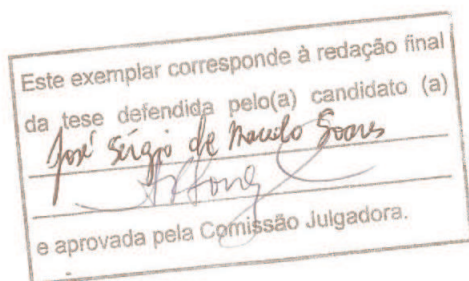


UNIVERSIDADE ESTADUAL DE CAMPINAS

José Sérgio de Macedo Soares

**“Modelagem Molecular e Caracterização Espectroscópica
de duas Proteínas de *Xylella fastidiosa* Potencialmente
Envolvidas com Fitopatogenicidade”**



Tese (ou Dissertação)
apresentada ao Instituto de
Biologia para obtenção do
Título de Mestre em Genética
e Biologia Molecular, na área
de Genética de
Microrganismos.

Orientador(a): Prof(a). Dr(a). Anete Pereira de Souza

Campinas, 2007

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

So11m

Soares, José Sérgio de Macedo

Modelagem molecular e caracterização espectroscópica de duas proteínas de *Xylella fastidiosa* potencialmente envolvidas com fitopatogenicidade / José Sérgio de Macedo Soares. – Campinas, SP: [s.n.], 2007.

Orientadora: Anete Pereira de Souza.

Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. *Xylella fastidiosa*. 2. Proteínas de membrana. 3. Dimetiladenosina transferase. 4. Proteínas - Purificação. 5. Genética - Expressão. I. Souza, Anete Pereira de. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Molecular modeling and spectroscopic characterization of two proteins from *Xylella fastidiosa* potentially involved with phytopathogenicity.

Palavras-chave em inglês: *Xylella fastidiosa*; Membrane proteins; Dimethyladenosine transferase; Proteins – Purification; Gene expression.

Área de concentração: Genética de Microorganismos.

Titulação: Mestre em Genética e Biologia Molecular.

Banca examinadora: Anete Pereira de Souza, Marco Aurélio Takita, Gonçalo Amarante Guimarães Pereira.

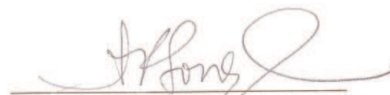
Data da defesa: 22/02/2007.

Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 22 de Fevereiro de 2007

BANCA EXAMINADORA

Prof(a). Dr(a) . Anete Pereira de Souza (Orientador(a))


Assinatura

Prof(a). Dr(a). Marco Aurélio Takita


Assinatura

Prof(a). Dr(a). Gonçalo Amarante Guimarães Pereira


Assinatura

Prof(a). Dr(a) . Beatriz Gomes Guimarães


Assinatura

Prof(a). Dr(a) . Ione Salgado Martins


Assinatura

Agradecimentos

À profa. Dra. Anete Pereira de Souza pela orientação e oportunidade concedida.

Ao Dr. e acima de tudo amigo Adriano Azzoni pelos ensinamentos e paciência no início do trabalho.

Aos professores Dr. Marco Aurélio Takita e Dra. Beatriz Gomes Guimarães pelas críticas e sugestões no exame de pré-banca da minha tese.

Aos professores Dr. Marco Aurélio Takita e Dr. Gonçalo Amarante Pereira pela participação na banca de tese.

Ao prof. Dr. Nilson Zanchin do Laboratório Nacional de Luz Sincroton pelas sugestões oferecidas ao longo do trabalho.

Ao Dr. Gustavo Henrique Ferreira do Laboratório Thomson do Instituto de Química da Universidade Estadual de Campinas pela grande amizade e ajuda com os experimentos envolvendo Espectrometria de Massas e Modelagem Molecular.

Ao prof. Dr. Carlos Henrique Ramos do Laboratório Nacional de Luz Sincroton pela ajuda com os experimentos de Dicroísmo Circular.

A técnica Maria Zilda pela assistência no laboratório. Sem a presença dela o trabalho jamais seria concretizado.

Aos meus colegas e companheiros do Laboratório de Análise Genética e Molecular e em especial a Dilaine Rose Schneider por todo o suporte, ensinamentos, e amizade. Sou grato a todos do laboratório pelo respeito e dedicação de todos no esforço diário em busca de resultados.

A minha gerente da IBM Regina Shizuno pela oportunidade e compreensão.

Aos colegas de trabalho da IBM por toda paciência e compreensão em inúmeras horas extras de trabalho realizadas.

Aos amigos de forma geral pela amizade e sinceridade que de certa forma contribuíram direta ou indiretamente para a realização deste trabalho.

Um agradecimento muito especial a minha companheira e namorada Cinthia Gonçalves Lenz Cesar por todo amor e carinho proporcionado, pela paciência demonstrada em vários momentos, sendo estes os melhores momentos da minha vida.

ÍNDICE

Resumo	1
Abstract	3
1. Introdução	5
2. Revisão Bibliográfica	
2.1. <i>Xylella fastidiosa</i> : patogenicidade e genômica	7
2.2. Proteína transmembrana XpsL de <i>Xylella fastidiosa</i>	10
2.2.1. Sistema de secreção do tipo II	11
2.2.2. Complexo da Membrana Interna	13
2.2.3. Biogênese do Pili do Complexo do tipo IV	17
2.2.4. Complexo da Membrana Externa	18
2.3. Proteína Dimetiladenosina transferase de <i>Xylella fastidiosa</i>	18
2.3.1. Metil Transferases (MTases)	20
2.3.2. Cofator S-adenosyl-L-methionine (SAM)	20
2.3.3. Adenina dimetil transferase de <i>Escherichia coli</i>	24
2.4. Análise Espectroscópica por Dicroísmo Circular	26
2.5. Espectrometria de Massas	28
2.5.1. MALDI-TOF	29
3. Objetivos	
3.1. Objetivos gerais	32
3.2. Objetivos específicos	32

4. Artigo 1: “Cloning, Expression and Purification of a transmembrane protein XpsL of secretion pathway of type II from the phytopathogen <i>Xylella fastidiosa</i> ”	33
5. Artigo 2: “Expression, Purification and Homology Modeling of a Dimethyladenosine transferase from the phytopathogen <i>Xylella fastidiosa</i> ”	52
6. Resultados Complementares	
6.1. Amplificação das <i>orfs</i> e subclonagem no vetor pGEM-T	80
6.2. Digestão enzimática utilizando endonucleases de restrição e ligação nos vetores pET28 e 29a	83
6.3. Expressão e purificação em baixa escala das proteínas recombinantes	85
6.4. Proteína XpsL expressa sem o peptídeo sinal (59 kDa)	86
6.5. Amplificação da <i>orf xf1524</i> sem a sequência referente ao peptídeo sinal	86
6.6. Expressão e purificação da proteína XpsL sem o peptídeo sinal	88
6.7. Expressão e purificação da proteína XpsL recombinante (61 kDa) adicionando detergente ao tampão de lise	89
7. Conclusões	91
8. Perspectivas	92
9. Referências Bibliográficas	93

Resumo

O fitopatogeno *Xylella fastidiosa* é uma bactéria limitada ao xilema que causa uma variedade de doenças economicamente importantes em plantas, incluindo uma doença chamada Clorose Variegada de Citrus (CVC). Recentemente, os dados de sequenciamento do DNA fornecidos pelo projeto genoma da *Xylella fastidiosa* permitiram uma aproximação da área genoma funcional para investigar a função de diversas proteínas baseadas na informação sobre suas estruturas. Avanços recentes na área genoma estrutural não somente nos ajudou a compreender funções da proteína, mas também causou um grande impacto na indústria farmacêutica. Nos últimos anos, o uso da informação estrutural de proteínas na pesquisa de descoberta de novas drogas tem amadurecido, e é usado agora em todos os níveis, variando da identificação e da seleção de alvos do genoma a candidatos apropriados. Esta emergente área é poderoso meio para compreender mais profundamente os mecanismos a patogenicidade da bactéria. Este trabalho visa adicionar novas informações sobre proteínas que podem ser relacionadas a patogenicidade de *Xylella fastidiosa*, necessárias para o desenvolvimento de novos métodos de combate à CVC.

A fim de recolher informações sobre as proteínas envolvidas nos mecanismos do patogenicidade da bactéria, nós escolhemos as *orf xf2148* e *orf xf1524* de *Xylella fastidiosa* para estudos de caracterização. A *orf xf2148*, apresenta 47% de similaridade com o gene *ksgA* de *Escherichia coli*. O gene *ksgA* codifica uma Dimetiladenosina transferase da classe de proteínas metiltransferases. Já a *orf xf1524*, apresenta 55% de similaridade com o gene *xpsL* de *Xanthomonas campestris*, que codifica uma proteína transmembrana da via de

secreção do tipo II. Ambas as proteínas foram clonadas no vetor pET32Xa/LIC e expressas na bactéria *Escherichia coli* linhagem BL21(DE3).

A proteína XPSL foi purificada através de uma cromatografia de afinidade (IMAC) e teve sua identidade determinada por SDS-PAGE e Espectrometria de Massas (MALDI-TOF). Para investigar a integridade estrutural da proteína XPSL purificada, a técnica espectroscópica de Dicroísmo Circular (CD) foi realizada. O espectro de CD da proteína XPSL mostrou sinais predominantemente de α -hélices indicando um perfil de estrutura secundária devidamente enovelada viável para estudos estruturais e funcionais.

A proteína XF2148, expressa em sua forma insolúvel, foi resolubilizada usando 8M de uréia como agente desnaturante. A proteína foi purificada através de uma cromatografia de troca iônica e sua pureza e identidade verificadas por SDS-PAGE e Espectrometria de Massas (MALDI-TOF). Seu correto enovelamento por espectroscopia de Dicroísmo Circular (CD) indicou uma composição predominantemente de α -hélices. O alinhamento da sequência primária e modelagem por homologia da proteína XF2148 com uma proteína de estrutura tridimensional resolvida, KsgA de *E. coli*, revelaram um local da molécula que envolve resíduos altamente conservados no domínio C-terminal e cinco dos oito motivos estruturais encontrados geralmente em Metiltransferases.

Abstract

The phytopathogen *Xylella fastidiosa* is a xylem-limited bacterium that causes a range of economically important plant diseases, including a serious disease of orange trees called citrus variegated chlorosis (CVC). Recently, DNA sequence data provided by the *Xylella fastidiosa* Brazilian genome project have allowed a functional genomic approach to investigate the function of several proteins based on the information about their structures. Recent advances in structural genomics not only help us to understand protein functions but also have a big impact on the pharmaceutical industry. In the last few years, the use of protein structural information in drug discovery research has matured, and it is now used at all levels, ranging from genomics target identification and selection to the final design of suitable drug candidates. This emergent area is a powerful means to more deeply understand the mechanisms of the bacterium pathogenicity. This work aims at adding new information on proteins that may be related to the *X. fastidiosa* pathogenesis, necessary for new approaches towards the combat of CVC.

In order to gather information about the proteins involved in the mechanisms of the bacterium pathogenicity, we chose *orf xf2148* and *orf xf1524* from *Xylella fastidiosa* for characterization studies. *Orf xf2148*, presents 47% of similarity with the gene *ksgA* of *Escherichia coli*. The *ksgA* gene encodes a dimethyladenosine transferase from the methyl transferase proteins class. On the other hand *orf xf1524*, presents 55% of similarity with the *xpsL* gene of *Xanthomonas campestris*, which encodes a transmembrane protein of the type II secretion machinery. Both proteins were cloned into the pET32Xa/LIC vector to overexpress the protein in *Escherichia coli* BL21(DE3).

The expressed XpsL protein was purified by immobilized metal affinity chromatography (IMAC) and had its identity determined by SDS-PAGE and Mass Spectrometry (MALDI-TOF). To investigate the structural integrity of the purified XPSL, the protein was analyzed by Circular Dichroism (CD) spectroscopy. The XpsL CD spectrum showed that it contains predominantly signal of α -helices suggesting that the XpsL recombinant protein maintains secondary structures, is viable to functional and structural studies and remained folded.

The XF2148 protein was expressed in the insoluble form and refolded using 8M Urea as denaturing agent. The protein was purified in one step by cation-exchange chromatography and its purity and identity were verified by SDS-PAGE and Mass Spectrometry (MALDI-TOF). Its correct folding was verified by Circular Dichroism Spectroscopy analysis that indicated a secondary structure composed mainly of α -helices. The alignment of the XF2148 primary sequence and homology modeling with one ksgA structure solved protein from *E. coli*, revealed a site involving highly conserved residues in the C-terminal domain and five of the eight structural motifs usually found in AdoMet-dependent methyltransferases.

1. Introdução

Xylella fastidiosa é uma bactéria que ocorre limitada aos vasos do xilema de um grande número de plantas cultivadas, árvores, ornamentais e outras plantas silvestres, em várias partes do mundo.

No Brasil, essa bactéria tem sido encontrada em varias plantas, porém as culturas de importância econômica mais afetadas são ameixeira, cafeeiro e citros. Destas, apenas em citros tem-se uma estimativa dos prejuízos causados pela bactéria. No estado de São Paulo, onde a citricultura é constituída por mais de 60 milhões de árvores, o prejuízo pode chegar a mais de um milhão de dólares, sendo que, aproximadamente 30% das plantas apresentam a bactéria (Monteiro *et al.*, 2001).

Apesar do grande número de estudos realizados com *Xylella fastidiosa* nestes últimos anos, pouco ainda se conhece a respeito de sua adesão, movimentação vaso a vaso, translocação sistêmica e expressão dos sintomas em plantas, além de seu comportamento no inseto vetor. Como ainda não há uma forma específica de combate à *Xylella fastidiosa*, os citricultores estão implantando em seus pomares as estratégias preventivas de combate a doença.

O explosivo crescimento da informação de seqüências genéticas tem oferecido coleções de seqüências de genes que codificam proteínas encontradas em muitos organismos (Brenner, 2001). Com o sequenciamento do genoma de *Xylella fastidiosa*, uma nova área de pesquisa vem sendo utilizada na tentativa de combater de forma mais eficaz doenças causadas por microrganismos, a área de Genoma Estrutural.

O Genoma Estrutural é uma consequência dos projetos genoma, que vêm permitindo o seqüenciamento genético do DNA de diversos organismos. A determinação estrutural de proteínas baseadas na informação de genomas tenta esclarecer a função e o mecanismo de ação de uma determinada proteína através de sua estrutura tridimensional.

A indústria farmacêutica e Química têm adotado a genômica estrutural como um meio de identificar novos alvos biológicos para desenvolver novas drogas a partir da estrutura tridimensional de proteínas no combate a doenças causadas por microrganismos.

Com o intuito de levantar informações estruturais preliminares por métodos espectroscópicos de dicroísmo circular, espectrometria de massas e modelagem tridimensional, uma proteína transmembrana da via de secreção do tipo II e uma enzima dimetiladenosina transferase foram estudadas de forma a contribuir com os conhecimentos do mecanismo de patogenicidade da bactéria *Xylella fastidiosa* na busca por um método mais eficiente de combate à doença.

2. Revisão Bibliográfica

2.1. *Xylella fastidiosa*: patogenicidade e genômica

A Clorose Variegada do Citros (CVC) foi primeiramente relatada em 1987 nos estados de São Paulo e Minas Gerais (Machado, 2003). Esta doença tem sido um dos maiores problemas da citricultura brasileira, atingindo até 43,56% dos pomares infectados pela bactéria responsável pelo aparecimento dos sintomas. Embora a doença esteja localizada na América do Sul, trata-se de uma potencial ameaça para todas as citriculturas do mundo, incluindo os Estados Unidos (Teixeira *et al.*, 2004).

A CVC ou amarelinho é uma anomalia causada pela bactéria *Xylella fastidiosa*, uma bactéria gram negativa, não flagelada, limitada ao xilema que infecta plantas de diversos grupos taxonômicos e causa doenças economicamente importantes (Lambais *et al.*, 2000). Os maiores sintomas da doença, como necrose e abscisão foliar, menores teores de zinco, potássio e declínio do vigor levando a morte da planta, sugerem uma disfunção do sistema de condução de água por oclusão vascular (Hopkins, 1989).

A *Xylella fastidiosa* é transmitida por mudas contaminadas e vetores homópteros adultos, conhecidos popularmente como “cigarrinhas”, que se alimentam de xilema. As “cigarrinhas” são capazes de transmitir a doença adquirindo a bactéria de árvores contaminadas e ao se alimentarem de árvores saudáveis, transmitem a doença, sendo que, atualmente, encontra-se disseminada em todas as regiões de citricultura do país (Roberto *et al.*, 1996). Entre as medidas mais importantes de manejo da doença está o controle de cigarrinhas no pomar (www.fundecitrus.com.br).

O explosivo crescimento da informação de seqüências genéticas tem oferecido coleções de seqüências de genes que codificam proteínas encontradas em muitos organismos (Brenner, 2001). Obtendo-se a biblioteca genômica e conhecendo-se sua seqüência de bases, torna-se possível, através de ferramentas moleculares, estudar as proteínas expressas por genes, bem como relacionar sua estrutura à sua função.

Umas das áreas de pesquisa mais populares na era pós-genômica refere-se a determinação estrutural de proteínas baseadas na informação de genomas. Muitos projetos visam determinar as estruturas de proteínas solúveis (Kyogoku *et al.*, 2003). Entre 25 a 30% das proteínas, no entanto, estão embebidas em membranas, incluindo receptores, canais, transportadores e conversores de energia, os quais possuem significativa função biológica. Suas estruturas devem ser determinadas não só para explicar suas funções biológicas, mas para desenvolver novas drogas e outros compostos de uso industrial devido ao fato de que 60% dos novos fármacos produzidos estão relacionados a proteínas de membrana. Porém estas proteínas apresentam-se insolúvel em soluções aquosas e são difíceis de expressar, purificar e cristalizar. Apenas 2% das estruturas de proteínas depositadas no banco de dados pertencem à categoria de proteínas de membranas (Kyogoku *et al.*, 2003).

A análise da seqüência do genoma da *Xylella fastidiosa* revela alta relação com muitos genes de outros patógenos, especialmente *Xanthomonas campestris*. O alto grau de relação sugere que os genes de *Xylella fastidiosa* e *Xanthomonas campestris* são ortólogos, visto que essas espécies são filogeneticamente próximas (Daniels & Dow, 2000). Em contraste, o tamanho relativo dos genomas de *X. fastidiosa* (~ 2.7Mb) e *X. campestris* (~ 5.5Mb), indica que existem grandes diferenças entre as duas bactérias. As diferenças no

tamanho do genoma provavelmente refletem as diferentes estratégias de doença e capacidade de crescimento entre os dois organismos (Daniels & Dow, 2000).

Os mecanismos envolvidos na interação planta-patógeno durante o processo de patogenicidade são complexos e ainda pouco entendidos. A elucidação da sequência genômica completa da *Xylella fastidiosa* (Simpson *et al.*, 2000) contribuiu significativamente na determinação de um número de potenciais mecanismos de patogenicidade desta bactéria.

O processo patológico da CVC está principalmente relacionado aos mecanismos de adesão, migração e toxicidade bacteriana e oclusão vascular devido a formação do biofilme, sendo este biofilme um importante fator de patogenicidade (de Souza *et al.*, 2005).

Estudos genéticos relacionados à formação do biofilme têm demonstrado que este é formado por múltiplos passos regulados por alguns genes com a sua expressão alterada em cada fase do desenvolvimento. Os passos iniciais da formação do biofilme envolvem a adesão da bactéria à superfície. Quando o biofilme atinge seu estágio maturo, com grande densidade celular, um sistema de sinal intracelular é ativado, permitindo a bactéria regular a expressão de um conjunto de genes específicos como fatores de virulência a genes associados à resistência microbiana. Essas características permitem às células do biofilme uma vantagem na adaptação e competição com o ambiente hospedeiro (de Souza *et al.*, 2005).

Tal como muitos fitopatógenos, esta bactéria produz uma matriz extracelular composta por polissacarídeos extracelulares (EPS), porém não foram encontrados os genes *gumI*, *gumL* e *gumG* que codificam enzimas envolvidas na síntese de EPS de outros fitopatógenos. Sua síntese em *Xylella fastidiosa* seria regulada por vários genes

independentes, organizados no cluster *rpf* (regulation pathogenicity factors) (Simpson *et al.*, 2000).

As fímbrias são estruturas importantes tanto para o processo de colonização do xilema, quanto para promover a adesão da bactéria no inseto vetor. Foram identificados genes que controlam a transcrição de subunidades da fímbria (*pilS* e *pilR*) e outros que codificam um sistema quimiostático traduzindo sinais do meio para o pili (*pilG*, *H*, *I*, *J* e *chpA*) em resposta ao hospedeiro (Simpson *et al.*, 2000).

Também foram identificados genes que codificam a celulase e seu precursor, poligalacturonase. Estas enzimas extracelulares seriam responsáveis pela migração bacteriana na planta, degradando a parede celular das células dos vasos do xilema, facilitando sua entrada e dispersão no hospedeiro (Simpson *et al.*, 2000).

