7.5 4.000
10/03
Recebido em

UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

**HrcA de *Caulobacter crescentus* e *Xylella fastidiosa*:
estudos comparativos de seqüências e desenvolvimento
de modelo estrutural.**

DISSERTAÇÃO DE MESTRADO

HUMBERTO RODRIGUEZ PEREZ

ORIENTADORA: PROF^a DR^a SUELY LOPES GOMES

SÃO PAULO
2002

T
5
P438h

DEDALUS - Acervo - CQ

30100005277

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Perez, Humberto Rodriguez

P438h HrcA de *Caulobacter crescentus* e *Xylella fastidiosa*: estudos comparativos de seqüências e desenvolvimento de modelo estrutural / Humberto Rodriguez Perez. -- São Paulo, 2002. 117p.

Dissertação (mestrado) – Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.

Orientador : Gomes, Suely Lopes

1. Proteína : Estrutura : Bioquímica 2. Espectroscopia : Bioquímica analítica 3. Proteína : Purificação : Biossíntese
I. T. II. Gomes, Suely Lopes, orientador.

574.19245 CDD

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

20.383
“HrcA de *Caulobacter crescentus* e *Xylella fastidiosa*: Estudos Comparativos de Sequências e Desenvolvimento de Modelo Estrutural”

HUMBERTO RODRIGUES PEREZ

Dissertação de Mestrado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Mestre em Ciências - Área: Bioquímica.

Aprovado por:

Profa. Dra. SUELY LOPES GOMES
IQ – USP
(Orientadora e Presidente)

Profa. Dra. WLADIA VIVIANI
IQ – USP

Profa. Dra. ANA PAULA ULIAN DE ARAUJO
IF – SC – USP

SÃO PAULO
28 DE NOVEMBRO 2002.

À minha família e especialmente à
minha mãe, que além de dar amor e
carinho soube ensinar seus filhos o
valor de se ter princípios.

AGRADECIMENTOS:

À Dra. Suely Lopes Gomes pela oportunidade de elaborar este trabalho, e pela orientação dedicada e competente.

A minha esposa e companheira Luciana, cuja força foi essencial na superação dos momentos difíceis.

Aos doutores Shirley Schreier, Cláudio Shida, Maurício S. Baptista, e aos doutorandos Michelle F. Susin e Fernando A. Genta, pelas frutíferas colaborações. Que elas possam continuar e gerar cada vez mais frutos por um longo tempo ainda.

A minha família, que a mais de dois mil quilômetros de distância proporcionou todo o apoio e suporte possível, e até o impossível, para que eu concretizasse este trabalho.

Aos amigos do laboratório, Sandra, Rita, Luciano Fietto, Michelle, Luciana, Cristina, Tie, Karina, Raphaela, Sílvia, Tonho e Terumi, pela orientação, ajuda, mas sobretudo pela sincera amizade.

Aos amigos Raquel, Alcides, Fabiana, Fernando, Juliana, Denise, Léo, Marcelo, Gilberto, Cris, PC, Ariel, e todos os demais conhecidos ao longo deste caminho, pela amizade e agradável convivência. Afinal a vida não é feita só de trabalho.

À FAPESP pelo suporte financeiro que possibilitou a realização deste trabalho.

