA. Another marker, a spectinomycin (Spc) resistance cassette from *Staphylococcus aureus*, was also used as a selectable marker in *Leptospira* spp. (1). Further studies have used Spc and Kan markers to screen for transformants resulting from plasmid replication or chromosomal integration in leptospires (8, 11). As the proportion of allelic-exchange mutants is low and as chromosomal integration generally occurs through a single recombination event, a plasmid containing the *rpsL* wild-type gene as a counterselectable marker in a streptomycin (Str)-resistant strain of *L. biflexa* (due to a mutation in *rpsL*) was also used to eliminate clones harboring the plasmid and/or clones that have integrated the plasmid through a single-cross-over event (17).

AQ: D

T1

Resistance to aminoglycosides and spectinomycin in *Leptospira* spp. Susceptibility testing of both saprophytic (*L. biflexa* serovar Patoc strain Patoc 1) and pathogenic (*L. interrogans* serovar Copenhageni strain Fiocruz L1-130, *L. interrogans* serogroup Canicola strain Kito, and *L. interrogans* serovar Lai strain 56601) *Leptospira* strains to aminoglycosides and Spc was performed to determine the MICs (Table 1). We then examined the frequency of appearance of spontaneous mu-

tants resistant to gentamicin (Gen), Kan, Spc, and Str (Table 2). Spontaneous mutations resulting in Str and Spc resistance arose *in vitro* at frequencies of approximately 10^{-9} to 10^{-10} (Table 2). The development of Spc resistance in *Leptospira* occurs at the level of detection in transformation assays in which 10^9 transformants are plated. The two copies of *rrs* (18) were amplified and sequenced in wild-type and Str- and Spc-resistant mutant strains (Table 3). The *rrs* sequences of three independent *L. interrogans* Spc-resistant isolates indicated that all had a C-to-U substitution in the two *rrs* copies at position 1192 (using *E. coli* as the reference for numbering). In contrast, the alleles identified in four *L. biflexa* Spc-resistant variants contained a G-to-U substitution in the 16S rRNA gene at position 1064 (Fig. 1B). The expression of the resistant phenotype may therefore require a homogeneous population of cellular antibiotic-resistant ribosomes (19). MICs for these Spc-resistant mutants with transversion G1064C and transition C1192U were greater than 100 μ g/ml, whereas MICs between 2 and 4 μ g/ml were observed for wild-type parents. Mutations at these positions confer Spc resistance in *E. coli* (12), *Neisseria* spp. (7), and *Nicotiana tabacum* chloroplasts (5).

Frequencies of spontaneous mutation in *rpsL*, which encodes the ribosomal protein S12, were on the order of 10^{-9} to 10^{-10} (Table 2), similar to the mutation frequency of *rpsL* in *E. coli* (4×10^{-10}) (10) and that in *Borrelia burgdorferi* (10^{-9}) (3). The rate of spontaneous resistance to Str for strains Patoc and Kito was about 10 times higher than that for strains Fiocruz

TABLE 1. Antibiotic susceptibility of *Leptospira* strains^a

Antibiotic	MIC (mg/ml) for <i>Leptospira</i> strain			
	Patoc wt	Lai wt	Kito wt	Fiocruz wt
Streptomycin	0.5	1	2	1
Kanamycin	4	4	5	4
Gentamicin	4	4	8	4
Spectinomycin	4	4	2	2

^a MICs were defined as the lowest concentration of compound to inhibit visible growth after 1 week of incubation in EMJH liquid medium (11) at 30°C. MICs for wild-type *L. biflexa* serovar Patoc strain Patoc 1 (Patoc wt), *L. interrogans* serovar Lai strain 56601 (Lai wt), *L. interrogans* serogroup Canicola strain Kito (Kito wt), and *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (Fiocruz wt) were assessed by a microdilution technique as previously indicated (13).

* Corresponding author. Mailing address: Unité de Biologie des Spirochètes, Institut Pasteur, 28 rue du docteur Roux, 75724 Paris Cedex 15, France. Phone: 33 (1) 45 68 83 68. Fax: 33 (1) 40 61 30 01. E-mail: mathieu.picardeau@pasteur.fr.

