

Débora Pires Paula

“CARACTERIZAÇÃO DA PROTEÍNA H-NS RELACIONADA À
PATOGENICIDADE DE *Xylella fastidiosa*”

2001.08.520

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
Débora Pires Paula
e aprovada pela Comissão Julgadora.

[Assinatura]

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

Orientadora: Profa. Dra. Anete Pereira de Souza

Campinas
2004

INIDADE BC
1ª CHAMADA UNICAMP
P281c
/ EX
OMBO BC/ 58459
ROC 16 - 117 - 04
: D X
PREÇO R\$ 11,00
DATA 22.06.09
4ª CPD

CM00198328-6

B ID 317712

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

P281c

Paula, Débora Pires

Caracterização da proteína H-NS relacionada à patogenicidade de *Xylella fastidiosa* / Débora Pires Paula.--
Campinas, SP: [s.n.], 2004.

Orientadora: Anete Pereira de Souza

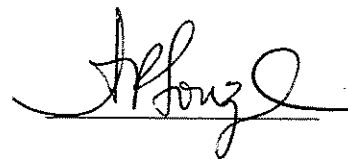
Tese (Doutorado) – Universidade Estadual de Campinas.
Instituto de Biologia.

1. Genomas. 2. Fitopatologia. 3. Patogenicidade. I. Souza, Anete Pereira de. II. Universidade Estadual de Campinas. Instituto de Biologia.
III. Título.

Campinas, 04 de março de 2004.

Banca Examinadora

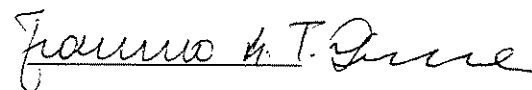
Profª. Dra. Anete Pereira de Souza (Orientadora)



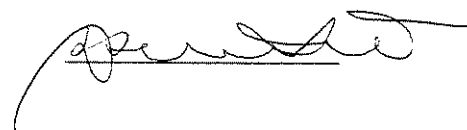
Prof. Dr. Michel Georges Albert Vincentz



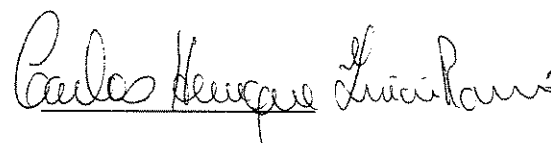
Prof. Dr. Francisco Benedito Teixeira Pessine



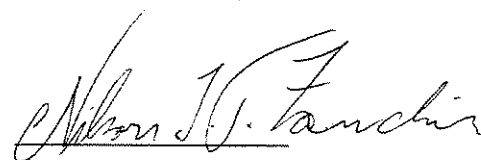
Profª. Dra. Maria Isabel Nogueira Cano



Prof. Dr. Carlos Henrique Inácio Ramos



Prof. Dr. Nilson Ivo Tonin Zanchin



DEDICO

A Deus, aos meus pais, irmãos, esposo e filho.

“Só se atiram pedras nas árvores que tem frutos.”

(Provérbio Chinês)

AGRADECIMENTOS

A realização deste trabalho tornou-se possível graças à colaboração direta e indireta de muitas pessoas. Manifesto minha gratidão a todas elas e de forma particular:

à professora Dra. Anete Pereira de Souza, pela orientação e por todas as oportunidades que me ofereceu;

aos professores Drs. Octavio Henrique Pavan e Adilson da Silva Leite[†];

aos professores Drs. Francisco Pessine, José Camillo Novello, Tomomasa Yano e Yoko Bomura Rosato, pelo empréstimo de equipamentos e pelo auxílio experimental;

à Rede de Biologia Molecular Estrutural (SMOLBnet), uma parceria da FAPESP com o Laboratório Nacional de Luz Sincrotron (LNLS), ligado ao Ministério da Ciência e Tecnologia, pela idealização e financiamento do projeto;

ao Centro de Biologia Molecular Estrutural (CEBiME – LNLS) e aos seus pesquisadores Drs.: Alberto Spisni, Beatriz G. Guimaraes, Carlos Henrique Inácio Ramos, Celso E. Benedetti, Javier Medrano, João Alexandre R. G. Barbosa, Jorg Kobarg, Nilson I. T. Zanchin, Rogério Meneghini, Thelma A. Pertinhez, pela utilização das instalações e equipamentos, fornecimento de materiais e pelo auxílio experimental; bem como à funcionária Rita de Cássia pelo auxílio experimental;

ao pessoal do Laboratório de Análise Genética e Molecular (Centro de Biologia Molecular e Engenharia Genética – CBMEG, Departamento de Genética e Evolução, IB/UNICAMP) Adriano, Alexandre, Antonio, Cleide, Dilaine, Fabíola, Karine, José Sérgio, Juverlande, Laura, Luciana Kauer, Luciana Lasry, Luciana Rossini, Maria Zilda, Milena, Prianda, Susely, Zildene e aos ex-colegas Andréia Navarro, Marcus, Kateli e Sibov;

ao meu amado esposo Tito por todo auxílio computacional e apoio emocional, bem como ao nosso amado filho Lucas pela compreensão e pelos momentos de muita alegria e descontração emprestados de sua infância;

e à Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP: 01/07533-7), pela bolsa de doutorado e apoio financeiro, respectivamente.

<i>Índice</i>	<i>página</i>
ABREVIACÕES E SÍMBOLOS	vii
1. RESUMO	xiii
1.1 ABSTRACT	xiv
2. INTRODUÇÃO	1
3. REVISÃO BIBLIOGRÁFICA	4
3.1 O fitopatógeno <i>Xylella fastidiosa</i>	6
3.2 Seleção da ORF XF0749 relacionada à patogenicidade	11
3.3 Função hipotética de regulação transcricional (H-NS)	13
4. ARTIGO 1	17
5. ARTIGO 2	25
6. RESULTADOS COMPLEMENTARES	55
6.1. Isoformas de H-NS	55
6.2. Ensaio de gel shift	57
6.3. Otimização da expressão heteróloga da ORF XF0749	60
6.4. RMN: 2D ^1H ^{15}N HSQC e medida da difusão em Gradiente de Campo Magnético Pulsado	65
6.5. Vida útil	74
6.6. Expressão heteróloga de outras ORFs relacionadas à patogenicidade de <i>X. fastidiosa</i>	77
7. CONCLUSÕES	80
8. PERSPECTIVAS	82
9. REFERÊNCIAS BIBLIOGRÁFICAS	85

ABREVIÇÕES E SÍMBOLOS

AFM	Microscopia de Força Atômica (<i>Atomic Force Microscopy</i>)
ÅKTA-FPLC	Palavra sueca para designar a atual plataforma de sistemas de purificação em <i>Fast Performance Liquid Chromatography</i>
AMPc, ATP	Adenosina-5'-monofosfato cíclico e adenosina-5'-trifosfato
APS	Persulfato de amônio
Bis-acrylamide	N,N'-metilenobisacrilamida
BL21(DE3)	Linhagem derivada da linhagem B de <i>Escherichia coli</i> que contém o fragmento do bacteriófago λ DE3 lisogênico, contendo a região de imunidade do fago 21 com o gene <i>lacI</i> , o promotor <i>lacUV5</i> , o início do gene <i>lacZ</i> e o gene para T7 RNA Polimerase.
BL21-SI	Linhagem BL21 em que a expressão heteróloga é induzível por cloreto de sódio (<i>Salt-Inducing</i>).
BL21(DE3)pLysS	Linhagem BL21 que contém o plasmídio que contém o gene T7 para co-expressão de lisozima.
BL21pT-GROE	Linhagem BL21 que contém o plasmídio para co-expressão das chaperoninas GroES e GroEL sob o controle do promotor T7.
BL21pT-TRX	Linhagem BL21 que contém o plasmídio para co-expressão de tireodoxina sob o controle do promotor T7.
BSA	Soro Albina Bovina (<i>Bovine Serum Albumine</i>)
CaCl₂	Cloreto de cálcio
CD	Dicroísmo Circular (<i>Circular dichroism</i>)
CHAPS	3-(3-cholamidopropyl)dimethylammonio-1-propane sulfonate
C-terminal	Extremidade proteica carboxi-terminal
CVC	Clorose Variegada dos Citros
D₂O	Água deuterada
DDT	Dicloro-Difenil-Tricloroetano
dGTP	Deoxiguanosina 5'-trifosfato

DH5α	Linhagem de <i>E. coli</i> comumente utilizada para transformação plasmídial por ser geneticamente modificada para não conter as DNAases <i>recA</i> ⁻ e <i>endA</i> ⁻ , conferindo alta eficiência de transformação e boa multiplicabilidade do plasmídial.
DLS	Espalhamento Dinâmico de Luz (<i>Dynamic Light Scattering</i>)
DTT	Ditiotreitol
DOSY	<i>Diffusion-Ordered Spectroscopy</i>
d_{prot}	Decaimento do espectro da proteína alvo
d_{ref}	Decaimento do espectro da molécula de referência
EcoRI	Primeira enzima de restrição identificada em <i>E. coli</i> , linhagem R
EDTA	Ácido Etileno Diaminotetra Acético
EMSA	<i>Electrophoretic Mobility Shift Assay</i>
F_0	Fluorescência na ausência do supressor
F	Fluorescência na presença do supressor
gi	Identificação gênica (número de acesso NCBI)
GroESL	Complexo das chaperoninas GroES e GroEL
GST	Glutathione-S-Transferase
HEPES	<i>N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid</i>
His-Tag	Resíduos de histidina como seqüência de fusão
HMFM	<i>Hogness Modified Freezing Medium</i>
H-NS	<i>Histone-like Nucleoid-Structuring Protein</i>
HSQC	<i>Heteronuclear Single Quantum Coherence</i>
HU	Proteína nucleóide <i>Heat-Unstable</i>
IMAC	<i>Immobilized Metal Affinity Chromatography</i>
IPG	<i>Immobilized pH Gradients</i>
IPTG	Isopropil-β-D-tiogalactopiranosídeo
K_A	Constante de equilíbrio de associação
K_1, K_2	Constantes de equilíbrio de dissociação
KCl	Cloreto de potássio

K_{SV}	Constante de supressão dinâmica de Stern-Volmer
LB	Luria-Bertani
LMN	<i>Low molecular-mass (Amersham Biosciences)</i>
MALDI-TOF	<i>Matrix-Assisted Laser-Desorption Ionisation Time-of-Flight</i>
MBP	<i>Maltose Binding Protein</i>
M9	Nome de um tipo específico de meio mínimo para crescimento de linhagens bacterianas compatíveis
M_r	Massa molecular relativa
MW	Massa molecular
N	Número de resíduos de uma cadeia qualquer de aminoácidos
NaCl	Cloreto de sódio
NaPi	Fosfato de sódio
NCBI	<i>National Center for Biotechnology Information</i>
NdeI	Primeira enzima de restrição identificada em <i>Neisseria denitrificans</i>
Ni-NTA	Resina de ácido nitriloacético com níquel adsorvido
NMR	Ressonância Magnética Nuclear
$^{15}\text{NH}_4\text{Cl}$	Cloreto de amônio com isótopo de nitrogênio 15
N-terminal	Extremidade proteica amino-terminal
O.D.	Densidade ótica
ORF	Seqüência abertura de leitura
PAGE	Eletroforese em gel de poliacrilamida
PCR	Reação em cadeia pela enzima Polimerase
PDB	<i>Protein Data Bank</i>
PEG	Polietileno Glicol
pET	Plasmídio comercial para expressão heteróloga sob o controle do forte promotor T7.
pET 32-Xa/LIC	Série do vetor pET que possui, entre outras características, uma seqüência de fusão de tireodoxina, sítio de clivagem das seqüências de fusão pela protease fator Xa e é <i>Ligation Independent Cloning</i> , ou seja, o vetor é fornecido linearizado, eliminando a necessidade de digestão por enzimas de restrição e de haver reações de ligação com a enzima Ligase.

pET 28a e 29a Série pET em que a sequência de fusão consiste em seis resíduos de histidina e sítio de clivagem pela protease Trombina.

PG-SLED *Pulse Gradient Stimulated Echo Longitudinal Encode-Decode*

PMSF *Phenylmethyl-Sulfonyl Fluoride*

proU Promotor do gene *proU*

pXF Plasmídios nativos de *Xylella fastidiosa*

Q *Quencher*

rbs Sítio de ligação ao ribossomo

R-250 A letra R indica a matiz avermelhada do corante (*reddish*), enquanto o número 250 indica o fator de intensidade de coloração do corante.

R_g Raio de giro

R_h Raio hidrodinâmico

R_h^{prot} Raio hidrodinâmico da proteína alvo

R_h^{ref} Raio hidrodinâmico da molécula de referência

RMN Ressonância Magnética Nuclear

SAXS *Small -Angle Scattering X-rays Studies*

SDS Sulfato de Sódio Dodecil

SMART *Simple Modular Architecture Research Tool*

SP Sulfo Propil

S-Tag Peptídio ribonuclease-S (15 aminoácidos) como sequência de fusão

TEMED N,N,N',N'-tetrametiletilenodiamina

Tm *Melting Temperature*

T7 Qualquer fragmento originado pelo bacteriófago T7

Tris Tris(hidroximetil)-aminoetano

TBE Tris-Borato EDTA

Tris-HCl Tris-Clorato

Trx-Tag Tioredoxina (109 aminoácidos) como proteína de fusão

UV	Ultravioleta
V	Constante de supressão estática
XF	<i>Xylella fastidiosa</i>
X-gal	5-bromo-4-cloro-3-indolil- β -D-galactosídio
xrvA	Regulador de virulência do <i>gene A</i> de <i>Xanthomonas oryzae</i>
YT	Meio rico contendo tripton e levedura
Gel 1 D	Eletroforese unidimensional
Gel 2 D	Eletroforese bidimensional
3 D	Estrutura tridimensional
[α-³²P]dNTP	Substituição do α -fosfato por ³² P em qualquer um dos quatro nucleotídeos (A, T, C, G) da extremidade 3' do DNA.

SÍMBOLOS DOS AMINOÁCIDOS

Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Cisteína	Cys	C
Glutamina ou glutamida	Gln	Q
Ácido glutâmico (glutamato)	Glu	E
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Treonina	Thr	T
Triptofano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

1. RESUMO

Xylella fastidiosa é um fitopatôgeno fastidioso hospedeiro do xilema de mais de 28 famílias de mono e dicotiledôneas, ocasionando grandes perdas econômicas em regiões de clima tropical e subtropical, principalmente nas Américas. No Brasil, é responsável pela grave doença denominada Clorose Variegada dos Citros (CVC). Sendo o primeiro fitopatôgeno a ter seu genoma sequenciado, tornou-se alvo da genômica funcional na busca pela compreensão de seu mecanismo de infecção e adaptação nos hospedeiros. Neste contexto, o presente trabalho estudou a ORF relacionada a patogenicidade XF0749 que, hipoteticamente, possui seqüência primária similar a proteínas da família de ligação a ácidos nucleicos H-NS de enterobactérias. No propósito de validar sua anotação gênica, a ORF foi clonada em vetor de expressão pET 32-Xa/LIC e sua proteína heteróloga expressa recombinada em larga escala na linhagem BL21(DE3). Após purificação proteica, fez-se a caracterização funcional e estrutural pela análise das mudanças conformacionais ao interagir com DNA (oligonucleotídeo de 14 pb) e ao sofrer desnaturação térmica e química por uréia, variação de pH e supressão pela acrilamida, monitoradas pelas técnicas espectroscópicas de Dicroísmo Circular e Fluorescência. Os experimentos de Ressonância Magnética Nuclear (RMN), para determinação da estrutura terciária, revelaram a necessidade da análise da estrutura quaternária por Espalhamento Dinâmico de Luz, por eletroforese em gel de poliacrilamida não desnaturante e por Dicroísmo Circular, para verificação da oligomerização da proteína. O presente trabalho propõe a validação da anotação gênica da ORF XF0749 de *X. fastidiosa* (clone 9a5c) pela caracterização funcional e estrutural de sua proteína ligante a DNA, assim como predito pela similaridade com os homólogos de estrutura primária.

1.1 ABSTRACT

X. fastidiosa is a fastidious phytopathogen that hosts the xylem of more than 28 families of mono and dycotiledones, causing great financial damage in tropical and subtropical regions, mainly in the Americas. In Brazil, it is responsible for a serious disease namely Citrus Variegated Chlorosis (CVC). With the objective of understanding its infection and mechanism on its hostess, it became a target of functional genomic, as it was the first phytopathogen to have its genome sequenced. Due to this the present work studied the pathogenicity related XF0749 ORF that, hypothetically has primary sequence similarity with DNA-binding proteins from enterobacteria H-NS family. In order to validate the genomic annotation, the ORF was cloned in the pET32-Xa/LIC expression vector and its recombinant protein expressed in on a large scale in the BL21(DE3) lineage. After protein purification, the functional and structural characterization was carried out by the analysis of the conformational changes when interaging with DNA (14 bp oligonucleotide) and under thermal and chemical unfolding denaturation by urea, pH variation and acrylamide quenching, monitored by the spectroscopic techniques of Circular Dichroism and Fluorescence. Nuclear Magnetic Resonance for tertiary structure determination revealed the necessity of quaternary structure analysis by Dynamic Light Scattering, non-denaturing PAGE electrophoresis and Dichroism Circular, in order to evaluate the degree of protein oligomerization. The present work validated the genetic annotation of the XF0749 ORF from *X. fastidiosa* (9a5c clone) by the functional and structural characterization of the its DNA-binding protein, as predicted by primary structure homologous similarity.

2. INTRODUÇÃO

Muitos genomas de arqueobactérias, bactérias e eucariotos têm sido sequenciados desde a década de 90, tais como o de *Drosophila melanogaster*, *Saccharomyces cerevisiae*, nematódio *Caenorhabditis elegans*, planta *Arabidopsis thaliana*, camundongo *Mus musculus* e várias bactérias (Suzuki *et al.*, 1992), e muito mais está previsto para ser completado nos próximos anos (Zarembinski *et al.*, 1998). Os dados gerados por tais estudos impulsionaram a entrada na chamada era pós-genômica ou genômica funcional onde, através das análises do transcriptoma¹ e do proteoma², aliado ao conhecimento da genômica estrutural³, efetua-se a caracterização molecular completa do organismo (Griffiths *et al.*, 2000). Assim, as informações geradas pelo sequenciamento genômico da linhagem 9a5c de *Xylella fastidiosa* (projeto ONSA/2000, Simpson *et al.* 2000) têm sido alvo da genômica funcional, na busca por uma melhor compreensão dos mecanismos de fitopatogenicidade e adaptabilidade do fitopatógeno, visando o seu combate.

A bactéria *X. fastidiosa* é um fastídio de xilema hospedeiro de mais de 28 famílias de mono e dicotiledôneas, sendo muitas de grande importância econômica (Purcell & Hopkins, 1996), como a cultura da laranja onde causa a doença Clorose Variegada dos Citros (CVC) no Brasil (Laranjeira, 1997; Lee *et al.*, 1992).

Os métodos de controle do fitopatógeno ainda se constituem em manejos paliativos como a poda dos galhos atingidos, erradicação definitiva da árvore contaminada, controle químico dos vetores e injeções de antibiótico na planta. Alguns dos defensivos agrícolas

¹ Compreende a gama de transcritos produzidos por um dado organismo.

² Conjunto completo de proteínas codificadas por um organismo.

³ Análise da estrutura do genoma: mapeamento físico e genético, bem como sua sequência de bases.

utilizados causam fitotoxicidade e apresentam alto custo (Raju & Wells, 1986). O estudo das funções proteicas através do conhecimento da estrutura terciária das proteínas do fitopatógeno podem implementar meios de investigação mais eficazes para seu controle, possibilitando o aumento na competitividade e produtividade da citricultura brasileira.

Apesar da grande importância do estudo das ORFs de *X. fastidiosa* para a melhor compreensão de seus processos celulares, torna-se impraticável determinar experimentalmente a estrutura tridimensional de todas as suas proteínas, devido a dificuldades nos processos de expressão proteica, purificação, cristalização e aquisição de dados por RMN devido à características intrínsecas de cada uma. Uma das abordagens utilizadas é a do estudo estrutural de proteínas cujas funções são pouco conhecidas. Para isto, baseando-se nos critérios descritos na seção 3.2 desta tese, selecionou-se a ORF XF0749 (gi 1360644) de um total de 147 ORFs de patogenicidade, produzindo-a em sistema heterólogo e, após sua purificação, utilizando-a para estudos funcionais e estruturais.

A ORF XF0749 codifica uma proteína citoplasmática de 134 aminoácidos (15 kDa) de sequência primária similar⁴ às sequências de domínios de ligação a ácidos nucleicos da família H-NS (*Histone-like Nucleoid-Structuring Protein*). Proteínas desta família organizam os ácidos nucleicos em estruturas nucleóides, sendo também responsáveis pela regulação transcricional de muitos genes não correlacionados, afetando a expressão gênica sob certas condições, seja pela ligação direta ao DNA ou pela alteração de sua topologia (compactação) (Azam *et al.*, 1999).

No presente trabalho de doutorado, teve-se como objetivos específicos a clonagem,

⁴ O termo “similaridade” tem uma conotação de “homologia” nesta tese pelo fato de que as proteínas similares encontradas pertencem a enterobactérias, portanto, partilhando o mesmo ancestral de *X. fastidiosa*.

expressão, purificação e a caracterização funcional e estrutural da proteína da ORF XF0749 relacionada à patogenicidade. Como objetivos gerais, visou-se a contribuição para o conhecimento dos mecanismos de patogenicidade de *X. fastidiosa*, bem como o fornecimento de subsídios ao estudo da estrutura tridimensional da proteína H-NS do fitopatógeno.

3. REVISÃO BIBLIOGRÁFICA

Em todos os genomas sequenciados uma larga porção das regiões codificantes codificam polipeptídeos de função bioquímica, biofísica e/ou celular desconhecidas. Para prever a função, a sequência de aminoácidos codificada pela ORF é inicialmente comparada com todas as sequências funcionais depositadas em bancos de dados (*p.ex. Protein Data Bank - PDB*). Caso exista uma identidade de sequência ou domínio de no mínimo 30 %, presume-se que as proteínas compartilham função similar. Entretanto, até 62 % das ORFs destes genomas compartilham pouca ou nenhuma identidade com qualquer sequência depositada e, então, são de função desconhecida (Zarembinski *et al.*, 1998).

Uma abordagem para determinar a função molecular é a determinação da estrutura tridimensional da proteína por cristalografia de raios-X ou por RMN. A estrutura tridimensional, ao invés da sequência primária de aminoácidos, é comparada com estruturas proteicas em bancos de dados. Se houver uma ou mais estruturas similares significativas (entre 30 – 50 %), a proteína hipotética é predita ter propriedades moleculares similares, podendo ser testado experimentalmente. Tal função molecular propicia a base para o conhecimento da função celular da proteína, sendo muito mais precisa que alinhamentos de estrutura primária. Muitas vezes proteínas de baixa similaridade de sequências primárias possuem alta similaridade de estrutura terciária ou de função molecular (Zarembinski *et al.*, 1998).

Existem muitos exemplos de estruturas tridimensionais altamente conservadas na evolução proteica (Chothia & Lesk, 1986). Por esta razão, a comparação de estruturas tridimensionais possibilitam a volta no tempo evolucionário e o estabelecimento da conexão entre famílias proteicas (Gibrat *et al.*, 1996). Embora o universo de sequências

primárias de proteínas distintas seja essencialmente ilimitado, o número de diferentes padrões de enovelamento para estas proteínas não o são. Proteínas de alta massa molecular comprimem invariavelmente um número de pequenos domínios, os quais são geralmente de 100 – 250 resíduos de aminoácidos. Embora proteínas com enorme variedade estrutural sejam geradas pela combinação de domínios individuais, o número de enovelamentos distintos aparentemente é limitado a poucos milhares. Isto tem conduzido esforços internacionais para desenvolver projetos sistemáticos de Biologia Molecular Estrutural tal que no *PDB* têm-se, no mínimo, um exemplo de cada tipo de enovelamento de domínio (Sali & Kuriyan, 1999).

Muito embora o conhecimento da estrutura terciária seja valioso para a determinação da função proteica, o atual conhecimento da relação entre estrutura e função ainda é insuficiente para deduzir sua função global unicamente por sua estrutura. Em todos os casos, faz-se necessário combinar estudos bioquímicos (*in vitro*) e celulares (*in vivo*) com as informações estruturais para completar a compreensão global do mecanismo celular proteico (Chothia & Lesk, 1986).

Algumas vezes os ensaios funcionais *in vitro* e/ou *in vivo* tornam-se possíveis antes que se tenha o conhecimento da estrutura terciária. Nestes casos, além da validação da anotação genômica da ORF do organismo sequenciado, estes estudos auxiliam na resolução da estrutura, fornecendo subsídios informativos sobre como a proteína poderia enovelar para desempenhar sua função e de como obtê-la nativa em altas concentrações.

De uma maneira ou de outra, o objetivo final é a elucidação da função proteica a partir de dados obtidos do genoma. Esta é uma forte tendência mundial, seguindo a corrida

funcional em diversos organismos, na busca pela compreensão e solução de muitas lacunas presentes nos mais diversos processos biológicos.

3.1 O fitopatógeno *Xylella fastidiosa*

A bactéria *X. fastidiosa* (Proteobacteria, Gamma proteobacteria, Xanthomonadales, Xanthomonadaceae) é um patógeno gram-negativo habitante de xilema e de crescimento fastidioso e lento. Tem forma bastonete, não flagelada, não forma esporos, medindo cerca de 0,3-0,5 μm de diâmetro e 1-5 μm de comprimento. Possui bastante similaridade com as bactérias do gênero *Xanthomonas*, agrupadas no mesmo subgrupo γ de eubactérias por Wells *et al.* (1987). Ela ocorre somente nos tecidos do xilema de plantas afetadas, de melhor crescimento a 20-25°C em pH 6,7-7,0 (Raju & Wells, 1986).

A bactéria ocorre principalmente em áreas de clima tropical ou subtropical, sendo raramente encontrada fora das Américas. Sua incidência diminui gradualmente conforme diminui a temperatura. Registros isolados foram observados apenas numa espécie de pereira em Taiwan (Leu & Su, 1993) e em outra de ameixeira na Índia (Jindal & Sharma, 1987).