Genes de fatores importantes de virulência em bactérias gram negativas pertencentes a uma família de toxinas (RTX) foram identificados. Tais toxinas lesam ou matam células vegetais, permitindo que o patógeno colonize a planta (Simpson *et al.*, 2000).

O conhecimento de tais genes contribuiu para o entendimento global dos mecanismos de patogenicidade.

2.2. Proteína transmembrana XpsL de *Xylella fastidiosa*

Partindo de um total de 147 *orfs* que possivelmente codificam genes envolvidos na patogenicidade em *Xylella fastidiosa* (Simpson *et al.*, 2000), 3 foram selecionadas para clonagem e expressão em *E. coli*. A *orf xf1524*, de 1155 bp, apresenta 55% de similaridade com o gene *xpsL* de *Xanthomonas campestris* (<http://aeg.lbi.ic.unicamp.br/xf>), o qual codifica uma proteína da via de secreção do tipo II de peso molecular equivalente a 43,4

kDa e 384 resíduos de aminoácidos. A XpsL é uma proteína de membrana citoplasmática e faz parte do complexo protéico da via de secreção do tipo II em *Xanthomonas campestris* pv. *campestris* (Tsai *et al.*, 2002). A via de secreção do tipo II é utilizada por bactérias gram negativas para secretar proteínas extracelulares (Lee *et al.*, 2001). Sendo a proteína XpsL um componente do mecanismo de secreção do tipo II, acredita-se que esta proteína crie um canal por onde passam as proteínas secretadas (Russel, 1998). As proteínas a serem secretadas são produzidas com peptídeos sinais na região N-terminal, o qual permite a translocação Sec dependente através da membrana citoplasmática. Em seguida o peptídeo sinal é removido sendo feito o enovelamento e liberação da proteína madura no espaço periplasmático. Na região periplasmática, proteínas podem sofrer modificações, tais como formação de pontes dissulfeto e aglomeração de subunidades, antes de serem translocadas através da membrana exterior pela via de secreção do tipo II. Esta via é altamente específica podendo distinguir proteínas a serem transportadas de proteínas próprias da região periplasmática (Sandkvist, 2001).

2.2.1. Sistema de secreção do tipo II

Secreção extracelular em bactérias gram negativas requer um sistema de transporte complexo para mover proteínas da região citoplasmática para a região extracelular. Várias vias de secreção altamente especializadas estão envolvidas neste propósito, incluindo o sistema de secreção do tipo II (T2S). A secreção de proteínas via sistema do tipo II ocorre em dois passos distintos genética e bioquimicamente. Primeiro, proteínas a serem secretadas são expressas com peptídeo sinal com o objetivo de se ligarem à maquinaria Sec ou Tat para assim atravessarem a membrana citoplasmática em direção ao periplasma

(Voulhoux *et al.*, 2001; de Keyzer *et al.*, 2003; Palmer *et al.*, 2005). Essa operação é sucedida pela remoção do peptídeo sinal e reconhecimento e translocação das proteínas enoveladas através da membrana por componentes da via de secreção do tipo II.

A via de secreção do tipo II foi descoberta primeiro em *Klebsiella oxytoca*, onde foi encontrada ser necessária à secreção da lipoproteína pululanase (d'Enfert *et al.*, 1987). Desde então, *Vibrio cholerae*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Burkholderia pseudomallei*, *Erwinia chrysanthemi* e *Xanthomonas campestris* são microrganismos que demonstraram fazer uso dessa via para a secreção e direcionamento de toxinas, proteases, celulases e lipases (Dow *et al.*, 1989; Filloux *et al.*, 1990; Reeves *et al.*, 1993; Sandkvist *et al.*, 1997; DeShazer *et al.*, 1999; Lathem *et al.*, 2002; Iwobi *et al.*, 2003; Rossier *et al.*, 2004).

O sistema de secreção do tipo II também está presente em diversas espécies já identificadas através do seqüenciamento genômico (Cianciotto, 2005). O número de genes identificados como sendo essenciais em diferentes organismos para o sistema de via de secreção do tipo II varia entre 12 a 15. Os genes homólogos e suas respectivas proteínas têm sido designados na maioria dos sistemas pelas letras de A a O e S.

Baseado em dados de diferentes experimentos, incluindo localização subcelular e interações proteína-proteína entre componentes do sistema de via de secreção do tipo II, foi proposto que proteínas desse sistema formam um complexo protéico que atravessa todo o envelope celular de bactérias gram negativas. Este complexo multi-protéico inclui componentes do citoplasma, um subcomplexo da membrana interna que atinge o compartimento periplasmático e um poro de secreção da membrana externa.

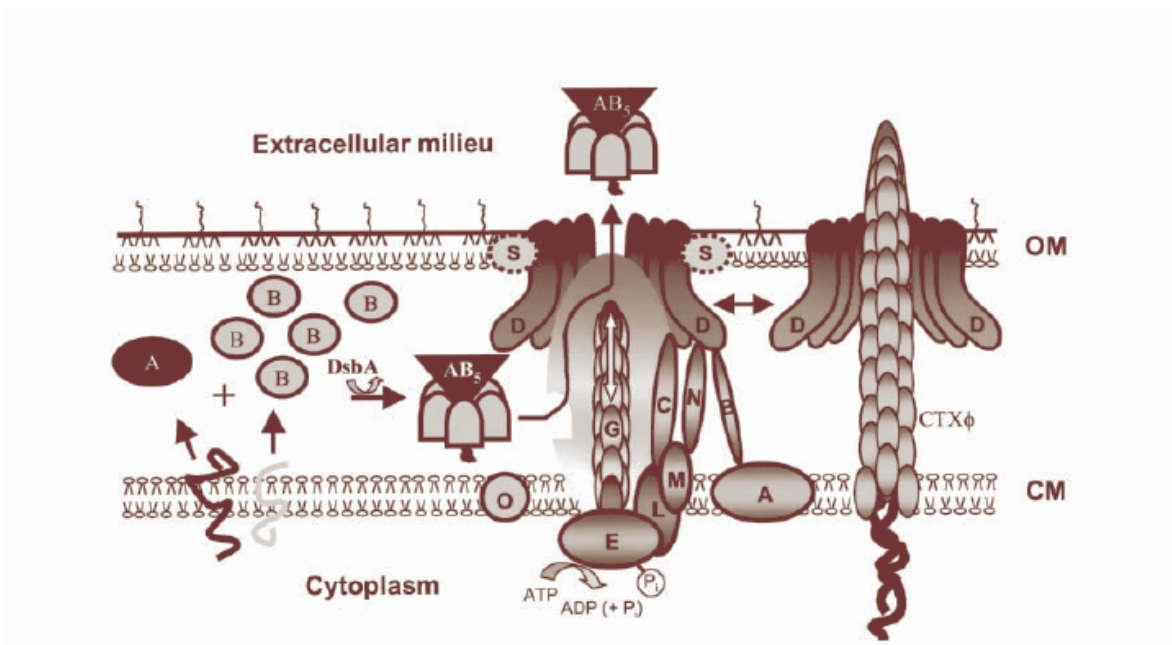


Figura 1. Via de secreção do tipo II em *V. cholerae*. Sandkvist, 2001.

2.2.2. Complexo da Membrana Interna

O sistema de transporte da via de secreção do tipo II requer a presença de uma proteína da família E. Esta proteína citoplasmática está associada à membrana citoplasmática via proteínas de membrana da família L e M. Proteínas da família E são membros de uma grande família de secreção de nucleosídeos trifosfatos, os quais estão relacionados à secreção de proteínas através da hidrólise de nucleosídeos trifosfatos (Planet *et al.*, 2001). Proteínas TS2:E são proteínas com múltiplos domínios. Todas as proteínas TS2:E têm sítios de ligação a ATP, chamados *Walker A boxes*, presentes em proteínas ATPases, ATP sintases e kinases. Recentemente foi demonstrado que a proteína EpsE purificada, uma TS2:E membro da família de *Vibrio cholerae*, é uma ATPase dependente de Mg^{2+} (Camberg & Sandkvist, 2005). Uma mutação introduzida no motivo de ligação de ATP na *Walker A box* reduz a atividade específica da ATPase in vitro e resulta em uma

proteína que não é capaz de auxiliar a secreção *in vivo* (Sandkvist *et al.*, 1995; Camberg & Sandkvist, 2005).

Interessantemente, EpsE apresenta atividade cooperativa com ATPase, e preparações de EpsE purificadas contêm uma pequena fração de oligômeros, especialmente hexâmeros, os quais aumentam a atividade específica da enzima (Camberg & Sandkvist, 2005). Este fato pode sugerir que a forma funcional desta proteína *in vivo* é hexamérica. Vários membros da família ATPase de secreção do tipo IV, como a HP0525 de *Helicobacter pylori* têm sido caracterizados como ATPases hexaméricas (Yeo *et al.*, 2000; Forest *et al.*, 2004).

Proteínas da família E da via de secreção do tipo II são dependentes das proteínas de membrana da família L (Sandkvist *et al.*, 1995; Ball *et al.*, 1999; Py *et al.*, 1999; Possot *et al.*, 2000). Proteínas da família L da via de secreção do tipo II consistem de um domínio citoplasmático, uma única hélice transmembrana e um pequeno domínio periplasmático. O domínio citoplasmático se liga à proteínas da família E na membrana citoplasmática, enquanto tanto a hélice transmembrana quanto o domínio periplasmático interagem com proteínas da família M da via de secreção do tipo II (Sandkvist *et al.*, 2000). Essa interação resulta em uma proteção mútua entre as proteínas da família L e M do complexo protéico contra uma ação proteolítica (Michel *et al.*, 1998; Sandkvist *et al.*, 1999).

A estrutura do domínio citoplasmático da família E de *Vibrio cholerae*, chamada de cyto-EpsL recentemente publicada, mostrou surpreendentemente uma homologia estrutural com a super família de proteínas actin-like ATPases (Abendroth *et al.*, 2004). Cyto-EpsL consiste de três domínios ricos em folhas β , com os domínios I e III da cyto-EpsL correspondendo aos domínios conservados 1A e 2A de actin-like ATPases. Cyto-EpsL não

contém os domínios não conservados 1B e 2B e como consequência, a proteína cyto-EpsL é incapaz de formar filamentos de actina.

Análises de estrutura e sequência sugerem que o domínio II da proteína cyto-EpsL está envolvido com interações proteína-proteína (Abendroth *et al.*, 2004). A ausência dos domínios 1B e 2B, assim como a presença de domínio II flexível, faz com que a cyto-EpsL seja única entre os membros conhecidos da super família das proteínas ATPase. Adicionalmente, diferente da vasta maioria de proteínas da família ATPases, tanto a sequência quanto a estrutura da cyto-EpsL revelaram um sítio de ligação a nucleotídeo (Abendroth *et al.*, 2004). Ao invés disso, essa proteína parece participar indiretamente da hidrólise de ATP através de sua interação com a ATPase EpsE da via de secreção do tipo II.

Estudos mutacionais têm mostrado que a interface entre as proteínas EpsE e EpsL é formado por 90 resíduos de aminoácidos da região N-terminal da EpsE e a região da porção citoplasmática da EpsL designada como os domínios II e III na estrutura de raio-X da cyto-EpsL (Sandkvist *et al.*, 2000; Abendroth *et al.*, 2004).

A estrutura em raio-X dos subdomínios N0 (região helical) e N1 (domínio central) da proteína XpsE da *Xanthomonas campestris* foi determinada em duas formas de cristal (Chen *et al.*, 2005). Em ambas as formas o subdomínio N1 adota o mesmo enovelamento $\alpha + \beta$ como visto para o N1 de EpsE em complexo com o domínio citoplasmático de EpsL. Este intrigante achado pode sugerir que o subdomínio N1 regula similarmente a atividade de ATPases da família E da via de secreção do tipo II. Para corroborar, a remoção dos 90 resíduos de aminoácidos da região N-terminal de EpsE resulta em um aumento de atividade da ATPase.

Nos dois cristais dos subdomínios N0-N1 de XpsE, 62 resíduos da região helical do subdomínio N0 adotam duas diferentes conformações, mas apesar dessa variabilidade

conformacional 36 resíduos da região N-terminal parecem indispensáveis para a funcionabilidade da via de secreção do tipo II em *Xanthomonas campestris*. Chen *et al.* formularam uma hipótese de que esses resíduos participam da ligação a XpsL e que as duas conformações estruturais podem representar estados de ligação e não ligação à região N-terminal de XpsE.

Os três componentes da via de secreção do tipo II T2S:M, T2S:L e T2S:E formam um complexo estável na membrana citoplasmática. T2S:M é uma pequena proteína da membrana interna com curto segmento citoplasmático, uma hélice transmembrana e um domínio na região C-terminal periplasmático. A homologia de seqüências entre membros da família M da via de secreção do tipo II concentra-se em prever a hélice transmembrana e dois blocos de resíduos de aminoácidos conservados presentes na região periplasmática da cadeia polipeptídica (Abendroth *et al.*, 2004b). A curta porção N-terminal da proteína mostra pouca homologia com outros membros na família. Estrutura de raio-X da maior porção do domínio periplasmático da proteína EpsM do *Vibrio cholerae* foi resolvida recentemente (Abendroth, *et al.*, 2004b). A proteína EpsM e o domínio periplasmático solúvel são capazes de formar dímeros em solução, e dois monômeros da região periplasmática da proteína EpsM estão presentes em cada unidade assimétrica do cristal. No modelo de dímero, o contato subunidade-subunidade ocorre na maioria através de interações entre os resíduos 123 e 135 da região C-terminal de cada monômero da proteína EpsM. Resíduos que são conservados apenas entre as homólogas mais próximas.

2.2.3. Biogênese do Pili do Complexo do tipo IV

Análises de sequência dos componentes da via de secreção do tipo II tem revelado homologia com vários componentes envolvidos na biogênese do Pili do complexo do tipo IV, sugerindo que a via de secreção do tipo II pode funcionar similarmente a maquinaria do complexo do tipo IV (Nunn, 1999; Burrows, 2005). As proteínas da família G, H, I, J e K demonstram algumas similaridades com as subunidades do complexo do tipo IV da formação do Pili, e como no caso das subunidades do Pili, são traduzidas com um peptídeo sinal especializado para exportação através da membrana citoplasmática. Além disso proteínas da família O possuem um alto grau de identidade com a prepili peptidase (EilD). PilD é uma protease bifuncional que cliva o peptídeo sinal da subunidade do complexo do tipo IV e metila a nova porção da região N-terminal (Nunn & Lory, 1993; LaPointe & Taylor, 2000).

Proteínas da família O realizam a mesma função em relação às proteínas da família G, H, I, J e K da via de secreção do tipo II. A notável similaridade com componentes do complexo do tipo IV formador de Pili tem levado as proteínas da família G, H, I, J e K da via de secreção do tipo II a receberem o termo *pseudopilis*. Mostrou-se recentemente que certas pseudopilis da via de secreção do tipo II são realmente capazes de formar pili, fibra que compreende o envelope celular sendo exposto na superfície da célula sob superexpressão (Sauvonnnet *et al.*, 2000; Durand *et al.*, 2003; Kohler *et al.*, 2004).

2.2.4. Complexo da Membrana Externa

Proteínas da família C interagem com proteínas da família L e M presentes na membrana interna e proteínas da família D na membrana externa. Especificamente a região de domínio periplasmático próxima da região em hélice da membrana tem demonstrado contribuir para interações com o subcomplexo das proteínas L e M (Gerard-Vincent *et al.*, 2002; Lee *et al.*, 2004), e a região C-terminal tem demonstrado em muitas espécies ser importante para a interação com as proteínas da família D (Bleves *et al.*, 1999; Possot *et al.*, 1999).

Embora os sistemas da via de secreção do tipo II sejam geralmente bem conservados, os componentes do sistema têm sido encontrados ser altamente específico a espécies. Proteínas da família C e D, em particular, não podem ser substituídas por homólogas mesmo das espécies mais próximas (de Groot *et al.*, 1991; Lindeberg *et al.*, 1996). Isso sugere que essas duas proteínas podem ser capazes de determinar a especificidade de substratos transportados pela via de secreção do tipo II.

2.3. Proteína Dimetiladenosina transferase de *Xylella fastidiosa*

A *orf xf2148*, de 873 bp, apresenta 47% de similaridade ao gene *ksgA* de *E. coli* (<http://aeg.lbi.ic.unicamp.br/xf>), o qual codifica uma proteína de peso molecular equivalente a 32 kDa e 290 resíduos de aminoácidos, cuja purificação e cristalização foi descrita por O'Farrell *et al.* (2003). Em *E. coli*, o gene *ksgA* codifica uma proteína que proporciona a bactéria uma resistência a kasugamicina, um antibiótico que pertence a família dos aminoglicosídeos (Lafontaine *et al.*, 1998). A proteína codificada pelo gene

ksgA é uma dimetiladenosina transferase de bactéria que converte duas adenosinas adjacentes em N⁶, N⁶ dimetiladenosina, usando S-adenosilmetionina (SAM) como doador de metil (O'Farrell *et al.*, 2003). S-adenosilmetionina é uma fonte de diversos grupos químicos usados na biossíntese e modificação de biomoléculas. A mais notável reação envolvendo S-adenosilmetionina, a transferência de grupos metil, é realizada por uma grande classe de enzimas, transferases dependentes de S-adenosilmetionina, a qual tem sido o foco de estudos estruturais e funcionais (Kosbial & Mushegian, 2005).

Um papel biológico relativamente bem entendido é a doação de grupos metil através de modificações covalentes de diferentes substratos, do mais simples como o arsênico oxidado, aos mais complexos como RNA ribossômico, RNA transportador e proteínas essenciais, cuja metilação pode servir como um sinal regulatório para a maturação e interações de controle com outras macromoléculas (Anantharaman *et al.*, 2002; Hopper & Phizicky, 2003). A enzima adenosil dimetiltransferase pode ligar-se ao 16S rRNA, assim como ao fragmento de RNA que contém os alvos de adenosina. A proteína permanece enzimaticamente inativa até a união parcial de uma pequena subunidade (Desai & Rife, 2006). A razão disso ainda não está clara, mas uma segunda interação, além do sítio de modificação, entre a adenosina dimetiltransferase e seu pré substrato 30S pode ser necessária para a ativação da proteína (O'Farrell *et al.*, 2004). A modificação das duas adenosinas adjacentes no rRNA é universalmente conservada (Van Knippenberg *et al.*, 1984 & O'Farrell *et al.*, 2006). Apesar deste alto nível de conservação, a enzima não é absolutamente essencial em bactérias (Lévêque *et al.*, 1990).

A perda de atividade é o modo mais comum de surgimento da resistência ao antibiótico kasugamicina em *E. coli* e em outras bactérias (O'Farrell *et al.*, 2004). A resistência ao antibiótico está associada à falta de dimetilação de resíduos de adenosinas em

células mutantes faltando a versão funcional da enzima. Além disso o preço pago pela perda de sua função é a menor taxa de crescimento, fidelidade traducional reduzida em *E. coli* e virulência reduzida no patógeno humano *Yersinia pseudotuberculosis* (Mecsas *et al.*, 2001).

2.3.1. Metil Transferases (MTases)

DNA metil transferases são enzimas encontradas em todos os grupos de organismos procarióticos e eucarióticos. Essas enzimas reconhecem uma sequência específica no DNA e transferem um grupo metil do substrato S-adenosyl-L-methionine (AboMet), um doador de grupos metil, para modificar suas sequências alvo.

Em bactéria, as proteínas do tipo DNA metil transferases estão geralmente associadas a sistemas de restrição, que geram uma especificidade e protegem a integridade do genoma da célula hospedeira contra a invasão de DNA estrangeiro (Noyer-Weidner & Trautner, 1993). A metilação da citosina ou adenina em ambas as fitas de um DNA específico, protege o DNA da clivagem por endonucleases de restrição que reconhecem a mesma sequência (Nelson *et al.*, 1993). Enzimas de restrição são classificadas em quatro tipos de acordo com sua estrutura molecular, modo de ação, estrutura do sítio de ligação e dependência de cofator.

2.3.2. Cofator S-adenosyl-L-methionine (SAM)

S-adenosyl-L-methionine (SAM) é uma fonte de diversos grupos químicos usados na biossíntese e modificação de quase todas as biomoléculas. A mais notável reação

envolvendo S-adenosyl-L-methionine (SAM), a transferência de grupos metil, é realizada por uma ampla classe de enzimas denominadas metil transferases dependentes de adenosyl-methionine. SAM é o intermediário metabólico essencial em cada forma de vida celular e em cada organismo há várias enzimas que utilizam SAM. Um papel biológico da S-adenosyl-L-methionine é a doação de grupos metil por modificações covalentes de diferentes substratos como RNA ribossômico, RNA transportador, e proteínas essenciais cuja metilação pode servir como um sinal regulatório para maturação e controle de interações com outras macromoléculas.

A maioria das metil transferases dependentes de SAM pertencem a classes de enzimas com enovelamento do tipo Rossmann, um dos mais comuns arranjos de estrutura espacial de proteínas observada em diversas famílias de enzima (Lesk, 1995).

Em seu arranjo básico o enovelamento do tipo Rossmann consiste na alternância em regiões de folha β e α hélice com uma região central formada por folhas β , e hélices preenchendo duas camadas, uma de cada lado do plano (Fig. 2).

