

UNIVERSIDADE ESTADUAL DE CAMPINAS



Anita Paula Testa Salmazo

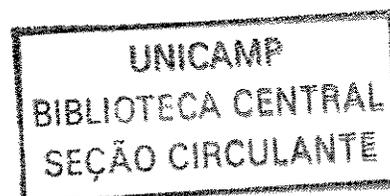
**“ESTUDOS DE RELAÇÃO ESTRUTURA-FUNÇÃO DE
PROTEÍNAS DA BACTÉRIA *Xylella fastidiosa*
ENVOLVIDAS EM PATOGENICIDADE E ADAPTAÇÃO”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
ANITA PAULA TESTA SALMAZO
[Assinatura]
e aprovada pela Comissão Julgadora.

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

Orientador: Dr. Francisco Javier Medrano Martín

Campinas
2004



UNIVERSIDADE	BC
Nº CHAMADA	
	I/UNICAMP
	Sa 35e
	EX
COMBO BC/	61231
DOC.	16-117-04
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
REÇO	11,00
DATA	28-11-04
Nº CPD	

h J d 335560

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

Sa35e

Salmazo, Anita Paula Testa

Estudos de relação estrutura-função de proteínas da bactéria *Xylella fastidiosa* envolvidas em patogenicidade e adaptação / Anita Paula Testa Salmazo. -- Campinas, SP: [s.n.], 2004.

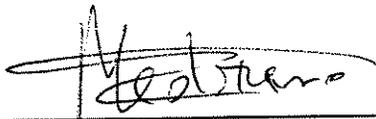
Orientador: Francisco Javier Medrano Martín

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

1. Proteínas – Estrutura I. Martín, Francisco Javier Medrano. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

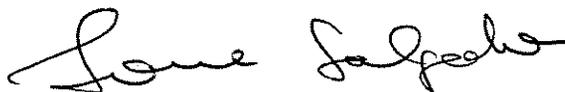
Campinas, 27 de julho de 2004.

Banca examinadora:



Dr. Francisco Javier Medrano Martín

Prof. Dr. Luis Eduardo Soares Netto



Profa. Dra. Ione Salgado

Dr. Marcos Antonio de Oliveira

Aos meus pais que me apoiam
em todos os momentos,
brigadão pela força (e que força!).

Às minhas irmãs Camila e Júlia
que mesmo a 1000km de
distância estão por perto,
sinto falta das brigas (nossa!!!).

À minha Pequena (Yasmim) que não só mudou minha vida
como também me fez rever vários conceitos e faz eu resgatar
o meu lado criança sempre que me esqueço dele.

A vida é aquilo que acontece enquanto
fazemos planos para o futuro.
(John Lennon)

Agradecimentos

À Bu, ô pessoa difícil de agradecer. Acho que se fosse possível traduzir não só a palavra amizade como também o sentimento em um só nome, sem dúvida alguma seria este: Bruna Guedes.

Ao Dr. Javier por me ensinar a andar sozinha.

À Dra. Beatriz Guimarães

À Dra. Thelma de Aguiar Pertinhez

Ao Dr. Luis Eduardo Soares Netto

À Dra. Ione Salgado

Ao Dr. João Alexandre Barbosa

Ao Dr. Gonçalo, muito importante na minha formação acadêmica.

Ao Dr. Marcos Antônio de Oliveira (meu Pai da Ciência) que me ensinou a dar os primeiros passos e é um amigo inenarrável.

Ao Dr. Sérgio Oyama Junior, Doutor não vem ao caso, amigo (Sérgio Augusto) que me levantava quando eu caía, meu companheiro no “Clube do cigarro” e me ajudou muito na parte de modelagem.

À Luciana, amiga desde a época de graduação e que sem seus sequenciamentos este trabalho não iria para lugar algum.

Ao grupo CPR: Aline, Andréia, Amanda, Camila, Deyse, Frances, Henrique, Karine, Kelly, Priscila, Rafael e Vera. Pela paciência e muitas risadas. Incluo aqui pessoas que não fazem parte deste “seleto” grupo, mas são importantes pelos mesmos motivos: Rosi, Celisa, Marina, Sílvia, Eugênia, Tereza, Adriana, Andrés, Thais, Leo, Wilian, Eurípedes, Buba, Sandra, Flávias, Marcelo, Patrícia, (...)

À Leandra pela ajuda indescritível.

Ao Fábio não só pela amizade e companherismo, mas também pela infinita ajuda em Física (cuidado com as retas paralelas!).

Ao Alexandre e Dario pela grande amizade descoberta a tempo e pelas noites em claro (aqui se encaixa o Fábio também!).

À Alessandra e Carlinha, grandes amigas, cúmplices e acessoras informais.

À Ana Deckman, além de ser do meu time (o das loiras), também pelas “canetadas” na minha tese.

Aos Amigos “extra bancada”: Rachel (meu orgulhinho), Poca (meu priminho LINDO), Nélio (Zé), Mytsuo (abra os zóios!), Ri (células anfílicas ???), Leandro (“dotô”), Regina e Paulo Nunes, Bandit, Tarcila, Estelinha, Guga, Flávia, Natália, Família Mostasso (...)

Ao Mamau, meu companheiro durante praticamente toda esta etapa. Mesmo de fora, na reta final estive ao meu lado à sua maneira.

À Família Fenocchi Guedes não só pelo Teto, mas também pelo carinho, apoio e por ter me dado mais uma família e completinha com direito a pai (Abel), mãe (Matilde), irmã (Bu), irmão (Guido), (...) e não podendo esquecer a Dona Nonna, é claro!

À Família Testa.

À Tia Rose que me ensinou a passar pelas tempestades da vida com galhardia.

À Família Salmazo.

Ao Arquiteto que sem saber apareceu na hora certa e me ajudou, também sem saber, a encarar muitas coisas de forma muito diferente do que eu estava acostumada até então.

Ao LNLS.

À FAPESP pelo apoio financeiro.

E por último, sem menos importância e sem dúvida, às bactérias ...

Resumo

O sistema regulador de dois componentes responde a mudanças no meio ambiente, regulando genes envolvidos na adaptação de bactérias. Este sistema, normalmente é formado por duas proteínas, uma proteína sensora que nota a mudança ambiental, e outra proteína reguladora que regula os genes necessários para a resposta adequada. Assim, visando caracterizar as proteínas Xf0389 (reguladora) e Xf0390 (sensora) de *Xylella fastidiosa*, experimentos desde amplificação gênica até purificação protéica foram iniciados.

Enquanto o gene Xf0390 foi amplificado e clonado em vetor pGEM-T-Easy, o gene Xf0389 foi inserido em vetor de clonagem e após a confirmação de similaridade do fragmento extraído do vetor com o gene descrito de *X. fastidiosa*, o fragmento foi subclonado em vetores de expressão (pET28a(+), pET29a(+)) e pGEX4T-3). A proteína foi expressa em diferentes linhagens de *Escherichia coli*, para obtenção da mesma em sua forma solúvel. Através de experimentos de dicroísmo circular, a proteína foi caracterizada como estável e predominantemente rica em α -hélice. Esta última característica da estrutura secundária também foi mostrada pelo modelo construído para a proteína Xf0389, baseado na estrutura tridimensional da proteína “DNA binding response regulator D” (DrrD) de *Thermotoga maritima* que pertence a subfamília OmpR/PhoB.

Abstract

The two-component regulatory system are signal transduction strategies used by bacteria in order to sense the surrounding the environment. This system is composed by two proteins, a transmembranic sensor protein (PhoQ) that interacts with the other one, a response regulator protein (PhoP) that once it has been phosphorylated, it generates a response by regulating the expression of several genes, involved in virulence and adaptation. In *Xylella fastidiosa* this system is formed, at least, by these two genes: Xf0389 (PhoP) and Xf0390 (PhoQ).

The Xf0390 was cloned only in pGEM-T-Easy, while the Xf0389 was also cloned into the expression vector pET28a(+) and the protein was expressed in *E. coli* C43 (DE3) strain. Circular dichroism spectra of the purified protein show a high content in alpha-helix.

A three-dimensional model of Xf0389 was built and the modeling was based on the protein DNA binding response regulator D (DrrD) from *Thermotoga maritima* (PDB accession code: 1kgs). Typically, response regulators have two domains: a N-terminal regulatory domain which is phosphorylated and a C-terminal effector domain which has a DNA binding motif. Based on the sequence of the effector domain, there are three subfamilies: NtrC, NarL/FixJ and OmpR/PhpB. Xf0389 has characteristics of the proteins belonging to the OmpR/PhoB subfamily.

Abreviaturas

CD: dicroísmo circular

DMSO: dimetil-sulfóxido

EDTA: etileno-diamino-tetracético

FPLC: “Fast Protein Liquid Chromatography”

IPTG: isopropil- β -D-tiogalactopiranosídeo

PCR: reação de polimerização em cadeia

PDB: “Protein Data Bank”

PEG: polietilenoglicol

PMSF: fenil-metil-sulfonil-fluorido

SDS-PAGE: eletroforese em gel de poli-acrilamida sob condições desnaturantes

X-gal: 5-bromo-4-cloro-3-indol- β -D-galactopiranosídeo

Glossário

supE44: supressor de tRNA glutamina

Δ *lacU169*: deleção número 169 que afeta os genes do operon lactose

(Φ 80 *lacZ* Δ M15): possui o profago 80; fusão que permite α -complementação com o N-terminal da β -galactosidase codificada

recA⁻: não permite recombinação homóloga

endA⁻: mutação de endonuclease I (melhor para preparação de plasmídios)

gyrA96: mutante para DNA girase

thi-1: célula necessita de tiamina para crescer em meio mínimo

relA1: RNA é sintetizado na ausência de síntese protéica

Δ (*lac-proAB*): célula necessita de prolina para crescer em meio mínimo

F' [*traD36 proAB⁺ lac^q lacZ* Δ M15]: o epissomo F' nessa linhagem carrega *lacZ* Δ M15, que permite α -complementação com o N-terminal da β -galactosidase codificada pelo λ ZAP

ompT⁻: não possui o gene que codifica a protease *ompT*, que é de membrana e pode clivar proteínas heterólogas durante a purificação

*hdsB*_B(*r_B⁻m_B*): não permite a clivagem do DNA por enzimas de restrição endógenas e não é protetor a metilação

gal: mutação que afeta o gene da galactose

dcm⁻: mutação da DNA citosina metilase, perda da metilação em resíduos de citosina nas seqüências CCAGG e CCAGG

λ (DE3): resistência/sensibilidade ao fago DE3

pLysS: T7 lisozima reduz expressão basal dos genes alvos e ainda estabiliza os vetores que expressam proteínas tóxicas em *E. coli*

(Cm^R): resistência a clorafenicol

lacY1: mutação *lacY*, perda da lactose permease

pRARE²: possui anticódons de mamíferos (raros em *E. coli*), que aumentam o nível de expressão protéica

(Ap^R): resistência a ampicilina

dcm⁺: metilação em resíduos de citosina nas seqüências CCAGG e CCAGG

Tet^R: resistência a tetraciclina

Índice

1) Introdução	5
1.1) <i>Xylella fastidiosa</i>	5
1.2) Patogenicidade	7
1.3) Sistema regulador de dois componentes	8
2) Justificativa	15
3) Objetivos	16
4) Materiais e métodos	17
4.1) Linhagens bacterianas	17
4.2) Predição de segmentos transmembrana	18
4.3) Amplificação dos genes Xf0389 e Xf0390 e amplificação por PCR de colônia	18
4.4) Clonagem em vetor pGEM-T-Easy	19
4.5) Sequenciamento	19
4.6) Subclonagem em vetores de expressão	20
4.7) Vetores de expressão	20
4.9) Preparação de bactérias competentes	20
4.10) Transformação de bactérias	21
4.11) Seleção de clones positivos mediante uso de antibióticos	21
4.12) Extração de plasmídios em mini escala	22
4.13) Extração de plasmídios em média escala	22
4.14) Eletroforese em gel de agarose	23
4.15) Ensaios de expressão de proteína em mini escala	23
4.16) Purificação da proteína codificada por Xf0389	23
4.17) Concentração da proteína	24
4.18) Determinação da concentração de proteína	24
4.19) Clivagem da cauda de histidina	25
4.20) Proteólise limitada	25
4.21) Eletroforese em gel de acrilamida	25
4.22) Dicroísmo circular	25
4.23) Modelagem molecular	26
4.23.1) Predição de estrutura secundária	26
4.23.2) Busca de estruturas conhecidas	26
4.23.3) Alinhamento de seqüências	27
4.23.4) Construção do modelo por homologia	27
4.23.5) Avaliação do modelo	27
4.23.6) Recursos computacionais	28

5)	Resultados	29
5.1)	Predição transmembrana da proteína Xf0390	29
5.2)	Amplificação e clonagem dos genes Xf0389 e Xf0390 de <i>Xylella fastidiosa</i>	29
5.3)	Subclonagem em vetores de expressão	30
5.4)	Ensaio de expressão	33
5.5)	Purificação de proteínas	34
5.6)	Caracterização estrutural	36
5.7)	Estabilidade térmica	37
5.8)	Construção do modelo por homologia	38
5.9)	Proteólise limitada	44
6)	Discussão	45
6.1)	Amplificação e clonagem dos genes Xf0389 e Xf0390	45
6.2)	Ensaio de expressão	45
6.3)	Purificação da proteína codificada por Xf0389	47
6.4)	Caracterização estrutural	47
6.5)	Estabilidade térmica	47
6.6)	Construção do modelo por homologia	48
6.7)	Proteólise limitada	50
6.8)	Estudo comparativo	51
6.8.1)	Xf0389	51
6.8.2)	Xf0390	52
7)	Conclusões	54
8)	Referências bibliográficas	55
	Apêndice I	62

Lista de figuras

Figura 1:	Esquema do mecanismo geral do sistema regulador de dois componentes	10
Figura 2:	Alinhamento entre proteínas reguladoras representantes das subfamílias OmpR/PhoB, NarL/FixJ e NtrC	12
Figura 3:	Predição de segmentos transmembrana da proteína Xf0390	29
Figura 4:	Produto de PCR dos genes Xf0389 e Xf0390 (região II)	30
Figura 5:	Produto de PCR do gene Xf0390 (região I)	30
Figura 6:	Produtos das reações de digestão dos vetores de clonagem	31
Figura 7:	Seleção de clones positivos dos vetores de expressão	32
Figura 8:	Seleção de clones positivos em diversas linhagens	32
Figura 9:	Ensaio de expressão da proteína codificada por Xf0389	34
Figura 10:	Cromatografia de afinidade a níquel	35
Figura 11:	Clivagem da cauda de histidina	36
Figura 12:	Espectro de CD	36
Figura 13:	Ensaio de estabilidade térmica	37
Figura 14:	Predição da estrutura secundária de Xf0389	38
Figura 15:	Alinhamento seqüencial e estrutural (Xf0389 x DrrD)	39
Figura 16:	Mapa de Ramachandran	41
Figura 17:	Detalhamento do mapa de Ramachandran	42
Figura 18:	Modelo tridimensional de Xf0389	42
Figura 19:	Representação da superfície eletrostática de Xf0389	43
Figura 20:	Proteólise limitada	44
Figura 21:	Domínio regulador da proteína Xf0389	49
Figura 22:	Sobreposição estrutural	50
Figura 23:	Detalhamento do alinhamento entre histidinas quinases	53

Lista de tabelas

Tabela 1:	Incidência da CVC nos estados de SP e MG nos anos de 1996 a 2001	6
Tabela 2:	Linhagens utilizadas para clonagem	17
Tabela 3:	Linhagens utilizadas para expressão	17
Tabela 4:	Antibióticos utilizados na seleção de clones positivos	21
Tabela 5:	Resumo das alternativas tentadas para obtenção da proteína Xf0389 em sua forma solúvel	33

1) Introdução

1.1) *Xylella fastidiosa*

Xylella fastidiosa é um bacilo gram-negativo que apresenta parede celular rugosa. É uma bactéria patogênica e se encontra no xilema das plantas infectadas, causando doenças como a clorose variegada de citros (CVC), também conhecida como amarelinho (Chang *et al.*, 1993) e a doença de Pierce (PD) na videira, sendo também encontrada na alfafa, amêndoa, café, olmo, sicômoro, carvalho, amoreira, pêssigo e ameixa (Purcell e Hopkins, 1996).

No Brasil, a CVC foi primeiramente detectada nos estados de São Paulo e Minas Gerais em 1987, e atualmente é considerada a mais devastadora doença da citricultura, responsável por grandes perdas econômicas. Os sintomas da doença são variegação visível nas folhas e produção de frutos sem valor comercial, por serem pequenos e duros (Roberto *et al.*, 2002). Após a infecção, o aparecimento dos sintomas pode levar de 6 a 14 meses, uma vez que o período de incubação da CVC não é bem conhecido (Bergamin Filho e Amorin, 2002).

A CVC é um grande problema no Brasil, onde se estima que 70 milhões de laranjeiras (36 % da plantação) estejam infectadas (McElrone e Forseth, 2001, www.fapesp.br). Esta doença afeta todas as variedades de laranjas doce (*Citrus sinensis* Osb.) e segundo as últimas pesquisas feitas pela *Fund for Citrus Plant Protection* (www.fundecitrus.com.br) em uma amostra de 1655 plantas, distribuída entre quatro variedades diferentes de laranjas, há um aumento da incidência de CVC nos estados de São Paulo e Minas Gerais ao longo dos anos de 1996 a 2001 (tabela 1).

Tabela 1: Incidência da CVC nos estados de SP e MG nos anos de 1996 a 2001.

Ano	Plantas com sintomas restritos a folhas (%)	Plantas com sintomas foliares e frutos miúdos (%)	Total (%)
1996	15,93	6,17	22,09
1997	22,89	9,39	32,28
1998	13,68	7,55	21,24
1999	20,95	15,13	36,08
2000	13,23	20,80	34,03
2001	12,37	24,07	36,44

A transmissão da doença de planta a planta é efetuada por insetos que se alimentam do xilema das mesmas. No Brasil, comprovadamente onze espécies de cigarrinhas estão envolvidas na transmissão da CVC, de forma que, ao se alimentar do xilema de uma planta doente, o vetor adquire a bactéria que se instala em seu aparelho bucal, transmitindo a doença a plantas saudáveis. O grande número de vetores, bem como a facilidade de disseminação da doença, dificultam as estratégias de controle da mesma, o qual atualmente está baseado na eliminação da planta ou de suas partes infectadas e no uso de inseticidas (Araújo *et al.*, 2002, www.fundecitrus.com.br).