O fitopatógeno é causador de doenças em muitas espécies de plantas, nas quais geralmente é letal num período de 2 a 3 anos (Raju & Wells, 1986). Entre estas encontram-se a alfafa e videira (Goheen *et al.*, 1973); pessegueiro (Hopkins & Mollenhauer, 1973); amendoeira (Mircetich *et al.*; 1976); ameixeira (Hopkins, 1989); cafeeiro (Paradela-Filho *et al.*, 1997); elmo; carvalho; vinca e a planta ornamental espiroleira (Purcell *et al.*, 1999). Duas espécies de gramíneas (*Echinochloa cruz-galli* e *Artemisia douglasiana*) e o

"blackberry" (*Rubus procerus*) também são hospedeiras mas não apresentam sintomas, sendo então consideradas reservatórios da bactéria (Hopkins & Purcell, 2002). No Brasil a Clorose Variegada dos Citros (CVC), popularmente conhecida como "Amarelinho", é considerada como a doença mais devastante em citros (Lee *et al.*, 1992) sendo um dos principais fatores responsáveis pela produtividade nacional da laranja não ser ainda maior (Laranjeira, 1997).

A doença abrange todas as variedades comerciais de laranja doce. Poucos meses após a infecção (6 a 8) a planta torna-se inteiramente tomada pela doença, causando rápida queda na produção de frutos de tamanho normal. A planta não morre, mas pára de crescer e de produzir frutas comercializáveis tanto para a indústria, como para o mercado de fruta fresca. A produtividade estabiliza-se por um ou dois anos após a infecção e, em seguida decresce continuamente, tornando-se antieconômica (Tubelis, 1992).

Os primeiros sintomas relacionados à doença são a redução foliar e variação conspícua nas folhas mais velhas, com áreas cloróticas na parte inferior à semelhança de sintomas de deficiência de zinco (Queiroz-Voltan & Paradela-Filho, 1999). Posteriormente ocorre a redução e o endurecimento dos frutos, com decréscimo significativo na produção (Fig. 1).

Estes sintomas têm sido observados desde árvores de viveiro a árvores com 7 a 10 anos de idade, sendo que árvores mais velhas que 15 anos são menos afetadas (Lee *et al.*, 1993).

O impacto econômico causado pela infecção na cultura da laranja tem causado grandes perdas para o Brasil, já que esta cultura sozinha é responsável por 18 % do PIB, e

por empregar cerca de 320 mil trabalhadores no campo e 100 mil na indústria (Lee *et al.*, 1992; www.fundecitrus.com.br).

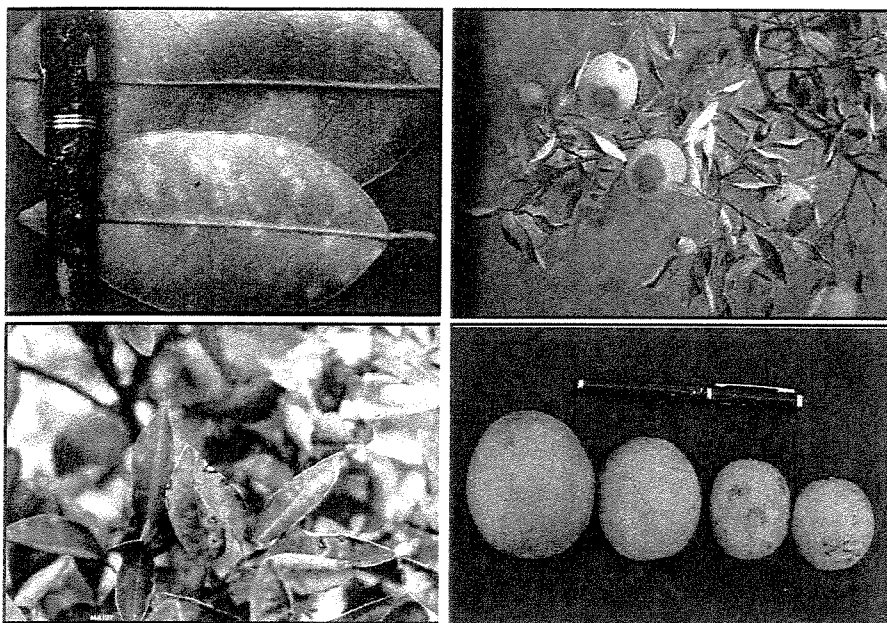


Fig. 1: Sintomas da Clorose Variegada dos Citros. Fonte: Fundecitrus (www.fundecitrus.com.br).

A CVC têm sido identificada na Argentina e no Brasil, onde vem ocasionando, desde a década de 80, prejuízos da ordem de 360 milhões de reais. Tal fato deve-se à sua rápida propagação, tendo contaminado 86 % dos municípios paulistas (principal Estado produtor nacional), devido à inexistência de um método definitivo de combate (Roberto *et al.*, 1996).

Desde seu surgimento no município paulista de Macaúbal (meados de 1987), a doença vem se propagando vertiginosamente, pondo em risco a sobrevivência da citricultura nas regiões Norte e Nordeste do Estado de São Paulo e na região do Triângulo Mineiro, no Estado de Minas Gerais (Tubelis, 1992). Sintomas da CVC já foram relatados na Bahia, Goiás, Paraná, Rio Grande do Sul, Rio de Janeiro, Santa Catarina, Sergipe e Distrito Federal (Tubelis *et al.*, 1993).

A bactéria é transmitida através de mudas contaminadas e por meio de vetores homópteros (Homoptera, Cicadellidae, subfamília Cicadillinae) que se alimentam do xilema. Atualmente são conhecidas dezesseis espécies de cigarrinhas⁵ capazes de transmitir a doença (Roberto *et al.*, 1996) (Fig. 2). A bactéria coloniza o cibário do inseto, mas não circula pelo vetor. Quando o inseto se alimenta do xilema, a bactéria é expelida para os traqueídeos da planta hospedeira. Embora o xilema seja um ambiente nutricionalmente pobre, contém uma grande variedade de aminoácidos monoméricos, ácidos orgânicos e íons, bastante utilizados por bactérias fastidiosas de xilema (Purcell & Hopkins, 1996).

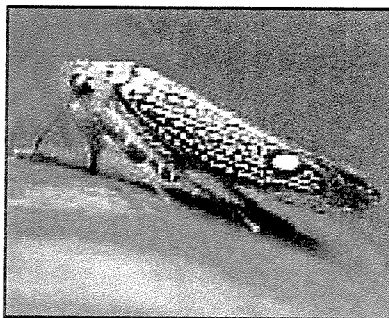


Fig. 2: Cigarrinha *Parathona gratiosa*, vetor de transmissão da doença Clorose Variegada dos Citros. Fonte: Fundecitrus (www.fundecitrus.com.br).

Os mecanismos de patogenicidade ainda são pouco conhecidos, sendo os mesmos atribuídos a fatores diversos, tais como: disfunção do sistema de condução de água na planta por oclusão vascular (Fig. 3) ocasionando estresse hídrico (Hopkins, 1989; Purcell & Hopkins, 1996), produção de fitotoxinas e desbalanço dos reguladores de crescimento da planta (Lee *et al.*, 1991; Lee *et al.*, 1993), e competição por nutrientes entre a planta e a bactéria, levando ao surgimento dos sintomas (Machado *et al.*, 2001).

⁵ As 16 espécies de cigarrinhas são: *Acrogonia terminalis*, *A. virescens*, *Bucephalogonia xanthophis*, *Carneocephala fulgida*, *Dilobopterus costalimai*, *Draeculocephala minerva*, *Ferrariana trivittata*, *Helochara communis*, *Homalodisca coagulata*, *H. ignorata*, *Macugonalia leucomelas*, *Oncometopia facialis*, *O. nigricans*, *Parathona gratiosa*, *Plesiommatia corniculata*, *Sonesimia grossa*.

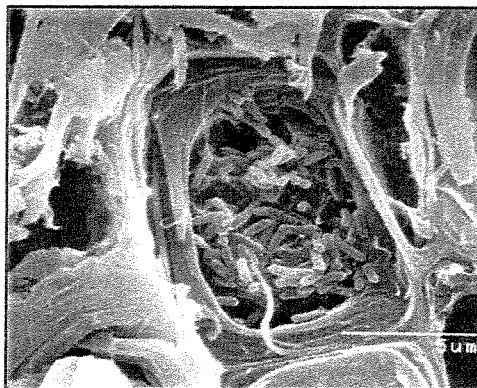


Fig. 3: Oclusão do xilema de citros por *Xylella fastidiosa*. Fonte: aeg.lbi.ic.unicamp.br/xf.

Os genes envolvidos com patogenicidade são o alvo no estudo de bactérias patogênicas e especialmente no caso de *X. fastidiosa*, essa característica foi uma das mais importantes para ter seu genoma sequenciado, já que é poderoso agente patológico de várias culturas, algumas de grande valor comercial (Simpson *et al.*, 2000).

Depois da publicação do genoma da linhagem 9a5c, isolada de citros infectado no Estado de São Paulo por Chang *et al.* (1993), uma linhagem isolada de videira foi sequenciada e outras duas linhagens isoladas de amêndoa e espiroideira tiveram seus genomas parcialmente sequenciados (Oliveira *et al.*, 2002; Bhattacharyya *et al.*, 2002 a, 2002 b). O isolado de citros foi a linhagem que apresentou as maiores diferenças em relação às demais. 13 % das ORFs não são encontradas nos outros genomas. Estas ORFs codificam diferentes moléculas, entre elas proteínas ribossomais, tyrosil tRNA sintetase, fatores de transcrição e transportadores. Estas diferenças funcionais podem sugerir mecanismos de controle biológico específicos (Bhattacharyya *et al.*, 2002 a). Uma comparação entre genomas de bactérias associadas a plantas revelou que *X. fastidiosa* possui a menor quantidade de G + C e a menor proporção de regiões codificadoras. Estes fatores levaram alguns autores a afirmar que seria um patógeno "mínimo" (Dow & Daniels,

2000; Van Sluys *et al.*, 2003), não possuindo certas vias metabólicas, impedindo a sua rápida adaptação a diferentes ambientes.

3.2 Seleção da ORF XF0749 relacionada à patogenicidade

Os dados de sequenciamento da linhagem 9a5c de *X. fastidiosa* (site: aeg.lbi.ic.unicamp.br/xf), revelaram que de um total de 2904 ORFs (genômicas e de dois plasmídios nativos pXF51 e pXF1,3), 1314 (cerca de 46 %) possuem função biológica conhecida. Outros 318 apresentaram similaridade com proteínas hipotéticas conservadas e 1107 não apresentaram homólogos no banco de dados. De acordo com a anotação genômica, as ORFs foram divididas em 9 categorias funcionais, dentre as quais estão os processos de metabolismo celular, transporte, biossíntese de aminoácidos e nucleotídeos, patogenicidade, virulência e adaptação (Simpson *et al.*, 2000).

Para a seleção da ORF estudada considerou-se seu provável significado biológico no processo de patogenicidade; seu valor de similaridade com outros fatores de virulência (o maior *e-value*⁶ sendo 4e-28 (47%) com a proteína hipotética xrvA de *Xanthomonas oryzae*, ainda não elucidada); não ser proteína integral de membrana; ter baixa massa molecular para estudos de RMN; além de possuir domínio homólogo ao de ligação a DNA de *E. coli* (*e-value* de 3e-23, cerca de 13%) já elucidado por RMN.

A ORF XF0749 codifica uma proteína citoplasmática de 134 aminoácidos (15 kDa) de domínio similar ao de ligação a ácidos nucleicos da família H-NS (*Histone-like*

⁶ O *e-value* representa o número de seqüências aleatórias com um escore maior ou igual a "X". Este valor relaciona o escore ("X") de um alinhamento entre uma seqüência fornecida e uma seqüência depositada num banco de dados, com os escores maiores ou similares aos alinhamentos que seriam esperados de uma busca para uma seqüência aleatória de tamanho equivalente.

Nucleoid-Structuring Protein). Proteínas desta família organizam os ácidos nucléicos em estruturas nucleóides, sendo também responsáveis pela regulação transcricional de muitos genes não correlacionados, afetando a expressão gênica sob certas condições, seja pela ligação direta no DNA ou pela alteração de sua compactação (Azam *et al.*, 1999).

A proteína H-NS de *E. coli* possui pI neutro (pI 7,5) e é termo-resistente (Spassky *et al.*, 1984). É considerada incomum por conter em seus poucos 137 resíduos de aminoácidos pelo menos dois domínios, um N-terminal, responsável pela oligomerização e outro C-terminal, de ligação ao DNA (Fig. 4) (Ueguchi *et al.*, 1997). Dados genéticos mutacionais sugerem que sua oligomerização é importante no desempenho de sua função (Falconi *et al.*, 1988; Friedrich *et al.*, 1988; Ueguchi *et al.*, 1996; Spurio *et al.*, 1997; Ueguchi *et al.*, 1997). Predominantemente, proteínas H-NS são encontradas como dímeros, sendo a interação do tipo hidrofóbica, mas formam tetrâmeros em alta concentração celular (micromolar) (Williams & Rimsky, 1997).

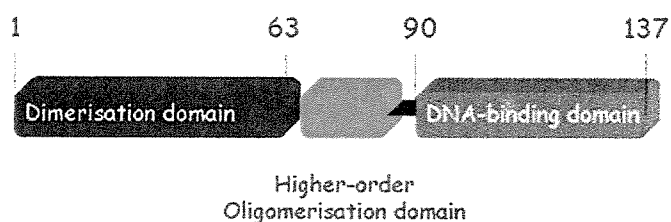


Fig. 4: Representação dos domínios N-terminal e C-terminal da proteína H-NS consenso de *E. coli*.

Apesar das propriedades intrigantes de proteínas H-NS, existe pouca informação a respeito de sua estrutura terciária. Devido a dificuldades de cristalização de proteínas ligantes a DNA, e da formação de oligômeros dificultando a análise por RMN, a única informação estrutural disponível é a do domínio C-terminal de ligação a DNA de *E. coli* por RMN. O mesmo domínio hipotético da ORF XF0749 de *X. fastidiosa* possui baixa



Protein A), Hfq (*Host Factor for Phage Q_β Replication*), IciA (*Inhibitor of Chromosome Initiation A*), StpA (*Suppressor of td Mutant Phenotype A*), Lrp (*Leucine-responsive Regulatory*) e Dps (*DNA-binding Protein from Starved Cells*).

Os complexos de nucleoproteínas são formados pela interação não-covalente da proteína com o DNA, resultando em uma estrutura dinâmica de organização adequada para responder rapidamente a mudanças ambientais quando a expressão gênica é modulada (Fig. 6) (Smyth *et al.*, 2000). Deste modo, têm-se a organização cromossomal em discretas estruturas dentro da célula (Ussery *et al.*, 1994; Smyth *et al.*, 2000).

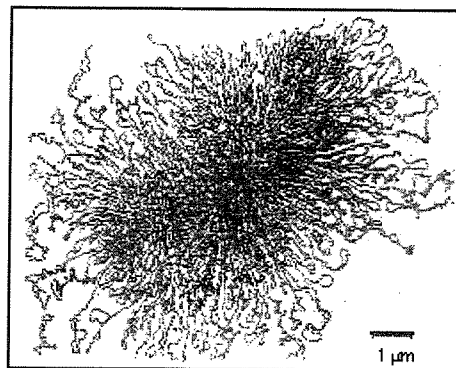


Fig. 6: Ilustração da estrutura nucleóide em procariotos.

Durante muitos anos não era claro se bactérias continham nucleossomos análogos aos das células eucarióticas. Contudo, muitas proteínas ligantes a DNA tem sido identificadas nestes microrganismos, as quais formam estruturas similares a cromatinas (Spassky *et al.*, 1984; Ussery *et al.*, 1994), com o nível de empacotamento e a concentração local de DNA (20-50 mg/ml) similar àquelas observadas numa intérfase nuclear eucariótica (Williams & Rimsky, 1997).

Uma das proteínas mais abundantes ligantes a DNA pertence à família H-NS, com cerca de 10^5 cópias por célula, correspondendo a uma molécula proteica para cada 400 pb de DNA cromossomal (Spassky *et al.*, 1984; Falconi *et al.*, 1988). A concentração celular

da proteína H-NS é dependente da fase de crescimento bacteriano, sendo sua expressão auto-regulada negativamente. A desrepressão do gene *hns*⁷ ocorre na fase estacionária ou em condições de choque-térmico na fase lag quando a temperatura abaixa de 37°C para 10°C (La Teana *et al.*, 1991; Ueguchi *et al.*, 1993). Em virtude da ação de forte compactação do DNA, a proteína H-NS em altas concentrações celulares é deletéria, tendo-se então, cuidado ao clonar sua ORF em plasmídios de alta multiplicidade e em expressá-la de maneira nativa em sistemas heterólogos (Spurio *et al.*, 1992).

Estudos mutacionais demonstraram que a proteína H-NS afeta a estrutura do DNA pela indução de forte compactação (Spassky *et al.*, 1984), pela alteração da super-helicidade (Tupper *et al.*, 1994) e pela indução de curvaturas (Spurio *et al.*, 1997). Ainda pela análise mutacional do gene *hns*, caracterizou-se seu efeito regulador na expressão de 35 genes envolvidos na adaptação bacteriana ao estresse ambiental, incluindo sua virulência ao hospedeiro (Bracco *et al.*, 1989; Yamada *et al.*, 1990; Owen-Huges *et al.*, 1992).

Baseado na falta de especificidade por um sítio específico de ligação, o mecanismo pelo qual a H-NS influencia a transcrição é provavelmente diferente dos típicos reguladores transcricionais de sequência específica. Estudos iniciais de atividade de ligação a ácidos nucleicos revelou sua habilidade em ligar a DNA e RNA, em escala decrescente de afinidade (Williams & Rimsky, 1997). Maior afinidade de ligação ocorre em regiões de DNA fita dupla curvadas, ricas em adenina e timina (Bracco *et al.*, 1989; Yamada *et al.*, 1990; Owen-Huges *et al.*, 1992; Smyth *et al.*, 2000).

⁷ Várias mutações, usando-se diferentes regimes de seleção, deram origem a diferentes designações do *locus* gênico de *hns*, por exemplo *osmZ*, *bglY*, *drdX*, *pilG* e *virR* (Ussery *et al.*, 1994).

Apesar dos extensos estudos mutacionais e de vários testes *in vitro*, o mecanismo completo de atuação de proteínas da família H-NS ainda não está completamente esclarecido. Neste sentido, proteínas H-NS de diversos microrganismos vem sendo estudadas⁸ na busca da elucidação de seu papel biológico. Contudo, dificuldades inerentes à proteínas H-NS, como curta vida útil, oligomerização e dificuldade de cristalização, tem dificultando o estudo pelas técnicas convencionais de cristalografia de raios-X e RMN. Acredita-se que tão logo estas dificuldades sejam solucionadas, a resolução de sua estrutura tridimensional aliada a estudos funcionais *in vivo*, fornecerão o entendimento global de seu mecanismo biológico.

⁸ A proteína H-NS apresenta conservação de resíduos em muitas enterobactérias (Bertin *et al.*, 1999): *Bordetella pertussis*, *E. coli*, *Erwinia chrysanthemi*, *H. influenzae*, *Proteus mirabilis*, *Salmonella typhimurium*, *Shigella flexneri*, *Rhodobacter capsultus*, *Vibrio cholerae*, *X. oryzae*, *Yersinia pestis*, etc.

4. ARTIGO 1

Expression and purification of a putative H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*

Débora Pires Paula,^a Adriano Rodrigues Azzoni,^a Susely Ferraz Siqueira,^a Cleide Ferreira Catani,^b Luciana Kauer Rosselli^a and Anete Pereira de Souza^a

^a *Genetic Engineering and Molecular Biology Center, Department of Genetic and Evolution, Institute of Biology, The State University of Campinas, C.P. 6010, Campinas – SP, Brazil*

^b *Department of Microbiology and Immunology, Institute of Biology, The State University of Campinas, C.P. 6010, Campinas – SP, Brazil*



Expression and purification of a putative H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*

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

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

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

Received 26 March 2003, and in revised form 12 June 2003

Abstract

The H-NS protein is one of the major constituents of the nucleoid structure that has been implicated in the DNA packaging and in the global regulation of gene expression. The study of this transcriptional regulator is an effort to fight *Xylella fastidiosa*, a citrus pathogen responsible for a range of economically important plant diseases, including the citrus variegated chlorosis (CVC). The putative H-NS ORF was cloned into a pET32-Xa/LIC vector in order to overexpress it coupled with fusion tags in *Escherichia coli* BL21(DE3). The expressed recombinant protein was purified by immobilized metal affinity chromatography (Ni-NTA resin) and its identity verified by mass spectrometry (MALDI-TOF). Final purification was performed by cation-exchange chromatography (SP Sepharose Fast Flow) and the purified protein was found as a single band on SDS-PAGE. The folding and its DNA binding activity were verified by circular dichroism and fluorescence spectroscopy, respectively.

© 2003 Elsevier Inc. All rights reserved.

Keywords: H-NS protein; *Xylella fastidiosa*; DNA-binding activity or functional genomic

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

From 147 pathogenicity predicted coding regions found by the *X. fastidiosa* genome project, the XF0749 ORF (gi 1360644) that codifies a protein of 134 amino acids (15 kDa), theoretical pI 9.33 and probable subcellular localization on the cytoplasm, was chosen. The putative function could be assigned via similarity with

characterized sequences in the databases. The comparative alignment revealed that the target protein has a common domain organization and composition of DNA binding proteins of 52 proteobacteria species. Some of these sequences are aligned in Fig. 1.

Twelve DNA binding proteins that form the nucleoprotein complexes responsible for global regulation in many crucial biological processes, such as replication, recombination, and transcription of 4000 non-correlated genes, either by binding directly to DNA or through changes in DNA topology, have been identified [5].

The codified protein from the XF0749 ORF seems to belong to the H-NS family whose primary structure is highly conserved in gram-negative bacteria [6]. H-NS is the most abundant protein of Enterobacteriaceae and has been implicated not only in the compact organization of the nucleoid structure, but also as a negative regulator of more than 35 unlinked gene expressions [5]. The DNA binding capacity is apparently non-specific, although H-NS protein preferentially binds to synthetic

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

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

		10	20	30	40	50	
		*..					

Fig. 1. Amino acid multiple-alignment sequence of the hypothetical protein from XF0749 *Xylella fastidiosa* (clone 9a5c) ORF. The sequences were retrieved from the National Library of Medicine NCBI Database. The consensus sequence is indicated above the aligned sequences and it refers to *E. coli* H-NS C-terminal domain. Conserved residues are in gray. Bold residues show conservative changes of residues in >50% of all sequences. Species abbreviations and accession numbers are as follows: Pm, *Proteus mirabilis* (gi 7341304); Haein, *Haemophilus influenzae* (gi 1170327); Rhoca, *Rhodobacter capsulatus* (gi 1170430); Bp, *Bordetella pertussis* (gi 2213665); Rs, *Rhodobacter sphaeroides* (gi 1835978); Ec, *E. coli* (gi 498871); xrvA, *Xanthomonas oryzae* (gi 1360644), and StpA_Ec, *E. coli* (gi 267038).

curved in vitro and AT-rich regions upstream of promoters [7], affecting the association of RNA polymerase and the formation of open complexes for transcription initiation [5]. H-NS functions in vivo as a repressor of its own transcription [8] and the derepression only occurs during cold shock [9] and when the culture enters the stationary phase (26,000 copies per cell against 5800 copies per cell at early logarithmic phase) [6].

Extensive genetic studies (mutational analyses) on H-NS suggested that it has at least three distinctive functional domains: transcriptional repression (relatively N-terminal), dimerization/oligomerization (relatively central), and DNA binding (relatively C-terminal) [10–13]. However, the knowledge about its global structure is very limited. The information available is the three-dimensional structure of a low resolution model of its C-terminal domain [14] and none of the several spontaneous *hns* alleles isolated so far have helped to shed light on the nature of the active sites of the protein [13]. This has led to efforts to combine functional studies (biochemical and cell biological) with three-dimensional structure determination to gain insights into the way the protein molecules work.

To address the relevant issue emphasized above, it is necessary to clarify the structural basis of the H-NS function. This work presents the expression, purification, and the DNA binding activity of a putative H-NS from *X. fastidiosa* in high purity and quantity, to further contribute to the clarification of the H-NS molecular mechanisms with the purpose of enabling advances on the CVC combat.

Materials and methods

Materials

Oligonucleotide primers were synthesized at Invitrogen Life Technologies (São Paulo, Brazil). The pET32-

Xa/LIC vector, the BL21(DE3) strain, the S-Tag rapid assay kit, and factor Xa protease were obtained from Novagen (Madison, WI). The Ni-NTA affinity resin was obtained from Qiagen (Hilden, Germany). The chromatography columns and the 1D and 2D protein markers were purchased from Bio-Rad (Hercules, CA). The molecular-mass marker (LMW), the IPG pH 6–11 gel strip, and the HiTrap Sulpho Propil Sepharose Fast Flow cation-exchange resin were purchased from Amersham-Pharmacia Biotech (Uppsala, Sweden). The protease inhibitors were purchased from Sigma Chemical (St. Louis, MO). The 0.45 m nitrocellulose membranes were obtained from Millipore (San Diego, CA). The dialyzed membranes of the molecular weight cutoff of 7 kDa were acquired from Pierce Biotechnology (Rockford, IL). All other reagents were of the highest commercially available grade.

Expression vector construction

The gene target from *X. fastidiosa* was amplified by polymerase chain reaction (PCR) using the purified genomic DNA as the template and cloned onto the expression vector pET32-Xa/LIC. A pair of PCR primers, 5'-GGTATTGAGGGTCGCATGAACGATACA-3' (sense) and 5'-AGAGGAGAGTTAGAGCCTCAGCCTGGG-3' (antisense), was designed in order to generate products with vector cohesive overhangs (in bold). The sense primer encodes the factor Xa recognition site and is in-frame with the open reading frame defined by the vector. The antisense primer encodes a stop codon. This vector is designed for expression of heterologous-protein fused to the 109 thioredoxin amino acids (to improve solubility), a six amino acid His-tag (to simplify the purification), and 15 amino acid S-tag (for detection and quantification) sequences upstream to the cloning site (Fig. 2). The fusion tags together have about 17 kDa and theoretical pI of 5.95, which can be removed by factor Xa cleavage [15].

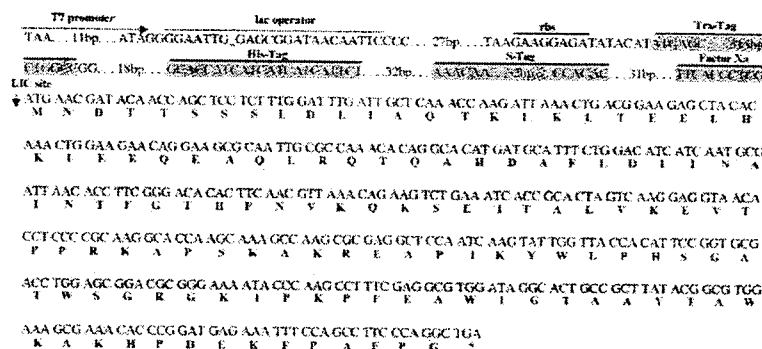


Fig. 2. Recombinant protein overexpressed from a pET32-Xa/LIC vector (Novagen): N-terminal fusion tags, factor Xa cleavage site, LIC site for insertion of heterologous-gene, and amino acid sequence deduced from XF0749 of *Xylella fastidiosa* (clone 9a5c) ORF. The C-terminal sequence is underlined.