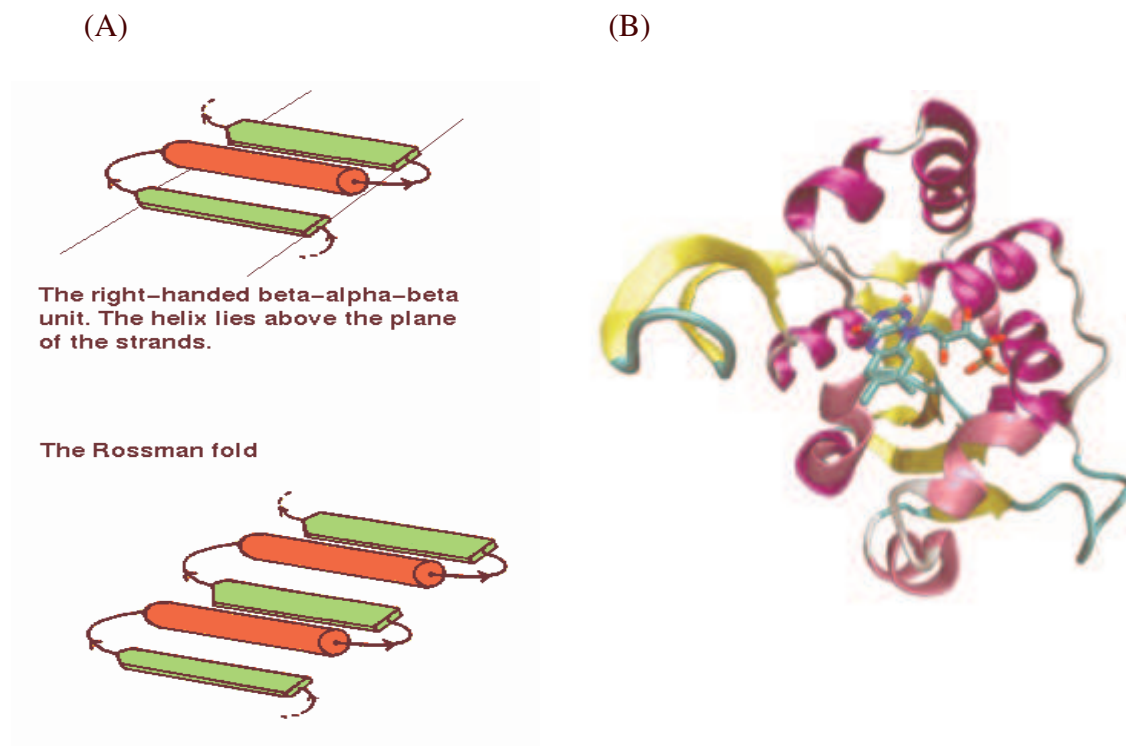


Figura 2. (A) Enovelamento do tipo Rossmann. A subunidade β - α - β - α - β presente em proteínas que se ligam a nucleotídeos é denominada Enovelamento do tipo Rossmann. (B) Exemplo de Enovelamento do tipo Rossmann. Domínio estrutural da proteína decarboxylase da bactéria *Staphylococcus epidermidis* (PDB 1G5Q). Rao & Rossmann, 1973.

A comparação de seqüências de metil transferases com o enovelamento do tipo Rossmann tem identificado um conjunto de cinco regiões de ligação SAM altamente conservadas. Essas regiões correspondem aos motivos I-V dos motivos inicialmente propostos para DNA:m5C MTases por Posfai *et al.*, 1989.

A primeira seqüência conservada (Motivo I) inclui em sua região N-terminal o consenso GxGxG, considerado o sítio de ligação à SAM de metil transferases dependentes de SAM. Nenhum dos três resíduos de glicina é universalmente conservado, mas as substituições são tipicamente por resíduos de aminoácidos com cadeias laterais pequenas,

ou com a cadeia principal propensa a dobrar. Isso corrobora com os dados de análise de sequência indicando que o consenso está localizado no “loop” conectando a primeira folha β e α hélice no centro do enovelamento do tipo Rossmann. No meio da folha β número 1 existe um resíduo ácido (D ou E) conservado. Um ou mais resíduos conservados carregados positivamente são encontrados próximos a região N-terminal. O motivo 2 compreende a folha β número 2.

Um resíduo ácido parcialmente conservado é comum na região C-terminal dessa folha. O Motivo 3 corresponde a folha β número 3. Um resíduo ácido parcialmente conservado também está localizado próximo a região C-terminal dessa folha.

O Motivo 4 consiste da folha β número 4 e os “loops” flanqueadores. Nesse motivo existe um resíduo bem preservado D/E/N, localizado na extremidade da região N-terminal dessa folha.

O Motivo 5 corresponde a hélice seguida da folha β presente no Motivo 4. Em algumas metil transferases esse domínio serve como sustentação da cadeia lateral aromática ou hidrofóbica que estabiliza a base adenina do AdoMet, mas em muitos casos têm sido demonstrado que esses resíduos não são essenciais para a atividade das metil transferases (Sankpal & Rao, 2002).

Finalmente, vários resíduos dos Motivos 1 ao 5 são conhecidos por fazer contato direto com a SAM. Em particular, um ou mais resíduos no “loop” GxGxG estão em contato com uma região SAM, enquanto que resíduos ácidos conservados do Motivo 2 formam pontes de hidrogênio com hidroxil ribose. Resíduos variáveis na região C-terminal do Motivo 2 interagem com uma base, enquanto que resíduos variáveis da região C-terminal do Motivo 4 parece ter contato com grupos amino como a metionina da molécula SAM (Fig. 3).

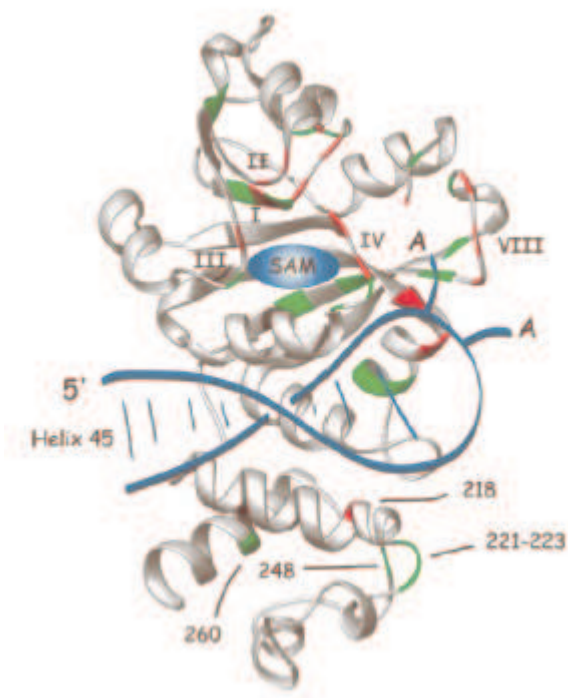


Figura 3. Modelo da enzima dimetiladenosina transferase (KsgA) de *E. coli* mostrando a região de ligação com o substrato SAM. O'Farrel *et al.*, 2004.

2.3.3. Adenina dimetil transferase de *Escherichia coli*

A enzima KsgA bacteriana de *E. coli* cataliza a transferência de um total de quatro grupos metil da SAM para duas adenosinas no RNA ribossômico 16S. Essa enzima e adenosina modificada resultante dessa catálise, parecem ser conservadas em todas as espécies de eubactéria, eucariotos, archaeobacteria e em organelas eucarióticas.

A biogênese ribossomal envolve uma via complicada, compreendendo muitos passos de processamento e modificação. Um exemplo dessa modificação é a metilação. Todos os organismos têm modificações no nucleotídeo e seus RNA ribossômicos, mas apenas duas modificações têm sido conservadas através da evolução. Uma modificação é a formação da pseudouridina no domínio 4 da maior subunidade do RNA ribossômico. Essa modificação é

encontrada somente em ribossomos citoplasmáticos, sendo realizada por diferentes sistemas em diferentes organismos (Ofengant, 2002). Outra modificação conservada evolutivamente é a dimetilação de duas adenosinas adjacentes no “loop” da hélice 45, próximo a região 3’ da pequena subunidade rRNA, para produzir N⁶, N⁶-dimetiladenosina. As únicas exceções conhecidas acontecem na subunidade 12S do rRNA mitocondrial em *Saccharomyces cerevisiae*, na qual não há grupo metil nas duas adenosinas (Klootwijk *et al.*, 1975), no rRNA do cloroplasto de *Euglena gracilis* na qual a adenosina próxima a região 5’ é dimetilada (Steege *et al.*, 1982; van Bull *et al.*, 1984), e a subunidade 16S de *Sulfolobus solfataricus* na qual contém apenas uma adenosina metilada (Noon *et al.*, 1998).

A hélice 45 do “loop” é uma das áreas mais conservadas da pequena subunidade do rRNA (van Knippenberg *et al.*, 1984; Cannone *et al.*, 2002).

A enzima KsgA ainda não foi caracterizada em archaeobacteria, mas análises de sequência revelam KsgAs potencialmente homólogas nesses organismos. Apesar do alto nível de conservação a enzima não é absolutamente essencial em bactérias. A perda de atividade da KsgA é o modo mais comum de aparecimento da resistência ao antibiótico kasugamicina em *E. coli* e outras bactérias (O’Farrell *et al.*, 2004). Ribossomos mutantes com ausência de dimetilações requerem quantidades maiores de certos fatores de tradução *in vitro* e são mais propensos a mutações. As bactérias que possuem esta enzima mutante parecem não ser afetadas, pelo menos sob condições de crescimento em laboratório (van Knippenberg *et al.*, 1986). Interessante notar que mutações no gene que codifica a proteína KsgA de *Yersinia pestis* resultaram em um patógeno menos virulento (Mecsas *et al.*, 2001).

Embora o rRNA 16S contenha adenosinas modificadas, o rRNA 16S não é substrato para a enzima KsgA (Helser *et al.*, 1972; Thammana & Held, 1974). A KsgA liga-se a um pequeno fragmento do rRNA contendo as bases de adenina alvo, mas não metila a

subunidade 16S livre. Subunidades 30S parcialmente formadas consistindo de rRNA 16S mais um subconjunto de proteínas ribossomais parecem ser o substrato mínimo para a enzima, e a subunidade 30S inteiramente formada serve como um eficiente substrato (Poldermans *et al.*, 1979).

2.4. Análise Espectroscópica por Dicroísmo Circular

O método espectroscópico de dicroísmo circular (CD) é bastante utilizado nas análises conformacionais de proteínas, onde o espectro obtido é determinado pela quantidade e composição de estrutura secundária (Fasman, 1996). Moléculas assimétricas em solução são opticamente ativas, isto é, quando elas interagem com uma luz circularmente polarizada elas absorvem luz polarizada à direita e à esquerda de maneira desigual. Esta diferença na absorção gera luz elipticamente polarizada que pode ser medida por um espectropolarímetro de CD, que converte este sinal em elipticidade (q) expresso em miligraus. O sinal de CD será positivo ou negativo dependendo da absorção preferencial da luz polarizada à direita ou à esquerda. Os cromóforos mais numerosos e característicos que proteínas e peptídeos possuem para a luz polarizada circularmente na faixa do UV-distante (170 a 260 nm), são as ligações peptídicas (Kelly e Price, 1997). Os espectros de CD produzidos por esses cromóforos são característicos para os diversos tipos de estrutura secundária. As faixas de absorção dos resíduos aromáticos no UV-distante são pequenas quando comparadas às das ligações peptídicas. Contudo, no UV-próximo (260 a 320 nm), o espectro de CD passa a ser dominado pela contribuição das cadeias laterais aromáticas, juntamente com as pontes dissulfeto e grupamentos prostéticos. Assim, enquanto o UV-distante detecta estrutura secundária, o UV-próximo é sensível à estrutura terciária, o que

compõe a base para a análise da conformação de proteínas através de dicroísmo circular (Kelly e Price 1997).

Experimentos de dicroísmo circular têm sido muito utilizados para a investigação de: estrutura secundária de peptídeos e proteínas, mudanças conformacionais em proteínas, desenovelamento de proteínas, desenho de peptídeos e proteínas, proteínas de membrana, interação entre domínios de proteínas e ligação de substratos (Kelly e Price, 2000). Por esta técnica ser sensível à quantidade total de estrutura secundária de uma proteína, ela é usada para investigar a quantidade de α -hélice, folha- β e estrutura randômica que compõe a proteína (Fasman, 1996). As α -hélices e folhas- β são estruturas bem compactas presentes nas proteínas enoveladas e a estrutura randômica está presente nas proteínas desenoveladas e em alças que conectam os elementos de estrutura secundária.

Proteínas que possuem sua estrutura secundária formada predominantemente por α -hélices exibem picos negativos a 208 e 222 nm e um pico positivo entre 190 e 195 nm. As proteínas formadas predominantemente por folhas- β apresentam espectro de CD contendo um pico negativo em aproximadamente 216 nm e um positivo a 196 nm. A estrutura randômica é representada pelo espectro contendo um pico negativo próximo de 195 nm e um pico positivo a 212 nm (Pelton e McLean, 2000). A Figura 3 traz a representação dos espectros de CD característicos de cada tipo de estrutura secundária.

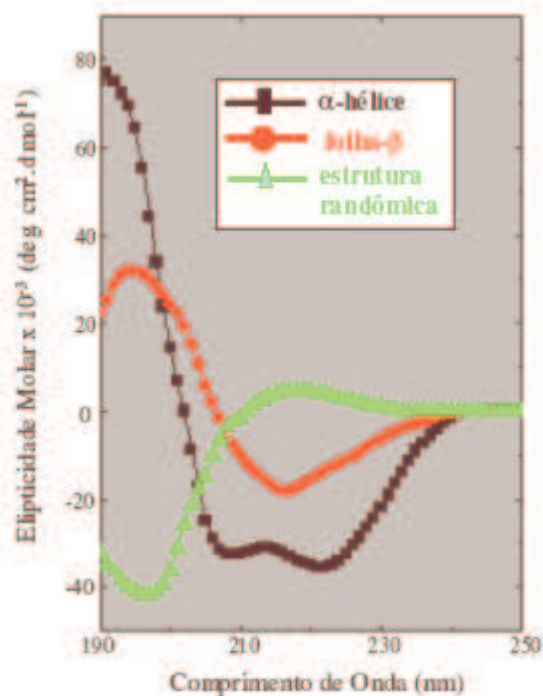


Figura 3. Representação dos espectros característicos de cada tipo de estrutura secundária obtidos por CD (http://www.ap-lab.com/circular_dichroism.htm).

2.5. Espectrometria de Massas

O espectrômetro de massas é um instrumento que permite determinar a massa molar de compostos eletricamente carregados, ou íons previamente formados (Cole, 1997). Estes íons são selecionados de acordo com a razão massa-carga (m/z), sendo m a massa em u (massa atômica unificada), definida como 1/12 da massa de um átomo do isótopo de ^{12}C , o qual foi designado como 12u por convenção (Gross, 2004).

As técnicas avançadas em espectrometria de massas diferem principalmente no modo de ionização das amostras. Essa característica permitiu o uso da técnica em muitas áreas, pois a análise de proteínas, peptídeos, açúcares, etc, que antes não eram possíveis de serem

detectados por técnicas antigas de ionização, como EI (electron ionization), agora são analisados rotineiramente.

2.5.1. MALDI-TOF (Matrix Assisted Laser Desorption/Ionization – Time of Flight)

A técnica de MALDI, introduzida por Hillenkamp *et al.* em 1985, a amostra contendo a espécie de interesse é misturada com uma matriz (geralmente um ácido orgânico aromático) formando uma “mistura sólida”. Um pulso de laser, com comprimento de onda próximo do UV, incide sobre essa mistura e a energia do laser é absorvida pela matriz, que tem o máximo de absorção perto do comprimento de onda do laser. Desse modo, a matriz evapora e o analito que estava incluso na matriz, agora se encontra na fase gasosa altamente energética atribuída à excitação eletrônica da molécula da matriz ao absorver a energia do laser. A formação dos íons (ionização) ocorre através da transferência de carga (ex: transferência de prótons) das moléculas da matriz para o composto que fica na forma de MH^+ . Os íons formados recebem uma alta energia cinética inicial (K) que os impulsiona para o analisador de massas Time-of-flight (TOF), onde são separados de acordo com o tempo de voo, considerando a distância na qual o íon se movimenta até atingir o detector. Utilizando a equação da energia $K = mV^2/2$, onde V é a velocidade do íon, calcula-se a massa molecular do composto. Os compostos são separados no analisador de acordo com sua razão m/z . Íons mais leves chegam mais rapidamente no detector, enquanto os íons mais pesados demoram mais tempo.

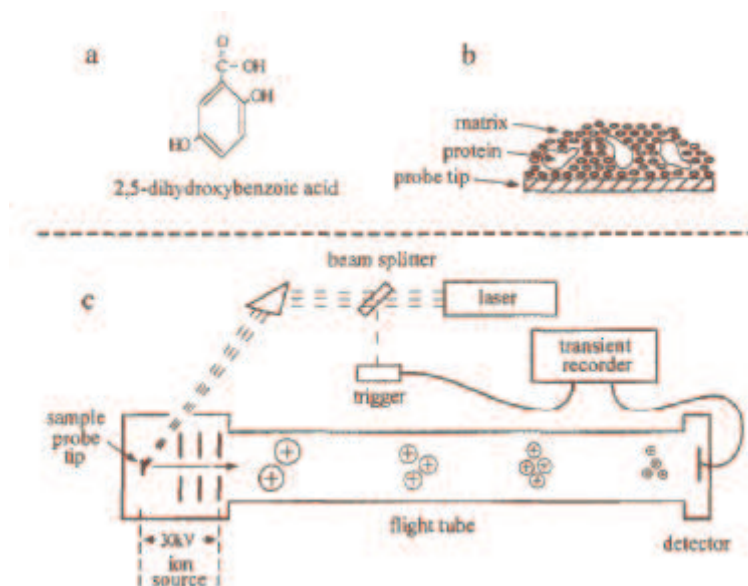


Figura 4. Componentes envolvidos em MALDI-TOF. (a) Estrutura da matriz utilizada. (b) Representação conceitual do analíto incluído na matriz. (c) Diagrama esquemático da técnica de MALDI acoplado ao analisador de massas Time of Flight (TOF). Watson *et al.*, 1995.

No MALDI, o analisador de massas pode operar em dois modos: modo linear e modo reflectron. O modo linear é utilizado para moléculas de grande massa molecular, como as proteínas, peptídeos e polímeros e o modo reflectron utilizado para moléculas de massa de até 7000 Da, como as micotoxinas. O modo reflectron possui a vantagem de apresentar uma maior resolução quando comparado ao modo linear, devido à presença de um conjunto de lentes com diferença de potencial crescente, propiciando o aumento de resolução. Os íons gerados por MALDI têm uma energia muito baixa, mas em MALDI os íons formados possuem poucas cargas e na maioria das vezes formam-se somente espécies monocarregadas. Essa característica é interessante porque facilita a interpretação dos resultados dos espectros de massas, mesmo para compostos de alta massa molecular.

Análise de proteínas por espectrometria de massas pode ser realizada usando fragmentação específica de seqüências de um peptídeo ou por *peptide mass fingerprinting*

(PMF), também conhecida como mapeamento de massas de peptídeo. O procedimento padrão para a identificação de proteínas inclui a separação de proteínas por um gel de eletroforese ou cromatografia líquida. Subseqüentemente as proteínas são clivadas com endoproteases, sendo tripsina a mais utilizada. Após a digestão, os peptídeos gerados são investigados através da determinação das massas moleculares. Para a identificação da proteína, as massas moleculares obtidas experimentalmente são comparadas com massas moleculares teóricas armazenadas em bancos de dados.

MALDI é a técnica mais comum para realizar *peptide mass fingerprinting* (PMF). MALDI é rápido, robusto, fácil de ser realizada, sensível, de grande acurácia e tolerante a certo nível de contaminantes.

3. Objetivos

3.1. Objetivos gerais

- Contribuir para o conhecimento dos mecanismos de patogenicidade de *Xylella fastidiosa*.
- Fornecer subsídios ao estudo da estrutura tridimensional de proteínas expressas por genes relacionados à patogenicidade.

3.2. Objetivos específicos

- Efetuar a expressão e purificação “in vitro” de dois genes de *X. fastidiosa*, potencialmente envolvido com a patogenicidade desta bactéria, através da utilização de vetores de expressão.
- Caracterizar as proteínas quanto às propriedades e funções através de ensaios bioquímicos pertinentes.
- Obter informações estruturais preliminares por métodos espectroscópicos como Dicroísmo Circular e Modelagem tridimensional.

**Cloning, Expression and Purification of a transmembrane protein XpsL
of secretion pathway of type II from the phytopathogen *Xylella fastidiosa***

**José Sérgio M. Soares ^a, Susely F. S. Tada ^a, Cleide F. Catani ^c, Luciana K. Rosselli ^a,
Antonio M. Saraiva ^a, Gustavo H. M. F. Souza ^b, Marcos N. Eberlin ^b and Anete P.
Souza ^{a+}**

^a *Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética e
Evolução, Instituto de Biologia, Universidade Estadual de Campinas, C.P. 6010, Campinas
– S.P., Brazil*

^b *Instituto Química, Universidade Estadual de Campinas, C.P. 6154, Campinas-S.P., Brazil*

^c *Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade
Estadual de Campinas , C.P. 6010, Campinas – S.P., Brazil*

⁺ Corresponding author. Fax: 55-19-35211090.