[∇] Published ahead of print on ●●●●●●●●.

TABLE 2. Frequency of antibiotic-resistant mutants^a

Strain	Mutation frequency			
	Kan	Str	Gen	Spc
<i>L. biflexa</i> serovar Patoc strain Patoc 1	<10 ⁻¹⁰	1.5 × 10 ⁻⁹	<10 ⁻¹⁰	2.5 × 10 ⁻⁹
<i>L. interrogans</i> serogroup Canicola strain Kito	<10 ⁻¹⁰	1.2 × 10 ⁻⁹	<10 ⁻¹⁰	4.5 × 10 ⁻⁹
<i>L. interrogans</i> serovar Copenhageni strain Fiocruz L1-130	<10 ⁻¹⁰	0.8 × 10 ⁻¹⁰	<10 ⁻¹⁰	1.2 × 10 ⁻¹⁰
<i>L. interrogans</i> serovar Lai strain 56601	<10 ⁻¹⁰	<10 ⁻¹⁰	<10 ⁻¹⁰	<10 ⁻¹⁰

^a Values are means for three independent experiments. Bacteria (10⁹ and 10¹⁰) were spread onto EMJH plates (11) supplemented with antibiotics (50 µg/ml of Str, Spc, Kan, or Gen) and incubated for 2 weeks (*L. biflexa*) or 5 weeks (*L. interrogans*) at 30°C.

L1-130 and 56601 (Table 2). Str-resistant mutants (MIC of >100 µg/ml) had mutations in codons 43 and 88 (AAA to AGA) of the *rpsL* gene, resulting in an amino acid change of lysine to arginine, formerly identified as being responsible for resistance to Str in other bacteria (6, 15). Although the Etest has not been validated for testing antimicrobial susceptibility in leptospires, this method also provides an indication of the *in vitro* susceptibility of *L. biflexa* strains (Fig. 1C).

Generating a gentamicin resistance marker. *L. interrogans* *flgB* and *hsp10* promoters were amplified with the FlgA/FlgC and HspA/HspC primer pairs (Table 3), respectively, and were inserted into the PvuII restriction site of the *E. coli*-*L. biflexa* shuttle vector pSLe94 (1), generating pSLe94PF and pSLe94PH, respectively. The promoterless Gen resistance marker was amplified by PCR with primers GTBam5 and GTXho3 (Table 3) from pBSV2G (4). The amplified product was then digested with BamHI and XhoI and inserted into the corresponding sites of pSLe94PF and pSLe94PH, resulting in pSLe94PfGenta and pSLe94PhGenta, respectively (Fig. 2). The plasmids were then transferred into *L. biflexa* through electroporation (11). The MIC for *L. biflexa* transformants expressing the Gen cassette via the two *L.*

interrogans promoters was ~20 µg/ml (versus 4 µg/ml for the *L. biflexa* wild-type strain) in both liquid and solid media; we then typically selected for transformants on solid media at a concentration of 10 µg/ml. At this concentration, spontaneous Gen-resistant mutants appeared at a frequency of 10⁻⁹ in *L. biflexa*.

We also used the *Himar1* transposon to deliver the Gen resistance cassette into the pathogen *L. interrogans*. The Gen resistance cassette was amplified from pSLE94PfGenta with FgAsc and GTAsc3 (Table 3). The PCR products were then purified, digested with AscI, and inserted into the AscI restriction site of the Kan-resistant transposon carried by pMKL (2). After a 6-week incubation, we selected transformants for further identification of the site of transposon insertion, as previously described (14). The MIC of Gen for an *L. interrogans* serovar Lai transformant that contained an insertion at position 3064024 in the intergenic region in the large chromosome (CI) was >50 µg/ml (versus 4 µg/ml for the wild-type strain) in liquid media, therefore demonstrating that this cassette confers resistance to Gen in pathogenic *Leptospira*.