Como revisado por Hopkins (1989), o mecanismo de patogenicidade das doenças causadas pela bactéria *X. fastidiosa* não está completamente elucidado, existindo três possibilidades: disfunção do sistema condutor de água, produção de fitotoxinas e desequilíbrio do sistema regulador de crescimento. A maioria dos sintomas da doença sugerem oclusão vascular gerando desbalanceamento hídrico e deficiência nutricional na planta. Esta oclusão vascular pode ser provocada por agregados de bactérias que se encontram presos à parede dos vasos por fímbrias, que se assemelham às fibras de polissacarídeos constituintes do glicocálix, estrutura esta responsável pela fixação de algumas bactérias em ambientes turbulentos, por exemplo ambientes de fluxo muito rápido, como o xilema das plantas. O glicocálix pode ter a função de retenção de nutrientes para o agregado bacteriano ou conservar e concentrar enzimas digestivas liberadas pela bactéria para ação nos tecidos da planta hospedeira, além de agir como proteção

contra mecanismos de defesa da planta. A manutenção deste agregado é um importante aspecto patogênico, assim a *X. fastidiosa* também produz uma matriz polimérica denominada biofilme, que mantém o agregado bacteriano unido e ligado à superfície do xilema ou no aparelho bucal da cigarrinha, garantindo sua sobrevivência e otimizando o aproveitamento dos nutrientes (Marques *et al.*, 2002 e Smolka *et al.*, 2003).

1.2) Patogenicidade

Os patógenos bacterianos infectam uma grande variedade de hospedeiros muito diferentes evolutivamente. Em todos os casos, o organismo patogênico tem que reconhecer, associar-se, aproveitar as reservas de nutrientes e combater os mecanismos de defesa do hospedeiro específico (Hopkins, 1989). Para atingir este fim, os patógenos possuem um arsenal de fatores relacionados à virulência.

Muitos patógenos causam doenças em apenas uma espécie ou em um número limitado de espécies hospedeiras, como consequência de uma história evolutiva comum. Entretanto, entende-se muito superficialmente como as interações entre patógeno e hospedeiro limitam a diversidade deste último e determinam sua resistência ou susceptibilidade. Apesar da idéia de que muitos componentes da virulência bacteriana sejam específicos para o hospedeiro, numerosos estudos têm demonstrado a existência de mecanismos comuns usados por diversas espécies bacterianas em diferentes hospedeiros (Finlay e Falkow, 1997). Estudos similares também revelaram a existência de características comuns nas respostas de defesa contra patógenos em plantas, insetos e mamíferos (Kopp e Medzhitov, 1999). É muito provável que os mecanismos de virulência dos patógenos e de defesa do hospedeiro possuam uma mesma origem evolutiva que é conservada.

Para compreender e combater os componentes patogênicos das bactérias é preciso conhecer a base molecular dos mecanismos de infecção e as interações patógeno-hospedeiro. O genoma completo e um mega-plasmídeo da bactéria *X. fastidiosa* foram seqüenciados recentemente (Simpson *et al.*, 2000) e na anotação do genoma foram identificados alguns genes que sugerem a existência de vários mecanismos de patogenicidade, tais como hidrólise da parede celular, adesão, migração intervascular e toxicidade. Todavia, a função destes genes deve ser

determinada experimentalmente. Uma forma eficaz para a obtenção destas informações é o estudo de seus produtos gênicos: as proteínas.

Dessa forma, pode-se observar que este fitopatógeno encontra-se bem adaptado a colonização do xilema de citrus e como o mecanismo de patogenicidade da *X. fastidiosa* não é bem conhecido, para o presente estudo, foram escolhidas proteínas envolvidas na adaptação da bactéria a mudanças ambientais e estas formam o sistema regulador de dois componentes.

1.3) Sistema regulador de dois componentes

Enquanto a cascata de sinalização em muitos eucariotos envolve proteínas quinases que se autofosforilam e fosforilam outras proteínas em resíduos de serina, treonina ou tirosina, regulando assim a atividade protéica, em procariotos o sistema de sinalização envolve outro esquema de fosforilação predominante. Este é conhecido como sistema regulador de dois componentes, composto por uma proteína histidina quinase e uma proteína reguladora que são fosforiladas em um resíduo de histidina e outro de aspartato, respectivamente (West e Stock, 2001). Porém, há bactérias que não possuem este tipo de sistema como, por exemplo, o *Mycoplasma spp.*, enquanto outras, como *Pseudomonas aeruginosa*, chegam a apresentar um número variável deste sistema (Rodrigue *et al.*, 2000), também foi descrito em nematódeos simbióticos de entomopatógenos, como *Photorhabdus luminescens* (Derzelle *et al.*, 2004).

Os diferentes sistemas existentes estão sintonizados com seus estímulos específicos, como a presença de fosfato, nitrogênio ou fontes de carbono, sendo que estes utilizam mecanismos homólogos de controle molecular (Volz e Matsumura, 1991).

Foi proposto que seu mecanismo geral de ação (figura 1) consista na percepção de sinais ambientais pela porção N-terminal da proteína sensora que, ao ser sensibilizada, transmita tais informações e module a atividade enzimática de sua porção C-terminal, após dimerização da mesma. Esta transdução de sinais resulta na transferência de um grupo fosfato de alta energia do resíduo de histidina de seu domínio C-terminal, para um resíduo de aspartato conservado na porção N-terminal da proteína reguladora (Waldburger e Sauer, 1996).

A transferência do grupo fosfato da histidina para o aspartato ativa a proteína reguladora, resultando em uma resposta ao estímulo externo. O mecanismo de autofosforilação do resíduo de histidina específico da proteína sensora é dependente de ATP e a catálise de transferência do grupo fosforil para o aspartato é feita pela própria proteína reguladora, assim como sua desfosforilação. Portanto, a proteína reguladora também apresenta uma atividade fosfatase, limitando seu tempo de ação que está associado a sua função fisiológica, sendo uma estratégia de regulação do sistema (Hess *et al.*, 1987). Porém, estudos recentes em *E. coli* têm mostrado que a proteína sensora também possui uma atividade fosfatase, já que esta desfosforilação envolve uma transferência reversa do fosfato do aspartato (57) conservado da proteína reguladora (PhoP) de volta para a histidina (243) da proteína sensora (PhoQ), pois mutantes portando valina no lugar da histidina não possuem atividade fosfatase (Castelli *et al.*, 2000).

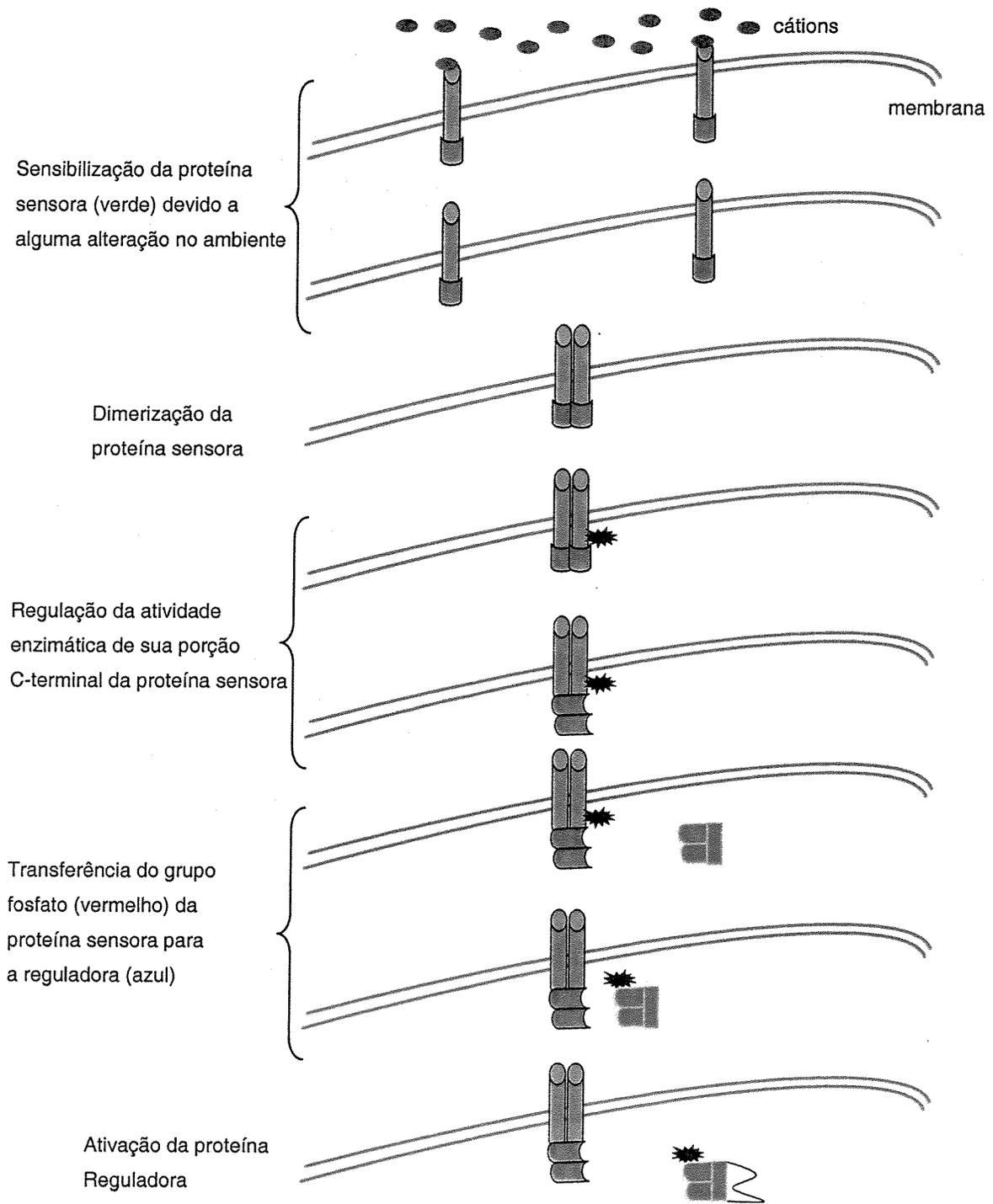


Figura 1: Esquema do mecanismo geral do sistema regulador de dois componentes

A proteína reguladora usualmente apresenta dois domínios estruturais: um N-terminal regulatório conservado e um C-terminal efetor variável que regula a transcrição de diversos genes. O domínio regulador possui três atividades: (i) interação com a histidina fosforilada da proteína sensora e transferência deste grupo para seu resíduo de aspartato; (ii) autodesfosforilação e (iii) regulação da atividade de seu domínio efetor de maneira dependente de fosforilação. Esta regulação pode ser feita através da desfosforilação da porção regulatória, assim inativando a porção efetora; pela ativação alostérica do domínio efetor através da fosforilação do domínio regulador; pela dimerização ou oligomerização da proteína reguladora fosforilada; e ainda pela interação da proteína reguladora com alvos protéicos heterólogos (Schenk e Snaar-Jagalska, 1999; West e Stock , 2001). Várias destas proteínas tiveram sua estrutura tridimensional resolvida (Baikalov *et al.*, 1996; Baikalov *et al.*, 1998; Pelton *et al.*, 1999; Okamura *et al.*, 2000; Buckler *et al.*, 2002) a fim de se obter um melhor entendimento de seus mecanismos de regulação no sistema, uma vez que a compreensão da indução por fosforilação é dificultada pela vida curta da ligação aspartato-fosfato, de poucos segundos a algumas horas (Feher e Cavangh, 1999). Observou-se através das estruturas resolvidas que todos os fatores reguladores apresentam um enovelamento similar, com vários resíduos conservados em seu domínio regulador, principalmente o aspartato que é fosforilado (Bourret *et al.*, 1990; Djordjeric e Stock, 1998; Stock *et al.*, 2000). Por outro lado, seu domínio efetor não possui elementos conservados, sendo esta característica utilizada para classificar estas proteínas em três subfamílias de reguladores de transcrição de acordo com a seqüência de seu domínio efetor: OmpR/PhoB, NarI/FixJ e NtrC (figura 2).



Figura 2: Alinhamento das seqüências de proteínas reguladoras representantes das três subfamílias, mostrando que o domínio regulador (N-terminal) apresenta vários resíduos conservados entre elas e que o domínio efetor (C-terminal) não apresenta resíduos conservados. O aspartato que é fosforilado para ativação da proteína está indicado em amarelo e em azul, outro aspartato conservado e importante para a conformação da proteína. O alinhamento foi gerado pelo programa *Multiple sequence alignment* (Corpet, 1988).

Com relação à proteína quinase, esta atua como um homodímero e foram definidas duas classes de enzimas de acordo com a organização de seus

domínios funcionais. As enzimas da classe I são mais comumente encontradas e possuem regiões transmembrânicas, de comprimento variável. Já as da classe II são citoplasmáticas, especializadas na resposta a motilidade da bactéria e se acoplam a proteínas de membrana (Foussard *et al.*, 2001).

O sistema regulador de dois componentes está envolvido na adaptação das bactérias às mudanças no meio ambiente e tem um papel fundamental na virulência de muitos microorganismos patogênicos (Hoch e Silhavy, 1995). Um exemplo é o envolvimento deste sistema na formação de biofilme (matriz de polissacarídeos e glicoproteínas) em linhagens patogênicas de *E. coli*, sendo que estudos recentes indicam a ação de OmpR (Dorel *et al.*, 1999). Outro exemplo é o sistema PhoPQ em *Salmonella typhimurium* que tanto ativa a expressão de mais de 40 polipeptídios essenciais à virulência e sobrevivência da bactéria no interior do macrófago em resposta a uma acidificação, como também é capaz de reprimir alguns genes (Aranda *et al.*, 1992; Garcia Véscosi *et al.*, 1994, Soncini *et al.*, 1996). Esta bactéria é um patógeno intracelular facultativo e causa febre tifóide em mamíferos. Dentro das células destes hospedeiros, PhoQ responde à ligação de cátions divalentes que atuam como um ligante sinalizador, induzindo mudanças conformacionais que afetam a atividade quinase ou fosfatase da porção citoplasmática. Quando a bactéria está sendo fagocitada pelo macrófago, os fatores de virulência estão reprimidos devido à alta concentração de íons de Ca^{2+} ou Mg^{2+} (passo 1, figura 1). Uma vez dentro do macrófago, a bactéria se aloja no fagossomo onde o nível destes cátions diminui (passo 2, figura 1), ativando o sistema, resultando na síntese de proteínas necessárias para a sobrevivência e virulência da bactéria (Garcia Vescovi *et al.*, 1996). Estudos demonstraram que este sistema de *S. typhimurium* está envolvido em mecanismos de defesa da bactéria a peptídeos antimicrobianos catiônicos, que são um dos mecanismos responsáveis pela defesa do hospedeiro (mamífero) contra infecções. A proteína efetora phoP, neste caso, ativa primeiramente o gene *pagB* que está organizado em um operon com *pmrA* e *pmrB*, o qual regula outros genes envolvidos na resistência aos peptídeos antimicrobianos (Gunn e Miller, 1996), sendo que alguns destes determinantes modificam os lipopolisacarídeos, enquanto outros codificam proteases extracitoplasmáticas com a capacidade de clivar os peptídeos antimicrobianos (Groisman *et al.*, 1997).

Várias bactérias não patogênicas também possuem este tipo de sistema, como por exemplo o sistema FeuPQ na bactéria *Rhizobium leguminosarum*, necessário na regulação de genes envolvidos na aquisição de Fe^{2+} (Yeoman *et al.*, 1997). Já em *Bacillus subtilis*, o sistema de dois componentes governa a adaptação a condições limitantes de fosfato (Hulett, 1996), enquanto que *E. coli*, possui um sistema que responde basicamente à falta de Mg^{2+} (Garcia Vescovi *et al.*, 1996) e outro sistema que responde a íons Ca^{2+} (Tisa e Adler, 1992).

A resposta a Ca^{2+} , também envolvida na quimiotaxia bacteriana, envolve a presença de, pelo menos, duas proteínas: CheA (quinase) e CheY (reguladora). A transiente CheY fosforilada (ativada) interage com a matriz flagelar causando uma rotação dos flagelos, mudando a célula de direção. O papel central de CheY foi muito estudado através de comportamento e caracterização bioquímica da fosforilação de mutantes e mutantes de matriz flagelar, a fim de se caracterizar a base bioquímica do sistema (Sanders *et al.*, 1989) e ainda, a estrutura tridimensional desta proteína foi resolvida tanto por métodos de difração de raios-X (Volz e Matsumura, 1991; Lee *et al.* 2001) como também por ressonância magnética nuclear (RMN; Bruix *et al.* 1993; Cho *et al.*, 2000), servindo como modelo para o entendimento da indução por fosforilação dos componentes reguladores destes sistemas (Volz, 1993).

No genoma da bactéria *Xylella fastidiosa*, foram localizados vários genes que apresentam homologia com os genes do sistema regulador de dois componentes. Assim, com o objetivo de aumentar o conhecimento em relação ao mecanismo de ação da bactéria e ao sistema regulador de dois componentes foram escolhidos para o presente trabalho o estudo dos genes de *X. fastidiosa* Xf0389 (phoP) que codifica para uma proteína reguladora de 25,3 kDa e o gene Xf0390 (phoQ) que codifica para uma proteína sensora de 53,6 kDa. Esta última apresenta características relacionadas a quinase da classe I, apresentando dois domínios transmembrânicos.

2) Justificativa

A clorose variegada de citrus (CVC) afeta a todas as variedades comerciais de laranja doce, sendo responsável por grandes perdas econômicas na produção de frutas cítricas do Brasil. O agente causador dessa doença é a bactéria *Xylella fastidiosa*, cujos mecanismos de patogenicidade e virulência são muito pouco conhecidos. Portanto, estudos em genômica funcional e estrutural de *X. fastidiosa* serão de grande importância para o melhor entendimento destes mecanismos, contribuindo significativamente para a descoberta de possíveis formas de tratamento da doença.

O sistema escolhido para este estudo pode estar envolvido na capacidade de adaptação e na virulência da bactéria fitopatogênica *X. fastidiosa*. O sistema regulador de dois componentes está, muito provavelmente, envolvido na detecção de mudanças no meio ambiente e na ativação de uma variedade de genes que permitem a colonização do xilema da planta ou do aparelho bucal das espécies de cigarrinha responsáveis pela transmissão da doença. Acredita-se que o conhecimento de sistemas envolvidos na patogenicidade levará a um melhor entendimento dos mecanismos utilizados pela bactéria na colonização, estabelecimento e crescimento dentro da planta hospedeira.

3) Objetivos

Um dos objetivos do Projeto SMOLBNet é a continuação dos projetos Genoma FAPESP com a resolução da estrutura de proteínas de importância biotecnológica. E como este projeto de mestrado está vinculado ao SMOLBNet, o objetivo deste é iniciar investigações sobre a estrutura de proteínas de sistemas relacionados a adaptação e virulência da bactéria *Xylella fastidiosa*, agente causador da clorose variegada de citrus.

Objetivos específicos:

1. Clonagem de dois genes do sistema regulador de dois componentes de *X. fastidiosa* (Xf0389 e Xf0390)
2. Expressão e purificação das proteínas codificadas por tais genes
3. Caracterização funcional e estrutural por meio de técnicas de cristalização e difração de raios X das proteínas em estudo.

Uma vez iniciado os experimentos e constatada a dificuldade de clonagem da histidina quinase (proteína sensora), os objetivos do projeto passaram a visar somente a proteína reguladora:

1. Clonagem do gene Xf0389
2. Expressão e purificação da proteína codificada por Xf0389
3. Caracterização estrutural da mesma.