E-mail address: anete@unicamp.br (Anete Pereira de Souza)

Abstract

Gram-negative bacteria use the type II secretion system to transport a large number of secreted proteins from the periplasmic space into the extracellular environment. Many of the secreted proteins are major virulence factors in plants and animals. XpsL is a protein of cytoplasmatic membrane and it is part of the necessary components for the secretion of type II from *Xylella fastidiosa*. Membrane proteins are hard to overexpress, purify and crystallize. Thus, only 2% of the structures of proteins deposited in the Protein Data Bank belong to the category of membrane proteins.

The gene that codifies the XpsL protein was cloned into a pET32Xa/LIC vector to overexpress the coupled with fusion tags in *Escherichia coli* BL21(DE3). The expressed XPSL protein was purified by immobilized metal affinity chromatography (IMAC) and had its identity determined by SDS-PAGE and Mass Spectrometry (MALDI-TOF). The correct folding of the purified recombinant protein was verified by circular dichroism spectroscopy.

Keywords: Membrane protein; pET32 Xa/LIC vector; Circular Dichroism; Mass Spectrometry.

Introduction

The phytopathogen *Xylella fastidiosa* is a xylem-limited bacterium that causes a range of economically important plant diseases, including a serious disease of orange trees called citrus variegated chlorosis (CVC) [1–3]. Recently, DNA sequence data provided by the *Xylella fastidiosa* Brazilian genome project have allowed a functional genomic approach to investigate the function of several proteins based on the information about their structures. This emergent area is a powerful means to more deeply understand the mechanisms of the bacterium pathogenicity [4].

Xylella fastidiosa (Xf) is a Gram-negative, rodshaped, non-flagellate, xylem-limited bacterium, and is the causal agent of several important crop diseases, including citrus variegated chlorosis (CVC) in orange trees [1]. Plants with CVC symptoms were first detected in Brazil in 1987 [3], and today, have been identified in more than 90% of the orchards in the state of São Paulo (Brazil), leading to the complete genomic sequencing of a 9a5c strain, which turned out to be the first plant bacterium to have its complete genome elucidated [4].

In order to gather information about the proteins involved in the mechanisms of the bacterium pathogenicity, we chose the *orf xf1524* which codifies the XpsL protein for the characterization studies.

The *orf xf1524*, of 1155 bp, presents 55% of similarity with the gene *xpsL* of *Xanthomonas campestris* (<http://aeg.lbi.ic.unicamp.br/xf>), which codifies a protein of secretion pathway of type II of molecular weight equivalent the 43,4 kDa and 384 amino acid residues. XpsL is a protein of cytoplasmatic membrane and it is part of the necessary

components for the secretion of type II in *Xanthomonas pv campestris*. [5]. The secretion pathway of type II is used by bacteria gram negative to secrete extracellular proteins [6]. Being XpsL a component of the secretion mechanism of type II, gives credit that this protein creates a channel where proteins are secreted [7]. One of the great challenges for molecular biologists is to learn how a protein sequence defines its three dimensional structure. For many years, the problem was even more difficult for membrane proteins because so little was known about what they looked like.

Membrane protein structural biology is a frontier area of modern biomedical research. Membrane protein structures constitute <1% of the entries in the Protein Data Bank (PDB). Disturbingly, only two of the current entries represent a membrane protein from human origin [8,9], whereas the majority of entries are of bacterial membrane proteins. Part of the reason why progress has been faster for bacterial membrane proteins stems from the fact that they can more easily be expressed in large quantities in bacterial hosts, and that they lack many of the posttranslational modifications that potentially complicate crystallization. Moreover, the fast pace at which bacterial genomes are sequenced provides an almost unlimited repertoire of target proteins including homologs from thermophilic bacteria that are often more stable during detergent solubilization, purification and crystallization [10].

Here, we report the cloning and overexpression the recombinant XpsL protein in *E. coli*. The purified recombinant XpsL protein exhibited a predominant signal of α -helices during the Dichroism Circular experiments suggesting that XpsL recombinant protein maintains secondary structures and is viable to functional and structural studies and remained folded. It also aims at adding new information on proteins that may be related to the *X. fastidiosa* pathogenesis, necessary for new approaches towards the combat of CVC.

Besides that many integral proteins are important in biomedical and biotechnological applications; the majority of drug targets are integral membrane proteins.

Material and Methods

Cloning *orf xf1524*

The *xpsL* gene of *Xylella fastidiosa* (ORF XF1524) was amplified by PCR using the purified genomic DNA as a template. The PCR was carried out using the following specific primers: forward primer 5'-GGTATTGAGGGTCGCTTGGGAAGAGGG-3' and reverse primer 5'-AGAGGAGAGTTAGAGCCTTAACGTTGTGCC -3'. The primers were designed to generate products with vector cohesive overhangs, for efficient cloning onto a pET32Xa/LIC vector. This vector is designed for the heterologous protein expression which is fused to thioredoxin (109 amino acids), His-tag (6 amino acids), and S-tag (15 amino acids) sequences upstream to the cloning site. The fusion tags together have 17.6kDa and can be removed by factor Xa cleavage [11]. The recombinant plasmids were used to transform competent *E. coli* DH5α cells, using the standard polyethylene glycol (PEG) method [12]. Cells were cultured overnight at 37°C in Luria-Bertani broth (LB) plates containing 100µg/mL ampicillin. The colonies were individually stored at -70°C in a permanent 2YT + HMFM broth [13] and checked by PCR.

Expression and purification of XpsL-His.tag fusion protein

The recombinant plasmid with *xpsL* insert was transformed into *E. coli* BL21 (DE3) cells by PEG method. The cells were grown in 2 ml of LB medium containing ampicillin (100µg/mL) at 37°C, 300rpm, for 18h. An aliquot of 1mL of the culture was used to inoculate 100ml of the same medium, maintained at the same culture conditions until OD600 reached 0.8. Expression of recombinant protein was induced by the addition of Lactose to a final concentration of 5,6mM and cultivated for 3h. The culture was harvested using 10min centrifugation at 6000g, 4°C, and stored for further use at -20°C. The bacterial pellet obtained from 100mL liquid culture was suspended in 22mL buffer A (50 mM Tris-HCl, pH 7.5; 300mM NaCl). The PMSF (1mM), protease inhibitor was added, as well lysozyme (1mg/mL). The cells were lysed by sonication on ice, in an ultrasonic cell disruptor (Cole-Parmer Instrument, Chicago, IL, USA). The obtained lysate was centrifuged at 27,000g for 20min at 4°C. The supernatant containing soluble cellular material (total volume 22mL) was used for purification procedure using immobilized metal affinity chromatography (IMAC). The column containing 1mL of Ni-NTA affinity resin was equilibrated with 10 column volumes (CV) from buffer A. The supernatant was loaded into the column by gravity. The column was washed with 10 CV of buffer B (50mM Tris-HCl, pH 7.5; 1.0M NaCl). Additional washing using 5 CV of buffer C (50mM Tris-HCl, pH 7.5; 300mM NaCl; and 5mM imidazole) was performed. The bounded XpsL-His.tag fusion protein was eluted with 4 CV of buffer D (50mM Tris-HCl; 300mM NaCl; pH 7.5 and 10mMimidazole). Fractions were assayed for total protein concentration and analyzed by SDS-PAGE.

Circular Dichroism spectroscopy

Circular Dichroism (CD) spectra of the purified protein in solution were measured with a Jasco J-810 Spectropolarimeter dichrograph (Japan Spectroscopic, Tokyo, Japan). The CD spectra were generated using a quartz cuvette with a cell path with a length of 1.0mm, containing 200 μ L XpsL-His.tag protein (0.1mg/mL) in 5mM Tris-HCl buffer, pH 7.5. Four scans in the 260–200nm range at a rate of 20nm/min at 20°C were recorded for sample and average measurement.

In-gel trypsin digestion

XpsL-His.tag protein isolated by SDS-PAGE was excised from the gel with a clean razor and prepared for complete trypsinolysis. The excised gel piece was diced into smaller pieces (1mm²) processed in one tube by destaining until clear 2x10min in potassium ferricyanide: sodium tiosulfate (1:1), and then washed 3x10min in water. Protein disulfide was reduced by incubation at 60 °C for 30 min with 10 μ l of 45 mM dithiothreitol in 100mM ammonium bicarbonate. The residual liquid was removed and then the proteins were alkylated by adding 100 μ l of 100mM iodoacetamide. Residual liquid was removed and gel pieces were washed with 500 μ l of 50mM ammonium bicarbonate in 50% acetonitrile. Gel pieces were dried under vacuum, and then digestion using trypsin was performed in the same tubes by adding 25mM ammonium bicarbonate, proportion of 0,2 μ g of trypsin/1 μ g sample (according to protein concentration). After digestion 18 h at 37°C, soluble trypsin fragments were recovered in the supernatant and combined with three

extractions of the gel pieces with 25µl of 60% acetonitrile 0.1% TFA. The pooled extracted fragments were dried under speed vacuum at 37 °C.

Purification of peptides by ZipTip_{µ-C18}

The sample was dissolved in 10µl of 2% TFA. ZipTip was washed three times with 15µl (0,1% TFA/50%acetonitrile), and then equilibrated four times with 15µl of 0,1% TFA. The sample was loaded onto ZipTip (three times up and down) and peptides bound onto ZipTip were washed three time with 0,1% TFA. The sample was eluted into 20 µl of 100% acetonitrile solution (three times up and down).

Peptide Fingerprint by Mass Spectrometry

Analysis of all samples obtained by tryptic digestion was prepared using the dried droplet method. This was carried out as follows. The samples were acidified by addition of 0.5 1 water/TFA 100:0,1 (V/V) and were left at room temperature for a few minutes, so that the droplet volume was reduced by evaporation. The matrix α -Cyano-4-hydroxycinnamic acid, was prepared followed by 1% (m/V) on H₂O/ACN 1:1 (V/V) (1uL) containing TFA 0,1%, was added and the sample was allowed to dry at room temperature. All measurements were performed on a MALDI-TOF (Waters - Micromass, UK) mass spectrometer in previously cleaned microplate. Analyses were performed of exclusively positively charged ions, in reflectron mode. For subsequent data evaluation with peptide identification, a software package Biolynx (Micromass- UK) was utilized.

Results and Discussion

Cloning and Expression of XpsL-His.tag protein

It is suggest that extracellular secretion in gram negative bacteria requires complex transport systems to move secreted proteins from the cytoplasm into the extracellular environment [14,15]. Several highly specialized pathways have evolved for this purpose, including the type II secretion system [14]. Due to the 55% of similarity with XpsL protein from *Xanthomonas campestris*, the XpsL protein from *Xylella fastidiosa* might be involved in the protein complex, interacting with other proteins like XpsE in the phytopathogen. Studies in *Xanthomas campestris* have shown the importance of the interaction between the XpsL-XpsE proteins [15].

The *xpsL* gene was amplified by PCR from the *Xylella fastidiosa* genomic DNA and cloned into the pET32Xa/LIC vector (Novagen). Once in the expression vector pET32Xa/LIC, the complete *xpsL* gene in frame with N-end histidine tag was transfered to the *E. coli* DH5 α strain and confirmed through DNA sequencing. After transformation of the *E. coli* BL21 (DE3) strain with pET32Xa/LIC-XpsL-His.tag, several recombinant clones were selected to test the expression levels of the recombinant protein. The expressed protein was detected by SDS–PAGE showing a low level of expression of XpsL-His.tag following the addition of inductor lactose. Recombinant XspL-His.tag protein was expressed partially soluble, yielding 100 μ g/mL of protein. Changes in expression condition (i.e., lowering the temperature to 28°C and lyse buffers containing detergents), even other

E. coli strains as BL21(DE3)pLysS, BL21(DE3)ptGroE and C43(DE3) did not result in the expression of a higher amount of soluble XpsL-His.tag protein.

Affinity purification of XpsL-His.tag protein

The XpsL-His.tag protein was recovered and purified from the lysed extract using immobilized metal affinity chromatography (Fig.1, lane 6). Carefully the protein extract was added into the Ni-NTA affinity column to prevent bubble formation. The reduction of non-specific interactions between extract proteins and the matrix was done washing the resin with high salt concentration and low imidazole concentration buffers. The elution of the recombinant protein was performed using a gradient of imidazole concentration, a procedure that considerably improved the final protein purity. In a single step of purification by affinity chromatography, amounts of almost pure XpsL-His.tag protein were obtained, as revealed through the SDS-PAGE analysis, elution of the nickel bounded XpsL-His.tag was achieved with 10mM imidazole in a 50mM Tris-HCl buffer, pH 7,5.

After the purification we tried to cleave the fusion protein using several proteases (factor Xa, Thrombin, Trypsin, Quimiotrypsin, Subtilisin), but none of these proteases worked. Since the membrane interior is nonpolar, we would expect the surface residues of transmembrane proteins to be predominantly composed of nonpolar side chains, the membrane interior is essentially void of water. Probably the proteases did not achieve the cleavage site.

Circular Dichroism spectroscopy analysis

Circular dichroism is an important technique to estimate secondary structure content, to validate protein quality, and for monitoring conformational changes of proteins. Circular dichroism signals only arise where absorption of radiation occurs, and thus spectral bands are easily assigned to distinct structural features of a molecule. An advantage of the Circular dichroism technique in studies of proteins is that complementary structural information can be obtained from a number of spectral regions [16]. To investigate the structural integrity of the purified XpsL-His.tag, the protein was analyzed by CD spectroscopy. The XpsL-His.tag CD spectrum showed that it contains predominant signal of α -helices (Fig. 3). Values of the CD signals at 208nm and 222nm estimates the α -helices content of peptides and proteins. The results suggest that XpsL-His.tag recombinant protein maintains secondary structures and is viable to functional and structural studies and remained folded, besides it is in agreement with secondary structure prediction from the PSIPRED program (<http://bioinf.cs.ucl.ac.uk/psipred>).

Mass Spectrometry Analysis

Protein identification by mass spectrometry (MS) can be performed using sequence-specific peptide fragmentation or peptide mass fingerprinting (PMF), also known as peptide mass mapping [17]. Matrix-assisted laser/desorption ionization mass spectrometry (MALDI-MS) is the most commonly used technique to perform PMF [18,19]. MALDI-MS is fast, robust, easy to perform, sensitive (low fmol range), accurate (low ppm range), tolerant to a certain level of various contaminants, and can be automated.

The identity of the resulting protein product was confirmed by trypsin digestion followed by MALDI-TOF analysis. The peptide profile found for the mass spectroscopy analysis of the digested XPSL-His.tag (“fingerprint”) was compared to a theoretical digestion of the protein and the result indicated a monoisotopic mass of 60,495 Da.

Conclusion

While the number of determined structures of soluble proteins is increasing dramatically, the structures of only about 40 membrane proteins have been solved so far at high resolution by means of either X-ray or electron crystallography.

In this work, the target protein was cloned, expressed in the soluble form and the expressed protein has established its methods of purification. The protein was purified in just one step of purification by immobilized metal affinity chromatography (agarose-Ni-NTA resin). Subsequently, its purity and identity were verified by SDS-PAGE and mass spectrometry. Its correct folding was verified by circular dichroism spectroscopy analysis that indicated a secondary structure composed mainly of α -helices, despite some short β -strands also seeming to be present.

Acknowledgments

The authors thank FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo, CNPq (Centro Nacional de Pesquisa e Desenvolvimento) and CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior) for their financial

support. J.S.M.S. received a fellowship from CNPq. L.K.R. and A.M.S received a PhD. fellowship from FAPESP, A.R.A., S.F.S., and C.F.C. received postdoc fellowships from FAPESP. We also thank the Structural Molecular Biology Center researchers from the National Synchrotron Light Laboratory for technical expertise, particularly Ph.D. Nilson Ivo Tonin Zanchin and Ph.D. Carlos Henrique Ramos (Biology Cellular and Molecular Laboratory) for equipment supports and collaboration.

List of Figures

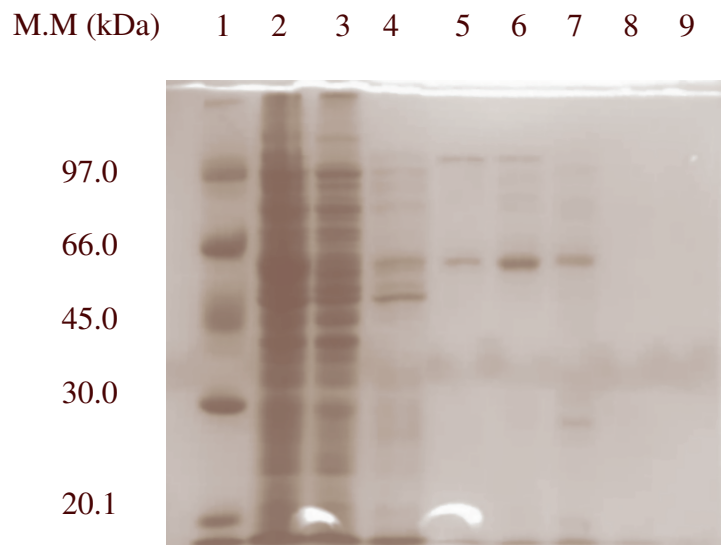


Figure 1. SDS-PAGE of the fractions collected during the purification steps of the recombinant XPSL. Legend: lane 1—molecular mass markers; lane 2—extract from bacterial lysis loaded into the Ni-NTA column; lane 3—flowthrough extract; lane 4 - column washed with lysis buffer plus 1M NaCl lane 5—fraction collected during the elution step at imidazole concentration of 5mM, containing the recombinant protein (XPSL+fusion tags), lane 6 - fraction collected during the elution step at imidazole concentration of 10mM, lane 7 - fraction collected during the elution step at imidazole concentration of 50mM, lane 8 - fraction collected during the elution step at imidazole concentration of 100mM, lane 9 - fraction collected during the elution step at imidazole concentration of 200mM. Separation gel of 12% acrylamide concentration.

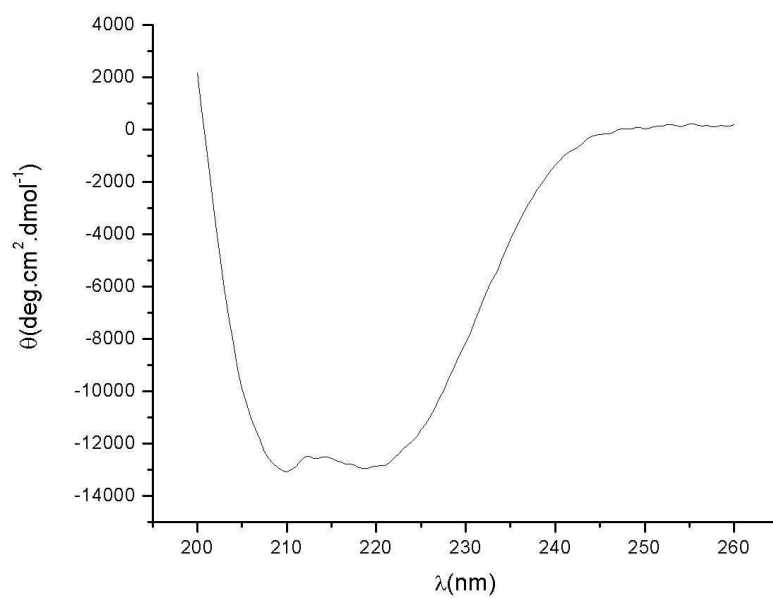


Figure 2. Circular dichroism spectrum of the purified XPSL-His.tag protein. The spectrum was generated using XPSL-His.tag protein concentration of 0.1mg/mL in a Tris-HCl buffer (5mM, pH 7.5). Four scans in a 260–200nm range at a rate of 20nm/min at 20⁰C.

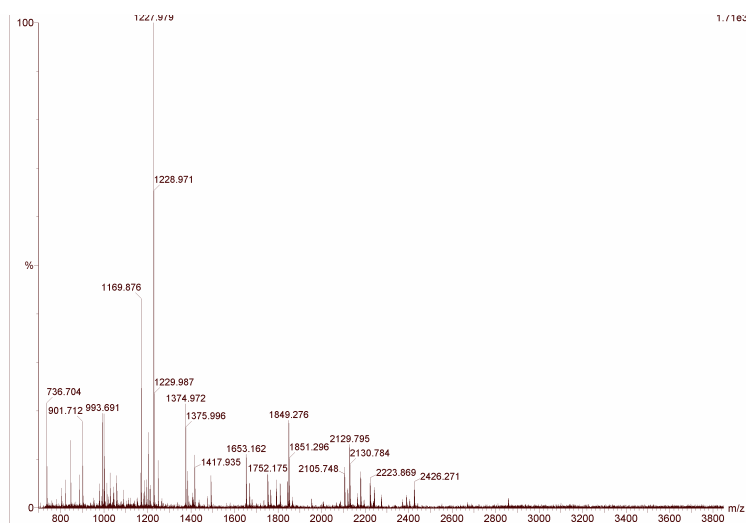


Figure 3. Ionized peptides spectrum from tryptic digestion of XPSL-His.tag protein (61kDa).