In conclusion, these results may facilitate the generation of

TABLE 3. Primers used in the study

Primer and function	Sequence (5'-3')	Target
<i>rrs</i> and <i>rpsL</i> sequencing in <i>Leptospira</i> spp.		
rpsLai1	ACTTCGCAACTTGCACGTCG	<i>L. interrogans rpsL</i>
rpsLai2	TGTGAATATGAAGTCGTTTC	<i>L. interrogans rpsL</i>
LA	CATGTTTAAAGACGCGCCGCC	<i>L. interrogans rrs₁</i> and <i>rrs₂</i>
LRS1	GACGGGGGTCCGCACAAGC	<i>L. interrogans rrs₁</i> and <i>rrs₂</i>
LRS2	AGGTCCCACACACTTTACAG	<i>L. interrogans rrs₁</i>
LRS3	TTTAAGTCCGAAAAGTAGT	<i>L. interrogans rrs₂</i>
rpsLPat1	ATGACGATTAGTCAACAATCG	<i>L. biflexa rpsL</i>
rpsLPat2	TTACCATCTACCATCAGG	<i>L. biflexa rpsL</i>
16SPat1	ATGTGACGGTTCCTGGTAGC	<i>L. biflexa rrs₁</i> and <i>rrs₂</i>
16SPat2	ACCCTGAGTAGCGAAGCGTA	<i>L. biflexa rrs₁</i> and <i>rrs₂</i>
16SPat3	ATGTGACGGTTCCTGGTAGC	<i>L. biflexa rrs₁</i>
16SPat4	ACCCTGAGTAGCGAAGCGTA	<i>L. biflexa rrs₂</i>
Generation of gentamicin cassette-containing plasmid constructs		
GTXho3	CCGCCTCGAGACGCGTAAGCCGATCTCGG	Gen resistance marker
FgAsc	TTGGCGCGCCATGAACTAATTGTACAGC	<i>L. interrogans PflgB</i>
GTAsc3	TTGGCGCGCCACGCGTAAGCCGATCTCGG	Gen resistance marker
GTBam5	GCGGGATCCTTACGCAGCAACGATG	Gen resistance marker
HspA	GAATTCTCTAAAAGTATGAATTCCTAC	<i>L. interrogans Phsp10</i>
HspC	CTCGAGCCCGGGGATCCCGTGATGGTGATGGTG ATGAATCGATGCCATAG	<i>L. interrogans Phsp10</i>
FlgA	GTAATCTTTTAAATTTAGCACTTC	<i>L. interrogans PflgB</i>
FlgC	CTCGAGCCCGGGGATCCCGTGATGGTGATGGTG ATGTTTCTCAAACATTA	<i>L. interrogans PflgB</i>