4) Materiais e métodos

4.1) Linhagens bacterianas

As linhagens de *Escherichia coli* utilizadas neste estudo e seus respectivos genótipos estão especificados nas tabelas 2 e 3.

Tabela 2: Linhagens utilizadas para clonagem e seus respectivos genótipos. (Sambrook e Russel, 2001).

Linhagens de <i>E. coli</i>	Genótipos
DH5α	<i>supE44 ΔlacU169 (Φ80 lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1</i>
JM109	<i>recA1 supE44 ΔlacU169 endA1 hsdR17 gyrA96 relA1 thi Δ(lac-proAB) F' [traD36 proAB⁺ lac⁺ lacZ ΔM15]</i>

Tabela 3: Linhagens utilizadas para expressão de proteínas e seus respectivos genótipos. (www.novagem.com.br; www.avidis.fr; www.invitrogen.com.br; Donahue e Bebee, 1999; Yasukawa *et al.*, 1995).

Linhagens de <i>E. coli</i>	Genótipos
BL21 SI	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm (proU, T7 RNAP, malQ-lacZ) Tet^S</i>
BL21 (DE3)	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3)</i>
BL21 (DE3) pLysS	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3) pLysS (Cm^R)</i>
BL21 (DE3) pLysE	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3) pLysE (Cm^R)</i>
BL21 (DE3) pTGROE	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm (Cm^R) pTGroE λ(DE3)</i>
Rosseta (DE3)	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm lacY1 pRARE² (Cm^R) λ(DE3)</i>
BL21 (DE3) pT-Trx	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm (Cm^R) λ(DE3) pT-Trx</i>
C41 (DE3)	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3)</i> (contendo no mínimo mais uma mutação não caracterizada em relação à BL21(DE3))
C43 (DE3)	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3)</i> (contendo no mínimo mais duas mutações não caracterizadas, em relação à BL21(DE3))
C41 (DE3) pAVD10	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3) pAV D 10 (Ap^R)</i> (contendo no mínimo mais uma mutação não caracterizada em relação à BL21(DE3))
C43 (DE3) pAVD10	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3) pAV D 10 (Ap^R)</i> (contendo no mínimo mais duas mutações não caracterizadas, em relação à BL21(DE3))
BL21 Star (DE3)	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) gal dcm λ(DE3) me131</i>
BL21 Codon plus (DE3) Ril	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) dcm⁺ Tet^r gal λ(DE3) endA Hte [argU ileY leuW Cm^R]</i>
BL21 Codon plus (DE3) RP	<i>E. coli</i> B F <i>ompT hsdS_B(r_B m_B) dcm⁺ Tet^r gal λ(DE3) endA Hte [argU proL Cm^R]</i>

4.2) Predição de segmentos transmembrana

A seqüência da proteína Xf0390 foi submetida a predição de segmento transmembrana por meio do programa DAS – *Transmembrane Prediction Server* (Cserzo *et al.*, 1997), acessível através da interface *web* localizada em <http://www.sbc.su.se/~miklos/DAS/>. Baseando-se na seqüência primária da proteína, o programa analisa os resíduos de aminoácidos em função de sua hidrofobicidade.

4.3) Amplificação dos genes Xf0389 e Xf0390 e amplificação por PCR de colônia

Os genes alvo foram amplificados por PCR, utilizando DNA genômico da bactéria *X. Fastidiosa*, gentilmente cedido ao nosso laboratório pela Profa. Dra. Anete Pereira de Souza (IB, UNICAMP). Foram desenhados oligonucleotídeos específicos a partir das seqüências gênicas da bactéria depositadas em um banco de dados (www.lbi.ic.unicamp.br/xf/), adicionando-se um sítio de restrição em cada extremidade para facilitar a subclonagem em vetores de expressão.

Oligonucleotídeos específicos:

→ Xf0389

Tamanho esperado para o fragmento amplificado: 684 pb

Sítios de restrição: NdeI/XhoI

389F 5' CATATGCGTATCCTTTTGGTCTGAAG 3'

389R 5' CTCGAGTTAATTTTCGTTGCGGGCAATTG 3'

→ Xf0390

1. Região I (equivalente aos aminoácidos 51 à 154)

Tamanho esperado para o fragmento amplificado: 310pb

Sítio de restrição: NheI/XhoI

390F1 5' GCTAGCGATACCGCTGAAAAGAATTTACG 3'

390R1 5' CTCGAGTTAAGCCGTGCGAAATACACGCAG 3'

2. Região II (equivalente aos aminoácidos 232 à 482)

Tamanho esperado para o fragmento amplificado: 800pb

Sítio de restrição: NdeI/XhoI

390F2 5' CATATGCGTCCAATAAGGCACGTCATTAAC 3'

390R2 5' CTCGAGTTATAACCCTGGCAATACCACCTC 3'

A reação de PCR foi feita para um volume final de 20 μ L, utilizando-se 1,25 mM MgCl₂, 0,25 mM dNTP's, 2,5 unidades *Taq Polymerase*, 1x tampão de PCR (10 mM Tris-HCl pH 8,0, 50 mM KCl, 0,08% Nonidet P-40), 10 pmol de cada oligo (F/R) e 10 a 20 ng de DNA genômico da bactéria. Os ciclos utilizados foram padronizados de acordo com a temperatura de anelamento.

Para as ampliações de DNA por PCR diretamente da colônia, utilizou-se oligonucleotídeos específicos e as reações foram feitas da forma descrita anteriormente, adicionando 1 μ L da mistura contendo a colônia, preparada a partir de uma diluição em 100 μ L de H₂O Milli Q, incubadas à 95°C por 5 min e depois centrifugadas (10000 g, 1 min, temperatura ambiente), promovendo a precipitação da membrana e outras impurezas que poderiam inibir a reação de amplificação.

4.4) *Clonagem em vetor de clonagem pGEM-T-Easy*

Os fragmentos amplificados foram purificados pelo Kit de purificação (Qiagen) e clonados em plasmídeo de acordo com instruções do fabricante (Promega). As colônias transformantes foram selecionadas em meio TB sólido (Terrific broth: 1,2% triptona, 2,4% extrato de levedura, 0,4% glicerol, 1% solução salina: 0,17 M KH₂PO₄, 0,72 M K₂HPO₄, 1,5% agar) contendo 100 μ g/mL ampicilina e 16 μ g/mL X-gal sendo posteriormente submetidos a reações de sequenciamento.

4.5) *Sequenciamento*

As reações de sequenciamento foram feitas com base na técnica de Sanger e colaboradores (1977), utilizando-se um seqüenciador automático de DNA ABI PRISM 377 Genetic Analyser (Applied Biosystems). Tanto as reações de sequenciamento quanto o ciclo das reações seguiram instruções do fabricante do *kit* para sequenciamento de DNA da empresa Applied Biosystem. Os oligonucleotídeos utilizados para a reação foram T7 ou SP6 para o vetor utilizado ou específicos aos genes (5 pmol), tendo como molde os vetores de clonagem e expressão ou produto de PCR de colônia, respectivamente.

4.6) Subclonagem em vetores de expressão

Após confirmação da identidade do inserto contido no vetor de clonagem, os plasmídios selecionados foram submetidos a reações de digestão conforme sítio de restrição presente nas extremidades 3' e 5' de cada oligonucleotídeo específico. Os insertos foram isolados do gel utilizando-se o Kit de extração (Qiagen) e posteriormente subclonados em vetores de expressão (pET28a(+), pET29a(+)) e pGEX4T-3), digeridos com as mesmas enzimas que formam os sítios de restrição do fragmento amplificado, com 1 unidade de T4 Ligase, de acordo com o kit de ligação do fabricante (Promega).

4.7) Vetores de expressão

Os vetores utilizados para subclonagem dos fragmentos gênicos foram pET28a(+), o qual possui uma seqüência codificadora de uma cauda de seis resíduos de histidina, fusionada à proteína expressa. O vetor pET29a(+) que expressa a proteína sem a cauda de histidina e pGEX4T-3, que gera a proteína alvo fusionada à Glutationa-S-Transferase (GST) de *Schistosoma japonicum*.

4.8) Princípio da purificação de DNA

Tanto as extrações do fragmento de DNA do gel de agarose quanto a purificação das reações de PCR em solução, foram baseadas no Kit da Qiagen que tem como princípio a absorção do DNA pela membrana de sílica da coluna em altas concentrações de sais e pH adequado (~7,5), enquanto as impurezas (ólígos, enzimas, agarose e brometo de etídio) passam através da membrana durante as lavagens e o DNA puro é eluído da membrana com 10 mM Tris-HCl pH 8,0 ou água Milli Q.

4.9) Preparação de bactérias competentes

As bactérias *E. coli* foram preparadas seguindo o protocolo baseado em cloreto de rubídio (RbCl) (adaptado de Maniatis *et al.*, 1986). Uma colônia isolada de bactéria é crescida durante a noite, sob agitação, à 37°C, em 5 mL de meio TB. A 3 mL do pré-inóculo foram adicionados 300 mL de TB e novamente incubados sob agitação (37°C) até atingir a OD₆₀₀ 0.2. Alcançada esta densidade ótica, as células foram centrifugadas (5000 g, 5 min, 4°C) e o sobrenadante descartado. As

células foram ressuspendidas em 60 mL de tampão de transformação (100 mM RbCl, 45 mM MnCl₂, 10 mM CaCl₂, 5 mM MgCl₂, 0.5 mM LiCl, 35 mM acetato de potássio pH 6,2 e 15% sacarose), e as células foram coletadas por centrifugação (5000 g, 5 min, 4°C), sendo ressuspendidas em 12 mL de tampão de transformação, adicionando-se 420 µL de DMSO (35 µL/mL de cultura) e alíquotas de 100 µL foram estocadas a uma temperatura de -80°C.

4.10) Transformação de bactérias

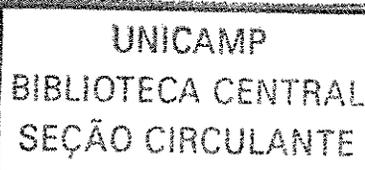
As bactérias competentes foram transformadas da seguinte maneira: 30 a 50 ng de vetor foi adicionado a 50 µL de células competentes e deixados em gelo por 30 min. Após este período, a amostra foi colocada à 42°C por 2 min e novamente em gelo por 5 min. Adicionou-se 1 mL de LB e a amostra foi incubada à 37°C por 50 min. Transcorrido este período, as células foram plaqueadas em meio TB sólido (1,5% agar) contendo antibiótico para seleção dos clones positivos.

4.11) Seleção de clones positivos mediante uso de antibióticos

Os antibióticos utilizados neste estudo para seleção dos clones positivos estão especificados na tabela 4.

Tabela 4: Antibióticos utilizados na seleção de clones positivos.

Ampicilina (100 µg/mL)	Kanamicina (50 µg/mL)	Ampicilina + Kanamicina	Cloranfenicol (34 µg/mL) + Kanamicina
Linhagens transformadas com pGEX4T-3	Linhagens transformadas com pET 28a(+) e pET 29a(+)	Linhagens C41 e C43 pAV10 transformadas com pET 28a(+) e pET 29a(+)	Linhagens BL21 (DE3): pLysS, pLysE, pTGROE, rosseta, pT-Trx, Ril e RP transformadas com pET 28a(+) e pET 29a(+)



4.12) *Extração de plasmídios em mini escala (miniprep)*

Esta extração foi baseada no método STET, fundamentado em lise alcalina (protocolo modificado de Sambrook e Russel, 2001), onde uma colônia isolada de bactéria transformada com vetores contendo o fragmento gênico, foi crescida durante a noite, sob agitação, à 37°C, em 5 mL de meio TB contendo antibiótico conforme a resistência de cada linhagem. Após este período, as células foram coletadas por meio de centrifugação (13000 g, 10 min, temperatura ambiente) e ressuspendidas em 300 µL de STET (8% sacarose, 1% tritron x-100, 50 mM EDTA, 50 mM Tris-HCl pH 8,0) e adicionados 20 µL de solução de lisozima (50 mM Tris-HCl pH 8,0, 10 mg/mL lisozima, 2 mg/mL RNase), sendo as amostras misturadas por inversão. Após 10 min de incubação à temperatura ambiente, as amostras foram incubadas a 100°C por 2 min e centrifugadas (13000 g, 10 min, temperatura ambiente). O sobrenadante foi transferido para outro tubo e para precipitação dos plasmídios, adicionou-se 325µL de solução de isopropanol (75% isopropanol, 1,25 M acetato de amônia), as amostras foram misturadas por inversão e centrifugadas (13000 g, 10 min, temperatura ambiente). O sobrenadante foi descartado e o precipitado seco em estufa (37°C), sendo posteriormente ressuspendido em 20 µL de 10 mM Tris-HCl pH 8,0.

4.13) *Extração de plasmídios em média escala (midiprep)*

A midiprep foi baseada em lise alcalina (protocolo modificado de Sambrook e Russel, 2001), onde uma colônia isolada de bactéria contendo o plasmídio foi crescida durante à noite, sob agitação, à 37°C, em 50 mL de meio TB contendo antibiótico. Após crescimento, as células foram coletadas por meio de centrifugação (4000 g, 10 min, 4°C) e o meio descartado. As células foram ressuspendidas em 5 mL de TEG: 50 mM glicose, 25 mM Tris-HCl pH 8,0, 10 mM EDTA pH 8,0, adicionou-se 4 mL de 0,2 N NaOH, 1% SDS e 3 mL de clorofórmio. As amostras foram misturadas por inversão e deixadas por 1min a temperatura ambiente. Depois, adicionou-se 3 mL de 3 M acetato potássio pH 5,5, as amostras foram misturadas por inversão e centrifugadas (4000 g, 5 min, 4°C). O sobrenadante foi transferido para outro tubo, adicionou-se etanol absoluto à -20°C (2 volumes) e as amostras foram incubadas à -20°C por 30 min a fim de precipitar os plasmídios. Após este período, as amostras foram centrifugadas (4000 g, 10

min, 4°C), o sobrenadante descartado e o precipitado foi seco em estufa (37°C). Depois foi ressuscitado em 400 µL de 10 mM Tris-HCl pH8,0, contendo 0,03 mg/mL RNase.

4.14) Eletroforese em gel de agarose

Os géis para análise de DNA foram de agarose 1% (peso/volume) em tampão TAE (40 mM Tris base, 5,7% ácido acético, 50 mM EDTA pH 8,0), com adição de 1,7 µg/mL brometo de etídeo, que permite a visualização das bandas de DNA sob luz ultravioleta.

4.15) Ensaio de expressão de proteína em mini escala

Uma colônia isolada das cepas para expressão, contendo o vetor com o fragmento gênico, foi crescida durante à noite, sob agitação, à 37°C, em 2 mL de meio TB líquido contendo antibiótico, conforme a cepa e o vetor utilizado. Após este período, 50 µL do pré-inóculo foram adicionados à 5 mL de meio TB líquido contendo antibiótico e as amostras foram incubadas em três diferentes temperaturas: 37°C, 30°C e 20°C até chegar à OD₆₀₀ entre 0.6 – 0.8. Alcançada esta densidade ótica, a expressão das proteínas em estudo foi induzida por meio de 300 mM NaCl para *E. coli* BL21 SI e 0,5 mM IPTG para as demais linhagens especificadas na tabela 3. A uma réplica de cada amostra induzida não se adicionou o indutor, como forma de controle da expressão. As duas culturas, induzida e não-induzida, foram incubadas por 4h nas temperaturas de 30°C e 37°C, e 16h para experimentos à 20°C. Posteriormente, as células foram ressuscitadas em H₂O Milli Q, lisadas por sonicação, centrifugadas (12000 g, 10 min, 4°C) e o precipitado (fração insolúvel) foi separado do sobrenadante (fração solúvel) para verificar a expressão e solubilidade das proteínas em questão.

4.16) Purificação de proteína codificada por Xf0389

Para obtenção da proteína pura, as células obtidas após a indução em larga escala (1 a 2 L) com 10 mM lactose, foram ressuscitadas em tampão A (50 mM fosfato de sódio pH 7,2, 100 mM NaCl, 5% glicerol, 5 mM imidazol) na proporção de 20 mL tampão A por litro de cultura induzida e as células foram lisadas por pressão (prensa francesa, 1250 psi). Após lise, adicionou-se à amostra

2% peso/volume de sulfato de estreptomicina, sendo a mistura incubada à 4°C por 30 min para precipitação do DNA. Passado este período, a amostra foi centrifugada (19000 g, 30 min, 4°C) e o sobrenadante contendo a proteína recombinante portadora de cauda de histidina foi coletado e utilizado nos ensaios de purificação, que foi feita por cromatografia de afinidade a metais, sendo níquel o metal utilizado (resina: Ni-NTA Super Flow, Qiagen e coluna: Hi-trap Chelating 1 mL, Amershan). A proteína foi eluída com quantidades crescentes de imidazol (tampão B: tampão A contendo 1 M imidazol).

Após purificação, a amostra foi dialisada em membrana de diálise (Spectra/Por, Spectrum) com poros de diâmetro que permitem a passagem de moléculas com uma massa molecular entre 12-14.000 Da, em tampão 10 mM Tris-HCl pH 7,5 contendo 100 mM NaCl, 20% glicerol e 10 mM β -mercaptoetanol, para remoção do imidazol e outros sais.

4.17) *Concentração da proteína*

Após diálise, a amostra foi concentrada por diálise inversa (PEG com peso molecular de 20.000 Da) ou por ultracentrifugação (filtro Centriprep YM-10 com celulose regenerada, que permite a passagem de moléculas com massa molecular até 10.000 Da, Millipore).

4.18) *Determinação da concentração de proteína*

A concentração de proteínas foi determinada espectroscopicamente com uso do equipamento DU640 Spectrophotometer (Beckman Coulter), em cubeta de quartzo de 1 cm de caminho ótico. A partir do coeficiente de extinção (ϵ), foi calculada a concentração de proteína final. A fórmula utilizada para o cálculo foi:

concentração da proteína em mg/mL = $A / \epsilon \times \text{diluição}$

Onde A = absorção à 280 nm obtida pelo espectroscópio.

$\epsilon_{280\text{nm}}$ teóricos para a proteína recombinante obtidos pelo programa ProtParam (Gill e von Hippel, 1989), acessível através da interface *web* localizada em <http://us.expasy.org/tools/protparam.html>:

$\epsilon_{280\text{nm}}$ (Xf0389) = $20.340 \text{ M}^{-1} \text{ cm}^{-1}$

4.19) *Clivagem da cauda de histidina da proteína*

À amostra de proteína adicionou-se trombina na proporção de 0,1, 0,2 e 1% peso/peso e as amostras foram incubadas à temperatura ambiente, durante à noite.

4.20) *Proteólise limitada*

A amostra de proteína foi submetida a proteólise por diversas proteases: tripsina, subtilisina, quimiotripsina e papaína na proporção de 1% peso/peso. Limitou-se o tempo de ação enzimática (5 min, 15 min, 30 min, 1 h, 2 h e 16 h) e as amostras foram incubadas a temperatura ambiente.