RESUMO

O gene *hrcA* é encontrado em quase todos os ramos da árvore filogenética das eubactérias, e seu produto, a proteína HrcA, funciona como repressor da expressão dos operons de choque térmico *groESL* e *dnaKJ*, ligando-se à sequência repetida invertida denominada CIRCE (*c*ontrolling *i*nverted *r*epet of *c*haperonin *e*xpression) presente na região regulatória destes operons. O sistema HrcA-CIRCE está, portanto, amplamente representado nas eubactérias. Particularmente, em *Caulobacter crescentus*, uma α -proteobactéria, este sistema está envolvido no controle da expressão do operon *groESL* durante o ciclo celular da bactéria. Conhecer a estrutura e as interações de HrcA é importante para entender este processo. Neste trabalho são apresentadas as análises de sequência das HrcA's de *C. crescentus* e de *Xylella fastidiosa*, uma proteobactéria do grupo γ , as quais são muito similares. Este estudo levou à proposta de um modelo estrutural com a delimitação dos domínios da proteína, os dobramentos de cada domínio, com base nas interações da HrcA de *C. crescentus* com o elemento CIRCE e ATP, que estão sendo caracterizadas em nosso laboratório, assim como a atribuição de aminoácidos e motivos conservados funcionais. Adicionalmente, embora a expressão da HrcA recombinante de *X. fastidiosa* não tenha tido sucesso, a HrcA recombinante de *C. crescentus* purificada tem se prestado aos ensaios espectroscópicos, ainda que tenha sido detectada uma microagregação que está sendo enfrentada com um protocolo de purificação baseado no uso de α ciclodextrina. Os estudos espectroscópicos preliminares da HrcA *C. crescentus* dão suporte ao modelo estrutural proposto.

ABSTRACT

The *hrcA* gene is found in almost all branches of the filogenetic tree of eubacteria, and its product, the protein HrcA, functions as a repressor regulating the expression of the heat shock operons *groESL* and *dnaKJ*, by binding to the inverted repeat sequence called CIRCE (*c*ontrolling *i*nverted *r*epet of *c*haperonin *e*xpression). The system HrcA-CIRCE, therefore, is widely represented in eubacteria. Specifically in *Caulobacter crescentus*, an α -proteobacterium, this system is involved in the cell-cycle control of *groESL* expression (Baldini *et al*, 1998). Knowledge of the structure of HrcA and its interactions is important to understand this process. This work presents the analysis of the sequences of HrcA from *C. crescentus* and *Xylella fastidiosa*, a proteobacterium of the γ group, which are very similar. A structural model has been proposed, with protein domain delimitation, specific domain folding, based on known interactions of *C. crescentus* HrcA with the CIRCE element and ATP, obtained in our laboratory, as well as assignment of functional residues and conserved motifs. Additionally, even though no success was obtained the expression of recombinant HrcA from *X. fastidiosa*, purified recombinant HrcA from *C. crescentus* has been shown to be suitable for spectroscopic studies, in spite of microagregation observed, which is being faced with a purification protocol based on the use of α cyclodextrin. The preliminary spectroscopic studies of HrcA from *C. crescentus* support the proposed structural model.

ABREVIATURAS

- [α - ^{32}P] dCTP - 5'-trifosfato de desoxicitidina marcado radiativamente na posição α com ^{32}P
- ATP - 5'-trifosfato de adenosina
- BCIP - 5-bromo-4-cloro-3-fosfato indol
- BSA - albulmina sérica bovina
- CAGE - Cooperação para Análise de Genes e sua Expressão
- CBS - Cistationeina Beta Sintase
- Ci, μCi - Curie, microCurie
- CIRCE - repetição invertida controladora da expressão de chaperoninas (Controlling Inverted Repeat of Chaperonin Expression)
- CRP - proteína repressora de catabólitos (Catabolite Repressor Protein)
- C-terminal - terminal carboxila
- Da, kDa - dalton, quilodalton
- DNA - ácido desoxirribonucléico
- dNTP - 5'-trifosfato de 2'-desoxinucleosídeo
- D.O. - densidade ótica
- DOC - desoxicolato de sódio
- DUF - domínio de função desconhecida (Domain of Unknown Function)
- EDTA - ácido etileno diamino tetraacético
- F, Fo - fluorescência, fluorescência inicial
- g, mg, ng - grama, miligrama, nanograma
- GCG - Genetics Computer Group
- Hsp - proteína de choque térmico (Heat Shock Protein)
- HTH - hélice alça hélice (Helix Turn Helix)
- IPTG - isopropil β -D-tiogalactosídeo
- kb - quilobases
- M, mM - molar, milimolar
- mm, nm - milímetro, nanometro
- min, seg - minuto, segundo
- ml, μl - mililitros, microlitros