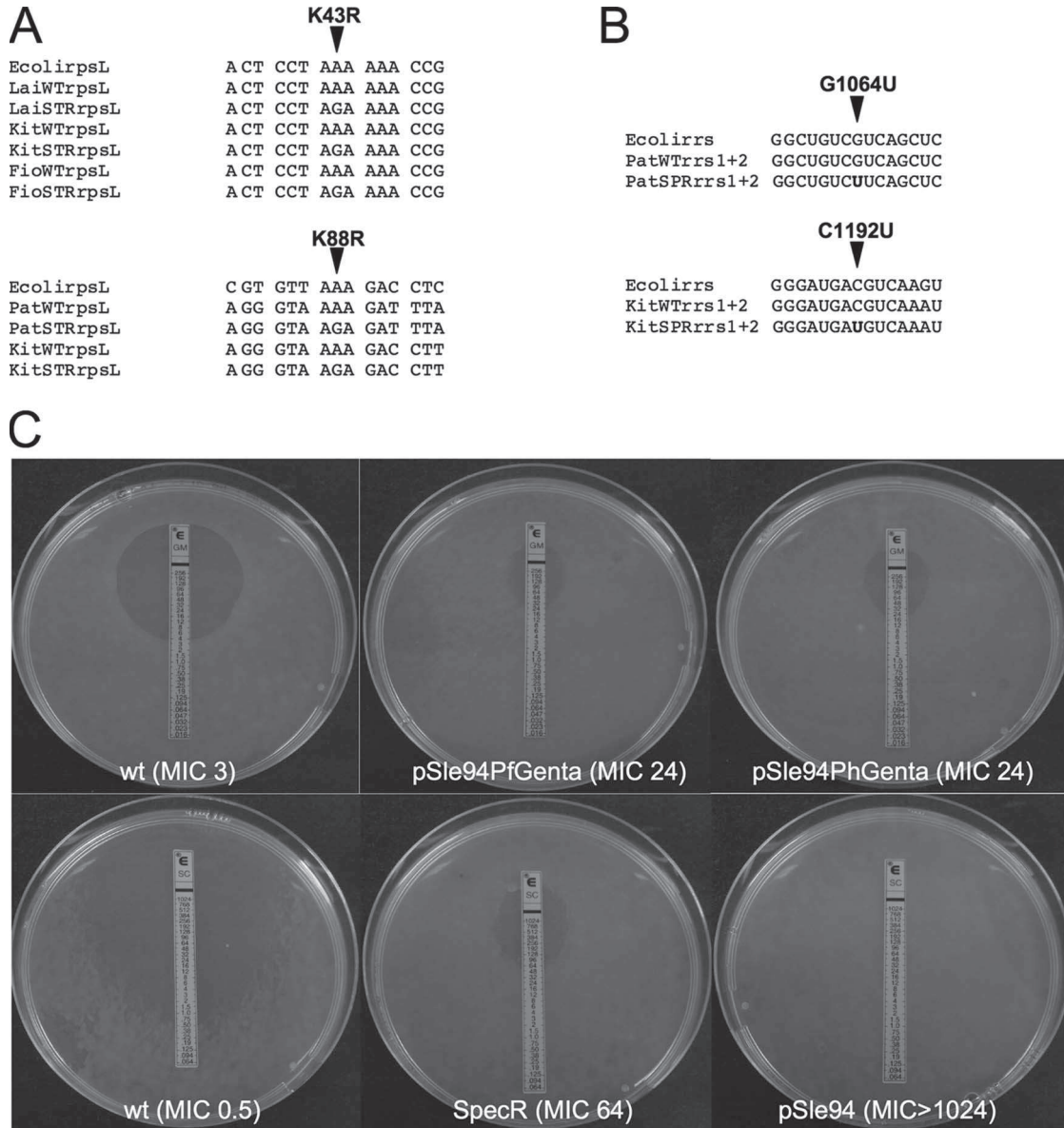


FIG. 1. Antibiotic resistance in *Leptospira* spp. Nucleotide alignments of the *rpsL* (A) and *ms* (B) genes of *E. coli* (Ecoli), *L. biflexa* (strain Patoc 1 [PatWT]), and *L. interrogans* (strains 56601 [LaiWT], Kito [KitWT], and Fiocruz L1-130 [FioWT]) wild-type strains and strains showing spontaneous resistance to Spc (SPR) and Str (STR). Boldface residues indicate the altered regions. (C) Antimicrobial susceptibility by Etest strips (AB Biodisk). Approximately 10^6 bacteria were spread onto EMJH plates. After overnight incubation, the strips were placed on the plates for 1 week at 30°C. The MIC was read as the concentration at which the border of the zone of growth inhibition intersected the scale in the strips. The results are representative of at least four independent observations.

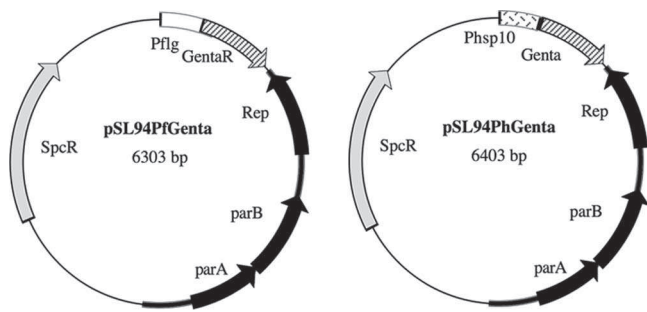


FIG. 2. Schematic diagram of plasmid constructs used to express the gentamicin resistance cassette (Genta). The determinants for replication in *L. biflexa* (*parA*, *parB*, and *rep*), as well as the Spc resistance cassette (SpcR) and *L. interrogans hsp10* (Phsp10) and *flgB* (Pflg) promoters, are indicated.

multiple knockout mutant strains or the use of compatible leptospiral vectors (18, 21).