The amplification protocol consisted of a 3 min denaturation at 94°C followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and 30 s and extension at 72°C for 2 min. The blunt PCR products were purified and treated with T4 DNA polymerase in the presence of the dGTP and DTT to generate the specific vector cohesive overhangs. The insert cloning in the linear vector was carried out according to the vector manufacturer protocol [15].

Escherichia coli DH5 α cells were transformed by recombinant vectors by the standard polyethyleneglycol (PEG) method [16] and plated overnight at 37°C on Luria–Bertani broth (LB) with 50 μ g/mL ampicillin. The colonies were individually stored in a permanent 2YT+HMFM broth at –70°C. The clones were checked by PCR and their sequences were verified by nucleotide sequencing. *E. coli* BL21(DE3) cells were transformed by positive recombinant vectors by the PEG method and used for protein expression.

Protein expression and purification

Escherichia coli BL21(DE3) cells transformed by recombinant vectors were cultured in 30 mL LB broth added to 50 μ g/mL ampicillin, grown overnight at 37°C and 300 rpm. This was then transferred to 3 L of LB broth with the same antibiotic. The culture was grown at 37°C and 300 rpm until OD₆₀₀ of 0.8, then induced with 1 mM IPTG for an additional 4 h. Cells were harvested by centrifugation at 2600g and 4°C for 10 min. Pelleted bacteria were suspended in 150 mL of 50 mM Tris–HCl buffer (pH 7.5) added to 300 mM NaCl (named buffer 1). Protease inhibitors: leupeptin, pepstatin A (both at a final concentration of 1 μ g/mL), and PMSF 1 mM, as well as 1 mg/mL of lysozyme were added. The suspension was incubated for 30 min at 4°C and sonicated 8 $\times$ 30 s at power 14 (70% cycle) in a Sonifier Misonix (Microson Ultrasonic Cell Disruptor XL). The lysed material was

clarified by centrifugation at 27,500g for 10 min at 4°C and the supernatant (soluble fraction) was collected to confirm recombinant protein expression by SDS–PAGE.

Clarified supernatant was passed per gravity through an IMAC (Immobilized Metal Affinity Chromatography) column packed with 1.0 mL of nickel-nitrilotriacetic acid (Ni-NTA) resin equilibrated with 10 column volumes (CV) of buffer 1. The resin was washed with 13 CV of a 50 mM Tris–HCl buffer (pH 7.5) added to a 1.0 M NaCl (buffer 2). Additional washing was performed with 13 CV from buffer 1 added to 5 mM of imidazole (buffer 3). Bound proteins were eluted with 6.6 CV from buffer 1 added to 200 mM imidazole (buffer 4). Fractions were used to estimate total protein concentration and analyzed by SDS–PAGE.

Fusion tags were separated from putative H-NS through factor Xa proteolysis (0.1 U/ μ L per 10 μ g recombinant protein). The reaction was performed at 15°C for 20 h in 50 mM Tris–HCl buffer (pH 7.8) added to 100 mM NaCl and 5 mM CaCl₂, corresponding to 10% of the total volume. The protease was inactivated by PMSF at 1 mM and the proteolysis products were analyzed by SDS–PAGE. The sample was dialyzed at 4°C against 20 mM NaPi buffer (pH 7.5) to prepare the sample for the following purification step.

The final purification was done by cation-exchange chromatography using a 1-mL Hitrap Sulpho Propil Sepharose Fast Flow in an ACTA–FPLC system. The sample was applied into the column equilibrated with 10 CV of 20 mM NaPi buffer (pH 7.5) (equilibration buffer). The unbound proteins were washed out with 10 CV of equilibration buffer and bound proteins were eluted using a linear NaCl gradient of 0–1.0 M in 10 CV of equilibration buffer. The flow rate used was 0.75 mL/min. The fractions were collected and analyzed by SDS–PAGE and quantified.

Protein analyses

The total protein concentration of the samples was estimated by S-Tag rapid assay kit [17] or according to Bradford assay [18]. The S-Tag fusion sequence from recombinant protein encodes a peptide that has a high affinity with the 104 amino acids S-protein, forming a complex with enzymatic activity which can be measured by reading the absorbance at 280 nm. More accurate estimates for purified H-NS were carried out based on UV absorbance spectroscopy using extinction coefficients from residues of tryptophan (Trp), tyrosine (Tyr), and cystine (disulfide bonds) [19]. Protein H-NS purity was monitored by SDS-PAGE performed using a 4% stacking gel and 13.5% separation gel under reducing conditions according to Laemmli [20] using Coomassie brilliant blue R-250 staining and destained with methanol/acetic acid.

Mass spectrometry

The recombinant H-NS mass spectrum analysis was carried out using matrix-assisted laser-desorption time-of-flight (MALDI-TOF) (2E, Micromass UK) on a Voyager DE-PRO MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA). The recombinant protein was deposited on a metal target as co-crystals with sinapic acid and the mass spectrum was determined in the positive ion mode. The masses of monoisotopic peaks with relative intensity over 2% of the most intense peak in the spectrum were used for *X. fastidiosa* (clone 9a5c) genome database searching using MS-Fit program [21].

Circular dichroism (CD) spectroscopy

The fully purified protein was used to assess the secondary structural integrity. Far-UV CD spectra were generated using a 1 mm pathlength cuvette containing 200 μ L H-NS protein sample at 54 μ M in 10 mM NaPi buffer (pH 7.0). The assays were carried out in a Jasco 810 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan). The sample was scanned eight times from 190 to 260 nm at 15 °C under a continuous nitrogen flush. The sample and buffer were filtered with a 0.45 μ m nitrocellulose membrane before the analysis. Molar ellipticity was calculated based on the number of amino acid residues (residual molar ellipticity). The composition of secondary structure was estimated using program CDNN [22] supplied with the instrument.

Analyzing the DNA binding activity by fluorescence spectroscopy

The 15 bp oligonucleotide duplex was prepared by resuspension of the purified single-stranded 5'-CA-

AAATATATTTTG-3' in 10 mM MgCl₂ added to 10 mM Tris-HCl (pH 7.6) (final volume to 50 μ L) followed by annealing at heating to 70 °C for 10 min and slowly cooling to room temperature. The duplex was purified by extraction with phenol:chloroform and precipitation with ethanol [16], and resuspended in 150 μ L of 1 mM NaPi buffer (pH 7.0). The concentration of duplex was determined using the calculated molar extinction coefficient from light absorption at 260 nm.

DNA binding studies were performed in 1 mM NaPi buffer (pH 7.0). Titrations were carried out by adding aliquots of concentrated protein stock solution to 1.5 ml containing a fixed concentration of 9 μ M of the 15 bp oligonucleotide duplex. Fluorescence emission spectra were measured on a ISSK2 Multifrequency Phase Fluorometer with a 300 W xenon arc lamp, using an excitation of 295 nm and emission spectra were recorded from 315 to 450 nm at 1 nm intervals with 1.2 s integration time. All experiments were performed using a 1 $\times$ 0.4 cm fluorescence cuvette at 15 °C and were corrected for buffer contribution and dilution. Corrections for the inner filter effect arising from DNA absorption were performed as described [23]. Spectra were recorded by averaging three scans.

Results and discussion

Expression and purification

The XF 0749 pathogenicity ORF was successfully cloned in the chosen vector and no difficulty was found in the transformation of the *E. coli* DH5 α and BL21(DE3) strains. The recombinant H-NS was expressed as soluble form and about 9 mg of protein per liter of culture was obtained (S-Tag quantification). No improvement in the expression level was verified when different expression strains, induction temperatures, and growth phase-induction were tested (data not shown). The expression of the recombinant protein as inclusion bodies was not observed.

Apparently, T7 promoter from the pET32-Xa/LIC vector might suffer transcriptional repression by the H-NS, even with fusion tags in the N-terminal extremity, because the DNA binding domain is in the free C-terminal portion. The H-NS inhibitory effect in *lacUV5*, *proV*, and in its own promoters as well as the non-specific DNA binding capacity in AT-rich regions upstream of another promoter is well known [24–26]. It has been demonstrated that the overproduction of the H-NS expressed from a high copy plasmid has a harmful effect on cell growth because of the extremal compaction of the nucleoid [26].

A high amount of the recombinant H-NS was recovered in the first purification step by IMAC. The

reduction of non-specific binding during the chromatography was done by washing the resin with high salt concentration and low imidazole concentration buffers (buffers 2 and 3, respectively). The elution of the recombinant protein was performed using 200 mM imidazole concentration. The SDS-PAGE analysis of the chromatography is shown in Fig. 3. The amount of recovered recombinant protein was estimated in 5 mg/L of culture and the analysis of the SDS-PAGE gel by optic densitometry in a GeneGenius Imager (Gateway, USA) indicated approximately 85% purity.

To separate the recombinant H-NS protein from N-terminal fusion tags, cleavage by factor Xa was performed. The fusion tags were successfully removed from H-NS and the proteolysis products could be observed on the SDS-PAGE as two new bands corresponding to their expected molecular masses (lane 5, Fig. 3). The H-NS remained stable without the fusion tags and there was no evidence of further and non-specific cleavage. Although the manufacturer recommended a cleavage temperature of 37 °C, the reaction was carried out at 15 °C to avoid any H-NS degradation.

A final ion-exchange chromatography step was required to remove the fusion tags and traces of impurities (including factor Xa), taking advantage of the high theoretical H-NS pI (9.3). The putative H-NS was eluted as a single peak around 550 mM NaCl. Lane 7 (Fig. 3) shows the single band of the purified H-NS after all purification steps. At this point, the total amount of pure putative H-NS obtained was approximately 3 mg, with an overall recovery of 24%. The recovery was calculated by taking into account that the recombinant protein is expressed as a fusion protein where 47% of the molecular mass corresponds to the recombinant H-NS.

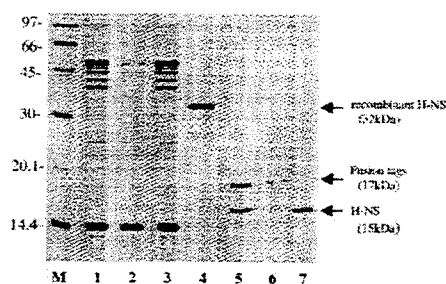


Fig. 3. SDS-PAGE (13.5% gel) of the expression, purification, and cleavage of putative H-NS from *Xylella fastidiosa*. Lane M, standard, 8.6 µg, (1) non-induced cells, 7.4 µg, (2) clarified supernatant after cells induction with IPTG, 9.3 µg, (3) flow-through from IMAC column, 8 µg, (4) recombinant H-NS eluted at 200 mM imidazole from IMAC column, 4.5 µg, (5) products after fusion tags cleavage, 3 µg, (6) flow-through from cation-exchange column, 1.1 µg, and (7) purified H-NS eluted within salt gradient from cation-exchange column, 1.5 µg.

Mass spectrometry and structural CD analysis

In order to verify the identity of the recombinant protein, it was analyzed by mass spectrometry. The recombinant protein spot of 2D gel showed that the protein had molecular mass of approximately 32.3 kDa and pI 7.0. These calculated pI and MW values found are in good agreement with those provided by the software Protein Parameters Tools (www.bm.expasy.org/tools/protparam.html) and with that apparent MW estimated for this protein band on the SDS-PAGE. The peptide profile was compared to the *X. fastidiosa* genome database and it was found that the highest similarity (43%) to the amino acid primary sequence corresponded to the protein studied. Since the protein studied still contained the fusion tags (17 kDa), the relative low similarity value was expected.

To investigate the structural integrity of the putative H-NS, the purified protein was studied by Far-UV CD spectroscopy. The CD spectrum analysis showed predominant signal of α -helices (peaks: positive at 190 nm and negative at 222 and 206 nm). The CD spectrum of unordered conformation, characterized by a strong negative CD signal just below 200 nm and a positive band at about 218 nm when the protein probably became denatured, could not be noted. The protein remained folded throughout the purification and it is suitable for functional analysis and further 3D structure determination.

The secondary-structure estimations on software CDNN [22] of α -helices, β -strands, β -sheets, β -turns, and unordered forms were 28.5, 13.3, 8, 19.6, and 29.7%, respectively. These estimates are not in agreement with secondary structure prediction of α -helices, β -strands, and β -turns of, respectively, 56.72, 43.28, and 0%, from the PredictProtein software (version 1.99.08) site <http://cubic.bio.columbia.edu> [27]. Some estimate and prediction errors may occur due to occasional difficulties in distinguishing between α -helices and β -strands [28].

Fluorescence analysis of the DNA binding activity

The putative DNA binding property was monitored by fluorescence based on tryptophanyl residues as a structural indicator [29]. Fluorescence spectroscopy provides a valuable technique for monitoring protein-DNA interactions [30] and is particularly useful for proteins containing four tryptophans, like the present protein. All the tryptophanyl residues are present in the C-terminal portion, which by primary sequence homology, is the DNA binding domain (residues 79–122) at positions 90, 98, 101, and 120. These studies demonstrate that the four tryptophanyl residues were affected upon DNA binding, as indicated in Fig. 4A. The intensity of the protein intrinsic fluorescence decreases after binding to 15 bp duplex. Probably, in the

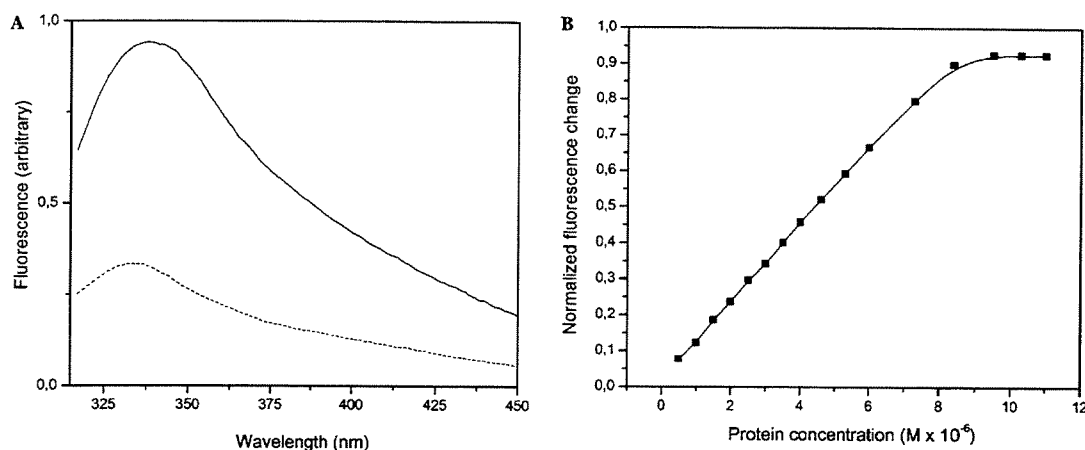


Fig. 4. Spectroscopic characterization of protein, DNA, and protein-DNA complex. (A) Fluorescence emission spectrum (excitation 295 nm) of free protein (solid line) and of protein saturated with DNA (dashed line). (B) Fluorescence titration curve. The curve was generated by subtracting fluorescence in the absence of DNA from that in the presence of DNA at each point during the titration of a concentrated stock solution of protein into a 9 μ M fixed concentration of 15 bp oligonucleotide duplex. The fitted curve corresponds to an equilibrium constant K_A of 1.2×10^6 M^{-1} . The experiments were recorded at 15 $^{\circ}C$ in 1 mM NaPi buffer (pH 7) and the data are plotted as the normalized change of the emission intensity.

DNA-bound state the exposure of the tryptophanyl residues to solvent decreases.

By monitoring the changes in fluorescence intensity during the titration of protein into fixed concentration of DNA, binding curves can be generated and the association constant (K_A) can be obtained by fitting the resulting curve. By subtracting such a curve from an identical fluorescence titration in the absence of DNA, a binding curve can be generated. The resulting experimental data are also shown in Fig. 4B. The best fit for the curve gives an equilibrium constant K_A of 1.2×10^6 M^{-1} at 15 $^{\circ}C$. This association constant value is of the same order of those obtained for other H-NS [30–33] and strongly suggests that the studied protein has DNA binding activity.

Conclusion

The host–pathogen interactions related to bacteria are dynamic and require the pathogen to adapt to a series of environments, many of these are hostile. To succeed, the bacteria brings into play a sophisticated network of gene-regulatory functions, most of which are focused on modulating the transcriptional profile of the cell. H-NS is a modulator through pleiotropic regulation of the genome stability, recombination, and gene expression.

The study presented here illustrates successful expression, purification, and DNA binding activity of a putative H-NS protein from the phytopathogen *X. fastidiosa*. The purity, identity, fold, and activity were verified by independent methods. The present method-

ology is suitable to supply high purity and quantity of functional H-NS, which is needed for three-dimensional structure determination. NMR studies are in progress in our laboratory in order to achieve a complete protein characterization and therefore to help in the knowledge of its biological role.

Acknowledgments

The authors thank FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 01/07533-7) and CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support. A.P.S. received a fellowship from CNPq. L.K.R. received Ms. and A.R.A., S.F.S., and C.F.C. received post-doc fellowships from FAPESP. D.P.P. received Ph.D. fellowship from CAPES. We also thank the Structural Molecular Biology Center researchers from the National Synchrotron Light Laboratory for technical expertise, particularly Ph.D. Nilson Ivo Tonin Zanchin and Ph.D. Thelma Aguiar Pertinhez (Biology Cellular and Molecular Laboratory); and researchers Ph.D. Tomomasa Yano, Ph.D. Yoko Bomura Rosato, and Ph.D. José Camillo Novello from the Institute of Biology, The State University of Campinas, for equipment supports.

References

- [1] M. Frome, A. Camargo, S. Heber, C. Czink, A.J.D. Simpson, J.D. Hoheisel, A.P. Souza, Mapping analysis of the *Xylella fastidiosa* genome, *Nucleic Acids Res.* 28 (2000) 3100–3104.

- [2] R.F. Lee, K.S. Derrick, M.J.G. Beretta, C.M. Chagas, V. Rosetti, Citrus variegated chlorosis: a new destructive disease of citrus in Brazil, *Citros Ind.* Oct. (1991) 12–15.
- [3] F.F. Laranjeira, Ten years of citrus variegated chlorosis: What do we know?, *Laranja* 18 (1997) 123–141.
- [4] A.J.G. Simpson et al., The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (2000) 151–157.
- [5] T.A. Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid, *J. Bacteriol.* 181 (1999) 6361–6370.
- [6] J.M. Lucht, P. Dersch, B. Kempf, E. Bremer, Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of *Escherichia coli*, *J. Biol. Chem.* 269 (1994) 6578–6586.
- [7] H. Yamada, S. Muramatsu, T. Mizuno, An *Escherichia coli* protein that preferentially binds to sharply curved DNA, *J. Biochem.* 108 (1990) 420–425.
- [8] C. Ueguchi, M. Kakeda, T. Mizuno, Autoregulatory expression of the *Escherichia coli* *hns* gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription, *Mol. Gen. Genet.* 236 (1993) 171–178.
- [9] A.L. Teana, A. Brandi, M. Falconi, R. Spurio, C.L. Pon, C.O. Gualerzi, Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS, *Proc. Natl. Acad. Sci. USA* 88 (1991) 10907–10911.
- [10] C. Ueguchi, T. Suzuki, T. Yoshida, K. Tanaka, T. Mizuno, Systematic mutational analysis revealing the functional domain organization of *Escherichia coli* nucleoid protein H-NS, *J. Mol. Biol.* 263 (1996) 149–162.
- [11] P. Bertin, N. Benhabiles, E. Krin, Ch. Laurent-Winter, E. Turlin, A. Thomas, A. Danchin, R. Brasseur, The structural and functional organization of H-NS-like proteins is evolutionarily conserved in gram-negative bacteria, *Mol. Microbiol.* 31 (1999) 319–329.
- [12] C.J. Dormann, J.C.D. Hinton, A. Free, Domain organization and oligomerization among H-NS-like associated nucleoid-associated proteins in bacteria, *Trends Microbiol.* 7 (1999) 124–128.
- [13] C. Ueguchi, C. Seto, T. Suzuki, T. Mizuno, Clarification of the dimerization domain and its functional significance for the *Escherichia coli* nucleoid protein H-NS, *J. Mol. Biol.* 274 (1997) 145–151.
- [14] H. Shindo, T. Iwaki, R. Ieda, H. Kurumizada, C. Ueguchi, T. Mizuno, S. Morikawa, H. Nakamura, H. Kuboniwa, Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*, *FEBS Lett.* 360 (1995) 125–131.
- [15] Xa/LIC Vector Kits Manual, Novagen, ninth ed., USA, 2000, 56 pp (www.novagen.com).
- [16] J. Sambrook, D.W. Russell, Plasmids and their usefulness in molecular cloning, *Molecular Cloning, A Laboratory Manual*, vol. 1, third ed., CSHL Press, New York, 2001, pp. 1.32–1.34.
- [17] F.M. Richards, H.W. Wyckoff, in: P.D. Boyer (Ed.), *The Enzymes*, vol. IV, Academic Press, New York, 1971, pp. 647–806.
- [18] M.M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [19] C.N. Pace, F.X. Schmidt, How to determine the molar absorbance coefficient of a protein, in: T.E. Creighton (Ed.), *Protein Structure, A Practical Approach*, second ed., IRL Press, New York, 1997, pp. 253–259, 383 pp.
- [20] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227 (1970) 680–685.
- [21] K.R. Clauser, P.R. Baker, A.L. Burlingame, Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching, *Anal. Chem.* 71 (1999) 2871–2882.
- [22] N. Sreerama, Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy, *Protein Sci.* 8 (1999) 370–380.
- [23] N.C. Price, C.M. Johnson, Proteinases as probes of conformation of soluble proteins. *Proteolytic Enzymes: A Practical Approach*, IRL Press, Oxford, 1989 (Chapter 9).
- [24] M. Goransson, B. Sonden, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, B.E. Uhlin, Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*, *Nature* 344 (1990) 682–685.
- [25] C. Ueguchi, T. Mizuno, The *Escherichia coli* nucleoid protein H-NS functions directly as a transcriptional repressor, *EMBO J.* 12 (1993) 1039–1046.
- [26] R. Spurio, M. Falconi, A. Brandi, C.L. Pon, C.O. Gualerzi, The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending, *EMBO J.* 16 (1997) 1795–1805.
- [27] B. Rost, PredictProtein, in: *Methods in Enzymology*, vol. 266, Academic Press, New York, 1996, pp. 525–539.
- [28] C. Branden, J. Tooze, *Introduction to Protein Structure*, second ed., Garland Publishing, New York, 1999, p. 410.
- [29] J.B.A. Ross, A.G. Szabo, C.W.V. Hogue, Enhancement of protein spectra with tryptophan analogs: fluorescence spectroscopy of protein–protein and protein–nucleic acid interactions, in: *Methods in Enzymology*, vol. 278, Academic Press, New York, 1997, pp. 151–189.
- [30] A. Chant, X. Provatoopoulou, I.W. Manfield, G.G. Kneale, Structural and functional characterization of the DNA binding domain of the *Aspergillus nidulans* gene regulatory protein AreA, *Biochim. Biophys. Acta* 36866 (2003) 1–6.
- [31] K. Friedrich, C.O. Gualerzi, M. Lammi, M.A. Losso, C.L. Pon, Proteins from the prokaryotic nucleoid, *FEBS Lett.* 229 (1988) 197–202.
- [32] S. Milev, A.A. Gorfe, A. Karshikoff, R.T. Clubb, H.R. Bosshard, I. Jelesarov, Energetics of sequence-specific protein–DNA association: binding of integrase Tn916 to its target DNA, *Biochemistry* 42 (2003) 3481–3491.
- [33] W. Yang, W. Zeng, D. Zhou, Y. Shi, Cloning expression, secondary structure characterization of HMG box 1 of hUBF from *E. coli* and its binding to DNA, *Biochim. Biophys. Acta* 1598 (2002) 147–155.

5. ARTIGO 2

Functional and secondary structure characterization of a H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*: a new case of functional genomic

Débora Pires Paula,^a Adriano Rodrigues Azzoni,^a Thelma de Aguiar Pertinhez,^b Alberto Spisni^{b,c} and Anete Pereira de Souza^{a,*}

^a *Genetic Engineering and Molecular Biology Center, Department of Genetic and Evolution, Institute of Biology, The State University of Campinas, C.P. 6010, Zip code 13083-970, Campinas – SP, Brazil*

^b *Center for Molecular and Structural Biology, National Synchrotron Light Laboratory, C.P. 6192, Zip code 13084-971, Campinas – SP, Brazil*

^c *Department of Experimental Medicine, Sec. Chemistry and Structural Biology, University of Parma, Via Volturno, 39, Zip code 43100, Parma - Italy*

* Corresponding author. Tel.: + 55-19-37881132. Fax: 55-19-37881089.

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

Keywords: Functional genomic; *Xylella fastidiosa*; H-NS; Secondary structure; Dicroism circular; Fluorescence.

Running head: Functional and structural characterization of H-NS from *Xylella fastidiosa*.

Submetido para publicação na revista *Biochimica et Biophysica Acta* (Section: Protein and Proteomics).

Abstract

H-NS is a major component of the bacterial nucleoid, involved in condensing and packaging DNA and modulating gene expression. From the phytopathogen *Xylella fastidiosa* the XF0749 ORF by primary sequence similarity presumably encodes an H-NS type protein, which was over-expressed and purified and then used for functional and structural studies in order to validate the genomic annotation from recent Brazilian sequencing data. The binding of putative H-NS to a 14 base pair oligonucleotide was studied by the spectroscopic techniques of circular dichroism and fluorescence. The secondary structure was analyzed by thermal and chemical denaturant unfolding of the protein also by the same spectroscopic techniques. Preliminary backbone assignment derived from 2D ^1H ^{15}N -HSQC NMR experiments is discussed in this work, as well as the investigation of the quaternary structure by dynamic light scattering, non-denaturing PAGE electrophoresis and circular dichroism. Overall, the data strongly suggest that the XF0749 protein belong to H-NS protein family. This finding might help to further understand the pathogenicity mechanisms from *X. fastidiosa* and thus design new and more specific molecules to combat the Citrus Variegated Chlorosis, a serious Brazilian citrus disease.