MW - massa molecular

Ni-NTA - complexo níquel-ácido nitrilotriacético

NTB - nitro-azul de tetrazólio

N-terminal - terminal amino

OB, *OB Fold* - [dobramento] ligador de oligonucleotídeos/oligossacarídeos

PCR - reação em cadeia da polimerase

PMSF - fluoreto de fenilmetilsulfonila

RNA, mRNA - ácido ribonucleico, RNA mensageiro

SDS - dodecil sulfato de sódio

SSC - Citrato trissódico-Cloreto de sódio

TBE - Tris-Borato-EDTA

TBS - TrisHCl-Cloreto de sódio

TBS-Leite - TBS-Leite desnatado

TBS-T - TBS-Tween20

TBS-TT - TBS-Tween20-Triton X-100

UDP - 5'-bifosfato de uridina

UV - ultra-violeta

x g - vezes a aceleração da gravidade

X-gal - 5-bromo-4-cloro-3- β -D-galactosídeo indol

ÍNDICE

	página
Resumo	I
Abstract	II
Abreviaturas	III
Índice	V
Introdução	1
O modelo <i>Caulobacter crescentus</i>	1
Proteínas de resposta ao choque térmico	4
Mecanismos de regulação da resposta ao choque térmico em bactérias	6
Regulação da resposta ao choque térmico em <i>Escherichia coli</i>	8
Regulação da resposta ao choque térmico em <i>Bacillus subtilis</i>	10
Regulação da resposta ao choque térmico em <i>Caulobacter crescentus</i>	11
O sistema CIRCE/HrcA	15
HrcA em <i>Xylella fastidiosa</i>	18
Estrutura das HrcA's: predições, oligomerias e domínios	19
Objetivos	24
Materiais e Métodos	25
Linhagens e plasmídeos	25

Meios e condições de	26
Amplificação de DNA por reação em cadeia da polimerase	26
Isolamento de DNA, digestão com enzimas de restrição, ligação, sequenciamento e transformações	28
Eletroforese de proteínas e "Western Blots"	29
Expressão e purificação da proteína HrcA-His	30
Marcação de sondas e hibridação	34
Dicroísmo circular	35
Fluorescência	36
Espalhamento de luz dinâmico com laser	36
FPLC	37
Métodos computacionais	38
Resultados e Discussão	41
Análises de seqüências: análise geral, motivo HTH, e domínio amino-terminal	41
Análises de seqüências: domínio carbóxi-terminal	62
Clonagem do gene <i>hrcA</i> de <i>Xyllela fastidiosa</i> em vetor de expressão de <i>E. coli</i>	76
Expressão e purificação de HrcA de <i>Xyllela fastidiosa</i>	78
Expressão e purificação de HrcA de <i>Caulobacter crescentus</i>	82
Dicroísmo circular	84
Espectros de fluorescência	92
Espalhamento de luz dinâmico	96
FPLC	99

Considerações Finais e Conclusão	102
Bibliografia	105
Anexos	115
Anexo 1	115
Anexo2	116
<i>Curriculum Vitae</i>	117

INTRODUÇÃO

O modelo *Caulobacter crescentus*

O grupo das caulobactérias, representado principalmente pelos gêneros *Caulobacter*, *Brevundimonas* e *Asticcacaulis*, é constituído por proteobactérias Gram-negativas do grupo α . São organismos aeróbios obrigatórios, encontrados em praticamente todos os ambientes aquáticos, desde solução intersticial de muitos tipos de solo, passando por mananciais de água doce, até habitats marinhos. Os ambientes onde as caulobactérias são usualmente encontradas têm pH próximo da neutralidade e são pobres em nutrientes. Elas são um dos principais componentes do biofilme que se forma na interface água-ar em água não perturbada. Por sua baixa exigência nutricional elas já foram isoladas até de reservatórios de água destilada (Poindexter, 1981). No seu habitat oligotrófico as caulobactérias são importantes na reciclagem de carbono, usando fontes deste elemento, encontradas em baixas concentrações no ambiente, para sua nutrição e liberando-o sob a forma de CO₂ posteriormente. Apesar de sua ampla distribuição e importância ecológica, a maior parte dos estudos sobre estas bactérias tem explorado seu peculiar ciclo de vida heteromórfico, que inclui uma divisão celular assimétrica que dá origem a células filhas distintas em forma e fisiologia (Poindexter, 1981). O grupo das caulobactérias tem como particularidade morfológica, em um dos estágios de seu ciclo celular, a presença de um apêndice derivado da parede celular denominado prosteca ou, quando há presença de material adesivo em sua extremidade, talo. O talo no grupo das caulobactérias apresenta bandas a intervalos irregulares e é constituído de parede celular, membrana plasmática, e citoplasma, sendo este desprovido de ribossomos ou DNA, tratando-se portanto de uma