References

- [1] M. Frome, A. Camargo, S. Heber, C. Czink, A.J.D. Simpson, J.D. Hoheisel, A.P. Souza, Mapping analysis of the *Xylella fastidiosa* genome, *Nucleic Acids Res.* 28 (2000) 3100–3104.
- [2] R.F. Lee, K.S. Derrick, M.J.G. Beretta, C.M. Chagas, V. Rosetti, Citrus variegated chlorosis: a new destructive disease of citrus in Brazil, *Citrus Ind. Oct.* (1991) 12–15.
- [3] F.F. Laranjeira, Ten years of citrus variegated chlorosis: What do we know?, *Laranja* 18 (1997) 123–141.
- [4] A.J.G Simpson *et al.*, The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (2000) 151–157.
- [5] Tsai, R.T.; Leu, W.M.; Chen, L.Y; Hu, N.T. A reversibly dissociable ternary complex formed by XpsL, XpsM and XpsN of the *Xanthomonas campestris* pv. *campestris* type II secretion apparatus. *Biochem. J.* (2002) 367: 865-871.
- [6] Lee, H.M.; Tyan, S.W.; Leu, W.M.; Chen, L.Y.; Chen, D.C.; HU, N.T. Involvement of the XpsN protein in formation of the XpsL-XpsM complex in *Xanthomonas campestris* pv. *campestris* type II secretion apparatus. *Journal of Bacteriology.* (2001) 183(2): 528-535.

- [7] Russell, M. Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *Journal of Molecular Biology*. (1998) 279: 485-499.
- [8] Murata, K. *et al.* Structural determinants of water permeation through aquaporin-1. *Nature*. (2000) 407, 599–605.
- [9] MacKenzie, K.R. *et al.* A transmembrane helix dimer: structure and implications. *Science*. (1997) 276, 131–133.
- [10] Fleishman, S.J.; Unger, V.M.; Ben, T.N. Transmembrane protein structures without X-rays. *Trends Biochem Sci*. (2006). 31(2): 106-113.
- [11] Xa/LIC Vector Kits Manual, Novagen, ninth ed., USA, 2000, 56 pp. Available from: www.novagen.com.
- [12] B.T. Kurien, R.H. Scofield. Polyethylene glycol-mediated bacterial colony transformation. *Biotechniques*. (1995). 18(6) : 1023-1026.
- [13] A.R. Azzoni, S.F.S. Tada, L.K. Rosselli, D.P. Paula, C.F. Catani, A.A. Sabino, J.A.R.G. Barbosa, B.G. Guimarães, M.N. Eberlin, F.J. Medrano, A.P. Souza. Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*. *Protein Expr. Purif*. (2004) 33: 297-303.

- [14] Johnson, T.L.; Abendroth, J.; Hol, W.G.J.; Sandkvist, M. Type II secretion: from structure to function. *FEMS Microbiol Lett.* (2006) 255, 175-186.
- [15] Chen,Y.; Shiue, S.J.; Huang C.W.; Chang J.L.; Chien Y.L.; Hu N.T.; Chan N.L. Structure and Function of the XpsE N-Terminal Domain, an Essential Component of the *Xanthomonas campestris* Type II Secretion System. *Journal of Biological Chemistry.* (2005) 280(51), 42356-42363.
- [16] Kelly, S.M.; Jess, T.J.; Price, N.C. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta.* (2005) 1751, 119-139.
- [17] Aebersold, R.; Goodlett, D.R. Mass Spectrometry in proteomics. *Chem Rev.* (2001) 101(2): 269-295.
- [19] Gevaert, K.; Vandekerckhove, J. Protein identification methods in proteomics. *Electrophoresis.* (2000) 21(6):1145-1154.
- [19] Henzel, W.J.; Watanabe, C.; Stults, J.T. Protein identification: the origins of peptide mass fingerprinting. *J Am Soc Mass Spectrom.* (2003) 14(9): 931-942.

Expression, Purification and Homology Modeling of a Dimethyladenosine transferase from the phytopathogen *Xylella fastidiosa*

José Sérgio M. Soares ^a, Susely F. S. Tada ^a, Cleide F. Catani ^c, Luciana K. Rosselli ^a, Antonio M. Saraiva ^a, Gustavo H. M. F. Souza ^b, Marcos N. Eberlin ^b and Anete P. Souza ^{a+}

^a *Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, C.P. 6010, Campinas – S.P., Brazil*

^b *Instituto Química, Universidade Estadual de Campinas, C.P. 6154, Campinas-S.P., Brazil*

^c *Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, C.P. 6010, Campinas – S.P., Brazil*

⁺ Corresponding author. Fax: 55-19-35211090.

E-mail address: anete@unicamp.br (Anete Pereira de Souza)

Abstract

Methyl transferases (MTases) catalyse the transfer of methyl groups from AdoMet (S-adenosyl-L-methionine) methylating DNA or RNA. DNA or RNA methylation has been reported to be essential for bacterial virulence, and it has been suggested that adenine MTases could be potential targets for both vaccines and antimicrobials.

The *orf* that codifies the XF2148 recombinant protein in *Xylella fastidiosa* was cloned into a pET32Xa/LIC vector to overexpress the protein with fusion tags in *Escherichia coli* BL21(DE3). The expressed protein in the insoluble form was refolded using 8M Urea as denaturing agent. Secondary structure contents and identity were verified by circular dichroism and mass spectrometry (MALDI-TOF) respectively. Finally a sequence comparison and homology modeling using the modeling package Modeller 7v7 were performed indicating a high level of structure similarity among XF2148 protein from *Xylella fastidiosa* and its homologous protein, KsgA from *Escherichia coli*.

Keywords: Circular Dichroism; Mass Spectrometry; pET32 Xa/LIC; homology modeling.

Introduction

The phytopathogen *Xylella fastidiosa* is a xylem-limited that causes a range of economically important plant diseases, including a serious disease of orange trees called citrus variegated chlorosis (CVC) [1–3]. Recently, DNA sequence data provided by the Brazilian *Xylella fastidiosa* genome project have allowed a functional genomic approach to investigate the function of several proteins based on the information about their structures. This emergent area is a powerful means to more deeply understand the mechanisms of the bacterium pathogenicity [4].

The *orf xf2148*, of 873 bp, presents 47% of similarity with the gene *ksgA* of *Escherichia coli* (<http://aeg.lbi.ic.unicamp.br/xf>). The *ksgA* gene codifies a dimethyladenosine transferase from the class of methyl transferase proteins whose purification and crystallization was described by O'Farrell *et al.* (2003)[5].

S-adenosyl methionine is a source of diverse chemical groups used in biosynthesis and modification of virtually every class of biomolecules. The most notable reaction requiring S-adenosyl methionine, transfer of methyl group, is performed by a large class of enzymes, S-adenosyl methionine- dependent methyltransferases, which have been the focus of considerable structure-function studies [6].

One relatively well-understood biological role of SAM is to donate methyl groups for covalent modification of different substrates as rRNA, tRNA, and essential proteins, whose methylation status can serve as a regulatory signal for maturation and control interactions with other macromolecules [7,8]. KsgA recognizes a complicated substrate. Although KsgA can bind naked 16S rRNA, it remains enzymatically inactive until a partially matured

small subunit has assembled [9]. The reason for this is unclear, but a second interaction, beyond the site of modification, between KsgA and its pre-30S substrate may be required for KsgA activation [10]. The modification of two adenosines in 16S/18S rRNA is universally conserved [7,12]. Despite the high level of conservation, the enzyme is not absolutely essential in bacteria [13]. Loss of KsgA activity is the most common mode of resistance to the antibiotic kasugamycin in *E. coli* and other bacteria [10]. The price paid for loss of KsgA function is a slower growth rate and reduced translational fidelity in *E. coli* and reduced virulence in the human pathogen, *Yersinia pseudotuberculosis* [14].

In order to gather information about the proteins involved in the mechanisms of the bacterium pathogenicity, we chose the *orfxf2148* for the characterization studies.

The XF2148 protein belongs the family of methyltransferases. MTases (methyltransferases) catalyse the transfer of methyl groups from AdoMet (S-adenosyl-L-methionine) methylating DNA or RNA. DNA or RNA methylation has been reported to be essential for bacterial virulence, and it has been suggested that adenine MTases could be potential targets for both vaccines and antimicrobials. Drugs that block adenine MTases could slow down bacterial growth and therefore drug-design initiatives could result in a whole new generation of antibiotics.

Here, we report the overexpression and comparative modeling the XF2148 protein from *Xylella fastidiosa*. The purified XF2148 protein exhibited a great similarity to its homologous protein, KsgA from *E. coli*. Its primary structure showed that the enzyme contains five of eight structural motifs usually found in AdoMet-dependent methyltransferases. In addition, comparative modeling using modeling package Modeller 7v7 showed seven β sheet structures which is a remarkable characteristic among these

MTases. Also, this work aims at adding new information on proteins that may be related to the *X. fastidiosa* pathogenesis, necessary for new approaches towards the combat of CVC.

Material and Methods

Cloning *orf xf2148*

The *orf xf2148* of *Xylella fastidiosa* was amplified by PCR using the purified genomic DNA as a template. The PCR was carried out using the following specific primers: forward primer 5'-GGTATTGAGGGTCGCATGATGGATCGC-3' and reverse primer 5'-AGAGGAGAGTTAGAGCCCTAATCATCAGG-3'. The primers were designed to generate products with vector cohesive overhangs, for efficient cloning onto a pET32Xa/LIC vector. This vector is designed for the heterologous protein expression and the target protein is fused to thioredoxin (109 amino acids), His-tag (6 amino acids), and S-tag (15 amino acids) sequences upstream to the cloning site. The fusion tags together have 17.6kDa and can be removed by factor Xa cleavage [15]. The recombinant plasmids were used to transform competent *E. coli* DH5 α cells, using a slightly different method to the standard polyethylene glycol (PEG) method [16]. Cells were cultured overnight at 37°C in Luria-Bertani broth (LB) plates containing 100 μ g/mL ampicillin. The colonies were individually stored at -70°C in a permanent 2YT + HMFM broth [17] and checked by PCR. The *orf xf2148* insert of pET32Xa/LIC plasmid was sequenced to verify the correct nucleotide sequence.

Expression of recombinant XF2148 fusion protein

The recombinant plasmid with *xf2148* insert was transformed into *E. coli* BL21 (DE3) cells by PEG method. The cells were grown in 2 ml of LB medium containing ampicillin (100µg/mL) at 37°C, 300rpm, for 18h. An aliquot of 1mL of the culture was used to inoculate 100ml of the same medium, maintained at the same culture conditions until OD600 reached 0.8. Expression of recombinant protein was induced by the addition of Lactose to a final concentration of 5,6mM and cultivated for 3h. The culture was harvested using 10min centrifugation at 6000g, 4°C, and stored for further use at -20°C. The bacterial pellet obtained from 100mL liquid culture was suspended in 22mL buffer A (50mM phosphate, 100mM NaCl, 5% glycerol, pH 8.0). The PMSF (1mM), protease inhibitor was added, as well lysozyme (1mg/mL). The cells were lysed by sonication on ice, in an ultrasonic cell disruptor (Cole-Parmer Instrument, Chicago, IL, USA). The obtained lysate was centrifuged at 27,000g for 20min at 4°C.

Refolding of recombinant XF2148 fusion protein

After a 3-h induction period, cells were harvested by centrifugation at 4000rpm for 10min and washed in buffer A (50mM phosphate, 100mM NaCl, 5% glycerol, pH 8.0). The washed cell pellets were lysed in buffer A by sonication (10 cycles of 15 s pulses). The insoluble fractions were washed with 0,1% Triton X-100 and 1M urea and the remaining pellet was collected by centrifugation at 15000rpm for 15min. To obtain a soluble recombinant ksgA protein, the protein in its denatured form was refolded by dialysis

against a refolding buffer. The isolated inclusion bodies were resuspended in solubilizing buffer B (8M Urea and 50mM phosphate, pH 8.0). The protein was refolded by dialyzing against buffer A overnight. The urea concentration of the dialysis was changed in just one overnight step at 4°C. The dialyzed sample was collected and concentrated by ultrafiltration using Millipore Amicon Ultra System. The concentrated sample was analyzed by SDS-PAGE and protein concentration was determined by spectrophotometer analysis.

Cleavage of XF2148 fusion protein

The recombinant XF2148 protein was separated from the fusion tags by Thrombin proteolysis (0,005 U/μl per 10μg of recombinant protein). The reaction was performed at 22°C for 2h in buffer 50mM sodium phosphate, 100mM NaCl, 5% glycerol pH 8.0 added to 2,5mM CaCl₂, corresponding to 10% of the total volume. The proteolysis products were analyzed by SDS-PAGE. The sample was dialyzed at 4°C overnight against buffer 50mM sodium phosphate, 5mM EDTA, 1% glycerol pH 8.0 to prepare the sample for the following purification step.

Purification of XF2148 protein

The purification was done by cation-exchange chromatography using a 1mL SP HiTrap from Invitrogen in an AKTA-FPLC system. The sample was applied into the column equilibrated with 10 column volumes (CV) of equilibration buffer (50mM phosphate, 50mM NaCl, 5% glycerol, 5mM EDTA, pH 8.0). The unbounded proteins were

washed out with 10 CV of equilibration buffer and bound proteins were eluted using a linear NaCl gradient of 0-1.0M in 10 CV of buffer 50mM phosphate, 500mM NaCl, 5% glycerol, 5mM EDTA, pH 8.0. The fractions were collected and analyzed by SDS-PAGE.

Circular Dichroism spectroscopy

Circular dichroism (CD) spectra of the purified protein in solution were measured with a Jasco J-810 Spectropolarimeter dichrograph (Japan Spectroscopic, Tokyo, Japan). The CD spectra were generated using a quartz cuvette with a cell path with a length of 1.0mm, containing 200 μ L of recombinant XF2148 protein (2.0mg/mL) in 5mM Tris-HCl buffer, pH 7.5. Four scans in the 260–200nm range at a rate of 20nm/min at 20°C were recorded for sample and average measurement.

In-gel trypsin digestion

XF2148 protein isolated by SDS-PAGE was excised from the gel with a clean razor and prepared for complete trypsinolysis. The excised gel piece was diced into smaller pieces (1mm²) processed in one tube by destaining until clear 2x10min in potassium ferricyanide: sodium tiosulfate (1:1), and then washed 3x10min in water. Protein disulfide was reduced by incubation at 60 °C for 30 min with 10 μ L of 45 mM dithiothreitol in 100mM ammonium bicarbonate. The residual liquid was removed and then the proteins were alkylated by adding 100 μ L of 100mM iodoacetamide. Residual liquid was removed and gel pieces were washed with 500 μ L of 50mM ammonium bicarbonate in 50% acetonitrile. Gel pieces were dried under vacuum, and then digestion using trypsin was

performed in the same tubes by adding 25mM ammonium bicarbonate, proportion of 0,2µg of trypsin/1µg sample (according to protein concentration). After digestion 18 h at 37°C, soluble trypsin fragments were recovered in the supernatant and combined with three extractions of the gel pieces with 25µl of 60% acetonitrile 0.1% TFA. The pooled extracted fragments were dried under speed vacuum at 37 °C.

Purification of peptides by ZipTip_{µ-C18}

The sample was dissolved in 10µl of 2% TFA. ZipTip was washed three times with 15µl (0,1% TFA/50%acetonitrile), and then equilibrated four times with 15µl of 0,1% TFA. The sample was loaded onto ZipTip (three times up and down) and peptides bound onto ZipTip were washed three time with 0,1% TFA. The sample was eluted into 20 µl of 100% acetonitrile solution (three times up and down).

Peptide Fingerprint by Mass Spectrometry

Analysis of all samples obtained by tryptic digestion was prepared using the dried droplet method. This was carried out as follows. The samples were acidified by addition of 0.5 1 water/TFA 100:0,1 (V/V) and were left at room temperature for a few minutes, so that the droplet volume was reduced by evaporation. The matrix α -Cyano-4-hydroxycinnamic acid, was prepared followed by 1% (m/V) on H₂O/ACN 1:1 (V/V) (1uL) containing TFA 0,1%, was added and the sample was allowed to dry at ambient temperature. All measurements were performed on a MALDI-TOF (Waters - Micromass, UK) mass spectrometer in previously cleaned microplate. Analyses were performed of exclusively

positively charged ions, in reflectron mode. For subsequent data evaluation with peptide identification, a software package Biolynx (Micromass- UK) was utilized.

XF2148 protein Homology Modeling

The linear amino acid sequences confirmed by mass spectrometry were used to construct models of XF2148 protein using as starting geometry coordinates those of the structure of an 2.1 angstrom crystal structure of KsgA: a universally conserved adenosine dimethyltransferase obtained at 2.1Å resolution (Protein Data Bank code 1QYR B chain; [10]). Optimal alignment of the XF2148 protein sequence was used to predict the structure using ESyPred3D and neural networks (<http://www.fundp.ac.be>; [18]). The 3D structure was built using the modeling package Modeller 7v7 [19]. Alignments for the best model template were obtained by combining, weighting and screening the results of several multiple alignments and were confirmed using the PDB-Blast server (<http://www.rcsb.org/pdb/>). The resulting structure file was examined and evaluated using the software RIBBONS [20] and the Swiss-PDBViewer v.3.51 (<http://us.expasy.org>). The theoretical pI of the proteins was calculated using the Expasy compute pI/MW tool (http://ca.expasy.org/tools/pi_tool.html).

Results and Discussion

Strategy for *orf xf2148* cloning and expression

The *orf xf2148* was amplified by PCR from the *X. fastidiosa* genomic DNA and cloned into the pET32Xa/LIC vector (Novagen). Once in the expression vector pET32Xa/LIC, the complete *orf xf2148* in frame with N-end histidine tag was transformed in *E. coli* DH5 α strain and confirmed through DNA sequencing. After transformation of the *E. coli* BL21 (DE3) strain with pET32Xa/LIC-XF2148, several recombinant clones were selected to test the expression levels of the recombinant protein. The expressed protein was detected by SDS–PAGE showing a high level of expression of recombinant XF2148 following the addition of inductor lactose. The pET32Xa/LIC-XF2148 construction was successfully transformed into the BL21(DE3) strain. Therefore, the overexpression level of the recombinant protein was tested under various conditions. In none of the cases the recombinant protein was expressed in the soluble form. SDS-PAGE analysis indicated that the recombinant XF2148 protein was expressed in the insoluble fraction as inclusion bodies.

Refolding of XF2148 protein

A consistent feature is that many disulfide bonded proteins accumulate in aggregates such as inclusion bodies when expressed at high non-physiological levels in bacteria. Inclusion bodies can be solubilized using agents such as 8M urea or 6M guanidine

hydrochloride. This protein species is then subjected to sequential dialysis to facilitate the movement from a denatured and reduced state to a folded state that also generates a soluble species. The REFOLD database (<http://www.refold.med.monash.edu.au>) shows that the majority of protein refolding techniques use dialysis or dilution procedures. These techniques expose a protein to conditions that promote partial denaturation but also help the molecule to slowly reach minimal energy states where correct folding can occur. However, the gradual removal of a denaturant such as guanidine hydrochloride is a critical factor in the refolding process. In this study, the protein was dialyzed in just one overnight step refolding. There was still loss of a substantial proportion of the recombinant XF2148 during the dialysis procedure consistent with the formation of incorrect disulfide bonds resulting in protein aggregates.

Recombinant XF2148 cleavage and purification

After the refolding and dialyze approximately 8mg of purified fusion recombinant protein per 0,5L of initial bacterial broth were recovered. The cleavage of the N-terminal fusion tags from the target protein was performed with success by thrombin proteolysis due to the poor efficiency of the factor Xa. After removing the fusion tag proteins a cation-exchange chromatography was performed to separate the fragment containing the tags and the target protein. These fusion tags were bounded in the column (Fig.2) and the XF2148 protein was eluted during the NaCl gradient, indicating that the separation was achieved. The final amount of XF2148 protein obtained was approximately 2mg per 0,5L of initial bacterial broth.

Sequence and secondary structure comparisons

The alignment of the XF2148 primary sequence with one ksgA structure solved protein from *E. coli*, revealed many interesting features (Fig.4). First, the two analyzed sequences have a similar molecular mass about 34 kDa. Second, a site involving highly conserved residues (R221, R222, K223 and R248) in the C-terminal domain is also observed in the XF2148 protein from *X. fastidiosa*. Moreover, the distributions of amino acid residues highly conserved are quite similar among the compared proteins. The predicted secondary structures contents are very close to both the XF2148 from *X. fastidiosa* and KsgA from *E. coli*, being that the seven β sheet structure is a remarkable characteristic among these proteins (Fig.5).