4.21) *Eletroforese em gel de poliacrilamida*

As amostras de proteína foram analisadas através de eletroforese em gel de poliacrilamida sob condições desnaturante segundo Laemmli (1970).

4.22) *Dicroísmo circular (CD)*

As medidas de CD foram realizadas em um espectropolarímetro J-810 (JASCO), cubeta de quartzo de 0.1 cm de caminho ótico e proteína diluída em tampão 10 mM Tris-HCl com pH 7,5 para uma concentração final de 0.4 mg/mL. Foram coletados espectros em UV próximo, em um intervalo de comprimento de onda entre 190 e 260 nm à 25°C, com uma velocidade de 20 nm/min, sendo feitas no total 5 leituras com resposta de 4 segundos em varredura contínua a partir dos quais uma média é tirada. Esta análise à 25°C foi feita com a finalidade de caracterização da estrutura secundária da proteína. Para o ensaio de estabilidade térmica, os espectros também foram coletados da mesma forma, porém com uma faixa de temperatura entre 20-90°C, sendo um espectro obtido a cada 5°C. A temperatura foi controlada através do controlador de temperatura Peltier Type Control System PFD425S (JASCO).

Para a amostra de proteína sem cauda de histidina, as medidas foram feitas da mesma maneira descrita anteriormente, e foram apenas coletados os dados para caracterização da estrutura secundária da proteína, onde a amostra apresentava uma concentração final de 0.1 mg/mL.

Subtraiu-se dos espectros de proteína o espectro do tampão utilizado como branco (1 mM Tris-HCl pH 7,5, 10 mM NaCl, 2% glicerol e 1 mM β -mercaptoetanol).

Os valores obtidos na leitura de CD (mdeg) foram convertidos para elipticidade molar residual ($[\theta]_{MRW}$) que é definida pela equação (Adler *et al.*, 1973):

$$[\theta]_{MRW} = \theta_{obs} \times MRW / 10 \times d \times C^n$$

Onde θ_{obs} é a elipsidade observada em graus, MRW é o peso molecular médio dos resíduos da proteína, d é o caminho óptico da cubeta em centímetros e C^n é a concentração da proteína em mg/mL.

O MRW é calculado pela seguinte fórmula:

$$MRW = \text{peso molecular da proteína} / \text{número de aminoácidos}$$

4.23) Modelagem molecular

Esta etapa do trabalho foi realizada sob orientação do Dr. Sérgio Oyama Junior, pós doutorando do Centro de Biologia Molecular e Estrutural do Laboratório Nacional de Luz Síncrotron.

4.23.1) Predição de estrutura secundária

A seqüência da proteína Xf0389 foi submetida a predição de estrutura secundária por meio do algoritmo PSIPRED (Jones, 1999; McGuffin *et al.*, 2000), acessível através da interface *web* localizada em <http://bioinf.cs.ucl.ac.uk/psiform.html>. Com base em técnicas computacionais de redes neurais e utilizando um banco de dados contendo informações estruturais e de seqüências, o algoritmo é capaz de prever com grande precisão a localização dos elementos de estruturas secundárias presentes nas seqüências submetidas.

4.23.2) Busca de estruturas conhecidas

Utilizou-se o programa BLASTp (Altschul *et al.*, 1990) na busca de seqüências similares a Xf0389, restringindo a busca por seqüências cujas estruturas tridimensionais já houvessem sido determinadas experimentalmente, assim limitando-se ao banco de dados PDB.

4.23.3) Alinhamento de seqüências

O alinhamento preliminar obtido pelo programa *Multiple sequence alignment* (Corpet, 1988) foi posteriormente aprimorado manualmente, levando-se em consideração as informações obtidas através da predição de estrutura secundária. Deste modo, adicionalmente ao alinhamento de seqüência foi realizado um alinhamento estrutural com as possíveis proteínas a serem utilizadas como molde para a construção do modelo.

4.23.4) Construção do modelo por homologia

Com base no alinhamento final, as coordenadas tridimensionais da estrutura-molde foram utilizadas como referência para a seqüência-alvo. Para tanto, utilizou-se o programa MODELLER (Sali *et al.*, 1993), que utiliza uma lista de restrições espaciais geradas a partir da estrutura-molde. Esta lista de restrições é aplicada durante os procedimentos de construção e refinamento do modelo, sendo uma estrutura gerada a cada etapa de *simulated annealing*. Esta técnica simula em computador o aquecimento até 1500K, de uma estrutura preliminar completamente estendida. Durante esta etapa, o aumento da temperatura resulta em um aumento da energia cinética dos átomos do sistema e desta forma, tem-se uma boa amostragem do espaço conformacional disponível para esta estrutura. Posteriormente, segue-se um resfriamento gradual até 300K. Concomitantemente, as restrições de distância derivadas da estrutura-molde são aplicadas, gerando a cada ciclo um novo modelo. Como foram realizados cinco ciclos, resultaram cinco modelos. Os termos de energia utilizados nos métodos de dinâmica molecular (*simulated annealing*) e minimização de energia (refinamento) baseiam-se no campo de força CHARMM (Brooks *et al.*, 1983).

4.23.5) Avaliação do modelo

As estruturas finais obtidas após a etapa de minimização de energia foram posteriormente avaliadas quanto à sua qualidade estereoquímica através do programa PROCHECK (Laskowski *et al.*, 1993).

4.23.6) Recursos computacionais

As etapas de construção do modelo por homologia, refinamento e avaliação do modelo foram realizadas em um multi-processador Compaq AlphaServer ES40. Os procedimentos de visualização e manipulação das estruturas foram realizados através da interface gráfica gerada pelo programa InsightII (Accelrys Inc.), instalado em uma estação gráfica Octane2 (Silicon Graphics Inc.). As imagens finais foram obtidas pelo programa WebLab Viewer (Accelrys Inc.), em ambiente Windows.

5) Resultados

5.1) Predição de segmentos transmembrana da proteína sensora

Como a proteína sensora possui regiões insolúveis, uma vez que está inserida na membrana bacteriana, a predição de fragmentos transmembrana foi feita através do programa DAS (Cserzo *et al.*, 1997). Obtiveram-se dois segmentos transmembrânicos, que correspondem aos resíduos de aminoácidos 1 a 50 e 155 a 231. A proteína também possui duas regiões extra-membranas, sendo a região I correspondente aos resíduos 51 à 154 que se encontra no periplasma bacteriano, enquanto a região II, se encontra no citoplasma e corresponde aos resíduos 232 à 482 (figura 3). Assim, por meio deste programa, delimitou-se as regiões I e II para clonagem da proteína sensora.

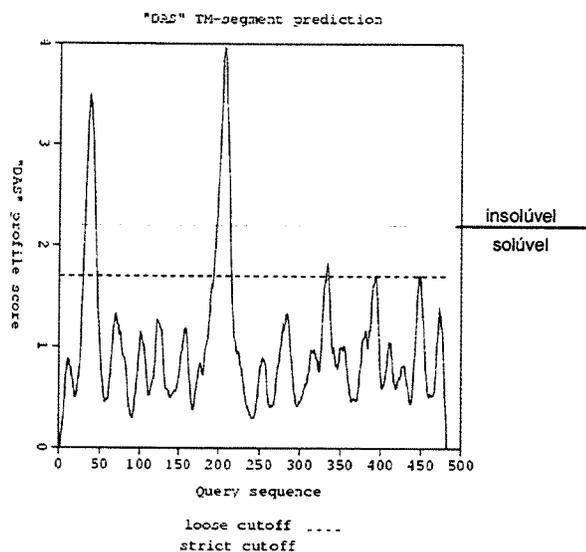


Figura 3 - Predição de segmentos transmembrana da proteína Xf0390. Regiões acima do limite indicado, correspondem a prováveis segmentos transmembrânicos. O gráfico mostra a hidrofobicidade em função da sequência de aminoácidos (Cserzo *et al.*, 1997).

5.2) Amplificação e clonagem dos genes Xf0389 e Xf0390 de *Xylella fastidiosa*

Os fragmentos amplificados por PCR correspondentes aos genes Xf0389 (que codifica para a proteína reguladora) e Xf0390 (que codifica para a proteína sensora) foram analisados através de eletroforese em gel de agarose. Os fragmentos advindos da reação de PCR que correspondem aos genes Xf0389 e região II de Xf0390 (figura 4) foram purificados e clonados em vetor pGEM-T-Easy, enquanto a região I de Xf0390 não foi clonada corretamente, pois observa-

se uma banda que apresenta aproximadamente 500pb a mais do que o esperado (310pb; figura 5). Isso pode ter ocorrido pelo fato de ter sido utilizado como molde para a reação DNA genômico e o oligonucleotídeo ter anelado em região não específica. Assim, seguiram-se experimentos com os plasmídios contendo o fragmento que codifica para a região II da proteína sensora e outros plasmídios contendo o fragmento gênico que codifica para a proteína reguladora.

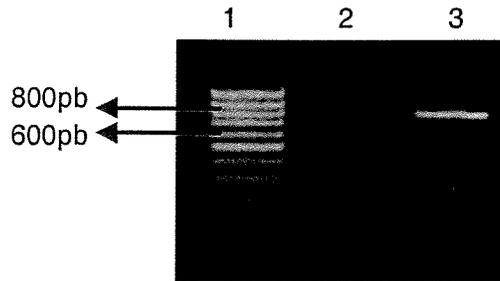


Figura 4 – Produto de PCR obtido a partir de DNA genômico de *X. fastidiosa* e oligonucleotídeos específicos. Coluna 1 - marcador de peso molecular; coluna 2 - Xf0389; coluna 3 - Xf0390 (região II). Gel de agarose a uma concentração de 1%.

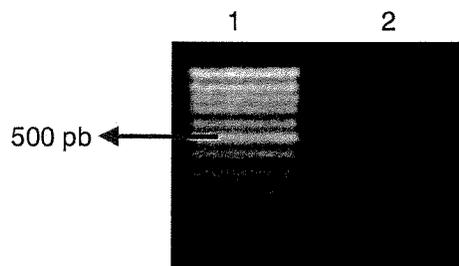


Figura 5 – Produto de PCR obtido a partir de DNA genômico de *X. fastidiosa* e oligonucleotídeos específicos para a região I de Xf0390. Coluna 1 - marcador de peso molecular, coluna 2 – Xf0390. Gel de agarose a uma concentração de 1%.

5.3) Subclonagem em vetores de expressão

Após sequenciamento e confirmação das seqüências gênicas, os vetores de clonagem foram submetidos a reações de digestão de acordo com o sítio de restrição presente em cada inserto e os fragmentos correspondentes aos genes foram extraídos do gel (figura 6).

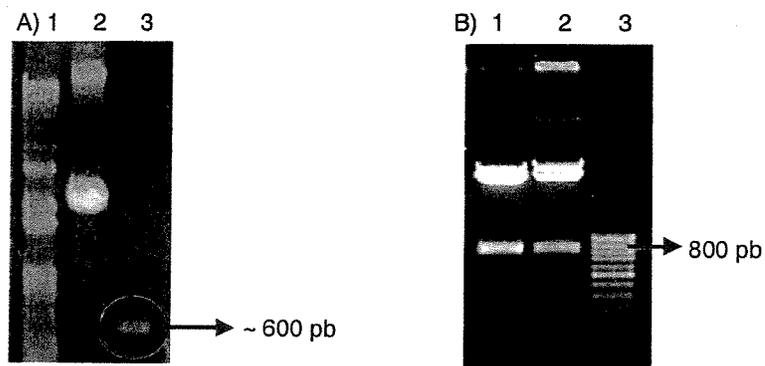


Figura 6 – Produtos das reações de digestão dos vetores de clonagem. **A) Xf0389**. Coluna 1- marcador de peso molecular, coluna 2 - pGEM-T-Easy, coluna 3 – fragmento flanqueado pelas enzimas NdeI/XhoI, purificado do gel. **B) Xf0390 (região II)**. Colunas 1 e 2- fragmento flanqueado pelas enzimas NdeI/XhoI, coluna 3- marcador de peso molecular. Gel de agarose a uma concentração de 1%.

A partir dos produtos da digestão do vetor de clonagem purificados, correspondentes aos insertos de aproximadamente 600pb e 800pb (Xf0389 e região II de Xf0390, respectivamente), foi tentada a subclonagem em vetores de expressão pET28a(+), pET29a(+) e pGEX4T-3, e cepas DH5 α de *E. coli* foram transformadas. Não foi possível subclonar a região II de Xf0390, talvez devido ao fato da proteína, provavelmente, ser tóxica para bactéria, ou ainda, o sítio de restrição do fragmento pode não ter sido amplificado corretamente, pois como visto após seleção por PCR de colônia ou reação de digestão, o fragmento que corresponde ao gene Xf0389 foi subclonado com êxito nos vetores descritos acima (figura 7), demonstrando que não havia problema com os plasmídios e estes clones positivos foram utilizados na transformação das linhagens bacterianas para expressão e as colônias selecionadas por PCR de colônia (figura 8). Assim, encerrando-se aqui os trabalhos experimentais com a proteína sensora (Xf0390).

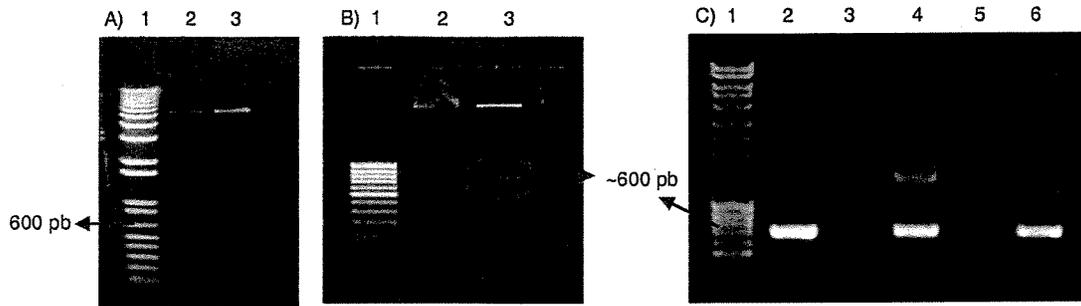


Figura 7: Seleção de clones positivos por meio de reações de digestão dos vetores de expressão: **A) Xf0389/pET29a(+)**. Coluna 1- marcador de peso molecular, colunas 2 e 3 - fragmento advindo da reação de digestão feita pelas enzimas NdeI/XhoI. **B) Xf0389/pGEX4T-3**. Coluna 1- marcador de peso molecular, colunas 2 e 3 - fragmento advindo da reação de digestão feita pelas enzimas NdeI/XhoI. **Seleção de clones positivos por meio de PCR de colônia: C) Xf0390/pET28a(+)**. Coluna 1- marcador de peso molecular, coluna 2- controle positivo (pGEM-T-Easy), coluna 3- controle negativo, colunas 4 e 6- produto de PCR correspondente ao gene em questão, coluna 5- pET28a(+) sem inserto (controle negativo). Gel de agarose a uma concentração de 1%.

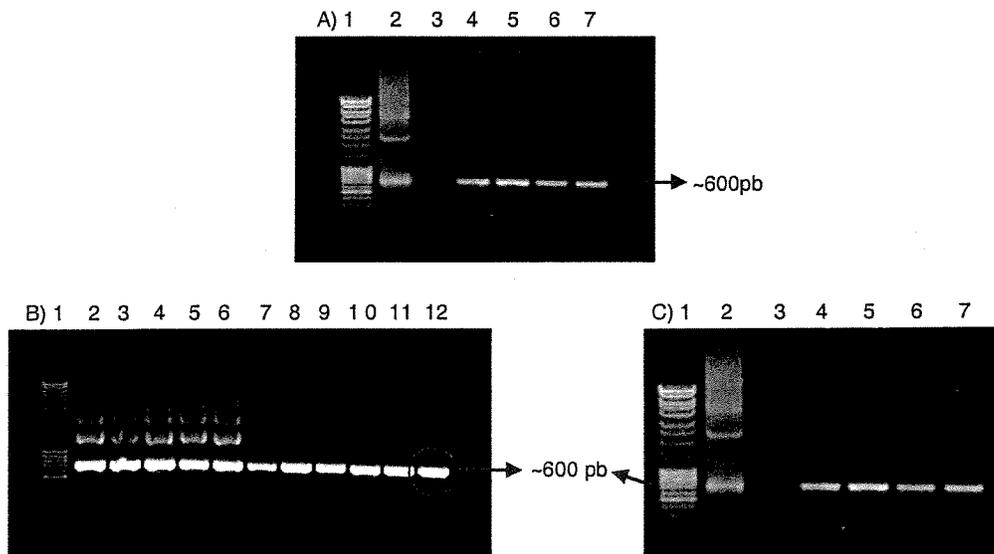


Figura 8: Seleção de clones positivos em diversas linhagens de *E. coli*, por meio de PCR de colônia. **A) Xf0389/pET29a(+)**. Coluna 1- marcador de peso molecular, coluna 2- controle positivo (pGEM-T-Easy com inserto), coluna 3- controle negativo: reação sem DNA, coluna 4- C41, coluna 5- rosseta (DE3), coluna 6- BL21 (DE3) pT-Trx e coluna 7- BL21 (DE3) pTGROE. **B) Xf0389/pET28a(+)**. Coluna 1- marcador de peso molecular; coluna 2- C41; coluna 3- BL21 (SI); coluna 4- C43; coluna 5- BL21 (DE3) pT-Trx; coluna 6- BL21 (DE3) Ril; coluna 7- BL21 (DE3) pLysE; coluna 8- BL21 (DE3); coluna 9- rosseta (DE3); coluna 10- BL21 (DE3) RP; coluna 11- BL21 (DE3) pLysS e coluna 12- BL21 (DE3) pTGROE. **C) Xf0389/pGEX4T-3**. Coluna 1- marcador de peso molecular, coluna 2- controle positivo (pGEM-T-Easy), coluna 3- controle negativo, coluna 4- BL21 (DE3) pLysS, coluna 5- BL21 (DE3) pLysE, coluna 6- BL21 (DE3) pTGROE e coluna 7- rosseta (DE3). Gel de agarose a uma concentração de 1%.

5.4) Ensaios de expressão

Tentou-se a expressão da proteína Xf0389 na sua forma solúvel *in vivo* seguindo diversas metodologias (tabela 5):

- 1) alterando a temperatura de indução,
- 2) alterando as linhagens de *E. coli*,
- 3) alterando os tampões de lise celular,
- 4) subclonagem em vetores alternativos.