estrutura viva; especula-se que *Brevundimonas diminuta*, que aparentemente não apresenta talo, tenha perdido a capacidade de formá-lo ao longo da sua evolução, ou tenha sido bloqueada permanentemente na fase móvel do ciclo celular (Stahl *et al.*, 1992).

Caulobacter crescentus é o principal modelo de estudo para diferenciação celular em bactérias Gram-negativas, cujos mecanismos moleculares vêm sendo extensivamente detalhados nos últimos anos. Trata-se de um organismo dulcícola, mesofílico e de vida livre que, para completar seu ciclo vital, passa necessariamente por eventos de diferenciação celular (figura 1). Em uma das fases do ciclo a célula é móvel (figura 1a), possuindo um único flagelo polar, fimbrias e capacidade quimiotática. Seu cromossomo encontra-se condensado, e portanto incapaz de se replicar. Após cerca de um terço do ciclo, sob o controle de sinais internos, a célula móvel perde o flagelo e as fimbrias, descondensa-se seu cromossomo, e inicia-se, no mesmo pólo da célula onde se encontrava anteriormente o flagelo, a formação do talo. A célula talo recém desenvolvida (figura 1b) passa então a duplicar seu cromossomo e aumentar de tamanho. Mais tarde, um novo flagelo é montado no pólo oposto ao do talo. Forma-se assim uma célula assimétrica denominada pré-divisional (figura 1c), que dá origem, por ocorrência de sua divisão, a duas células filhas morfológica e fisiologicamente diferentes, sendo uma delas uma nova célula móvel e outra uma nova célula talo sésil. A duração do ciclo é de aproximadamente três horas a 30°C em meio definido (revisado por Shapiro, 1994).

C. crescentus é um organismo bem adequado como modelo de estudo para os fenômenos que ocorrem durante o ciclo celular principalmente porque as células móveis podem ser facilmente separadas das demais em gradiente de densidade, o que possibilita

um método conveniente de sincronização de uma população desta bactéria pelo período de um ciclo (Dingwall *et al.*, 1990).

Figura 1. Ciclo celular de *Caulobacter crescentus*. (a) célula móvel; (b) célula talo; (c) célula pré-divisional (extraído e adaptado de Kelly *et al.*, 1998).

A diferenciação celular em *C. crescentus* envolve a transcrição temporal e espacialmente controlada de vários genes, a translocação e restrição de certas proteínas a regiões específicas da célula, e a proteólise regulada. O evento melhor estudado deste fenômeno é a cascata regulatória da biossíntese do flagelo: os genes envolvidos nesta cascata podem ser classificados em quatro classes, sendo que a expressão dos genes classificados no topo da hierarquia é necessária para a expressão dos genes classificados mais abaixo (revisado em Brun *et al.*, 1994, e Gober & Marques, 1995). O único gene da