We thank P. Stewart for the generous gift of the gentamicin-resistant plasmid and S. Brémont for susceptibility testing of *Leptospira*.

REFERENCES

1. Bauby, H., I. Saint Girons, and M. Picardeau. 2003. Construction and complementation of the first auxotrophic mutant in the spirochaete *Leptospira meyeri*. *Microbiology* **149**:689–693.
2. Bourhy, P., H. Louvel, I. Saint Girons, and M. Picardeau. 2005. Random insertional mutagenesis of *Leptospira interrogans*, the agent of leptospirosis, using a *mariner* transposon. *J. Bacteriol.* **187**:3255–3258.
3. Criswell, D., V. L. Tobiason, J. S. Lodmell, and D. S. Samuels. 2006. Mutations conferring aminoglycoside and spectinomycin resistance in *Borrelia burgdorferi*. *Antimicrob. Agents Chemother.* **50**:445–452.

4. Elias, A. F., P. E. Stewart, D. Grimm, M. J. Caimano, C. H. Eggers, K. Tilly, J. L. Bono, D. R. Akins, J. D. Radolf, T. G. Schwan, and P. Rosa. 2002. Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: implications for mutagenesis in an infectious strain background. *Infect. Immun.* **70**:2139–2150.
5. Fromm, H., M. Edelman, D. Aviv, and E. Galun. 1987. The molecular basis for rDNA-dependent spectinomycin resistance in *Nicotiana* chloroplasts. *EMBO J.* **6**:3233–3237.
6. Funatsu, G., and H. G. Wittman. 1972. Ribosomal proteins. XXXIII. Location of amino-acid replacements in protein S12 isolated from *Escherichia coli* mutants resistant to streptomycin. *J. Mol. Biol.* **68**:547–550.
7. Galimand, M., G. Gerbaud, and P. Courvalin. 2000. Spectinomycin resistance in *Neisseria* spp. due to mutations in 16S rRNA. *Antimicrob. Agents Chemother.* **44**:1365–1366.
8. Ko, A. I., C. Goarant, and M. Picardeau. 2009. *Leptospira*: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat. Rev. Microbiol.* **7**:736–747.
9. Lederberg, J. 1951. Streptomycin resistance: a genetically recessive mutation. *J. Bacteriol.* **61**:549–550.
10. Levin, B. R., V. Perrot, and N. Walker. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **154**:985–997.
11. Louvel, H., and M. Picardeau. 2007. Genetic manipulation of *Leptospira biflexa*. J. Wiley and Sons, Hoboken, NJ.
12. Makosky, P. C., and A. E. Dahlberg. 1987. Spectinomycin resistance at site 1192 in 16S ribosomal RNA of *E. coli*: an analysis of three mutants. *Biochimie* **69**:885–889.
13. Murray, C. K., and D. R. Hospenthal. 2004. Broth microdilution susceptibility testing for *Leptospira* spp. *Antimicrob. Agents Chemother.* **48**:1548–1552.
14. Murray, G. L., V. Morel, G. M. Cerqueira, J. Croda, A. Srikram, R. Henry, A. I. Ko, O. A. Dellagostin, D. M. Bulach, R. Sermiswan, B. Adler, and M. Picardeau. 2009. Genome-wide transposon mutagenesis in pathogenic *Leptospira* spp. *Infect. Immun.* **77**:810–816.
15. Okamoto-Hosoya, Y., T. Hosaka, and K. Ochi. 2003. An aberrant protein synthesis activity is linked with antibiotic overproduction in *rpsL* mutants of *Streptomyces coelicolor* A3(2). *Microbiology* **149**:3299–3309.
16. Picardeau, M. 2008. Conjugative transfer between *Escherichia coli* and *Leptospira* spp. as a new genetic tool. *Appl. Environ. Microbiol.* **74**:319–322.
17. Picardeau, M., A. Brenot, and I. Saint Girons. 2001. First evidence for gene replacement in *Leptospira* spp. Inactivation of *L. biflexa* *flaB* results in non-motile mutants deficient in endoflagella. *Mol. Microbiol.* **40**:189–199.
18. Picardeau, M., D. M. Bulach, C. Bouchier, R. L. Zuerner, N. Zidane, P. J. Wilson, S. Creno, E. S. Kuczek, S. Bommezzadri, J. C. Davis, A. McGrath, M. J. Johnson, C. Boursaux-Eude, T. Seemann, Z. Rouy, R. L. Coppel, J. I. Rood, A. Lajus, J. K. Davies, C. Médigue, and B. Adler. 2008. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS One* **3**:e1607.
19. Recht, M. I., and J. D. Puglisi. 2001. Aminoglycoside resistance with homogeneous and heterogeneous populations of antibiotic-resistant ribosomes. *Antimicrob. Agents Chemother.* **45**:2414–2419.
20. Reyat, J. M., V. Pelicic, B. Gicquel, and R. Rappuoli. 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* **66**:4011–4017.
21. Saint Girons, I., P. Bourhy, C. Ottone, M. Picardeau, D. Yelton, R. W. Hendrix, P. Glaser, and N. Charon. 2000. The LE1 bacteriophage replicates as a plasmid within *Leptospira biflexa*: construction of an *L. biflexa*-*Escherichia coli* shuttle vector. *J. Bacteriol.* **182**:5700–5705.
22. Saint Girons, I., D. Margarita, P. Amouriaux, and G. Baranton. 1990. First isolation of bacteriophages for a spirochaete: potential genetic tools for *Leptospira*. *Res. Microbiol.* **141**:1131–1138.