A proteína foi expressa parcialmente na fração solúvel somente com o gene clonado em pET28a(+), transformado em linhagem de *E. coli* C43 (DE3) (figura 9).

	pGEX4T-3	pET29a(+)	pET28a(+)
BL21 SI	○	○	○
BL21 (DE3)	○	○	○
BL21 (DE3) pLysS	○ ○	○	○
BL21 (DE3) pLysE	○ ○ ○	○ ○	○
BL21 (DE3) pTGROE	○ ○ ○	○ ○ ○	○
Rosseta (DE3)	○ ○ ○	○ ○ ○	○
BL21 (DE3) pT-Trx	○ ○	○ ○ ○	○
C41 (DE3)			○
C43 (DE3)			○
C41 (DE3) pAVD10		○ ○	
C43 (DE3) pAVD10		○ ○	
BL21 Star (DE3)			○
BL21 Codon plus (DE3) Ril			○
BL21 Codon plus (DE3) RP			○

Tabela 5: Resumo das alternativas tentadas para obtenção da proteína Xf0389 em sua forma solúvel (Markides, 1996). A linhagens bacterianas e vetores utilizados estão descritos na primeira coluna e na primeira linha, respectivamente. Os ○ representam as linhagens em que a proteína não foi expressa; ○ a proteína foi expressa na fração insolúvel quando induzida à 37, 30 e 20°C, ressuspensando as células em H₂O; ○ a proteína foi expressa na fração insolúvel quando induzida à 37°C, ressuspensando as células em H₂O; ○ a proteína foi expressa na fração insolúvel quando induzida à 37°C, ressuspensando as células em tampão 0.1 M Tris-HCl contendo 1% triton x-100; ○ a proteína foi expressa na fração insolúvel quando induzida à 37°C, ressuspensando as células em 0.1 M de acetato de sódio pH 5.0 e ○ proteína expressa na fração solúvel do lisado celular quando induzida à 37 e 30°C.

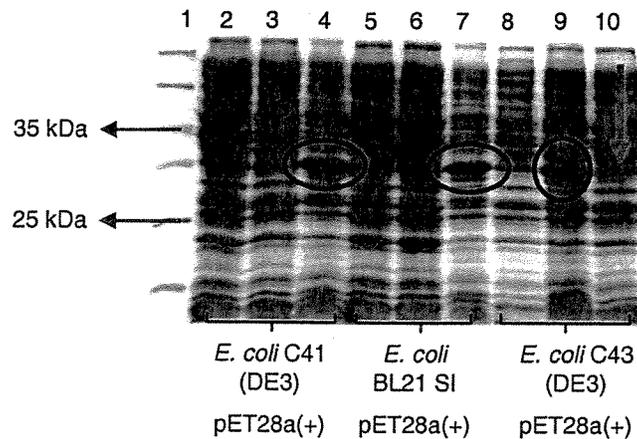


Figura 9: Ensaio de expressão da proteína codificada por Xf0389, a 37°C. Coluna 1- marcador de peso molecular, colunas 2, 5 e 8- sobrenadante da amostra não induzida, colunas 3, 6 e 10- sobrenadante da amostra induzida, colunas 4, 7 e 9- precipitado da amostra induzida. As linhagens e vetor utilizado estão especificados no gel. Com o círculo é mostrada a proteína na fração insolúvel e com a seta, a proteína na fração solúvel. SDS-PAGE a uma concentração de 12,5%.

5.5) Purificação de proteínas

Após indução em larga escala e lise celular, o extrato protéico contendo a proteína recombinante com cauda de histidina foi submetido a cromatografia de afinidade a metal, em coluna contendo níquel, pré-equilibrada com tampão A. Como resultado, a proteína começou a ser eluída a partir de 10 mM de imidazol, sendo desprendida do Ni^{2+} em maior quantidade nas concentrações de 50 e 100mM de imidazol, porém ainda impura mesmo com as lavagens iniciais. Também foi visualizada uma possível banda de degradação protéica, assim, foi adicionado 2 mM PMSF. E como foi obtido um nível significativo de impurezas e degradação residual, procedimentos visando otimizar a purificação foram conduzidos em um equipamento de FPLC (Pharmacia) (figura 10).

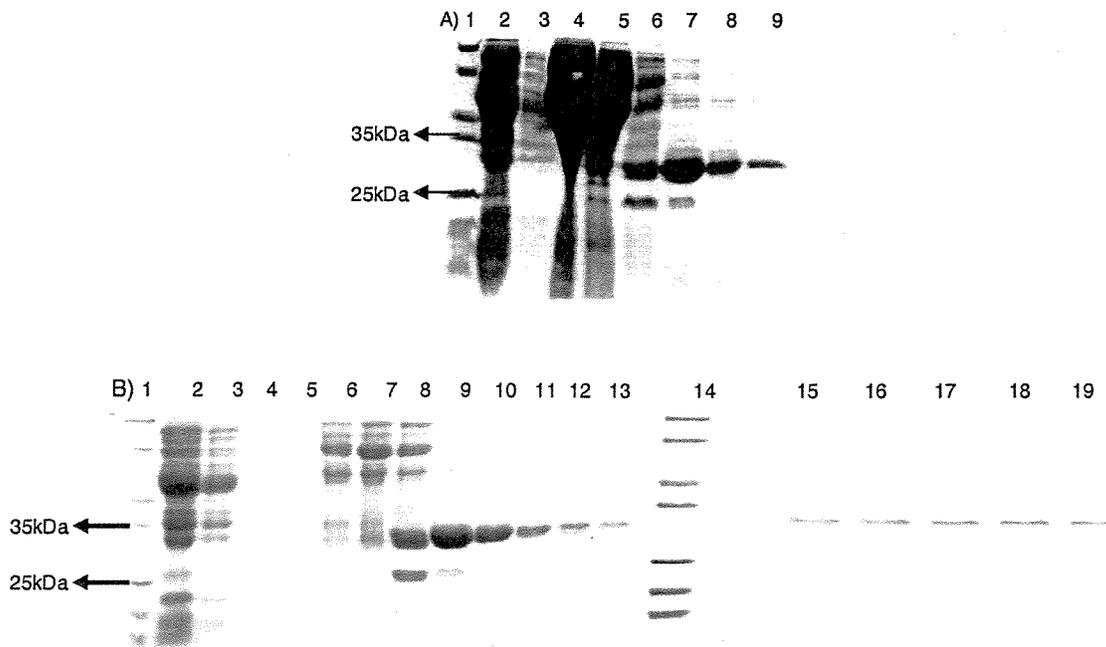


Figura 10: Purificação da proteína codificada por Xf0389 por cromatografia de afinidade a Ni^{2+} , feita pelo equipamento de FPLC . A) Indução da proteína à 37°C . Coluna 1- marcador de peso molecular; coluna 2- extrato não purificado; colunas 3 a 5- lavagens; colunas 6 a 9- frações correspondentes ao pico protéico obtido por meio do espectro fornecido pelo FPLC. B) Indução da proteína à 30°C . Colunas 1 e 14- marcador de peso molecular; coluna 2- extrato não purificado; colunas 3 a 8- lavagens; colunas 9 a 13 e 15 a 19- frações correspondentes ao pico protéico obtido por meio do espectro fornecido pelo FPLC. SDS-PAGE a uma concentração de 12,5%.

As frações correspondentes à proteína parcialmente pura foram reunidas e submetidas à proteólise com trombina devido ao sítio de clivagem existente entre a proteína recombinante e a cauda de histidina fornecida pelo vetor de expressão em questão. Foram testadas várias concentrações diferentes de protease/proteína e o resultado (figura 11) foi obtido satisfatoriamente apenas na proporção de 1% peso/peso (coluna 6).

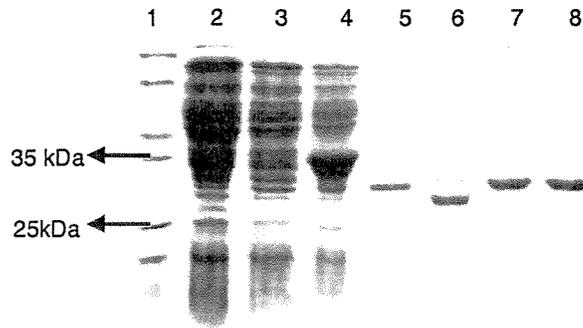


Figura 11: Clivagem da cauda de histidina dada pelo vetor pET28a(+), com trombina. Coluna 1- marcador de peso molecular; coluna 2- sobrenadante da amostra não induzida; coluna 3- sobrenadante da amostra induzida; coluna 4- precipitado da amostra induzida; coluna 5- proteína purificada como controle da proteólise; coluna 6- proteólise da proteína com trombina na proporção de 1% peso/peso; coluna 7- proporção de 0,2%, coluna 8- proporção de 0,1%. SDS-PAGE a uma concentração de 15%.

5.6) Caracterização estrutural

Tanto a amostra de proteína de Xf0389 contendo a cauda de histidina quanto a amostra da proteína sem a cauda de histidina foram submetidas a ensaios de dicroísmo circular que foram feitos com o objetivo de observar o conteúdo de estrutura secundária da proteína alvo (figura 12). E como observado nos espectros, a diferença entre suas estrutura não é significativa, sendo que ambas apresentam um alto conteúdo de α -hélice.

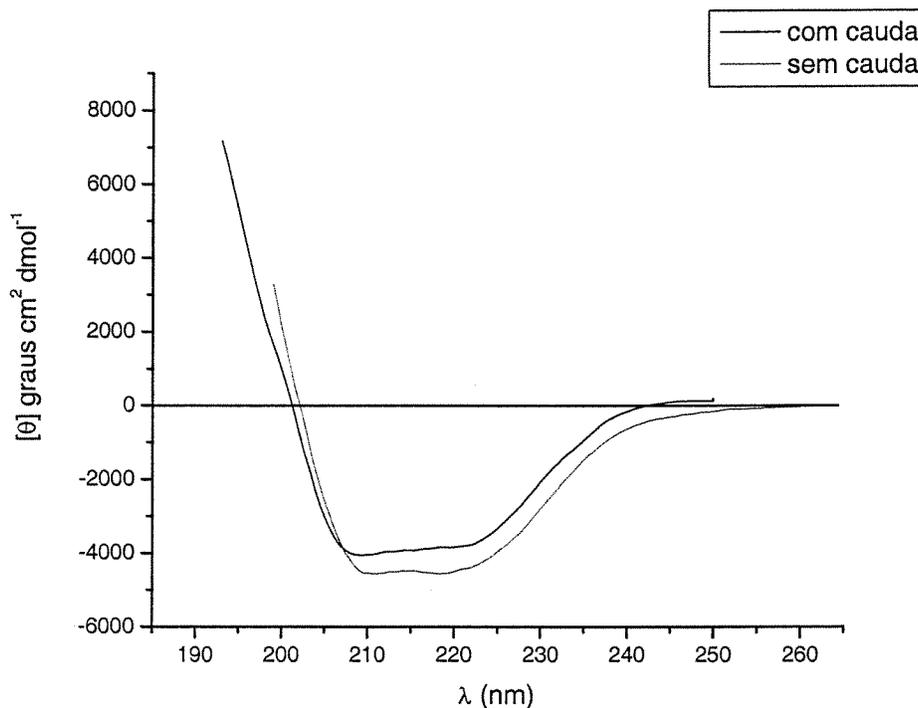


Figura 12: Espectro de CD da proteína reguladora codificada pelo gene Xf0389.

5.7) Estabilidade térmica

A fim de caracterizar a proteína por sua estabilidade térmica, foi realizado um ensaio de CD em uma faixa de temperatura entre 20 e 90°C (figura 13). E aparentemente é possível dizer que a proteína seja estável termodinamicamente, uma vez que ela começa a perder sua estrutura secundária entre 60 e 65°C, porém não ocorreu um total desnovelamento protéico.

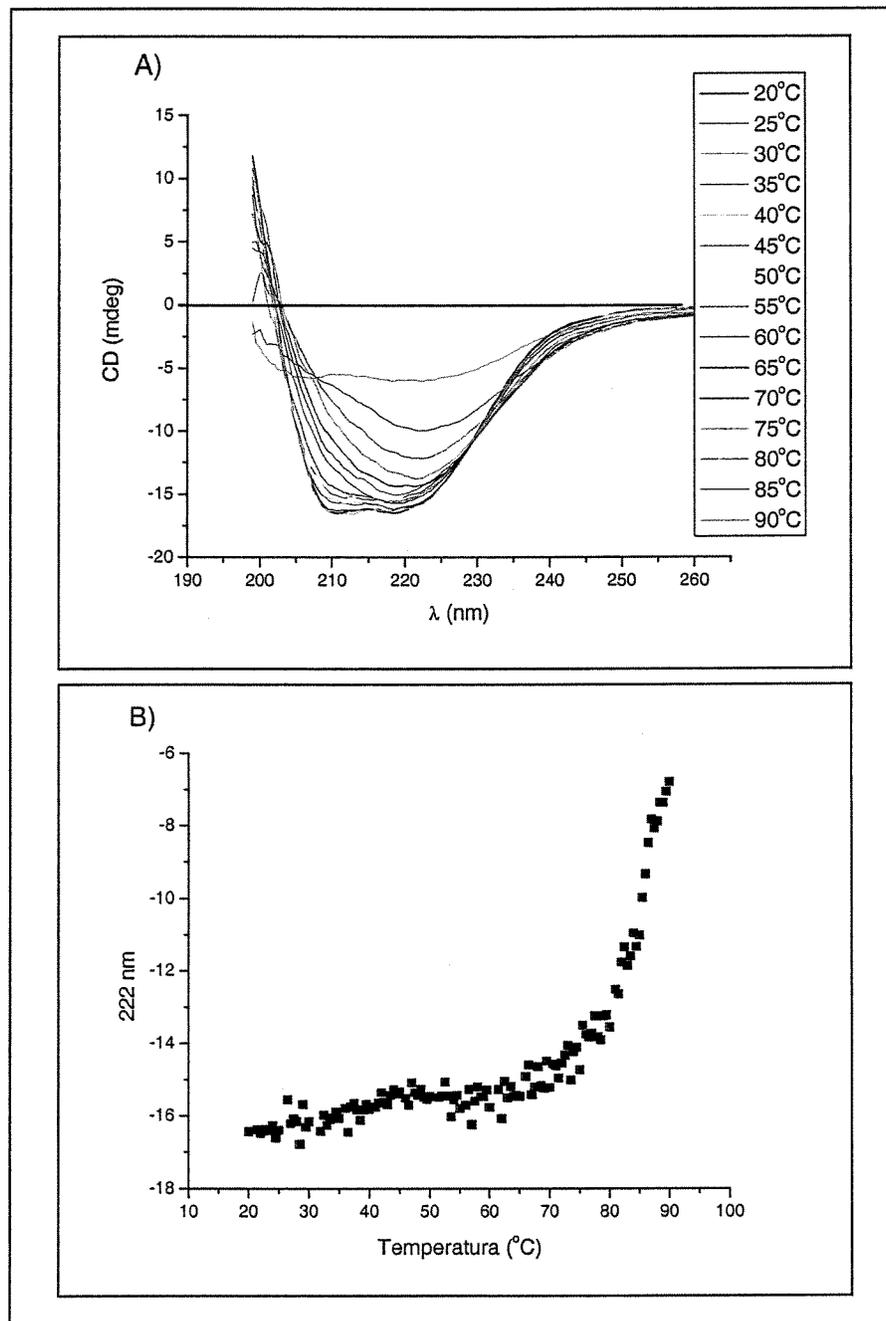


Figura 13: Ensaio de estabilidade térmica da proteína Xf0389. A) Espectro de CD com faixa de temperatura entre 20 e 90°C, em diferentes comprimentos de onda. B) Espectro de agregação protéica em função da temperatura. Comprimento de onda e temperatura estão especificados na figura.

Considerando os elementos de estrutura secundária preditos para a proteína alvo, as proteínas DrrD e Xf0389 foram alinhadas utilizando o programa Multiple sequence alignment (Corpet, 1988) e posteriormente foi feito um alinhamento de forma manual (figura 15), onde levou-se em conta tanto a seqüência quanto a estrutura secundária das proteína-molde e proteína-alvo, conseguindo um aumento de identidade para 44,5%. Assim, após definida a estrutura-molde, foram calculados cinco modelos teóricos, pelo programa MODELLER (Sali *et al.*, 1993), que posteriormente foram submetidos a uma avaliação da sua qualidade estereoquímica, permitindo a seleção de um modelo. Esta análise estereoquímica foi feita pelo programa PROCHECK (Laskowski et al., 1993) e os resultados são apresentados nas figuras 16 e 17.

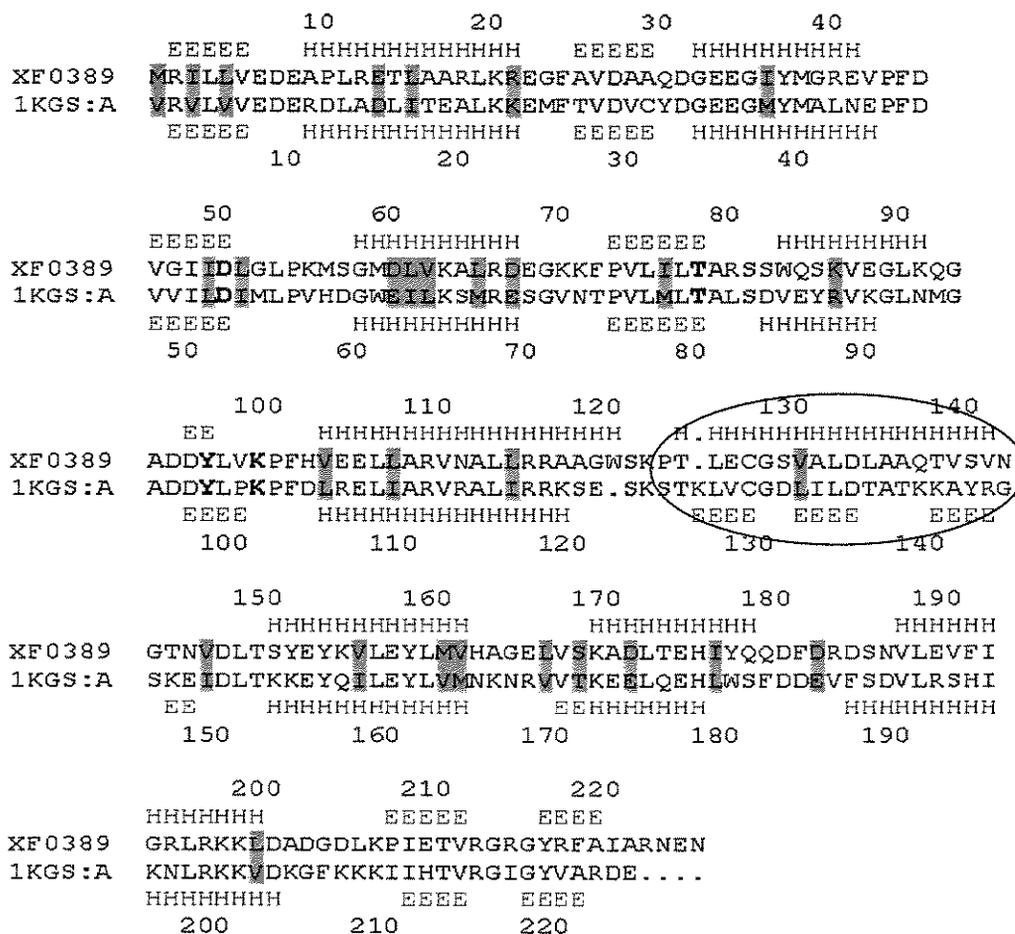


Figura 15: Alinhamento seqüencial e estrutural das proteínas Xf0389 de *X. fastidiosa* e DrrD de *T. maritima*, feito manualmente, depois de alinhamento feito pelo programa *Multiple sequence alignment* (Corpet, 1988), apresentando 44.5% de identidade. Marcados em amarelo estão os aminoácidos idênticos e em azul, os similares, que apresentam características químicas conservadas (13.2%). As letras acima da seqüência de Xf0389 indicam a predição da estrutura secundária (PSIPRED) e abaixo de 1KGS, estrutura determinada por difração de raios X de DrrD (Buckler et al., 2002), onde H indica α -hélice e E, fita- β . Em negrito estão os resíduos importantes para a conformação da proteína. O círculo vermelho mostra a diferença entre a estrutura secundária da proteína-molde (DrrD) em relação a predição da estrutura secundária da proteína-alvo (Xf0389).