classe I identificado até o momento é o que codifica a proteína regulatória CtrA (Quon *et al.*, 1996), que se liga a sítios específicos no DNA, sendo sua atividade controlada por fosforilação e proteólise (Wu *et al.*, 1998; Jenal & Fuchs, 1998). Os genes da classe II possuem seqüências reconhecidas por CtrA em suas regiões regulatórias e codificam os fatores de transcrição σ^{54} e FlbD, as proteínas que formam a base da estrutura do flagelo, além de proteínas envolvidas na montagem do flagelo. Os genes da classe III, transcritos pela RNA polimerase $E\sigma^{54}$, codificam proteínas que compõem a região intermediária do flagelo, enquanto os genes da classe IV codificam as proteínas do gancho e as flagelinas. A expressão dos genes da classe II é necessária para a transcrição dos genes da classe III, assim como a expressão destes últimos é condição obrigatória para que sejam transcritos os genes da classe IV (revisado em Gober & Marques, 1995). Além da cascata de regulação de genes envolvidos na biossíntese do flagelo, CtrA também participa da regulação da transcrição de genes essenciais para o prosseguimento do ciclo celular, como a DNA metilase *ccrM* e *hemC* por exemplo, e impede a replicação do cromossomo na célula móvel (Quon *et al.*, 1996). Entre as proteínas que têm sua expressão e/ou localização controladas ao longo do ciclo celular de *C. crescentus* estão as chamadas proteínas de choque térmico.

Proteínas de resposta ao choque térmico

A indução da síntese de certas proteínas em resposta ao choque térmico ou a outros tipos de estresse é um fenômeno que ocorre na maioria dos organismos. Muitas destas proteínas possuem seqüências de aminoácidos bastante conservadas evolutivamente, tendo sido denominadas proteínas de choque térmico ou Hsp's (*heat shock proteins*) (Lindquist &

Craig, 1998). A indução destas proteínas em condições de estresse permite que a célula possa se recuperar da desnaturação protéica, prevenindo-a por possibilitar que cadeias desnaturadas de proteína possam voltar à conformação nativa, impedindo assim a formação de agregados protéicos, ou mesmo por degradar aquelas irreversivelmente desnaturadas. As principais classes de Hsp's são as Hsp90, Hsp70 e Hsp60. As classes Hsp60 e Hsp70 estão envolvidas com o enovelamento correto de proteínas nascentes ou na prevenção da desnaturação e ainda com a montagem de estruturas multiméricas (Craig *et al.*, 1993), o que conferiu a elas a denominação de “cicerones moleculares”, ou mais popularmente, por anglicismo, chaperones moleculares. Por conseguinte, várias Hsp's são chaperones moleculares enquanto outras são proteases.

Apesar dos chaperones moleculares terem sido identificados inicialmente devido à indução por choque térmico ou outras formas de estresse, sabe-se hoje que eles são essenciais em condições fisiológicas, uma vez que são capazes de se ligar a outras proteínas que se encontrem em formas não nativas, assistindo seu enovelamento até que alcancem a conformação funcional, em geral às custas de ATP (revisado em Bukau & Horwich, 1998). As famílias mais estudadas de chaperones moleculares são a da Hsp70 (DnaK em *E. coli*) e a das chaperoninas (GroEL em *E. coli*), mas o termo é aplicável também às Hsp100 e Hsp90, e às Hsp's de baixo peso molecular.

DnaK age em conjunto com as proteínas de choque térmico DnaJ e GrpE. O substrato a ser enovelado passa por vários ciclos subsequentes de ligação/liberação, geralmente na estequiometria de uma molécula de DnaK para cada molécula de substrato. Os substratos liberados podem chegar à conformação nativa, entrar em um novo ciclo de ligação a DnaK, ligar-se a outros chaperones ou proteases, ou formarem agregados (Bukau

& Horwich, 1998). O sistema DnaK/DnaJ/GrpE pode interagir com peptídeos assim que estes emergem do ribossomo, a fim de prevenir sua agregação ou dobramento incorreto.

A classe das chaperoninas forma estruturas multiméricas relativamente grandes, compostas por 14, 16, ou 18 subunidades, e pode ser dividida em dois grupos. O primeiro é representado por GroEL, presente no citossol de eubactérias, e Hsp60, encontrada em organelas de provável origem endossimbionte, tais como mitocôndrias e cloroplastos; os membros deste grupo apresentam alto grau de similaridade. Já o segundo grupo é representado por chaperoninas contendo TCP-1 (CCT) do citossol de eucariotos, e TF55 em arqueobactérias, que apresentam menor similaridade de sequência (Braig, 1998).