A.4. Referências bibliográficas

AURAN, N. E., JOHNSON, R. C. and RITZI, D. M. Isolation of the outer sheath of *Leptospira* and its immunogenic properties in hamsters. *Infection and immunity*, v. 5, p. 968-975, 1972.

BEEMAN, R. W. and STAUTH, D. M. Rapid cloning of insect transposon insertion junctions using 'universal' PCR. *Insect molecular biology*, v. 6, p. 83-88, 1997.

BOURHY, P., et al. Random insertional mutagenesis of *Leptospira interrogans*, the agent of leptospirosis, using a mariner transposon. *Journal of bacteriology*, v. 187, p. 3255-3258, 2005.

BROMLEY, D. B. and CHARON, N. W. Axial filament involvement in the motility of *Leptospira interrogans*. *Journal of bacteriology*, v. 137, p. 1406-1412, 1979.

CHIANG, S. L. and RUBIN, E. J. Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene*, v. 296, p. 179-185, 2002.

KAN, W. and WOLGEMUTH, C. W. The shape and dynamics of the Leptospiraceae. *Biophysical journal*, v. 93, p. 54-61, 2007.

LOUVEL, H., BETTON, J. M. and PICARDEAU, M. Heme rescues a two-component system *Leptospira biflexa* mutant. *BMC microbiology*, v. 8, p. 25, 2008.

LOUVEL, H. and PICARDEAU, M. Genetic manipulation of *Leptospira biflexa*. *Current protocols in microbiology*, v. Chapter 12, p. Unit 12E 14, 2007.

LOUVEL, H., SAINT GIRONS, I. and PICARDEAU, M. Isolation and characterization of FecA- and FeoB-mediated iron acquisition systems of the spirochete *Leptospira biflexa* by random insertional mutagenesis. *Journal of bacteriology*, v. 187, p. 3249-3254, 2005.

PICARDEAU, M. Conjugative transfer between *Escherichia coli* and *Leptospira* spp. as a new genetic tool. *Applied and environmental microbiology*, v. 74, p. 319-322, 2008.