1. Introduction

Both eukaryotic and prokaryotic organisms are faced with the problem of condensing a large quantity of genetic material into a nucleus, or cell, whilst retaining biological activity. Given that chromosomal replication occurs throughout most of the cell cycle, that about 20% of genes are transcribed at any one time, these genes being dispersed in the chromosome and which can be switched on or off almost instantaneously, chromosomal packaging would appear to be an intractable problem [1]. Non-covalent binding of proteins to DNA forms a dynamic structure and organization able to respond rapidly to changes in environmental conditions when gene expression is modulated [2]. From this interaction a structure called nucleoid is formed, which plays a role in organizing the chromosomal DNA into a discrete structure within the cell [1,2]. It has been clear for many years that bacteria do not have nucleosomes analogous to those of eukaryotic cells. However, several abundant DNA-binding proteins have been identified which are believed to form chromatin-like structures [1,3].

The H-NS is a nucleoid-associated protein that together with HU (Heat-Unstable nucleoid protein) is one the major constituent of the nucleoid. This protein has been implicated not only in the compact organization of the nucleoid structure, but also as a global regulator of gene expression [4]. In particular, H-NS negatively regulates the transcription of a number of apparently unlinked genes in the chromosome (about 30), including some virulent genes [5-9]. However, its mode of action appears to be uniquely different from those of many other well-documented sequence-specific transcriptional repressors, because H-NS is a very abundant (10^5 copies.cell⁻¹) and apparently non-specific

DNA-binding protein, giving preference to intrinsic A-T rich and curvature sequences [10-13].

Most of the H-NS protein from different species is composed of 137 amino acids ($M_r \sim 15.4$ kDa) and is heat-stable with pI about 7.5 (although it migrates in the pI 5.6 under denaturing conditions) [14]. Only recently, a heat unstable H-NS form has been identified in the antarctic *Psychrobacter* spp. [15]. The protein binding to DNA occurs at least in part, via the C-terminal fragment of the molecule (last 47 residues) which is able to interact with DNA as an isolated domain, albeit very weakly [2,16]. The interaction of H-NS with DNA is dependent on the formation of high order structures, generated by protein oligomerization by means of strong hydrophobic interaction [17], a process which was shown to influence the interaction of this protein with DNA and is attributive to N-terminal fragment (first 80 residues) with coiled-coil domain property [4,18]. However, a well defined portion is not yet known to be responsible for that process [19].

The primary sequence of the putative *X. fastidiosa* (9a5c clone) H-NS from XF0749 ORF (134 amino acids) has already been elucidated, over-expressed, purified and analyzed by MALDI-TOF [20]. Here we report data on the structural and functional features of XF0749 protein that demonstrate indeed it is an H-NS protein. This kind of work follows a worldly context of the emerging functional genomic, where the function is studied from its genic sequence. The insight gained from such analysis may be useful to further understand the pathogenical mechanisms of this phytopathogen, and consequently, to contribute to combat a serious Brazilian citrus disease, the Citrus Variegated Chlorosis.

2. Materials and methods

2.1. Protein expression and purification

The protein from *X. fastidiosa* (XF0749 ORF) was over-expressed and purified as described by [20], utilizing *E. coli* BL21(DE3)pLysS strain transformed by recombinant pET32-Xa/LIC (Novagen, Madison-WI) and purified by Immobilized Metal Affinity Chromatography of nickel-nitrilotriacetic acid (Ni-NTA) following cation-exchange chromatography in Sulpho Propil Sepharose Fast Flow after fusion protein cleavage. The size and homogeneity (> 95%) of the purified protein was confirmed by denaturing SDS-PAGE and staining with Coomassie brilliant blue R-250. The identity and molecular mass was confirmed using MALDI-TOF mass spectrometry (Micromass, UK) and the folding was confirmed by Far-UV CD spectroscopy.

2.2. Preparation of oligonucleotide duplex

Single-stranded purified oligonucleotides were synthesized by MWGInc Biotech (USA). For 14 bp oligonucleotide preparation, the oligo 5'-CAAAATATATTTTG-3' was resuspended in 10 mM magnesium chloride added to 10 mM Tris-clorate (pH 7.6) (final volume to 50 µl) and annealed by heating it up to 70°C for 10 min and slowly cooling it to room temperature. The duplex was precipitated with 3 M sodium acetate (pH 5.8) and ethanol 100% [21], and resuspended in 20µl of 1 mM sodium phosphate buffer (pH 7.0). The concentration of duplex was determined using the calculated molar extinction coefficient from light absorption at 260 nm. This 14 bp oligonucleotide was used for all DNA binding experiments.

2.3. Circular Dichroism Measurements

The XF0749 protein was extensively dialyzed against 1 mM sodium phosphate buffer (pH 7.0) prior to analysis. All measurements were carried out on a Jasco 810 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan) equipped with a Peltier Jasco PFD-425S for temperature control. One millimeter or 10 mm quartz cells were used. Spectra were collected from 195-260 nm, unless otherwise specified, and averaged over four scans at 20°C, a 0.5 nm step resolution, 50 nm.min⁻¹ scan rate and 1 nm bandwidth, were generally used. The CD data were expressed in degree cm².dmol⁻¹ per mean residue, based on a molecular mass of 15,036.1 Da calculated for this protein of 134 amino acids.

The pH titration curves were obtained by following the elliptical change in range 3 to 11.5. The protein stability curves were obtained upon continuous heating (1°C.min⁻¹) or cooling in the temperature range of 5°C-85°C. The reversibility chance to unfolding was also evaluated by repeated scans. The data are presented as the change of the molar ellipticity per residue at 222 nm and they were analyzed as described by [22]. In both experiments, pH titration and thermal stability, the protein concentration was 15 µM.

The protein-DNA complex was formed by mixing 30 µM of protein to 15 µM of 14 bp oligonucleotide described previously. The complex was incubated in a 1 mM NaPi buffer (pH 7.0) at 25°C for 30 min prior to collecting the data.

The CD spectra in function of protein concentration were recorded using different path-length quartz cells, 1 or 5 mm, depending on the protein concentration. The estimation of secondary structure content was performed using a DICROPROT program (2.6 version) [23].

2.4. Fluorescence Spectroscopy

Fluorescence emission spectra were measured on a Hitachi F2000 Fluorometer (300 W xenon arc lamp) using an excitation at 295 nm and emission spectra were recorded from 310 to 400 nm at 1 nm intervals with 1.2 s integration time. All experiments were performed using a 1.5 ml fluorescence quartz cell in 1 mM sodium phosphate buffer (pH 7.0) at 25°C. The intensity values were normalized for the dilution factor. The four tryptophans at positions 90, 98, 101 and 120, were used as a structural indicator, all of them presumably by primary sequence homology present in the C-terminal DNA binding domain (residues prediction 79 to 122).

For binding experiments, 1-10 μ M 14 bp oligonucleotide was titrated in 4 μ M of XF0749 protein solution. The mixture was incubated at 25°C in for 30 min prior to measurement. Corrections for the inner filter effect arising from DNA absorption were performed as described [22].

Denaturing by urea was monitored by the fluorescence emission maximum, after 8 μ M protein were allowed to reach equilibrium for 1 h with the denaturant concentration from 0 to 8 M urea. The fluorescence spectra were recorded as above. Corrections were made for the buffer contribution for each urea concentration.

Quenching titration was carried out by adding increasing amounts of a 5 M stock solution of acrylamide to 8 μ M of protein solution up to 400 mM acrylamide. The signal of reference containing buffer and quencher was subtracted from each point. The classical Stern-Volmer equation relates the drop in fluorescence to the concentration of a collisional quencher as shown in Equation 1 [24]:

$$\frac{F_0}{F} = 1 + K_{sv} [Q] \quad (\text{Eq. 1})$$

where F_0 is the fluorescence in the absence and F is the fluorescence in the presence of the quencher (Q). K_{sv} is the Stern-Volmer constant for the collisional quenching process. According to the fluorescence quenching data obtained from acrylamide quenching, the following modified Stern-Volmer equation was used to estimate the relevant quenching constants, as described by [25]:

$$\frac{F_0}{F} = (1 + K_{sv} [Q]) e^{V[Q]} \quad (\text{Eq. 2})$$

where K_{sv} is the dynamic quenching constant and V is the additional static quenching constant. To determine K_{sv} and V , the collected data were fitted to Equation 2.

2.5. NMR experiments

The uniformly ^{15}N -labeled putative H-NS was obtained from over-expression cultures in minimal M9/glycerol medium with $^{15}\text{NH}_4\text{Cl}$ (99%; Isotec, Inc.) as the unique nitrogen source. The purified-labeled protein was dialysed into 10 mM ammonium bicarbonate solution (pH 6.5) and concentrated by lyophilization. Ammonium bicarbonate is a volatile salt and evaporates after lyophilization, but its presence during the dialysis process maintains the ionic strength, preventing protein precipitation. Approximately 3 mg of protein (about 500 μM) was resuspended in 450 μl of 90% H_2O /10% D_2O and 1.0 mM DDT (pH 6.5). The 2D ^1H ^{15}N -HSQC (Heteronuclear Single Quantum Coherence) NMR spectra experiments were performed at 25°C.

All NMR spectra were recorded on a Varian Inova AS 600, operating at 599.683 MHz spectrometer at 14.1 Tesla for ^1H frequency, in the phase sensitive mode. Sequential ^{15}N -HSQC were collected at various times and the disappearing peaks were followed. The NMR data were processed using the nmrPIPE/nmrVIEW software [26,27].

2.6. *Dynamic Light Scattering*

Dynamic light scattering was carried out using a DynaPro-LSR equipment, with a temperature controlled micro sampler at 25°C. The sample concentration was 1 mg/ml on 40 mM sodium phosphate buffer (pH 7.5) added to 100 mM sodium chloride.

2.7. *Non-denaturing PAGE electrophoresis*

The XF0749 protein was dialyzed in a 40 mM HEPES buffer (pH 8.0) and thus incubated in the same HEPES buffer in three different compositions: 1) without salt (buffer A); 2) Added to 100 mM sodium chloride (buffer B); 3) Added to 100 mM sodium chloride, 60 mM potassium chloride and 60 mM ammonium chloride (buffer C). After 30 min incubation at room temperature, the 10 μl mixture protein (30 μM) it was by addition of the 10 μl loading buffer of 100 mM Tris-clorate (pH 6.8) adding 0.1% bromophenol blue (w/v) and 20% glycerol (v/v). The final protein concentration in total volume of 20 μl was 15 μM . The sample was then loaded onto a non-denaturing 13.5% polyacrylamide gel. Electrophoresis was carried out in a run buffer composed of 25 mM Tris-clorate (pH 8.3) added to 190 mM glycine and 1 mM EDTA, using a 14 x 16 cm x 1 mm Hoefer apparatus (Scientific Instruments, San Francisco, CA, USA). The gel was stained with Coomassie

brilliant blue R-250 and destained with a 30% methanol-10% acetic acid solution before being photographed.

To test the effect of the DNA on the protein self-association, 7.5 μM of the 14 bp oligonucleotide was incubated with 15 μM of the target protein at 30 min at room temperature in the same A, B and C buffers and loaded onto the non-denaturing PAGE.

3. Results and discussion

In the first part of this section, we describe functional analysis of the protein by circular dichroism and fluorescence spectroscopy. The second part concerning the secondary structure characterization by thermal and by chemical denaturants unfolding by the same spectroscopy techniques mentioned previously. The last part is relative to the quaternary structure study by dynamic light scattering, non-denaturing PAGE electrophoresis and circular dichroism.

3.1. Functional analysis of the DNA binding activity

Circular Dichroism spectroscopy provides a valuable technique for monitoring protein-DNA interactions. The CD spectra of protein-nucleic acid complexes may be divided into two spectral regions: the region above 250 nm where only the nucleic acid and protein aromatic residues contribute, and the region below 250 nm where the protein secondary structure can be indicated by content of α -helix, β -sheet, turns and loops [28].

The CD signature of B-form DNA as read from longer to shorter wavelength (320 to 200 nm) is a positive peak centered at 275 nm, a negative band about 240 nm, with zero being around 258 nm [28] (Fig. 1 A, line black). The XF0749 protein CD spectra in 1 mM

sodium phosphate buffer (pH 7.0) showed two negative peaks (208 and 222 nm) related to a predominant signal of α -helix (Fig. 1 A, line red). This secondary structure shape is similar to what observed for H-NS protein from other species, the peaks of which have different intensities. When a protein and a DNA are mixed but do not interact with each other, is expected a CD spectrum just like the sum of the signal from both macromolecules. However, when the macromolecules do interact, the CD spectrum of the mixture differs from the theoretical sum of the protein and the DNA spectra, indicating that there were conformational changes in the CD spectra [28].

The difference between the CD spectrum of the protein-DNA mixture and the profile obtained by the convolution of the spectra of the single components, Fig. 1 A (lines green and blue), clearly indicates that the protein does bind to the 14 bp oligonucleotide. The protein in the DNA-bound state presented a secondary structure conformation more stabilized, this is confirmed both by the perfect α -helical CD spectra where the negative bands have the same intensity (Fig. 1 A, line blue) and by the increasing of the melting temperature⁹ (T_m) from 52°C, in the free form, to 62°C in complexed form (Fig. 1 B). In the complexed form, the protein became more stable to thermal denaturation. This T_m found for free putative *X. fastidiosa* H-NS is in good agreement with those H-NS from other species, like 57°C from *E. coli* H-NS [25] and 58°C from *Salmonella typhimurium* [2].

Figure 1 C reports the CD spectra for the free protein showing that at 85°C the protein is not unfolded yet, and, after cooling, it reverts to a stable characterized by a CD

⁹ T_m is the temperature at which 50% of the protein is inactivated during reversible heat denaturation.

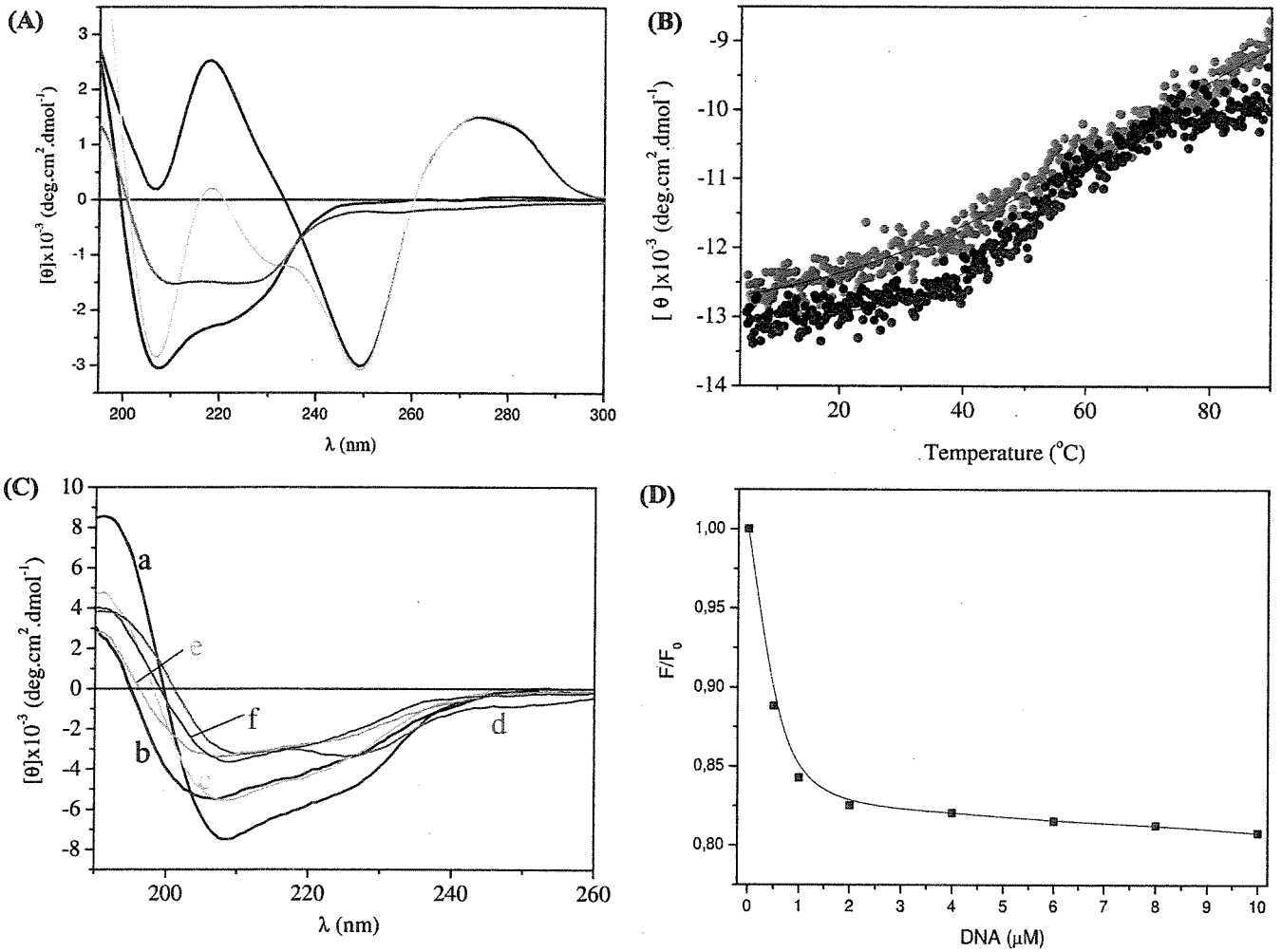


Fig. 1. Spectroscopic characterization of protein, DNA and protein-DNA complex. **(A)** CD spectra of protein (red line), DNA (black line), 2:1 protein-DNA complex (blue line), and calculated sum of spectra of free protein and free DNA (green line). **(B)** Thermal unfolding curves: red symbols represent the increased in temperature in free protein and blue symbols refer to complexed protein. The changes in molar ellipticity as a function of temperature measured at 222 nm. **(C)** Difference spectra at 5°C and 85°C: *a.* free protein at 5°C (black line); *b.* free protein at 85°C (red line); *c.* free protein reversibility for 5°C (green line); *d.* complex at 5°C (blue line); *e.* complex at 85°C (olive line); *f.* complex reversibility for 5°C (magenta line). **(D)** Protein (4 μM) fluorescence ratio with titration of the 14 bp oligonucleotide.

spectrum very similar to the one obtained for the native state, though less intense. Similarly, when the protein is complexed with the 14 bp oligonucleotide upon reaching 85°C, the CD spectrum turns out to be similar to the one of the free protein at the end of the

temperature cycle. The reversibility to the complexed form apparently is much greater than to the free protein.

The presence of tryptophan intrinsic probes makes the XF0749 protein a candidate for fluorescence studies and so was also utilized to monitor binding protein to an oligonucleotide duplex. Addition of 14 bp oligonucleotide to XF0749 protein decreased the intensity of the protein intrinsic fluorescence, with inducing a blue shift of the emission maximum [20]. Figure 1 D shows the behavior of the F/F_0 ratio upon titration with the 14 bp oligonucleotide. Saturation is reached at approximately a protein/DNA molar ratio of 2.

Values reported for the stoichiometry of H-NS binding to DNA have varied from 1 dimer of H-NS per 160 bp DNA [14] to 1 dimer per 12 bp [29]. These differences may reflect experimental methods and conditions, but probably also reflect the unusual nature of H-NS binding: binding is highly sensitive to factors such as ionic strength; the dissociation time for the H-NS-DNA complex is relatively short [14]; at various DNA-H-NS ratios different complexes appear to form [14,30], and at least under some conditions, there may be cooperative binding.

3.2. Secondary structure characterization

The pH dependence of the CD spectra of the XF0749 protein in a wide pH range are reported in Fig. 2 A. The protein is reasonably stable in the pH interval 5-10.5 where only a slight increase of the ellipticity is observed on going from pH 5 to 7.6. This conformational change could be attributed to the deprotonation of the 5 His residues distributed through the primary sequence. Interestingly a more relevant change is observed at pH 3 and 11.5 where either all residues are protonated or all ionized. In both cases we observed a shift of the

negative band from 208 to 205 nm suggesting destabilization of the protein secondary structure. The total number of negatively charged residues (Asp and Glu), about 15, is close to the total number of positively charged residues (Arg and Lys), which is 19. The protonation and ionization balance could prevent a dramatic alteration of the hydrogen bonds and salt bridge patterns which possibly would result in the protein denaturing face up to pH extremes.

The behavior of the denaturant-induced unfolding curves describes a simple two-state transition between the folded and unfolded states. Such study may provide information about the properties of partially folded intermediate states, and therefore be relevant in the understanding of the pathway of folding [31].

Plotting the wavelength of the emission maximum peak as a function of urea concentration shows the transition with a midpoint at around 3.5 M urea (Fig. 2 B). The fluorescence emission spectra excited at 295 nm has a peak centered at 342 nm, however between 1 and 7 M urea, there is a blue shift in maximum emission to 355 nm and an increase of the fluorescence intensity emission (data not shown). Probably the tryptophans moved from an apolar environment in the folded protein to a polar environment in the unfolded protein. Protein fluorescence spectroscopy showed that the DNA binding domain was folded and could be denatured by high concentrations of urea (about 7 M), exhibiting a sharp transition.

To provide information concerning the accessibility of the tryptophans, the sensitivity of the intrinsic fluorescence was tested by the addition of an electrically neutral and strong widely used quencher, acrylamide. Comparison with quenching data on other proteins helps to determine the degree of exposure of the fluorophor.

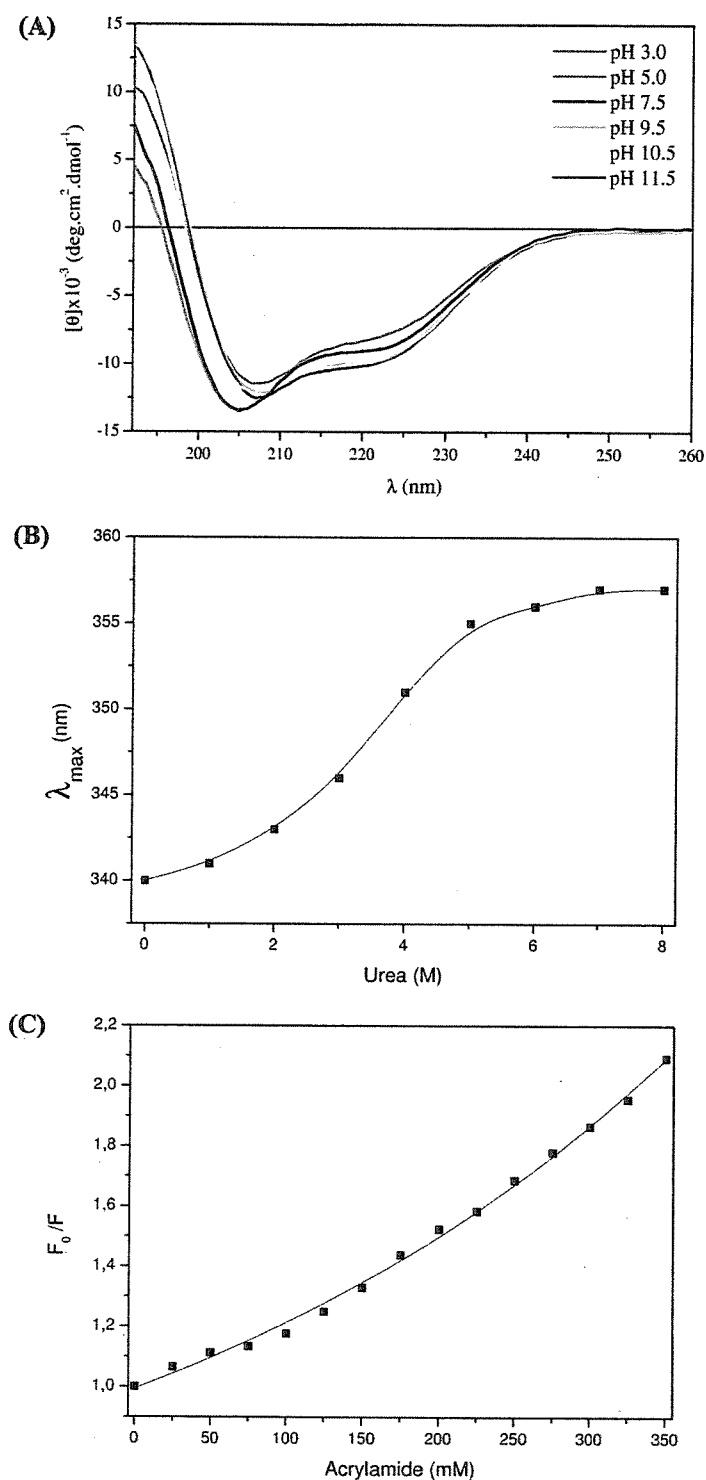


Fig. 2. Monitoring of the spectra of the XF0749 protein at 15 μM in 1 mM sodium phosphate buffer (pH 7.5) at 25°C, facing to different denaturants conditions: **(A)** CD spectra in pH range; **(B)** Maximum emission wavelength protein fluorescence upon urea titration; **(C)** Stern-Volmer plot of the acrylamide quenching of the fluorescence emission.

The Stern-Volmer plots of the collected acrylamide quenching data followed an upward function (Fig. 2 C), which could be fitted according to Equation 2 (described in the Materials and methods section). The acrylamide quenching data yielded apparent static and dynamic quenching constants of $V = 2.1 \text{ M}^{-1}$ and $K_{SV} = 3.1 \times 10^{-3} \text{ M}^{-1}$, respectively. Such magnitude of static and dynamic acrylamide quenching constants obtained indicate that the microenvironment of the putative H-NS tryptophans at positions 90, 98, 101 and 120 are relatively accessible in the apolar microenvironment. When treated with acrylamide, the fluorescence intensity dropped and in the concentration of 200 mM, the intensity declinment was 34%. Moreover, the increase in the fluorescence under urea denaturation indicated that the tryptophans are not external. However, the four tryptophans can be distinctly accessible for quenching.

3.3. Analyzing the tertiary and quaternary structure

Nuclear Magnetic Resonance (NMR) spectra of full-length H-NS showed a limited number of observable peaks, consistent with a significant line broadening effect causing a large number of peaks to become invisible (data not shown). This effect usually results from large molecules for which the tumbling times are extended or when the molecule is partially unfolded. However, XF0749 protein has only 134 residues and thus, the first hypothesis is only feasible if the protein had been in a self-associated state. The dispersed signal spectra corresponding to the four tryptophans, present only in the C-terminal region, allowed the inference that these dispersed spectra regions correspond to a possible C-terminal portion, suggesting that the C-terminus is highly mobile, whereas the N-terminal region in homologous exists in a self-associated state. Within the context of the intact

protein, this can only occur if these domains are separated by a highly flexible amino acid linker.

These findings are in full agreement with those of the 2D ^1H , ^{15}N -HSQC NMR spectra from the homologous ($3.8\text{e-}2$ *e-value*) H-NS from *Salmonella typhimurium* [2], suggesting that the same phenomena may be occurring to both H-NS proteins. The authors studied three truncated forms (two N-terminally and one C-terminally) of the *Salmonella typhimurium* H-NS and it was found that the C-terminal domain present the same tertiary structure as in isolation or in intact protein. Besides, the first 88 residues from *Salmonella typhimurium* H-NS do not give rise to significant NMR resonance.