Both are members of the group of methyltransferases based on the order of conserved structural motifs [21]. XF2148 protein clearly demonstrates five of the eight structural motifs usually found in AdoMet-dependent methyltransferases. Residues in these motifs have been shown to be involved in SAM binding and interaction with the target adenosine base [22,23]

Motif I contains a canonical GXG sequence, which, in other Mtases, forms part of the binding pocket for S-AdoMet. Motif II consists of an Asp or Glu residue (E66 in KsgA) followed by a hydrophobic residue (L67). Motif IV contains residues that are important in catalysis. The consensus sequence in DNA amino MTases is (D/N)PPY and the XF2148 protein sequence is NLPY, residues 113–116. The first residue may form a hydrogen bond to the amino group in the target base and may play a role in its deprotonation. Mutation of this residue is poorly tolerated in DNA MTases [24-27]. Motif VI contains three adjacent

hydrophobic residues at the C-terminal end (F138, M139 and L140) that are suggested to be involved in proper positioning of the target adenosine base. Motif VIII contains a very highly conserved Phe residue. This residue is involved in stabilization of the target base through stacking interactions.

Circular Dichroism spectroscopy analysis

In order to investigate the structural integrity and secondary structure contents of the purified XF2148, the protein was analyzed by CD spectroscopy. Circular dichroism is an important technique to estimate secondary structure content, to validate protein quality, and for monitoring conformational changes of proteins. The XF2148 CD spectrum showed that it contains predominant signal of α -helices (Fig. 2). Values of the CD signals at 208nm and 222nm estimates the α -helices content of peptides and proteins. The results suggest that XF2148 protein maintains secondary structures and is viable to functional and structural studies and remained folded, besides it is in agreement with secondary structure prediction from the PSIPRED program (<http://bioinf.cs.ucl.ac.uk/psipred>)

Mass Spectrometry Analysis

The identity of the resulting protein product was confirmed by trypsin digestion followed by MALDI-TOF analysis. The peptide profile found for the mass spectroscopy analysis of the digested XF2148 protein (“fingerprint”) was compared to a theoretical digestion of the protein and the result indicated a monoisotopic mass of 35,062 Da (Fig.3).

Homology Modeling of XF2148 protein

Knowledge of three-dimensional protein structure is crucial to answering many biological questions; however, the rapidly growing number of sequenced genes and genomes is heavily outpacing the number of experimentally determined structures. Comparative modeling is based on the general observation that evolutionarily related sequences have similar three-dimensional structures [28]. As a consequence, a three-dimensional model of a protein of interest (target) can be built from related protein(s) of known structure (templates) that share statistically significant sequence similarity.

Modeller implements an automated approach to comparative protein structure modeling by satisfaction of spatial restraints. Briefly, the core modeling procedure begins with an alignment of the sequence to be modeled (target) with related known 3D structures (templates). This alignment is usually the input to the program. The output is a 3D model for the target sequence containing all main chain and side chain non hydrogen atoms. Given an alignment, the model is obtained without any user intervention. First, many distance and dihedral angle restraints on the target sequence are calculated from its alignment with template 3D structures. The form of these restraints was obtained from a statistical analysis of the relationships between many pairs of homologous structures.

To elucidate the structural basis of KsgA protein from *Xylella fastidiosa*, we built a homology structural model for XF2148 protein (Fig.6) based on the crystal structure that showed high sequence identity with XF2148 – KsgA from *Escherichia coli* (PDB accession code 1QYR B chain, 47% identity). When the modeled XF2148 structure was superimposed onto the structure of KsgA from *E. coli*, all the residues lining the protein sequence were superimposed well.

Conclusion

We have shown in this paper that the gene coding XF2148 protein from *X. fastidiosa* was expressed in the insoluble form and solubilized using 8M Urea as denaturating agent. The protein was purified in one step by cation-exchange chromatography and its purity and identity were verified by SDS-PAGE and Mass Spectrometry (MALDI-TOF). Its correct folding was verified by Circular Dichroism Spectroscopy analysis that indicated a secondary structure composed mainly of α -helices. The alignment of the XF2148 primary sequence with one KsgA structure solved protein from *E. coli*, revealed a site involving highly conserved residues (R221, R222, K223 and R248) in the C-terminal domain. The distributions of amino acid residues are highly conserved among the compared proteins. Five of the eight structural motifs usually found in AdoMet-dependent methyltransferases are presented in XF2148 protein. According to sequence comparison and homology modeling XF2148 protein contains seven β sheet structure which is a remarkable characteristic among these proteins.

Acknowledgments

The authors thank FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo, CNPq (Centro Nacional de Pesquisa e Desenvolvimento) and CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support. A.P.S. received a fellowship from CNPq. L.K.R. received Ms. And A.R.A., S.F.S., and C.F.C. received postdoc fellowships from FAPESP. D.P.P. received Ph.D. fellowship from CAPES. We also thank the Structural Molecular Biology Center researchers from the National Synchrotron Light Laboratory for technical expertise, particularly Ph.D. Nilson Ivo Tonin Zanchin and Ph.D. Carlos Henrique Ramos (Biology Cellular and Molecular Laboratory); and researchers for equipment supports and collaboration.

List of Figures

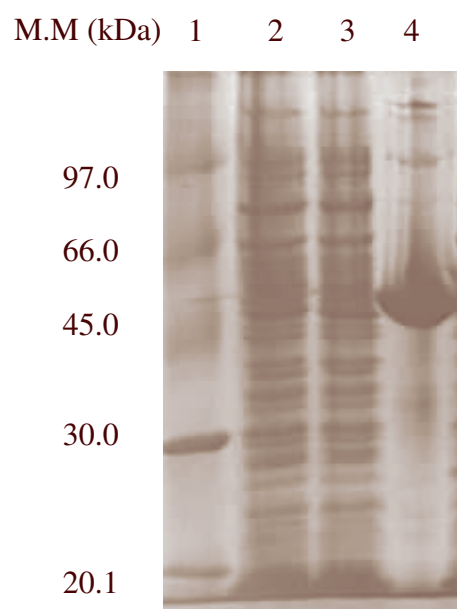


Figure 1. SDS-PAGE of the fractions collected after the expression of the recombinant XPSL. Legend: lane 1—molecular mass markers; 2 - extract from bacterial lysis with no lactose induction, 3 - soluble fraction from bacterial lysis, 4 - insoluble fraction from bacterial lysis. Separation gel of 12% acrylamide concentration.

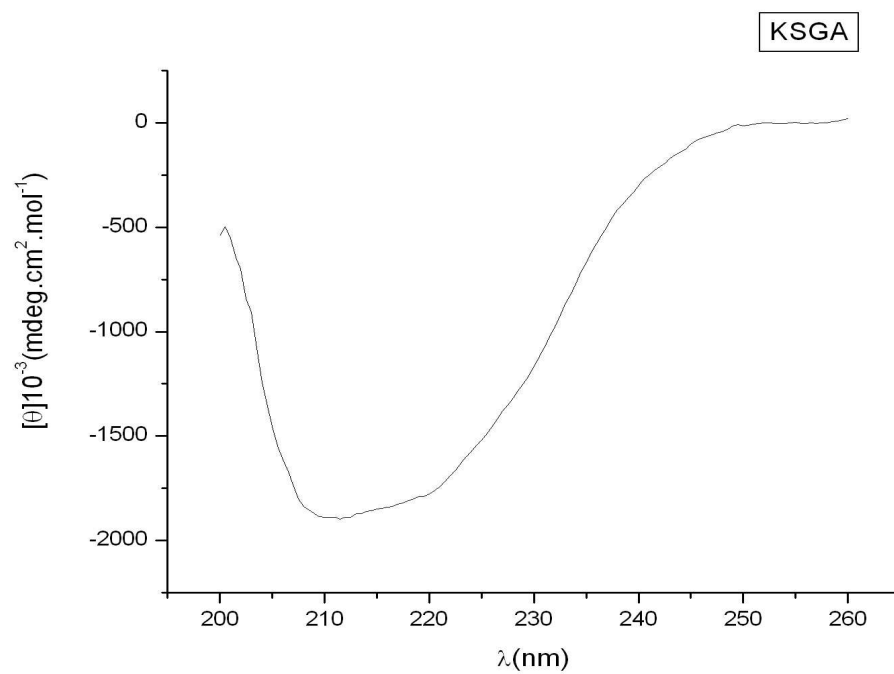


Figure 2. Circular dichroism spectrum of the purified XF2148 protein. The spectrum was generated using XF2148 concentration of 0.1mg/mL in a Tris-HCl buffer (5mM, pH 7.5). Four scans in a 260–200nm range at a rate of 20nm/min at 20°C.

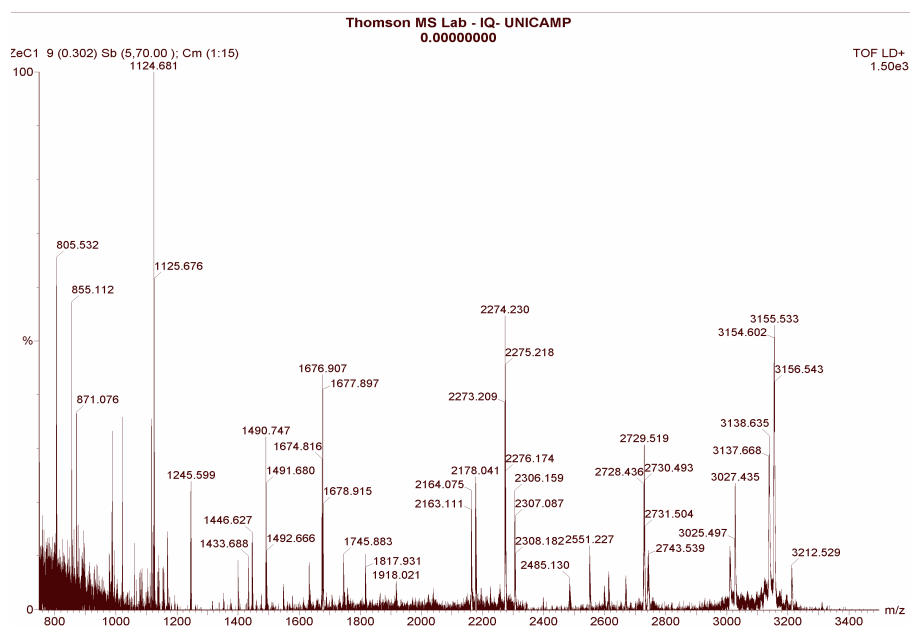


Figure 3. Ionized peptides spectrum from tryptic digestion of XF2148 protein (35kDa).

XF 2148	QHFLVD RYYIDRIIHAITPQPN DHIVEIGPGQGATTLPLLKCCGSLTAIELDRDLIAPLTAAATPLGKLDIIHRDVLTVDLSILA : 85
KSGA_chain_	QNFELNDQFVIDSIVSAINPQKGQAMVEIGPGLAALTEPVGERLDQLTVIELDRDLAARLQTHPFLGPKLTIYQQDAMTFNFGELA : 85
XF 2148	KPGNKKLR LVGNLPYNISSPIIEHVLQQAATIADMHEMLQKEVVD RMAAPPGSKVYGRLSVMLQAWCEVTTMFVVPDDAFQPPPK : 170
KSGA_chain_B	EKMGPPLRVFGNLPYNIISTELMEHLFSYTDATADMHEMLQKEVVNRLVAGPNSKAYGRLSVMAQYYCNVIPVLEVPPSAETPPPK : 170
XF 2148	VNSAITRLVPRDPTTIRIADTKRESDIVRAAFGQRRKTLRNSLADICTPAHFEHAGIRTNARAEQLEVTEFIALAN--AKDT : 250
KSGA_chain_B	VDSAVVRLVEHATMPHPVKDVRVLSRITTEAFNQRRKTI RNSLGNLFSVEVLTGMGIDPAMRAENISVAQYQCMANYLAENA : 252

Figura 4. Comparison of amino acid sequences of XF2148 protein from *X. fastidiosa* with KsgA protein from *E. coli* (PDB 1QYR). Letters shaded in black indicate amino acids that are identical in at least two aligned sequences

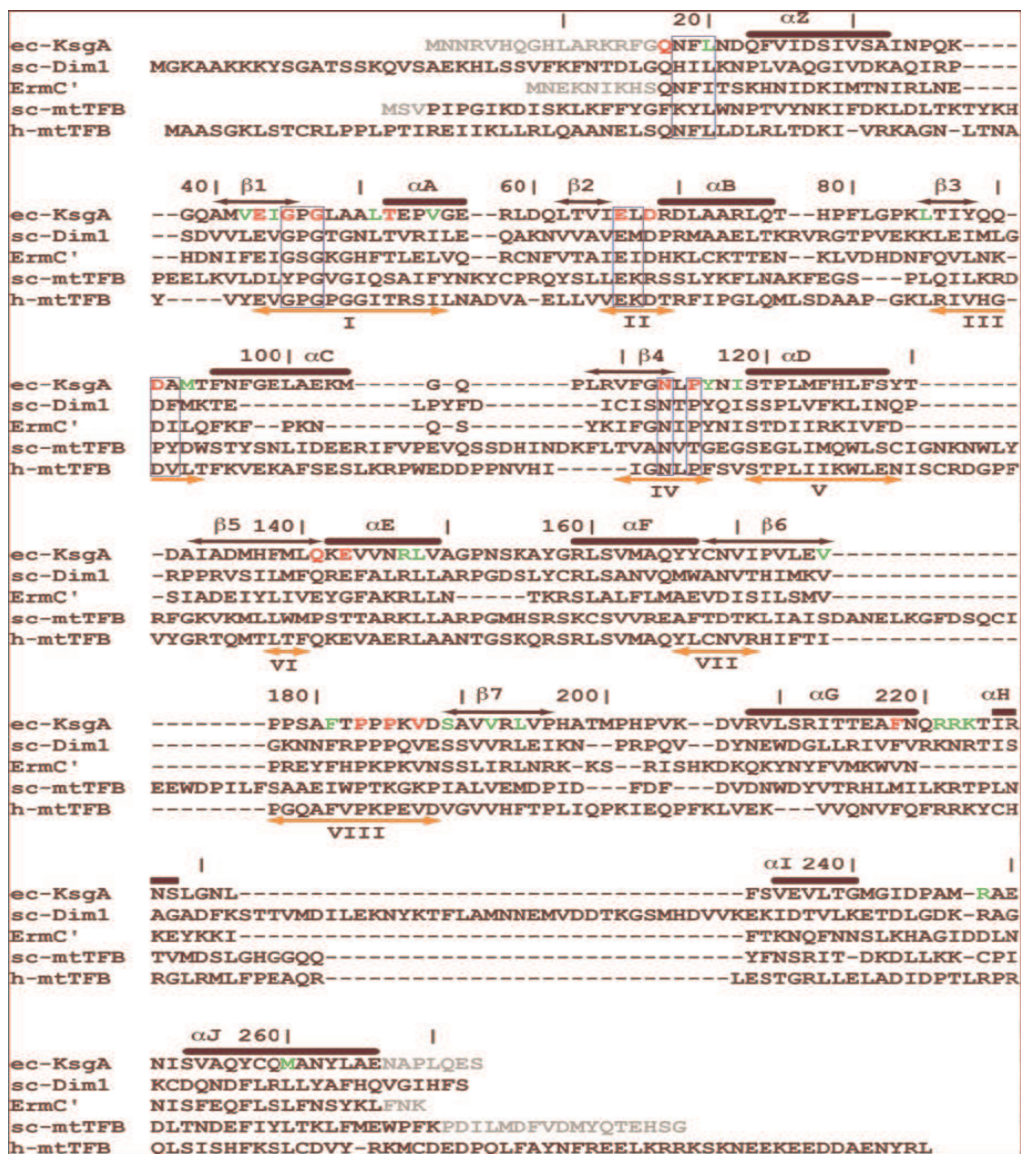


Figure 5. Structure-based sequence alignment of KsgA, Dim1, ErmC0, h-mtTFB, and sc-mtTFB (GenBank accession nos. P06992, P41819, P13956, NP_057104, and P14908). The α -helices and β -strands are indicated by black barrels and double-headed arrows, respectively, above the alignment. Brown double-headed arrows below the alignment indicate eight motifs common to SAM-dependent MTases. Red letters designate residues that are absolutely conserved in KsgA enzymes, while green letters designate residues that are highly conserved. Boxed residues are those that, in ErmC0, make direct contact with the SAM molecule. Figure from H.C. O'Farrell, J.N. Scarsdale, J.P. Rife. Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli*. *J. Mol. Biol.* (2004). 339: 337–353.

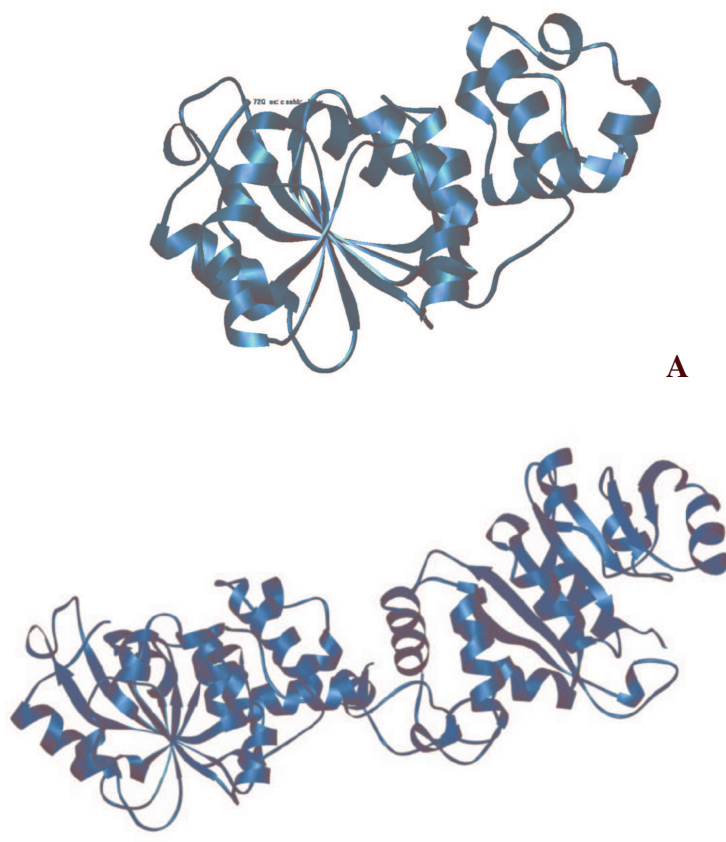


Figure 6. (A) Homology structural model for XF2148 protein from *Xylella fastidiosa* based on the crystal structure of ksgA protein from *Escherichia coli*. (B) structure of an 2.1 angstrom crystal structure of ksgA: a universally conserved adenosine dimethyltransferase from *Escherichia coli* obtained at 2.1Å resolution (Protein Data Bank code 1QYR B).

References

- [1] M. Frome, A. Camargo, S. Heber, C. Czink, A.J.D. Simpson, J.D. Hoheisel, A.P. Souza. Mapping analysis of the *Xylella fastidiosa* genome, Nucleic Acids Res. (2000) 28: 3100–3104.
- [2] R.F. Lee, K.S. Derrick, M.J.G. Beretta, C.M. Chagas, V. Rosetti, Citros variegated chlorosis: a new destructive disease of citros in Brazil, Citros Ind. Oct. (1991) 12–15.
- [3] F.F. Laranjeira, Ten years of citrus variegated chlorosis: What do we know? Laranja 18 (1997) 18:123–141.
- [4] A.J.G Simpson *et al.*, The genome sequence of the plant pathogen *Xylella fastidiosa*, Nature 406 (2000) 151–157.
- [5] H.C. O’Farrel, F.N. Musayev, J.N. Scarsdale, H.T. Wright, J.P. Rife. Crystallization and preliminary X-ray diffraction analysis of KsgA, a universally conserved RNA adenine dimethyltransferase in Escherichia coli. Acta Crystallogr D Biol Crystallogr. (2003). 59: 1490-1492.
- [6] P.Z. Kozbial, A.R. Mushegian. Natural History of S-adenosylmethionine-binding proteins. (2005). 14: 5-19.

- [7] V. Anantharaman, E.V. Koonin, L. Aravind. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* (2002). 30:1427-1464.
- [8] A.K. Hopper, E.M. Phizicky. tRNA transfers to the limelight. *Genes Dev.* (2003) 17:162-180.
- [9] P.M. Desai, J.P. Rife. The adenosine dimethyltransferase KsgA recognizes a specific conformational state of the 30S ribosomal subunit. *Archives of Biochemistry and Biophysics.* (2006). 449: 57-63.
- [10] H.C. O'Farrell, J.N. Scarsdale, J.P. Rife. Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli*. *J. Mol. Biol.* (2004). 339: 337–353.
- [11] P.H. Van Knippenberg, J.M. Van Kimmenade, H.A. Heus. Phylogeny of the conserved 3' terminal structure of the RNA of small ribosomal subunits. *Nucleic Acids Res.* (1984) 12: 2595–2604.
- [12] H.C. O'Farrell, N. Pulicherla, P.M. Desai, J.P. Rife. Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* (2006) 12:725–733.

[13] Lévêque, F., Blanchin-Roland, S., Fayat, G., Plateau, P. & Blanquet, S. Design and characterization of *Escherichia coli* mutants devoid of Ap4N-hydrolase activity. *J. Mol. Biol.* (1990) 212:319–329.