Dentre os modelos gerados, foi escolhido aquele que, por sua qualidade estereoquímica, apresentou 99,5% de seus resíduos distribuídos entre as regiões mais favoráveis, possuindo 95% dos resíduos de aminoácidos, e permitida com 3%. Nas regiões generosamente permitidas encontra-se 1,5% de resíduos e apenas 0,5% localizou-se em região não permitida, que representa apenas um resíduo de lisina (K56, figura 17).

Também foram levados em consideração os contatos impróprios entre os átomos e se o comprimento das ligações, os ângulos das ligações e os grupos planares estavam dentro dos limites permitidos pela estrutura-molde e foi visto que o modelo (figura 18) apresenta uma boa qualidade uma vez que possui apenas 0,4% de contato impróprio entre os átomos, não contando os de hidrogênio e 99,2% do comprimento das ligações, 94% dos ângulos entre as ligações atômicas e 100% dos grupos planares se encontram dentro dos limites permitidos pelas coordenadas dadas pela proteína-molde. A superfície eletrostática da proteína é apresentada na figura 19.

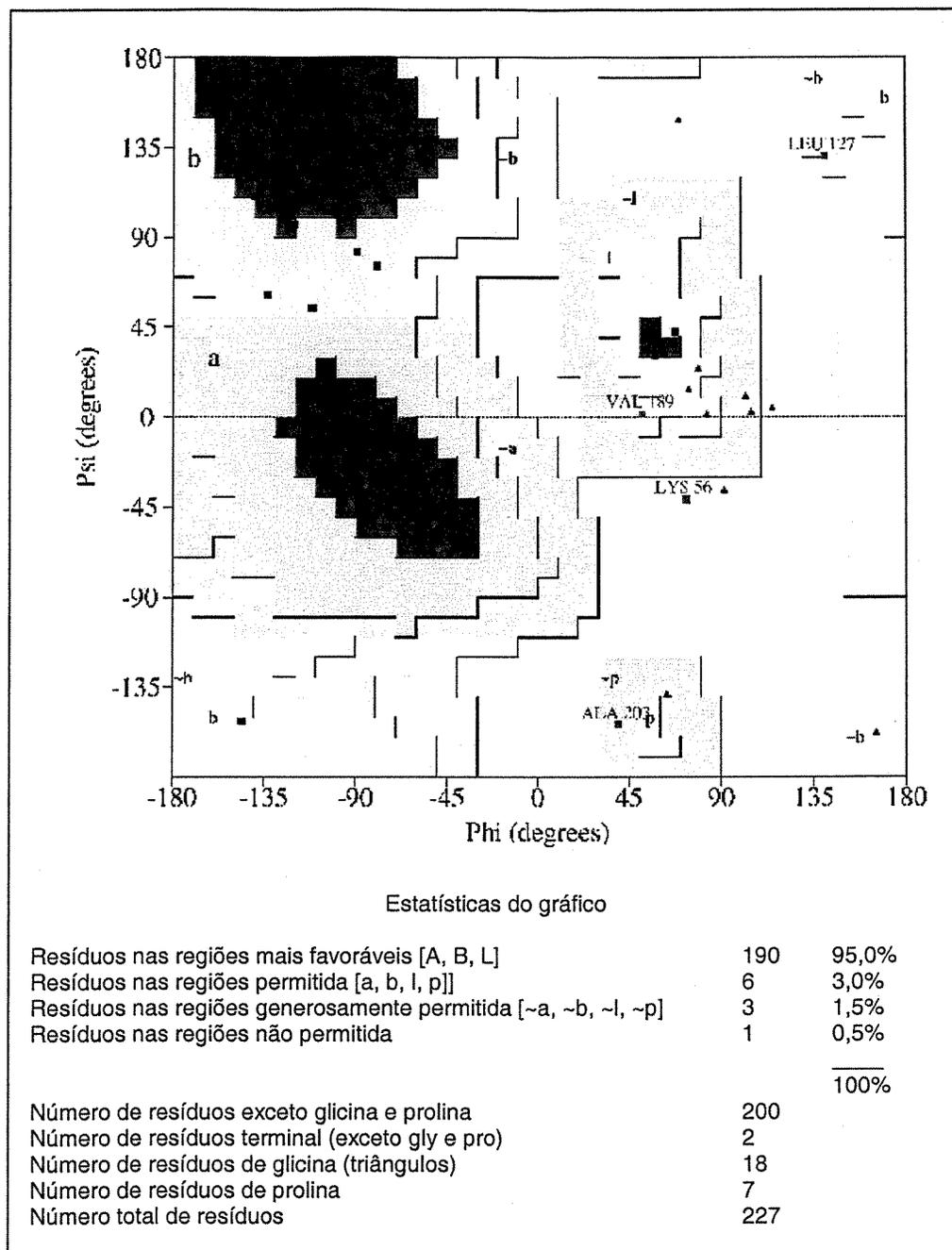


Figura 16: Mapa de Ramachandran mostrando a avaliação da qualidade estereoquímica do modelo de Xf0389, feita pelo programa PROCHEK (Laskowski et al., 1993). Os resíduos especificados se encontram em regiões generosamente permitidas ou não permitidas, equivalendo a apenas 2% do total de resíduos existentes. A análise estatística foi baseada em 118 estruturas com resolução de no mínimo 2 Å e fator R de no máximo 20%, sendo esperada boa qualidade do modelo se o mesmo apresentar mais de 90% dos resíduos em regiões mais favoráveis.

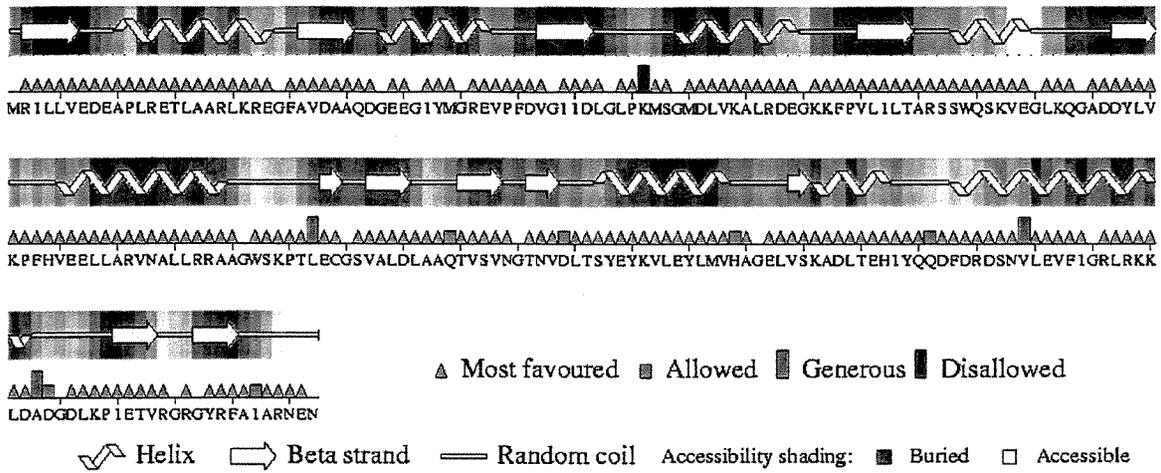


Figura 17: Detalhamento do mapa de Ramachandran por resíduo de aminoácidos, mostrando também a descrição da estrutura secundária do modelo e acessibilidade das regiões ao solvente.

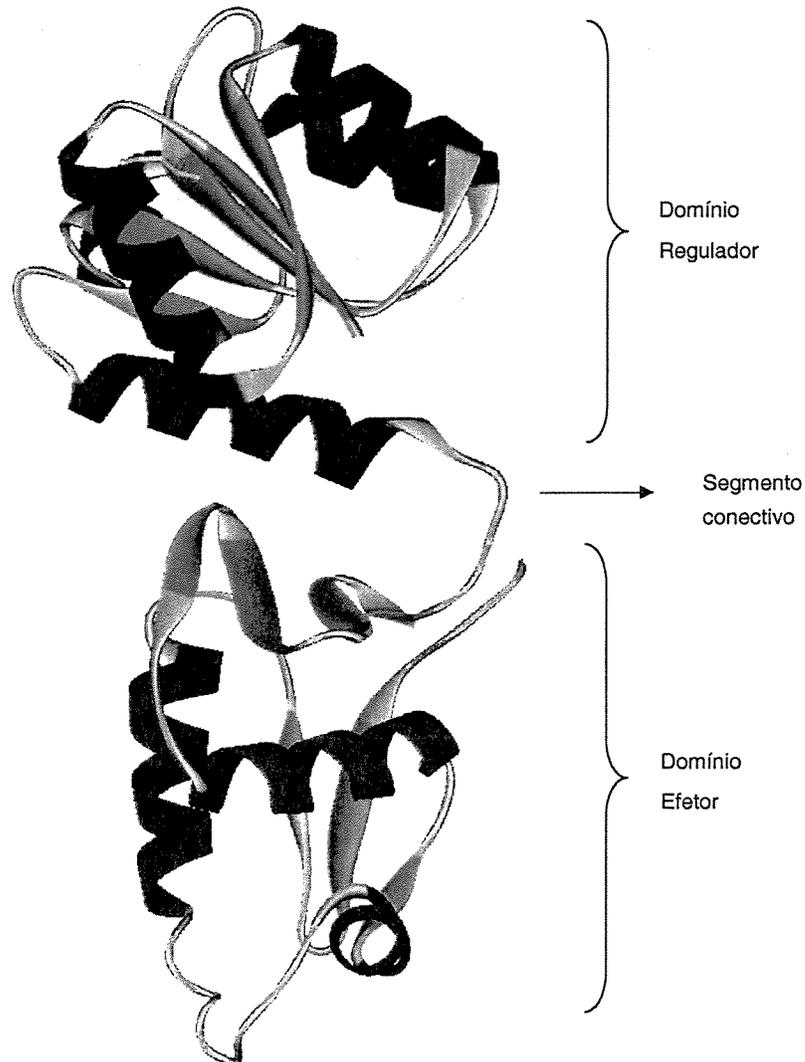


Figura 18: Modelo tridimensional da proteína Xf0389. Os domínios regulador e efetor estão indicados na figura e em amarelo é mostrado o aspartato (51) conservado nas proteínas reguladoras pertencentes ao sistema regulador de dois componentes.

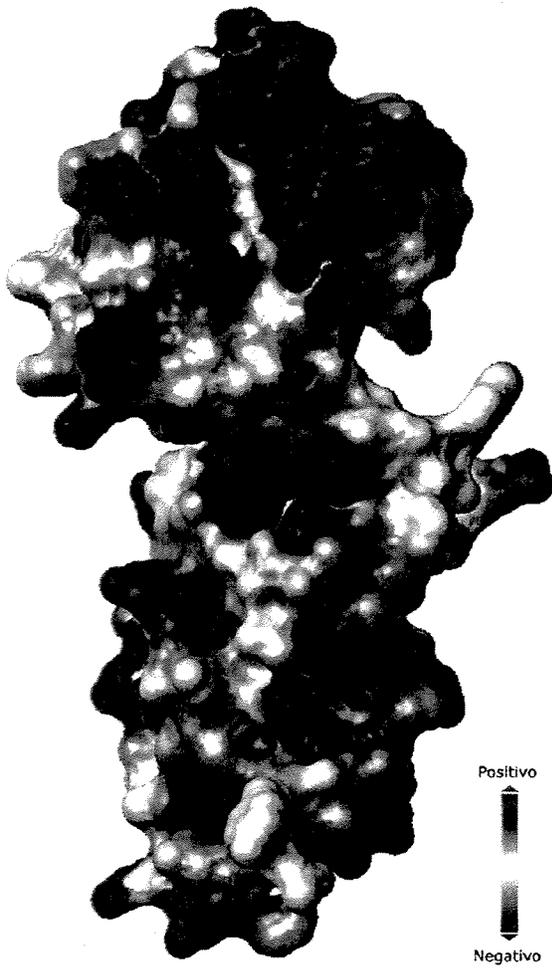


Figura 19: Representação da superfície eletrostática da proteína reguladora, gerada pelo programa MOLMOL (Koradi *et al.*, 1996). Em azul são mostradas as regiões que apresentam carga positiva, em branco as regiões eletrostaticamente neutra e em vermelho, carga negativa.

5.9) Proteólise limitada

Ainda caracterizando a estrutura da proteína Xf0389 e objetivando analisar suas seqüências mais expostas, a amostra foi submetida à proteólise limitada (figura 20) por diversas proteases, adicionadas separadamente na proporção de 1% peso/peso, sendo elas: tripsina, subtilisina, quimiotripsina e papaína.

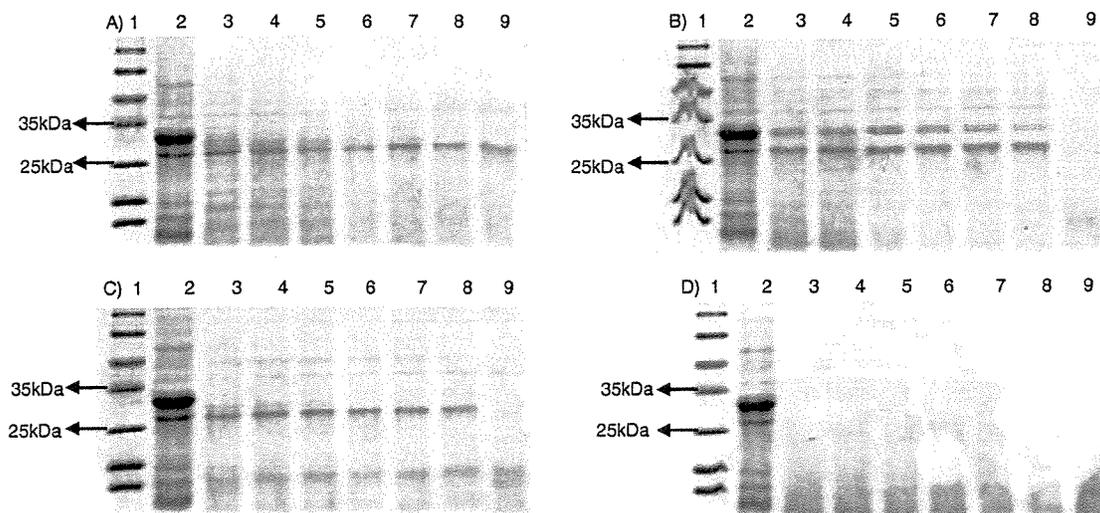


Figura 20: Proteólise limitada, incubando a amostra à temperatura ambiente. Coluna 1- marcador de peso molecular, coluna 2- controle: proteína sem adição de protease, coluna 3- 5 min de reação, coluna 4- 15 min, coluna 5- 30 min, coluna 6- 1 h, coluna 7- 2 h, coluna 8- 4 h, coluna 9- 16 h. A) Tripsina. B) Subtilisina. C) Quimiotripsina. D) Papaína. SDS-PAGE a uma concentração de 15%.

6) Discussão

6.1) Amplificação e clonagem dos genes Xf0389 e Xf0390

Como foi observado na figura 4, os genes Xf0389 e Xf0390 (região II), foram amplificados nos tamanhos correspondentes ao esperado, sendo que o gene Xf0389 possui 684 pares de bases e Xf0390 (região II) 800 pb, como pode ser visualizado pelas bandas no gel de agarose. Por outro lado, a região I (periplasmática) definida para a proteína Xf0390 não teve seu fragmento gênico amplificado no seu tamanho correspondente, uma vez que ela deveria apresentar um tamanho de aproximadamente 310 pb e que não foi evidenciado pela reação de PCR (figura 5), pois os óligos devem ter anelado inespecificamente, pois foi utilizado DNA genômico bacteriano como molde da reação. Porém, esta região não vem a ser relevante ao presente estudo, já que a região II da proteína, correspondente ao domínio citoplasmático, possui a histidina que é fosforilada e responsável pela interação com a proteína reguladora (Xf0389). Assim, devido a este problema na clonagem, a região I da proteína sensora não foi mais trabalhada experimentalmente, equanto seguiu-se experimentos com os fragmentos amplificados de Xf0389 e da região II de Xf0390.

6.2) Ensaio de expressão

Para o teste de expressão das proteínas Xf0389 e região II de Xf0390, foram utilizados inicialmente os vetores:

- ✓ pET29a(+), que devido aos sítios de restrições inseridos nos fragmentos gênicos, não fornece a cauda de histidina que pode interferir na solubilidade da proteína,
- ✓ pGEX4T-3, que promove a fusão da proteína em estudo à GST de *Schistosoma japonicum*, que pode evitar a formação de corpos de inclusão e facilitar a purificação da proteína (Hannig e Makrides, 1998).

Novamente houve dificuldades em relação aos experimentos com Xf0390, pois, desta vez, não foi possível subclonar a região II em nenhum dos vetores descrito a cima, talvez por problemas no sítio de restrição, assim não se seguiram experimentos com tal região da proteína sensora e somente para a proteína reguladora (Xf0389) foi possível levar os objetivos adiante, como se segue.

Ambos vetores, contendo o fragmento gênico Xf0389, foram inicialmente utilizados na transformação de *E. coli* BL21 SI e *E. coli* BL21 (DE3). Em *E. coli* BL21 (SI) não houve expressão; porém, em *E. coli* BL21 (DE3) tanto a proteína de aproximadamente 25kDa (equivalente à Xf0389) quanto a de aproximadamente 55kDa (30 kDa da GST + 25 kDa da Xf0389) foram expressas na fração insolúvel do lisado celular.

Assim, na tentativa de tornar a proteína solúvel *in vivo* (revisado por Markides, 1996) foram feitos ensaios de expressão protéica alterando diversas condições, como:

- ✓ Temperatura: 37, 30 e 20°C, pois a redução da temperatura diminui a velocidade de expressão da proteína, podendo torná-la solúvel;
- ✓ Linhagens: transformaram-se diferentes linhagens;
- ✓ pH: alterou-se o pH do tampão de lise celular de duas formas:
 - (1) para pH 5,0, sendo as células quebradas por sonicação em 0,1 M de acetato de sódio, ao invés de H₂O Milli Q.
 - (2) as células foram ressuspensas em tampão 0,1 M Tris-HCl pH 7,5 contendo 1% Triton x-100, sendo que a presença do detergente Triton X-100 solubiliza parte da membrana, podendo aumentar a fração solúvel da proteína.