The NMR result for the XF0749 protein is also in good agreement with the prediction of domain by program SMART (Simple Modular Architecture Research Tool: <http://smart.embl-heidelberg.de/>) that infer at least two domains for the studied protein: one N-terminal (coiled-coil, residues 8 to 39) and another C-terminal (DNA-binding, residues 79 to 122). The N-terminal region is a domain of 31 amino acids that displays features to assume a coiled-coil conformation. The C-terminal domain consists of 43 amino acids (79-122) and is similar (about 13%) to the C-terminal domain of 47 residues from *E. coli* H-NS, which had the structure solved by NMR [32]. In the homologues, the N-terminal domain *per se* did not contain structural information, but structure predictions suggest that it probably would be a coiled-coil (or superhelix) domain, responsible for oligomerization [33,34].

In this context, we submitted the primary sequence of the studied protein to the following programs: NNPredict (www.cmpchem.ucsf.edu/cgi-bin/nnpredict.pl), COILS Prediction [35] and Helical Wheel Representation from O'Neil and Grisham, Java Applet

(*cti.itc.Virginia.EDU/~cmg/Demo/wheel/wheelApp.html*). Computer-based structural prediction programs [36] showed that the N-terminal have a tendency to adopt a coiled-coil motif (Fig. 3 A), starting with Leu in position 9. The presence of four heptad repeats in N-terminal region and sequence similarity (22-29%) to well-known coiled-coil proteins, such as myosin, keratin, and dystrophin [1], led us to believe that this protein could be mediated by interactions between the α -helices. This coiled-coil conformation presumably provides the structural basis for self-association.

The XF0749 protein contains a larger number of charged residues, like most other DNA-binding proteins and eukaryotic histones, these are predominantly basic rather than acidic. It is not known which residues contribute to DNA binding, although there is a stretch of basic residues toward the C-terminus that might be involved in DNA binding. The conserved basic residues, possibly involved in DNA binding, are presented in Fig. 3 A.

In the Fig. 3 B, the amphipathic helix by helical wheel representation of residues 9 to 39 can be visualized. An heptad (orange residues) that form a hydrophobic region capable of mediating intermolecular reactions, can also be noticed. The N-terminal portion has a high α -helical content (residues 9 to 39) and contains heptad repeats capable of forming an amphipathic α -helical coiled-coil structure. In this way, it is believed that they can aggregate with other hydrophobic surfaces in a strong involvement of hydrophobic interactions in the overall oligomer assembly. They can form a parallel coiled-coil of alpha helices from two polypeptide chains, which hold them together.



Fig. 3. Prediction for putative *X. fastidiosa* H-NS of: **(A)** Coiled-coil structure in the N-terminal region using COILS Prediction. The coiled-coil domain is displayed by a periodicity represented by repeated units composed of seven amino acid length (heptad *a-g*), starting with *d* for the first Leu or Ile. Also is shown the secondary structure composed by helix (H), β -strand (E) and no prediction (-) (using NNPredict program). In bold are the conserved basic residues possibly involved in DNA binding. **(B)** Helical wheel representation of the residues in the position 9 to 39. The amino acid side chains are projected down the axis (the axis of an alpha helix, orthogonal to the paper plane). The filling patterns of the residues in the wheel indicate the nature of each residue.

These models are in close consistency with the formation oligomer in the homologue. It is well know that Leucine Zipper has a dimerization domain occurring

mostly in regulatory proteins, where the coiled-coil structure of a Leucine Zipper is required for dimerization and can be predicted with reasonable accuracy by existing algorithms, such as the programs used in this work [36].

Combining the NMR results with all precedent predictions and with the similarity with other H-NS proteins, we proposed that the N-terminal domain might primarily serve to multimerize. To test this hypothesis and predictions experimentally, XF0749 protein by Dynamic Light Scattering (DLS), by non-denaturing PAGE electrophoresis and by Circular Dichroism in function of the protein concentration were analysed.

The DLS analysis under native conditions provided an estimate of the molecular weight of 76 kDa, confirming that XF0749 protein was at least tetrameric at 1 mg/ml in 40 mM sodium phosphate buffer (pH 7.5) added to 100 mM sodium chloride at 25°C. Errors can occur due to artifacts caused by formation of aggregates.

It is well established that H-NS exists as a homodimer and/or even an oligomer in solution where it depends on its concentration ($\text{dimer} \leq 10^{-9} \text{ M}$) and that the ability to form a dimer is required for the H-NS to function as a transcriptional repressor [4,17,33]. Recently it was found that the oligomerization in the H-NS family is strongly affected by physical (temperature) and chemical (cation) effectors [16]. Beside this, the predominant form of H-NS in a wide range solution is tetrameric, presumably at a high concentration ($\geq 10^{-6} \text{ M}$) comparable to its intracellular concentration. Most of the temperature effects on the oligomerization equilibrium of H-NS occur below the critical temperature of 25-30°C, while above this temperature there are only marginal effects on the equilibrium governing the quaternary structure of the protein [16].

In this sense, to more accurately determine the oligomeric state, the non-denaturing PAGE electrophoresis of the XF0749 protein was used to evaluate the oligomerization state in the presence and absence of salt (monovalent cations) and with oligonucleotide duplex.

The Fig. 4 results indicate that the absence of salts the protein is mainly monomeric. On the other hand, monovalent cations and/or DNA induce the formation of oligomers, thus suggesting that hydrophobic interactions may be involved. The predominant form in the presence of those cofactors/substrate appears to be the tetrameric one.

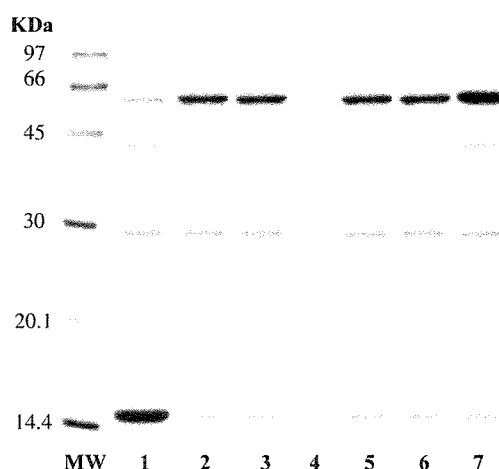


Fig. 4. Coomassie blue-stained non-denaturing polyacrylamide gel shows the analysis of protein-protein interactions. Lane M, low-molecular-weight protein markers (Amersham Pharmacia Biotech, Uppsala, Sweden); Lane 1, protein in buffer A; Lane 2, protein in buffer B; Lane 3, protein in buffer C; Lane 5, protein in buffer A added to 7.5 μ M oligonucleotide duplex; Lane 6, protein in buffer B added to 7.5 μ M oligonucleotide duplex; Lane 7, protein in buffer C added to 7.5 μ M oligonucleotide duplex.

Having found that in the absence of salts the protein is mainly monomeric, we therefore applied a complementary CD techniques to examine the self-associated state under equilibrium conditions over a concentration range from the low 0.15 μ M to 30 μ M, in 1 mM sodium phosphate buffer (pH 7.0) (Fig. 5). Has can be seen the protein shows a limited concentration dependence of its folding state. In Table 1 are reported the percentage of α -helix secondary structure evaluated by CD deconvolution analysis using the program DICROPROT program [23]. Changes in the CD spectra demonstrate that the self-

association of XF0749 protein is coupled to the formation of a α -helical structure. This is consistent with a coiled-coil structure. At concentration above 10 μM , just minor change in CD spectra was observed, indicating that a defined structural state had been attained.

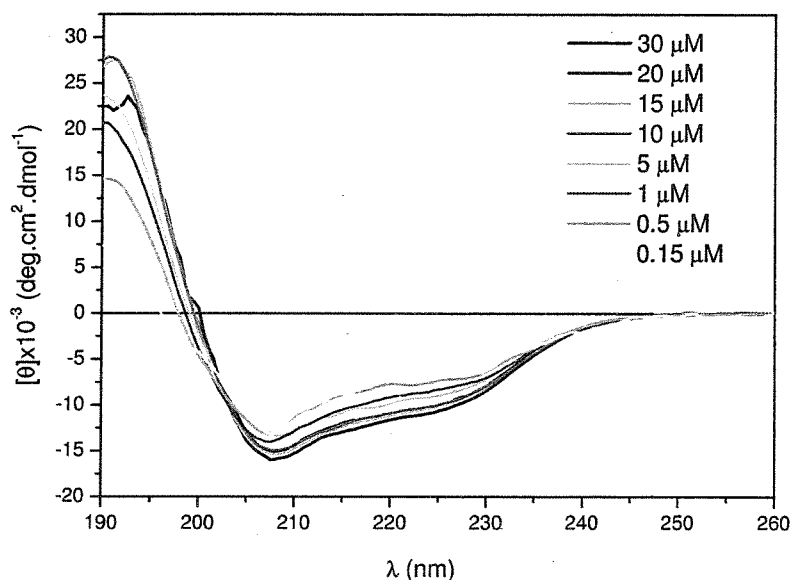


Fig. 5: Circular dichroism spectra at various concentration of the XF0749 protein in 1 mM sodium phosphate buffer (pH 7.0) at 25°C.

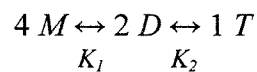
Table 1. Deconvolution of the circular dichroism data of the XF0749 protein in 1mM NaPi (pH 7) at 25°C. Dicroprot program (2.6 version) [23].

Concentration (μM)	α -Helix (%)			
	Chang	Chen	Fasman	Yang
30 ^a	33	31	41	23
20 ^a	31	28	37	20
15 ^a	31	29	39	20
10 ^a	30	28	37	19
5 ^a	27	24	33	16
1 ^b	26	22	34	15
0.5 ^b	24	19	31	12
0.15 ^b	24	17	28	13

a. Quartz cell 1 mm.

b. Quartz cell 5 mm.

These analyses have shown first that the XF0749 protein undergoes self-association and second that the oligomerization process is affected by the presence of cations and DNA. From all quaternary study data, relationships have been established with NMR results in which it could be indicated that the N-terminal was possibly unstructured in the free monomeric protein, justifying the poor dispersion of the NMR spectra signals. The last finding appears to have relevant significance to rationalize the *in vivo* properties and behavior of the H-NS, which is known to modulate gene expression in response to external conditions such as temperature, growth phase control and osmotic regulation [16,37]. Changes in such environmental parameters shift the oligomerization equilibrium toward their dissociation. Earlier biophysical experiments [37] were also satisfactorily accounted for by using a model implying the existence of two coupled equilibrium of the following types:



where M represents a monomer, D is a dimer, and T is a tetramer. The corresponding dissociation equilibrium constants K_1 ($\sim 10^{-7}$ M) and K_2 ($\sim 10^{-5}$ M) are sensitive to ionic strength and to temperature [16]. It is clear that a mechanistic model is needed to explain how the association interfaces involved in the oligomerization of H-NS in solution are rearranged during the formation of specific and nonspecific assemblies on a given DNA template.

The high concentration of H-NS in bacterial cells means that the prevalence of high order oligomers will provide a scaffold around which DNA can interact and facilitate its compaction [2].

4. Conclusion

A structural investigation of the XF0749 protein and its DNA interaction is reported herewith. The *in vitro* analysis of the non-specific DNA binding characteristics was successfully monitored by circular dichroism and fluorescence spectroscopy. The CD study demonstrated that the XF0749 protein does interact with synthetic DNA 14 bp oligonucleotide due conformational changes. In addition, the red shift of the cross-over associated to the fact that the 222/208 ration becomes < 1 when the protein is in the presence of the DNA fragment, Fig. 1 A, can be associated to the formation of a coiled-coil structure. In fact, when heated at 85°C it undergoes a sort of reversible conformational change characterized by the fact that upon cooling the shape of the spectra is quite comparable with the native state, only present in a lower molar ellipticity. Interestingly, when the same thermal cycle is carried out on the complex with DNA we find it in a very similar state. This might indicate that the protein tends to oligomerize in some higher aggregated state.

The DNA binding activity was also verified by fluorescence analyses where the DNA functioned as a quencher, decreasing the intrinsic fluorescence providing an indication that the Trp residues in the C-terminal region are involved in the DNA-binding. The data indicate a protein/DNA molar ratio of 2.

The spectroscopic techniques were also used for secondary structure study. The XF0749 protein appears to be quite stable in a ample pH range. The transition temperature of 52°C is consistent with that related to homologous, indicating that this protein is reasonably thermal resistant. From the fluorescence experiments, the denaturation effect of urea and quencher by acrylamide was studied. The results reveal that the C-terminal (DNA-

binding domain where all the tryptophans are) was folded and that they probably are located in an apolar region.

An attempt was made to obtain a full-length NMR spectra, but only a limited number of peaks were dispersed, suggesting that it was similar to what occurred in the that homologous, where the protein oligomerized in a higher concentration or that the protein was partially denatured due to the absence of a cofactor/substrate to stabilize it.

Structure predictions suggest the presence of two domains responsible for XF0749 protein function. Indeed, the N-terminal domain probably has a high α -helical content and contains heptad repeats capable of forming an amphipathic α -helical coiled-coil structure. This would be consistent with the observation that this domain has a role in protein-protein interactions. DLS, non-denaturing PAGE electrophoresis data indicate that the XF0749 protein is able to oligomerize to tetramer in a high ionic strength solution or in the presence of DNA. At low ionic strength the protein does not present any significant oligomerization concentration dependence.

We strongly believe that the C-terminal (43 residues) from XF0749 protein does not participate in oligomerization and form a free mobile DNA-binding domain, presenting dispersed peaks in the 2D ^1H ^{15}N HSQC NMR spectra, like related to [2] and [35].

In the present paper, using predictions and experimental studies, evidence is provided to put forth that XF0749 protein seems to be an H-NS like protein, comprising a large and widespread family of functionally and structurally related DNA binding proteins. Such a conclusion validates the genomic annotation of the XF0749 ORF from *X. fastidiosa* (9a5c clone) and corroborates the homologous predictions, based on the primary sequence.

Acknowledgements

The authors would like to thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 01/07533-7) and CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support. A.P.S. received a fellowship from CNPq. A.R.A received a post-doctorate fellowship from FAPESP. D.P.P. received a Ph.D. fellowship from CAPES. We would also like to thank the researchers Francisco Benedito Teixeira Pessine Ph.D. from the Institute of Chemistry, The State University of Campinas, and Francisco Javier Medrano Ph.D. and Rita de Cássia Rossi Figueredo Ms. technician from the Center for Molecular and Structural Biology, National Synchrotron Light Laboratory, for technical expertise.

References

- [1] D.W. Ussery, J.C. Hinton, B.J. Jordi, P.E. Granum, A. Seirafi, R.J. Stephen, A.E. Tupper, G. Berridge, J.M. Sidebotham, C.F. Higgins, The chromatin-associated protein H-NS, *Biochimie* 76 (1994) 968-980.
- [2] C.P. Smyth, T. Lundbäck, D. Renzoni, G. Siligardi, R. Beavil, M. Layton, J.M. Sidebotham, J.C.D. Hinton, P.C. Driscoll, C.F. Higgins, J.E. Ladbury, Oligomerization of the chromatin-structuring protein H-NS, *Mol. Microbiol.* 36 (2000) 962-972.
- [3] A. Spassky, H.C. Buc, Physico-chemical properties of a DNA binding protein: *E. coli* factor H1, *Eur. J. Biochem.* 81 (1977) 79-90.
- [4] C. Ueguchi, S. Tomomi, T. Suzuki, T. Mizuno, Clarification of the dimerization domain and its functional significance for the *E. coli* nucleoid protein H-NS, *J. Mol. Biol.* 274 (1997) 145-151.
- [5] P. Bertin, N. Benhabiles, E. Krin, Ch. Laurent-Winter, E. Turlin, A. Thomas, A. Danchin, R. Brasseur, The structural and functional organization of H-NS-like proteins

- is evolutionarily conserved in gram-negative bacteria, *Mol. Microbiol.* 31 (1999) 319-329.
- [6] P. Dersch, K. Schmid, E. Bremer, Synthesis of the *E. coli* K12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation, *Mol. Microbiol.* 8 (1993) 875-889.
- [7] M. Falconi, N.P. Higgins, R. Spurio, C.L. Pon, C.O. Gualerzi, Expression of the gene encoding the major bacterial nucleotide protein H-NS is subject to transcriptional auto-repression, *Mol. Microbiol.* 10 (1993) 273-282.
- [8] T. Yoshida, C. Ueguchi, H. Yamada, T. Mizuno, Function of the *E. coli* nucleoid protein, H-NS: molecular analysis of a subset of proteins whose expression is enhanced in a *hns* deletion mutant, *Mol. Gen. Genet.* 237 (1993) 113-122.
- [9] A. Zhang, S. Rimsky, M.E. Reaban, H. Buc, M. Belfort, *E. coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics, *EMBO J.* 15 (1996) 1340-1349.
- [10] C.F. Higgins, C.J. Dorman, D.A. Stirling, L. Waddell, I.R. Booth, G. May, E. Bremer, A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*, *Cell* 52 (1988) 569-584.
- [11] L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann, H. Buc, Synthetic curved DNA sequences can act as transcriptional activators in *E. coli*, *EMBO J.* 13 (1989) 4289-4296.
- [12] H. Yamada, S. Muramatsu, T. Mizuno, An *E. coli* protein that preferentially binds to sharply curved DNA, *J. Biochem.* 108 (1990) 420-425.
- [13] T.A. Owen-Hughes, G.D. Pavitt, D.S. Santos, J.M. Sidebotham, C.S.J. Hulton, J.C.D. Hinton, C.F. Higgins, The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression, *Cell* 71 (1992) 255-265.
- [14] A. Spassky, S. Rimsky, H. Garreau, H. Buc, H1a, an *E. coli* DNA binding protein which accumulates in stationary phase, strongly compacts DNA *in vitro*, *Nucleic Acids Res.* 12 (1984) 5321-5340.

- [15] C. Tendeng, E. Krin, O.A. Soutourina, A. Marini, A. Danchini, P. N. Bertin. A novel H-NS protein from the antartics Psychrophilic bacterium reveals a crucial role for the N-terminal domain in thermal stability, *Biolo. Chem.* 278 (2003) 18754-18760.
- [16] S. Ceschini, G. Lupidi, M. Coletta, C.L. Pon, E. Fioretti, M. Angeletti, Multimeric self-assembly equilibrium involving the Histone-like protein H-NS, *J. Biol. Chem.* 275 (2000) 729-734.
- [17] M. Falconi, M.T. Gualtieri, A. La Teana, M.A. Losso, C.L. Pon, Proteins from the prokaryotic nucleoid: primary and quaternary structure of the 15-kD *E. coli* DNA binding protein H-NS, *Mol. Microbiol.* 2 (1988) 323-329.
- [18] R. Spurio, M. Falconi, A. Brandi, C.L. Pon, C.O. Gualerzi, The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending, *EMBO J.* 16 (1997) 1795-1805.
- [19] D. Esposito, A. Petrovich, R. Harris, S. Ono, J.F. Eccleston, A. Mbaali, I. Haq, C.F. Higgins, J.C.D Hinton, P.C. Driscoll, J.E. Ladbury, H-NS oligomerization domain structure reveals the mechanism for the high order self-association of the intact protein, *J. Mole. Biol.* 324 (2003) 841-850.
- [20] D.P. Paula, A.R. Azzoni, S.F. Siqueira, C.F. Catani, L.K. Rosselli, A.P. Souza, Expression and purification of a putative H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*, *Protein Expres. Purif.* 32 (2003) 61-67.
- [21] J. Sambrook, D.W. Russell, Plasmids and their usefulness in molecular cloning, in: *Molecular Cloning, a Laboratory Manual*, 3^a ed., vol. 1, pp. 1.32-1.34, CSHL Press, New York, 2001.
- [22] S. Milev, A.A. Gorfe, A. Karshikoff, R.T. Clubb, H.R. Bosshard, I. Jelesaroy, Energetics of sequence-specific protein-DNA association: binding of integrase Tn916 to its target DNA, *Biochemistry* 42 (2003) 3481-3491.
- [23] G. Deléage, C. Geourjon, An interactive graphic program for calculating the secondary structures content of proteins from circular dichroism spectrum, *Comp. Appl. Biosc.* 9 (1993) 197-199.

- [24] I.M. Grodin, S.I. Vavilov, Über die wirkungssphäre der ausloschens vargangeinsen fluoreszierenden flüssigkeiten, *Z. Phys.* (1931) 69, 100.
- [25] D. Tippner, R. Wagner, Fluorescence analysis of the *E. coli* transcription regulator, H-NS reveals two distinguishable complexes dependent on binding to specific or nonspecific DNA sites, *J. Biol. Chem.* 38 (1995) 22243-22247.
- [26] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277-293.
- [27] B.A. Johnson, R.A. Blevins, NMRView: A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR* 4 (1994) 603-614.
- [28] A. Rodger, B. Nordén, Circular dichroism of biomolecules: DNA-ligand interactions, in: *Circular dichroism and Linear dichroism*, 1st edition, Oxford University Press, Oxford, pp. 15-32, 1997.
- [29] C.O. Gualerzi, M.A. Losso, M. Lammi, K. Friedrich, R.T. Pawlik, M.A. Canonaco, G. Gianfranceschi, A. Pingoud, C.L. Pon, in: C.O. Gualerzi, C.L. Pon, *Proteins from the prokaryotic nucleoid: structural and functional characterization of the E. coli DNA binding proteins NS (HU) and H-NS*, *Bacteriol. Chromatin*, vol 3, Springer, 1986.
- [30] A.E. Tupper, H.T. Owen, D.W. Ussery, D.S. Santos, D.J. Ferguson, J.M. Sidebotham, J.C. Hinton, C.F. Higgins, The chromatin-associated protein H-NS alters DNA topology in vitro, *EMBO J.* 16 (1994) 1795-1805.
- [31] S.M. Kelly, N.C. Price, The application of circular dichroism to studies of protein folding and unfolding, *Biochim. Biophys. Acta.* 1338 (1997) 161-185.
- [32] H. Shindo, A. Ohnuki, H. Ginba, E. Kato, C. Ueguchi, T. Mizuno, T. Yamazaki, Identification of the DNA binding surface of H-NS protein from *E. coli* by heteronuclear NMR spectroscopy, *FEBS Lett.* 455 (1999) 63-69.
- [33] C. Ueguchi, S. Tomomi, T. Yoshida, K. Tanaka, T. Mizuno, Systematic mutational analysis revealing the functional domain organization of *E. coli* nucleoid protein H-NS, *J. Mol. Biol.* 263 (1996) 149-162.

- [34] R.M. Williams, S. Rimsky, H. Buc, Probing the structure, function and interactions of the *E. coli* H-NS and StpA proteins by using dominant negative derivatives, J. Bacteriol. 178 (1996) 4335-4343.
- [35] A. Lupas, Coil 2.2 Manual, 1996 ([ftp: //ftp.biochem.mpg.de/coils](http://ftp.biochem.mpg.de/coils)).
- [36] E. Bornberg-Bauer, E. Rivals, M. Vingron, Computational Approaches to Identify Leucine Zippers, Nucleic Acids Res. 26 (1998) 2740-2746.
- [37] C. Badaut, R. Williams, V. Arluison, E. Bouffartigues, B. Robert, H. Buc, S. Rimsky, The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA, J. Biol. Chem. 277 (2002) 41657-41666.

6. RESULTADOS COMPLEMENTARES

Durante o desenvolvimento do doutorado alguns experimentos não foram relatados nos artigos 1 e 2 gerados nesta tese. São eles:

6.1. Isoformas de H-NS

Três isoformas da proteína XF0749, tanto para a proteína recombinante, quanto para a nativa purificada, foram identificadas no gel bidimensional preliminar ao MALDI-TOF.

A eletroforese bidimensional foi realizada no Laboratório de Química de Proteínas (LAQUIP), Departamento de Bioquímica, IB – UNICAMP. A proteína nativa purificada (60 µg) foi precipitada com acetona/etanol/ácido acético na proporção (50% : 50% : 0,1%) a – 20°C, correspondendo a três vezes o volume inicial da amostra. Incubou-se por 15 min em gelo e centrifugou-se a 14000 rpm por 10 min. A seguir, o precipitado foi ressuspendido no tampão de amostra composto de 7 M de uréia, 2 M tiouréia, 4 % (v/v) CHAPS, 0,75% (v/v) anfólitos (*IPGbuffer*), 70 mM DTT e 0,002% de azul de bromofenol. Agitou-se a amostra por 1 min em vortex e, após 10 min, novamente agitou-se por 1 min. Centrifugou-se a 14000 rpm por 10 min. A amostra foi aplicada no gel 1 D de focalização isoeletrica, utilizando-se um *IPGstrip* de 6-11, e posteriormente no gel 2 D de SDS-PAGE (12,5%). A coloração do gel foi com *Coomassie blue* R-250 e descoloração em ácido acético/metanol (Fig. 7).

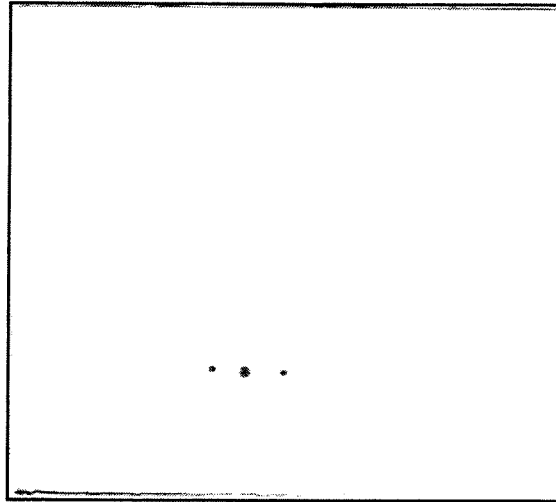


Fig. 7. Eletroforese bidimensional da proteína XF0749 nativa.

A detecção destas isoformas sugerem que a proteína XF0749 sofre algum tipo de modificação pós-traducional, mesmo tratando-se de um sistema heterólogo. Estes dados são consistentes com aqueles observados para a proteína H-NS de *E. coli* por Spassky *et al.* (1984) e Ussery *et al.* (1994). Estes autores observaram a existência de três isoformas denominadas H1a, H1b e H1c, com pI diferindo de aproximadamente 0,1 unidade entre si. Estes autores discutem que cada isoforma poderia apresentar alguma diferença na atividade biológica e, assim, influenciar no mecanismo pelo qual a proteína H-NS modula a expressão gênica. Contudo, nenhuma evidência de que as isoformas tenham de fato algum significado biológico foi comprovada. Até o momento, ainda não foi realizada a purificação de cada isoforma separadamente para testar suas propriedades de ligação ao DNA. Apenas foi analisada a incorporação de fosfato marcado em H-NS, mas a ausência de detecção da incorporação descartou a hipótese de modificação por fosforilação (Ussery *et al.*, 1994).