[14] J. Mecsas, I. Bilis, S. Falkow. Identification of attenuated *Yersinia pseudotuberculosis* strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. *Infect. Immun.* (2001). 69: 2779–2787.

[15] Xa/LIC Vector Kits Manual, Novagen, ninth ed., USA, 2000, 56 pp. Available from: www.novagen.com.

[16] B.T. Kurien, R.H. Scofield. Polyethylene glycol-mediated bacterial colony transformation. *Biotechniques.* (1995). 18(6) : 1023-1026.

[17] A.R. Azzoni, S.F.S. Tada, L.K. Rosselli, D.P. Paula, C.F. Catani, A.A. Sabino, J.A.R.G. Barbosa, B.G. Guimarães, M.N. Eberlin, F.J. Medrano, A.P. Souza. Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*. *Protein Expr. Purif.* (2004) 33: 297-303.

[18] C. Lambert, N. Leonard, X. De Bolle, E. Depiereux, ESyPred3D: prediction of proteins 3D structures, *Bioinformatics.* 18 (2002) 1250-1256.

[19] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234 (1993) 779-815.

- [20] M. Carson, RIBBONS, *Methods Enzymol.* 277 (1997) 493-505.
- [21]. Malone, T., Blumenthal, R. M. & Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* (1995) 253, 618–632.
- [22]. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. & Saenger, W. Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. *Proc. Natl Acad. Sci. USA.* (1994) 91, 10957–10961.
- [23] Schluckebier, G., Labahn, J., Granzin, J. & Saenger, W. M.TaqI: possible catalysis via cation– π interactions in N-specific DNA methyltransferases. *Biol. Chem.* (1998) 379, 389–400.
- [24] Guyot, J. B., Grassi, J., Hahn, U. & Guschlbauer, W. The role of the preserved sequences of Dam methylase. *Nucl. Acids Res.* (1993).21, 3183–3190.
- [25] Willcock, D. F., Dryden, D. T. & Murray, N. E. A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* (1994). 13, 3902–3908.
- [26] Kong, H., Smith, C. L. Substrate DNA and cofactor regulate the activities of a multi-functional restriction-modification enzyme, BcgI. *Nucl. Acids Res.* (1997). 15, 3687–3692.

[27] Roth, M., Helm-Kruse, S., Friedrich, T. & Jeltsch, A. Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase. *J. Biol. Chem.* (1998). 273, 17333–17342.

[28] Chothia, C. Lesk, A.M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* (1986) 5:823-826.

6. Resultados Complementares

6.1. Amplificação das *orfs* e subclonagem no vetor pGEM-T

As duas *orfs* em estudo (*orf xf1524* e *orf xf2148*) foram amplificadas a partir do DNA genômico da bactéria *Xylella fastidiosa* utilizando oligonucleotídeos flanqueadores. Os produtos destas amplificações serviram como molde para novas amplificações, utilizando-se os oligonucleotídeos específicos para cada tipo de vetor de expressão. Para otimizar cada reação de PCR, várias temperaturas de anelamento foram testadas de acordo com cada oligo. As figuras 5.1 e 5.2 mostram as condições iniciais propostas para amplificação de cada *orf* com seus respectivos oligos para os vetores pET28 e 29a (Novagen, EUA). Em algumas condições de amplificação fez-se necessário o uso do aditivo glicerol. A partir do resultado visualizado em gel de agarose 1% (tampão TBE 1X), escolheu-se a melhor temperatura de anelamento para o par de oligonucleotídeos.

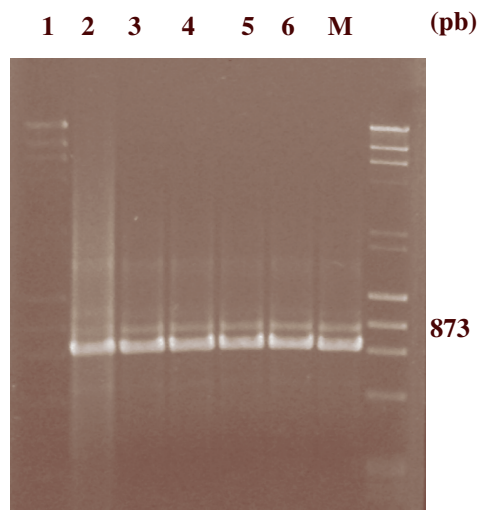


Figura 5.1. Foto de gel de agarose a 1% contendo os produtos de PCR da *orf xf2148* utilizando oligos específicos para o vetor pET28a. As linhas 1 a 6 representam diferentes temperaturas de anelamento. M – marcador de peso molecular λ /Hind III e ϕ x/174; 1 – 55°C; 2 – 56°C; 3 – 57,6°C; 4 – 58,6°C; 5 – 60,1°C; 6 – 61°C.

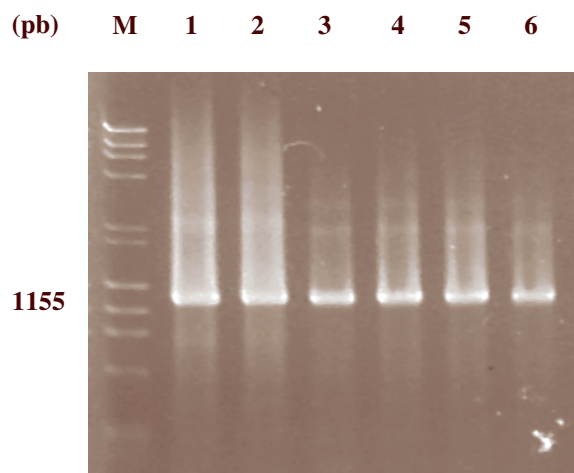


Figura 5.2. Foto do gel de agarose a 1% contendo os produtos de amplificação da *orf xf1524* utilizando oligonucleotídeos específicos para o vetor pET28a. As linhas 1 a 6 representam diferentes temperaturas de anelamento. M – marcador de peso molecular ϕ x 174/ λ Hind III; 1 – 56,3°C; 2 – 57,3°C; 3 – 59,8°C; 4 – 61°C; 5 – 61,8°C; 6 – 63°C.

Já as figuras 5.3 e 5.4 mostram as amplificações específicas para as duas *orfs*, fazendo-se uso dos oligonucleotídeos específicos para o vetor pET29a (Novagen, EUA).

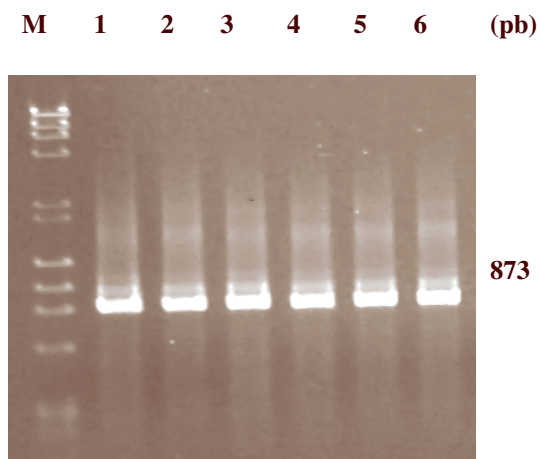


Figura 5.3. Foto de gel de agarose a 1% contendo os produtos de amplificação da *orf xf2148* utilizando oligonucleotídeos específicos para o vetor pET29a e aditivo glicerol a 5%. As linhas 1 a 6 representam diferentes temperaturas de anelamento. M – marcador de peso molecular ϕ x 174/ λ *Hind III*; 1 – 57°C; 2 – 58°C; 3 – 59,6°C; 4 – 60,6°C; 5 – 61,5°C; 6 – 62,6°C.

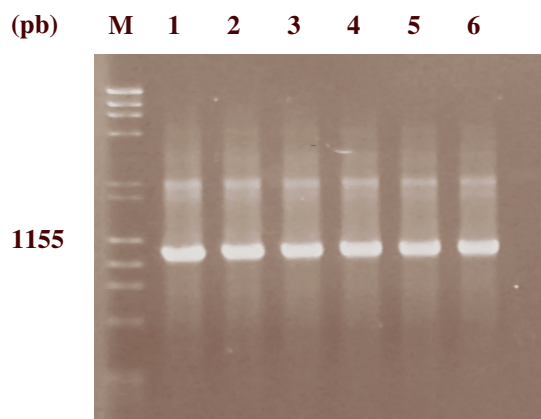


Figura 5.4. Foto do gel de agarose a 1% contendo os produtos de amplificação da *orf xf1524* utilizando oligonucleotídeos específicos para o vetor pET29a. As linhas 1 a 6 representam diferentes

temperaturas de anelamento. M – marcador de peso molecular ϕ x 174/ λ Hind III; 1 – 57°C; 2 – 58°C; 3 – 59,6°C; 4 – 60,6°C; 5 – 61,5°C; 6 – 62,6°C.

Nota-se que em todos os géis não houve apenas uma banda, sendo isso resultado de uma inespecificidade dos oligonucleotídeos na temperatura de anelamento escolhida. Para que não houvesse coleta de material indesejado, os fragmentos de interesse foram purificados através de eletroeluição e subclonados no vetor de clonagem pGEM-T. Uma vez feita a ligação do inserto no vetor, realizou-se a transformação dos plasmídeos recombinantes em células competentes de *E. coli* DH5 α de forma eficaz, já que todas as placas contendo meio LB sólido apresentaram alto índice de colônias brancas e pouca azuis. As colônias positivas foram devidamente armazenadas e confirmadas por sequenciamento.

6.2. Digestão enzimática utilizando endonucleases de restrição e ligação nos vetores pET28 e 29a

As reações de digestão para clonagem nos vetores pET28 e 29a das duas *orfs* em estudo (*orf xf2148* e *orf xf 1524*) subclonadas em pGEM-T foram feitas com as enzimas apresentadas na tabela 1. Todas as digestões foram realizadas com sucesso havendo a liberação dos insertos de interesse do plasmídeo pGEM-T. As ligações seguindo a metodologia do manual do sistema pET também foram realizadas com êxito e verificadas por PCR de “colônia”, mas ainda não por sequenciamento.

Tabela 1. Lista de oligonucleotídeos e enzimas de restrição para os vetores pET28 e 29.

Oligo	Seqüência (5'→3')	Endonuclease de Restrição	Vetor de Expressão
XF1524 flanq. direto	GGTGGCGCCGAAGATTCAGATT ACG	—	—
XF1524 flanq. reverso	GCCGATGCAAGCTCAACCGAAG ACT	—	—
XF1524 direto p28	CATATGTTGGGAAGAGGGTATC GC	NdeI	pET28a
XF1524 reverso p28	CTCGAGTTAACGTTGTGCCTTGC C	XhoI	pET28a
XF1524 direto p29	CATATGTTGGGAAGAGGGTATC GC	NdeI	pET29a
XF1524 reverso p29	CTCGAGACGTTGTGCCTTGCCAT C	XhoI	pET29a
XF2148 flanq. direto	GCAGCACCAAGCACCATTAT	—	—
XF2148 flanq. reverso	TAGCGGCCTTCATTACGATTAG	—	—
XF2148 direto p28	ATCTCAATCGCTAGCGTGCATAC	NheI	pET28a
XF2148 reverso p28	ACTCGAGTTAAGTGTCTTGGCA TTG	XhoI	pET28a
XF2148 direto p29	GATATCGTGCATACGCCACACTC C	EcoRV	pET29a

XF2148 reverso p29	CTCGAGAGTGTCTTGGCATTGG	XhoI	pET29a
	C		

6.3. Expressão e purificação em baixa escala das proteínas recombinantes

A expressão das duas proteínas recombinantes em estudo foi feita a partir do vetor pET32 Xa/LIC (Novagen, EUA). Para a proteína codificada pela *orf xf2148*, foram escolhidas as linhagens de *E. coli* BL21(DE3), BL21(DE3)ptGroE e BL21(DE3)pLysS para os testes iniciais. A BL21(DE3) apresentou bons resultados com outras proteínas em estudo em nosso grupo de pesquisa e tornou-se assim a primeira opção para os testes de expressão e solubilidade das novas proteínas em estudo. A linhagem BL21(DE3) é caracterizada pela presença do fago lambda DE3 que carrega o gene que codifica a enzima T7 RNA Polimerase, a qual está sobre o controle do promotor lacUV5. Além disso, a BL21(DE3) é deficiente em duas proteases (lon e OmpT) resultando em uma redução na degradação de proteínas heterólogas. Já a BL21(DE3)ptGroE apresenta as mesmas características da BL21(DE3) acrescentando-se o fato de esta linhagem possuir um plasmídeo que expressa uma chaperonina, a qual auxilia no enovelamento protéico e, conseqüentemente, aumenta a solubilidade das proteínas heterólogas. A linhagem BL21(DE3)pLysS possui um plasmídeo que estabiliza a expressão dos genes alvo. O plasmídeo codifica a T7 lisozima, que atua como inibitor natural da T7 RNA polimerase além de auxiliar na extração da proteína, fragilizando a membrana da célula. Os testes iniciais demonstraram a expressão das proteínas em sua forma insolúvel. Os corpos de inclusão responsáveis pela precipitação da proteína podem ocorrer devido a um enovelamento protéico incorreto imposto pelas condições de indução (Trésaugues *et al.*,

2004). As expressões de ambas as proteínas obtiveram melhores resultados no vetor pET32Xa/LIC enquanto os resultados de expressão de ambas as proteínas em estudo nos vetores pET28a e pET29a mostraram um menor rendimento protéico e uma maior insolubilidade.

6.4. Proteína XpsL expressa sem o peptídeo sinal (59 kDa)

Pelas dificuldades encontradas em se trabalhar com uma proteína de membrana, optou-se pela tentativa de clonar, expressar, purificar e caracterizar a proteína XpsL sem o peptídeo sinal responsável pelo ancoramento da proteína na membrana citoplasmática. Este teste foi realizado na tentativa de se obter maior solubilidade e rendimento da proteína para os estudos de caracterização estrutural e funcional. Para tanto, foram desenhados oligonucleotídeos que se anelassem na sequência da *orf* posteriormente a sequência do peptídeo sinal, acrescentando-se o códon de iniciação ATG para o início da transcrição.

6.5. Amplificação da *orf xf1524* sem a sequência referente ao peptídeo sinal

O fragmento em estudo foi amplificado a partir do DNA genômico da bactéria *Xylella fastidiosa* utilizando-se oligonucleotídeos flanqueadores. O produto desta PCR serviu como molde para uma nova amplificação, mas desta vez utilizando-se oligonucleotídeos específicos ao gene (Figura 5.5) para inserção do mesmo no vetor de expressão pET32Xa/LIC (Novagen, EUA).

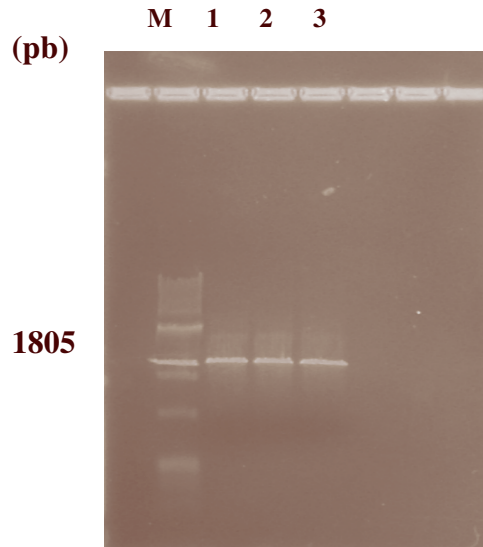


Figura 5.5 Foto do gel de agarose a 1% contendo os produtos de PCR da região compreendida pela *orf xf1524* utilizando oligos flanqueadores. M – marcador de peso molecular λ /Hind III e ϕ x/174; 1 – Amplificação da região compreendida pela *orf xf1524*. 2 e 3 - duplicatas.

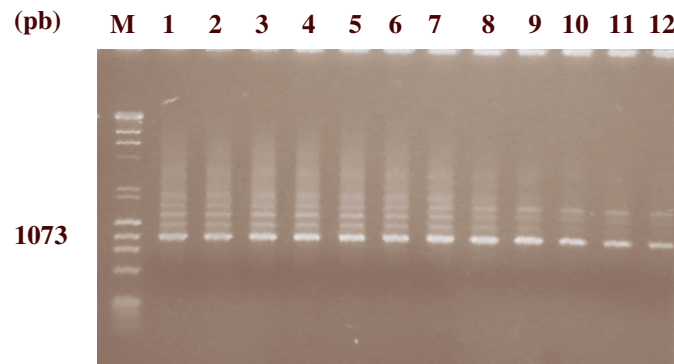


Figura 5.6. Foto do gel de agarose a 1% contendo os produtos de PCR da *orf xf1524* sem a sequência referente ao peptídeo sinal utilizando oligos específicos para o vetor pET32Xa/LIC. As linhas 1 a 12 representam diferentes temperaturas de anelamento. M – marcador de peso molecular λ /Hind III e ϕ x/174; 1 – 54°C; 2 – 54,2°C; 3 – 54,8°C; 4 – 55,5°C; 5 – 56,6°C; 6 – 57,9°C; 7- 59,4°C; 8- 60,7°C; 9- 61,7°C; 10- 62,3°C; 11- 62,8°C; 12- 63°C.

Para que não houvesse contaminação, os fragmentos de interesse foram purificados através de eletroeluição e clonados no vetor de expressão pET32Xa/LIC. Uma vez feita a

ligação do inserto no vetor, realizou-se a transformação dos plasmídeos recombinantes em células competentes de *E. coli* DH5 α . As colônias positivas foram devidamente armazenadas em meio de cultura HMFMM e confirmadas por sequenciamento de nucleotídeos. Logo em seguida, os plasmídeos foram extraídos das células DH5 α e transformadas na linhagem BL21(DE3) utilizada para expressão de proteínas.

6.6. Expressão e purificação da proteína XpsL sem o peptídeo sinal

A proteína XpsL sem o peptídeo sinal fusionada (59kDa) foi expressa na linhagem de *E. coli* BL21(DE3) utilizando o vetor pET32Xa/LIC, devido aos bons resultados de expressão obtidos anteriormente com esta linhagem. Porém nos primeiros ensaios a proteína foi expressa em sua forma insolúvel (Figura 5.7), contrariando os resultados esperados. Mesmo assim, uma cromatografia de afinidade em coluna convencional de gravidade contendo a resina de agarose Ni²⁺ (Ni-NTA, Qiagen) foi realizada com o intuito de se obter uma fração solúvel da proteína XpsL sem o peptídeo sinal. Corroborando os resultados de expressão preliminares, a cromatografia de afinidade da fração solúvel da expressão da proteína de membrana sem o peptídeo sinal não mostrou nenhuma banda, contrariando as expectativas geradas antes da realização do experimento. O peptídeo sinal é uma porção da proteína localizada na região N-terminal caracterizada por uma alta hidrofobicidade devido ao alto teor de aminoácidos hidrofóbicos em sua sequência.

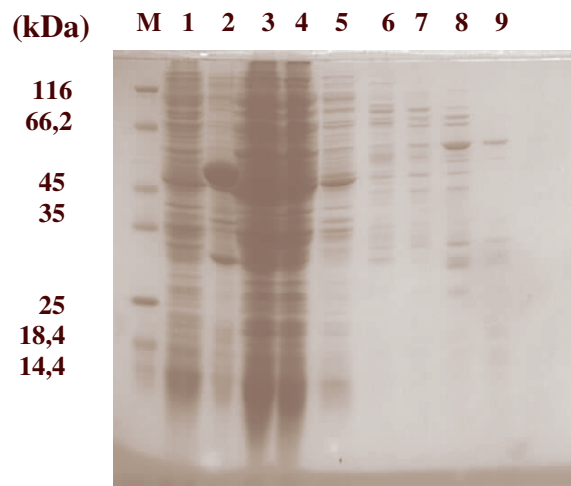


Figura 5.7. Gel SDS-PAGE mostrando os resultados de expressão e purificação em coluna de afinidade contendo resina de agarose Ni^{2+} (Ni-NTA, Qiagen) da proteína XpsL sem o peptídeo sinal no vetor pET32Xa/LIC. M – marcador de peso molecular; 1 – extrato sem indução com lactose; 2 – extrato da fração insolúvel; 3 – Extrato da fração solúvel; 4 – lavagem da coluna com tampão 50mM Tris-HCl pH 7,5 1M NaCl; 5 – eluição com 5mM de imidazol; 6 – eluição com 10mM de imidazol; 7 - eluição com 50mM de imidazol; 8 – eluição com 100mM de imidazol; 9 – eluição com 200mM de imidazol. Gel de poliacrilamida concentração 13,5%; voltagem,160V.

6.7. Expressão e purificação da proteína XpsL recombinante (61 kDa) adicionando detergente ao tampão de lise

Devido ao caráter anfipático da proteína transmembrana XpsL recombinante, testes de expressão em pequena escala foram realizados adicionando-se detergentes ao tampão de lise para tentar aumentar o rendimento e solubilidade da proteína. Em nenhum dos dois experimentos obteve-se resultados satisfatórios. Em ambos os casos, a proteína XpsL recombinante não foi expressa na fração solúvel. Esperava-se que a adição dos detergentes

Tween 20 1% (não iônico) e Octyl Glucopyranoside 2% (não ionico) no tampão de lise da proteína transmembrana XpsL pudesse contribuir para uma maior rendimento da fração solúvel do extrato protéico, pelo fato da proteína possuir regiões em sua sequência tanto hidrofóbicas como hidrofílicas, apresentando portanto um caráter anfipático.