Todas estas alterações nas condições de expressão protéica, porém, não apresentam resultados satisfatórios, pois a proteína permaneceu na porção insolúvel do lisado celular, formando corpos de inclusão (tabela 5).

Contudo, partindo dos resultados descritos anteriormente e mantendo o objetivo de obtenção da proteína solúvel *in vivo*, foi tentada a subclonagem do fragmento gênico que codifica a proteína alvo em vetor pET28a(+) e os resultados foram satisfatórios.

Todas estas tentativas de se obter a proteína solúvel *in vivo* se dá ao fato de que se for realizado experimentos de renaturação, não há experimentos viáveis a fim de analisar se a mesma se encontra enovelada, pois seria necessária a consolidação de ensaios funcionais, o que é impossível no momento já que não é sabida a seqüência de ligação da mesma ao DNA.

Assim, o ensaio de expressão realizado em mini escala da proteína subclonada em pET28a(+) foi feito inicialmente à 37°C e apresentou alguns resultados insatisfatórios, pois a proteína continuava formando corpos de inclusão

ou não estava sendo expressa (tabela 5), com exceção de apenas uma linhagem, a C43 (DE3), em que a proteína apresentou-se parcialmente em sua forma solúvel, apesar de também ter formado corpos de inclusão. Tendo sido então a proteína induzida em larga escala e a fração solúvel submetida à purificação (figura 9).

6.3) Purificação da proteína codificada por Xf0389

A proteína foi purificada parcialmente pois além de apresentar outras bandas incluindo a banda de indução (figura 10), a amostra apresentou grande quantidade de DNA quando medida a absorbância. Foram feitas algumas tentativas para retirar o DNA, como utilização de sulfato de estreptomicina (2% peso/volume), sulfato de amônio (50% de saturação) e aumento da concentração de sal, mas nenhum destes métodos foi eficaz.

6.4) Caracterização estrutural

Os dados de CD mostraram que a cauda de histidina não alterou as estruturas secundárias da proteína em questão e mesmo a amostra contendo DNA, é possível ver que Xf0389 mantém seu enovelamento (figura 12), pois o espectro mostrou bandas negativas nos comprimentos de onda de 208 e 222 nm, o que caracteriza proteínas predominantemente constituídas por α -hélice (Kelly e Price, 2000).

6.5) Estabilidade térmica

O aquecimento da amostra, não gerou uma total perda da estrutura secundária da proteína e também não ocorreu a formação de *random coil*, caracterizado por uma banda negativa a 200 nm (figura 13A). Ainda, não foi possível determinar a temperatura de desnaturação (*melting temperature*) de Xf0389 já que, dentro do limite de temperatura imposto pelo sistema (até 90°C), não foi possível alcançar uma total desnaturação da amostra (figura 13B).

6.6) Construção do modelo por homologia

Como o objetivo inicial de resolução da estrutura cristalográfica não pode ser cumprido devido a baixa eficiência tanto da expressão quanto da purificação, já que foi conseguido apenas 2 mg/L de cultura induzida, enquanto, segundo a literatura, proteínas homólogas a esta foi cristalizada a uma concentração entre 15 a 20 mg/ml (Baikalov *et al.*, 1996; Baikalov *et al.*, 1998; Buckler *et al.*, 2002; Shin *et al.*, 2002; Robinson *et al.*, 2003), partiu-se para a caracterização estrutural através de técnicas de modelagem molecular por homologia *in silico*. Este procedimento foi realizado somente para a proteína reguladora (Xf0389), pois não foi encontrada nenhuma outra proteína que apresentasse sua estrutura resolvida com identidade de seqüência significativa à proteína sensora (Xf0390).

Assim, após escolhida a estrutura que apresentava significativa identidade de seqüência com a proteína Xf0389, ambas foram alinhadas seqüencial e estruturalmente e como pode ser visto na figura 15, a única diferença obtida no alinhamento estrutural entre a proteína-alvo (reguladora) e a estrutura-molde (DrrD) residiu entre os resíduos 126 ao 146, em que o segmento correspondente a fita- β na estrutura de DrrD, corresponde a uma α -hélice segundo a predição de Xf0389. No entanto, observou-se que a confiabilidade da predição para esta região se encontra muito baixa e como o alinhamento apresentou resíduos conservados, essa diferença não é muito relevante para a qualidade final do modelo.

A proteína Xf0389 possui um grande conteúdo de α -hélice como mostrado nos dados de CD, sendo reforçado pelo modelo que mostra 41,8% de α -hélice e 21% de fita- β . O modelo também apresentou os dois domínios (regulador e efetor) característicos de proteínas que fazem parte da subfamília OmpR/PhoB, a qual a proteína DrrD pertence. A proteína alvo ainda apresenta em sua região reguladora o aspartato (53) que é fosforilado pela proteína histidina quinase, fazendo com que a proteína adquira sua forma ativa. É importante salientar que este aspartato é conservado em todas as proteínas pertencentes a este tipo de sistema. Além disso, a proteína alvo também possui outros resíduos conservados, como a treonina (T79), tirosina (Y98) e lisina (K101), que são importantes para sua conformação e função, pois formam o bolsão ácido ao redor do aspartato (D53) acceptor do grupo fosforil (Buckler *et al.*, 2002; mostrado em negrito na figura 15 e

figura 21). Normalmente a lisina 101 é seguida de uma prolina, gerando uma ligação na conformação cis, o que é observado em outras estruturas já determinadas para este domínio e encontrada também na proteína Xf0389 (Djordjevic *et al.*, 1998). Como não há na literatura estrutura resolvida com o fosfato ligado, não é possível discutir quanto ao tipo de ligação que o fosfato faz com o aspartato.

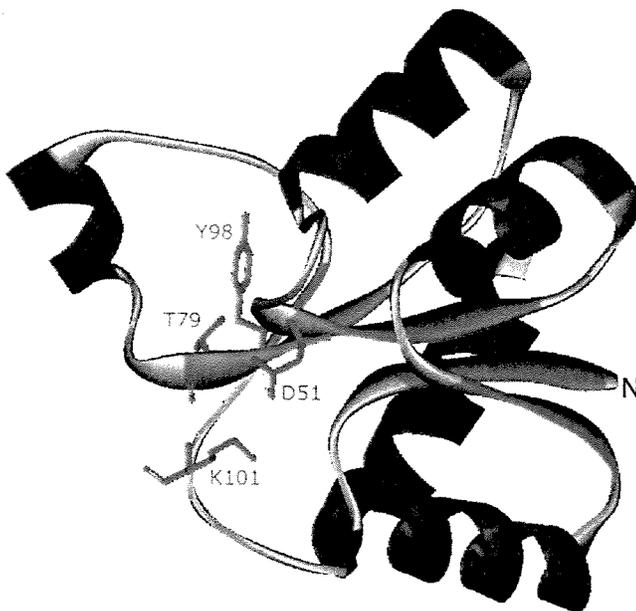


Figura 21: Domínio regulador da proteína Xf0389 e os resíduos que são conservados em diversas proteínas da subfamília OmpR/PhoB, com exceção do aspartato (D51) que é absolutamente conservado em qualquer proteína reguladora deste sistema.

Analisando o modelo e sua qualidade estereoquímica, foi mostrado que apenas uma lisina (K56; figura 17) se encontra em região não permitida, o que mostra que este resíduo não se encaixou perfeitamente no modelo, porém, não há comprometimento da qualidade estereoquímica do modelo, visto que este resíduo não está localizado em segmentos de estrutura secundária e sim em uma alça. Além disso, mais de 95% dos resíduos encontram-se nas regiões mais favoráveis do mapa de Ramachandran, o que atesta a qualidade estereoquímica do modelo. Também foi mostrado, como esperado, que a maior parte das regiões de estruturas secundárias são menos acessíveis ao solvente.

Ainda seria possível inferir, a partir do alinhamento estrutural entre as proteínas Xf0389, OmpR e phoB (figura 22), que provavelmente o domínio de

ligação da proteína ao DNA seria: 189 VLEVFIGRLRKKL 201, equivalente a terceira α -hélice do domínio efetor (Buckler *et al.*, 2002).

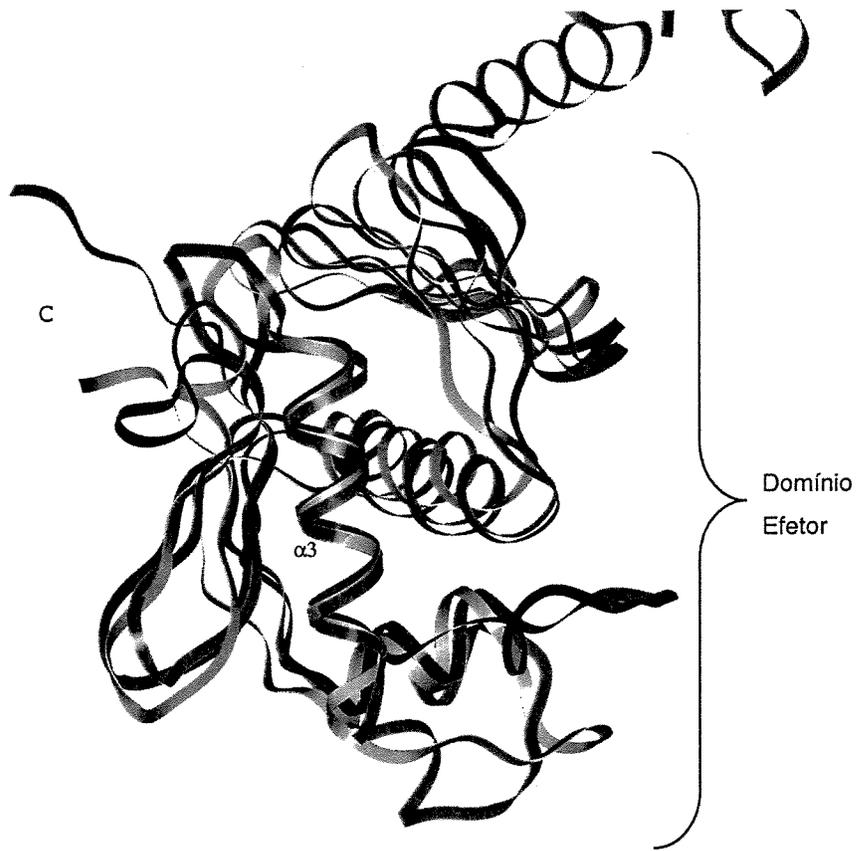


Figura 22: Sobreposição estrutural dos traços dos C- α correspondentes aos domínios efetores das proteínas Xf0389 (azul), OmpR (vermelho) e PhoB (laranja). As proteínas OmpR e PhoB, ambas de *E. coli*, foram resolvidas por cristalografia e ressonância magnética nuclear, respectivamente. α_3 representa a α -hélice que é o domínio de ligação ao DNA em OmpR e PhoB (Buckler *et al.*, 2002) e que foi usada como referência para a sobreposição.

6.7) Proteólise limitada

A partir da proteólise limitada (figura 20), observou-se que provavelmente a porção mais exposta da proteína é o fragmento da cauda de histidina, já que o tamanho que a proteína se apresentou após a proteólise com tripsina, subtilisina e quimiotripsina, era equivalente a ausência desta cauda (aproximadamente 25 kDa). E apesar do fragmento conectivo ser flexível, pela dificuldade de obtenção de dados de difração (Buckler *et al.*, 2002), este segmento não se apresentou suficientemente exposto à proteólise por tais enzimas.

6.8) Estudo comparativo

O sistema regulador de dois componentes PhoPQ está envolvido em diversos processos, entre os quais (Groisman, 2001):

- ✓ Operon PhoP-PhoQ medeia a adaptação a ambientes limitantes de Mg^{2+} ,
- ✓ Governa a virulência em varias espécies gram-negativas,
- ✓ Muitos genes regulados por PhoP são espécies específicas e conferem propriedades únicas ao microorganismo,
- ✓ Governa a modificação de muitos componentes do envelope celular bacteriano.

6.8.1) Xf0389

PhoP é a proteína reguladora do sistema regulador de dois componentes em *Salmonella sp.* e responde a mudanças na concentração de Mg^{2+} a qual PhoQ está exposta, dependendo da ativação por fosforilação (Castelli *et al.*, 2000). Vários genes são regulados por este tipo de mecanismo como alguns que estão estritamente presentes em *Salmonella* por terem sido incorporados ao cromossomo da bactéria por transferência horizontal, como: *pagCD* (patogenicidade tipo 3; Miller *et al.*, 1989), *mgtCB* (envolvidos no transporte de Mg^{2+}), *phoN* (codifica uma acido-fosfatase periplasmática não específica), *pcgL* (D-alanina-D-alanina peptidase periplasmática), *pagL* (catalisa a deacetilação de precursores de lipídio A), entre outros (revisado por Lejona *et al.*, 2003). E também regula genes que são similares em seqüência com genes de outras bactérias, como *mgtA* (o produto é um transportador de Mg^{2+} tipo P, necessário para o crescimento da bactéria em baixas concentrações deste cátion; Tao *et al.*, 1995), *syfB* (codifica uma lipoproteína externa; Soncini *et al.*, 1996), *pagP* (produto responsável pela acetilação de lipídio A e resistência a específicos peptídeos antimicrobianos; Guo *et al.*, 1998) e *pmrD* (ativação do operon PmrAB, outro sistema regulador de dois componentes, que controla a modificação da carga de lipolissacarídeos em resposta a concentração de Fe^{2+} e resistência a peptídeos antimicrobianos; Gunn *et al.*, 1998; Kox *et al.*, 2000).

Devido a similaridade de Xf0389 a PhoP e a fim de verificar se seria possível inferir a regulação desta proteína a alguns genes de *X. fastidiosa in silico*,

utilizou-se o programa tBLASTn (Altschul *et al.*, 1997) para alinhar as proteínas codificadas pelos genes *mgtA*, *syfB*, *pagP* e *pmrD* contra o genoma de *X. fastidiosa* e os resultados não foram satisfatórios pelo fato destes genes não apresentarem identidade significativa. Assim, não foi possível postular os genes regulados por este sistema em *X. fastidiosa*. E mesmo sabendo que a seqüência destes genes a qual a proteína reguladora se liga seja (T/G)GTTTA, localizada 25 pb antes do sítio de início da transcrição em *E. coli* e *Salmonella* (Soncini *et al.*, 1995; Kato *et al.*, 1999) em *X. fastidiosa* não foi possível prever se tal seqüência seja conservada pelo fato de não se saber quais genes são regulados por Xf0389.

6.8.2) Xf0390

As proteínas sensoras, como dito, são divididas em duas classes (I e II) enzimáticas, sendo que por homologia de seqüência a histidinas quinases de *Salmonella sp.* e *E. coli*, a proteína Xf0390 foi encaixada na classe I. Tais enzimas possuem duas regiões transmembrânicas, um C-terminal citoplasmático que porta o resíduo de histidina conservado, que é autofosforilado e um domínio periplasmático que contém vários resíduos acídicos envolvidos na percepção de cátions divalentes, nos dois microorganismos citados anteriormente (revisado por Groisman, 2001). Xf0390 também é similar a PhoQ de *Providencia stuartii* que aparentemente não responde a íons divalentes, mas sim a algum outro sinal não identificado ainda (Rather *et al.*, 1998).

Experimentos de substituição de aminoácidos mostraram que o *cluster* de resíduos acídicos (149 EDDDDAE 155) na região periplasmática de PhoQ em *E. coli* é responsável pela ligação de Mg^{2+} pois quando tais resíduos são substituídos por resíduos não carregados (QNNNNAQ), altera a habilidade de ligação da proteína a Mg^{2+} ou Ca^{2+} (Waldburger e Sauer, 1996).

Como visto na figura 23, este domínio também é conservado em *S. typhimurium* e não é conservado tanto em *P. stuartii* quanto em *X. fastidiosa*. Assim, como mostrado para *P. Stuartii* (Rather *et al.*, 1998), talvez Xf0390 não responda a concentrações de cátions divalentes, mesmo a bactéria habitando o xilema das plantas. Porém, segundo Groisman (2001), talvez não seja o ideal se apoiar somente neste fato, uma vez que nem todas as bactérias gram-negativas apresentam tal domínio conservado.

A)	151	200
XF0390	QVDGKLG	RVYRYG..MGLIWN
phoQ_Ecoli	QQLQEVRE	DDDDAEMTHSVAVNVVYPATSRMPKLTIVVVDTIPVELKSSYM
PHOQ_SALTY	EKLKEVRE	DDDDAEMTHSVAVNIYPATARMPLTIVVVDTIPIELKRSYM
P_stuartii	SKLDDMNHNDEP..LTHSVSVNQYGATENLPPLTIVVVDTLPQDLQKTGL	
Consensus	q.l.....d.....ths!	vn.y.at...P.ltiVvvDt.p..lq....
B)	251	300
XF0390	EQHPPELEPLTQ	SINAFIESERENLDRQRNTLADLAHSLKTPLAVLRTQL
phoQ_Ecoli	PATTREL	TSLVRNLNRLKSERERYDKYRTTLTDLTHSLKTPLAVLQSTL
PHOQ_SALTY	PETTREL	TSLVRNLNQLKSERERYNKYRTTLTDLTHSLKTPLAVLQSTL
P_stuartii	ENPPYELRGLVRNLNLLTNERKRY	SKYRTTLDLTHSLKTPLAVLQSTL
Consensus	e..p.EL..LvrnlN.l	l.sERerydKyRtTL.DLthSLKTPLAVLqstL

Figura 23: Detalhamento do alinhamento entre as histidinas quinases de *X. fastidiosa*, *E. coli*, *S. typhimurium* e *P. stuartii* respectivamente, gerado pelo programa *Multiple Sequence Alignment* (Corpet, 1988). A). O domínio ácido de ligação a cátions divalentes é mostrado em azul. B) Em amarelo é mostrado o resíduo de histidina conservado entre tais proteínas, responsável pela fosforilação e transferência deste grupo para o aspartato da proteína reguladora. O alinhamento completo é mostrado no apêndice I.

7) Conclusões

Foi um trabalho pioneiro com objetivos de caracterização estrutural e funcional para as proteínas Xf0389 e Xf0390 que formam um dos sistemas de dois componentes de *X. fastidiosa*.

Os genes Xf0389 e Xf0390 foram clonados em vetor pGEM-T-Easy e após confirmação da seqüência, o gene Xf0389 foi subclonado satisfatoriamente em vetores de expressão (pET28a(+), pET29a(+)) e pGEX4T-3). Após transformação de cepas para expressão, o gene foi induzido e sua proteína foi obtida na sua forma solúvel em *E. coli* C43 (DE3) com vetor pET28a(+). Assim, foi desenvolvido protocolo para cromatografia de afinidade por metal para a proteína Xf0389. A cromatografia ainda não foi totalmente padronizada pelo fato da proteína se apresentar parcialmente pura, porém para os presentes ensaios foi eficiente. A proteína também foi caracterizada por meio de CD e ensaio de estabilidade térmica.

A utilização de técnicas de modelagem molecular, particularmente a construção de modelos por homologia, auxiliou na interpretação dos dados experimentais obtidos, de forma a permitir uma caracterização estrutural preliminar da proteína Xf0389 e também na identificação da subfamília a qual a proteína pertence.

8) Referências Bibliográficas

Adler, A. J., Greenfield, N. J. and Fasman, G. D. (1973). Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol* **27**, 675-735.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-10.

Altschul S. F., Stephen F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.

Aranda, C. M. A., Swanson, J.A., Loomis, W.P. and Miller, S.I. (1992) Salmonella typhimurium activates virulence gene transcription within acidified macrophage phagosomes. *Proc Natl Acad Sci. USA* **89**, 10079-83.

Araújo, W. L., Marcon, J., Maccheroni, W. J., van Elsas, J. D., van Vuurde, J. W. L. and Azevedo, J. L. (2002) Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. *Appl Environ Microbiol* **68**, 4906-14.

Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P. and Dickerson, R.E. (1996) Structure of the Escherichia coli response regulator NarL. *Biochemistry* **35**, 11053-61.

Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Cascio, D., Gunsalus, R.P. and Dickerson, R.E. (1998) NarL dimerization? Suggestive evidence from a new crystal form. *Biochemistry* **37**, 3665-76.

Bergamin Filho, A. e Amorim, L. (2002) Doenças com período de incubação variável em função da fenologia do hospedeiro. *Fitopatol Bras* **27**, 561-65.

Bourret, R. B., Hess, J. F. and Simon, M. I. (1990) Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. *Proc Natl Acad Sci USA* **87**, 41-5.

Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* **4**, 187-217.

Bruix, M., Pascual, J., Santoro, J., Prieto, J., Serrano, L. and Rico, M. (1993) 1H- and 15N-NMR assignment and solution structure of the chemotactic Escherichia coli Che Y protein. *Eur J Biochem* **215**, 573-85.

Buckler, D. R., Zhou, Y. and Stock, A. M. (2002) Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from *Thermotoga maritima*. *Structure* **10**, 153-64.

Castelli, M. E., Garcia Vescovi, E. and Soncini, F. C. (2000) The phosphatase activity is the target for Mg²⁺ regulation of the sensor protein PhoQ in Salmonella. *J Biol Chem* **275**, 22948-54.

Chang, C. J., Garnier, M., Zreik, L., Rosseti, V. and Bové, J.M. (1993) Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Curr Microbiol* **27**, 137-142.

Cho, H. S., Lee, S-Y, Yan, D., Pan, X., Parkinson, J. S., Kustu, S., Wemmer, D. E. and Pelton, J. G. (2000) NMR structure of activated CheY. *J Mol Biol* **297**, 543-51.

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

Cserzo, M., Wallin, E., Simon, I., Von Heijne, G. and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10**, 673-6.

Derzelle, S., Turlin, E., Duchaud, E., Pages, S., Kunst, F., Givaudan, A. and Danchin, A. (2004) The PhoP-PhoQ two-component regulatory system of *Photobacterium luminescens* is essential for virulence in insects. *J Bacteriol* **186**, 1270-9.

Djordjevic, S. and Stock, A. M. (1998) Structural analysis of bacterial chemotaxis proteins: components of a dynamic signaling system. *J Struct Biol* **124**, 189-200.

Djordjevic, S., Goudreau, P. N., Xu, Q. and Stock, A. M. (1998) Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proc Natl Acad Sci USA* **95**, 1381-6.

Donahue, R. A. And Bebee, R. L. (1999) *BL21-SI* Competent cells for protein expression in *E. coli*. *Focus* **21**, 49-51.

Dorel, C., Vidal, O, Prigent-Combaret, C., Vallet, I. and Lejeune, P. (1999) Involvement of the Cpx signal transduction pathway of *E. coli* in biofilm formation. *FEMS Microbiol Lett* **178**, 169-75.

Feher, V. A and Cavanagh, J. (1999) Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature* **400**, 289-93.

Finlay, B.B. and Falkow S. (1997) Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* **61**, 136-69.

Foussard, M., Cabantous, S., Pédelacq, J., Guillet, V., Tranier, S., Mourey L., Birck, C. and Samara J. (2001) The molecular puzzle of two-component signaling cascades. *Microbes Infect* **3**, 417-24.

- Garcia Vescovi, E., Soncini, F. C. and Groisman, E.A. (1994) The role of the PhoP/PhoQ regulon in Salmonella virulence. *Res Microbiol* **145**, 473-80.
- Garcia Vescovi, E., Soncini, F. C. and Groisman, E. A. (1996) Mg²⁺ as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* **84**, 165-74.
- Gill S. C. and von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.
- Groisman, E. A., Kayser, J. and Soncini, F. C. (1997) Regulation of polymyxin resistance and adaptation to low-Mg²⁺ environments. *J Bacteriol* **179**, 7040-5.
- Groisman, E. A. (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. *J Bacteriol* **183**, 1835-42.
- Gunn, J. S. and Miller, S. I. (1996) PhoP-PhoQ activates transcription of pmrAB, encoding a two-component regulatory system involved in Salmonella typhimurium antimicrobial peptide resistance. *J Bacteriol* **178**, 6857-64.
- Gunn, J. S., Lim, K. B., Krueger, J., Kim, K., Guo, L., Hackett, M. and Miller, S. I. (1998) PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol Microbiol* **27**, 1171-82.
- Guo, L., Lim, K. B., Poduje, C. M., Daniel, M., Gunn, J. S., Hackett, M. and Miller, S. I. (1998) Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* **95**, 189-98.
- Hannig, H. and Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends Biotechnol* **16**, 54-60.
- Hess, J. F., Oosawa, K., Matsumura, P. and Simon M. I. (1987) Protein phosphorylation is involved in bacterial chemotaxis. *Proc Natl Acaad Sci USA* **84**, 7609-13.
- Hoch, J. A. and Silvahy, T. J. (1995) Two components signal transduction. American Society for Microbiology, Washington, DC.
- Hopkins, D.L. (1989) *Xylella fastidiosa*: xylem-limited bacterial pathogen of plants. *Ann Rev Phytopathol* **27**, 271-90.
- Hulett, F. M. (1996) The signal-transduction network for Pho regulation in Bacillus subtilis. *Mol Microbiol* **19**, 933-9.
- Jones D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202.

- Kato, A., Tanabe, H. and Utsumi, R. (1999) Molecular characterization of the PhoP-PhoQ two-component system in *Escherichia coli* K-12: identification of extracellular Mg²⁺-responsive promoters. *J Bacteriol* **181**, 5516-20.
- Kelly, S. M. and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* **1**, 349-84.
- Kopp, E.B. and Medzhitov, R. (1999) The Toll-receptor family and control of innate immunity. *Curr Opin Immunol* **11**, 13-8.
- Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics* **14**, 51-5.
- Kox, L. F. F., Wösten, M. M. S. M. and Groisman, E. A. (2000) A small protein that mediates the activation of a two-component system by another two-component system. *Embo J* **19**, 1861-72.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *J. Appl. Cryst* **26**: 283-90.
- Lee, S-Y, Cho, H.S., Pelton, J.G., Yan D., Berry, E.A. and Wemmer D.E. (2001) Crystal structure of activated CheY. Comparison with other activated receiver domains. *J Biol Chem* **276**, 16425-31.
- Lejona, S., Aguirre, A., Cabeza, M. L., Vescovi, E. G. and Soncini, F. C. (2003) Molecular characterization of the Mg²⁺-responsive PhoP-PhoQ regulon in *Salmonella enterica*. *J Bacteriol* **185**, 6287-94.
- Makrides, S. C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* **60**, 512-38.
- Maniatis, T., Sambrook, J. and Russel, D. W. (1986) *Molecular Cloning: a laboratory manual*.
- Marques, L. L. R., Ceri, H., Manfio, G. P., Reid, D. M. and Olson, M. E. (2002) Characterization of biofilm formation by *Xylella fastidiosa* *in vitro*. *Plant Disease* **86**, 633-38.
- McElrone, A. J., Sherald J. L. and Forseth, I. N. (2001) Effects of water stress on symptomatology and growth of *Parthenocissus quinquefolia* infected by *Xylella fastidiosa*. *Plant Disease* **85**, 1160-64.
- McGuffin, L. J., Bryson, K. and Jones, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-5.

- Miller, S. I., Kukral, A. M. and Mekalanos J. J. (1989) A two-component regulatory system (phoP phoQ) controls *Salmonella typhimurium* virulence. *Proc Natl Acad Sci USA* **86**, 5054-8.
- Okamura, H., Hanaoka, S., Nagadoi, A, Makino, K. and Nishimura, Y. (2000) Structural comparison of the PhoB and OmpR DNA-binding/transactivation domains and the arrangement of PhoB molecules on the phosphate box. *J Mol Biol* **295**, 1225-36.
- Pelton, J.G., Kustu, S. and Wemmer D.E. (1999) Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. *J Mol Biol* **292**, 1095-110.
- Purcell, A.H. and Hopkins, D.L. (1996) Fastidious xylem-limited bacterial plant pathogens. *Annu Rev Phytopathol* **34**, 131-51.
- Rather, P. N., Paradise, M. R., Parojcic, M. M. And Patel, S. (1998) A regulatory cascade involving AarG, a putative sensor kinase, controls the expression of the 2'-N-acetyltransferase and an intrinsic multiple antibiotic resistance (Mar) response in *Providencia stuartii*. *Mol Microbiol* **28**, 1345-53.
- Roberto, S. R., Farias, P. R. e Bergamin Filho, A. (2002) Geostatistical analysis of spatial dynamics of citrus variegated chlorosis. *Fitopatol Bras* **27**, 599-604.
- Robinson, V. L., Wu, T. and Stock, A. M. (2003) Structural analysis of the domain interface in DrrB, a response regulator of the OmpR/PhoB subfamily. *J Bacteriol* **185**, 4186-94.
- Rodrigue A., Quentin, Y., Lazdunski, A., Méjean, V. and Foglino, M. (2000) Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol* **8**, 498-504.
- Šali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815.
- Sambrook, J. and Russel, D. W. (2001) *Molecular Cloning: a laboratory manual* 3^{ed}: A1.16.
- Sanders, D. A, Gillece-Castro, B.L. Stock A M., Burlingame, A L. and Koshland D. E. Jr. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *J Biol Chem* **264**, 21770-8.
- Sanger, F., Nicklen, S. and Plaxton, W. C. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463-7.
- Schenk P. W. and Snaar-Jagalska, B. E. (1998) Signal perception and transduction: the role of proteins kinases. *Biochimica et Biophysica Acta* **1449**, 1-24.

Shin, D. H., Yokota, H., Kim, R. and Kim, S. (2002) Crystal structure of conserved hypothetical protein Aq1575 from *Aquifex aeolicus*. *PNAS* **99**, 7980-85.

Simpson, A. G. F., Reinach, F. C., Arruda, P. (...) and Setubal, J. C. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151-59.

Smolka, M. B., Martins, D., Winck, F. V., Santoro, C. E., Castellari, R. R., Ferrari, F., Brum, I. J., Galembeck, E., Filho, H. D. C., Machado, M. A., Marangoni, S. and Novello, J. C. (2003) Proteome analysis of the plant pathogen *Xylella fastidiosa* reveals major cellular and extracellular proteins and a peculiar codon bias distribution. *Proteomics* **3**, 224-37.

Soncini F. C., Garcia Vescovi, E. G., and Groisman, E. A. (1995) Transcriptional autoregulation of the *Salmonella typhimurium* *phoPQ* operon. *J Bacteriol* **177**, 4364-71.

Soncini F. C., Garcia Vescovi, E. G., Solomon F. and Groisman, E. A. (1996) Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. *J Bacteriol* **178**, 5092-9.

Stock, A M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochemistry*, **69**, 183-215.

Tao, T., Snavely, M. D., Farr, S. G. and Maguire, M. E. (1995) Magnesium transport in *Salmonella typhimurium*: *mgtA* encodes a P-type ATPase and is regulated by Mg²⁺ in a manner similar to that of the *mgtB* P-type ATPase. *J Bacteriol* **177**, 2654-62.

Tisa, L. S. and Adler, J. (1992) Calcium ions are involved in *Escherichia coli* chemotaxis. *Proc Natl Acad Sci USA* **89**, 11804-8.

Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein Expr Purif* **28**, 1-8.

Volz, K. and Matsumura P. (1991) Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J Biol Chem* **266**, 15511-9.

Volz, K. (1993) Structural conservation in the CheY superfamily. *Biochemistry* **32**, 11741-53.

Waldburger, C. D. and Sauer, R. T. (1996) Signal detection by the PhoQ sensor-transmitter. Characterization of the sensor domain and a response-impaired mutant that identifies ligand-binding determinants. *J Biol Chem* **271**, 26630-6.

West, A. H. and Stock, A. M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-76.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *J Biol Chem* **270**, 25328-31.

Yeoman, K. H., Delgado, M. J., Wexler, M., Downie, J. A. and Johnston, A. W. (1997) High affinity iron acquisition in *Rhizobium leguminosarum* requires the *cycHJKL* operon and the *feuPQ* gene products, which belong to the family of two-component transcriptional regulators. *Microbiology* **143**, 127-34.

Apêndice I

Alinhamento completo entre as proteínas histidina quinase de *X. Fastidiosa*, *E. coli*, *S. typhimurium* e *P. stuartii*, gerado pelo programa *Multiple Sequence Alignment* (Corpet, 1988).

```

1      10      20      30      40      50      60      70      80      90      100     110     120     130
XF0390 MLGKRALLSRLMHSNGFQPRSLQRKQLFHRSFSLVFLRLAGVALDARFADTAEKLRERIKAVATAYSSVEFLRGG-----SLYSSDRPPPPHVPGGGLYS--EVIIPANRHSLSSEGPLIP
EcoLi  HKKLLRFFPLSLRVRFLLAARVVLVLSLRYGVAVLIGYSYDFDKITFRLLRGSNLFYTLAKHEKNNKLVAVLPEKIDKQSPHLLYDFENGLLAKRQDVPVHLKALQDPAHLKSNGEN
SALTY  HKKFRARHFLPLSLRVRFLLAARVVLVLSLRYGVAVLIGYSYDFDKITFRLLRGSNLFYTLAKHEKNNKLVAVLPEKIDKQSPHLLYDFENGLLAKRQDVPVHLKALQDPAHLKSNGEN
PstuarLi HRLKFKFKPLSLRVRFLCATSRVILLALILSYGIVAVLIGYSYDFDKITFRLLRGSNLFYTLAKHEKNNKLVAVLPEKIDKQSPHLLYDFENGLLAKRQDVPVHLKALQDPAHLKSNGEN
Consensus .....n.....f.PLSLrARL.Rhs.v.lal.l.gd.va.g.g..sf#kkt..iLR..snl.yslv.u.n#k.....p.ni..#PT....%Re.G..Lr.qr#ypu..n.ir..alk..g.l.

131     140     150     160     170     180     190     200     210     220     230     240     250     260
XF0390 PYGPLLKARBEVEFGPFETIVDQGLGRVRYG--HGLIANGTGNPHDEFPYIIVYHEDERALLGAILVFRVAVFVLSAGETVLLIARFLLQ--HSLRPTIHWINELAKVQRGQRHSCQHPPELEP
EcoLi  EIEADVNDISLLSGDHSIQQLQEVREDDDAEMTHSVAVAVPATSRAPKLIIVVVDIIPVELKSSYVHSAETVYLSANLLIPLLAHVAHNSLRPTIENLAKREVRELENNKRELLNPAITRELTSL
SALTY  EIEADVNDISLLSGDHSIQQLQEVREDDDAEMTHSVAVAVPATSRAPKLIIVVVDIIPVELKSSYVHSAETVYLSANLLIPLLAHVAHNSLRPTIENLAKREVRELENNKRELLNPAITRELTSL
PstuarLi EIDIDIGEIRLLIQNPFNSKLDNNHNDPEP--LINSYSVNDYGTENLPPLIIVVVDIIPVELKSSYVHSAETVYLSANLLIPLLAHVAHNSLRPTIENLAKREVRELENNKRELLNPAITRELTSL
Consensus eI.....at...lI.g...e...q..l.....d.....chsi.vn.g.at...P.Ib.IVvDk.P..Iq.....vu..uf.YoLshnl.I.vip.lIuIaa..HSLRPTI..I..#...l#..g.r#.#$.e..p.EI..l

261     270     280     290     300     310     320     330     340     350     360     370     380     390
XF0390 IOSTINAFIESERENLRDRNTLRULAHSLKTPILAVLRITQLDGRASENELREE-----LDVQLRQVNNLVSYQLARASSGKLFSPVVI--TMTFRETIVRGLKRYARAKVYLCEFEIDPKACFYDEPGD
EcoLi  VKNLNRLLKSERERYKYRITLTLHSLKTPILAVLRITQLDGRASENELREE-----LDVQLRQVNNLVSYQLARASSGKLFSPVVI--TMTFRETIVRGLKRYARAKVYLCEFEIDPKACFYDEPGD
SALTY  VKNLNRLLKSERERYKYRITLTLHSLKTPILAVLRITQLDGRASENELREE-----LDVQLRQVNNLVSYQLARASSGKLFSPVVI--TMTFRETIVRGLKRYARAKVYLCEFEIDPKACFYDEPGD
PstuarLi VKNLNRLLKSERERYKYRITLTLHSLKTPILAVLRITQLDGRASENELREE-----LDVQLRQVNNLVSYQLARASSGKLFSPVVI--TMTFRETIVRGLKRYARAKVYLCEFEIDPKACFYDEPGD
Consensus vrnIN..I..sERerydkyRtL..DLtHSLKTPILAVLRITQLDGRASENELREE-----LDVQLRQVNNLVSYQLARASSGKLFSPVVI--TMTFRETIVRGLKRYARAKVYLCEFEIDPKACFYDEPGD

391     400     410     420     430     440     450     460     470     480     490     500     505
XF0390 LQELLNLLNNAFKHAI SRVLLTASPGQVGTTRPGLVLSVEDDGGPIPLEEVSKILQRGVRRDERVHGIGLSTVQILYKGYRGLQVTRSDLELGGACFEVYLLPPL
EcoLi  FVEVAGVLDNACKCYCLEFVEISAKRTDE-----HLVYVEDDGGPIPLSKREYIEIRGURVDILRPGVGVGLAVAREITTEYVEKQIVRAGES--HLGGARHEVIFGRHSHPKDDE
SALTY  FVEVAGVLDNACKCYCLEFVEISAKRTDDE-----HLVYVEDDGGPIPLSKREYIEIRGURVDILRPGVGVGLAVAREITTEYVEKQIVRAGES--HLGGARHEVIFGRHSHPKDDE
PstuarLi FVEVAGVLDNACKCYCLEFVEVTRVGVGD-----HLVITVDGDPGVPESQRNLIRRGURVDILRPGVGLGLSTVAREITDQVDDITSDS--PLGGARHEVIFGRHSHPKDDE
Consensus f.EvEgN.#HNaKycLefVe.t.fr.g##.....bl.vI.VADDDGPGIPLs.r.e..f.RqR.DcLrPqG.Gl.s.a.#i...q.Y.G.i.....S.LGGAr.EVvI.rq.....


```