6.2. Ensaios de gel shift

Um dos procedimentos mais comumente utilizados para a detecção de proteínas ligantes a DNA é o Ensaio da Mudança da Mobilidade Eletroforética (*Electrophoretic Mobility Shift Assay* - EMSA), também conhecido como Retardamento em Gel ou *Gel Shift*. A interação entre a proteína e o fragmento de DNA resulta num complexo de maior massa molecular com menor mobilidade no gel, comparado à mobilidade do fragmento de DNA livre (Sambrook & Russel, 2001).

A escolha das sondas de DNA foi baseada na provável afinidade de ligação da proteína XF0749 às regiões promotoras, naturalmente ricas em seqüências de A e T, onde ocorre a regulação transcricional nos homólogos.

Utilizou-se como sondas de DNA o promotor *lacI* de pUC18 e o fragmento correspondente à provável região promotora da ORF XF0749 de *X. fastidiosa*. Estes fragmentos foram obtidos pela amplificação através de PCR (vetor pUC18 e DNA genômico de *X. fastidiosa* purificados), após desenho dos primers específicos. Os primers para amplificação do fragmento de 81 pb correspondente ao promotor *lacI* de pUC18 são: *sense* 5'-ACCGAGCTCGAATTCGTAATCATGG-3' e *antisense* 5'-GTTAC GTCATCATTAGGCACCCCA-3', e os primers para amplificação do fragmento de 181 pb correspondente à provável região promotora da ORF XF0749 são: *sense* 5'-GGCATTGATATCCATCGGCAGATAA-3' e *antisense* 5'-TCATAGTTTCCGTGTGG GGTGTCTTA-3'.

Cada sonda foi purificada pelo método padrão de extração com fenol:clorofórmio e precipitação com acetato de sódio e etanol (Sambrook & Russel, 2001) e então ressuspensas no tampão de ligação 40 mM HEPES (pH 7,9) com 60 mM cloreto de

potássio, 1 mM EDTA e DTT, 5 mM acetato de magnésio, 100 mM cloreto de sódio e 0,02 mM acetato de zinco. As concentrações dos fragmentos foram determinadas usando-se seus coeficientes de extinção molar pela absorção a 260 nm.

As reações de ligação consistiram de 20 µl de volume final em quantidades equimolares de 40 µg de DNA e concentrações crescentes de proteína de 0; 0,2; 0,6; 1; 1,4 µM. As reações foram incubadas a 15°C em banho-maria por 30 min seguido pela adição de *loading buffer* (0,25 % de azul de bromofenol e 15 % Ficoll-400 em tampão Tris-glicina 1X composto por 25 mM Tris-clorato pH 8,3 adicionado a 190 mM glicina e 1 mM EDTA) correspondente a 10% do volume total de reação. A eletroforese foi realizada a 4°C em gel não desnaturante 10% (gel concentrador de 4%) na proporção de 29:1 (acrilamida:bis), composto de tampão Tris-glicina 1 X, 10% de APS e TEMED. O tampão de corrida da eletroforese foi o mesmo Tris-glicina 1 X, usando uma placa de dimensões de 10 x 16 cm x 1,5 mm do sistema Hoefer (*Scientific Instruments*, São Francisco, CA - USA). Antes de carregar a amostra no gel, fez-se uma pré-corrida por 60 min a 120 V. A eletroforese ocorreu por 4 h, inicialmente a 200 V e após as amostras entrarem no gel, abaixou-se a voltagem para 120 V. Os fragmentos de DNA foram corados em 0,5 µg/ml de solução de brometo de etídio em tampão Tris-glicina 1X (pH 8,3) por 30 min e visualizados em transiluminador com lâmpada de UV, sendo então o gel fotografado no equipamento *GeneGenius Imager* (Gateway, USA).

A Fig. 8 representa os resultados relativos aos ensaios de gel shift. Observa-se que para os fragmentos de DNA na ausência de proteína, as bandas correspondem exatamente ao tamanho de pares de bases esperado para os respectivos fragmentos. Contudo, a medida que a concentração da proteína XF0749 era crescente, diminuía-se gradativamente a

quantidade de DNA livre, sugerindo que parte dos fragmentos estavam complexados com a proteína, entretanto a detecção por brometo de etídio foi insuficiente devido a baixa sensibilidade para corar o DNA complexado.

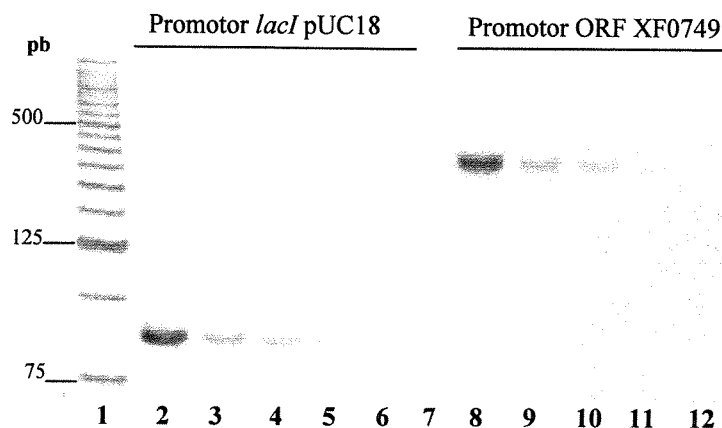


Fig. 8. Ensaio de *gel shift* para a interação entre a proteína XF0749 e os fragmentos de DNA correspondentes ao promotor *lacI* do vetor pUC18 (81 pb) e ao promotor hipotético *hns* da ORF XF0749 (181 pb). Trilha 1, Ladder de 25 pb (*Invitrogen* do Brasil, São Paulo - SP). Trilhas 2 a 6, concentrações crescentes de proteína: 0; 0,2; 0,6; 1; 1,4 μ M para reação com o fragmento de 81 pb. Trilhas 8 a 12 correspondem às mesmas concentrações crescentes de proteína, mas para reação com o fragmento de 181 pb.

Quantidades maiores de fragmentos de DNA e de proteína foram testadas, mas a visualização dos complexos ainda não se tornou possível. Acredita-se que deva ser feito a marcação radioativa da extremidade 3' dos fragmentos de DNA com [α - 32 P]dNTPs usando *Klenow fragment* de *E. coli* DNA Polimerase I e detecção por autoradiografia. Outras variáveis também foram testadas sem sucesso, tais como a adição de 0,1 mg/ml de BSA e de 5% (v/v) glicerol no tampão de ligação, tampão de corrida TBE 1 X e corrida a 150 V.

Dada a confirmação da interação proteica com DNA por meio das técnicas espectroscópicas de dicroísmo circular e fluorescência, não seguiu-se adiante com os ensaios de *gel shift* pois conduziriam ao resultado já obtido.

6.3. Otimização da expressão heteróloga da ORF XF0749

Com a finalidade de obter-se o máximo de proteína heteróloga marcada com o isótopo de ^{15}N para os ensaios de RMN (2 D ^1H ^{15}N HSQC), buscou-se a otimização da expressão e da obtenção da mesma, baseando-se na literatura dos homólogos onde há relatos de testes de indução da expressão em diferentes condições: na presença do antibiótico rifampicina; indução após choque térmico; em diferentes fases do crescimento populacional bacteriano; indução em diferentes linhagens e ainda em meios de cultura de diferente força iônica. Todas estas condições estão detalhadas nos parágrafos que se seguem.

Almeida *et al.* (2001) testaram o antibiótico rifampicina para a marcação isotópica de proteínas heterólogas para estudos de RMN. Segundo os autores, a expressão em meio mínimo com $^{15}\text{NH}_4\text{Cl}$ na presença de rifampicina possibilita a marcação exclusiva da proteína heteróloga devido a inibição da síntese proteica celular pela ligação do antibiótico à subunidade β da RNA polimerase bacteriana, impedindo a iniciação da transcrição. Entretanto, a rifampicina não inibe a T7 RNA polimerase de vetores tipo pET. A rifampicina a 200 $\mu\text{g/ml}$ recém-preparada foi adicionada ao meio de cultura entre 5 a 20 min da adição do indutor da expressão.

O nível de expressão de H-NS aumenta quando a célula entra na fase estacionária (Ueguchi *et al.*, 1993), embora este aumento da expressão de H-NS não tenha sido detectado por Hinton *et al.* (1992) e Yasuzawa *et al.* (1992) ao testarem as mesmas condições. La Teana *et al.* (1991) também detectou aumento na expressão de H-NS de três a quatro vezes durante o crescimento na fase lag ao haver a mudança da temperatura de crescimento 37°C para 10°C.

Grossman *et al.* (1998) relataram a instabilidade dos sistemas de expressão pET que permitem a desrepressão do operon *lac* em resposta à limitação nutricional do meio quando a população bacteriana entra na fase estacionária em meios complexos. A diminuição gradativa da glicose no meio de cultura faz com que a inibição da enzima adenilato ciclase, pelo subproduto catabólico da glicose, reduza continuamente e, em consequência, haja a reação catalisada pela adenilato ciclase da conversão do ATP em AMP cíclico (AMPc). O AMPc, por sua vez, atua como um efêtor alostérico da proteína ativadora do metabolismo (CAP), formando o complexo AMPc/CAP. Este complexo se liga ao sítio CAP do DNA, aumentando a afinidade da RNA polimerase ao promotor cromossomal *lacUV5*, e assim estimulando a transcrição do gene *l* da T7 RNA polimerase¹⁰. Uma vez expressa, a enzima T7 RNA polimerase se liga ao promotor T7 plasmidial e inicia a transcrição da proteína heteróloga (Fig. 9).

No caso da proteína XF0749, a transcrição basal compromete a produção total da proteína heteróloga, devido à propriedade regulatória repressora desta proteína. Para proteínas reguladoras como esta, ou outras de caráter tóxico, é extremamente desejável o controle rígido de sua transcrição. Em vista disto, testou-se a expressão na linhagem de *E. coli* BL21(DE3)pLysS e, assim como Grossman *et al.* (1998), fez-se a peletação celular na fase log e a ressuspensão em meio complexo fresco (Para 1 L de meio 4 YT têm-se: 32 g de bacto triptona, 20 g de bacto extrato de levedura e 5 g de cloreto de sódio pH 6), antes da indução da expressão.

¹⁰ O controle positivo pelo AMPc atua como um sistema de controle adicional ao sistema negativo *lac* repressor-operador.

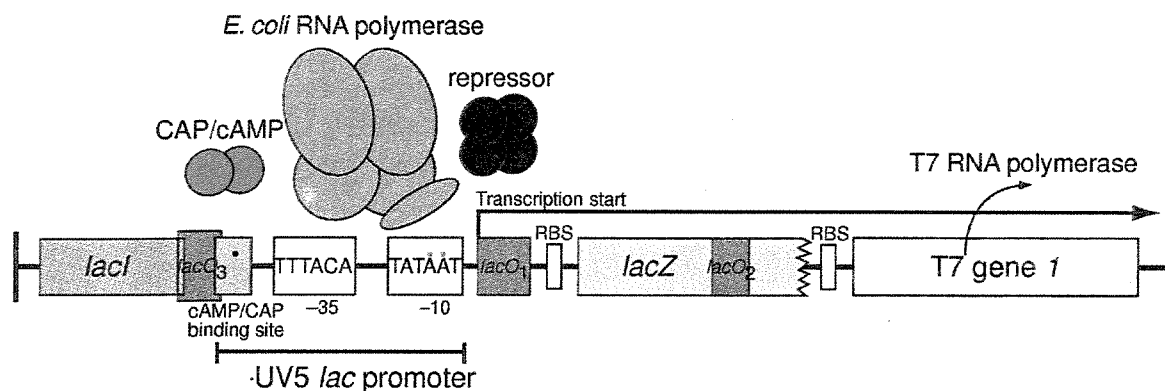


Fig. 9. Controle transcricional do gene *l* T7 (codifica a T7 RNA Polimerase) em λDE3 lisogênicos. A transcrição do gene *l* T7 em sistemas de expressão tipo pET é controlada pelo promotor *lacUV5*. O repressor *lac* (codificado pelo gene *lacI*) liga-se ao operador (gene *lacO*) inibindo a transcrição pela RNA Polimerase de *E. coli*. O indutor IPTG liga-se ao repressor, reduzindo sua afinidade pelo operador, desinibindo a transcrição. Quando o nível de AMPc está elevado, isto é, na ausência de glicose, o complexo AMPc/CAP é formado e liga-se *upstream* do promotor para estimulação da transcrição. Fonte: Novy, R.; Morris, B. Use of glucose to control basal expression in the pET systems. *in* *Novations* 13.

A linhagem BL21(DE3)pLysS possui um plasmídio compatível com os vetores comerciais (pET, p.ex.), o qual produz uma pequena quantidade de lisozima T7, inibidor natural da enzima T7 RNA polimerase, impedindo a transcrição basal do vetor. Na presença de indutor (IPTG ou lactose) o gene da T7 RNA polimerase é expressa em grande quantidade, suprimindo o efeito da inibição pela lisozima T7 (Fig. 10) (Moffart & Studier, 1987). Entretanto, há o aumento do tempo lag entre a indução do gene T7 RNA polimerase e a expressão máxima da proteína heteróloga (Sambrook & Russel, 2001). Em vista disto, duas curvas de sobrevivência tiveram de ser determinadas, uma para a linhagem BL21(DE3) e outra para a linhagem BL21(DE3)pLysS, ambas transformadas com o vetor recombinante para a ORF XF0749, cultivadas a 300 rpm e a 37°C em meio LB contendo 50 µg/µl de ampicilina, sem realizar indução da expressão heteróloga (curvas não apresentadas).

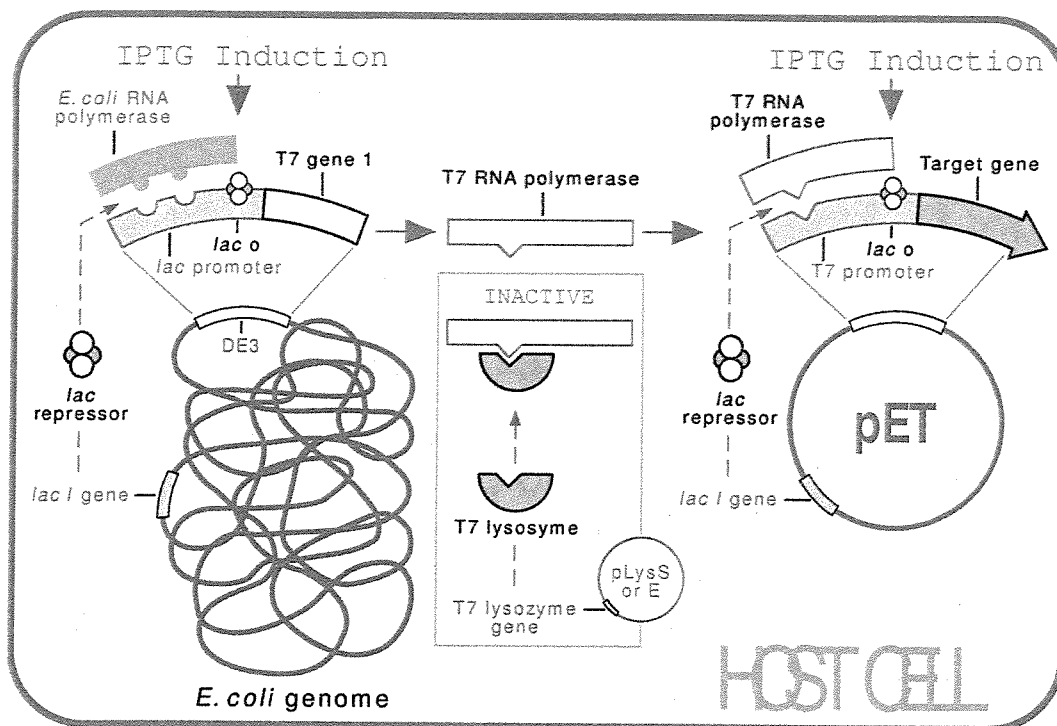


Fig. 10. Elementos de controle da transcrição do sistema pET. Fonte: Mierendorf, R.; Yeager, K.; Novy, R. The pET systems: your choice for expression. *inNovations*, vol. 1, n. 1, may 1994.

Baseando-se nos trabalhos referidos neste tópico, testou-se seis condições de expressão para a proteína XF0749. Todos os tratamentos foram realizados a 300 rpm em meio LB (a exceção do tratamento de peletação para ressuspensão em meio 4 YT), utilizando a linhagem BL21(DE3) (a exceção do tratamento em BL21(DE3)pLysS), e adicionando-se 5,6 mM de lactose para indução plasmidial. O Quadro 1 descreve resumidamente cada tratamento.

Conforme ilustrado na Fig. 11, o melhor nível de expressão para a proteína hipotética H-NS de *X. fastidiosa* ocorreu com a linhagem BL21(DE3)pLysS, evidenciando a importância da supressão da transcrição basal da proteína alvo, devido a sua provável propriedade bioquímica, mesmo estando recombinada em sua porção N-terminal.

QUADRO 1. Testes para otimização da expressão heteróloga da ORF XF0749 clonada em pET 32-Xa/LIC.

TRATAMENTO	LINHAGEM	TEMPERATURA (°C)
1. Est: Indução na fase estacionária	BL21	37
2. Indução após peletação a 6000 rpm e ressuspensão em meio 4 YT	BL21	37
3. Indução na fase lag, seguida de choque térmico	BL21	37 → 15
4. Indução na fase lag com rifampicina, seguida de choque térmico	BL21	37 → 15
5. Indução na fase log, seguida de choque térmico	BL21	37 → 15
6. Indução na fase log em BL21(DE3)pLysS	pLysS	37
7. Controle: Indução na fase log em BL21(DE3)	BL21	37

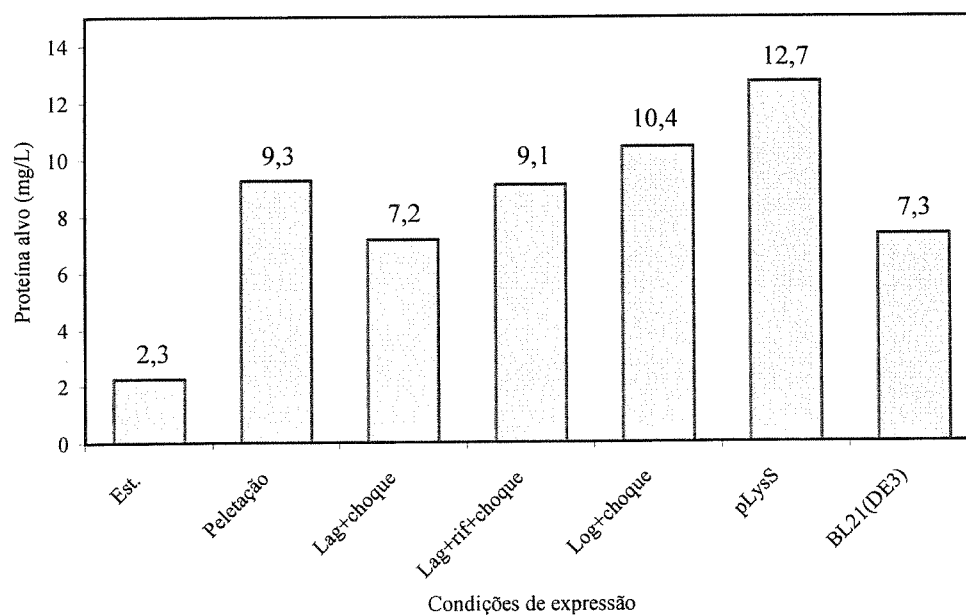


Fig. 11. Expressão da proteína XF0749 recombinante: produção por célula até O.D.₆₀₀ = 1,6 (fase estacionária). A proteína alvo foi detectada e quantificada do extrato clarificado utilizando-se o kit *S-Tag Rapid Assay* (Novagen, WI - USA), tendo-se a vantagem da proteína estar recombinada ao peptídeo S na extremidade N-terminal (vide artigo 1).

Os tratamentos de choque térmico nas fases log e lag + rif após indução, bem como a peletação da cultura para ressuspensão em meio complexo antes da indução, tiveram melhores valores de expressão heteróloga em relação ao controle.

6.4. Espectros de RMN: 2 D ^1H ^{15}N HSQC e medida da difusão em Gradiente de Campo Magnético Pulsado

Conforme discutido no artigo 2 desta tese, não foi possível prosseguir a coleta de dados para a resolução da estrutura tridimensional da proteína XF0749 por RMN 2 D ^1H ^{15}N HSQC devido ao padrão de dispersão insatisfatório dos picos proveniente de uma proteína parcialmente desenovelada ou oligomerizada (Fig. 12). Para a averiguação do estado de oligomerização proteica nas mesmas condições experimentais do primeiro ensaio de RMN 2 D ^1H ^{15}N HSQC, fez-se a medida da difusão proteica em Gradiente de Campo Magnético Pulsado por espectroscopia DOSY (*Diffusion-ordered spectroscopy*) em PG-SLED (*Pulse Gradient Stimulated Echo Longitudinal Encode-Decode*), como descrito por Wilkins *et al.* (1999). A medida do coeficiente de difusão proteica permite a estimativa precisa da massa molecular da proteína alvo através da medição de seu raio hidrodinâmico.

A preparação da amostra foi efetuada como relatado no artigo 2 (tópico de RMN), à exceção que não houve marcação isotópica com ^{15}N . A ressuspensão da proteína liofilizada foi feita em solução contendo 0,01% de dioxano, utilizado como padrão interno de viscosidade (Wilkins *et al.*, 1999). Para repetir as mesmas condições experimentais, a medição foi efetuada a 25°C na ausência de cátions monovalentes e DNA.

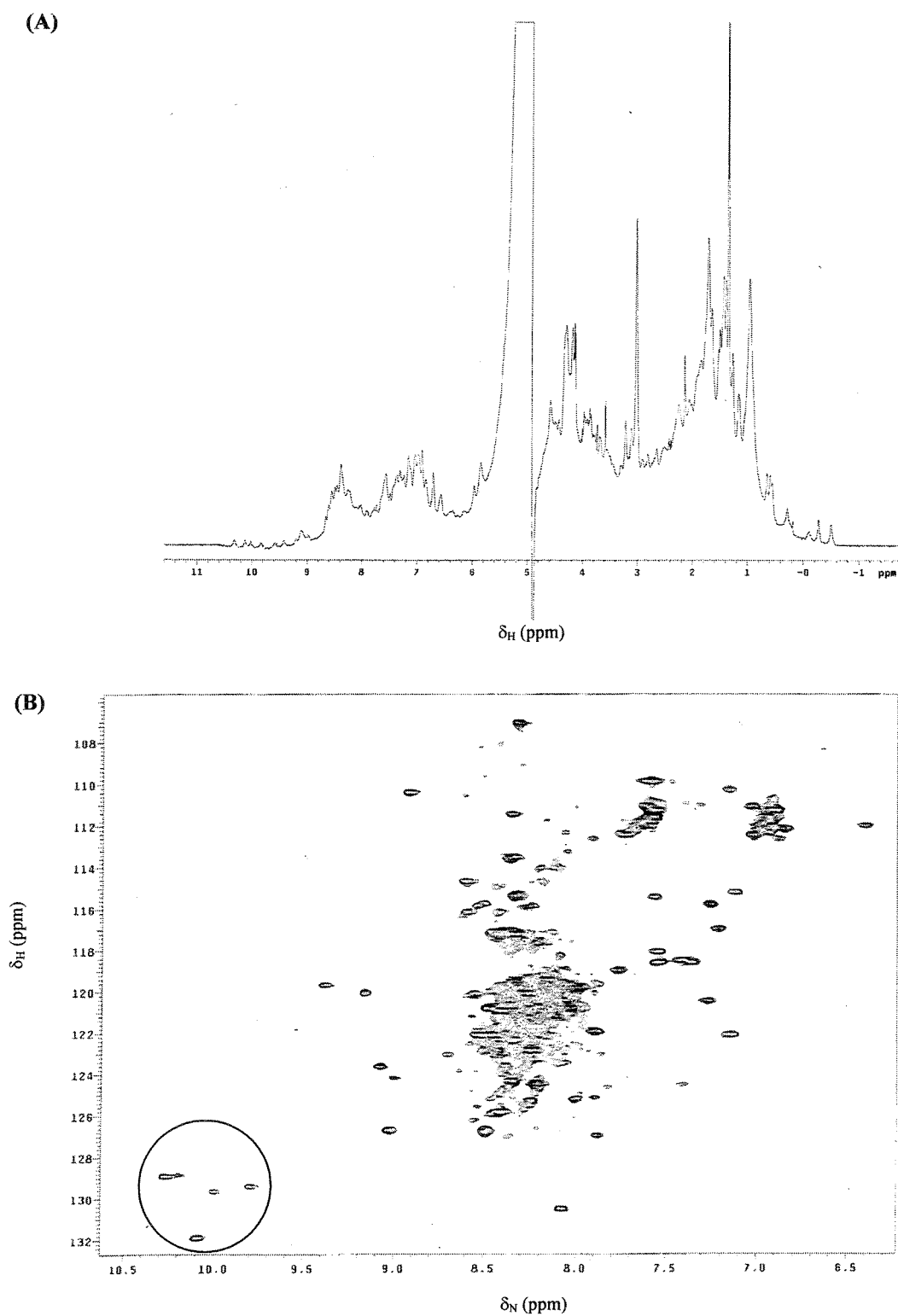


Fig. 12. Espectros da proteína XF0749: **(A)** Espectro de RMN 1D ^1H ; **(B)** Espectro RMN 2D ^1H ^{15}N -HSQC. Os quatro picos no círculo correspondem ao sinal dos quatro resíduos de triptofano do domínio C-terminal.

O valor do raio hidrodinâmico foi calculado das medidas do coeficiente de difusão pela taxa de decaimento dos espectros para a proteína alvo (d_{prot}) e para a molécula de referência (d_{ref}). O raio hidrodinâmico pôde então ser estimado pela fórmula:

$$R_h^{prot} = \frac{d_{ref}(R_h^{ref})}{d_{prot}} \quad (1)$$

onde R_h^{prot} corresponde ao raio hidrodinâmico da proteína e R_h^{ref} o raio da molécula de dioxano.

Para a determinação do raio do dioxano utilizado na equação 1, fez-se a calibração utilizando-se o raio de giro da lisozima de galinha por *Small-Angle Scattering X-rays Studies* (SAXS, 15,3 Å). O raio de giro (R_g) é a raiz quadrática média da distância de todos os átomos na molécula em relação ao seu centro de massa, enquanto o raio hidrodinâmico (R_h) pode ser definido como o raio de uma esfera que tem o mesmo coeficiente de difusão da molécula de proteína (Wilkins *et al.*, 1999).

A análise da medida do raio hidrodinâmico, para proteínas esféricas de estrutura conhecida, indica que existe uma forte correlação entre as dimensões moleculares e o comprimento da cadeia de polipeptídeo (Wilkins *et al.*, 1999). Estes parâmetros podem ser ajustados na equação empírica:

$$R_h = (4,75 \pm 1,11)N^{0,29 \pm 0,02} \quad (2)$$

onde N é o número de resíduos na cadeia e R_h é o raio hidrodinâmico (em Å).

O Quadro 2 descreve os valores teóricos do raio hidrodinâmico da proteína XF0749 ajustados na equação 2. Cada medida foi repetida cinco vezes para melhorar a relação sinal/ruído. O processamento dos dados foi efetuado no *software* Felix 2,3 da *Biosym Technologies* (San Diego, CA - USA).

QUADRO 2. Valores teóricos do raio hidrodinâmico (R_h) da proteína XF0749.

COMPRIMENTO DA CADEIA (N)	RAIO HIDRODINÂMICO TEÓRICO (R_h) EM Å
134 (monômero)	19,7
268 (dímero)	24,0
402 (trímero)	27,0
536 (tetramero)	29,4

A medida do raio hidrodinâmico (eq. 1) para a proteína X0749 foi de $R_h = 18,54$ Å. Já o raio hidrodinâmico teórico (eq. 2) para um monômero de 134 resíduos de aminoácidos foi de $R_h = 19,66$ Å (Quadro 2). Comparando-se o resultado experimental com os valores teóricos descritos no Quadro 2, pode-se verificar que a proteína se apresentava na forma monomérica na concentração de 500 μ M, a 25°C, na ausência de cátions monovalentes e DNA, ou seja, mesmas condições experimentais do ensaio de 2D ^1H ^{15}N HSQC.

Este resultado de difusão em Gradiente de Campo Magnético Pulsado revelou a necessidade de realizar testes complementares a fim de se explicar por que o espectro de RMN 2D ^1H ^{15}N HSQC da forma monomérica da proteína XF0749 não se apresentava de forma dispersa. A proteína alvo poderia estar parcialmente desenhovelada; contudo, os testes espectroscópicos de CD e Fluorescência não possuem sensibilidade para revelar o grau de desenhovelamento, tal como o de dispersão dos picos de RMN 2D ^1H ^{15}N HSQC.

Contudo, a proteína alvo revelou-se oligomerizada nos ensaios de Espalhamento Dinâmico de Luz (solução tamponante a 40 mM de fosfato de sódio pH 7,0 acrescido de 100 mM de cloreto de sódio), de eletroforese em gel de poliacrilamida não desnaturante (presença de cátions monovalentes e DNA) e de variação da concentração proteica em CD.

Além disto, a proteína XF0749 apresentou atividade de ligação ao DNA nos testes funcionais por CD e Fluorescência (vide artigo 2).

A literatura dos homólogos de H-NS relata que além da concentração proteica, outros fatores poderiam estar relacionados e serem essenciais para a oligomerização sendo eles: a presença de cátions monovalentes e a temperatura, em condições semelhantes ao ambiente celular em que a proteína exerce naturalmente seu papel biológico (Spassky *et al.*, 1984; Falconi *et al.*, 1988; Ueguchi *et al.*, 1996; Ceschini *et al.*, 2000).

Posteriormente descobriu-se que para o desempenho de sua função de interação com o DNA, o estado de oligomerização era crucial (Badaut *et al.*, 2002), podendo também a presença do DNA ser um dos fatores que estabilizam e/ou direcionam sua oligomerização (efeito de cooperatividade). Diante dos resultados experimentais de difusão em Gradiente de Campo Magnético Pulsado e dos obtidos no artigo 2 desta tese, fazia-se necessário realizar novos experimentos de difusão com a proteína pura na presença de sal monovalente e de DNA. Todavia não foi possível a execução destes testes dada à dificuldade de resolubilização proteica após liofilização. Acredita-se que a proteína XF0749 não seja suficientemente solúvel na alta concentração requerida para os ensaios de RMN. Aliado a isto, a paralização do equipamento de espectrometria de ressonância a 600 MHz devido a vazamento de hélio, e a sobrecarga na utilização do espectrômetro de 500 MHz, nos encorajou a optar apenas pelos testes de eletroforese em gel de poliacrilamida não desnaturante, tal como descrito no artigo 2 desta tese. Os resultados deste experimento não nos revelaram a massa molecular nos diferentes estados de oligomerização com a mesma precisão e elegância que seriam obtidos pelos experimentos de difusão em RMN, mas permitiu elucidar quais fatores combinados estabilizam a proteína alvo.

De posse de tais conhecimentos, mesmo que se tentasse novamente obter o espectro de RMN 2D ^1H ^{15}N HSQC da proteína marcada, colocando-a em presença de sal e DNA (oligonucleotídeo de 14 pb, $M_r \sim 6,6$ kDa), não seria possível obter um bom padrão de dispersão dos picos pois este seria de uma proteína maior que 15 kDa, devido à sua oligomerização, ultrapassando o limite de resolução do equipamento de RMN. Ainda assim, tentou-se obter o espectro de RMN 2D ^1H ^{15}N HSQC da proteína na presença de sais e DNA, a fim de se monitorar a alteração no padrão de dispersão dos sinais atribuídos aos quatro triptofanos da região C-terminal que, por homologia, corresponde ao domínio de ligação ao DNA. Caso o padrão de dispersão se alterasse, poder-se-ia então confirmar que os sinais eram provenientes da região C-terminal de ligação ao DNA e, assim, prosseguir com a resolução da estrutura tridimensional do domínio, mesmo que a dispersão das demais regiões da molécula não fosse satisfatória. Contudo, observou-se que a proteína marcada estocada a 4°C por três meses, havia perdido todo o sinal de ressonância, impossibilitando a análise, conduzindo-nos, entretanto, a um outro estudo: o da vida útil proteica. Ao medir novamente a concentração proteica a 280 nm, constatou-se uma redução significativa na leitura da absorção. Provavelmente, a proteína estaria se agregando, mas mantendo-se em suspensão, uma vez que não foi observada precipitação.

Paralelamente, começou-se a observar a inconsistência do valor da T_m obtida nos experimentos de desnaturação térmica monitorados por CD, para proteínas de diferentes lotes. Primeiramente constatou-se para um lote de proteína pura armazenada 4°C por três semanas uma T_m de 21°C. Ao se repetir o mesmo experimento para uma proteína pura recém-sintetizada (cinco dias armazenada a 4°C), a T_m foi de 52°C, corroborando os resultados dos homólogos para proteínas da família H-NS, que são termo-estáveis (vide

artigo 2). Esta diferença na estabilidade térmica proteica poderia estar ocorrendo pela perda de estrutura do lote de proteína pura após três semanas de armazenamento a 4°C. Novos lotes de proteína pura fresca foram sintetizados e submetidos ao experimento de desnaturação térmica por CD para confirmar a T_m , sendo confirmada a temperatura de 52°C em todos eles. Para garantir a confiabilidade e repetibilidade dos testes estruturais e funcionais, sempre utilizou-se proteína pura fresca e ainda fez-se a monitoração da vida útil da proteína pela varredura de seu espectro de absorção ao ultra-violeta na faixa de 210 a 340 nm (vide próximo tópico).

A resolução da parte enovelada da proteína correspondente à região C-terminal poderia ainda ser realizada por dois tipos de abordagens. Uma seria a separação do domínio C-terminal por proteólise limitada. O domínio C-terminal poderia ser obtido pela proteólise limitada quando a proteína recombinante estivesse adsorvida à coluna de afinidade Ni-NTA, de maneira similar ao descrito por Smyth *et al.* (2000). O domínio C-terminal seria então eluído no tampão de lavagem (*flow-through*) da resina e purificado posteriormente utilizando-se outros métodos cromatográficos. A massa molecular e a homogeneidade do domínio poderiam ser confirmados por SDS-PAGE e sua identidade confirmada por espectrometria de massa e/ou sequenciamento N-terminal. Contudo, dada à incerteza da clivagem do fragmento proteico na posição desejada, acredita-se que o rendimento final possa ser muito baixo. A proteína XF0749 possui ao todo 10 sítios teóricos de clivagem para tripsina, sendo que para outras proteases o número de sítios teóricos aumenta ainda mais. No entanto, Shindo *et al.* (1995) obtiveram o domínio C-terminal, por hidrólise da proteína H-NS de *E. coli* com tripsina bovina, em quantidades suficientes para resolução tridimensional por RMN (*2 D Proton NMR Spectroscopy* a 27°C).

O segundo tipo de abordagem consiste na expressão de proteínas mutantes de deleção N-terminal (predito corresponder dos resíduos 1 ao 78). Shindo *et al.* (1999) relataram que a proteína mutante H-NS⁶⁰⁻¹³⁷ apresentou total reversibilidade à desnaturação térmica e menor degradação após armazenamento por várias semanas à temperatura ambiente. Contudo, Smyth *et al.* (2000) verificaram que o domínio C-terminal isoladamente apresentou diminuição na atividade de ligação ao DNA. Tendo-se em vista esta abordagem, desenhou-se primers para duas proteínas mutantes de deleção N-terminal da ORF XF0749 de *X. fastidiosa* correspondendo a H-NS_Xf⁵⁶⁻¹³⁴ e H-NS_Xf⁶⁸⁻¹³⁴ (Fig. 13).

Os primers foram desenhados de modo a deixar alguns resíduos de aminoácidos excedentes na porção N-terminal do produto de PCR para assegurar a integridade do domínio C-terminal, pois apesar da predição prever o início do domínio C-terminal no resíduo Ser79, para algumas proteínas homólogas, o início deste domínio ocorre em posições anteriores (resíduo 56 para a proteína EcKorB de *E. coli*; 57 para a proteína HvrA de *Rhodobacter capsulatus*; 59 para a proteína RsSPB de *Rhodobacter sphaeroides*) (descrição dos primers no Quadro 3). A partir dos produtos de PCR de ambas versões de proteínas mutantes de deleção N-terminal, podem-se empregar todas as técnicas de biologia molecular, purificação proteica e caracterização funcional e estrutural, tal como utilizadas para a proteína íntegra.

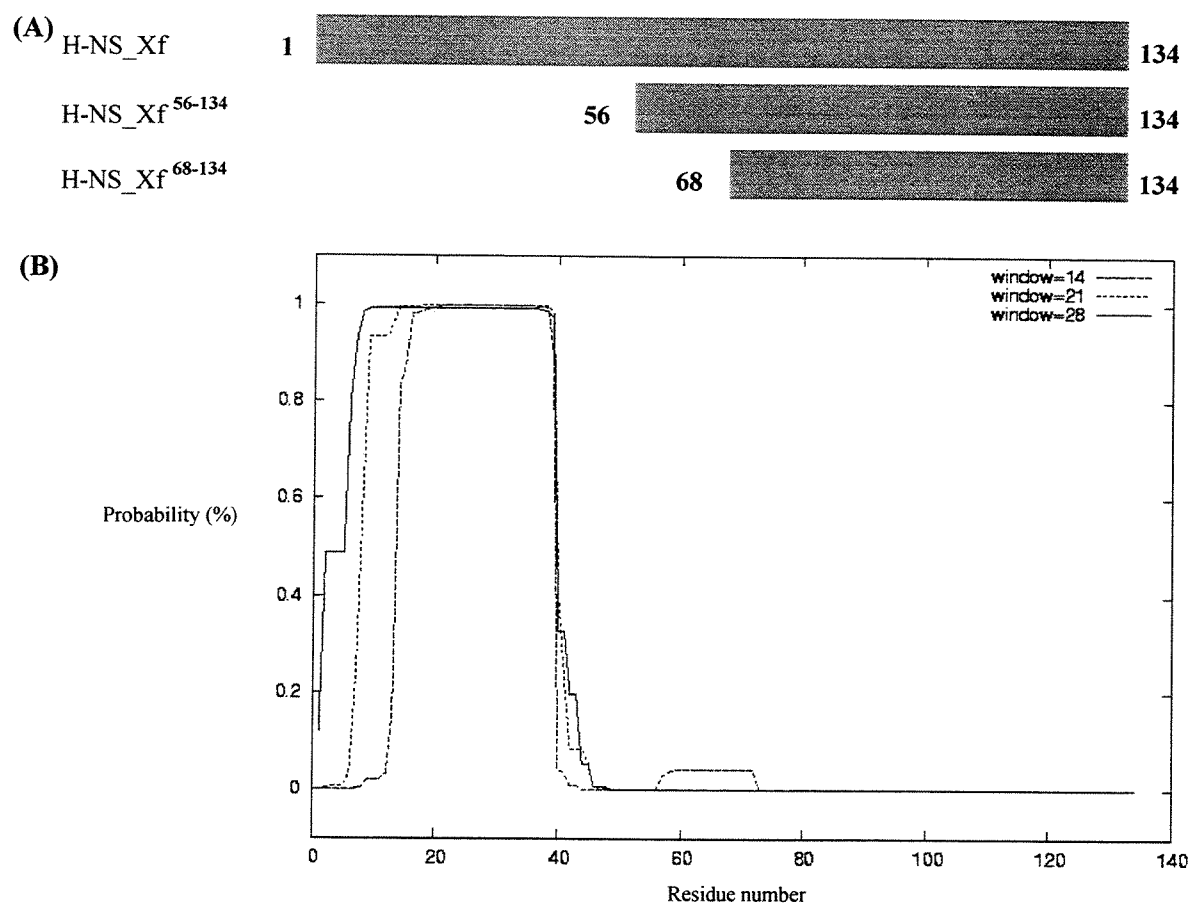


Fig. 13. (A) Ilustração da proteína XF0749 íntegra e das duas versões mutantes de deleção N-terminal; **(B)** Predição da estrutura *coiled-coil* usando-se o programa COILS (versão 2.1). O termo “window =14” se refere a representação de 14 resíduos de aminoácidos e assim por diante.

QUADRO 3. Primers para produção do domínio C-terminal da ORF XF0749. O nome do primer informa o vetor de clonagem (pET 32-Xa/LIC ou pET 28a/29a), o sentido da amplificação (“a” para *sense* e “b” para *antisense*), o aminoácido N-terminal (Phe56 e Val68) e os sítios de restrição *NdeI* (*sense*) e *EcoRI* (*antisense*).

NOME DO PRIMER	SEQÜÊNCIA DO PRIMER (5'→ 3')
XF05pET32a_Cdomain(Phe)	GGT ATT GAG GGT CGC TTC AAC GTT AAA CAG AAG
XF05pET32a_Cdomain(Val)	GGT ATT GAG GGT CGC GTC AAG GAG GTA ACACCT
XF05pET32b_Cdomain	AGA GGA GAG TTA GAG CCT CAG CCT GGG AAGGCT GG
XF05NdeI_Cdomain(Phe)	CGG CAG CCA TAT TTC AAC GTT AAA CAG AAG
XF05NdeI_Cdomain(Val)	CGG CAG CCA TAT GTC AAG GAG GTA ACA CCT
XF05EcoRI_Cdomain	GAG CTC GAA TTC TCA GCC TGG GAA GGC TGG

Não foi dada continuidade à obtenção do domínio C-terminal dado o elevado tempo de execução que excederia o prazo de 48 meses para a conclusão da tese.

6.5. Vida útil

Todos os aminoácidos absorvem na faixa UV de 180 a 220 nm, mas a cistina e os resíduos aromáticos fenilalanina (Phe), tirosina (Tyr) e triptofano (Try), absorvem significativamente luz UV a: 240 nm (cistina), 260 nm (Phe) e 280 nm (Tyr e Trp, principalmente) (Pace & Schmid, 1997). Contudo, absorção entre 310 e 340 nm começa a aumentar quando inicia-se agregação proteica, devido ao efeito do espalhamento de luz pelos agregados.

Para o teste de vida útil da proteína XF0749, aliquotou-se um lote de proteína pura de 15 μ M recém-obtida (cinco dias), em três soluções: 1) Água milli-Q; 2) 1 mM de tampão fosfato de sódio (pH 7); 3) 40 mM de tampão HEPES (pH 7,9) com 60 mM cloreto de potássio, 1 mM EDTA e DTT, 5 mM acetato de magnésio, 100 mM cloreto de sódio e 0,02 mM acetato de zinco. As três alíquotas foram estocadas a 4°C e fez-se a medida diária da absorção proteica por um mês no mesmo horário e equipamento (*DU640 Spectrophotometer Beckman Coulter*), subtraindo-se sempre o branco. O resultado desta monitoração apresenta-se na Fig. 14.

À medida que o tempo de armazenamento proteico a 4°C aumenta, há o deslocamento gradativo de absorção para a faixa de 310 a 340 nm, indicando a formação de agregados devido a perda de estrutura proteica com o tempo. O aumento da agregação não foi acompanhada pela precipitação visível da proteína, para o tempo de um mês de monitoração do experimento.

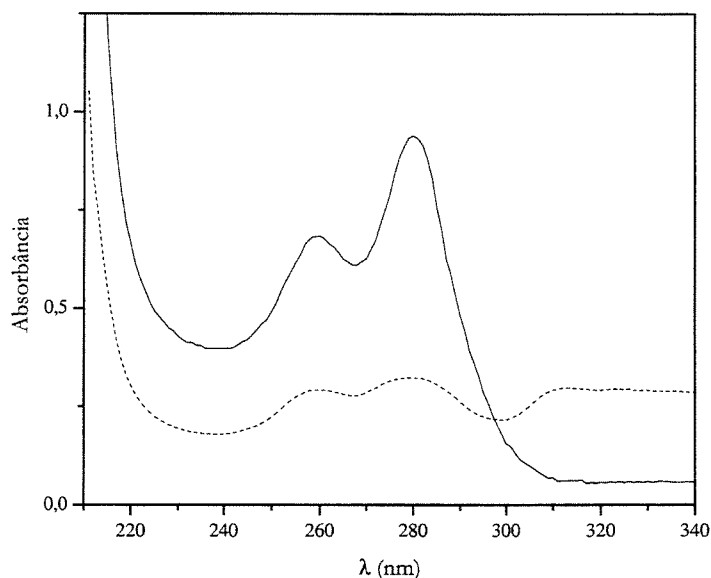


Fig. 14. Espectro de absorção da proteína XF0749 em tampão 1 mM de fosfato de sódio (pH 7) para avaliação de sua vida útil. A linha cheia representa o espectro da proteína recém-sintetizada (cinco dias) e a tracejada representa a última medida de absorção após um mês de armazenamento a 4°C.

O armazenamento a 4°C não conservou a estrutura proteica a longo prazo, independentemente da solução onde a proteína estivesse dissolvida. O armazenamento por períodos prolongados deve ser realizado na forma de pó seco pela liofilização ou na forma congelada a - 20°C. Contudo, ambas as formas apresentam desvantagens. Todas as vezes que a proteína pura foi submetida à liofilização, houve perda proteica pela não resolubilização de boa parte da amostra. No congelamento a - 20°C observou-se relativa perda de atividade (aproximadamente 10%) ao repetir-se os testes funcionais por espectroscopia de fluorescência (dados não apresentados) para a mesma amostra antes e após o congelamento. No entanto, acredita-se que a forma de armazenamento a longo prazo pelo congelamento a - 20°C seja a mais adequada para esta proteína. Não foram realizados testes de armazenamento por precipitação com sulfato de amônio (*salting-out*).

A ausência de pontes de dissulfeto na estrutura terciária da proteína XF0749, por não haver resíduos de cisteína na sequência primária, pode ser um dos fatores que favorece a instabilidade proteica. As pontes de dissulfeto reduzem o número das conformações intermediárias desenoveladas, aumentando a estabilidade do estado proteico nativo (Branden & Tooze, 1999). Entretanto, existem inúmeros relatos na literatura de proteínas estáveis sem pontes de dissulfeto, incluindo algumas de organismos termófilos.

Outra hipótese poderia ser a da co-purificação da proteína XF0749 com alguma protease de *E. coli* que, por estar em baixa concentração, não foi detectada em SDS-PAGE e nem no espectro de RMN 1 D. Dado ao grande número de sítios teóricos de clivagem na sequência primária da proteína XF0749 por diferentes proteases, a protease poderia atuar continuamente, justificando os resultados de: baixa vida útil (com perda de absorbância a 280 nm); diminuição gradual de sua concentração; não repetibilidade do valor da T_m nos ensaios de desnaturação térmica para proteínas de diferentes tempos de armazenamento; presença de bandas de massa molecular menor que 15 kDa em SDS-PAGE após armazenamento a 4°C ou após vários ciclos de descongelamento e congelamento a - 20°C; menor valor de massa molecular (cerca de 14,2 kDa) ao tentar-se fazer MALDI-TOF da proteína nativa e ao tentar estimar-se a massa molecular por cromatografia de Filtração em Gel. Uma maneira de se prevenir e/ou confirmar esta hipótese seria o armazenamento da proteína XF0749 na presença de algum inibidor de protease e, então, reavaliar todos os itens mencionados acima.

Experimentos como estes de vida útil e de otimização do armazenamento proteico, não foram encontrados na literatura para os homólogos. A única ressalva sobre precipitação proteica foi feita por Ueguchi *et al.* (1996), onde foi feito o relato de precipitação no

processo de diálise, sendo a proteína precipitada ressuspensa em tampão de agregação 10 mM de fosfato de sódio (pH 6) com 2 mM β -mercaptoetanol e ainda assim utilizada para testes funcionais. Também ocorreu a precipitação da proteína XF0749 nas trocas por diálise do tampão 50 mM Tris-clorato (pH 7,5) contendo 281,1 mM de cloreto de sódio, 182 mM de imidazol e 4,54 mM de cloreto de cálcio pelo tampão 20 mM de fosfato de sódio (pH 7,8). Esta precipitação foi atenuada pela retirada prévia do cloreto de cálcio pela diálise com 50 mM Tris-clorato (pH 7,5) antes da diálise definitiva com tampão fosfato. A proteína precipitada apresentou-se enovelada ao analisar seu espectro de estrutura secundária por CD (dado não apresentado), porém não foi utilizada para estudos estruturais e funcionais.

Estes resultados de vida útil desencorajam experimentos de cristalografia de raios-X para resolução da estrutura tridimensional e justificam a ausência na literatura de relatos de estudos de cristalografia para proteínas da família H-NS.

6.6. Expressão heteróloga de outras ORFs relacionadas à patogenicidade de X. fastidiosa

Além da ORF XF0749, o presente trabalho de doutorado inicialmente partiu da escolha de outras seis ORFs de patogenicidade descritas resumidamente no Quadro 4.

A numeração do nome da ORF corresponde à sua posição no mapa gênico do cromossomo principal da bactéria.

Todas as ORFs foram clonadas com sucesso no vetor de expressão pET 32-Xa/LIC. Não houve nenhuma dificuldade na transformação das linhagens de *E. coli* DH5 α , BL21(DE3) e BL21(DE3)pLysS. Do total das sete ORFs relacionadas à patogenicidade

preliminarmente selecionadas, apenas a ORF XF0749 expressou proteína solúvel. Todas as demais apresentaram baixa expressão heteróloga ou perfil insolúvel (Fig. 15 A).

QUADRO 4. ORFs de patogenicidade de *X. fastidiosa* selecionadas inicialmente para estudos funcionais.

ORF	PROVÁVEL FUNÇÃO PROTEICA	MASSA (kDa)
XF0175	Hemolisina III	23,8
XF0432	Família de virulência BrkB (<i>Bordetella resistant to killing</i>)	29,5
XF0591	Família gênica de fator de virulência	26,4
XF0754	Família gênica de virulência <i>acvB</i>	26,8
XF1020	Classe de proteínas PR (<i>Pathogenesis Related protein</i>)	18,7
XF2148	Dimetiladenosina transferase	32

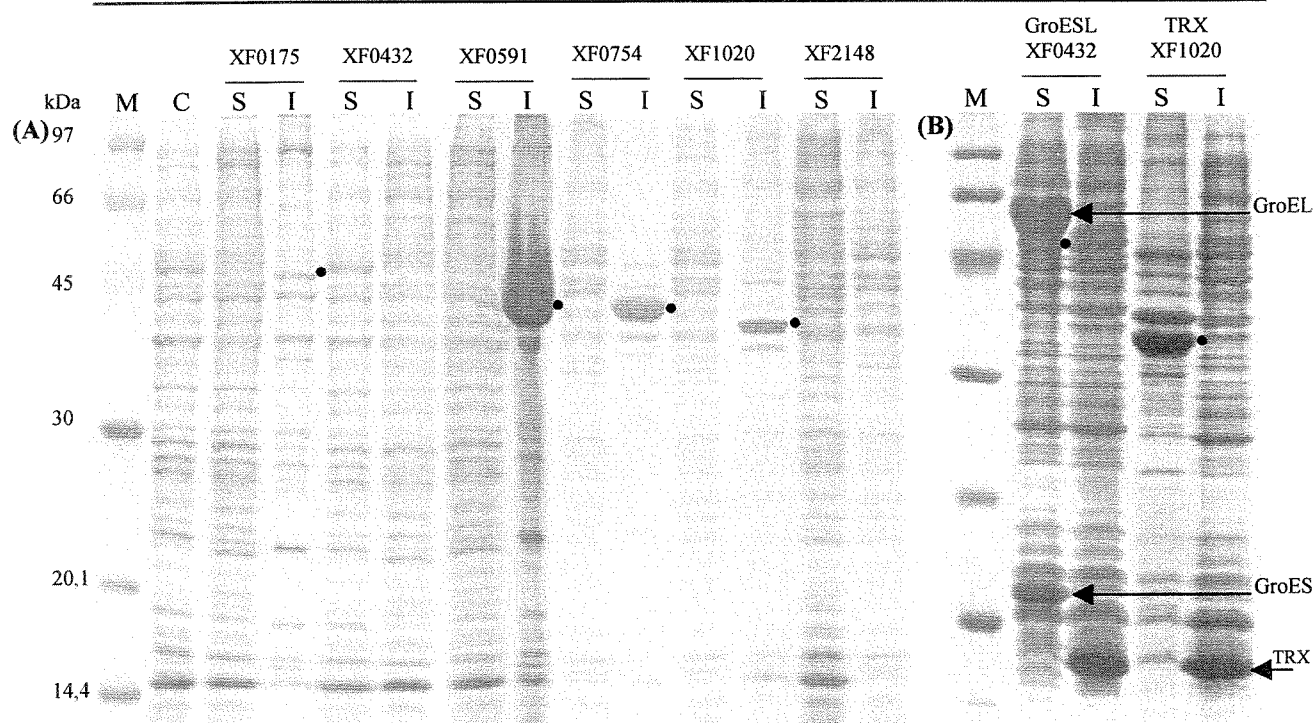


Fig. 15. SDS-PAGE (12%) da expressão heteróloga de seis ORFs relacionadas à patogenicidade de *X. fastidiosa*. **(A)** Trilha 1, marcador de massa molecular LMW (*Amersham Biosciences*), e trilha 2, extrato clarificado sem indução da expressão da proteína recombinante (controle). Trilhas 3 a 15, testes de expressão de cada ORF em *E. coli* BL21(DE3) induzido com 5,6 mM de lactose por 2 h a 37°C a 300 rpm, fração solúvel (S) e insolúvel (I), respectivamente. **(B)** Trilha 1, LMW. Trilhas 2 a 5, testes de expressão nas linhagens de *E. coli* BL21pT-GROE e BL21pT-TRX induzido com 5,6 mM de lactose por 2 h a 37°C a 300 rpm, fração solúvel (S) e insolúvel (I). Aplicou-se 6 µl de cada amostra no gel.

Na tentativa de obter maior número de proteínas solúveis e/ou de maior nível de expressão, testou-se outras linhagens de expressão de *E. coli* (BL21-SI, BL21pT-GROE e BL21pT-TRX), indução em temperaturas mais baixas (28 e 15°C), e em diferentes quantidades de indutor (tanto IPTG quanto lactose). Para a maioria das ORFs não obteve-se sucesso nestes testes (dados não apresentados). Apenas as ORFs XF0432 e XF1020 tiveram uma quantidade diferencial de expressão heteróloga ao se utilizar as linhagens BL21pT-GROE e BL21pT-TRX, respectivamente (Fig. 15 B). Estas linhagens contém o vetor pACYC-T7 resistente ao antibiótico cloranfenicol modificado para a co-expressão das chaperoninas GroES e GroEL, no caso da linhagem BL21pT-GROE, e de tireodoxina, no caso da linhagem BL21pT-TRX (Yasukawa *et al.*, 1995).

Tanto as chaperoninas GroEL e GroES, quanto a tireodoxina, auxiliam no enovelamento de proteínas citoplasmáticas. As chaperoninas formam um complexo que interage transitoriamente com alguns polipeptídeos de conformação intermediária desenovelada *in vivo*, promovendo sua correta isomerização e destinação subcelular. A formação de pontes de dissulfeto é desfavorecida no ambiente citoplasmático oxidativo de linhagens tipo selvagem de *E. coli*. Em linhagens de *E. coli* geneticamente modificadas, existem no mínimo cinco proteínas (tioredoxina 1 e 2, e glutaredoxina 1, 2 e 3, produtos dos genes *trxA*, *trxC*, *grxA*, *grxB* e *grxC*, respectivamente) envolvidas na redução das pontes de dissulfeto de enzimas heterólogas citoplasmáticas (Baneyx, 1999).

7. CONCLUSÕES

- A ORF XF0749 relacionada à patogenicidade de *X. fastidiosa* (clone 9a5c) codifica uma proteína citoplasmática de 134 resíduos de aminoácidos (15036,1 kDa) e pI 8. Testes funcionais e de estrutura secundária demonstram que a proteína XF0749 possui atividade de ligação a DNA, podendo pertencer à família *Histone-like Nucleoid Structuring Protein*, conforme predito pelos homólogos de estrutura primária.
- A proteína XF0749 revelou-se termo-resistente à desnaturação térmica e encontra-se predominantemente na forma tetramérica quando em condições bioquímicas semelhantes ao seu ambiente *in vivo*. O tipo de interação, tal como para proteínas desta família, parece ser hidrofóbica pelo potencial de formação da estrutura α -hélice anfipática *coiled-coil* no domínio N-terminal (resíduos 9 a 39).
- A proteína possui pouca sensibilidade à variações de pH, mas desnaturou a 7 M de uréia, indicando que seu domínio C-terminal de ligação ao DNA encontrava-se enovelado. Os resíduos de triptofano neste domínio parecem estar contidos numa região hidrofóbica, porém acessível aos supressores de DNA e acrilamida.
- A proteína XF0749 não foi passível de resolução da estrutura tridimensional por RMN nem por cristalografia de raios-X, haja visto que possui a propriedade de se oligomerizar e possui vida útil curta, a não ser que se fizessem testes de prolongamento da vida útil da proteína na presença de inibidor de protease e/ou de DNA.

- Em razão das evidências teóricas e experimentais apresentadas nesta tese, propomos a validação funcional e da anotação genômica da ORF XF0749 relacionada à patogenicidade de *X. fastidiosa* sendo que, de agora em diante, faremos referência a esta proteína como H-NS_Xf.

8. PERSPECTIVAS

O presente trabalho de doutorado sugere a possibilidade de se obter informação sobre a acessibilidade e o microambiente de cada um dos quatro Trp de HNS_Xf, realizando-se experimentos de Tempo-de-vida em Fluorescência (*Fluorescence Lifetime*), tais como *Time-Resolved Emission Spectroscopy* e/ou *Time-Resolved Anisotropy*. Outras análises da proteína H-NS_Xf podem ainda ser feitas por *Differential Scanning Calorimetry* (DSC) e *Chemical Cross-linking*.

O trabalho sugere também a obtenção da estrutura tridimensional do domínio C-terminal de ligação ao DNA (resíduos 79 a 122), por RMN 2 D ^1H ^{15}N HSQC, em colaboração com os pesquisadores do CEBiME - LNLS. A partir da estrutura tridimensional do domínio, poder-se-ia pela engenharia de proteínas, via mutagênese sítio-dirigida da ORF XF0749, modificar a estrutura terciária do domínio no propósito de melhor compreender a relação entre sua conformação e propriedade biológica.

A técnica de Microscopia de Força Atômica (*Atomic Force Microscopy* - AFM) está sendo otimizada para a visualização da compactação do plasmídeo relaxado pUC18 pela proteína H-NS_Xf, de acordo com Dame *et al.* (2000). Este trabalho está sendo realizado em parceria entre o Laboratório de Análise Genética e Molecular (CBMEG), Departamento de Genética e Evolução, Instituto de Biologia/UNICAMP e o Laboratório de Pesquisa em Dispositivos, Departamento de Física Aplicada, Instituto de Física Gleb Wataghin/UNICAMP.

O estudo funcional *in vivo* da proteína H-NS_Xf como fator transcricional pode ser realizado utilizando-se plasmídeo clonado com os genes repórteres para beta-galactosidase

(Ueguchi *et al.*, 1993) ou para luciferase (Owen-Hughes *et al.*, 1992). A avaliação da patogenicidade proteica pode ser realizada por meio da cultura de células de citros inoculada com *E. coli* que produza a proteína nativa. Em ambas as abordagens, uma linhagem de expressão de *E. coli* (p.ex. BL21(DE3)pLysS) deve conter um vetor para expressão citoplasmática da proteína XF0749 nativa (p.ex. o pET 29, primer reverso com *stop codon*).

Nos ensaios de plasmídio contendo genes repórteres, deve-se fazer a co-transformação bacteriana com qualquer plasmídio que contenha os genes *lacZ* (Fig. 16) ou *luxAB downstream* de algum promotor {p.ex. promotor *lac* – pUC; promotor *proU* – pBR322; promotor *T7* – pGEM-T; promotor *hns* – vetor pHOP-11 desenvolvido por Tanaka *et al.* (1994)}. A medida da atividade da beta-galactosidase pode ser feita de acordo com Miller (1972) e a da luciferase de acordo com Owen-Hughes *et al.* (1992).

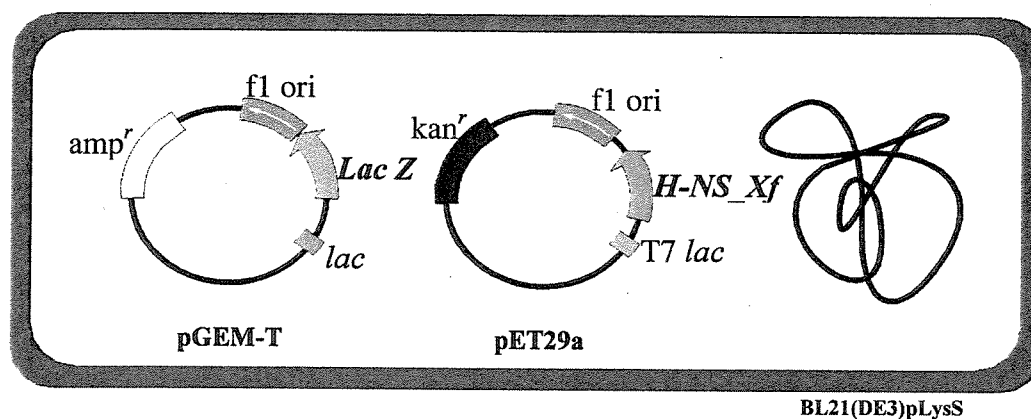


Fig. 16: Ilustração da linhagem BL21(DE3)pLysS co-transformada com plasmídios para expressão da proteína heteróloga e da enzima beta-galactosidase, para avaliação da atividade da proteína H-NS_Xf como fator transcricional.

O teste de patogenicidade da proteína H-NS_Xf na cultura de células de citros pode ser viabilizado em colaboração com o Laboratório de Bioquímica de Plantas, Departamento de Bioquímica, Instituto de Biologia/UNICAMP.

Pode-se ainda dar continuidade ao estudo funcional das seis ORFs relacionadas à patogenicidade preliminarmente selecionadas (Quadro 4), empregando-se outras estratégias de expressão, tais como mudança do vetor e tentativa de solubilização das proteínas insolúveis, testando-se protocolos com agentes desnaturantes e redutores. Outros vetores de expressão heteróloga podem melhorar a solubilidade proteica por conterem menor sequência de fusão (His₆ pET 28) ou sequências de fusão para peptídeos sinais N-terminal para secreção no espaço periplasmático (série pET 39 e 40b). Neste último caso, facilita-se a purificação e favorece-se o enovelamento por conter enzimas locais que catalizam a formação e isomerização das pontes de dissulfeto. Além disto, pode-se reduzir proteólises e prevenir o acúmulo de proteínas de caráter tóxico em alta concentração citoplasmática. A purificação proteica também pode ser favorecida utilizando-se vetores que contenham sequências de fusão de alta afinidade a determinados ligantes, tais como GST da séries pET 41a-c e 42a-c ou pGEX, e MBP da série pMAL.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, F. C. L.; Amorim, G. C.; Moreau, V. H.; Sousa, V. O.; Creazola, A. T.; Américo, T. A.; Pais, A. P. N.; Leite, A.; Netto, L. E. S.; Giordano, R. J.; Valente, A. P.** 2001. Selectively labeling the heterologous protein in *E. coli* for NMR studies: A strategy to speed up NMR spectroscopy. *J. Mag. Reson.* 148: 142-146.
- Azam, T. A.; Iwata, A.; Nishimura, A.; Ueda, S.; Ishihama, A.** 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181: 6361-6370.
- Badaut, C.; Williams, R.; Arluison, V.; Bouffartigues, E.; Robert, B.; Buc, H.; Rimsky, V.** 2002. The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *J. Biol. Chem.* 277: 41657-41666.
- Baneyx, F.** 1999. Recombinant protein expression in *E. coli*. *Current Opinion in Biotechnology*, 10: 411-421.
- Bertin, P.; Benhabiles, N.; Krin, E.; Laurent-Winter, C.; Tendeng, C.; Turlin, E.; Thomas, A.; Danchin, A.; Brasseur, R.** 1999. The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. *Mol. Microbiol.* 31: 319-329.
- Bhattacharyya, A.; Stilwagen, S.; Ivanova, N.; D'Souza, M.; Bernal, A.; Lykidis, A.; Kapatral, V.; Anderson, I.; Larsen, N.; Los, T.; Reznik, G.; Selkov, E.; Walunas, T. L.; Feil, H.; Feil, W. S.; Purcell, A.; Lassez, J.-L.; Hawkins, T. L.; Haselkorn, R.; Overbeek, R.; Predki, P. F.; Kyrpides, N. C.** 2002 a. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc. Natl. Acad. Sci. USA.* 99: 12403-12408.
- Bhattacharyya, A.; Stilwagen, S.; Ivanova, N.; D'Souza, M.; Bernal, A.; Lykidis, A.; Kapatral, V.; Anderson, I.; Larsen, N.; Los, T.; Reznik, G.; Selkov, E.; Walunas, T. L.; Feil, H.; Feil, W. S.; Purcell, A.; Lassez, J.-L.; Hawkins, T. L.; Haselkorn, R.; Overbeek, R.; Predki, P. F.; Kyrpides, N. C.** 2002 b. Draft sequencing and comparative

- genomics of *Xylella fastidiosa* strains reveal novel biological insights. *Gen. Res.* 12: 1556-1563.
- Bracco, L.; Kotlarz, D.; Kolb, A.; Diekmann, S.; Buc, H.** 1989. Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*. *EMBO J.* 8 : 4289-4296.
- Braden, C.; Tooze, J.** Prediction, engineering and design of protein structures. In: Introduction to protein structure. 2nd ed., Chapter 17, Garland Publising, New York, 1999, pp. 347-380.
- Ceschini, S.; Lupidi, G.; Coletta, M.; Pon, C.L.; Fioretti, E.; Angeletti, M.** 2000. Multimeric self-assembly equilibrium involving the Histone-like protein H-NS. *J. Biol. Chem.* 275: 729-734.
- Chang, C. J.; Garnier, M.; Zreik, L.; Rossetti, V.; Bove, 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*. *Cur. Microbiol.* 27: 137 – 142.
- Chothia, C.; Lesk, A. M.** 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5: 823-826.
- Dame, R. T.; Wyman, C.; Goosen, N.** 2000. H-NS mediated compaction of DNA visualized by Atomic Force Microscopy. *Nucleic Acids Res.* 28: 3504-3510.
- Dersch, P.; Schmid, K.; Bremer, E.** 1993. Synthesis of the *Escherichia coli* K12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. *Mol. Microbiol.* 8: 875-889.
- Dow, J. M.; Daniels, M. J.** 2000. *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* 17: 263-271.
- Falconi, M.; Gualtieri, M. T.; La Teana, A.; Losso, M. A.; Pon, C. L.** 1988. Proteins from the prokaryotic nucleoid: primary and quarternary structure of the 15-kD *Escherichia coli* DNA binding protein H-NS. *Mol. Microbiol.* 2: 323-329.

- Falconi, M.; Higgins, N. P.; Spurio, R.; Pon, C. L.; Gualerzi, C. O.** 1993. Expression of the gene encoding the major bacterial nucleotide protein H-NS is subject to transcriptional auto-repression. *Mol. Microbiol.* 10: 273-282.
- Friedrich, K.; Gualerzi, C. O.; Lammi, M.; Losso, M. A.; Pon, C. L.** 1988. Proteins from the prokaryotic nucleoid. *FEBS Lett.* 229: 197-202.
- Gibrat, J. F.; Madej, T.; Bryant, S. H.** 1996. Surprising similarities in structure comparison. *Curr. Opin. Struc. Biol.* 6: 377-385.
- Goheen, A. C.; Nyland, G.; Lowe, S. K.** 1973. Association of a rickettsia-like organism with Pierce's disease of grapevines and alfalfa dwarf and heat therapy of disease in grapevines. *Phytopathology* 63: 341-345.
- Griffiths, A. J. F.; Miller, J. H.; Suzuki, D. T.; Lewontin, R. C.; Gelbart, W. M.** Genomics. In: **An introduction to genetic analysis**. 7thed., Chapter 14, W. H. Freeman, New York, 2000, pp. 436-460.
- Grossman, T. H.; Kawasaki, E. S.; Punreddy, S. R.; Osburne, M. S.** 1998. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* 209: 95-103.
- Hopkins, D. L.** 1989. *Xylella fastidiosa* – Xylem-limited bacterial pathogen of plants. *Annu. Rev. Phytopathol.* 27: 271-290.
- Hopkins, D. L.; Mollenhauer, H. H.** 1973. Rickettsia-like bacterium associated with Pierce's Disease of grapes. *Science* 179: 298-300.
- Hopkins, D. L.; Purcell, A. H.** 2002. *Xylella fastidiosa*: cause of Pierce's disease of grapevine and other emergent diseases. *Plant Dis.* 86: 1056-1066.
- Hulton, C. S. J.; Seirafi, A.; Hinton, J. C. D.; Sidebotham, J. M.; Waddell, L.; Pavitt, G. D.; Owen-Hughes, T.; Spassky, A.; Buc, H.; Higgins, C. F.** 1990. Histone-like protein H1 (H-NS), DNA supercoiling and gene expression in bacteria. *Cell* 63: 631-642.
- Jindal, K. K.; Sharma, R. C.** 1987. Outbreaks and new records: India almond leaf scorch, a new disease from India. *FAO Plant Protec. B.* 35: 64-65.

- La Teana, A.; Falconi, M.; Scarlato, V.; Lammi, M.; Pon, C. L.** 1986. Characterization of the structural genes for the DNA-binding protein H-NS in Enterobacteriaceae. *FEBS Lett.* 244: 34-38.
- La Teana, A.; Brandi, A.; Spurio, R. M.; Pon, C. L.; Gualerzi, C. O.** 1991. Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS. *Proc. Natl. Acad. Sci. USA.* 88: 10907-10911.
- Laranjeira, F. F.** 1997. Ten years of Citrus Variegated Chlorosis: What do we know? *Laranja* 18: 123-141.
- Lee, R. F. ; Beretta, M. J. G.; Derrick, K. S.; Hooker, M. E.** 1992. Development of a serological assay for citrus variegated chlorosis: a new disease of citrus in Brazil. *Proc. Fla. State Hortic. Soc.* 105: 32-35.
- Lee, R. F.; Beretta, M. J. G.; Hartung, J. H., Hooker, M. E. and Derrick, K. S.** 1993. Citrus Variegated Chlorosis: confirmation of *Xylella fastidiosa* as the causal agent. *Sum. Phytopathol.* 19: 123-125.
- Lee, R. F.; Derrick, K. S.; Beretta, M. J. G.; Chagas, C. M.; Rosetti V.** 1991. Citrus Variegated Chlorosis: a new destructive disease of citrus in Brazil. *Citrus Ind.* pp. 12-15.
- Leu, L. S.; Su, C. C.** 1993. Isolation, cultivation and pathogenicity of *Xylella fastidiosa*, the causal bacterium of pear leaf scorch in Taiwan. *Plant Dis.* 77: 642-646.
- Machado, M.; Souza, A. A.; Coletta Filho, H. D.; Kuramae, E. E.; Takita, M. A.** 2001. Genome and pathogenicity of *Xylella fastidiosa*. *Mol. Biol. Today* 2: 33-43.
- Miller, J. H.** Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. 1972. New York, USA, pp. 352-355.
- Mircetich, S. M.; Lowe, S. K.; Moller, W. J.; Nyland, G.** 1976. Etiology of almond leaf scorch disease and transmission of causal agent. *Phytopathology* 66: 17-24.
- Moffart, B. A.; Studier, F. W.** 1987. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189: 113-130.
- Oliveira, R. C.; Yanai, G.; Muto, N. H.; Leite, D. B.; Souza, A. A.; Coletta-Filho, H. D.; Machado, M. A.; Nunes, L. R.** 2002. Competitive hybridization on spotted

- microarrays as a tool to conduct comparative genomic analysis of *Xylella fastidiosa* strains. *FEMS Microbiol. Lett.* 216: 15-21.
- Owen-Hughes, T. A.; Pavitt, G. D.; Santos, D. S.; Sidebotham, J. M.; Hulton, C. S. J.; Hinton, J. C. D.; Higgins, C. F.** 1992. The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* 71: 255-265.
- Pace, C. N.; Schmid, F. X.** 1997. How to determine the molar absorbance coefficient of a protein in "Protein structure, a practical approach", Creighton, T. E. 2nd ed., IRL Press, New York, pp. 253-259.
- Paradela-Filho, O.; Sugimori, M. H.; Ribeiro, I. J. A.; Garcia Junior, A.; Beretta, M. J. G.; Harakawa, R.; Machado, M. A.; Laranjeira, F. F.; Rodrigues Neto, J.; Beriam, L. O.** 1997. Constatação de *Xylella fastidiosa* em cafeeiros no Brasil. *Sum. Phytopathol.* 23: 46-49.
- Purcell, A. H.; Hopkins, D. L.** 1996. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* 34: 131 – 151.
- Purcell, A. H.; Saunders, S. R.; Henderson, M.; Grebus, M. E.; Henry, M. J.** 1999. Causal role of *Xylella fastidiosa* in oleander leaf scorch disease. *Phytopathology* 89: 53-58.
- Queiroz-Voltan, R. B.; Paradela-Filho, O.** 1999. Caracterização de estruturas anatômicas de citros infectados com *Xylella fastidiosa*. *Laranja* 20: 55-76.
- Raju, B. C.; Wells, J. M.** 1986. Diseases caused by fastidious xylem-limited bacteria and strategies for management. *Plant Dis.* 70: 182-186.
- Roberto, S. R.; Coutinho, A.; Delima, J. E. O.; Miranda, V. S.; Carlos, E. F.** 1996. Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Dilobopterus costalimai*, *Ocrogonia terminalis* e *Oncometopia facielis* em citros. *Fitopatologia Brasileira* 21: 517-518.
- Sali, A.; Kuriyan, J.** 1999. Challenges at the frontiers of structural biology. *Trends in Cell Biol.* 9: M20-M24.

- Sambrook, J.; Russell, D.W.** Plasmids and their usefulness in molecular cloning, in: *Molecular Cloning, a Laboratory Manual*, 3rd ed., vol. 1, CSHL Press, New York, 2001, pp. 1.32-1.34.
- Shindo, H.; Ohnuki, A.; Ginba, H.; Kato, V; Ueguchi, C.; Mizuno, T.; Yamazaki, T.** 1999. Identification of the DNA binding surface of H-NS protein from *E. coli* by heteronuclear NMR spectroscopy. *FEBS Lett.* 455: 63-69.
- Simpson, A. J. G. et al.** 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406: 151-157.
- Smyth, C. P.; Lundbäck, T.; Renzoni, D.; Siligardi, G.; Beavil, R.; Layton, M.; Sidebotham, J. M.; Hinton, J. C. D; Driscoll, P. C.; Higgins, C. F.; Ladbury, J. E.** 2000. Oligomerization of the chromatin-structuring protein H-NS. *Mol. Microbiol.* 36: 962-972.
- Spassky, A.; Rimsky, S.; Garreau, H.; Buc, H.** 1984. H1a, an *E. coli* DNA binding protein which accumulates in stationary phase, strongly compacts DNA *in vitro*. *Nucleic Acids Res.* 12: 5321-5340.
- Spurio, R.; Durrenberger, M.; Falconi, M.; La Teana, A.; Pon, C. L.; Gualerzi, C. O.** 1992. Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Mol. Gen. Genet.* 231: 201-211.
- Spurio, R.; Falconi, M.; Brandi, A.; Pon, C. L.; Gualerzi, C. O.** 1997. The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO J.* 16: 1795-1805.
- Suzuki, D. T.; Griffiths, A. J. F.; Miller, J. H.; Lewontin, R. C.** 1992. A estrutura e função dos cromossomos em eucariontes. In: **Introdução à genética**. Cap. 14. Guanabara Koogan. 4^a ed., Rio de Janeiro – RJ, pp. 297-316.
- Tanaka, K.; Yamada, H.; Yoshida, T.; Mizuno, T.** 1991. Overproduction and rapid purification of the *E. coli* Histone-like protein, H-NS. *Agric. Biol. Chem.* 55: 3139-3141.
- Tubelis, A.** 1992. Difusão da clorose em pomares de São Paulo e Minas Gerais. *Inform. Cooper.* 72: 24-30.

- Tubelis, A.; Barros, J. C.; Leite, R. M. V. B. C.** 1993. Difusão da Clorose Variegada dos Citros em pomares comerciais de laranja no Brasil. *Laranja* 14: 239-254.
- Ueguchi, C.; Kakeda, M.; Mizuno, T.** 1993. Autoregulatory expression of the *Escherichia coli hns* gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription. *Mol. Gen. Genet.* 236: 171-178.
- Ueguchi, C.; Suzuki, T.; Yoshida, T.; Tanaka, K.; Mizuno, T.** 1996. Systematic mutational analysis revealing the functional domain organization of *Escherichia coli* nucleoid protein H-NS. *J. Mol. Biol.* 263: 149-162.
- Ueguchi, C., Seto, C.; Suzuki, T.; Mizuno, T.** 1997. Clarification of the dimerization domain and its functional significance for the *Escherichia coli* nucleoid protein H-NS. *J. Mol. Biol.* 274: 145-151.
- Ussery, D. W.; Hinton, J. C.; Jordi, B. J.; Granum, P. E; Seirafi, A.; Stephen, R. J.; Tupper, A. E.; Berridge, G.; Sidebotham, J. M.; Higgins, C. F.** 1994. The chromatin-associated protein H-NS. *Biochimie* 76: 968-980.
- Van Sluys et al.** 2003. Comparative analyses of the complete genome sequences of Pierce's disease and Citrus Variegated Chlorosis strains of *X. fastidiosa*. *J. Bacteriol.* 185: 1018-1026.
- Varshavsky, A. J.; Nedospasov, S. A.; Bakayev, V. V.; Bakayeva, T. G.; Georgiev, G. P.** 1977. Histone-like proteins in the purified *Escherichia coli* deoxyribonucleo-protein. *Nucleic Acids Res.* 4: 2725-2745.
- Wells, J. M.; Raju, B. C.; Hung, H. Y.; Weisburg, W. G.; Mandelcopaul, L.; Brenner, D. J.** 1987. *Xylella fastidiosa*, Gram negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. *Int. J. Syst. Bacteriol.* 37: 136 – 143.
- Williams, R.M.; Rimsky, S.** 1997. Molecular aspects of the *E. coli* nucleoid protein, H-NS: a central controller of gene regulatory networks. *FEMS Microbiol. Lett.* 156: 175-185.

- Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J.** 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 38: 16424-16431.
- Yamada, H.; Muramatsu, S.; Mizuno, T.** 1990. An *Escherichia coli* protein that preferentially binds to sharply curved DNA. *J. Biochem.* 108: 420-425.
- Yasukawa, T.; Kanei-Ishii C.; Maekawa T.; Fujimoto J.; Yamamoto T.; Ishii S.** 1995. Increase of solubility of foreign proteins in *E. coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* 270: 25328-25331.
- Zarembinski, T. I.; Hung, L. W.; Mueller-Dieckmann, H. J.; Kim, K. K.; Yokota, H.; Kim, R.; Kim, S. H.** 1998. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc. Natl. Acad. Sci. USA.* 95: 15189-15193.