7. Conclusões

✓ Diferentemente de outras proteínas de membrana relatadas na literatura, a proteína XpsL de *Xylella fastidiosa* foi expressa parcialmente solúvel, no vetor pET32Xa/LIC em BL21(DE3) meio aquoso sem a adição de qualquer detergente. Segundo dados de Dicroísmo Circular a proteína apresentou-se enovelada e estável.

✓ A enzima Dimetiladenosina transferase de *Xylella fastidiosa* foi resolubilizada a partir da fração insolúvel expressa em linhagem BL21(DE3) de *E. coli* apresentando grande similaridade em termos de alinhamento de sequência com a sua proteína homóloga KsgA de *E. coli*. Segundo os dados de alinhamento e modelagem molecular, a enzima dimetiladenosina transferase de *X. fastidiosa* apresentou cinco dos oito motivos característicos da família de enzimas Metiltransferases dependentes do substrato S-Adenosil-L-Metionina e três resíduos de aminoácidos conservados evolutivamente no domínio C-terminal (R221, R222, K223 and R248).

8. Perspectivas

✓ Para a proteína de membrana XpsL de *Xylella fastidiosa* novos vetores de expressão, como por exemplo o vetor PSV 282 que possui a maltose binding protein (MPB) e um sítio de clivagem para a protease TEV, podem ser testados no intuito de aumentar o rendimento protéico e possibilitar a excisão da proteína de fusão.

✓ Para a enzima Dimetiladenosina transferase um experimento de SAXS (Small Angle X-ray light Scattering) pode ser realizado com o intuito de complementar os dados estruturais já obtidos pelo alinhamento de sequência e modelagem molecular com a proteína homóloga KsgA de *E. coli*.

✓ A expressão em larga escala da Dimetiladenosina transferase poderá ser feita para obter-se uma maior concentração da proteína a fim de se realizarem ensaios cristalográficos.

✓ Ensaios funcionais através de Espectrometria de Massas e Ressonância Magnética Nuclear (NMR) poderiam ser realizados para verificar a função de metilação por parte da enzima Dimetiladenosina transferase.

9. Referências Bibliográficas

Abendroth, J., Bagdasarian, M., Sandkvist, M., Hol, W.G.J. (2004a) The structure of the cytoplasmic domain of epsL, an inner membrane component of the type II secretion system of *Vibrio cholerae*: an unusual member of the *actin-like* ATPase superfamily. *J Mol Biol*, **344**: 619-633.

Abendroth, J., Rice, A.E., McLuskey, K., Bagdasarian, M., Hol, W.G.J. (2004b) The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from *Vibrio cholerae*: the simplest version of the ferredoxin fold. *J Mol Biol*, **338**: 585-596.

Anantharaman, V., Koonin, E.V., Aravind, L. (2002) SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol*, **4**: 71-75.

Ball, G., Chapon-Herve, V., Bleves, S., Michel, G., Bally, M. (1999) Assembly of XcpR in the cytoplasmic membrane is required for extracellular protein secretion in *Pseudomonas aeruginosa*. *J Bacteriol*, **181**: 382-388.

Bleves, S., Gerard-Vincent, M., Lazdunski, A., Filloux, A. (1999) Structure-function analysis of XcpP, a component involved in general secretory pathway-dependent protein secretion in *Pseudomonas aeruginosa*. *J Bacteriol*, **181**: 4012-4019.

Brenner, S.E. (2001). A tour of structural genomics. *Nature Reviews*. **2**: 801-809.

Burrows, L.L. (2005) Weapons of mass retraction. *Mol Microbiol*, **57**: 878-888.

Camberg, J.L., Sandkvist, M. (2005) Molecular analysis of the *Vibrio cholerae* type II secretion ATPase EpsE. *J Bacteriol*, **187**: 249-256.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collet, J.R., D'Souza, L.M., Du, Y. et al. (2002) The comparative RNA Web (CRW) Site: an online data-base of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**: 2.

Chen, Y., Shiue, S.J., Huang, C.W., Chang, J.L., Chien, Y.L., Hu, N.T., Chan, N.L. (2005) Structure and function of the XpsE N-terminal domain, an essential component of the *Xanthomonas campestris* type II secretion system. *J Biol Chem*, **280**: 42356-42363.

Cianciotto, N.P. (2005) Type II secretion: a protein secretion system for all seasons. *Trends Microbiol*, **13**: 581-588.

Cole, R. B. (1997). Electrospray ionization Mass Spectrometry-Fundamentals, Instrumentation & Applications. New York: John Wiley&sons.

d'Enfert, C., Ryter, A., Pugsley, A.P. (1987). Cloning and expression in *Escherichia coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase. *EMBO J*, **6**: 3531-3538.

Daniels, M.J.; Dow, J.M. (2000). *Xylella* genomics and bacterial pathogenicity to plants. *Yeast*. **17**: 263-271.

de Groot, A., Filloux, A., Tommassen, J. (1991). Conservation of xcp genes, involved in the two-step protein secretion process, in different *Pseudomonas* species and other gram-negative bacteria. *Mol Gen Genet*, **229**: 278-284.

de Keyser, J., van der Does, C., Driessen, A.J. (2003). The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci*, **60**: 2034-2052.

- De Souza, A.A, Takita, M.A, Pereira, E.O, Filho, H.D.V, Machado, M.A. (2005). Expression of pathogenicity related genes of *Xylella fastidiosa* in vitro and in planta. *Current Microbiology*, **50**: 223-228.
- Desai, P.M, Rife, J.P. (2006). The adenosine dimethyltransferase KsgA recognizes a specific conformational state of the 30S ribosomal subunit. *Arch Biochem Biophys*, **449**(1-2):57-63.
- DeShazer, D., Brett, P.J., Burtnick, M.N., Woods, D.E. (1999). Molecular characterization of genetic loci required for secretion of exoproducts in *Burkholderia pseudomallei*. *J Bacteriol*, **181**: 4661-4664.
- Dow, J.M., Daniels, M.J., Dums, F., Turner, P.C., Gpugh, C. (1989). Genetic and biochemical analysis of protein export from *Xanthomonas campestris*. *J Cell Sci Suppl*, **11**: 59-72.
- Durand, E., Bernadac, A., Ball, G., Lazdunski, A., Sturgis, J.N., Filloux, A. (2003). Type II protein secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *J Bacteriol*, **185**: 2749-2758.
- Fasman, G. D. (1996). Circular Dichroism and the Conformational Analysis of Biomolecules. New York: Plenum Press.
- Filloux, A., Bally, G., Akrim, M., Tommassen, J., Lazdunski, A. (1990). Protein secretion in gram-negative bacteria: transport across the outer membrane involves common mechanisms in different bacteria. *EMBO J*, **9**: 4323-4329.
- Forest, K.T., Satyshur, K.A., Worzalla, G.A., Hansen, J.K., Herdendorf, T.J. (2004). The pilus-retraction protein PilT: ultrastructure of the biological assembly. *Acta Crystallogr D Biol Crystallogr*, **60**: 978-982.

Gerard-Vincent, M., Robert, V., Ball, G., Bleves, S., Michel, G.P.F., Lazdunski, A., Filloux, A. (2002). Identification of XcpP domains that confer functionality and specificity to the *Pseudomonas aeruginosa* type II secretion apparatus. *Mol Microbiol*, **44**: 1651-1665.

Gross, J. H. (2004). Mass Spectrometry- A Textbook. Berlin: Springer.

Helser, T.L., Davies, J.E., Dahlberg, J.E. (1972) Mechanism of kasugamycin resistance in *Escherichia coli*. *Nature New Biol.*, **235**: 6-9.

Hopkins, D.L. (1989). *Xylella fastidiosa*: a xylem-limited bacterial pathogen of plants. *Annu. Rev. Phytopathol.* **27**: 271-290.

Hopper, A.K., Phizicky, E.M. (2003). tRNA transfers to the limelight. *Genes Dev*, **17**: 162-180.

Iwobi, A., Heesemann, J., Garcia, E., Igwe, E., Noelting, C., Rakin, A. (2003). Novel virulence-associated type II secretion system unique to high-pathogenicity *Yersinia enterocolitica*. *Infect Immun*, **71**: 1872-1879.

Kelly, S. M., Price, N. C. (1997). The application of circular dichroism to studies of protein folding and unfolding. *Biochimica et Biophysica Acta*, **1338**: 161-185.

Kelly, S. M., Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, **1**: 349-384.

Klootwijk, J., Klein, I., Grivell, L.A. (1975). Minimal post-transcriptional modification of yeast mitochondrial ribosomal RNA. *J Mol. Biol.*, **97**: 337-350.

Köhler, R., Schafer, K., Muller, S., Vignon, G., Diederichs, K., Philippsen, A., Ringer, P., Pugsley, A.P., Engel, A., Welte, W. (2004). Structure and assembly of the pseudopilin PulG. *Mol Microbiol*, **54**: 647-664.

Kozbial, P.Z, Mushegian, A.R. (2005). Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol*, **14**:5-19.

Kyogoku. Y, Fujiyoshi. Y, Shimada. I, Nakamura. H, Tsukihara. T, Akutsu. H, Odahara. T, Okada. T, Nomura. N. (2003). Structural genomics of membrane proteins. *Acc Chem Res*, **36**(3):199-206.

Lafontaine, D.L.J., Preiss, T., Tollervey, D. (1998). Yeast 18S dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Mol. Cell. Biol.*, **18**: 2360-2370.

Lambais, M.R, Goldman, M.H.S, Camargo, L.E.A, Goldman, G.H. (2000). A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity. *Current Opinion in Microbiology*, **3**: 459-462.

LaPointe, C.F., Taylor, R.K. (2000). The type IV prepilin peptidases comprise a novel family of aspartic acid proteases. *J Biol Chem*, **275**: 1502-1510.

Lathem, W.W., Gryns, T.E., Witowski, S.E., Torres, A.G., Kaper, J.B., Tarr, P.I., Welch, R.A. (2002). StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol Microbiol*, **45**: 277-288.

Lee, H.M, Tyan, S.W, Leu, W.M, Chen, L.Y, Chen, D.C, HU, N.T. (2001). Involvement of the XpsN protein in formation of the XpsL-XpsM complex in *Xanthomonas campestris* pv. *campestris* type II secretion apparatus. *Journal of Bacteriology*, **183**(2): 528-535.

Lee, H.M., Chen, J.R., Lee, H.L., Leu, W.M., Chen, L.Y., Hu, N.T. (2004). Functional dissection of the XpsN (GspC) protein of the *Xanthomonas campestris* pv. *campestris* type II secretion machinery. *J Bacteriol*, **186**: 2946-2955.

Lesk, A.M. (1995) NAD- binding of dehydrogenases. *Curr Opin Struct Biol*, **5**: 775-783

Lévêque, F., Blanchin-Roland, S., Fayat, G., Plateau, P., Blanquet, S. (1990) Design and characterization of *Escherichia coli* mutants of Ap₄N-hydrolase activity. *J. Mol. Biol.*, **212**: 319-329.

Lindeberg, M., Salmond, G.P., Collmer, A. (1996). Complementation of deletion mutations in a cloned functional cluster of *Erwinia chrysanthemi* out genes with *Erwinia carotovora* out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway. *Mol Microbiol*, **20**: 175-190.

Machado, M.A. Xylella fastidiosa and the Citros Variegated Chlorosis (CVC), a new destructive citros disease in Brazil. Disponível em: <http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>, acessado em: 05 de Agosto de 2003.

Mecsas J., Bólis, I., Falkow S. (2001). Identification of attenuated *Yersinia pseudotuberculosis* strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. *Infect. Immun.*, **69**: 2779-2787.

Michel, G., Bleves, S., Ball, G., Lazdunski, A., Filloux, A. (1998). Mutual stabilization of the XcpZ and XcpY components of the secretory apparatus in *Pseudomonas aeruginosa*. *Microbiol*, **144**: 3379-3386.

Nelson, M., Raschke, E., McClelland, M. (1993). Effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res*, **21(13)**: 3139-3154.

Noon, K.R., Bruenger, E., McCloskey, J.A. (1998). Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J. Bacteriol*, **180**: 2883-2888.

Noyer-Weidner, M., Trautner, T.A. (1993). Methylation of DNA in prokaryotes. *EXS*, **64**: 39-108

Nunn, D. (1999). Bacterial type II protein export and pilus biogenesis: more than just homologies? *Trends Cell Biol*, **9**: 402-408.

Nunn, D.N., Lory, S. (1993). Cleavage, methylation, and localization of the *Pseudomonas aeruginosa* export proteins XcpT, XcpU, XcpV, and XcpW. *J Bacteriol*, **175**: 4375-4382.

O'Farrell, H.C., Musayev, F.N., Scarsdale, J.N., Wright, H.T., Rife, J.P. (2003). Crystallization and preliminary X-ray diffraction analysis of KsgA, a universally conserved RNA adenine dimethyltransferase in *Escherichia coli*. *Acta Crystallog.sect. D*, **59**: 1490-1492.

O'Farrell, H.C.; Scarsdale, J.N.; Rife, J.P. (2004). Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli*. *J.Mol.Biol*, **339**(2): 337-53.

O'Farrell, H.C, Pulicherla, N, Desai, P.M, Rife, J.P. (2006). Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA*, **12**(5):725-733.

Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Letters*, **514**: 17-25.

Palmer, T., Sargent, F., Berks, B.C. (2005). Export of complex cofactor-containing proteins by the bacterial Tat pathway. *Trends Microbiol*, **13**: 175-180.

Pelton, J. T., McLean, L. R. (2000). Spectroscopic Methods for Analysis of Protein Secondary Structure. *Analytical Biochemistry*, **277**: 167-176.

Planet, P.J., Kachlany, S.C., DeSalle, R., Figurski, D.H. (2001). Phylogeny of genes for secretion NTPases: identification of the widespread *tadA* subfamily and development of a diagnostic key for gene classification. *Proc Natl Acad Sci USA*, **98**: 2503-2508.

Poldermans, B., van Buul, C.P.J.J., van Knippenberg, P.H. (1979). Studies on the function of two adjacent N⁶, N⁶-dimethyladenosines near the 3' end of 16S ribosomal RNA of *Escherichia coli*. II. The effect of the absence of the methyl groups on initiation of protein biosynthesis. *J. Biol. Chem.*, **254**: 9090-9094.

Possot, O.M., Gerard-Vincent, M., Pugsley, A.P. (1999). Membrane association and multimerization of secreton component PulC. *J Bacteriol*, **181**: 4004-4011.

Possot, O.M., Vignon, G., Bomchil, N., Ebel, F., Pugsley, A.P. (2000). Multiple interactions between pullulanase secreton components involved in stabilization and cytoplasmic membrane association of PulE. *J Bacteriol*, **182**: 2142-2152.

Py, B., Louseau, L., Barras, F. (1999). Assembly of the type II secretion machinery of *Erwinia chrysanthemi*: direct interaction and associated conformational change between OutE, the putative ATP-binding component and the membrane protein OutL. *J Mol Biol*, **289**: 659-670.

Rao, S.T, Rossmann, M.G. (1973). Comparison of super-secondary structures in proteins. *J Mol Biol*. **76(2)**:241-256.

Reeves, P.J., Whitcombe, D., Wharam, S., Gibson, M., Allison, G., Bunce, N., Barallon, R., Douglas, P., Mulholland, V., Stevens, S. (1993). Molecular cloning and characterization of 13 out genes from *Erwinia carotovora* subspecies *carotovora*: genes encoding members of a general secretion pathway (GPS) widespread in gram-negative bacteria. *Mol Microbiol*, **8**: 443-456.

- Roberto, S.R, Coutinho, A, Delima, J.E.O, Miranda, V.S, Carlos, E.F. (1996). Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Dilobopterus costalimai*, *Ocrogonia terminalis* e *Oncometopia facielis* em citros. *Fitopatologia Brasileira*. **21**: 517-518.
- Rossier, O., Starkenburg, S.R., Ciancionotto, N.P. (2004). *Legionella pneumophila* type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia. *Infect Immun*, **72**: 310-321.
- Russel, M. (1998). Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *J. Mol. Biol.*, **279**: 485-499.
- Sandkvist, M. (2001). Biology of type II secretion. *Mol. Microbiol*, **40**: 271-283.
- Sandkvist, M., Bagdasarian, M., Howard, S.P., Dirita, V.J. (1995). Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in *Vibrio cholerae*. *EMBO J*, **14**: 1664-1673.
- Sandkvist, M., Hough, L.P., Bagdasarian, M.M., Bagdasarian, M. (1999). Direct interaction of the EpsL and EpsM proteins of the general secretion apparatus in *Vibrio cholerae*. *J Bacteriol*, **181**: 3129-3135.
- Sandkvist, M., Keith, J.M., Bagdasarian, M., Howard, S.P. (2000). Two regions of EpsL involved in species-specific protein-protein interactions with EpsE and EpsM of the general secretion pathway in *Vibrio cholerae*. *J Bacteriol*, **182**: 742-748.
- Sandkvist, M., Michel, L.O., Hough, L.P., Morales, V.M., Bagdasarian, M., Koomey, M., Dirita, V.J., Bagdasarian, M. (1997). General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*. *J Bacteriol*, **179**: 6994-7003.

Sankpal, U.T., Rao, D.N. (2002). Mutational analysis of conserved residues in HhaI DNA methyltransferase. *Nucleic Acids Res*, **30**: 2628-2638.

Sauvonnnet, N., Vignon, G., Pugsley, A.P., Gounon, P. (2000). Pilus formation and protein secretion by the same machinery in *Escherichia coli*. *EMBO J*, **19**: 2221-2228.

Simpson, A.J.G. et al. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*. **406**: 151-157.

Steege, D.A., Graves, M.C., Spremulli, L.L. (1982). *Euglena gracilis* chloroplast small subunit rRNA. *J. Biol. Chem.*, **257**: 10430-10439.

Teixeira, D.C, Rocha, S.R.P, Santos, M.A, Mariano, A. G, Li, W.B, Monteiro, P.B. (2004). A suitable *Xylella fastidiosa* CVC strain for post-genome studies. *Current Microbiology*. **49**: 396-399.

Thammana, P., Held, W. (1974) Methylation of 16S RNA during ribosome assembly in vitro. *Nature*, **251**: 682-686.

Tsai, R.T, Leu, W.M, Chen, L.Y, Hu, N.T. (2002). A reversibly dissociable ternary complex formed by XpsL, XpsM and XpsN of the *Xanthomonas campestris* pv. *campestris* type II secretion apparatus. *Biochem. J*, **367**: 865-871.

van Bull, C.P.J.J., Hamersma, M., Visser, W., van Knippenberg, P.H. (1984). Partial methylation of two adjacent adenosines in ribosomes from *Euglena gracilis* chloroplasts suggests evolutionary loss of an intermediate stage in the methyl-transfer reaction. *Nucl. Acids Res*, **12**: 9205-9208.

van Knippenberg, P.H. (1986). Structural and functional aspects of the N⁶, N⁶ dimethyladenosines in 16S ribosomal RNA. In Structure, Function and Genetics of Ribosomes (Hardesty, B., Kremaer, G., eds). pp. 412-424, Springer, Berlin

Van Knippenberg, P.H., van Kimmenade, J.M.A., Heus, H.A. (1984) Phylogeny of the conserved 3' terminal structure of the RNA of small ribosomal subunits. *Nucl.Acids Res.*, **12**: 2595-2604.

Voulhoux, R., Ball, G., Ize, B., Vasil, M.L., Lazdunski, A., Wu, L.F., Filloux, A. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J*, **20**: 6735-6741.

Yeo, H.J., Savvides, S.N., Herr, A.B., Lanka, E., Waksman, G. (2000). Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol Cell*, **6**(6):1461-72.

DECLARAÇÃO

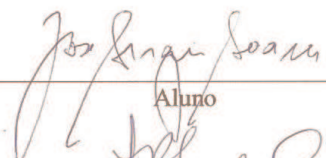
Declaro para os devidos fins que o conteúdo de minha dissertação/ tese de mestrado/doutorado intitulada Modelagem Molecular e Caracterização Espectroscópica de duas Proteínas de *Xylella fastidiosa* Potencialmente Envolvidas com Patogenicidade:

() não se enquadra no Artigo 1º, § 3º da Informação CCPG 002/06, referente a bioética e biossegurança.

(x) está inserido no Projeto CIBio (Protocolo nº 15/2006), intitulado Determinação de Estruturas Tridimensionais de Proteínas Relacionadas a Patogenicidade de *Xylella fastidiosa*.

() tem autorização da Comissão de Ética em Experimentação Animal (Protocolo nº _____).

() tem autorização do Comitê de Ética para Pesquisa com Seres Humanos (?) (Protocolo nº _____).



Aluno



Orientador

Para uso da Comissão ou Comitê pertinente:

() Deferido () Indeferido

Nome:
Função: