

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA



Luciana Kauer Rosselli

"CARACTERIZAÇÃO DE UMA ALDO-CETO REDUTASE RELACIONADA À
PATOGENICIDADE DE *Xylella fastidiosa*"

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
Luciana Kauer Rosselli
[Assinatura]
e aprovada pela Comissão Julgadora.

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

Orientadora: Profa. Dra. Anete Pereira de Souza

Co-orientador: Prof. Dr. Francisco Javier Martin Medrano

Campinas
2007

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

R734c

Rosselli, Luciana Kauer

Caracterização de uma aldo-ceto redutase relacionada à
patogenicidade de *Xylella fastidiosa* / Luciana Kauer
Rosselli. – Campinas, SP: [s.n.], 2007.

Orientadores: Anete Pereira de Souza, Francisco Javier
Martin Medrano.

Tese (doutorado) – Universidade Estadual de
Campinas, Instituto de Biologia.

1. Aldo-ceto redutase. 2. Proteína AKR13B1. 3.
Xylella fastidiosa. 4. Gene XF1729. I. Souza, Anete
Pereira de. II. Medrano, Francisco Javier Martin. III.
Universidade Estadual de Campinas. Instituto de Biologia.
IV. Título.

(rcdt/ib)

Título em inglês: Characterization of a pathogenicity related Aldo-keto reductase from *Xylella fastidiosa*.

Palavras-chave em inglês: Aldo-keto reductase; AKR13B1 protein; *Xylella fastidiosa*; XF1729 Gene.

Área de concentração: Genética de Microorganismos.

Titulação: Doutora em Genética e Biologia Molecular.

Banca examinadora: Anete Pereira de Souza, Jörg Kobarg, Marco Aurélio Takita, Alessandra Alves de Souza, Ana Carolina Zeri.

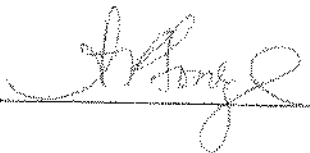
Data da defesa: 26/04/2007.

Programa de Pós-Graduação: Genética e Biologia Molecular.

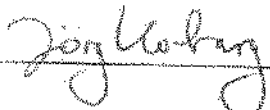
Campinas, 26 de abril de 2007.

Banca Examinadora

Profa. Dra. Anete Pereira de Souza (Orientadora)



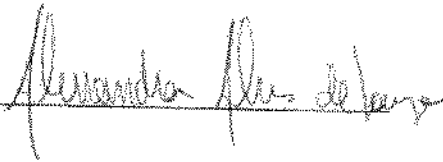
Prof. Dr. Jörg Kobarg



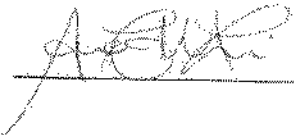
Prof. Dr. Marco Aurélio Takita



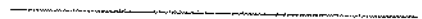
Profa. Dra. Alessandra Alves de Souza



Profa. Dra. Ana Carolina Zeri



Prof. Dr. José Camillo Novello



Profa. Dra. Maria Isabel Nogueira Cano



Prof. Dr. Marcelo Menossi Teixeira



Levando com paciência todos os sofrimentos da vida; submetendo, com o auxílio de Deus, nossos baixos instintos ao império da razão, podemos seguramente confiar que, se houvermos combatido o bom combate, consumado a carreira, e conservado a fé, como o fez o Apóstolo: Nosso Senhor, que é juiz justo, dar-nos-á naquele dia a coroa da justiça, que também nos está preparada.

*Catecismo de Trento
Parte segunda, cap. II, p. 212, 1962, Ed. Vozes
(II Timóteo IV, 7)*

OFEREÇO

Ad Deum qui laetificat juventutem meam.

DEDICO

Ao meu marido *Marcelo*,
A cada um de nossos filhos *Gabriel, André,*
Isabel, Maria Teresa e Maria Luisa.

AGRADEÇO

Aos meus pais *Pierre* e *Maria Ivone*, por me transmitirem a vida.

Ao meu marido *Marcelo*, pela fiel e inestimável amizade no cumprimento de nossos deveres
como pais e profissionais.

À professora *Anete*, pela confiança e excelente orientação.

Aos colegas do grupo de pesquisa, pelo produtivo trabalho em equipe.

Aos meus amigos do grupo São Nicolau de Flüe, pela verdadeira amizade.

Aos pesquisadores do LNLS, pelas colaborações realizadas.

À *Maria Zilda*, *Juverlande*, *Tânia* e *Sandra*, pelo apoio técnico de excelente qualidade.

À FAPESP, pelo auxílio financeiro concedido.

ÍNDICE

Prefácio	xii
Resumo	xiv
Abstract	xvi
1. Introdução	01
1.1. A <i>Xylella fastidiosa</i> e a Clorose Variegada dos Citros (CVC)	03
1.2. Os Mecanismos de Patogenicidade de <i>X. fastidiosa</i>	06
1.3. A <i>ORF</i> XF1729 de <i>X. fastidiosa</i>	10
1.4. As Aldo-ceto Redutases nos Microorganismos	11
2. Objetivos	15
2.1. Objetivos Gerais	15
2.2. Objetivos Específicos	15
3. Metodologia	16
3.1. Caracterização Enzimática	16
3.2. Técnicas Utilizadas na Caracterização Estrutural de Proteínas	21
3.2.1. Dicroísmo Circular	21
3.2.2. Espalhamento de Raios-X a Baixo Ângulo	24
4. Manuscrito	27
5. Resultados Complementares	36
5.1. Testes de Cristalização da Proteína XF1729	36
5.2. Ensaio de Expressão e Purificação da proteína codificada pela <i>ORF</i> XF1934	38
5.3. Ensaio de Expressão da proteína codificada pela <i>ORF</i> XF1532	41
6. Perspectivas	44

6.1. Proteína XF1729 (AKR13B1)	44
6.2. Proteína XF1934	44
6.3. Proteína XF1532	45
7. Conclusões	46
8. Referências Bibliográficas	49
9. Anexos	55
9.1. Anexo I	56
9.2. Anexo II	64

ÍNDICE DE FIGURAS

Figura 1 – Sintomas da Clorose Variegada dos Citros	04
Figura 2 – Colonização da bactéria <i>Xylella fastidiosa</i> nos vasos do xilema da planta	05
Figura 3 – Gráfico de Michelis-Menten	18
Figura 4A – Polarização elíptica da luz	23
Figura 4B – Representação dos espectros de CD característicos	23
Figura 5 – Esferolitas da proteína XF1729 encontradas no tampão 1,6 Hexenodiol pHs 7,5; 8,0; 8,5 e 9,0, respectivamente	36
Figura 6 – Fotos dos cristais obtidos nos ensaios cristalográficos da proteína XF1729 complexada ao cofator NADPH	38
Figura 7 – Gel de SDS-PAGE 12% apresentando os resultados da expressão e purificação de 500 ml de cultura da proteína heteróloga XF1934 (~40 kDa)	39
Figura 8 – Gel SDS-PAGE 12% apresentando os resultados do ensaio de hidrólise da proteína XF1934 utilizando fator Xa, a temperatura ambiente	40
Figura 9 – Gel SDS-PAGE 12% apresentando o padrão de expressão da proteína heteróloga XF1532 nas frações solúvel e insolúvel. A linhagem testada foi BL21(DE3) a 37°C e 1mM de IPTG	42
Figura 10 – Gel SDS-PAGE 12% apresentando o padrão de expressão da proteína heteróloga XF1532 nas frações solúvel e insolúvel. A linhagem testada foi BL21(DE3)pLysS a 37°C e 1mM de IPTG	43
Figura 11 – O motivo ($\alpha\beta$) ₈ -Barrel das aldo-ceto redutases (AKRs)	47

ÍNDICE DE EQUAÇÕES

Equação 1 – Equação modelo de Michaelis-Menten para enzimas	17
Equação 2 – Equação de Michaelis-Menten	19
Equação 3 – Equação derivada da equação de Michaelis-Menten	19

PREFÁCIO

Os resultados obtidos durante o desenvolvimento deste trabalho de tese serão apresentados na forma de artigos científicos. O principal artigo deste trabalho, intitulado “A new member of the aldo-keto reductase family from the plant pathogen *Xylella fastidiosa*”, publicado na revista *Archives of Biochemistry and Biophysics* 453 (2006) 143-150, descreve a caracterização enzimática e estrutural a baixa resolução da proteína codificada pela ORF XF1729 de *X. fastidiosa*.

Como resultados complementares, apresentamos os estudos de expressão e purificação das proteínas codificadas pelas ORFs XF1532 e XF1934 envolvidas na patogenicidade de *X. fastidiosa*, além dos ensaios cristalográficos realizados para a proteína XF1729.

Em anexo, podem ser verificados dois trabalhos realizados por nosso grupo de pesquisa, caracterizando duas outras proteínas de *X. fastidiosa*. O primeiro artigo, “Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*” publicado na revista *Protein Expression & Purification* 33 (2004) 297-303, apresenta os estudos de expressão, purificação e caracterização da atividade de uma nova chaperonina, HSP17.9. O segundo artigo, “Cloning, expression and purification of the virulence protein D from *Xylella fastidiosa*” também publicado na revista *Protein Expression & Purification* 37 (2004) 320-326, descreve os estudos de expressão, purificação e caracterização da estrutura secundária da proteína VapD, a qual está relacionada à virulência de diversos patógenos e, no entanto, sua função permanece desconhecida.

Gostaríamos ainda de mencionar a participação no curso “*Biopolymer in materials and Life Sciences*” realizado na Universidade de Potsdam, Alemanha, no período de 11 a 29 de setembro de 2006. O conteúdo do curso, além de fazer parte do escopo do tema da tese,

enriqueceu consideravelmente minha formação na área de bioquímica de proteínas. Durante o curso, inúmeros tópicos referentes à produção e à purificação de proteínas foram abordados permitindo um aprofundamento no assunto.

RESUMO

O programa de seqüenciamento genômico da bactéria *Xylella fastidiosa* gerou um enorme número de *ORFs* ("Open Reading Frames" – quadro de leitura aberto ou genes putativos) pertencentes às categorias de patogenicidade, virulência e adaptação deste importante fitopatógeno. Uma destas *ORFs* (XF1729) foi anotada como sendo uma fenilacetaldéido desidrogenase de 31,4 kDa. No entanto, a análise de sua seqüência primária de aminoácidos revelou similaridades com proteínas pertencentes à superfamília de aldo-ceto redutases. As proteínas desta superfamília são oxidoredutases NADPH-dependentes relacionadas funcionalmente e estruturalmente. Neste trabalho, a seqüência similar de *X. fastidiosa* foi clonada no vetor de expressão pET32Xa/LIC com o objetivo de super expressar a proteína recombinante fusionada a seis resíduos de histidina em *Escherichia coli* BL21(DE3). A proteína expressa na fração solúvel foi purificada por cromatografia de afinidade em metal imobilizado (resina Agarose-IDA-Ni). O conteúdo de sua estrutura secundária foi verificado por espectroscopia de dicroísmo circular. As medidas de espalhamento de raios-X a baixo ângulo (SAXS) forneceram os parâmetros estruturais gerais (raio de giro de $27,5 \pm 0,8$ Å e máxima dimensão de 90 Å) e indicaram que a proteína apresenta-se como um monômero em solução. Além disto, os cálculos estruturais *ab initio* mostraram que a proteína apresenta algumas similaridades com uma aldo-ceto redutase previamente cristalizada. A proteína XF1729 purificada foi capaz de catalisar a redução dos substratos DL-gliceraldeído (K_{cat} $2.26 s^{-1}$, K_m 8.20 ± 0.98 mM) e 2-nitrobenzaldeído (K_{cat} $11.74 s^{-1}$, K_m 0.14 ± 0.04 mM) na presença de NADPH. A seqüência de aminoácidos da proteína XF1729 apresentou mais alta identidade (maior que 40%) com inúmeras proteínas de função desconhecida. Entre as *AKRs* identificadas, encontramos aproximadamente 29% de identidade com YakC (AKR13) de *Schizosaccharomyces*

pombe, 30% e 28% com AKR11A e AKR11B, ambas de *Bacillus subtilis*, respectivamente. Os resultados estabeleceram a proteína XF1729 como um novo membro da superfamília das AKRs, da nova sub-família AKR13B1. Finalmente, os experimentos de caracterização por cromatografia de gel filtração indicaram que a proteína apresenta uma forma alongada, gerando um peso molecular aparente maior que o esperado.

Além da proteína AKR13B1, selecionamos mais duas proteínas codificadas pelas ORFs XF1934 e XF1532 para estudos de estrutura e função. A proteína HetI (XF1934), de 22,4 kDa, apresenta similaridade com proteínas da família phosphopantetheinyl transferase, sendo necessária à síntese de ácido graxo via acetato na bactéria. A proteína XF1532, de 36,8 kDa, é similar a reguladores transcricionais de estresse oxidativo, sendo sua sequência similar (44%) à proteína *OxyR* de *Escherichia coli*, a qual pertence à família LysR de reguladores transcricionais. As duas ORFs foram clonadas e expressas em *Escherichia coli*, no entanto não foi possível obter a proteína na fração solúvel, impossibilitando o seguimento dos estudos de caracterização estrutural e funcional. A descrição da metodologia utilizada para produção destas proteínas e os resultados preliminares obtidos estão descritos no item Resultados Complementares desta tese.

Um amplo benefício no conhecimento dos mecanismos de patogenicidade da *X. fastidiosa* tem sido esperado, desde o seqüenciamento completo de seu genoma. Neste sentido, passos iniciais foram dados no sentido de se caracterizar a função das proteínas codificadas por seus genes.

ABSTRACT

The *Xylella fastidiosa* genome program generated a large number of gene sequences that belong to pathogenicity, virulence and adaptation categories from this important plant pathogen. One of these genes (XF1729) was described in the genome annotation as being a phenylacetaldehyde dehydrogenase. However, the XF1729 primary sequence analysis showed similarities to Aldo-keto reductase superfamily proteins. The AKRs are NADPH-dependent oxidoreductases structurally and functionally related. In this work, the similar sequence XF1729 from *Xylella fastidiosa* was cloned onto the pET32Xa/LIC vector in order to over-express a recombinant His-Tag fusion protein in *E. coli* BL21(DE3). The expressed protein in the soluble fraction was purified by immobilized metal affinity chromatography (agarose-IDA-Ni resin). Secondary structure contents were verified by circular dichroism spectroscopy. Small-Angle X-Ray Scattering (SAXS) measurements furnish general structural parameters (the particle radius of gyration of $27.5 \pm 0.8 \text{ \AA}$ and the particle maximum dimension of $\sim 90 \text{ \AA}$) and provide a strong indication that the protein has a monomeric form in solution. Also, *ab initio* calculations show that the protein has some similarities with a previously crystallized aldo-keto reductase protein. The recombinant XF1729 purified to homogeneity catalyzed the reduction of DL-glyceraldehyde (Kcat 2.26 s^{-1} , Km $8.20 \pm 0.98 \text{ mM}$) and 2-nitrobenzaldehyde (Kcat 11.74 s^{-1} , Km $0.14 \pm 0.04 \text{ mM}$) in the presence of NADPH. The amino acid sequence deduced from XF1729 showed the highest identity (40% or higher) with several functionally unknown proteins. Among the identified AKRs, we found approximately 29% of identity with YakC (AKR13) from *Schizosaccharomyces pombe*, 30% and 28% with AKR11A and AKR11B, both from *Bacillus subtilis*, respectively. The results establish XF1729 as the new member of AKR family, AKR13B1. Finally, the first characterization by gel filtration chromatography assays indicates

that the protein has an elongated shape, which generates an apparent higher molecular weight. Since *Xylella fastidiosa* 9a5c strain (associated with CVC) is the first plant pathogen to be fully sequenced, a large benefit for the whole field of disease research can be expected. An initial step has been taken towards the characterization the protein function encoded by its genes.

1. INTRODUÇÃO

A bactéria *Xylella fastidiosa* é o agente causador de diversas doenças em plantas. No Brasil é responsável pela doença denominada Clorose Variegada dos Citros (CVC) (Lee *et al.*, 1991), a qual prejudica uma das culturas mais importantes economicamente em nosso país, a da laranja. Tendo em vista o enorme prejuízo causado pela CVC à citricultura brasileira e, com o intuito de capacitar e qualificar pesquisadores brasileiros, foi criada pela FAPESP em 1997 a Organização para Seqüenciamento e Análise de Nucleotídeos (ONSA), uma rede de laboratórios responsável pelo seqüenciamento completo do genoma da *X. fastidiosa*. Sendo o primeiro fitopatógeno a ter seu genoma completamente seqüenciado (Simpson *et al.*, 2000), tornou-se alvo da genômica funcional e estrutural na busca pela compreensão de seu mecanismo de infecção e adaptação nos hospedeiros, através da caracterização das proteínas codificadas pelos seus genes.

A estrutura e função de uma proteína estão intimamente correlacionadas. O conhecimento da estrutura protéica é uma valiosa informação para a determinação tanto da função protéica, quanto da relação existente entre ambas. Deste conceito deriva a grande importância dos estudos para a determinação da estrutura tridimensional de proteínas. No entanto, sabe-se que apenas o conhecimento estrutural é insuficiente para deduzir a função global da proteína de interesse na célula, fazendo-se necessário combinar estudos bioquímicos (*in vitro*) e celulares (*in vivo*) com informações estruturais para completar a compreensão global do mecanismo celular protéico (Chothia & Lesk, 1986).

Para se caracterizar a estrutura de uma proteína ao nível atômico, há dois métodos complementares – a Cristalografia de raios-X e a Espectroscopia de Ressonância Magnética Nuclear (RMN), além do Espalhamento de raios-X a baixo ângulo (SAXS) que permite o

conhecimento da estrutura tridimensional protéica a baixa resolução. Porém, ao contrário dos outros projetos genoma, em que as informações são geradas em alta velocidade e precisão, o genoma estrutural é muito mais minucioso e laborioso. Para se chegar à resolução da estrutura tridimensional de uma proteína faz-se necessário atravessar inúmeros gargalos durante o percurso, que vão desde a clonagem do gene alvo até propriamente a resolução da estrutura tridimensional da proteína codificada.

Neste contexto, baseando-se nos dados gerados pelo seqüenciamento genômico da *X. fastidiosa* (linhagem 9a5c) e na predição da função dada por similaridade a outros genes conhecidos, nosso grupo de pesquisa selecionou um conjunto de genes potencialmente envolvidos no processo de patogenicidade desta bactéria e, dentre elas a *ORF XF1729*, objeto de estudo desta tese de doutorado. O grupo de pesquisa vinculado ao programa da Rede de Biologia Molecular Estrutural da FAPESP (SMOLBnet) tinha por objetivo o estudo estrutural e funcional de tais proteínas. Desde o início do projeto muito dos objetivos propostos foram alcançados. Um laboratório para produção e purificação das proteínas recombinantes destinadas aos estudos de caracterização estrutural e funcional foi implantado, permitindo aos componentes do grupo o desenvolvimento deste tipo de estudo com segurança e velocidade. Além disso, o grupo apresentou um contínuo e progressivo avanço tendo estabelecido os protocolos de obtenção de quatro proteínas de *Xylella fastidiosa* destinadas a ensaios estruturais e funcionais, como devidamente descrito nos trabalhos publicados (Paula *et al.*, 2004; Azzoni *et al.*, 2004; Catani *et al.*, 2004; Rosselli *et al.*, 2006).

A *ORF XF1729* selecionada para estudo codifica uma proteína de 291 resíduos de aminoácidos, com um peso molecular predito de 31.4kDa e um pI teórico de 4.99. De acordo com a anotação genômica, o produto deste gene codifica uma fenilacetaldéido desidrogenase pertencente à categoria de produção de toxina e detoxificação (classificação atribuída na

anotação do projeto genoma da *Xylella fastidiosa*). No entanto, o alinhamento comparativo de sua sequência de aminoácidos revelou alta similaridade a oxidoredutases encontradas em diversas espécies. Além disso, o alinhamento também revelou a presença de um domínio comum com as proteínas pertencentes à superfamília das aldo-ceto redutases (*AKR superfamily*). Diante destas informações, iniciou-se o trabalho de caracterização desta *ORF*.

Apesar do grande número de informações geradas pelo seqüenciamento completo de seu genoma, as proteínas de *X. fastidiosa* são pobremente caracterizadas através de estudos bioquímicos, estruturais e funcionais. Desta forma, este trabalho de doutorado pretende apresentar os estudos de expressão e purificação protéica, bem como de caracterização funcional e estrutural da proteína codificada pela *ORF* XF1729 de *X. fastidiosa*.

1.1. A *XYLELLA FASTIDIOSA* E A CLOROSE VARIEGADA DOS CITROS (CVC)

A *Xylella fastidiosa* é uma bactéria gram-negativa que vive limitada aos vasos do xilema de seus hospedeiros. Esta bactéria, com formato de bastonete, sem flagelo e de tamanho variável, pode chegar até 0,7 µm de diâmetro e 3 µm de comprimento (Wells *et al.*, 1987). Em meio de cultura, também caracteriza-se pelo crescimento lento, com colônias circulares, discretas, medindo até 0,6 mm de diâmetro após 10 dias de incubação a 28°C, podendo alcançar 1,5 mm depois de 30 dias, com variações decorrentes do meio de cultura utilizado (Coletta Filho & Machado, 2002).

A transmissão da *X. fastidiosa* se dá através de cigarrinhas sugadoras de xilema, pertencentes à família Cicadellidae (Purcell *et al.*, 1979), podendo também ser transmitida através do uso de porta enxertos contaminados (He *et al.*, 2000). A bactéria sobrevive no lúmen

do canal alimentar das cigarrinhas transmissoras e nos vasos do xilema da planta, fechando o ciclo infectante após a sua alimentação (Hopkins *et al.*, 1995).

A CVC é uma doença que ataca todas as variedades comerciais de laranja doce (Carvalho *et al.*, 1996). Os principais sintomas da CVC em plantas de laranja doce são o aparecimento de manchas cloróticas na parte dorsal das folhas correspondendo a pontos marrons com aspecto de goma na superfície ventral. As plantas severamente atacadas apresentam um aspecto de debilidade geral com ramos envassourados, uma superbrotção das gemas e folhas murchas nas horas mais quentes do dia, mesmo em condições de umidade. Posteriormente, verifica-se uma drástica redução no tamanho dos frutos, tornando-os enrijecidos e impróprios para consumo (Hopkins, 1989) (Fig. 1).

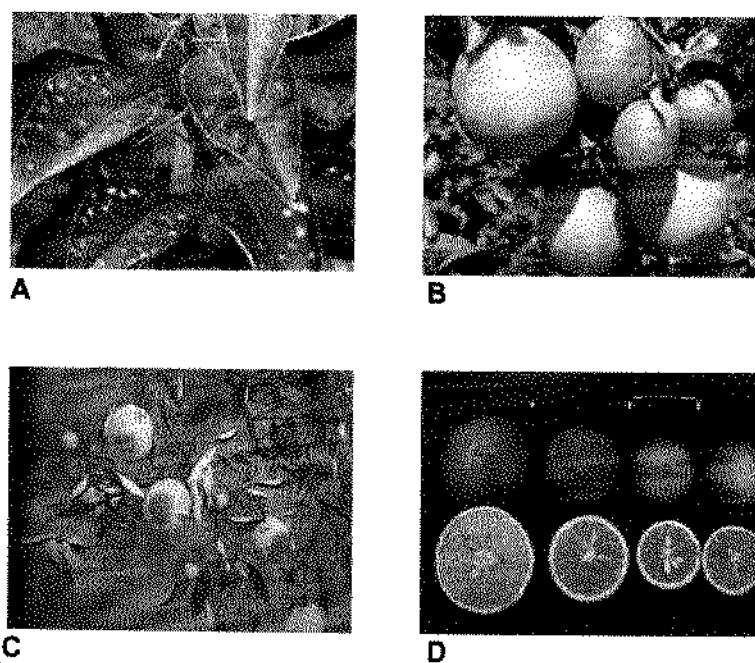


Figura 1: Sintomas da Clorose Variegada dos Citros. **A)** Manchas cloróticas na superfície das folhas. **B)** Frutos sadios ao lado de frutos doentes. **C)** Sintoma de murcha em folhas. **D)** Redução do tamanho dos frutos. (Fonte: www.fundecitrus.com.br)

Esses sintomas indicam a ocorrência da oclusão dos vasos do xilema da planta, ocasionando principalmente o bloqueio do transporte de água e nutrientes. Neste caso, observa-se uma eficiente capacidade de aderência e multiplicação da *X. fastidiosa* nos vasos condutores de seiva bruta, como apresentado na Figura 2. (Machado *et al.*, 1994; McElrone *et al.*, 2001; Machado *et al.*, 2001).

A bactéria depois de instalada no xilema da planta (Fig. 2), passa a se multiplicar e a produzir substâncias extracelulares. Entre estas substâncias, uma grande quantidade de polissacarídeos extracelulares é secretada, os quais contribuem para a formação de um biofilme na superfície dos vasos do xilema. Sobre esse material depositam-se mais bactérias que resultam em grandes agregados que podem interromper o fluxo regular de água e nutrientes através destes vasos. Esses agregados celulares podem, em tese, contribuir para concentrar atividade de exoenzimas bacterianas, auxiliar na captura de nutrientes, acumular substâncias tóxicas, manter a agregação de células e proteger contra as flutuações ambientais (Michelmore, 2000).

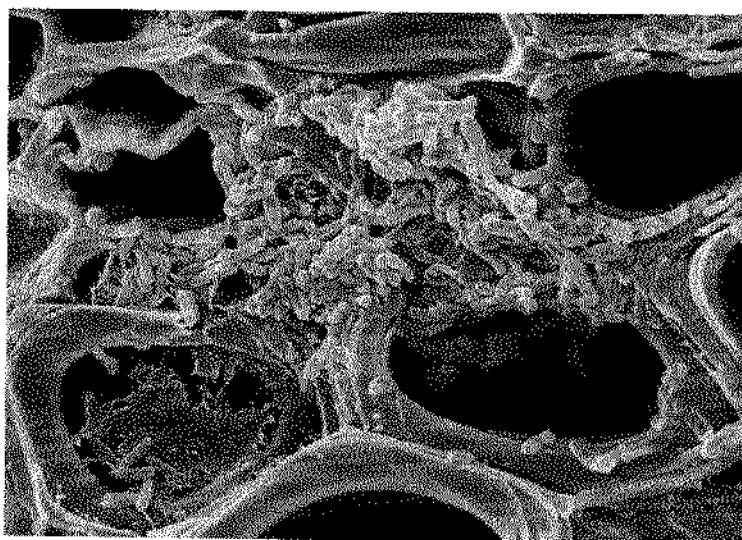


Figura 2: Colonização da bactéria *Xylella fastidiosa* nos vasos do xilema da planta. (Fonte: <http://aeg.lbi.ic.unicamp.br/xf/>)

A *X. fastidiosa*, além de causar doença em citros, também ataca diversos hospedeiros como: alfafa e videira (Goheen *et al.*, 1973); pessegueiro (Hopkins *et al.*, 1973; Nyland *et al.*, 1973); amendoeira (Mircetich *et al.*, 1976); ameixeira (Hopkins, 1989); e cafeeiros (Paradela Filho *et al.*, 1997). No Brasil, a CVC foi detectada primeiramente no ano de 1987 e, dez anos depois, já comprometia cerca de 90% dos pomares em idade produtiva nas regiões do Estado de São Paulo e do sul do triângulo mineiro. Atualmente cerca de 45% dos pomares encontram-se plantas com sintomas de CVC (www.fundecitrus.com.br). Os principais danos são decorrentes do não aproveitamento comercial dos frutos que, por apresentarem tamanho reduzido, não são aceitos para o consumo de mesa e oferece baixa quantidade de suco, principal produto citrícola brasileiro de exportação. A CVC possui especial importância no Brasil visto o peso relegado à indústria cítrica. Em 2000, no estado de São Paulo, 35% de 200 milhões de laranjeiras apresentavam sintomas de CVC, representando uma perda direta de mais de 100 milhões de dólares (www.fundecitrus.com.br).

1.2. OS MECANISMOS DE PATOGENICIDADE DE *X. FASTIDIOSA*

A mencionada importância da indústria cítrica no país, bem como os danos causados pela CVC nos pomares brasileiros, aliados à possibilidade de se desenvolver novas estratégias de controle baseadas no conhecimento dos mecanismos de patogenicidade, resultaram no desenvolvimento de um extenso programa de pesquisa sobre *X. fastidiosa*. Tal programa teve início com o seqüenciamento completo do genoma de *X. fastidiosa* realizado por um consórcio de pesquisadores brasileiros (Simpson *et al.*, 2000).

A partir da seqüência genômica de *X. fastidiosa*, foi traçado todo o perfil básico para a sobrevivência dessa bactéria como o seu metabolismo energético, a síntese de aminoácidos,

nucleotídeos e lipídeos e os mecanismos de transcrição, tradução e reparo. Mais importante que esse perfil, o sequenciamento do genoma de *X. fastidiosa* também permitiu a formulação de hipóteses mais elaboradas sobre os mecanismos de patogenicidade dessa bactéria (Dow & Daniels, 2000; Keen *et al.*, 2000; Simpson *et al.*, 2000; Silva *et al.*, 2001).

Os possíveis mecanismos de patogenicidade revelados pela sequência genômica de *X. fastidiosa* são: produção de toxinas, secreção de enzimas celulolíticas e proteolíticas, sequestro de íons e a formação de biofilmes que bloqueiam o fluxo de seiva nos vasos xilemáticos. Associados a estes mecanismos, um enorme número de proteínas estão envolvidas na detoxificação, no bloqueio da entrada de drogas e no descarte ativo de toxinas que entram na célula garantindo sua sobrevivência no hospedeiro, protegendo-a de ataques de outros microorganismos que habitam no xilema, ou do próprio hospedeiro.

Anteriormente à análise genômica, as hipóteses pelas quais a *X. fastidiosa* causava doenças em seus inúmeros hospedeiros eram vagas e pouco exploradas. Hopkins (1989) cita a oclusão dos vasos xilemáticos pelas bactérias, tiloses e gomas; a produção de fitotoxinas e o desbalanço de reguladores de crescimento como os possíveis mecanismos de patogenicidade dessa bactéria. Os procedimentos comumente utilizados no estudo da interação planta - patógeno, em nível molecular, consistiam na obtenção de linhagens mutantes, que tivessem sofrido alteração na expressão de alguma função ligada à patogenicidade (Daniels *et al.*, 1987). Porém, nenhum experimento foi capaz de prova ou excluir qualquer uma dessas teorias.

Desta forma, com a análise da sequência genômica completa foi possível fazer algumas deduções a respeito dos mecanismos de adesão, migração e toxicidade, os quais estão envolvidos com o processo patológico de *X. fastidiosa*.

O genoma da *X. fastidiosa* 9a5c possui um cromossomo com 2.679.305 pares de bases e dois plasmídeos denominados pXF51 e pXF1.3. Do total de 2.884 quadros de leitura aberta

(ORFs), apenas 1.314 delas puderam ter sua função predita a partir da predição de similaridade a outros genes de função conhecida (Simpson *et al.*, 2000).

Talvez a primeira característica marcante encontrada na análise genômica seja a ausência dos genes de avirulência (*avr*), bem como dos genes que compõem o sistema de secreção do tipo III, necessário aos genes *avr*. A ausência destes genes pode ser em parte explicada pelo modo de vida e transmissão da *X. fastidiosa* que não requer a invasão e/ou lise das células do hospedeiro para a sua infecção ou sobrevivência (Dow *et al.*, 2000; Lambais *et al.*, 2000; Simpson *et al.*, 2000).

No genoma da *X. fastidiosa* também se identificou a presença de genes pertencentes a uma família de toxinas (RTX), as quais são largamente distribuídas entre bactérias gram-negativas patogênicas. As toxinas lesam ou matam células vegetais permitindo com que o patógeno colonize a planta.

Outra estratégia crucial para que a *X. fastidiosa* colonize efetivamente a planta é a migração. Para tal, a bactéria produz enzimas extracelulares que degradam a parede celular das células dos vasos do xilema, facilitando a entrada do patógeno ou sua dispersão no hospedeiro. Foram identificados genes em *X. fastidiosa* que codificam a celulase e seu precursor, poligalacturonase (Simpson *et al.*, 2000). A ausência de enzimas pectinolíticas funcionais no genoma de *X. fastidiosa* se deve a sua forma de vida, que não requer a invasão e destruição dos tecidos do hospedeiro (Keen *et al.*, 2000; Simpson *et al.*, 2000). A redução ou ausência desses mecanismos enzimáticos de invasão ou degradação pode ser uma forma de evitar os sistemas de defesa da planta, que são disparados quando esta detecta a lise de suas células (Keen *et al.*, 2000).

Muitos fitopatógenos produzem uma matriz extracelular composta por polissacarídeos extracelulares (EPS). Esta matriz também é característica de *X. fastidiosa* (Chagas *et al.*, 1992),

embora, diferentemente de *Xanthomonas campestris* pv *campestris*, não encontra-se no genoma de *X. fastidiosa* alguns dos genes que codificam enzimas envolvidas na síntese de EPS (*gumI*, *gumL* e *gumG*), sugerindo que em *X. fastidiosa* estes exopolissacarídeos sejam menos viscosos.

A síntese destas enzimas de EPS é regulada de uma maneira complexa por vários genes independentes, organizados no *cluster rpf* (*regulation pathogenicity factor*). Em *Xanthomonas* sp., utilizando-se mutagênese por transposon (Tang *et al.*, 1991) foi verificada a existência de pelo menos oito genes (*rpfA-H*) que levaram à redução dos níveis de todas as enzimas extracelulares e EPS. Sabe-se que *X. fastidiosa* também contém genes *rpfA*, *rpfB*, *rpfC* e *rpfF*, sugerindo que ambas bactérias regulem a síntese de fatores patogênicos EPS por mecanismos similares.

O bloqueio dos vasos xilêmicos pelo biofilme produzido pela *X. fastidiosa* é o mecanismo de patogênese mais sugerido e pesquisado. Além da síntese do EPS denominado “goma fastidiana”, a formação do biofilme envolve a adesão das bactérias ao substrato e agregação entre elas.

As fimbrias de *X. fastidiosa* observadas ao microscópio eletrônico (Raju & Wells, 1986) são estruturas importantes tanto para o processo de colonização do xilema da planta, quanto para promover a adesão da bactéria ao hospedar o inseto vetor (Brlansky *et al.*, 1983). Foram identificados genes que controlam a transcrição de subunidades da fimbria (*pilS* e *pilR*) e outros que codificam um sistema quimiostático traduzindo sinais do meio para a maquinaria do pili (*pilG*, *H*, *I*, *J* e *chpA*) em resposta ao hospedeiro (Simpson *et al.*, 2000).

Além dessas duas formas de adesão, produção de EPS e fimbria, as quais contribuem para o agrupamento da bactéria e adesão da mesma à parede do xilema, também foram identificados genes de alta similaridade com genes de outros patógenos. Estes genes codificam adesinas. As adesinas são proteínas que se localizam na membrana externa das células e

promovem a adesão a superfícies. As adesinas, até o seqüenciamento do genoma da *X. fastidiosa*, eram associadas apenas a patógenos animais. Sua presença em um fitopatógeno aumenta o número de evidências sobre a generalidade dos mecanismos de patogênese bacteriana, independentemente da natureza do hospedeiro (Simpson *et al.*, 2000).

A presença de genes relacionados à patogenicidade no genoma da *X. fastidiosa* contribuiu para se generalizar os mecanismos de patogenicidade, no entanto os processos são complexos e ainda pouco entendidos.

1.3. A ORF XF1729 DE *X. FASTIDIOSA*

De acordo com a anotação genômica, a ORF XF1729 de 876 pares de bases, codifica uma proteína de 291 resíduos de aminoácidos, com um peso molecular predito de 31.4kDa, um pI teórico de 4.99. Esta proteína pertence à categoria de produção de toxina e detoxificação e está anotada como sendo uma fenilacetaldéido desidrognase (<http://aeg.lbi.ic.unicamp.br/xf/>).

No entanto, a sua função putativa pode ser identificada via similaridade com seqüências caracterizadas em banco de dados. O alinhamento comparativo de sua seqüência de aminoácidos revelou que a proteína é similar às putativas oxidorreduases encontradas em diversas espécies. Além disso, o alinhamento também revelou que a proteína apresenta um domínio comum com as proteínas pertencentes à superfamília das aldo-ceto redutases (*AKR superfamily*).

A ORF XF1729 está localizada, juntamente com outras diferentes e putativas desidrogenases/ceto-redutases (Nunes *et al.*, 2003), em uma grande ilha genômica específica de *X. fastidiosa* causadora de CVC (chamada GI₂), a qual parece ser importante na adaptação ao ambiente e na patogenicidade desta bactéria (Bhattacharyya *et al.*, 2002; Oliveira *et al.*, 2002). A bactéria requer proteção contra a produção de espécies de oxigênios reativos, os quais são

produzidos pela planta em resposta às injúrias. Para se prevenir contra estes oxigênios reativos, a *X. fastidiosa* apresenta seis genes envolvidos no processo de oxido-redução (Moreira *et al.*, 2004).

1.4. AS ALDO-CETO REDUTASES NOS MICROORGANISMOS

Aldeídos e cetonas são intermediários importantes em muitas vias metabólicas como metabolismo de açúcar, biossíntese de esteróide, metabolismo de aminoácidos e biossíntese de metabólitos secundários. Também encontramos em células de microorganismos produtos aldeídicos de peroxidação lipídica formados após a exposição a agentes oxidantes e metais pesados (Avery, 2001) bem como, diretamente de fonte exógena, incluindo fitoalexina de plantas, toxinas e derivados sintéticos do meio. Por causa da reatividade do grupo carbonil, muitos aldeídos e cetonas são mutagênicos e representam um perigo em potencial para a célula microbiana (Marnett *et al.*, 1985). A habilidade de metabolizar estes compostos confere uma vantagem seletiva a qualquer microorganismo quando exposto aos mesmos.

Aldeídos e cetonas podem ser metabolizados por inúmeros sistemas enzimáticos incluindo aldeído desidrogenases que podem oxidar aldeídos à ácidos e as bem caracterizadas álcool desidrogenases, as quais incluem as famílias das desidrogenases/reduases de cadeia curta e média (Reid & Fewson, 1994; Jornvall *et al.*, 1999).

O número e a diversidade estrutural destas famílias de enzimas podem ser considerados suficiente para a célula ser capaz de reduzir aldeídos e cetonas. No entanto, membros de uma outra família de enzimas, Aldo-ceto reduases (AKR), estão presentes em muitos microorganismos (Kataoka, *et al.*, 1992) e também são capazes de reduzir estes mesmos substratos, sugerindo uma sobreposição de função.

As Aldo-ceto redutases são uma superfamília de enzimas relacionadas que compartilham características estruturais e funcionais. Com a rápida expansão desta superfamília criou-se a necessidade de se estabelecer uma nomenclatura sistemática baseado na similaridade da sequência de aminoácidos (Jez *et al.*, 1997), a qual é periodicamente atualizada (Jez & Penning, 2001). Determinar a habilidade e a função destas enzimas no metabolismo do carbonil é um importante objetivo, porém a dificuldade consiste na grande quantidade de substratos da maioria das AKRs e na típica presença de mais de uma AKR em cada célula.

Apesar de uma considerável sobreposição funcional, a estrutura tri-dimensional das AKRs difere consideravelmente das proteínas pertencente à família das desidrogenases/redutases de cadeia curta (Jornvall *et al.*, 1995). Na maioria dos casos, as AKR são monoméricas, com exceção da enzima dimérica AKR7 de mamífero e de algumas xyloses redutases de levedura. Em microorganismos, a AKR5D (2,5 diceto-D-gluconico ácido redutase - 2,5-DKGR) de *Corynebacterium* sp. é capaz de catalisar a redução estereoespecífica do 2,5-diceto-D-gluconato à 2-ceto-L-gluconato, um precursor na produção industrial de vitamina C (L-ascorbato) (Khurana *et al.*, 1998). Em levedura, a AKR3A1 (enzima Gcylp) é capaz de catalisar a redução de inúmeros substratos aldeídos, incluindo o DL-gliceraldeído, ainda que sua função fisiológica seja desconhecida (Hur & Wilson, 2000). Ambas enzimas são estruturalmente similares às enzimas AKRs de mamíferos, compartilhando o mesmo enovelamento do tipo $(\alpha/\beta)_8$ *barrel*, além de apresentarem alta similaridade de aminoácidos no sítio ativo.

Informações estruturais têm revelado detalhes importantes sobre o mecanismo catalítico, o qual envolve a tétrede de aminoácidos Tyr, His, Asp e Lys. Parece que o mecanismo é semelhante na maioria das AKR que possuem estes quatro aminoácidos.

O modo pelo qual a AKR liga ao cofator NADPH em uma conformação estendida, sem um enovelamento do tipo "Rossmann" também parece ser similar na maioria das AKRs. Em 2,5-

DKGR, a ligação do cofator NADPH causa mudanças conformacionais na estrutura da enzima e isto parece ser essencial para a subsequente ligação ao substrato (Sanli & Blaber, 2001). Mutações sítio dirigidas da AKR de mamífero revelaram que a região C-terminal é importante para determinar a especificidade do substrato à enzima.

Muitas AKRs tem sido estudadas com respeito à atividade enzimática catalítica. Outras enzimas microbianas NADPH-depentes com atividade de redutase para aldeídos e cetonas têm sido purificadas e testadas, mas sem qualquer informação de sequência, não sendo possível saber se elas pertencem à família das AKR ou das álcool desidrogenases. Muitas enzimas AKR que têm sido caracterizadas mostram uma distinta preferência pelo NADPH como cofator, no entanto, algumas xyloses redutases de levedura podem usar NADH e NADPH como cofator, com igual eficiência. A variedade de substratos que uma AKR pode reduzir é muito ampla, mas há claras diferenças na afinidade a eles. As xyloses redutases de levedura formam a subfamília AKR2B (Lee, 1988). Estas enzimas catalisam a redução de xylose à xylitol, como parte da via metabólica da xylulose.

Devido à sobreposição de substratos específicos das enzimas AKR tem sido difícil caracterizar uma enzima individualmente. Além disso, pode existir mais de uma enzima dentro da célula capaz de reduzir um aldeído ou uma cetona. Por exemplo, tomando-se por base as informações genômicas, pode-se dizer que um procarioto como a *E. coli* possui 6 AKRs (Blattner *et al.*, 1997) e um eucarioto com a levedura, possui mais de 14 AKRs (Goffeau *et al.*, 1996). Ainda assim, pode haver a complementação de outras desidrogenases/redutases, as quais podem agir com grande similaridade funcional no metabolismo dos aldeídos.

A questão fundamental consiste em saber o motivo pelo qual as células apresentam tantas enzimas capazes de desempenhar aparentemente papéis similares. Acredita-se que apesar de existir uma considerável sobreposição de funções, cada enzima está envolvida em uma particular

proposta, e sua habilidade em ter um grande número de substratos deve-se ao fato da enzima ter dentro daquele grupo, um substrato preferido.

2. OBJETIVOS

2.1. OBJETIVOS GERAIS

- Explorar a funcionalidade das proteínas relacionadas à patogenicidade da bactéria *X. fastidiosa*.
- Fornecer subsídios para a resolução da estrutura tridimensional da proteína em estudo.

2.2. OBJETIVOS ESPECÍFICOS

- Clonagem, expressão em *Escherichia coli* e purificação da proteína codificada pela *ORF* XF1729.
- Caracterização funcional e estrutural da proteína codificada por esta *ORF*.

3. METODOLOGIA

3.1. CARACTERIZAÇÃO ENZIMÁTICA

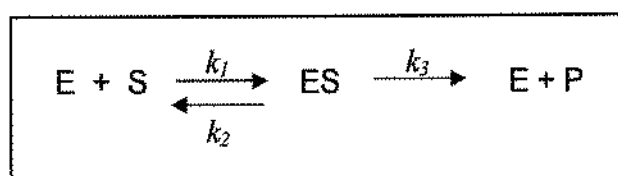
Como sabido, as enzimas são proteínas especializadas na catálise de reações biológicas. Elas estão entre as biomoléculas mais notáveis devido à sua extraordinária especificidade e poder catalítico. Praticamente todas as reações que caracterizam o metabolismo celular são catalisadas por enzimas. Como catalisadores celulares extremamente poderosos, as enzimas aceleram a velocidade de uma reação, sem no entanto participar dela como reagente ou produto. As enzimas atuam ainda como reguladoras deste conjunto complexo de reações, sendo consideradas as unidades funcionais do metabolismo celular.

As enzimas podem ser classificadas de acordo com vários critérios, entretanto a nomenclatura das enzimas é normatizada por um comitê especializado, o *Nomenclature Committee of the International Union of Biochemistry and Molecular Biology* (NC-IUBMB). De acordo com este comitê as enzimas são divididas em seis grandes classes, conforme a reação química catalisada: (1) Oxidorredutases; (2) Transferases; (3) Hidrolases; (4) Liasas; (5) Isomerases; (6) Ligases. As oxidorredutases são enzimas que catalisam reações de oxidação-redução (reações de transferência de elétrons). Nesta classe, incluem-se as redutases, oxidases, desidrogenases, oxigenases, peroxidases e hidrogenases.

A cinética enzimática é a parte da enzimologia que estuda a velocidade das reações catalisadas por enzimas, e os fatores que influenciam nesta velocidade. A cinética de uma enzima é estudada avaliando-se a quantidade de produto formado ou a quantidade de substrato consumido por unidade de tempo de reação (Lehninger *et al.*, 1993).

As enzimas aceleram a velocidade de uma reação por diminuir a energia livre de ativação da mesma, sem alterar a termodinâmica da reação, ou seja, a energia dos reagentes e produtos da reação enzimática e de sua equivalente não enzimática é idêntica. As enzimas aceleram as reações por prover uma nova via de reação na qual o estado de transição tem uma energia livre mais baixa e, portanto é mais acessível do que na reação não catalisada. A primeira etapa na catálise é a formação de um complexo enzima-substrato (ES). Para superar a energia de ativação em uma reação, a enzima liga-se à molécula de seu substrato formando um estado de transição (ES), sendo este obrigatório no processo catalítico enzimático. O estado de transição (ES) é sempre um composto instável e de alta energia, ligado com altíssima afinidade ao sítio catalítico da enzima. O reconhecimento do substrato pela enzima é um processo dinâmico, acompanhado de mudanças conformacionais no centro ativo.

O modelo de Michaelis-Menten explica as propriedades cinéticas de algumas enzimas. Neste modelo, uma enzima (E) combina-se com um substrato (S) formando um complexo enzima-substrato (ES), que pode seguir para formar um produto (P), ou dissociar-se em E e S. Conforme esquematizado na equação 1, onde k_1 , k_2 e k_3 são as constantes cinéticas:



Equação 1: Equação modelo de Michaelis-Menten para enzimas.

Um dos fatores que afetam a velocidade de uma reação catalisada por uma enzima purificada é a concentração do substrato presente [S]. Os estudos da concentração de substrato são difíceis de serem avaliadas devido ao fato da [S] variar durante o curso de uma reação à

medida que o substrato é transformado em produto. Porém, a medição da velocidade inicial (V_0) é uma alternativa para simplificar o problema. Em um experimento cinético a $[S]$ é sempre muito maior que a concentração da enzima $[E]$, e se o tempo da reação é suficientemente curto, as mudanças da $[S]$ serão negligenciáveis, podendo ser consideradas uma constante.

O efeito provocado em V_0 pela variação da $[S]$, quando a concentração da enzima é mantida constante está apresentado na Figura 3. Em concentrações pequenas do substrato, V_0 aumenta quase linearmente como os aumentos de $[S]$. Em concentrações maiores de substrato V_0 aumenta por incrementos menores em resposta aos aumentos da $[S]$. Finalmente é alcançado um ponto acima do qual ocorrem apenas aumentos insignificantes em V_0 , mesmo diante de aumentos em $[S]$, sendo este ponto chamado de velocidade máxima ($V_{\text{máx}}$).

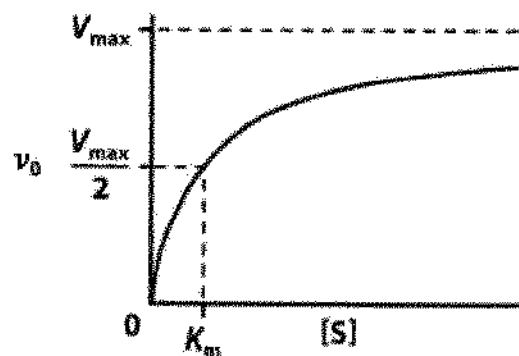


Figura 3: Gráfico de Michaelis-Menten: Gráfico da velocidade inicial de reação, V_0 , em função da concentração do substrato, $[S]$, para uma enzima que obedece à cinética de Michaelis-Menten, sendo V_{max} a velocidade máxima e K_m a constante de Michaelis.

A Figura 3 mostra a relação entre $[S]$ e V_0 para uma reação enzimática. A curva que expressa esta relação possui a mesma forma para a maioria das enzimas, sendo expressa pela equação de Michaelis-Menten, derivada por estes pesquisadores partindo da hipótese básica de

que, nas reações enzimáticas, o passo limitante da velocidade é a quebra do complexo ES para formar o produto e a enzima livre.

A Equação de Michaelis-Menten (Equação 2) é a equação da velocidade para uma reação catalisada enzimaticamente e com um único substrato. É uma expressão da relação quantitativa entre V_0 , velocidade inicial máxima $V_{m\acute{a}x}$ e a concentração inicial do substrato $[S]$, todas relacionadas através da constante de Michaelis-Menten, K_m . A constante de Michaelis-Menten é uma constante dinâmica, ou de pseudoequilíbrio, que expressa a relação entre as concentrações reais no estado estacionário ao invés de concentrações no equilíbrio.

$$V_0 = \frac{V_{m\acute{a}x}[S]}{K_m + [S]}$$

Equação 2: Equação de Michaelis-Menten.

Uma relação numérica importante emerge da Equação de Michaelis-Menten no caso especial quando V_0 é exatamente a metade de $V_{m\acute{a}x}$ (Equação 3).

$$K_m = [S] \longrightarrow V_0 = \frac{1}{2} V_{m\acute{a}x}$$

Equação 3: Equação derivada da equação de Michaelis-Menten.

Nota-se que K_m é equivalente à concentração do substrato na qual V_0 é igual à metade de $V_{m\acute{a}x}$, indicando a afinidade de uma enzima pelo seu substrato. Quanto menor for o valor de K_m , maior será a afinidade da enzima pelo substrato. A constante cinética k_3 ou K_{cat} expressa o número de moléculas de substrato transformadas em produto por unidade de tempo, em um só centro catalítico quando a enzima está completamente saturada pelo substrato. Cada enzima tem valores ótimos de K_{cat} e K_m que refletem o ambiente celular, a concentração do substrato normalmente encontrado *in vivo* pela enzima e a química da reação que está sendo catalisada.

A determinação da cinética das enzimas é importante como medida da velocidade da reação e de como esta velocidade é afetada pelas condições químicas e físicas. Os estudos cinéticos fornecem informações sobre os mecanismos básicos de catálise e outros fatores que caracterizam a enzima (Lee, 1992).

Para a caracterização de oxidoredutases que utilizam NADPH para catalisar a redução de diversos substratos, os estudos de atividade enzimática são medidos através do decrescimento da absorbância a 340nm, decorrente da oxidação do NADPH a NADP⁺. O valor do coeficiente de extinção molar para o NADPH é de $6220M^{-1}cm^{-1}$ a 340 nm. A taxa da reação é determinada mantendo-se fixa as concentrações da enzima e cofator e, variando as concentrações do substrato. Os dados obtidos da leitura da absorbância a 340 nm, pelo tempo, para cada concentração do substrato, são representados graficamente e calculada a velocidade inicial da reação (V_0) pela equação de regressão linear ($y = ax + b$), sendo $a = V_0$. Desta forma é possível determinar o efeito do substrato na velocidade inicial da reação, podendo-se calcular a velocidade máxima ($V_{m\acute{a}x}$) e obter a constante de Michaelis-Menten (K_m) para cada substrato testado. Segundo Naqui & Chance a cinética da maioria das reações catalisadas por enzimas obedece à equação de Michaelis-Menten.

3.2. TÉCNICAS UTILIZADAS NA CARACTERIZAÇÃO ESTRUTURAL DE PROTEÍNAS

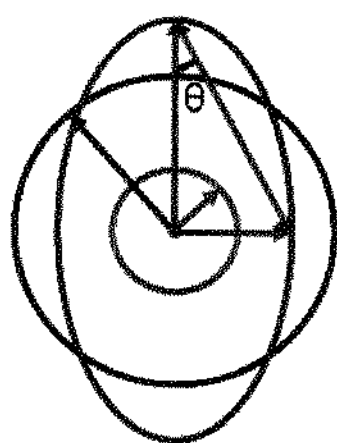
3.2.1. DICROÍSMO CIRCULAR

O Dicroísmo Circular (CD) tem sido reconhecido como uma valiosa técnica espectroscópica para examinar a estrutura de proteínas em solução. O princípio desta técnica se deve ao fato de que as moléculas opticamente ativas interagem com a luz polarizada provocando alteração na polarização da luz incidente. Pode-se dizer que a luz polarizada apresenta dois componentes circularmente polarizados de igual magnitude, sendo um à esquerda (E) e outro à direita (D). A espectroscopia de dicroísmo circular detecta a absorção diferencial das duas componentes. Se após a passagem através da amostra examinada, as componentes E e D não são absorvidas ou são absorvidas em igual extensão, a resultante de E e D regenera a radiação polarizada no plano inicial. No entanto, se E e D são absorvidas em diferentes extensões, pode-se dizer que a radiação resultante possui uma polarização elíptica (Fig. 4A) (Kelly *et al.*, 2005). Os equipamentos de CD, espectropolarímetros, medem a diferença da absorção de luz circularmente polarizada à direita e à esquerda ($\Delta A = A_E - A_D$), convertendo o sinal em termos da elipticidade (θ) expressa pela fórmula $\theta = 32.98 \cdot \Delta A$ (graus) (Woody, 1995).

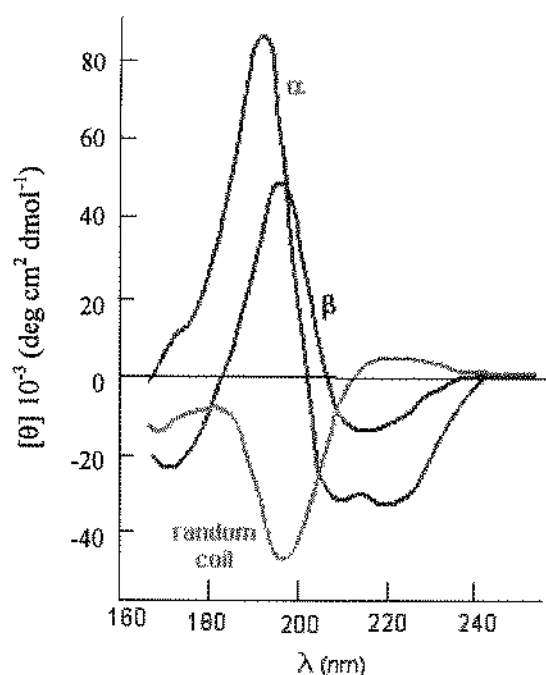
O sinal de CD é observado quando o cromóforo é quiral (opticamente ativo) devido ao fato de (a) ser intrinsecamente quiral por causa de sua estrutura, por exemplo, um átomo de C com quatro diferentes ligantes ou uma ponte de dissulfeto, a qual é quiral devido ao ângulo diedral de sua cadeia de átomos; (b) estar covalentemente ligado ao centro quiral na molécula; (c) estar localizado em um ambiente assimétrico em virtude da estrutura tridimensional adotada pela molécula (Kelly & Price, 2000). No caso das proteínas, os cromóforos de interesse incluem

a ligação peptídica (absorção abaixo de 240nm), aminoácidos aromáticos da cadeia lateral (absorção entre 260 e 320nm) e, as pontes de dissulfeto (tênue banda de absorção por volta de 260nm) (Kelly & Price, 1997). O sinal de CD somente aumenta quando a absorção da radiação ocorre e, devido a este motivo as bandas espectrais são facilmente verificadas para distinguir características estruturais da molécula. Um avanço da técnica de CD no estudo de proteínas reside no fato de obtermos informações complementares sobre a estrutura protéica de diferentes regiões espectrais, por exemplo, a composição da estrutura secundária.

Os diferentes tipos de estrutura secundária (α -hélice, folha- β e estrutura randômica) que compõem a molécula apresentam um espectro de CD característico na faixa do UV-distante (Fasman, 1996), devido aos ângulos ϕ e ψ do carbono- α da cadeia polipeptídica. As α -hélices e folhas- β são estruturas bem compactas presentes nas proteínas enoveladas. Já as estruturas randômicas estão presentes tanto em proteínas desenoveladas quanto nas alças flexíveis de conexão entre elementos de estrutura secundária. A composição predominante da estrutura secundária de uma proteína pode ser verificada pela simples observação espectral. As proteínas que possuem sua estrutura secundária formada predominantemente por α -hélices exibem bandas espectrais em 192, 208 e 222 nm, aproximadamente. As proteínas formadas predominantemente por folhas- β apresentam um espectro de CD contendo bandas espectrais em 196 e 216 nm, aproximadamente. A estrutura randômica é representada pelo espectro contendo bandas espectrais próximo de 195 e 212 nm (Pelton & McLean, 2000). A Figura 4B apresenta o espectro de CD característico de cada estrutura secundária.



A



B

Figura 4: **A)** Polarização elíptica da Luz (roxo) é composta pela contribuição diferencial da luz circularmente polarizada à esquerda (vermelho) e à direita (azul). **B)** Representação dos espectros de CD característicos para cada tipo de estrutura secundária: α -hélices (vermelho); folhas- β (azul) e estrutura randômica (verde).

A estimativa da composição da estrutura secundária da proteína a partir do espectro de CD continua sendo uma tarefa empírica, apesar dos métodos propostos variarem da simples observação à utilização de algoritmos sofisticados. Atualmente, foram desenvolvidos inúmeros algoritmos, os quais utilizam as informações geradas pelos bancos de dados de espectros de proteínas com estrutura cristalográfica determinada e o espectro de CD obtido na faixa de UV-distante da proteína em estudo. Dentre os algoritmos amplamente utilizados para esta análise estão: SELCON (Sreerama & Woody, 1993), VARSLC (Manavalan & Johnson Jr., 1987), CDSSTR (Johnson, 1999), K2d (Andrade *et al.*, 1993) e CONTIN (Provencher & Glöckner, 1981).

3.2.2. ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO

O espalhamento de raios-X a baixo ângulo (SAXS) é uma técnica experimental que permite o estudo da estrutura e interações de macromoléculas biológicas em solução. Esta técnica pode ser usada para estudar proteínas, ácidos nucleicos e seus complexos em uma variedade de condições, desde altamente desnaturante até próximo da fisiológica sem a necessidade de cristalizar a amostra e sem as limitações de peso molecular inerentes à técnica de RMN. A crescente importância do SAXS como uma ferramenta em biologia estrutural se reflete no número de publicações relacionadas por ano que triplicaram na última década. Além disto, os progressos na instrumentação de SAXS e os novos métodos de análise, que melhoram substancialmente a resolução e a realidade dos modelos estruturais, estão fazendo do método uma importante ferramenta para o estudo estrutural (Grossmann, 2007).

O princípio básico do SAXS é o espalhamento elástico de fótons de raios-X de moléculas em solução. Assim, a intensidade espalhada é registrada em função do ângulo de espalhamento (Svergun *et al.*, 2001). O perfil de espalhamento fornece informações sobre conformação e estrutura global das moléculas estudadas.

Os primeiros estudos de moléculas biológicas através da técnica de SAXS limitavam-se à obtenção de alguns parâmetros moleculares como: peso molecular, Raio de giro (R_g) e distância intramolecular máxima (D_{max}). Na década passada, o desenvolvimento de algoritmos permitiu a construção *ab initio* dos mapas de densidade eletrônica de baixa resolução tridimensional a partir do perfil de espalhamento unidimensional (Walther *et al.*, 2000; Svergun *et al.*, 2001), obtendo-se informações estruturais além de simples parâmetros. Recentemente, foram criadas ferramentas para modelar proteínas e complexos moleculares a partir das informações obtidas por SAXS,

basendo-se nas estruturas conhecidas ou parcialmente conhecidas através de experimentos de alta resolução.

Os parâmetros freqüentemente extraídos de um perfil de SAXS de uma biomolécula em solução são o R_g e a intensidade de espalhamento direta $I(0)$. Estes parâmetros são obtidos da fórmula de Guinier, $I(q) \sim I(0) \exp(-q^2 R_g^2/3)$, para pequenos momentos de transferência q , ($q = 4\pi \sin(\theta)/\lambda$), onde 2θ é o ângulo de espalhamento e λ é o comprimento de onda de raios-X, plotando $\ln(I(q))$ por q^2 e fitando o ângulo e a intercessão. O raio de giro (R_g) é um modelo de caracterização do tamanho da molécula e a intensidade de espalhamento direta ($I(0)$) pode ser dada em função do peso molecular, através da relação: $I(0) = kc(\Delta\rho)^2(PM)^2$, onde k é a constante (determinada a partir das medidas de proteínas de peso molecular e concentração conhecida), c é a concentração da macromolécula, $\Delta\rho$ é a variação no contraste da densidade eletrônica da molécula e PM é o peso molecular (Lipfert & Doniach, 2007).

O perfil de espalhamento também pode ser descrito em termos da função de distribuição de distâncias $p(r)$ (Glatter & Kratky, 1982). Esta função possui uma definição geométrica que pode ser entendida dividindo-se uma partícula em uma grande número de elementos de volumes pequenos e idênticos. A função $p(r)$ descreve a distribuição das distâncias r que podem ser encontradas a partir da combinação de qualquer par de elementos de volume. A função $p(r)$ pode ser obtida através da transformada indireta de Fourier a partir do perfil de espalhamento usando o software GNOM (Svergun, 1992), por exemplo. Em princípio, a função de distribuição de distâncias $p(r)$ contém a mesma informação da intensidade de espalhamento $I(s)$, mas a representação do espaço real é mais intuitiva e a informação sobre a forma da partícula pode freqüentemente ser deduzida diretamente por uma inspeção visual da $p(r)$. No entanto, atualmente, o uso da função $p(r)$ para ajudar a visualizar a forma da molécula tem sido

substituída por algoritmos que fornecem o mapa da densidade eletrônica da molécula a partir do perfil de SAXS unidimensional.

A visualização e manipulação de estruturas atômica e modelos de baixa resolução, a partir de modelos de alta resolução, pode ser obtida através de programas CRY SOL (Svergun *et al.*, 1995) e GNOM (Svergun, 1992). Porém, quando não há modelos da alta resolução para comparação, parte-se para uma análise mais cuidadosa, utilizando-se a função $p(r)$ para inferir a curva experimental dada pelos programas CREDO (Petoukhov *et al.*, 2002), DAMAVER (Svergun *et al.*, 2001), DAMIM (Svergun, 1999) e outros.

4. MANUSCRITO

**"A new member of the aldo-keto reductase family from the plant pathogen *Xylella*
fastidiosa"**

Luciana K. Rosselli, Cristiano L.P. Oliveira, Adriano R. Azzoni, Susely F.S. Tada, Cleide F. Catani, Antônio M. Saraiva, José Sérgio M. Soares, Francisco J. Medrano, Iris L. Torriani e Anete P. de Souza

Publicado na revista

Archives of Biochemistry and Biophysics 453 (2006) 143-150



A new member of the aldo–keto reductase family from the plant pathogen *Xylella fastidiosa*

Luciana K. Rosselli^a, Cristiano L.P. Oliveira^{b,c}, Adriano R. Azzoni^a, Susely F.S. Tada^a,
Cleide F. Catani^d, Antonio M. Saraiva^a, José Sérgio M. Soares^a, Francisco J. Medrano^c,
Iris L. Torriani^{b,c}, 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, 13083-875 Campinas, SP, Brazil

^b Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Campinas, SP, Brazil

^c Laboratório Nacional de Luz Síncrotron, C.P. 6192, Campinas, SP, Brazil

^d Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, C.P. 6010, Campinas, SP, Brazil

Received 18 April 2006, and in revised form 7 July 2006

Available online 31 July 2006

Abstract

The *Xylella fastidiosa* genome program generated a large number of gene sequences that belong to pathogenicity, virulence and adaptation categories from this important plant pathogen. One of these genes (XF1729) encodes a protein similar to a superfamily of aldo–keto reductase together with a number of structurally and functionally related NADPH-dependent oxidoreductases. In this work, the similar sequence XF1729 from *X. fastidiosa* was cloned onto the pET32Xa/LIC vector in order to overexpress a recombinant His-tag fusion protein in *Escherichia coli* BL21(DE3). The expressed protein in the soluble fraction was purified by immobilized metal affinity chromatography (agarose–IDA–Ni resin). Secondary structure contents were verified by circular dichroism spectroscopy. Small angle X-ray scattering (SAXS) measurements furnish general structural parameters and provide a strong indication that the protein has a monomeric form in solution. Also, ab initio calculations show that the protein has some similarities with a previously crystallized aldo–keto reductase protein. The recombinant XF1729 purified to homogeneity catalyzed the reduction of DL-glyceraldehyde (K_{cat} 2.26 s⁻¹, K_m 8.20 ± 0.98 mM) and 2-nitrobenzaldehyde (K_{cat} 11.74 s⁻¹, K_m 0.14 ± 0.04 mM) in the presence of NADPH. The amino acid sequence deduced from XF1729 showed the highest identity (40% or higher) with several functional unknown proteins. Among the identified AKRs, we found approximately 29% of identity with YakC (AKR13), 30 and 28% with AKR11A and AKR11B, respectively. The results establish XF1729 as the new member of AKR family, AKR13B1. Finally, the first characterization by gel filtration chromatography assays indicates that the protein has an elongated shape, which generates an apparent higher molecular weight. The study of this protein is an effort to fight *X. fastidiosa*, which causes tremendous losses in many economically important plants.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Aldo–keto reductase; *Xylella fastidiosa*; XF1729 protein; XF1729 gene

The bacteria *Xylella fastidiosa* [1] is the causal agent of a number of economically important crop diseases, including Citrus Variegated Chlorosis (CVC)¹, affecting mainly sweet oranges [2]. Although its complete genome is already

sequenced [3], *X. fastidiosa* is very poorly characterized by biochemical approaches at a protein level. Since *X. fastidiosa* 9a5c strain (associated with CVC) is the first plant pathogen to be fully sequenced, a large benefit for the whole field of disease research in plants can be expected. An

* Corresponding author. Fax: +55 19 37881089.

E-mail address: anete@unicamp.br (A.P. Souza).

¹ Abbreviations used: CVC, Citrus Variegated Chlorosis; AKR, aldo–keto reductase; PMSF, phenylmethylsulfonyl fluoride; CHCA, α -cyano-4-hydroxycinnamic

acid; PCR, polymerase chain reaction; SDS–PAGE, SDS–polyacrylamide gel electrophoresis; MALDI–TOF, matrix-assisted laser-desorption time-of-flight; CD, circular dichroism; PEG, polyethylene glycol.

initial step towards characterizing the function of several proteins has been taken by functional and structural genomic projects.

In order to gather information about the proteins involved in the mechanisms of the bacterium pathogenicity, we chose the XF1729 ORF (gi 9106797) for the characterization studies. This target ORF is located, together with other different and putative dehydrogenases/keto-reductases [4], in a large citrus specific genomic island (named GI₂) that seems to be important in the environmental adaptation and pathogenicity of this bacterium [5,6]. The bacteria require protection against the production of reactive oxygen species (ROS) by the plant in response to injuries. To prevent damage from ROS, *X. fastidiosa* contain six genes involved with oxide reduction [7].

The XF1729 ORF codifies a protein of 291 amino acids, with a predicted molecular weight of 31.4 kDa, theoretical pI of 4.99 that probably belongs to the toxin production and detoxification category. Its putative function could be assigned via similarity to characterized sequences in the databases. The comparative amino acids alignment shows that the target protein is similar to putative oxidoreductases found in several species and revealed that the protein has a common domain with aldo-keto reductase (AKR) superfamily proteins. However, this ORF product was first described in the genome annotation as being a phenylacet-aldehyde dehydrogenase due its similarity with the PadA enzyme from *Escherichia coli*. This enzyme belongs to another protein family (aldehyde dehydrogenase superfamily) that efficiently oxidizes only phenylacetaldehyde-like aromatic aldehydes and is a homodimer that uses NADP⁺ as a coenzyme [8,9].

The AKRs are a growing superfamily of oxidoreductases that are found across biological kingdoms and phyla. Found in prokaryotes and eukaryotes including yeast,

plant, amphibia and mammals. Although several dimers and tetramers have been characterized, members of the superfamily are typically monomeric (α/β)₈-barrel proteins, approximately 35 kDa in mass, which use NAD(P)(H) to catalyze the reduction of aldehydes and ketones, monosaccharides, ketosteroids and prostaglandins and other specific substrates [10,11]. Currently, the superfamily contain more than 120 proteins that are distributed throughout 14 families (AKR1–AKR14) on the basis of sequence identity: The criteria follow that members within a family have less than 40% amino acid sequence identity with other families and that members within a subfamily have greater than 60% sequence identity [11].

Comparing the XF1729 protein with its similar sequences, we found the highest identity (~65%) with a putative oxidoreductase from *Deinococcus radiodurans* [12] and approximately 29% of identity with YakC from *Schizosaccharomyces pombe* that is the founding member of 13th AKR family [13]. However, family 13 of the aldo-keto reductases currently consists of one poorly characterized protein and several unknown function proteins were found to belong to this family. Apart from this, we found two solved structure proteins – the vegetative protein 147 (PDB entry 1PYF) and the general stress protein 69 (PDB entry 1PZI) – from *Bacillus subtilis* that show an approximate identity of 30 and 28% with the XF1729 target protein, respectively. Both proteins are members of the AKR11 family [14,15] and the alignment of these sequences is shown in Fig. 1.

Here, we report the cloning and overexpression the recombinant protein in *E. coli*. The purified XF1729 protein exhibited a typical aldo-keto reductase activity, being able to reduce DL-glyceraldehyde and 2-nitrobenzaldehyde in the presence of NADPH. Its primary structure and activity assay showed that the enzyme is a new member of the

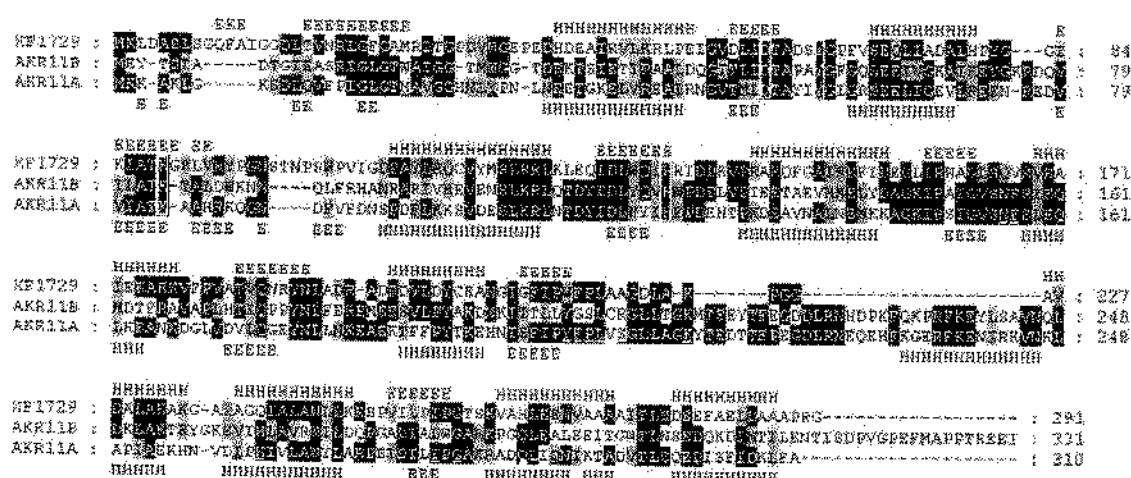


Fig. 1. Comparison of amino acid sequences of the XF1729 protein from *X. fastidiosa* (XF1729, gi 15838330) with the vegetative protein 147 (AKR11A, gi 47168571) and the general stress protein 69 (AKR11B, gi 47168574) from *Bacillus subtilis*. The multiple sequence alignment was carried out using the CLUSTALW and GENEDOC programs. Letters shaded in black indicate amino acids that are identical in at least two aligned sequences. Letters shaded in gray indicate similar amino acids. Vertical rectangles indicate the residues belonging to the catalytic tetrad present in all the sequences. Underlined letters show the residues involved in NADPH cofactor binding. Secondary structure elements (E = β -sheet, H = α -helix) are indicated above the predict structure deduced from the XF1729 sequence (by PSIPRED program) and below the AKR11A crystallographic structure (PDB entry 1PYF) obtained from its sequence.

AKR 13 family, classified as AKR13B1.² Although, in the genome annotation, the protein is described as belonging to the aldehyde dehydrogenase superfamily, its function has not been assigned (data not shown). In addition, modeling approaches based on SAXS data give the indication of the general shape of this protein, permitting the identification of similar structural features with proteins of the AKR11 family. 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.

Materials and methods

Materials

The oligonucleotide primers were synthesized at Invitrogen Life Technologies (São Paulo, Brazil). The pET32-Xa/LIC vector and the BL21(DE3) strain were obtained from Novagen (Madison, WI). The Chelating Sepharose Fast Flow (agarose-IDA-Ni) resin affinity resin was obtained from Qiagen (Hilden, Germany). The Coomassie blue reagent for total soluble protein determination was purchased from Bio-Rad (Hercules, CA). The molecular-mass marker (LMW) and the Tricorn 10/100 (10 × 100 mm) chromatography column were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). The protease inhibitor phenylmethylsulphonyl fluoride (PMSF), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α -cyano-4-hydroxycinnamic acid (CHCA), lysozyme, and the enzyme trypsin were purchased from Sigma Chemical (St. Louis, MO). All other chemical reagents used were of the highest commercially grade available.

Expression vector construction

The target ORF (XF1729) amplification for cloning into pET32Xa/LIC was done in two steps by polymerase chain reaction (PCR): the initial amplification of target ORF from *X. fastidiosa* 9a5c strain was carried out using the purified genomic DNA as the template and the 5'-GCG GTCGGCAGCGTGTCTG-3' (sense) and 5'-GCGGGC CTTTGTCGTTTTCTG-3' (antisense) primers. The amplification protocol consisted of a 3 min denaturation at 94 °C followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min and 30 s and final extension at 72 °C for 2 min. This first PCR product was used, as a template, for the specific amplification with a pair of PCR primers, 5'-GGTATTGAGGGTTCGCATGAACTTGAT-3' (sense) and 5'-AGAGGAGAGTTAGAGCCTCAACCGC GTGG-3' (antisense), specifically designed to be cloned into pET32Xa/LIC. The primers were designed to generate products with vector cohesive overhangs (in bold). In this second amplification, we used the same amplification pro-

ocol, except for the annealing temperature that was 55 °C. The amplified DNA fragments were analyzed on 1% agarose gel. The blunt-ended PCR products were purified, treated and cloned in the linearized pET32Xa/LIC vector according to the vector manufacturer's protocol [16]. The pET32Xa/LIC vector is designed for the expression of the recombinant protein fused to the 109 amino acid thioredoxin (11.7 kDa), a six amino acid His-tag and 15 amino acid S-tag sequences upstream to the cloning site. The fusion tags together have a molecular mass of 17.6 kDa, a theoretical pI of 5.95, and they can be removed from the recombinant target protein by protease cleavage using the factor Xa [16].

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 [17]. The standard method describes a heat shock step when cells are kept at 42 °C for 30 s. The high temperature step was substituted by letting the cells stand at room temperature for 10 min. Cells were cultured overnight at 37 °C in Luria-Bertani broth (LB) plates containing 50 μ g/mL ampicillin. The colonies were individually stored at -70 °C in a permanent 2YT+HMFM broth [18] and checked by PCR. The ORF XF1729 insert of pET 32Xa/LIC plasmid was sequenced to verify the correct nucleotide sequence.

Protein expression and purification

The *E. coli* strain BL21(DE3) cells were transformed with recombinant plasmid and cultured in 3.0 mL LB broth added to 50 μ g/mL ampicillin, grown overnight at 37 °C and 300 rpm. This pre-innoculum was then transferred to 5.0 L of LB broth containing ampicillin at the same concentration and grown to OD₆₀₀ of 0.8. At this point, the cells were induced with an addition of 5.6 mM lactose, at 37 °C, for 4 h and with shaking at 300 rpm. The cell pellets were resuspended in adsorption buffer (300 mM NaCl, 50 mM Tris-HCl, pH 7.5) containing PMSF and lysozyme to concentrations of 1.0 mM and 1.0 mg/mL, respectively. Then the suspension was sonicated, clarified by centrifugation at 27,500g and the clear supernatant was used for protein purification.

The two steps of chromatography purification were carried out in the Tricorn 10/100 column (Amersham Biosciences, USA), packed with 8 mL agarose-IDA-Ni resin, using an AKTA-FPLC System. In the first step, the column was equilibrated with an adsorption buffer, containing 1 M NaCl, and the clarified sample was applied to the column. The resin was washed with this buffer, containing 5 mM imidazole, and subsequently the heterolog protein was eluted using a gradient (1–100%) of adsorption buffer, containing 300 mM NaCl and 300 mM imidazole. The sample containing the recombinant XF1729 was incubated with trypsin (1%, w/w – 0.1 μ g per 10 μ g recombinant protein) at 25 °C for 20 min to separate the fusion tags of the protein. The protease was inactivated by PMSF at 1 mM and the sample was dialyzed against the adsorption buffer to

² Dr.T.M. Penning, a founding member of the nomenclature system for the AKR superfamily, personal communication.

prepare for the next purification step. The final step of purification was performed equilibrating the column with the adsorption buffer, containing 1 M NaCl. The column was loaded with the sample (containing XF1729 and the cleaved fusion protein) and the flowthrough proteins (XF1729) were collected. Bound proteins were eluted using a gradient (1–100%) of adsorption buffer, containing 300 mM NaCl and 300 mM imidazole.

Molecular weight and protein concentration determination

The purity and molecular weight were estimated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) was performed using a 4% stacking gel and a 12% separation gel according to Laemmli [19]. The use of gel filtration chromatography for the determination of the molecular mass, shape and oligomeric structure of the purified protein was performed using a pre-packed Superdex 75 HR10/30 column from Pharmacia Biotech (Uppsala, Sweden). The preparation of a calibration curve was carried out using the Low Molecular Weight Gel Filtration Calibration Kit, also from Pharmacia Biotech. The column was equilibrated with 50 mM Tris–HCl, pH 7.5, containing 300 mM NaCl. The protein sample at a concentration of about 2 mg/mL was injected using a 100 μ L sample loop. The protein elution volume needed for molecular mass determination was obtained from the UV chromatogram recorded by the AKTA–FPLC system used.

The total soluble protein concentration was assayed according to the method presented by Bradford (1976), using bovine albumin serum as the protein standard [20]. More accurate estimations for purified XF 1729 were made based on absorbance at 280 nm, using a calculated extinction coefficient of 1.190 g L⁻¹ cm⁻¹, based on the method described by Pace and Schmidt [21].

Identity confirmation of the XF1729 protein

In order to verify the correct amino acid sequence of the XF1729 protein, mass spectrometry and N-terminal sequencing were carried. The purified protein XF1729 mass spectrometric analysis was carried out using matrix-assisted laser-desorption time-of-flight (MALDI-TOF) MALDI-LR mass spectrometer (Micromass, USA). The trypsin-digested protein and intact protein were analyzed in the spectrometer using reflecting and linear modes, respectively, according to the equipment manufacturer's instructions. The matrixes used for trypsin-digested (peptides) and intact protein samples were CHCA and sinapinic acid, respectively. The masses of monoisotopic peaks with relative intensity higher than 5% of the most intense peak in the spectrum were used for comparison to a theoretical digestion of the protein by trypsin. This was carried out using the MS-Digest program (<http://prospector.ucsf.edu>). The N-terminal sequencing was done with the purified protein loaded in a 12% SDS–polyacrylamide gel, electroblotted onto a PVDF membrane (Bio-Rad, Hercules, CA) using

the transfer buffer (0.29% glycine; 0.58% Tris–HCl; 0.037% SDS; 20% methanol) and stained with Ponceau corant. A portion of the stained 31 kDa protein was excised and loaded onto a Protein Sequencer PPSQ-23A (Shimadzu, Japan) according to the equipment manufacture's protocols.

Circular dichroism (CD) spectroscopy

In order to estimate the protein's secondary structure, the CD spectra of the fully purified protein, free of highly scattering particles, was studied. Far-UV CD spectra were generated using a 1 mm pathlength cuvette containing 200 μ L XF1729 protein sample at 0.1 mg/mL in 5 mM Tris–HCl buffer pH 7.5 at 20 °C. The assays were carried out in a Jasco 810 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan). The spectrum was presented as an average of four scans recorded from 190 to 250 nm, at a rate of 20 nm/min.

Enzyme assay

The enzymatic activity was measured spectrophotometrically in a thermostated Hewlett–Packard 8453 spectrophotometer at 25 °C. Aldo–keto reductase activity was determined by measuring the oxidation of NADPH at 340 nm using a molar extinction coefficient of 6220 M⁻¹ cm⁻¹ and DL-glyceraldehyde as the substrate. The assay conditions were as follows: 50 mM K₂HPO₄, 10 mM KCl and 0.5 mM EDTA at pH 7.0. The concentration of NADPH was kept constant at 250 μ M in all experiments and the concentration of purified XF1729 was 0.48 μ M. Twenty measurements were obtained in a substrate concentration range of 0.5–80 mM for DL-glyceraldehyde and 0.02–1 mM for 2-nitrobenzaldehyde. The values for K_m and K_{cat} were determined by a non-linear curve fitting of the data to the Michaelis–Menten equation with the program Origin (Microcal).

Small angle X-ray scattering

Small angle X-ray scattering experiments were performed at the SAXS beamline of the Laboratório Nacional de Luz Síncrotron (LNLS) in Campinas, Brazil. The wavelength used in the experiments was 1.488 Å, and the experimental setup included a temperature-controlled (15 °C) capillary tube sample holder [22] and a linear position sensitive detector. Solutions of purified XF1729 protein at a concentration of 2.39 mg/mL were used in the experiments. The buffer solution used was 10 mM Tris–HCl, pH 7.5. A 10 mg/mL albumin protein solution (67 kDa), measured in the same experimental conditions, was used as a secondary standard for molecular weight determination. Two samples to detector distances were used (933 and 336 mm), which gives a measured q range ($q = (4\pi/\lambda) \sin(\theta)$, λ = wavelength and 2θ = scattering angle) equal to 0.0112 Å⁻¹ < q < 0.2226 Å⁻¹. The data treatment of the scattering intensity was carried out using

the software package TRAT1D [23], performing the usual correction for detector homogeneity, incident beam intensity, sample absorption, blank subtraction, intensity averaging and the experimental error values. Data analysis and model calculations were performed using the GNOM [24], GASBOR [25], DAMAVER [26] and CRY SOL [27] computer programs.

Results and discussion

Strategy for XF1729 cloning and expression

The pET32Xa/LIC-XF1729 construction was successfully transformed into the BL21(DE3) strain. Therefore, the overexpression level of the recombinant protein was tested under various conditions. The induction of the protein expression was most efficient when induced for 4 h at 37 °C with a lactose concentration of 5.6 mM. Approximately 6 mg of soluble protein was obtained from 1 L of the induced bacterial culture. However, SDS-PAGE analysis indicated that the recombinant XF1729 protein was partly expressed in the soluble fraction and partly as inclusion bodies.

Protein purification and molecular weight determination of XF1729

A high amount of relatively pure recombinant protein containing the His-tag was recovered from the lysed extract in the first purification step using immobilized metal affinity chromatography (Fig. 2, lanes 2–4). The elution fractions contained the recombinant protein and a few impurities which could be completely eliminated in the last purification step. Approximately 3 mg of purified fusion recombinant protein per liter of initial bacterial broth were recovered at this step.

Subsequently, in the first purification, the cleavage of the N-terminal fusion tags from the target protein was performed by trypsin proteolysis due to the poor efficiency of the factor Xa. The fusion tags were successfully removed from XF1729 by trypsin cleavage. Analysis on SDS-PAGE of the proteolysis products are shown in Fig. 2 (lane 5) as two bands: the XF1729 (31 kDa) and the fusion tags (17.6 kDa). The N-terminal sequence of the cleaved protein was found to be: MKLDASLSGQFAIGG confirming its correct identity and the cleaved site of trypsin. The N-terminal sequencing was very important because it showed that trypsin protease completely removed the fusion tag proteins.

The second chromatographic step was performed to separate the fragment containing the tags and the target protein. In this case, the purified XF1729 was collected in the flowthrough fractions (Fig. 2, lane 6). The contaminant protein was also separated from the target protein during this step (Fig. 2, lane 7). The result indicated that the separation was achieved, since the flowthrough fractions were virtually free from proteins containing His-tag. These fusion tags were bounded in the column (Fig. 2, lane 7). The final amount of purified recombinant XF1729 obtained was approximately 1.6 mg/L of initial bacterial broth. Finally, the molecular mass of the purified protein determined by mass spectrometry was 31,023 Da, which is close to the theoretical value calculated from the primary amino-acid sequence (31,442 Da).

Sequence and secondary structure comparisons

The alignment of the XF1729 primary sequence with two structure solved proteins from *B. subtilis*, AKR11A and AKR11B, revealed many interesting features (Fig. 1). First, the three analyzed sequences have a similar molecular mass of about 35 kDa. Second, the active site architecture

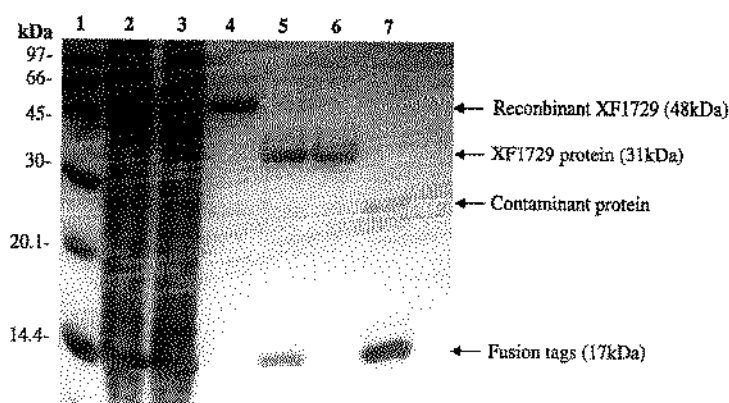


Fig. 2. SDS-PAGE of the fractions collected during the purification steps of the recombinant XF1729. Lane 1, molecular mass markers; lane 2, extract from bacterial lysis loaded onto the affinity chromatography column; lane 3, flow-through extract; lane 4, fraction collected during the elution step at imidazole concentration of approximately 100 mM, containing the recombinant protein (XF1729 + fusion tags); and lane 5, proteins resulted from limited trypsin digestion step for the XF1729 separation from fusion tags. This fraction was loaded onto the affinity chromatography column to promote final purification of the XF1729; lane 6, flowthrough fraction of the final purification step containing the XF1729 protein; lane 7, proteins eluted at 200 mM imidazole concentration containing the fusion tags and contaminant protein. Separation gel of 12% acrylamide concentration.

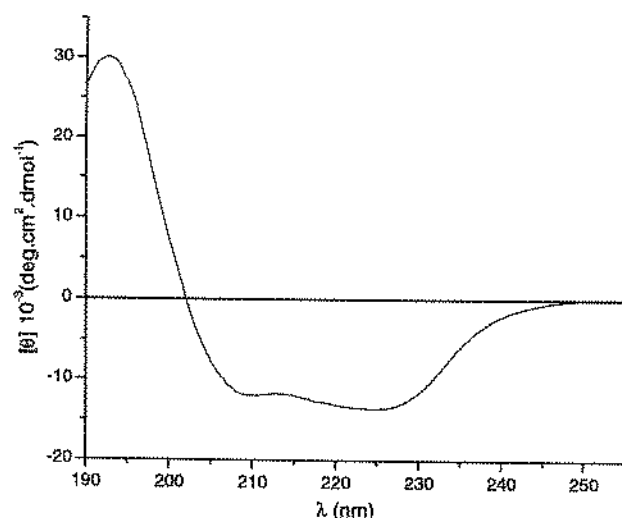


Fig. 3. Circular dichroism spectrum of the purified recombinant XF1729. The spectrum was recorded at 20 °C and a protein concentration of 0.1 mg/mL in 5 mM Tris buffer, pH 7.0. Data were collected using an average of 4 scans per replicate.

of the AKRs making up the catalytic tetrad Asp-53, Tyr-58, Lys-84 and His-125 (numbered according to the AKR11A) is observed also in the XF1729 protein. Moreover, the distributions of essential amino acid residues involved in cofactor binding are quite similar among the compared proteins. Third, the secondary structures predict contents are very close to both the AKR11A and AKR11B, being that the structural motif core (α/β) barrel is a notable characteristic among members of AKR family.

In order to investigate the structural integrity and secondary structure contents of the purified XF1729, the protein was analyzed by CD spectroscopy. The CD spectrum result is shown in Fig. 3 and indicates that the recombinant XF1729 contains a substantial amount of secondary structure. The protein presented a predominant signal of α -helices, with a positive band around 192 nm and negative ones at 209 and 222 nm. This result is in agreement with secondary structure prediction from the PSIPRED program (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>) that estimated 41.2% of α -helices and 19.6% β -strands for the XF1729 protein.

Kinetic results

The kinetic results show that the XF1729 has aldo-keto reductase activity and can use NADPH as a cosubstrate. The NADPH-dependent reductase activity of XF1729 has been confirmed with DL-glyceraldehyde and 2-nitrobenzal-

dehyde substrate (Table 1). However, the catalytic efficiency (K_{cat}/K_m) for 2-nitrobenzaldehyde substrate was more than 250-fold higher than that for DL-glyceraldehyde substrate. Some AKRs have been shown to use NADH, however, no activity was observed in XF1729 when assayed with this potential cosubstrate. Furthermore, XF1729 is unable to oxidize glucose with NADP as a cosubstrate.

Thus, comparing the activity with other AKRs by kinetics parameters, we verified that the XF1729 protein (classified as AKR13B1) is able to reduce the standard AKR substrate DL-glyceraldehyde like both the AKR11A and AKR11B enzymes. However, the YakC, a founding member of a 13th AKR family, showed no measurable activity with this substrate, but in contrast, the same enzyme exhibited an excellent activity with 2-nitrobenzaldehyde as did XF1729.

SAXS results

The experimental scattering intensity as a function of the modulus of the momentum transfer vector q is shown in Fig. 4A. The theoretical fit was done using the GNOM soft-

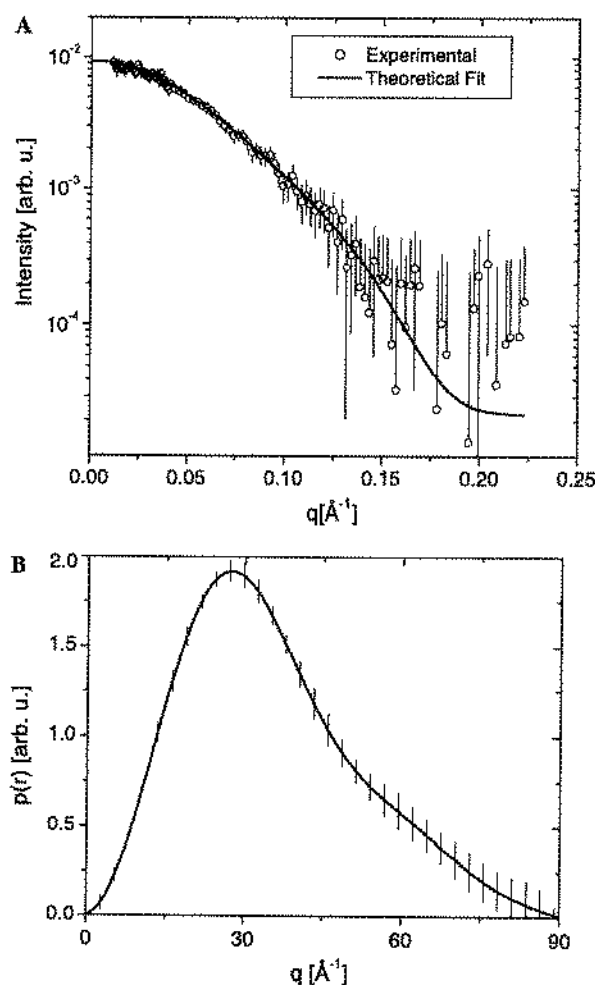


Fig. 4. (A) Experimental scattering intensity and theoretical fit (B) Pair distance distribution function $p(r)$ calculated from the scattering intensity.

Table 1
Kinetic parameters of XF1729 protein

Substrate	K_m (mM)	K_{cat} (s^{-1})
DL-Glyceraldehyde	8.20 ± 0.98	2.56
2-Nitrobenzaldehyde	0.15 ± 0.004	11.74
NADPH	0.012 ± 0.002	2.08

ware [24]. This program also calculates the inverse Fourier transform of the scattering intensity giving the pair distance distribution function $p(r)$ shown in Fig. 4B. From the comparison of the extrapolated zero angle scattering intensity value, given by the theoretical fit for the XF1729 protein and albumin scattering intensity, we obtained a value of (34 ± 2) kDa for the XF1729 protein molecular weight. This value is in very good agreement with the expected molecular weight (31.4 kDa) indicating that the protein is in monomeric form in solution. Also, from the theoretical fit, we obtained (27.5 ± 0.8) and $\sim 90 \text{ \AA}$ values for the particle radius of gyration and particle maximum dimension, respectively. The $p(r)$ function gives the indication that the protein possesses quite an elongated prolate shape.

We performed ab initio calculations to obtain more information about the protein shape. The modeling strategy used was based on chain like models. Starting from a spherical arrangement of a dummy backbone with the same number of protein residues, the GASBOR program uses simulated annealing optimization to search for the residues conformation that best fits the experimental data. From the low information nature of the SAXS technique, it is not possible to retrieve a unique solution from this modeling procedure. Due to this, we performed an average of the ab initio models in order to retrieve the most probable configuration, using the DAMAVER program [26]. In this procedure, the ab initio models are compared with each other by the SUPCOMB alignment program [28] and the models that have higher similarity are averaged. The most probable configuration is space-filled with a close packing of spheres. From a set of 10 independent models, we retrieved a most probable solution, shown in Fig. 5A, from the averaging process with spherical subunits radii of $2.25 \text{ \AA}$. As previously indicated by the $p(r)$ function, the protein has an anisotropic shape, but from the model we can see that it has a transversal diameter of $\sim 30 \text{ \AA}$ with a well-defined heavier domain in its conformation. One of the fits is shown in Fig. 5B (solid line). In all cases, the ab initio calculations show a very good fit with $\chi < 0.5$ (data not shown).

In order to compare the XF1729 protein with its similar sequences, we choose the AKR11A crystallized protein with the highest alignment score ($\sim 30\%$ sequence identity – PDB entry 1PYF) and compared it with our results. Using the atomic coordinates from a PDB file, the CRY-SOL program [27] was used to fit the experimental SAXS data, varying the protein excluded volume and hydration shell electronic contrast. The resulting fit of the XF1729 scattering data by the 1PYF crystallographic structure is shown in Fig. 5B (dashed line). The differences between the scattering intensity calculated for the crystallographic structure and that of the experimental data are the origin of some differences in the three-dimensional structure of the XF1729 protein and AKR11A. Fits using proteins with lower identity scores give similar (or worse) results (data not shown). The structural differences can be better seen in Fig. 5C where the resulting ab initio model (transparent spheres) was superposed with the 1PYF crystallo-

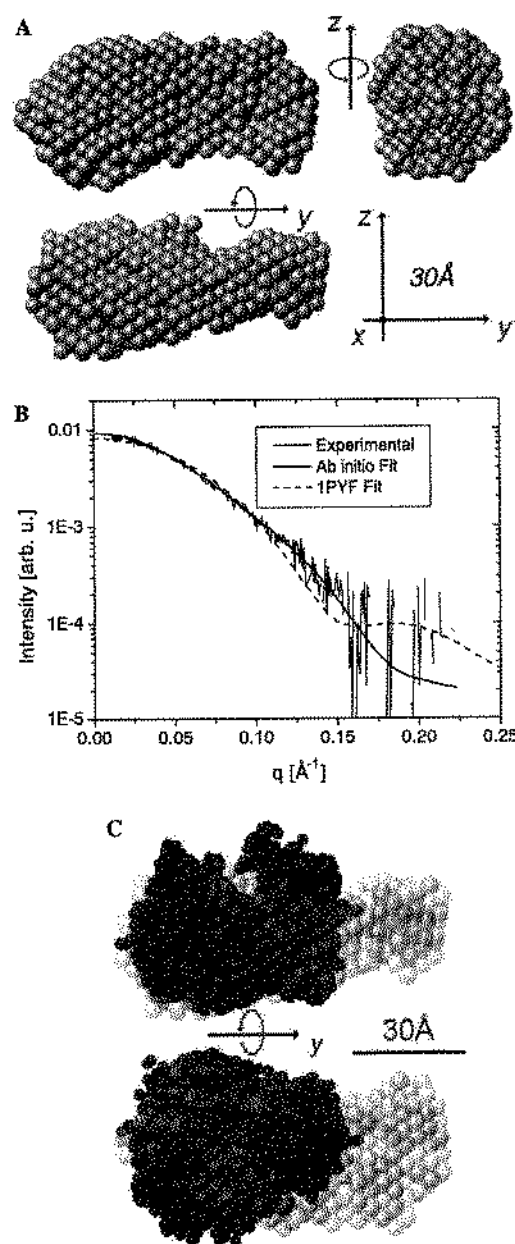


Fig. 5. (A) Most probable model for XF1729 protein given by a spatial average of 10 independent ab initio model calculations. (B) Fit by one model (solid line) and by the crystallographic structure 1PYF (dashed line). (C) Superposition of the crystallographic structure from 1PYF (solid spheres) and the SAXS model for XF1729 (transparent spheres).

graphic structure (solid spheres). It is interesting to note that the above mentioned heavier domain of the ab initio model is very similar to part of the crystallographic structure, differing in the final portion where the XF1729 ab initio model presents a longitudinal arrangement contrasting with the more globular arrangement of the crystallographic structure. This test may not be conclusive, but it is quite interesting to note that the proteins have a high shape similarity even for a low ($\sim 30\%$) sequence identity.

Oligomerization analysis

It is known that AKRs are typically monomeric proteins of about 35 kDa, but dimeric and tetrameric members do occur. In view of this fact, we investigated the oligomeric structure of the XF1729 protein. The molecular weight determined by gel filtration chromatography was 62 kDa, corresponding to double the molecular weight found in mass spectroscopy assays (~31 kDa), data not shown. Although the results seem to indicate that the protein forms oligomers in the observed conditions (50 mM Tris-HCl, pH 7.5), it is well known that elongated proteins give higher molecular weight values in these kinds of experiments. In this way, we have another indication that the XF1729 protein has a prolate shape, in agreement with the SAXS results.

Finally, our results showed that the protein is quite stable at room temperature and remained folded throughout the purification process, being suitable for crystallization studies aiming at three dimensional structure determinations. Also, from the scattering results and model calculations, we have the indication that the general protein shape is in some aspects similar to the proteins from AKR 11 family.

Conclusion

This is the first report of a putative aldo-keto reductase from the *X. fastidiosa* plant pathogen. In this work, the target protein was cloned, expressed in the soluble form and the expressed protein established its methods of purification. The protein was purified in two steps by immobilized metal affinity chromatography (agarose-IDA-Ni resin). Subsequently, its purity and identity were verified by SDS-PAGE, N-terminal sequencing 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.

Small angle X-ray scattering measurements furnish the general structural parameters for this protein (radius of gyration, maximum dimension and molecular weight) leading to the conclusion that the protein is a monomer in the solution. This indicates that the solution state of the XF1729 protein is different from the PadA similar enzyme from *E. coli*, which forms dimers under similar conditions. In addition, the initial characterization by gel filtration chromatography assays showed that the protein has an elongated shape, in agreement with the SAXS results. Moreover, the kinetics results show that the enzyme has an aldo-keto reductase activity allowing it to be classified as the AKR13B1. Several unknown function proteins were found to belong to this family, therefore the natural substrate of these enzymes, including XF1729, should be elucidated.

Acknowledgments

The authors thank Dr. H.S.S. de Araújo from Universidade Federal de São Carlos for the N-terminal sequence

analysis; Dr. M.N. Eberlin from Universidade Estadual de Campinas (UNICAMP) for the MALDI-TOF mass spectrometry; R.C. Sassonia M.Sc., and Dr. M.G. Oliveira for the use of the spectrophotometer and the Structural Molecular Biology Center (CEBIME) researchers from the National Synchrotron Light Laboratory (LNLS) for technical expertise. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 01/07533-7). L.K.R., C.L.P.O., A.R.A., S.F.S.T., C.F.C. and A.M.S. were recipients of fellowships from FAPESP and J.S.M.S. was recipient of fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A.P.S. is recipient of a research fellowship from CNPq.

References

- [1] J.M. Wells, B.C. Raju, H. Hung, W.G. Weisburg, L. Mandelco-Paul, D.J. Brenner, *Int. J. Syst. Bact. Apr.* (1987) 136–143.
- [2] R.F. Lee, K.S. Derrick, M.J.G. Beretta, C.M. Chagas, V. Rosetti, *Citros Ind. Oct.* (1991) 12–15.
- [3] A.J.G. Simpson, F.C. Reinach, P. Arruda, et al., *Nature* 406 (2000) 151–157.
- [4] L.R. Nunes, Y.B. Rosato, N.H. Muto, G.M. Yanai, V.S. Silva, D.B. Leite, E.R. Gonçalves, A.A. Souza, H.D. Coletta-Filho, M.A. Machado, S.A. Lopes, R.C. Oliveira, *Genome Res.* 13 (2003) 570–578.
- [5] A. Bhattacharyya, S. Stilwagen, G. Reznik, et al., *Genome Res.* 12 (2002) 1556–1563.
- [6] R.C. Oliveira, G.M. Yanai, N.H. Muto, D.B. Leite, A.A. de Souza, H.D. Coletta Filho, M.A. Machado, L.R. Nunes, *FEMS Microbiol. Lett.* 216 (2002) 15–21.
- [7] L.M. Moreira, R.F. de Souza, N.F. Almeida Jr., J.C. Setubal, J.C.F. Oliveira, L.R. Furlan, J.A. Ferro, A.C.R. da Silva, *Annu. Rev. Phytopathol.* 42 (2004) 163–184.
- [8] A. Ferrández, M.A. Prieto, J.L. Garcia, E. Diaz, *FEBS Lett.* 406 (1997) 23–27.
- [9] S.P. Hanlon, T.K. Hill, M.A. Flavell, J.M. Stringfellow, R.A. Cooper, *Microbiology* 143 (1997) 513–518.
- [10] J.M. Jez, T.M. Penning, *Chem.-Biol. Interact.* 130–132 (2001) 499–525.
- [11] D. Hyndman, D.R. Bauman, V.V. Heredia, T.M. Penning, *Chem.-Biol. Interact.* 143–144 (2003) 621–631.
- [12] O. White, J.A. Eisen, J.F. Heidelberg, et al., *Science* 286 (1999) 1571–1577.
- [13] T. Morita, T. Huruta, M. Ashiuchi, T. Yagi, *J. Biochem.* 132 (2002) 635–641.
- [14] A. Ehrensberger, D.K. Wilson, *Acta Crystallogr. D* 59 (2003) 375–377.
- [15] A. Ehrensberger, D.K. Wilson, *J. Mol. Biol.* 337 (2004) 661–673.
- [16] Xa/LIC Vector Kits Manual, ninth ed., Novagen, USA, 2000, 56 pp (www.novagen.com).
- [17] B.T. Kurien, R.H. Scofield, *BioTechniques* 18 (6) (1995) 1023–1026.
- [18] 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, *Protein Expr. Purif.* 33 (2004) 297–303.
- [19] U.K. Laemmli, *Nature* 227 (1970) 680–685.
- [20] M.M. Bradford, *Anal. Biochem.* 72 (1976) 248–254.
- [21] C.N. Pace, F.X. Schmidt, in: T.E. Creighton (Ed.), *Protein Structure, A Practical Approach*, IRL Press, New York, 1997, pp. 253–259.
- [22] L.P. Cavalcanti, I.L. Torriani, T.S. Plivelic, C.L.P. Oliveira, Kellermann, R. Neuenschwander, *Rev. Sci. Instr.* 75 (2004) 4541–4546.
- [23] C.L.P. Oliveira, TRAT1D – Computer Program for SAXS Data Treatment. LNLS Technical Manual MT 01/2003.
- [24] V. Semenyuk, D.I. Svergun, *J. Appl. Crystallogr.* 24 (1991) 537–540.
- [25] D.I. Svergun, M.V. Petoukhov, M.H.J. Koch, *Biophys. J.* 80 (2001) 2946–2953.
- [26] V.V. Volkov, D.I. Svergun, *J. Appl. Cryst.* m36 (2003) 860–864.
- [27] D.I. Svergun, C. Barberato, M.H.J. Koch, *J. Appl. Cryst.* 28 (1995) 768–773.
- [28] M.B. Kozin, D.I. Svergun, *J. Appl. Cryst.* 34 (2001) 33–41.

5. RESULTADOS COMPLEMENTARES

5.1. TESTES DE CRISTALIZAÇÃO DA PROTEÍNA XF1729

Amostras da proteína XF1729 purificada foram submetidas a testes iniciais de cristalização utilizando-se os *kits* comerciais *Crystal Screen I e II* (Hampton Research) através do método de *sitting drop*, a temperatura de 18°C. Em uma das condições do *kit Crystal Screen II* testadas (1,6 Hexanodiol, pH 4,5) observou-se o surgimento de esferolitas, representando um ponto de partida para o refinamento com a finalidade de obter cristais. Para tanto, iniciou-se o primeiro refinamento da condição em que foram encontradas as esferolitas, fazendo-se um gradiente de pH de 4 a 9. Após 48 horas as placas foram inspecionadas, verificando a presença de esferolitas em todas as condições de pH testadas, embora no intervalo de pH entre 7,5 e 9,0 (Fig. 5) a gota estivesse translúcida, com esferolitas, mas sem precipitado ao redor, indicando que a proteína se comporta melhor em pHs maiores.

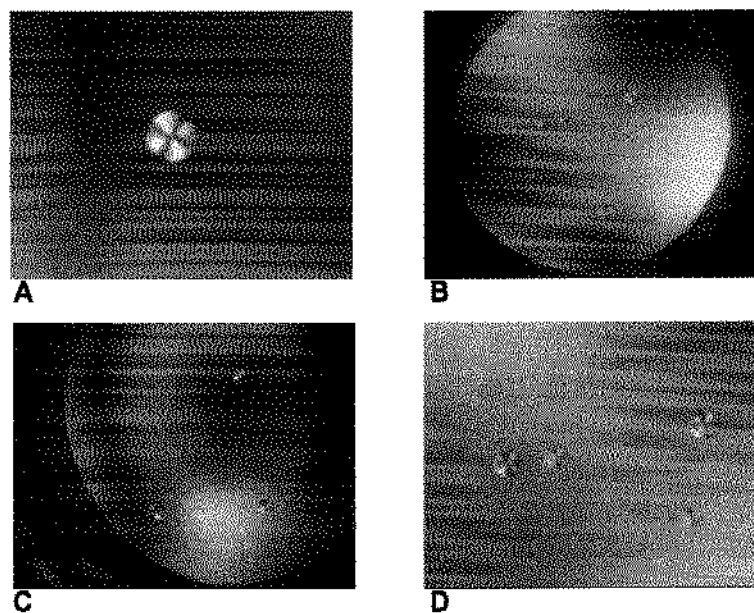


Figura 5: Esferolitas da proteína XF1729 encontradas no tampão 1,6 Hexenodiol A) pH 7,5 B) pH 8,0 C) pH 8,5 D) pH 9,0.

Amostras da proteína XF1729, purificada e complexada ao cofator NADPH, foram submetidas a testes de cristalização utilizando-se os *kits* comerciais *Crystal Screen I e II* (Hampton Research), *JB Screen* de 1 a 10 (Jena Bioscience) e *Salt Rx I e II* (Hampton Research), através da técnica de difusão de vapor por *hanging-drop*. A temperatura de cristalização testada foi de 18°C.

Todos os testes foram feitos adicionando-se 1,5 µl da proteína a 12 mg/ml, 0,4 µl de NADPH a 50mM e 1,5 µl da solução do poço para a formação da gota. A gota foi suspensa sobre 100 µl de solução contida no poço.

Após a realização dos testes de cristalização da proteína complexada ao NADPH, observou-se a formação de cristais nas seguintes condições:

- a) Kit *JB Screen 9*, condição C3: 1,6 M de tartarato de sódio/potássio, 0,1 M de Na MES, pH 6,5;
- b) Kit *Salt Rx I*, condição D3: 1,2 M de citrato de sódio, 0,1 M de Bis-Tris propano, pH 7;
- c) Kit *Salt Rx II*, condição B4: 1,8 M de fosfato de sódio/potássio, pH 8,2;
- d) Kit *JB Screen 3*, condição D6: 32% PEG 4000, 0,8 M de cloreto de lítio, 0,1 M de Tris-HCl, pH 8,5.

Os cristais obtidos nas condições a, b, e c foram corados com azul de metileno (comercialmente conhecido por IZIT), confirmando que eram cristais de proteína e não de sal (Fig. 6). Na condição d, obtivemos cristais de sal identificados pela não coloração pelo corante (vide Fig. 6G). No entanto, não observamos em nenhuma das condições a formação de um único cristal, necessário para a coleta dos dados de difração de raios-X. A qualidade dos cristais obtidos dever ser otimizada através da variação de temperatura, pH, concentração de sais e precipitantes, a fim de obter um monocristal adequado para os testes de difração.

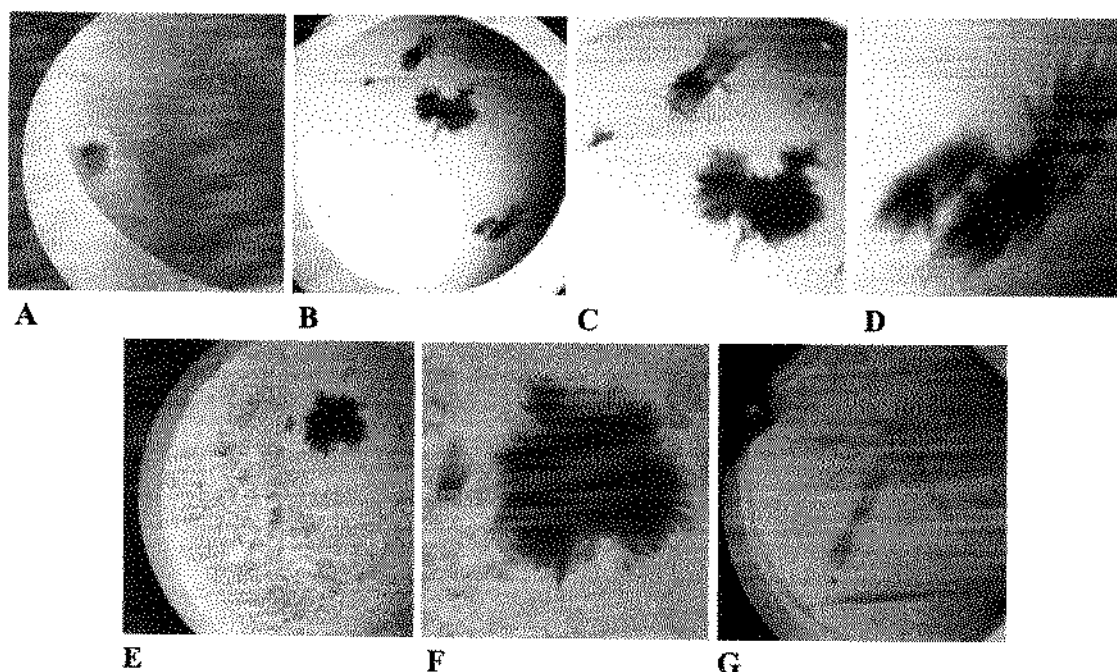


Figura 6: Fotos dos cristais obtidos nos ensaios cristalográficos da proteína XF1729 complexada ao cofator NADPH. Cristais de proteína corados com IZIT: A-F. **A)** Condição D3, kit *Salt RX I*. **B)** Condição B4, kit *Salt RX II*. **C)** Ampliação dos cristais situados na parte superior da foto B. **D)** Ampliação do cristal situado na parte inferior da foto B. **E)** Condição C3 *JB Screen 9*. **F)** Ampliação do cristal da foto E. **G)** Cristal de sal obtido na condição D6, kit *JB Screen 3*. Observa-se que a adição do IZIT não corou o cristal.

5.2. ENSAIOS DE EXPRESSÃO E PURIFICAÇÃO DA PROTEÍNA CODIFICADA PELA ORF XF1934

A ORF XF1934, de 595 pb, codifica uma proteína HetI de 22,4 kDa e 197 resíduos de aminoácidos. Estudos mostram que esta proteína requerida para a manutenção do crescimento vegetativo (Black & Wolk, 1994), uma vez que desempenha um papel fundamental na síntese na da membrana biológica da célula.. Um trabalho recente de resolução da estrutura tridimensional *in silico* mostra que esta enzima HetI é uma das duas enzimas essenciais para síntese de ácidos graxos via acetato na bactéria (Osiro *et al.*, 2004).

A sequência correspondente a *ORF* XF1934 foi clonada no vetor pET32Xa/LIC (Novagen). A indução da expressão da proteína heteróloga foi feita em linhagem de *E. coli* BL21(DE3) a temperatura de 37°C, sob agitação de 300 rpm e 5,6 mM de lactose. Após 4 horas de indução a cultura foi centrifugada e as células separadas para a extração protéica. As células foram ressuspensas em tampão (Tris HCl 50 mM, NaCl 300 mM, pH 7,5) e rompidas através da adição de lisozima (1 mg/ml) e sonicação, adicionando-se ao extrato inibidor de protease PMSF (Phenylmethylsulfonylfluoride) a 1mM. O extrato foi centrifugado e o sobrenadante (extrato solúvel) utilizado para carregar 1,5 ml de resina agarose-Ni (Ni-NTA, Quiagen, EUA). A proteína heteróloga foi capturada pela cromatografia de afinidade, como mostrado na Figura 7, sendo eluída desde concentrações mais baixas de imidazol, juntamente com grande quantidade de contaminantes, até concentrações mais altas de imidazol em que a proteína é eluída com menos contaminantes.

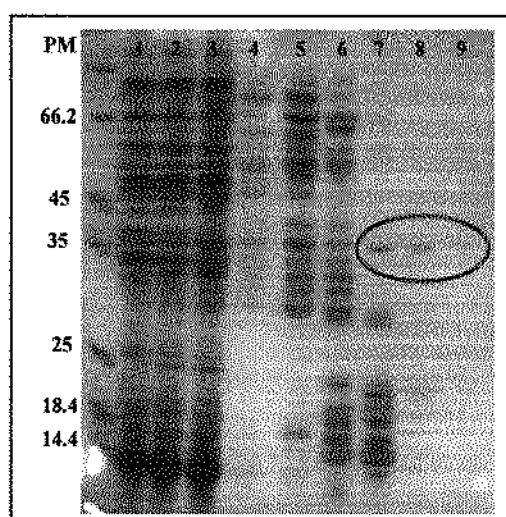


Figura 7: Gel SDS-PAGE 12% apresentando os resultados da expressão e purificação de 500 ml de cultura da proteína heteróloga XF1934 (~40kDa). Legenda: PM: marcador de peso molecular em kDa; 1: Extrato da lise bacteriana; 2 e 3: Extrato que passou pela coluna de afinidade; 4: Lavagem com tampão sem imidazol; 5: Eluição com 10 mM de imidazol; 6: Eluição com 50 mM de imidazol; 7: Eluição com 100 mM de imidazol; 8 e 9: Eluição com 200 mM de imidazol. Proteína indicada pelo círculo vermelho.

A amostra eluída com 200mM de imidazol foi dialisada e utilizada para os testes de proteólise para a remoção da proteína de fusão de aproximadamente 17 kDa. O ensaio de proteólise realizado testou a eficiência da enzima fator Xa (Novagen) a 0,1; 0,2 e 0,5 U/ μ l. A reação foi feita em tampão Tris HCl 50 mM, NaCl 100 mM, CaCl₂ 5 mM, pH 7,8, contendo 10 μ g da proteína heteróloga XF1934 e fator Xa para cada concentração testada. Aliquotas de cada reação foram retiradas nos intervalos de tempo de 2, 8 e 16 horas e analisadas em gel de SDS-PAGE (Fig. 8). O teste apresentado mostrou que a protease (fator Xa) cliva inespecificamente a proteína heteróloga, verificando-se a formação de um produto de maior peso molecular que o esperado.

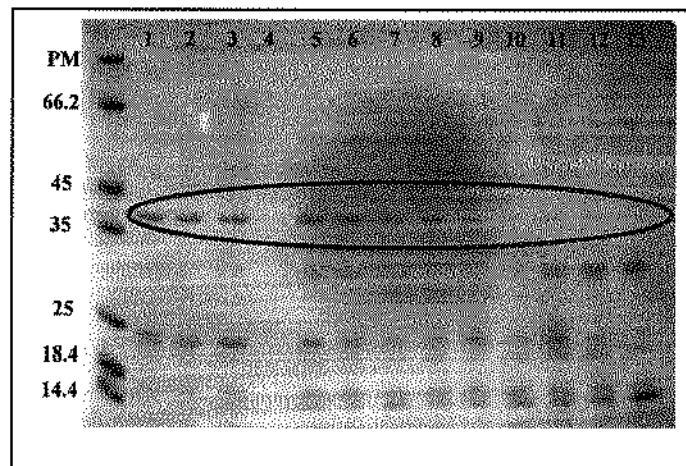


Figura 8: Gel SDS-PAGE 12% apresentando os resultados do ensaio de hidrólise da proteína XF1934 utilizando fator Xa, a temperatura ambiente. Foram estudadas três concentrações de enzima e três diferentes tempos de reação de hidrólise. Legenda: **PM:** marcador de peso molecular em kDa; **1, 2 e 3:** Controle negativo contendo proteína sem a protease, nos tempos 2, 8 e 16 h, respectivamente; **4:** poço vazio; **5, 6 e 7:** Proteína na presença de 0,1 U/ μ l de fator Xa, nos tempos 2, 8 e 16 h, respectivamente; **8, 9 e 10:** Proteína na presença de 0,2 U/ μ l de fator Xa, nos tempos 2, 8 e 16 h, respectivamente; **11, 12 e 13:** Proteína na presença de 0,5 U/ μ l de fator Xa, nos tempos 2, 8 e 16 h, respectivamente. O círculo azul indica a proteína heteróloga XF1934, não clivada.

5.3. ENSAIOS DE EXPRESSÃO DA PROTEÍNA CODIFICADA PELA ORF XF1532

A proteína codificada pela ORF XF1532 apresenta 325 resíduos de aminoácidos em sua sequência e um peso molecular predito de 36,8 kDa. O alinhamento de sua sequência primária revelou 74% de similaridade com o produto de gene *oxyR* de *Xanthomonas campestris*, conforme análise extraída do programa BLASTP (www.ncbi.nlm.nih.gov). Este gene (*oxyR*) codifica um regulador transcricional de estresse oxidativo. A proteína OxyR de *E. coli* pertence à família LysR de reguladores transcricionais (Tao *et al.*, 1989), já tendo sido purificada e caracterizada (Tao *et al.*, 1991) e resolvida sua estrutura tridimensional (Choi *et al.*, 2001). No entanto, sua similaridade com a proteína codificada pela ORF XF1532 de *X. fastidiosa* é inferior a 44%. A família de proteínas LysR se caracteriza por se ligar ao DNA e a proteína OxyR apresenta uma função regulatória, sendo um ativador transcricional de genes importantes na defesa contra o estresse oxidativo. Na presença de peróxido de hidrogênio, OxyR induz a expressão de inúmeros genes envolvidos no processo de detoxificação da bactéria.

A ORF XF1532 foi clonada no vetor pET32Xa/LIC (Novagen) e a expressão da proteína heteróloga foi testada em diferentes linhagens de *E. coli*: BL21(DE3), BL21(DE3)pLysS e BL21-SI, consideradas linhagens padrão de teste.

Inicialmente, testou-se a expressão da proteína heteróloga XF1532, de aproximadamente 55 kDa, na linhagem de expressão BL21(DE3). As células contendo o plasmídeo foram cultivadas à 37°C em meio LB com ampicilina até atingir uma absorbância (A_{600nm}) de 0,6. Neste ponto foi adicionado o indutor IPTG na concentração final de 1mM. As células foram coletadas neste momento e após 1, 2, 3 e 4 horas de indução. Em seguida realizou-se a preparação dos extratos (solúvel e insolúvel) para avaliação dos níveis de expressão em géis de poliacrilamida

(SDS-PAGE). Nestas condições verificou-se um alto nível de expressão da proteína heteróloga XF1532 (aproximadamente 55 kDa) no extrato insolúvel, como apresentado na Figura 9.

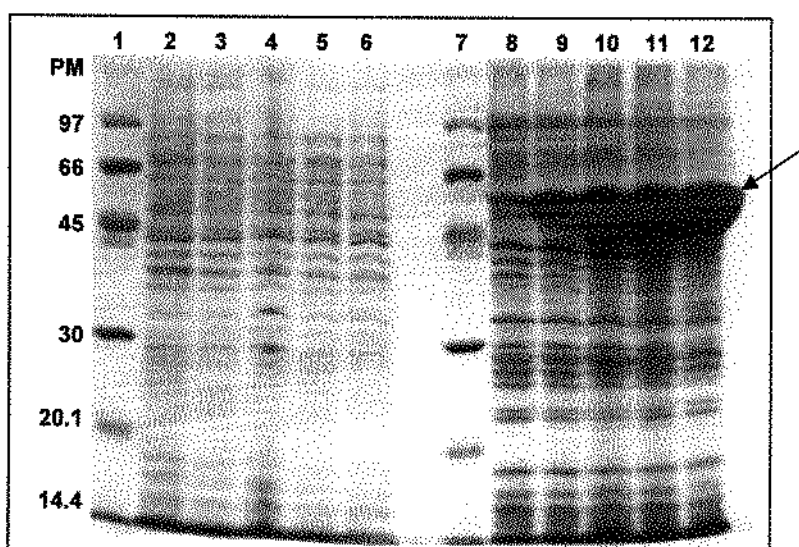


Figura 9: Gel SDS-PAGE 12% apresentando o padrão de expressão da proteína heteróloga XF1532 nas frações solúvel e insolúvel. A linhagem testada foi BL21(DE3) a 37°C e 1mM de IPTG. Legenda: PM: marcador de peso molecular em kDa; 1 e 7: marcador de peso molecular; 2, 3, 4, 5 e 6: Extrato da fração solúvel em 0, 1, 2, 3 e 4 horas de indução, respectivamente; 8, 9, 10, 11 e 12: Extrato da fração insolúvel em 0, 1, 2, 3 e 4 horas de indução, respectivamente. A seta indica a proteína heteróloga XF1532 (~55 kDa) expressa na fração insolúvel.

Em seguida, testou-se esta mesma linhagem de *E.coli* BL21(DE3), porém crescendo-a a uma temperatura mais baixa (25°C), com o intuito de fazer com que a bactéria produzisse mais lentamente a proteína heteróloga e assim tornasse a proteína solúvel, não havendo a formação de corpúsculos de inclusão. No entanto, as proteínas continuaram insolúveis. Novos testes de expressão em baixa escala foram realizados na linhagem hospedeira BL21(DE3)pLysS, induzida com 1 mM de IPTG por 2 e 4 horas, a 37°C. Verificou-se também a expressão das proteínas desejadas apenas no extrato insolúvel, embora em níveis mais baixos (Fig. 10).

Por último, testou-se a linhagem BL21-SI utilizando um gradiente de concentração do indutor (100, 300 e 500 mM de NaCl) a temperatura de 28°C e a 15°C (*overnight*), mas nenhuma alteração significativa foi verificada em nenhuma das condições testadas.

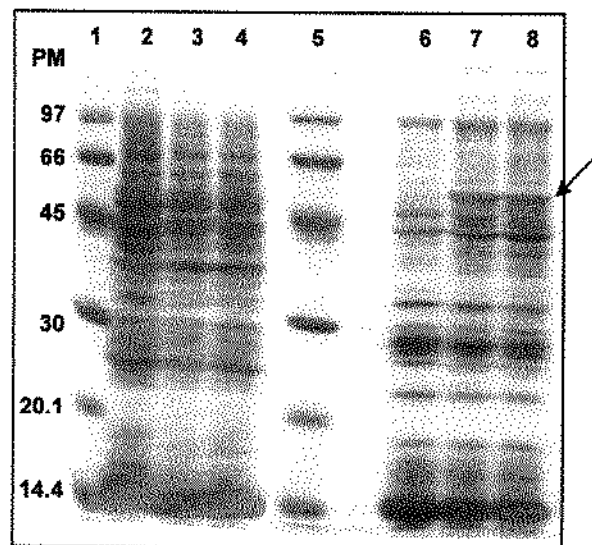


Figura 10: Gel SDS-PAGE 12% apresentando o padrão de expressão da proteína heteróloga XF1532 nas frações solúvel e insolúvel. A linhagem testada foi BL21(DE3)pLysS a 37°C e 1mM de IPTG. Legenda: **PM:** marcador de peso molecular em kDa; **1 e 5:** marcador de peso molecular; **2, 3 e 4:** Extrato da fração solúvel em 0, 2 e 4 horas de indução, respectivamente; **6, 7 e 8:** Extrato da fração insolúvel em 0, 2 e 4 horas de indução, respectivamente. A seta indica a proteína heteróloga XF1532 (~55 kDa) expressa em menor quantidade na fração insolúvel.

6. PERSPECTIVAS

6.1. PROTEÍNA XF1729 (AKR13B1)

- Refinamento das condições em que foram encontrados cristais protéicos com a finalidade de obter um monocristal adequado para a difração de raios-X, coletando-se os dados e parâmetros estatísticos necessários à resolução da estrutura tridimensional.

- Ensaio funcionais *in vivo* através do uso de anticorpos policlonais, avaliando as condições em que a enzima é expressa.

- Estudo dos possíveis substratos naturais

6.2. PROTEÍNA XF1934

- Ensaio de novas condições de clivagem do fator Xa, de modo a se obter a remoção correta e completa da proteína de fusão.

- Experimentos de proteólise limitada a fim de se detectar a presença de algum domínio estável, resistente à proteólise, na proteína em estudo.

- Concluir seu processo de purificação com o objetivo de produzir grande quantidade de proteína com alto grau de pureza para os testes de cristalização.

- Produção de anticorpos policlonais para estudos funcionais *in vivo*.

- Caracterização bioquímica funcional para elucidar seu papel no metabolismo de *Xylella fastidiosa*.

6.3. PROTEÍNA XF1532

– Ensaios de extração em condições desnaturantes da proteína XF1532-Trx-His-S-Tag, purificação e reenovelamento.

– Clonagem em outros vetores de expressão, como por exemplo: pET28a, pSV282 e pETSUMO para expressão em diferentes linhagens de *E. coli*, testando a solubilidade da proteína fusionada a 6 resíduos de histidina (pET28a), à MBP – *Maltose Binding Protein* (pSV282), à proteína SUMO (pETSUMO), respectivamente.

7. CONCLUSÕES

As AKRs (*Aldo-Keto Reductases*) são uma crescente família de oxidorreduções NADPH-dependentes (Jez *et al*, 1997) encontradas ao longo de todos os reinos e filos biológicos. São encontradas tanto em procariotos quanto em eucariotos, incluindo leveduras, plantas, anfíbios e mamíferos. Apesar de muitos dímeros e tetrâmeros terem sido caracterizados, os membros desta superfamília são tipicamente proteínas monoméricas, de aproximadamente 35kDa, que apresentam o domínio $(\alpha\beta)_8$ -barrel (Fig. 11) e que utilizam NAD(P)(H) para catalisar a redução de aldeídos, cetonas, monossacarídeos, cetoesteróides, prostaglandinas e outros substratos específicos. Além disto, o sítio catalítico destas enzimas contém uma tétrade característica formada por uma tirosina, uma lisina, um aspartato e uma histidina (Jez & Penning, 2001).

Originalmente, os membros desta superfamília de enzimas eram identificados através dos nomes dados às enzimas, os quais correspondiam à sua especificidade ao substrato. No entanto, este sistema adotado gerou inúmeros problemas, sendo então criado em 1996 um sistema de nomenclatura para organizar os membros desta superfamília. Este novo sistema de nomenclatura baseado na identidade da sequência de aminoácidos foi aceito no *Eigth International Symposium on Enzymology and Molecular Biology of Carbonyl Metabolism* (Jez *et al*, 1997), sendo um método consistente para registrar e estabelecer novos membros à superfamília AKR. O web site www.med.upenn.edu/akr promove a atualização da superfamília AKR, onde as informações de alinhamento de sequências, nomes comuns, estrutura e potenciais membros são avaliados (Hydman *et al.*, 2003).

Atualmente, a superfamília AKR contém mais de 120 membros que estão distribuídos em 14 famílias (AKR1-AKR14). A família AKR1 é a maior família com um enorme número de

membros bem caracterizados. Em contrapartida, uma das mais recentes famílias, AKR13, é pobremente caracterizada através de estudos estruturais e seus substratos naturais ainda permanecem desconhecidos, sendo representada por uma aldose redutase de *Schizosaccharomyces pombe* (Morita *et al*, 2002).

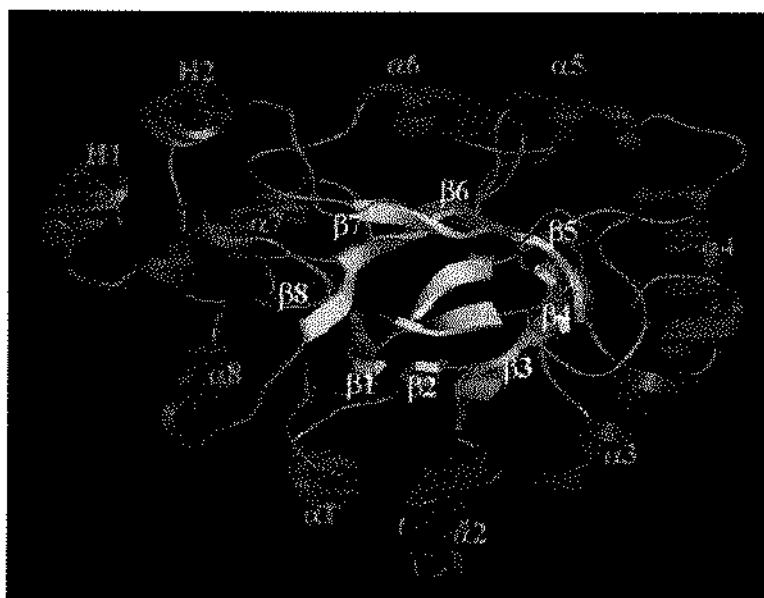


Figura 11: O enovelamento $(\alpha\beta)_8$ -Barrel das aldo-ceto redutases (AKRs). (Fonte: http://www.med.upenn.edu/akr/akr_graphic.shtml)

O trabalho realizado relata a primeira oxidoredutase da bactéria fitopatogênica *Xylella fastidiosa*. O estudo de clonagem, expressão em *Escherichia coli* e purificação da proteína codificada pela ORF XF1729 foi desenvolvido para sua caracterização funcional e estrutural de baixa resolução. O alinhamento da proteína XF1729 com aldo-ceto redutases de estrutura terciária conhecida permitiu desvendar características interessantes: Primeiro, a semelhança de peso molecular (~35 kDa); Segundo, a arquitetura similar do sítio catalítico com a presença da tétrade catalítica (Asp, Tyr, Lys e His) e a distribuição dos aminoácidos envolvidos na ligação ao cofator; Terceiro, a composição da estrutura secundária predita ser extremamente próxima à

composição das AKRs, a qual caracteriza-se pela presença do enovelamento estrutural do tipo $(\alpha\beta)_8$ -Barrel. Além disto, a composição da estrutura secundária da proteína XF1729, estimada por espectroscopia de dicroísmo circular, convalidou tais análises teóricas. As medidas de espalhamento de raios-X a baixo ângulo (SAXS) forneceram os parâmetros estruturais, levando a concluir que a proteína apresenta-se como um monômero em solução, tal como a grande maioria das AKRs. Por fim, os resultados de cinética enzimática mostraram que a proteína apresenta típica atividade de aldo-ceto redutase, permitindo ser classificada como AKR13B1.

Em suma, o trabalho realizado permitiu concluir que:

- O genoma da *X. fastidiosa* codifica uma AKR funcional;
- A proteína codificada pela ORF XF1729 de *X. fastidiosa* é uma enzima pertencente à superfamília das aldo-ceto redutases (*AKR Superfamily*);
- Os resultados obtidos sugerem correções na anotação genômica para esta ORF. Uma vez que o produto gênico da ORF XF1729 está anotado no banco de dados do genoma da *X. fastidiosa* como sendo uma fenilacetaldeído desidrogenase;
- A proteína XF1729 é um membro fundador da sub-família B, sendo classificada como AKR13B1.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Andrade M.A., Chacón P., Merelo J.J., Morán F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised neural network. *Protein Engeneering*. 6:383-390.
- Avery, S.V. (2001) Metal toxicity in yeasts and the role of oxidative stress. *Advances in Applied Microbiology* 49:111-142.
- Azzoni A.R., Tada S.F.S., Rosselli L.K., Paula D.P., Catani C.F., Sabino A.A., Eberlin M.N., Barbosa J.A.R.G., Guimarães B.G., Medrano F.J., Souza A.P. (2004) Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*. *Protein Expression & Purification*. 33:297-303.
- Bhattacharyya A., Stilwagen S., Reznik G., Feil H., Feil W.S., *et al.* (2002). Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. *Genome Research*. 12:1556-63.
- Black T.A., Wolk C.P. (1994) Analysis of a *hef* mutation in *Anabaena* sp. Strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing. *Journal of Bacteriology* 181(20):6254-6263.
- Blattner, F. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453-1474.
- Brlansky R.H., Timmer I.W., French W.J., McCoy R.E. (1983) Colonization of the sharpshooter vectors, *Oncometopia igricans* and *Homalodisca coagulata*, by xylem-limited bacteria. *Phytopathology* 73:530-535.
- Carvalho M.L.V., Bereta M.J.G., Chagas C.M., Harakawa R., Palazzo D.A., Rosseti V. (1996) Clorose variegada dos citrinos: etiologia e manejo. *São Paulo: Instituto Biológico*, 1996. 13p (Boletim Técnico, 3).
- Catani C.F., Azzoni A.R., Paula D.P., Siqueira S.F., Rosselli L.K., Souza A.P., Yano T. (2004) Expression and purification of the virulence-associated protein D of *Xylella fastidiosa*. *Protein Expression & Purification* 37:320-6.
- Chagas C.M., Rossetti V., Beretta M.J.G. (1992) Electron microscopy studies of a xylem – limited bacterium in sweet orange affected with citrus variegated chlorosis disease in Brazil. *Journal of Phytopathology*. 134:306-312.
- Choi H., Kim S., Mukhopadhyay P., Cho S., Woo J., Storz G., Ryu S. (2001) Structural basis of the redox switch in the OxyR transcription factor. *Cell* 105:103-113.
- Chothia C., Lesk A.M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*. 5:823-826.

- Coletta Filho H.D., Machado M.A. (2002) Evaluation of the genetic structure of *Xylella fastidiosa* populations from different citrus sinensis varieties. *Applied and Environmental Microbiology* 68(8):3731-3736.
- Daniels M.J., Collinge D.B., Dow J. M., Osbourn A.E., Roberts I.N. (1987) Molecular genetics of pathogenicity in phytopathogenic bacteria. *Annual Review of Phytopathology* 25:353-359.
- Dow J.M., Daniels M.J. (2000) *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* 17:263- 271.
- Dow J.M., Feng J.X., Barber C.E., Tang J.L., Daniels M.J. (2000) Novel genes involved in the regulation of the pathogenicity factor production within the rpf gene cluster of *Xanthomonas campestris*. *Microbiology* 146(4):885-891.
- Fasman G.D. (1996) Circular dichroism and the conformational analysis of biomolecules. New York: Plenum Press.
- Glatter O., Kratky O. (1982) Small angle X-ray scattering. Londres: Academic Press.
- Goffeau A. et al. (1996) Life with 6000 genes. *Science* 274:563-567.
- Goheen A.C., Nyland G., Lowe S.K. (1973) Association of a rickettsia-like organism with Pierce's disease of grapevines and alfalfa dwarf and heat therapy of the disease in grapevines. *Phytopathology* 63:341-345.
- Grossmann J.G. (2007) Biological solution scattering: recent achievements and future challenges. *Journal of Applied Crystallography* 40.
- He C.X., Li W.B., Ayres A.J. (2000) Distribution of *Xylella fastidiosa* in citrus rootstocks and transmission of citrus variegated chlorosis between sweet orange plants through natural root grafts. *Plant Disease* 84(6):622-626.
- Hindman D., Bauman D.R., Heredia V.V., Penning T.M. (2003) The aldo-keto reductase superfamily homepage. *Chemico-Biological Interactions* 143-144:621-631.
- Hopkins D.L. (1989) *Xylella fastidiosa*: xylem-limited bacterial pathogen of plants. *Annual Review of Phytopathology* 27: 271-290.
- Hopkins D.L., Mollenhauer H.A., French W.J. (1973) Occurrence of a rickettsia-like bacterium in the xylem of peach trees with phony disease. *Phytopathology* 63:1422-1423.
- Hopkins D.L., Thompson C.M., Wichman R.L., et al. (1995) Effect of inoculation of mature citrus trees in the grove with *Xylella fastidiosa* on citrus blight incidence. *Proc. Florida. State Horticulture Society* 108: 103-106.
- Hur E., Wilson, D.K. (2000) Crystallization and aldo-keto reductase activity of Gcylp from *Saccharomyces cerevisiae*. *Acta crystallographica. Section D, Biological crystallography* 56:763-765.

- Jez J.M., Flynn T.G., Penning T.M. (1997) A new nomenclature for the aldo-keto reductase superfamily. *Biochemical Pharmacology* 54:639-647.
- Jez J.M., Penning T.M. (2001) The aldo-keto reductase (AKR) superfamily: an update. *Chemico-Biological Interactions* 130-132:499-525.
- Johnson W.C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function, and Genetics*. 35:307-312.
- Jornvall H., Hoog J., Persson B. (1999) SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Letter* 445:261-264.
- Jornvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D. (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 34:6003-6013.
- Kataoka M., Shimizu S., Yamada, H. (1992) Distribution and immunological characterization of microbial aldehyde reductases. *Archives of Microbiology* 157:279-283.
- Keen N.T., Dumenyo C.K., Yang C., Cooksey D.A. (2000) From rags to riches: insights from the first genomic sequence of a plant pathogenic bacterium. *Genome Biology* 1(3):1019.1-1019.4.
- Kelly S.M., Jess T.J., Price N.C. (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta* 1751:119-139.
- 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.
- Khurana S., Powers D.B., Anderson S., Blaber M. (1998) Crystal structure of 2,3-diketo-D-gluconic acid reductase A complexed with NADPH at 2.1-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* 95:6768-6773.
- Lambais M.R., Goldman M.H., Camargo L.E., Goldman G.H. (2000) A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity. *Current in Microbiology* 3(5):459-462.
- Lee H. (1998) The structure and function of yeast xylose (aldose) reductases. *Yeast* 14: 977-984.
- Lee J.M. (1992) Biochemical Engineering. New Jersey: Prentice Hall.
- Lee R.F., Derrick K.S., Beretta M.J.G., Chagas C.M., Rosetti V. (1991) Citrus variegated chlorosis: a new destructive disease of citrus in Brazil. *Citrus Industry* 72(October):12-15.
- Lehninger A.L., Nelson D.L., Cox M.M. (1993) Principles of biochemistry. Second edition. New York: Worth Publishers.

- Lipfert J., Doniach S. (2007) Small-angle X-ray scattering from RNA, protein, and protein complexes. *The Annual Review of Biophysics and Biomolecular Structure* 36:307-327.
- Machado E.C., Quaggio J.A., Lagôa A.M.M.A., Ticelli M., Furlani P.R. (1994). Gas exchange and water relations of orange trees with citrus variegated chlorosis. *Revista Brasileira de Fisiologia Vegetal* 6:53-57.
- Machado M.A., Souza A.A., Coletta Filho H.D., Izioka E.E.K., Takita M.A. (2001) Genome and pathogenicity of *Xylella fastidiosa*. *Molecular Biology Today* 2(3):33-43.
- Manavalan P., Johnson Jr. W.C. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Analytical Biochemistry* 167:76-85.
- Marnett L.J., Hurd H.K., Hollstein M.C., Levin D.E., Esterbauer H., Ames B.N. (1985) Naturally occurring carbonyl compounds are mutagens in Salmonella tester strain TA104. *Mutation Research* 148:25-34.
- McElrone A.J., Sherald J.L., Forseth I.N. (2001) Effects of water stress on symptomatology and growth of *Parthenocissus quinquefolia* infected by *Xylella fastidiosa*. *Plant Disease* 85(11): 1160-1164.
- Michelmore R. (2000) Genomic approaches to plant disease resistance. *Current Opinion in Plant Biology* 3:125-31.
- Mircetich S.M., Lowe S.K., Moller W.J., Nyland G. (1976) Etiology of almond leaf scorch disease and transmission of the causal agent. *Phytopathology* 66:17-24.
- Moreira L.M., de Souza R.F., Almeida Jr N.F., Setubal J.C., Oliveira J.C.F., Furlan L.R., Ferro J.A. and da Silva A.C.R. (2004) Comparative genomics analyses of citrus-associated bacteria. *Annual Review of Phytopathology* 42:163-84.
- Morita T., Huruta T., Ashiuchi M., Yagi T. (2002) Characterization of recombinant YakC of *Schizosaccharomyces pombe* showing YakC defines a new family of aldo-keto reductases. *The Journal of Biochemistry* 132:635-641.
- Naqui A., Chance B.A. (1984) Semi-integrated method for the determination of the enzymes kinetic parameters and graphical representation of the Michaelis-Menten equation. *Anal. Bioch.* 141:179-183.
- Nunes L.R., Rosato Y.B., Muto N.H., Yanai G.M., Silva V.S., Leite D.B., Gonçalves E.R., Souza A.A., Coletta-Filho H.D., Machado M.A., Lopes S.A., Oliveira R.C. (2003) Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. *Genome Research* 13:570-578.
- Nyland G., Goheen A.C., Lowe S.K., Kirkpatrick H.C. (1973) The ultrastructure of a rickettsia-like organism from a peach tree affected with phony disease. *Phytopathology* 63:1255-1258.

- Oliveira R.C., Yanai G.M., Muto N.H., Leite D.B., Souza A.A., Coletta-Filho H.D., Machado M.A., Nunes L.R. (2002) Competitive hybridization on spotted microarrays as a tool to conduct comparative genomic analyses of *Xylella fastidiosa* strains. *FEMS Microbiology Letters* 216:15-21.
- Osiro D., Muniz J.R.C., Coleta Filho H.D., de Souza A.A., Machado M.A., Garratt R.C., Colnago L.A. (2004) Fatty acid synthesis in *Xylella fastidiosa*: correlations between genome studies, ¹³C NMR data, and molecular models. *Biochemical and Biophysical Research Communications* 323:987-995.
- Paradela Filho O., Sugimori M.H., Ribeiro I.J.A., Garcia Jr. A., Beretta M.J.G., Harakawa R., Machado M.A., Laranjeira F.F., Rodrigues Neto J., Beriam L.O.S. (1997) Constatação de *Xylella fastidiosa* em cafeeiro no Brasil. *Summa Phytopathologica* 23:46-49.
- Paula D.P., Azzoni A.R., Siqueira S.F., Catani C.F., Rosselli L.K., Souza A.P. (2003) Expression and purification of a putative H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*. *Protein Expression & Purification*. 32:61-67.
- Pelton J.T., McLean L.R. (2000) Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry* 277:167-176.
- Petoukhov M.V., Eady N.A.J., Brown K.A., Svergun D.I. (2002) Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophysical Journal* 83(6):3113-3125.
- Provencher S.W., Glöckner J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20:33-37.
- Purcell H.A., Finlay A.H. (1979) Evidence for noncirculative transmission of Pierce's disease bacterium by sharpshooter leafhoppers. *Phytopathology* 69:393-395.
- Raju C. B., Wells J. M. (1986) Diseases caused by fastidious xylem-limited bacteria. *Plant Disease* 70:182-186.
- Reid M.F., Fewson C.A. (1994) Molecular characterization of microbial alcohol dehydrogenases. *Critical Reviews in Microbiology* 20:13-56.
- Rosselli L.K., Oliveira C.L.P., Azzoni A.R., Tada S.F.S., Catani C.F., Medrano F.J., Torriani I.L., Souza A.P. (2006) A new member of the aldo-keto reductase family from the plant pathogen *Xylella fastidiosa*. *Archives of Biochemistry and Biophysics*. 453:143-150.
- Sanli G., Blaber, M. (2001) Structural assembly of the active site in an aldo-keto reductase by NADPH cofactor. *Journal of Molecular Biology*. 22:1209-1218.
- Silva F.R. da, Vettore A.L., Kemper E.L., Leite A., Arruda P. (2001) Fastidian gum: the *Xylella fastidiosa* exopolissaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiology Letters* 203:165-171.

- Simpson A.J.G., Reinach F.C., Arruda P. *et al.* (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406:151-157.
- Sreerama N., Woody R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Analytical Biochemistry* 209:32-44.
- Svergun D.I. (1992) Determination of the regularization parameters in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* 25:495-503.
- Svergun D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* 76:2879-2886.
- Svergun D.I., Petoukhov M.V., Koch M.H.J. (1995) CRY SOL – A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* 28:768-773.
- Svergun D.I., Petoukhov M.V., Koch M.H.J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* 80:2946-2953.
- Tang J.L., Liu Y.N., Barber C.E., Dow J.M., Wootton J.C., Daniels M.J. (1991) Genetic and molecular analysis of a cluster of rpf genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. *Molecular & General Genetics* 226:409-417.
- Tao K., Makino K., Yonei S., Nakata A., Shinagawa H. (1989) Molecular cloning and nucleotide sequencing of *oxyR*, the positive regulatory of a regulon for an adaptative response to oxidative stress in *Escherichia coli*: homologies between OxyR protein and a family of bacterial activator proteins. *Molecular General Genetics*. 218:371-376.
- Tao K., Makino K., Yonei S., Nakata A., Shinagawa H. (1991) Purification an characterization of the *Escherichia coli* OxyR protein, the positive regulator for a hydrogen peroxide-inducible regulon. *Journal of Biochemistry*. 109:262-266.
- Walther D., Cohen F.E., Doniach S. (2000) Reconstruction of low resolution three-dimensional density maps from one-dimensional small angle X-ray scattering data for biomolecules in solution. *Journal of Applied Crystallography* 33:350-363.
- Wells J.M., Raju B.C., Hung H-Y., Weisberg W.G., Mandelco-Paul L., Brenner D.J. (1987) *Xylella fastidiosa* gen. nov., sp. nov.: gram negative, xylem-limited fastidious plant bacteria related to *Xanthomonas spp.* *International Journal of Systematic Bacteriology* 37(2):136-143.
- Woody R.W. (1995) Circular dichroism. *Methods Enzymol.* 246:34-71.

9. ANEXOS

9.1. ANEXO I

– "Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*"

9.2. ANEXO II

– "Cloning, expression and purification of the virulence protein D from *Xylella fastidiosa*"

"Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*"

Adriano R. Azzoni, Susely F.S. Tada, Luciana K. Rosselli, Débora P. Paula, Cleide F. Catani, Adão A. Sabino, João A.R.G. Barbosa, Beatriz G. Guimarães, Marcos N. Eberlin, Francisco J. Medrano e Anete P. de Souza.

Publicado na revista

Protein Expression & Purification 33 (2004) 297-303



Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*

Adriano R. Azzoni,^{a,*} Susely F.S. Tada,^{a,1} Luciana K. Rosselli,^a Débora P. Paula,^a Cleide F. Catani,^b Adão A. Sabino,^c João A.R.G. Barbosa,^d Beatriz G. Guimarães,^d Marcos N. Eberlin,^c Francisco J. Medrano,^d 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, SP, Brazil

^b Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, C.P. 6010, Campinas, SP, Brazil

^c Instituto de Química, Universidade Estadual de Campinas, C.P. 6154, Campinas – SP, Brazil

^d Departamento de Cristalografia de Proteínas, Laboratório Nacional de Luz Síncrotron, C.P. 6192, Campinas, SP, Brazil

Received 26 August 2003, and in revised form 9 October 2003

Abstract

The small heat shock proteins (smHSPs) belong to a family of proteins that function as molecular chaperones by preventing protein aggregation and are also known to contain a conserved region termed α -crystallin domain. Here, we report the expression, purification, and partial characterization of a novel smHSP (HSP17.9) from the phytopathogen *Xylella fastidiosa*, causal agent of the citrus variegated chlorosis (CVC). The gene was cloned into a pET32-Xa/LIC vector to over-express the protein coupled with fusion tags in *Escherichia coli* BL21(DE3). The expressed HSP17.9 was purified by immobilized metal affinity chromatography (IMAC) and had its identity determined by mass spectrometry (MALDI-TOF). The correct folding of the purified recombinant protein was verified by circular dichroism spectroscopy. Finally, the HSP17.9 protein also proved to efficiently prevent induced aggregation of insulin, strongly indicating a chaperone-like activity.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Small heat shock protein; *Xylella fastidiosa* smHSP; Expression and purification

The heat shock proteins belong to the class of molecular chaperones defined as “cellular proteins which mediate the correct folding of other polypeptides, and in some cases their assembly into oligomeric structures, but which are not components of the final functional structures” [1]. This implies that the main function of chaperones is to transiently interact with proteins avoiding aggregations that are deleterious to the cell. Heat shock proteins are produced by the cells under stress conditions, such as high temperatures, but are also developmentally regulated under physiological condition [2]. They are divided into families according to their size and amino-acid sequence. The small heat shock proteins (smHSPs) belong to a structurally divergent group within the chaperone super-

family, with molecular masses ranging from 12 to 43 kDa [3]. They are also known to contain an evolutionarily conserved region termed α -crystallin domain, located toward the C-terminus of the protein monomer that proved to be important for oligomer formation, thermotolerance, and chaperone activity [1,2]. Circular dichroism analysis of smHSPs indicates that their secondary structures are composed mainly of β -sheets and the C-terminus extensions are likely to be highly flexible [4].

A common feature of small heat shock proteins is their tendency to assemble into large oligomeric complexes, ranging from 145 kDa to 10 MDa [1,5]. Recent studies have demonstrated that oligomers exhibit a dynamic equilibrium with constituent subunits, which is important but not sufficient itself for the chaperone activity [2,6]. The simplest quaternary structure presented in the literature is the 145 kDa trimers of trimers formed by *Mycobacterium tuberculosis* HSP16.3 [5]. Despite the

* Corresponding author. Fax: +55-19-37881089.

E-mail address: azzoni@unicamp.br (A.R. Azzoni).

¹ Both authors contributed equally to this work.

interesting properties of the smHSPs and their role in the mechanisms of protein folding, there is little literature on protein structural information, especially three-dimensional structures. One of the few and best-characterized smHSP structures is the *Methanococcus jannaschii* HSP16.5 that has been crystallized revealing a monomer containing a composite β -sandwich structure. It also revealed a highly ordered oligomeric structure composed of 24 monomers forming a hollow spherical complex of octahedral symmetry [7].

In this work, we present the expression, purification, and partial characterization of a novel smHSP from the phytopathogen *Xylella fastidiosa*. The *X. fastidiosa* is a xylem-limited bacterium responsible for a variety of economically important plant diseases, including the citrus variegated chlorosis (CVC) that causes tremendous damage to the citrus cultures in southern Brazil [8–11]. The work presented here aims to contribute to increasing the knowledge of the biological role of smHSPs, presenting a method for the production of high amounts of relatively pure protein that are needed to determine the three-dimensional structure. It also aims at adding new information about proteins that may be related to *X. fastidiosa* pathogenesis, necessary for new approaches towards CVC combat.

Materials and methods

Materials

The oligonucleotide primers were synthesized at Invitrogen Life Technologies (São Paulo, Brazil). The pET32-Xa/LIC vector, the BL21(DE3) strain, protease factor Xa, and anti-His-tag anti-body were obtained from Novagen (Madison, WI). The Ni-NTA affinity resin was obtained from Qiagen (Hilden, Germany). The Coomassie blue reagent for total soluble protein determination was purchased from Bio-Rad (Hercules, CA). The molecular mass marker (LMW) and the HR5/5 (50 × 5 mm) chromatography column were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). The protease inhibitor phenylmethylsulfonyl fluoride (PMSF), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α -cyano-4-hydroxycinnamic acid (CHCA), lysozyme, bovine serum albumin (BSA), and the enzyme trypsin were purchased from Sigma Chemical (St. Louis, MO). The high purity bovine insulin (97% purity) was obtained from Biobrás S.A., (Montes Claros, Brazil). All other chemicals used were of at least reagent grade.

Methods

Expression vector construction

The target ORF (XF2234) from *X. fastidiosa* was amplified by polymerase chain reaction (PCR) using

genomic DNA as the template and cloned into the pET32-Xa/LIC expression vector. A pair of PCR primers, 5'-GGTATTGAGGGTCGCATGAACGTTG TTCG-3' (sense) and 5'-AGAGGAGAGTTAGAGCC TCACATTGACTG-3' (antisense), was designed to generate products with vector cohesive overhangs. The amplification protocol consisted of a 3 min denaturation at 94 °C followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min and 30 s, and extension at 72 °C for 2 min. The blunt PCR products were purified and treated with T4 DNA polymerase in the presence of the dGTP and DTT to generate the specific vector cohesive overhangs. The insert cloning in the linearized vector was carried out according to the vector manufacturer's protocol [6]. The pET32-Xa/LIC vector is designed for the expression of the recombinant protein fused to the 109 amino acid thioredoxin (11.7 kDa), a 6 amino acid His-tag, and 15 amino acid S-tag sequences upstream of the cloning site. The fusion tags together have a molecular mass of 17.6 kDa, a theoretical pI of 5.95, and they can be removed from the recombinant target protein by protease cleavage using factor Xa.

Competent *Escherichia coli* DH5 α cells were transformed with recombinant plasmids by using a slightly different method of the standard polyethylene glycol (PEG) method [12]. The standard method describes a heat shock step when cells are kept at 42 °C for 30 s. The high temperature step was substituted by letting the cells stand at room temperature for 10 min. Cells were cultured overnight at 37 °C in Luria-Bertani broth (LB) plates containing 50 μ g/mL ampicillin. The colonies were individually stored at –70 °C in a permanent 2YT+HMFM broth, which contained 16 g tryptone, 10 g yeast extract, 5 g NaCl, 0.076 g MgSO₄·7H₂O, 0.45 g dehydrated sodium citrate, 0.90 g (NH₄)₂SO₄, 1.8 g KH₂PO₄, 4.7 g K₂HPO₄, and 44 mL glycerol, per liter. The clones were checked by PCR and their sequences were verified by nucleotide sequencing. *E. coli* BL21(DE3) cells were transformed with positive recombinant plasmid using the PEG method and used for protein expression.

smHSP over-expression and purification

Escherichia coli BL21(DE3) cells transformed by recombinant plasmids were cultured in 10 mL LB broth added to 50 μ g/mL ampicillin and grown overnight at 37 °C and 300 rpm. This pre-inoculum was then transferred to 1.0 liter LB broth containing ampicillin at the same concentration. The culture was grown at 37 °C and 300 rpm until OD₆₀₀ of 0.8 when cells were induced with 1 mM IPTG for an additional 2 h. The cells were harvested by centrifugation at 2600g and 4 °C for 10 min. The pellet containing the bacteria was suspended in 44 mL of 50 mM Tris-HCl buffer, pH 7.5, containing

300 mM NaCl (adsorption buffer). The protease inhibitor PMSF and lysozyme were added to final concentrations of 1.0 mM and 1.0 mg/mL, respectively. The suspension was incubated for 30 min at 4 °C and sonicated 8 times for 15 s at 70% of the maximum power in a Sonifier Misonix (Microson Ultrasonic Cell Disruptor XL). The lysed material was clarified by centrifugation at 27,500g for 15 min at 4 °C and the supernatant was collected to confirm recombinant protein expression by SDS-PAGE.

Small-scale purification of the target protein was done by loading 6 mL of the clarified supernatant onto an HR5/5 column, packed with 1.0 mL nickel-nitrilotriacetic acid (Ni-NTA) resin, using an AKTA-FPLC system. The resin was washed with 10 CV of adsorption buffer containing 1.0 M NaCl. The adsorbed proteins were eluted from the column using an imidazole gradient (from 0 to 200 mM in 10 CV) in adsorption buffer. Fractions containing 1.0 CV of volume were collected and a flow rate of 0.75 mL/min was used during all the chromatographic steps. Fractions were assayed for total protein concentration and analyzed by SDS-PAGE. The recombinant HSP17.9 was separated from the fusion tags by factor Xa proteolysis (0.1 U/ μ L per 10 μ g of recombinant protein). The hydrolysis was performed at 25 °C for 16 h according to the protease supplier's instructions. The protease was inactivated by PMSF at 1 mM and the proteolysis products were analyzed by SDS-PAGE. The sample was dialyzed against adsorption buffer containing 10 mM imidazole to prepare for the next purification step.

The final purification step was performed using the same FPLC system and column/resin mentioned above. The sample containing the HSP17.9 and the cleaved fusion protein was loaded into the column equilibrated with the same dialysis buffer. The flowthrough proteins (HSP17.9) were collected and washed out with 10 CV of the same buffer. Bound proteins were eluted with equilibration buffer containing 200 mM imidazole. The flowrate used was 0.75 mL/min. The fractions containing 1 CV of volume were collected and analyzed by SDS-PAGE and Western blot, and had their total soluble protein concentration determined. Large-scale purification was performed by scaling up the same protocol described above, except for the column/resin used. In this case, clarified extract from 5.0 liters of bacterial culture was loaded into an HR10/10 column packed with 8.0 mL of agarose-IDA-Ni²⁺ resin. All chromatographic steps were carried out at 2.0 mL/min.

Analytical assays

The total soluble protein concentration was assayed according to the method presented by Bradford [13], using bovine serum albumin as the protein standard. More accurate estimations for purified HSP17.9 were

made based on absorbance at 280 nm, using a calculated extinction coefficient of 0.852 g/L/cm, based on the method described by Pace and Schmidt [14]. Protein purity during the purification step was monitored by SDS-PAGE performed using a 4% stacking gel and a 12% separation gel according to Laemmli [15] and stained using the Coomassie brilliant blue R-250 staining technique. Native gel electrophoresis was carried out using the same protocol described above, subtracting the detergent sodium dodecyl sulfate (SDS) from all the buffers used. For Western blot, protein samples from the second chromatographic step were run on 12% gel SDS-PAGE and transferred to nitrocellulose membranes (Hybond-N, Amersham, USA). Detection was carried out using rabbit anti-His-tag antibody and a conjugated goat anti-rabbit IgG alkaline phosphatase (Sigma, St. Louis, MO) followed by chemiluminescent detection with CSPD (Tropix, USA).

Mass spectrometry

The recombinant HSP17.9 mass spectrometric analysis was done in a matrix-assisted laser-desorption time-of-flight (MALDI-TOF) M@LDI-LR mass spectrometer (Micromass, USA). The trypsin-digested protein and intact protein were analyzed in the spectrometer using reflecting and linear modes, respectively, according to the equipment manufacturer's instructions. The matrixes used for trypsin-digested (peptides) and intact protein samples were CHCA and sinapinic acid, respectively. The masses of monoisotopic peaks with relative intensity higher than 5% of the most intense peak in the spectrum were used for comparison to a theoretical digestion of the protein by trypsin. This was carried out using the MS-Digest program ([www.http://prospector.ucsf.edu](http://prospector.ucsf.edu)).

Circular dichroism spectroscopy

The circular dichroism (CD) spectroscopy of the fully purified protein was studied to assess the secondary structural integrity. The CD spectra were generated using a 1 mm pathlength cuvette containing 200 μ L HSP17.9 protein sample at 0.1 mg/mL in 5 mM Tris-HCl buffer, pH 7.5, at 20 °C. The assays were carried out in a Jasco 810 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan). The spectrum was presented as an average of four scans recorded from 190 to 250 nm, at a rate of 20 nm/min.

Chaperone-like activities study

The inhibition of the aggregation of bovine insulin (B chain) by HSP17.9 was studied according to the method presented by Horwitz et al. [16]. The method is based on the reduction of insulin disulfide bonds by dithiothreitol

and aggregation of insulin B chain followed by the increase of the turbidity from light scattering at 360 nm in a Hitachi U-3000 spectrophotometer. All experiments were carried out at 37 °C.

Results and discussion

Expression and purification of HSP17.9

The *X. fastidiosa* Genome Program generated a large amount of gene sequences from this important plant pathogen [11]. One of these genes encodes a 17.9 kDa protein similar to a family of small heat shock proteins. The authors in the literature suggest that the minimal functional unit of these kinds of proteins consists of a core region of about 85 amino acids called the α -crystallin domain. Fig. 1 presents the sequence alignment of the *X. fastidiosa* protein, called HSP17.9, with three other proteins that may belong to the same family. The strongest similarity (26%) was found with the SP21 protein (Q06823) from *Stigmatella aurantiaca*. The alignment also compares the HSP17.9 with the *M. jannaschii* HSP16.5 (Q57733) and human α -crystallin (B53814). As expected, the highest number of hits was found in the α -crystallin domain, the highly conserved domain of smHSPs.

In our work, the ORF (XF2234) encoding the HSP17.9 protein was successfully cloned in the pET32 Xa/LIC expression vector and no difficulty was found to transform the *E. coli* BL21(DE3) strain for expression studies. Dose dependence and time course studies of the induction of the recombinant protein expression, analyzed by SDS-PAGE, led to an IPTG concentration of 1 mM and induction time of 2 h at 37 °C. The expression of the fusion protein containing the HSP17.9 was considered high: 9–10 mg of soluble protein/liter of bacterial broth. High amounts of relatively pure recombinant protein containing the His-tag were recovered and

purified from the lysed extract by using immobilized metal affinity chromatography (Fig. 2, lanes 2–4). 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. Elution fractions containing the recombinant protein with a level of purity higher than 90% (analyzed by optic densitometry) were separated and combined as

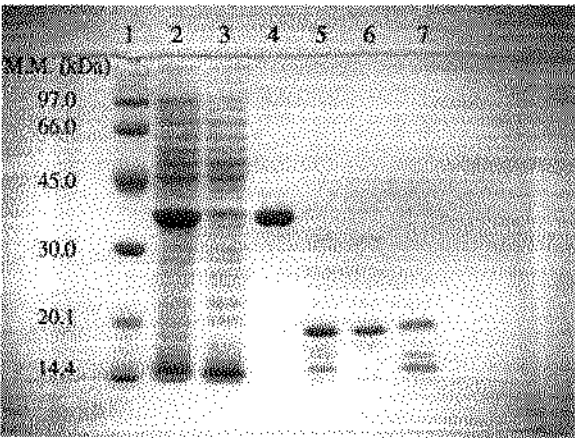


Fig. 2. SDS-PAGE of the fractions collected during the purification steps of the recombinant HSP17.9. Legend: lane 1—molecular mass markers; lane 2—extract from bacterial lysis loaded into the Ni-NTA column, 11.0 µg; lane 3—flowthrough extract, 7.0 µg; lane 4—fraction collected during the elution step at imidazole concentration of 200 mM, containing the recombinant protein (HSP17.9 + fusion tags), 4.0 µg; and lane 5—proteins resulted from factor Xa cleavage step for HSP17.9 separation from fusion tags, 3.2 µg. This fraction was loaded into the Ni-NTA column to promote final purification of the HSP17.9. Lane 6—flowthrough fraction of the final purification step containing the HSP17.9, 2.0 µg. Lane 7—proteins eluted at 200 mM imidazole concentration containing the fusion tags and impurities resulted from non-specific factor Xa cleavage, 2.3 µg. Separation gel of 12% acrylamide concentration.

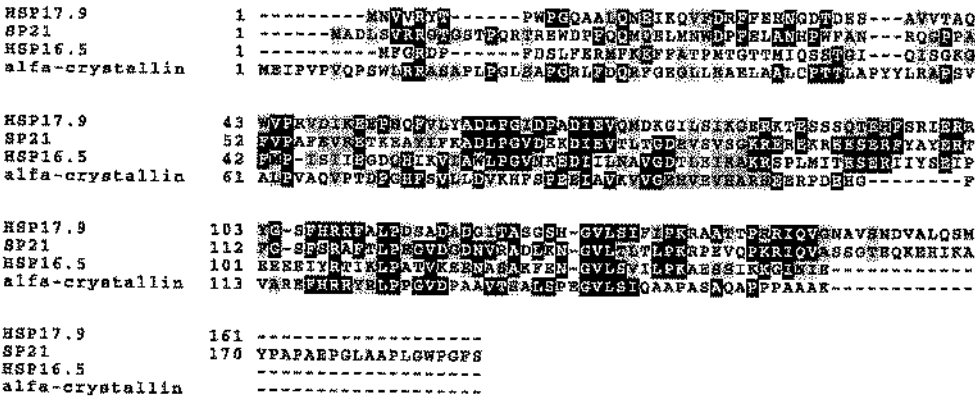


Fig. 1. Comparison of the *X. fastidiosa* HSP17.9 with other smHSPs. The multiple sequence alignment of the HSP17.9 with *S. aurantiaca* SP21 (Q06823), *M. jannaschii* HSP16.5 (Q57733), and human α -crystallin (B53814) was done using the CLUSTALW program. Letters shaded in black indicate amino acids that are identical in at least two of the smHSPs. Conserved substitutions are shaded in gray.

a pool for further purification steps. Approximately 6 mg of purified fusion recombinant protein/liter of initial bacterial broth was recovered at this step.

To remove the N-terminal fusion tags from the target protein, a cleavage by factor Xa protease was performed. The fusion tags were used so as to improve the recombinant protein solubility, also allowing ease of detection and purification from the extract. The fusion tags were successfully removed from HSP17.9 by cleavage and few non-specific cleavages were noted. The analysis on SDS-PAGE of the proteolysis products is shown in Fig. 2 (lane 5). Since the recombinant HSP17.9 and the fusion tags contain almost identical molecular masses (17.9 and 17.6 kDa, respectively), they are seen on the gel as a unique protein band. The cleaved recombinant HSP17.9 remained stable in the solution, however, some tendency towards aggregation and precipitation was noticed when the protein was dialyzed against water at a low temperature (4 °C).

A final chromatographic step on the IMAC resin was performed to separate the fragment containing the tags and the target protein. In this case, the purified recombinant HSP17.9 was collected in the flowthrough fractions (Fig. 2, lane 6). Few amounts of non-cleaved protein were also separated from the target protein during this step. Since both proteins contain similar molecular masses, merged as a single band in SDS-PAGE, the fractions collected during the chromatography were analyzed by Western blot using anti-His-tag antibody. The result indicated that the separation was achieved, since the flowthrough fractions were virtually free from proteins containing His-tag (result not shown). The final amount of purified recombinant HSP17.9 obtained was approximately 1.6 mg/liter of initial bacterial broth. More detailed data on the purification of the HSP17.9 are presented in Table 1.

Mass spectrometry and oligomerization analysis

The purified recombinant HSP17.9 was analyzed by mass spectrometry to verify the protein molecular mass

and identity. The analysis resulted in a molecular mass of 17,729 Da, which is a value close to the theoretical molecular mass calculated from the primary amino-acid sequence: 17,858 Da. The peptide profile found for the mass spectroscopy analysis of the digested HSP17.9 ("fingerprint") was compared to a theoretical digestion of the protein. Five different peptides were found the molecular masses of which matched (difference lower than 1.0 Da) those of the expected peptides, confirming the protein identity (data not shown).

The HSP17.9 tendency to assemble into oligomeric complexes was also detected by native gel electrophoresis and gel filtration chromatography analysis. Despite the results for the determination of the exact molecular mass of the protein quaternary structure still not being conclusive, maybe due to a natural instability of such complexes, there is no doubt that the protein tends to form large oligomeric complexes in solution. Recent studies have suggested that the assemblies of smHSPs high level structures are dynamic and shaped by the frequent exchange of subunits [1]. The studies have also suggested that these structures are important but not sufficient for the chaperone-like activity [2,6].

Circular dichroism spectroscopy analysis

To investigate the structural integrity of the purified HSP17.9, the protein was analyzed by CD spectroscopy. The CD spectrum, which resulted, is shown in Fig. 3 and indicated that the recombinant HSP17.9 contains a substantial amount of secondary structure. The protein presented a predominant signal of β -strands, with a maximum in ellipticity at 197 nm and a minimum at 213 nm. This result is coherent with the secondary-structure predictions and data presented in the literature for other smHSP structures, all of them making up a majority of β -strands. One of the best structurally characterized smHSPs in the literature, the *M. jannaschii* HSP16.5, was found to have its monomer formed by nine β -strands, two short 3_{10} -helices, and one short β -strand [7]. Finally, our results indicated that the

Table 1
Purification of recombinant HSP17.9 from *E. coli* extract

Step	Volume (mL)	Protein (mg)	HSP17.9 (mg)	HSP 17.9 yield (%)	HSP 17.9 purity (%)	Purification factor
Extraction	44	79	9.0	100	11	1
First IMAC ^a						
Flowthrough	84	68	0	0	—	—
Eluate	80	6.7	6.0	67	90	8.2
Second IMAC						
Load ^b	32	5.7	2.3	26	40	3.6
Flowthrough	37	3.5	0	0	—	—
Eluate	10	1.8	1.6	18	89	8.1

^a Immobilized metal ion chromatography.

^b This step is preceded by a cleavage step using Factor Xa protease. The cleaved HSP17.9 usually represents about 40% of the resulting total protein in solution.

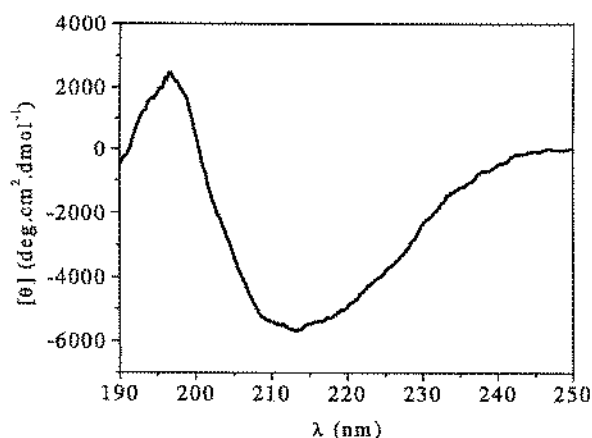


Fig. 3. Circular dichroism spectrum of the purified recombinant HSP17.9. The spectrum was recorded at 20 °C and a protein concentration of 0.1 mg/mL in 5 mM Tris buffer, pH 7.0. Data were collected using an average of four scans per replicate.

protein is quite stable at room temperature and remained folded throughout the purification process, being suitable for crystallization studies aiming at three dimensional structure determinations.

Chaperone-like activity study

It is well known that some proteins can be unfolded by reducing the interchain disulfide bond. In the case of insulin, the reduction of the disulfide bonds by DTT will lead to the aggregation and precipitation of the B chain while the A chain remains in solution. In this way, the chaperone-like activity of the recombinant HSP17.9 could be indicated by its characteristic of preventing insulin B chain aggregation upon reducing environmental condition. Since the HSP17.9 does not possess any disulfide bond, it is not affected in this condition as insulin. Fig. 4 presents the results found for insulin aggregation in the presence of different mass concentrations of HSP17.9. Significant prevention of insulin aggregation was noted even at low HSP17.9 concentrations (0.02 mg/mL for an insulin concentration of 0.4 mg/mL) strongly indicating a chaperone-like activity. Almost 40% reduction of aggregate formation was detected when a final concentration of 0.08 mg/mL of HSP17.9 was used (20% of the insulin concentration).

Finally, the different effects caused by the presence of HSP17.9, and the control proteins bovine serum albumin (BSA) and lysozyme on the induced aggregation of insulin are illustrated in Fig. 5. The HSP17.9 presented a chaperone-like activity effectively preventing insulin aggregation, whereas BSA interaction seems only to retard it. The lysozyme also aggregated during the assay, causing a shift in the measured light scattering, the opposite effect of HSP17.9.

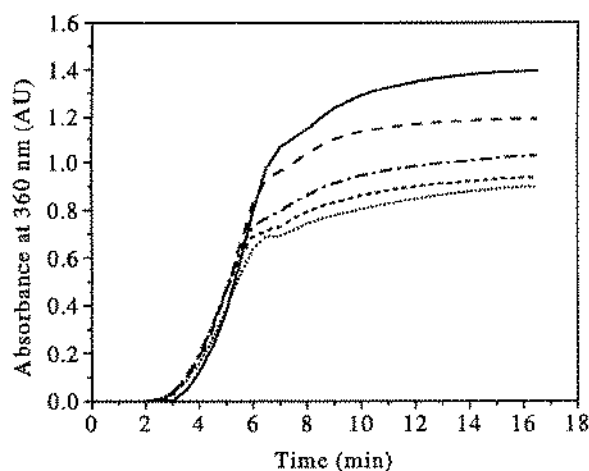


Fig. 4. Profile of insulin aggregation induced by dithiothreitol (DTT) in the presence of different concentrations of the recombinant HSP17.9 protein. Legend: solid line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and absence of HSP17.9; dashed line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.02 mg/mL HSP17.9; dash-dot line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.04 mg/mL HSP17.9; short dash line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.06 mg/mL HSP17.9; short dot line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.08 mg/mL of HSP17.9. Experiments were performed in a 50 mM NaPi buffer, pH 7.5, containing 100 mM NaCl at a temperature of 37 °C.

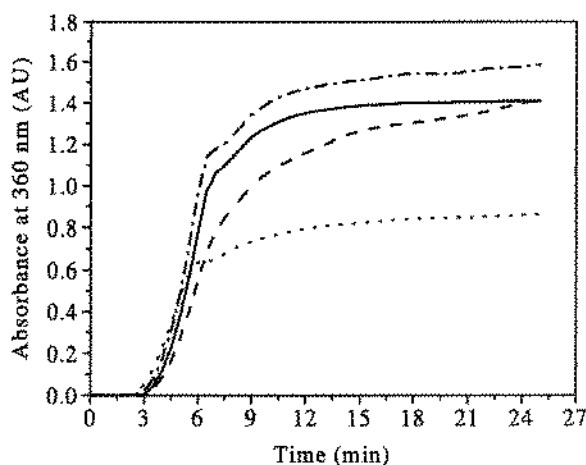


Fig. 5. Profile of insulin aggregation induced by dithiothreitol (DTT) in the presence of recombinant HSP17.9, bovine serum albumin (BSA), and lysozyme. Legend: solid line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT (control); dashed line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.1 mg/mL BSA; dash-dot line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.1 mg/mL lysozyme; and short dot line, 0.4 mg/mL insulin solution in the presence of 20 mM DTT and 0.1 mg/mL HSP17.9. Experiments were performed in a 50 mM NaPi buffer, pH 7.5, containing 100 mM NaCl at a temperature of 37 °C.

Conclusion

This work presented the cloning, expression, purification, and partial characterization of a novel smHSP

protein from the plant pathogen *X. fastidiosa*. The protein was purified using two steps of affinity chromatography on Ni-NTA resin, and its purity and identity were verified by SDS-PAGE, Western blot, and mass spectrometry. As expected for smHSP proteins, the CD analysis of the HSP17.9 indicated a secondary structure composed mainly of β -strands, despite some short α -helices also seeming to be present. The HSP17.9 proved to efficiently avoid chemically induced insulin aggregation, strongly indicating a chaperone-like activity. Finally, this work presents new information about a protein the biological role of which may be related to the *X. fastidiosa* pathogenicity. Crystallization studies of the protein are currently in progress in our laboratory for future three-dimensional structure characterization by X-ray crystallography.

Acknowledgments

The authors thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 01/07533-7) and the Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their financial support. L.K.L., A.R.A., C.F.C., and S.F.S.T. are recipients of fellowships from FAPESP, and D.P.P. is from CAPES. We would also like to thank Professor Dr. Nilson Zanchin from the Laboratório Nacional de Luz Síncrotron, Professor Dr. Everson Alves Miranda from the Faculdade de Engenharia Química—UNICAMP, and Professor Dr. Tomomasa Yano, Professor Dra. Yoko Bomura Rosato, Professor Dr. Adilson Leite, and Professor Dr. José Camillo Novello from the Instituto de Biologia, Universidade Estadual de Campinas—UNICAMP for equipment support.

References

- [1] F. Narberhaus, α -crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network, *Microb. Mol. Biol. Rev.* 66 (2002) 64–96.
- [2] J.A. Crack, M. Mansour, Y. Sun, T.H. MacRae, Functional analysis of a small heat shock/ α -crystallin protein from *Artemia franciscana*, *Eur. J. Biochem.* 269 (2002) 933–942.
- [3] W.W. de Jong, G.J. Caspers, Genealogy of the α -crystallinC small heat shock protein superfamily, *Int. J. Biol. Macromol.* 22 (1998) 151–162.
- [4] M.R. Leroux, B.J. Ma, G. Batelier, R. Melki, E.P.M. Candido, Unique structural features of a novel class of small heat shock proteins, *J. Biol. Chem.* 272 (19) (1997) 12847–12853.
- [5] Z. Chang, T.P. Primm, J. Jakana, I.H. Lee, I. Serysheva, W. Chiu, H.F. Gilbert, F.A. Quiocho, Mycobacterium tuberculosis 16-kDa antigen (Hsp16.3) function as an oligomeric structure in vitro to suppress thermal aggregation, *J. Biol. Chem.* 271 (2) (1996) 7218–7223.
- [6] M.P. Bova, H.S. Mchaourab, Y. Han, B.K.-K. Fung, Subunit exchange of small heat shock proteins. Analysis of oligomers formation of α A-crystallin and Hsp27 by fluorescence resonance energy transfer and site-direct truncations, *J. Biol. Chem.* 275 (2000) 1035–1042.
- [7] K.K. Kim, R. Kim, S.-H. Kim, Crystal structure of a small heat-shock protein, *Lett. Nat.* 394 (1998) 595–599.
- [8] 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 (16) (2000) 3100–3104.
- [9] 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, *Citros Ind.* (October) (1991) 12–15.
- [10] F.F. Laranjeira, Ten years of citrus variegated chlorosis: what do we know?, *Laranja* 18 (1997) 123–141.
- [11] A.J.G. Simpson et al., The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (6792) (2000) 151–157.
- [12] B.T. Kurien, R.H. Scofield, Polyethylene glycol-mediated bacterial colony transformation, *BioTechniques* 18 (6) (1995) 1023–1026.
- [13] M.M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [14] C.N. Pace, F.X. Schmidt, How to determine the molar absorbance coefficient of a protein, in: T.E. Creighton (Ed.), *Protein Structure, A Practical Approach*, second ed., IRL Press, New York, 1997, pp. 253–259, 383 pp.
- [15] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227 (1970) 680–685.
- [16] J. Horwitz, Q.-L. Huang, L. Ding, M.P. Bova, Lens α -crystallin: chaperone-like properties, *Methods Enzymol.* 290 (1998) 365–383.

"Cloning, expression and purification of the virulence protein D from *Xylella fastidiosa*"

Cleide F. Catani, Adriano R. Azzoni, Débora P. Paula, Susely F.S. Tada, Luciana K. Rosselli,
Anete P. de Souza e Tomomasa Yano.

Publicado na revista

Protein Expression & Purification 37 (2004) 320-326



Cloning, expression, and purification of the virulence-associated protein D from *Xylella fastidiosa*

Cleide Ferreira Catani^{a,*}, Adriano Rodrigues Azzoni^b, Débora Pires Paula^b,
Susely Ferraz Siqueira Tada^b, Luciana Kauer Rosselli^b, Anete Pereira de Souza^b,
Tomomasa Yano^a

^a Department of Microbiology and Immunology, Institute of Biology, The State University of Campinas, CP 6010, Campinas, SP, Brazil

^b Department of Genetic and Evolution, Genetic Engineering and Molecular Biology Center, Institute of Biology,
The State University of Campinas, CP 6010, 13083970 Campinas, SP, Brazil

Received 18 February 2004, and in revised form 30 June 2004

Available online 12 August 2004

Abstract

In this study, an efficient expression system, based on the pET32Xa/LIC vector, for producing a *Xylella fastidiosa* virulence-associated protein D, found to have a strong similarity to *Riemerella anatipestifer* and *Actinobacillus actinomycetencomitans* VapD protein, is presented. The protein has a molecular mass of 17.637 Da and a calculated pI of 5.49. The selected XF0052 gene was cloned in the pET32Xa/LIC vector and the plasmid was transformed into *Escherichia coli* BL21 (DE3) strain at 37°C, with an induction time of 2 h and 1 mM IPTG concentration. The protein present in the soluble fraction was purified by immobilized metal affinity chromatography (IMAC), and had its identity determined by mass spectrometry (MALDI-TOF) and N-terminal sequencing. The purified protein was found as a single band on SDS-PAGE and its correct folding was verified by circular dichroism spectroscopy.

© 2004 Elsevier Inc. All rights reserved.

Keywords: VapD protein; Expression and purification; *Xylella fastidiosa*

Xylella fastidiosa (Xf) is a Gram-negative, rod-shaped, non-flagellate, xylem-limited bacterium, and is the causal agent of several important crop diseases, including citrus variegated chlorosis (CVC) in orange trees [1], Pierce's disease (PD) in vineyards, phony peach disease (PP), periwinkle wilt, and leaf scorch diseases in plum, elm, maple, oak, sycamore, and coffee [2]. 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].

The existence of different disease symptoms observed in a wide range of plant hosts [5], possibly associated with genetically distinct Xf strains, has led researchers to hypothesize that genome comparisons among Xf strains may uncover important information regarding genes involved in the interaction with specific hosts. Thus, the genomes of two other Xf strains, isolated from oleander and almond trees, were sequenced and annotated [6], and a fourth strain, isolated from grapevines and responsible for PD in California, has also been sequenced [7].

Clearly, on the basis of both DNA and protein sequence conservation, the almond and oleander strains are more closely related to one another than either is to the citrus 9a5c strain [6]. More than 88% of the ORFs from the almond–oleander genomes are in common protein families, as compared to 87% between the 9a5c strain

* Corresponding author. Fax: +55-19-37881089.
E-mail address: cleidec@yahoo.com (C.F. Catani).

and oleander (Ann1 strain) and 86% between the 9a5c strain and almond (Dixon strain) [6]. Comparative analysis with the sequenced 9a5c strain and Temecula strain, isolated from infected grapevine with Pierce's disease (PD), revealed that 98% of the *Xf* Temecula genes are shared with the 9a5c genes. Furthermore, the average amino acid identity of ORFs in the strains is 95.7% [7].

Analysis of the *Xf* genomic sequence [4] revealed a number of genes the products of which have very high amino acid sequence related to the gene products in *Xanthomonas* [8]. This relatedness between *Xf* and *Xanthomonas* is not surprising, since these bacterial species are closely related. The phylogenetic analyses place *Xanthomonas campestris* pv. *campestris* (Xcc-strain AT CC33913) and *Xanthomonas axonopodis* pv. *citri* (Xac-strain 306) [9] at the base of the γ -subdivision of protobacteria, in the same clade as *X. fastidiosa*. The high degree of relatedness strongly suggests that *Xanthomonas* and *Xf* genes are orthologous. In contrast, the relative size of the genomes of *Xf* (2.7 Mb) and *Xanthomonas* (5.0 Mb for Xcc and 5.2 Mb for Xac) indicates that there are distinct differences in the complement of genes carried by the two bacteria [8].

This presumed evolutionary genome reduction may be related to the fact that *X. fastidiosa* does not appear to be found in several alternative habitats, in contrast to *Xanthomonas*. However, *X. fastidiosa* presents several genes known previously only in animal pathogens, arguing for the acquisition of additional functions that may facilitate its attachment and growth in insect vectors [10].

The 9a5c strain genome has a circular chromosome of 2.7 Mb, and two plasmids of about 51 and 1.3 kb [11]. From the 2904 predicted open reading frames (ORFs), 47% showed strong homology to known proteins, allowing putative functions to be assigned. Putative pathogenicity, virulence, and adaptation functions were assigned to 147 ORFs (5.1%) [11].

Plasmid pXF51 contains 64 ORFs, of which 5 encode proteins involved in replication or plasmid stability and 20 encode proteins involved in replication or plasmid transfer [4]. One of the ORFs (gene XFa0052) encodes a protein similar to the virulence-associated protein D (VapD) [5,12], found in many other bacterial pathogens [13]. In *Xf* strains, the XFa0052 gene is also present in SJ strain (Citrus SJ), B14 strain (Citrus X1), and C13 strain (Coffee), but is missing in the *Xf* strains from plum, mulberry, grape, elm, almond, and ragweed [5]. In 306 strain (Xac) and ATCC 33913 strain (Xcc) of *Xanthomonas*, the virulence VapD factor is missing [14].

In the study presented by Cao and Cover [15], the *vapD* gene was detected in 61.3% of *Helicobacter pylori* strains and was determined to be closely related to ORFs of similar size found in several different bacterial species. In *A. actinomycetenumitans* [16], *Riemerella anatipestifer* [17], *Rhodococcus equi* [18,19], and *Neisseria gonorrhoeae* [20], similar to *X. fastidiosa*, *vapD* has been found

in plasmids, whereas in *Haemophilus influenzae*, this gene is chromosomal [21]. In *Dichelobacter nodosus*, a Gram-negative anaerobic organism that causes ovine foot rot, *vapD*, has been identified in both chromosomal and plasmid locations [22] and has been shown to encode a translated product of unknown function [13].

The *D. nodosus* chromosomal *vapD* gene is part of a larger *vap* region (designated *vapABCD*) that is found in virulent isolates significantly more frequently than in benign isolates [13]. The VapD protein was gel purified, shown to contain the predicted amino-terminal sequence, and used to produce rabbit antibodies. Western blots (immunoblots) showed that all of the *D. nodosus* strains tested that contained the *vap* region produced the VapD protein [23].

The *R. equi* strains isolated from pneumonic foals typically contain a large plasmid of about 85–90 kb with a 27.5 kb pathogenicity island containing seven *vap* genes. Plasmid-cured isogenic mutants of virulent strains lose their ability to survive on alveolar macrophages and fail to induce pneumonia in foals [18,19]. Furthermore, pneumonic foals had significantly more antibodies against VapD and VapE than did healthy exposed foals. This may indicate a difference in the expression of these two Vap proteins during persistent infections.

There is only one report in the literature presenting the expression and purification of these kinds of proteins [23]. In this case, the authors presented the expression of seven members of the Vap family (VapA and VapC to VapH) from *R. equi* produced in *Escherichia coli* as glutathione *S*-transferase fusion proteins. The proteins were purified by glutathione-Sepharose 4B beads or just concentrated by a protein agglutination protocol to access the immune response of pneumonic or health foals and adult horses.

Despite the fact that several studies emphasize the importance of VapD type proteins in bacteria pathogenesis, biochemical data describing the biological role of these proteins are rare in literature, making the study of VapD proteins a difficult task. In this report, the cloning, expression, and purification of a protein related to virulence D (VapD) from phytopathogen *X. fastidiosa* are presented. The recombinant protein was expressed in *E. coli* and purified for crystallographic studies. The determination of the three-dimensional structure could contribute towards the understanding of the importance of Vap proteins in *X. fastidiosa* pathogenesis.

Materials and methods

Materials

Bacterial strains, plasmids, enzymes, and reagent

The reagents for PCR and oligonucleotide primers were obtained from Invitrogen Life Technologies (São

Paulo, Brazil). The pET32Xa/LIC vector and the BL21(DE3) strain were obtained from Novagen (Madison, WI). The QIAquick PCR purification kit and Ni-NTA (nickel-nitrilotri-acetic acid) affinity resin were obtained from Qiagen (Hilden, Germany). The chromatography columns and the Coomassie blue reagent for total soluble protein determination were purchased from Bio-Rad (Hercules, CA). The molecular-mass marker (LMW) was purchased from Amersham-Pharmacia Biotech (Uppsala, Sweden). The agarose, protease inhibitors, 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α -cyano-4-hydroxycinnamic acid (CHCA), lysozyme, bovine serum albumin (BSA), and trypsin were purchased from Sigma Chemical (St. Louis, MO).

Methods

PCR amplification

The *vapD* gene of *X. fastidiosa* (ORF XFa0052) 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'-GGTATTGAGGGT CGCATGATGGATCGC-3' and reverse primer 5'-AG AGGAGAGTTAGAGCCCTAATCATCAGG-3', with stop codon (bold nucleotides in reverse primer). The primers were designed to generate products with vector cohesive overhangs (underlined), 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 (Fig. 2). The fusion tags together (Trx) have 17.6 kDa and can be removed by factor Xa cleavage [24].

The PCRs were carried out in 25 μ L volumes containing 2.5 μ L of 10 $\times$ reaction buffer, 1 μ L dNTPs (10 mM), 2.5 μ L of each primer (1 pmol/ μ L), 2.0 μ L DNA template (8.0 ng/ μ L), and 1 U *Taq* polymerase. The amplifications were carried out using the thermal profile starting 3 min denaturation at 94°C, followed by 30 cycles consisting of 94°C (1 min), 49°C (1 min and 30 s), and 72°C (2 min) with an additional extension time at 72°C (7 min). After amplification, 10 μ L samples were subjected to electrophoresis in a standard 1% agarose gel to confirm the presence of amplified products (488 bp).

Cloning

The PCR products were purified and cloned in the vector according to the manufacturer's protocol [24]. *E. coli* DH5 α was used for the propagation of recombinant plasmids. The competent DH5 α cells were transformed with recombinant vectors by using the polyethylene glycol (PEG) method [25]. Cells were cultured overnight at 37°C in Luria-Bertani broth (LB) plates containing 50 μ g/mL ampicillin. The colonies were maintained at -80°C and checked by PCR. The *vapD*

insert of pET32Xa/LIC plasmid was sequenced to verify the correct nucleotide sequencing.

Expression and purification of *VapD* fusion protein

The recombinant plasmid with *vapD* insert was transformed into *E. coli* BL21 (DE3) cells by PEG method. The cells were grown in LB medium containing ampicillin (50 μ g/mL) at 37°C, 300 rpm, for 18 h. An aliquot of 4 mL of the culture was used to inoculate 1 L 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 IPTG to a final concentration of 1 mM and cultivated for 2 h. The culture was harvested using 10 min centrifugation at 6000g, 4°C, and stored for further use at -20°C.

The bacterial pellet obtained from 1 L liquid culture was suspended in 45 mL buffer A (50 mM Tris-HCl, pH 7.5; 300 mM NaCl). The PMSF (1 mM), leupeptin (1 μ g/mL), and pepstatin A (1 μ g/mL) protease inhibitors were added, as well lysozyme (1 mg/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 20 min at 4°C. The supernatant containing soluble cellular material (total volume 42 mL) was used for purification procedure using immobilized metal affinity chromatography (IMAC). The column containing 6 mL 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 (50 mM Tris-HCl, pH 7.5; 1.0 M NaCl). Additional washing using 5 CV of buffer C (50 mM Tris-HCl, pH 7.5; 300 mM NaCl; and 5 mM imidazole) was performed. The bounded *VapD* fusion protein was eluted with 5 CV of buffer D (50 mM Tris-HCl, pH 7.5; 300 mM NaCl; and 200 mM imidazole). Fractions were assayed for total protein concentration and analyzed by SDS-PAGE.

Cleavage of Trx-*VapD* fusion protein

The *VapD* protein was cleaved from the Trx-*VapD* fusion protein using the protease trypsin. An aliquot of trypsin solution at a concentration of 1 mg/mL was added to a Trx-*VapD* fusion protein solution at a final mass ratio of 1%. The mixture was incubated at room temperature for 1 h. The protease was inactivated by PMSF at a final concentration of 1 mM. The proteolysis products were analyzed by SDS-PAGE and the samples were submitted to dialysis against buffer C to prepare for the next purification step.

Purification of *VapD* protein

Separation of the cleavage products was performed using the same Ni-NTA column mentioned above.

The affinity resin was equilibrated with 10 CV of buffer C. The sample with cleaved recombinant protein was loaded into the column by gravity. The unbound proteins were washed out with 5 CV of buffer C. The bounded Trx fusion protein was eluted with 5 CV of buffer A, containing 50mM imidazole. Fractions were assayed for total protein concentration and analyzed by SDS-PAGE.

Analytical assays

To estimate the amount of recovered recombinant protein in the first chromatography, the SDS-PAGE gel was analyzed by optic densitometry in a GeneGenius Imager (Gateway, USA). The protein concentration was assayed according to the Bradford method [26], using bovine serum albumin (BSA) as the protein standard. More accurate estimates for purified VapD concentration were carried out based on the method described by Pace and Schmidt [27]. This method is based on UV absorbance spectroscopy at 280nm using extinction coefficients from residues of tryptophan, tyrosine, and disulfide bonds. The calculated extinction coefficient of VapD protein was $0.916\text{gL}^{-1}\text{cm}^{-1}$. The sequence of purification was monitored by SDS-PAGE performed using a 4% stacking gel and 13.5% separation gel under reducing conditions [28]. The gels were stained using Coomassie Brilliant Blue R-250 and destained with a 30% methanol and 10% acetic acid solution.

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 μL VapD protein (0.1mg/mL) in 5mM Tris-HCl buffer, pH 7.5. Four scans in the 250–190nm range at a rate of 20nm/min at 20°C were recorded for sample and average measurement.

Mass spectrometry

The VapD protein mass spectroscopy analysis was carried out in a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MALDI-LR mass spectrometer (Micromass, USA). The trypsin-digested VapD protein was analyzed in the spectrometer using the reflecting mode, and intact protein was analyzed using the linear mode, according to the equipment manufacturer's instructions. The matrixes used for trypsin-digested protein (peptides) and intact protein samples were CHCA and sinapinic acid, respectively. The masses of monoisotopic peaks, with relative intensity higher than 5% of the most intense peak in the spectrum, were used for comparison to a theoretical digestion of the protein by trypsin, using the MS-Digest program (<http://prospector.ucsf.edu>).

N-terminal sequence

The purified protein was loaded in a 12% SDS-polyacrylamide gel, electroblotted onto a PVDF membrane (Bio-Rad, Hercules CA) using the transfer buffer (0.29% glycine; 0.58% Tris-HCl; 0.037% SDS; and 20% methanol) and stained with Ponceau dye. A portion of the stained VapD protein was excised and loaded into a Protein Sequencer PPSQ-23A (Shimadzu, Japan) according to the equipment manufacturer's protocols.

Results and discussion

Expression and purification of the VapD protein

The VapD protein belongs to a family of virulence-associated proteins (Vap family) with an unknown function. The comparative alignment revealed that the VapD protein of *X. fastidiosa* has a stronger similarity with the N-terminal sequence of other proteins that belong to the VapD protein family (Fig. 1).

VapD.XF	2	MNDKCLVDESLI...
VapD.AA	1	...
VapD.RA2	1	...
VapD.HP	1	...
VapD.RA1	1	...
VapD.XF	51	...
VapD.AA	51	...
VapD.RA2	58	...
VapD.HP	56	...
VapD.RA1	49	...
VapD.XF	121	GLTSEKTYRRTIGQKXPENATIALPNDP...
VapD.AA	111	GLSADKRYRRTVSKQFSLQIFPKKTLQNNQ...
VapD.RA2		...
VapD.HP		...
VapD.RA1		...

Fig. 1. Comparison of amino acid sequences of the virulence-associated protein of *X. fastidiosa* (VapD) to that of *A. actinomycetemcomitans* VapD homolog (VapD. AA, GenBank D88189), *R. anatipestifer* (VapD. RA1 and VapD. RA2, GenBank L22308), and *H. pylori* (VapD. HP, GenBank U94318). Letters shaded in black indicate amino acids that are identical in at least two of the VapDs. Conserved substitutions are shaded in gray and gaps are indicated by dashes that were introduced to allow optimal alignment.

In this work, the gene for VapD from *X. fastidiosa* was amplified by PCR using primers designed to clone onto the pET32Xa/LIC vector. The gene was easily cloned onto the vector and successfully transformed in *E. coli* DH5 α strain. The sequence of the gene was verified by nucleotide sequencing and transformed into BL21(DE3) strain. Time course studies for the induction of protein expression using IPTG, analyzed by SDS–PAGE, led to a maximum expression of Trx–VapD at an induction time of 2 h at 37°C. The recombinant VapD protein had high expression in soluble form, expressing 13 mg of protein per liter of culture. However, the Trx–VapD fusion protein was also visualized in the pellet, indicating that part of the protein was also expressed as an insoluble form.

Crude bacterial protein extract from lysed cells was purified using the Ni–NTA affinity column (Fig. 3, lanes 2 and 4). A high amount of Trx–VapD fusion proteins was obtained at this first purification step taking advantage of the His.tag fusion protein (Fig. 3, lane 4). The SDS–PAGE result indicated a protein with an apparent molecular weight of about 35 kDa (corresponding to the Trx–VapD fusion protein) with purity level higher than 90%. In this step, the amount of recombinant protein was 12 mg/L of initial culture. Following the vector manufacturer’s instructions, the recombinant protein was incubated with factor Xa in order to separate VapD protein from fusion tags [24]. Although different temperatures, incubation time, and factor Xa concentration were tested, only unspecific cleavage of the recombinant protein was noted. After exhaustive tests using different proteases, the best cleavage results were found using trypsin. The protease hydrolyzed the Trx–VapD fusion, resulting in two fragments of 19 kDa (VapD protein) and 15 kDa (Trx tag) (Fig. 3, line 5), which were subsequently separated on the Ni–NTA column. As a result, the Trx tag carrier was bound to the column (Fig. 3, lane 7) while the VapD protein passed through the column

and was collected in the flowthrough fractions (Fig. 3, lane 6). The final amount of the highly purified VapD protein was 6.5 mg/L of the initial culture.

Mass spectrometry and N-terminal sequence analysis

The identity of the resulting protein product was confirmed by trypsin digestion followed by MALDI–TOF analysis. The analyses of intact protein showed a molecular mass of 19,481 Da, a value 1.843 Da above the theoretical molecular mass calculated from primary amino acid sequence (17,638 Da) provided by the software Protein Parameters Tools (www.expasy.org/cgi-bin/protoparam). This indicates that the trypsin did not hydrolyze the fusion protein after the arginine residue present at the LIC site, as indicated in Fig. 2. Since trypsin is known to hydrolyze proteins preferentially at the carboxyl side

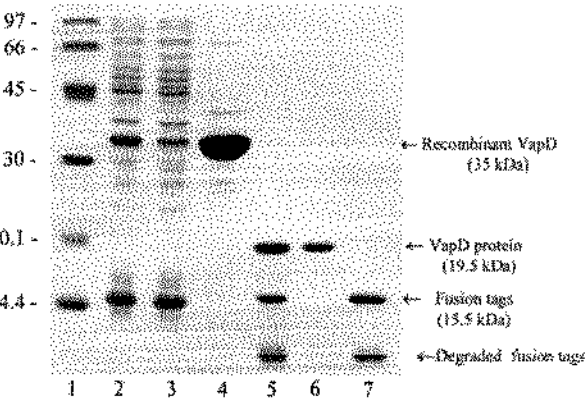


Fig. 3. Purification of the VapD protein expressed in *E. coli* BL21(DE3), 13.5% SDS–PAGE gel. Lane 1, molecular weight markers; lane 2, cells over producing Trx–VapD fusion protein; lane 3, unbound fraction after passage over affinity column; lane 4, fraction eluted with 200 mM imidazole from affinity column; lane 5, eluted fraction cleaved by trypsin; lane 6, VapD protein purified using affinity column; and lane 7, Trx tag fraction eluted with 50 mM imidazole.

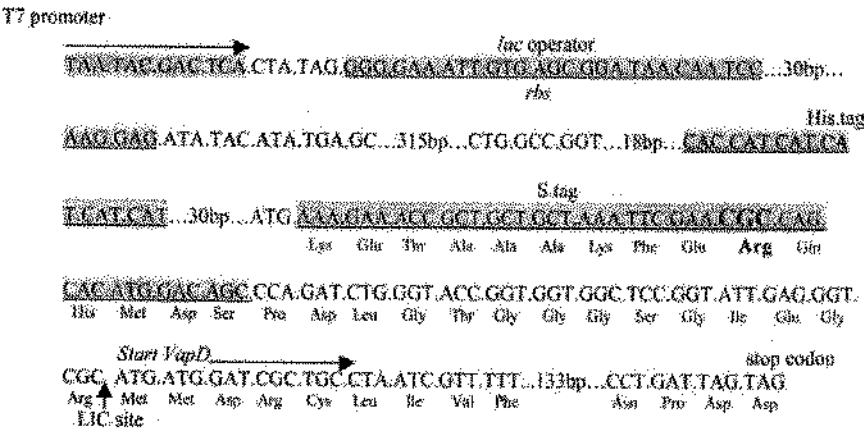


Fig. 2. Nucleotide and amino acid sequences of the pET 32Xa/LIC plasmid near the site of the *vapD* gene insertion. The site of amino acid arginine cleaved by trypsin is bolded.

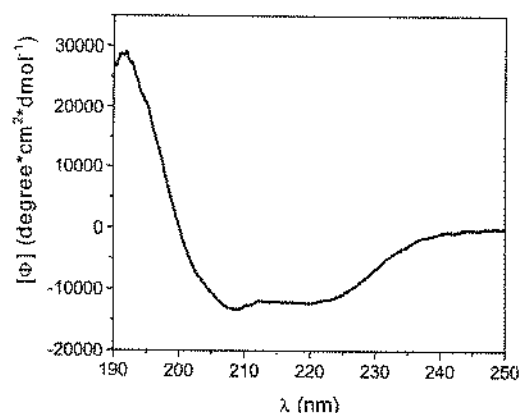


Fig. 4. Circular dichroism spectrum of the purified VapD protein. The spectrum was generated using VapD protein concentration of 0.1 mg/mL in a Tris-HCl buffer (5 mM, pH 7.5). Four scans in a 250–190 nm range at a rate of 20 nm/min at 20°C.

of arginine or lysine residues, the analysis of the protein sequence next to the LIC site may indicate the trypsin hydrolysis site. Fig. 2 presents the amino acid sequence next to the LIC site and the probable trypsin cleavage site (an arginine residue). The site is located 20 residues far from the initial protein methionine adding an extra tail of approximately 1.85 kDa (theoretical molecular weight of 19.49 kDa). This explains the result found by MALDI-TOF spectrometric analysis (19.481 Da). The N-terminal sequencing of the purified protein (Gln-His-Met-Asp-Ser-Pro-Asp) confirmed that the hydrolysis occurred precisely at the arginine residue, at a distance of 20 amino acids from the target protein (Fig. 2).

The peptide profile found for the mass spectroscopy analysis of the digested VapD was compared to the theoretical digestion of the protein (VapD + 19 aa) using the MS-Digest program. Nine different peptides were found, the molecular masses of which matched (difference lower than 1.0 Da) those of the expected peptides (data not shown). This result confirms the protein identity and indicates the cleavage site.

Circular dichroism analysis

To investigate the integrity of the purified VapD protein, it was analyzed by CD spectroscopy. The protein presented a predominant signal of α -helices, with a maximum in ellipticity at 190 nm and minimum at 208 and 222 nm (Fig. 4). This result is in agreement with secondary structure prediction from the PSIPRED program (<http://bioinf.cs.ucl.ac.uk/psipred>). The random coil conformation, with minimum signal at 195 nm and maximum at 212 nm, could not be noted. This result indicated that the protein was folded and stable, being appropriate for crystallographic studies. From our knowledge, this is the first preliminary structural analysis of a VapD protein.

Conclusion

The original goal of the project, the cloning, expression, and purification of a VapD protein in sufficient amount and purity for structural studies by protein crystallography, was fully achieved. Over-expression as a fusion protein with large soluble tag proved to be a proper method to obtain a high yield of the recombinant protein. Unfortunately, attempts to separate VapD protein from Trx tag failed, since the factor Xa was not able to cleave the recombinant protein at the expected site. Although the VapD protein is still associated with a 19 amino acid peptide, crystallization studies of this protein are currently in progress in our laboratory. The resolution of the three-dimensional structure could help in the knowledge of the Vap proteins in the pathogenesis of *X. fastidiosa*.

To our knowledge, there are no genetic experiments underway with *vapD* mutants of *X. fastidiosa*. Since it is known that *R. equi* plasmid-cured mutants for *vapD* lose the ability to survive in macrophages and fail to induce pneumonia in foals [29,30], a similar experiment for *X. fastidiosa* (using a cured pXF51 plasmid) could be valuable to understand the biological role of this protein in the CVC disease.

Acknowledgments

The authors thank Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo), Process # 01/07533-7 and Process # 01/08002-5 for their financial support. Thanks also to the Laboratory of Crystallography of LNLS (National Synchrotron Light Laboratory) for technical expertise, particularly Francisco Javier Medrano, Ph.D., as well as to researchers Heloisa Sobreiro Selistre de Araújo, Ph.D., Departamento de Ciências Fisiológicas, Universidade Federal de São Carlos for N-terminal sequence analysis, and Carlos Eberlin, Ph.D., Instituto de Química, Universidade Estadual de Campinas, for MALDI-TOF mass spectrometry.

References

- [1] M.R. Lambais, M.H.S. Goldman, L.E.A. Camargo, G.H. Goldman, A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity, *Cur. Opin. Microbiol.* 3 (2000) 459–462.
- [2] D.L. Hopkins, A.H. Purcell, *Xylella fastidiosa*: cause of Pierce's disease of grapes and other emergent diseases, *Plant Dis.* 86 (2002) 1056–1066.
- [3] M.J.G. Beretta, R.F. Lee, K.S. Derrick, C.L. Davis, G.A. Barthe, Culture and serology of a *Xylella fastidiosa* associated with citrus variegated in Brazil, *Proc. Int. Soc.* 2 (1992) 830–831.
- [4] A.J.G. Simpson, F.C. Reinach, P. Arruda, et al., The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (2000) 151–158.
- [5] L.R. Nunes, Y.B. Rosato, N.H. Muto, G.M. Yanai, V.S. da Silva, D.B. Leite, E.R. Gonçalves, A.A. de Souza, H.D.

- Coletta-Filho, M.A. Machado, S.A. Lopes, R.C. de Oliveira, Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements, *Gen. Res.* 13 (2003) 570–578.
- [6] A. Bhattacharyya, S. Stilwagen, G. Reznik, H. Feil, et al., Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights, *Genome Res.* 12 (2002) 3707–3714.
- [7] M.A. Van Sluys, M.C. de Oliveira, C.B. Monteiro-Vitorello, C.Y. Miyaki, et al., Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*, *J. Bacteriol.* 185 (2003) 1018–1026.
- [8] N.T. Keen, C.K. Dumenyo, C.-H. Yang, D.A. Cooksey, From rags to riches: insights from the first genomic sequence of a plant pathogenic bacterium, *Genome Biol.* 3 (2000) 1019.1–1019.4.
- [9] A.C.R. da Silva, J.A. Ferro, F.C. Reinach, C.S. Farah, et al., Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities, *Nature* 417 (2002) 459–463.
- [10] J.M. Dow, M.J. Daniels, *Xylella* genomics and bacterial pathogenicity to plants, *Yeast* 17 (2000) 263–271.
- [11] L.M. Scarpari, M.R. Lambais, D.S. Silva, D.M. Carraro, H. Carrer, Expression of putative pathogenicity-related genes in *Xylella fastidiosa* grown at low and high cell density conditions in vitro, *FEMS Microbiol. Lett.* 222 (2003) 83–92.
- [12] M.V. Marques, A.M. da Silva, S.L. Gomes, Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*, *Plasmid* 45 (2001) 184–199.
- [13] M.E. Katz, R.A. Strugnelli, J.I. Rood, Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*, *Infect. Immun.* 60 (1992) 4586–4592.
- [14] <http://cancer.lbi.ic.unicamp.br/xanthomonas/>, Accession to the genome sequences from *Xanthomonas axonopodis* pv. *citri* (strain 306) and *Xanthomonas campestris* pv. *campestris* (strain ATCC33913).
- [15] P. Cao, T.L. Cover, High-level genetic diversity in the *vapD* chromosomal region of *Helicobacter pylori*, *J. Bacteriol.* 179 (1997) 2852–2856.
- [16] D.M. Galli, D.J. LeBlanc, Characterization of pVT736-1, a rolling-circle plasmid from the Gram-negative bacterium *Actinobacillus actinomycetencomitans*, *Plasmid* 31 (1994) 148–157.
- [17] S.-C. Weng, W.-H. Lin, Y.-F. Chang, C.-F. Chang, Identification of a virulence-associated protein homolog gene and *ISRal* in a plasmid of *Riemerella anatipestifer*, *FEMS Microbiol. Lett.* 179 (1999) 11–19.
- [18] S. Takai, S.A. Hines, T. Sekizaki, V.M. Nicholson, D.A. Alperin, M. Osaki, D. Takamatsu, M. Nakamura, H. Dan, J.F. Prescott, DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103, *Infect. Immun.* 68 (2000) 6840–6847.
- [19] S. Benoit, A. Benachour, S. Taouji, Y. Auffray, A. Hartke, Induction of *vap* genes encoded by the virulence plasmid of *Rhodococcus equi* during acid tolerance response, *Res. Microbiol.* 152 (2001) 439–449.
- [20] C. Korch, P. Hagblom, H. Öhman, M. Göransson, S. Normark, Cryptic plasmid of *Neisseria gonorrhoeae*: complete nucleotide sequence and genetic organization, *J. Bacteriol.* 163 (1985) 430–438.
- [21] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* 269 (1995) 496–512.
- [22] M.E. Katz, C.L. Wright, T.S. Gartside, B.F. Cheetham, C.V. Doidge, E.K. Moses, J.I. Rood, Genetic organization of the duplicated *vap* region of the *Dichelobacter nodosus* genome, *J. Bacteriol.* 176 (1994) 2663–2669.
- [23] K.E. Hooper-McGrevy, B.N. Wilkie, J.F. Prescott, Immunoglobulin G subisotype responses of pneumonic and healthy, exposed foals and adult horses to *Rhodococcus equi* virulence-associated proteins, *Clin. Diagn. Lab. Immunol.* 10 (2003) 345–351.
- [24] *Xa/LIC Vector Kits Manual*, Novagen, ninth ed., USA, 2000, 56 pp. Available from: <www.novagen.com>.
- [25] J. Sambrook, D.W. Russell, Plasmids and their usefulness in molecular cloning, in: *Molecular Cloning, A Laboratory Manual*, third ed., vol. 1, CSHL Press, New York, 2001, pp. 1.32–1.34.
- [26] M.M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [27] C.N. Pace, F.X. Schmidt, How to determine the molar absorbance coefficient of a protein, in: T.E. Creighton (Ed.), *Protein Structure, A Practical Approach*, second ed., IRL Press, New York, 1997, pp. 253–259.
- [28] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227 (1970) 680–685.
- [29] M.K. Hondalus, D.M. Mosser, Survival and replication of *Rhodococcus equi* in macrophages, *Infect. Immun.* 62 (1994) 4167–4175.
- [30] S. Giguere, M.K. Hondalus, J.A. Yager, P. Darrah, D.M. Mosser, J.F. Prescott, Role of the 85-kilobase plasmid and plasmid-encoded virulence-associated protein A in intracellular survival and virulence of *Rhodococcus equi*, *Infect. Immun.* 67 (1999) 3548–3557.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação/ tese de doutorado intitulada "Caracterização de uma Aldo-ceto Redutase relacionada à Patogenicidade de *Xylella fastidiosa*":

() 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 15/2003), intitulado: Determinação de estruturas tridimensionais de proteínas relacionadas à 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: Luciana Kauer Rosselli



Orientador: Profª. Dra. Anete Pereira de Souza

Para uso da Comissão ou Comitê pertinente:

(X) Deferido () Indeferido



Nome: Prof. Dr. Marcelo Menossi Teixeira
Função: Presidente do Cíbio/CBMEG