PICARDEAU, M., BRENOT, A. and SAINT GIRONS, I. First evidence for gene replacement in *Leptospira* spp. Inactivation of *L. biflexa* flaB results in non-motile mutants deficient in endoflagella. *Molecular microbiology*, v. 40, p. 189-199, 2001.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha tese de Doutorado intitulada "Estudos estruturais de proteínas de *Leptospira interrogans* sorovar Copenhageni potencialmente localizadas no envelope celular":

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

(x) tem autorização da(s) seguinte(s) Comissão(ões) de Bioética ou Biossegurança*:

CIBio, sob Protocolo(s) nº BGG 04.01.

* Caso a Comissão seja externa à UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.

Pirila Ol. de Giuseppe

Aluno:

Beatriz Alves Pinheiro

Orientador:

Para uso da Comissão ou Comitê pertinente:

(x) Deferido () Indeferido

Nome:

Função:

Helena Coutinho F. de Oliveira
Prof. Dra. HELENA COUTINHO F. DE OLIVEIRA
Presidente
Comissão Interna de Biossegurança
CIBio/IB - UNICAMP

Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança da ABTLuS – Associação Brasileira de Tecnologia de Luz Síncrotron

Título do projeto: Estudos Estruturais de Proteínas de *Leptospira interrogans* sorovar copenhageni potencialmente imunogênicas

Pesquisador responsável: Beatriz Gomes Guimarães

Experimentador: Priscila Oliveira de Giuseppe

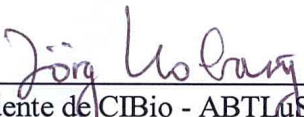
Nível do treinamento do experimentador: []-Iniciação científica, []-mestrado, []-doutorado, [X]-doutorado direto, []-pós-doutorado, []-nível técnico, []-outro, especifique: _____

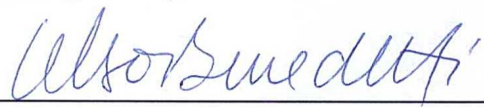
Resumo do projeto:


A leptospirose é uma zoonose de distribuição mundial que, na última década, emergiu como um sério problema de saúde pública. Mais de 200 linhagens da espécie *Leptospira interrogans* já foram reconhecidas como seus agentes causadores. Em humanos, a forma severa é frequentemente causada por membros do sorogrupo Icterohaemorrhagiae dentre eles a linhagem copenhageni. Recentemente, o genoma desta foi seqüenciado, e a partir dele um grande número de proteínas localizadas na superfície celular foi identificado, dentre as quais uma considerável parcela não apresenta função molecular conhecida. A predição da função a partir da estrutura tridimensional da proteína é o método geralmente empregado quando esta não pode ser inferida a partir de comparação sequencial. A estrutura tridimensional de uma proteína pode revelar vários indícios funcionais, que são obtidos, basicamente, por homologias estruturais e pela presença de ligantes na proteína. Além disso, o conhecimento da estrutura tridimensional pode levar ao estabelecimento de relações evolutivas e à determinação de mecanismos de interações entre domínios protéicos. Dentro desse contexto, esse projeto visa à resolução da estrutura cristalográfica por difração de raios X das proteínas LIC10793, LIC12922 e LIC10494 de *Leptospira interrogans* sorovar copenhageni. Tais proteínas são potencialmente localizadas, respectivamente, na membrana interna, no periplasma e na membrana externa da parede celular, são reconhecidas por anticorpos de pacientes acometidos pela leptospirose e não são funcionalmente caracterizadas. Sendo assim, acredita-se que a partir da estrutura tridimensional das mesmas possam ser extraídos indícios funcionais que poderão ter implicações na elucidação dos mecanismos moleculares envolvidos na interação entre este patógeno e seus hospedeiros.

A CIBio analisou este projeto em reunião realizada no dia: 4.9.2007.

Parecer final: [X]-projeto aprovado, []-projeto recusado, []-projeto com deficiências, favor comentários abaixo:


 Presidente de CIBio - ABTLuS
 Prof. Dr. Jörg Kobarg


 Membro da CIBio - ABTLuS
 Prof. Dr. Celso Eduardo Benedetti


 Membro da CIBio - ABTLuS
 Prof. Dr. Nilson Ivo Tonin Zanchin