Em *E. coli*, como na maioria das bactérias, existe apenas um operon que codifica GroEL, chamado *groESL* (por codificar também a co-chaperonina GroES). Entretanto, em algumas bactérias Gram-positivas e cianobactérias está presente, além deste operon, um gene *groEL* isolado. Por outro lado, em *Sinorhizobium meliloti* existem, além de um gene *groEL* isolado, dois operons *groESL* sendo um deles plasmidial. *Bradyrhizobium japonicum* apresenta 5 cópias de *groESL* (Fischer *et al.*, 1993). *Caulobacter crescentus* apresenta apenas um operon *groESL* (Avedissian & Gomes, 1996).

Mecanismos de regulação da resposta ao choque térmico em bactérias

Os sistemas de regulação da resposta ao choque térmico em bactérias podem ser controlados em nível de transcrição, através de fatores sigma alternativos ou por repressores de transcrição, ou por mecanismos pós-transcrição, como estabilização de RNA's mensageiros.

Os mecanismos de controle em nível de transcrição mais bem caracterizados são os que envolvem o fator σ^{32} de *E. coli*, e aqueles dependentes da sequência repetida invertida CIRCE (*controlling inverted repeat of chaperone expression*, Zuber & Schmann, 1994), à qual se liga o repressor HrcA (*heat regulator at CIRCE*, Roberts *et al.*, 1996). O elemento CIRCE foi inicialmente descrito em *Bacillus subtilis* e micobactérias e posteriormente foi encontrado à montante de operons *groE* e *dnaK* de diversas bactérias filogeneticamente distintas (muitos arrolados em Hecker *et al.*, 1996, e Segal & Ron, 1996a), em uma lista que só tende a aumentar com o crescente número de genomas bacterianos sendo finalizados. A sequência CIRCE apresenta um alto grau de conservação, com consenso TTAGCACTC-N₉-GAGTGCTAA (Zuber & Schmann, 1994, Segal & Ron, 1996a), e está sempre associada a promotores com o fator sigma vegetativo. O único operon estudado em que CIRCE está associado a um promotor regulado por σ^{32} é *groESL* de *C. crescentus* (Avedissian & Gomes, 1996, Baldini *et al.*, 1998), embora o operon *groESL* de *Bordetella pertussis* também apresente simultaneamente um promotor de σ^{32} e a sequência repetida invertida CIRCE (Fernandez & Weiss, 1995), porém nenhum estudo de regulação da expressão de *groESL* foi realizado ainda nesta bactéria. Entretanto existem fortes indícios de que ao menos duas outras α proteobactérias também possuam simultaneamente um promotor funcional de σ^{32} e CIRCE como elemento de ligação de HrcA na região regulatória do operon *groESL*, sendo elas *Agrobacterium tumefaciens* (Nakahigashi *et al.*, 1999) e *Rhodobacter sphaeroides* (Lee *et al.*, 1997, Karls *et al.*, 1998).

Quanto aos mecanismos de controle pós-transcrição, há evidências de que ocorram ao menos dois tipos de controle. Em *A. tumefaciens* e *B. subtilis*, a sequência invertida repetida CIRCE é transcrita e contribui para uma menor meia vida do RNA mensageiro em

condições normais (Segal & Ron, 1996b e Yuan & wong, 1995, respectivamente). O segundo tipo de controle é apresentado pelo mRNA proveniente do operon *groESL* em *A. tumefaciens*; este pode ser clivado em um ponto específico entre as regiões codificadoras *groEL* e *groES*, em um processo dependente da temperatura, levando a rápida degradação do transcrito resultante para *groES*, desestabilizado pela presença da sequência CIRCE, enquanto aquele de *groEL* é estável. Deste modo o processo cria um mecanismo de expressão diferencial dos dois genes (Segal & Ron, 1995). Mecanismo idêntico foi encontrado no processamento e expressão diferencial dos transcritos primários do operon *dnaK* de *B. subtilis*, que também apresenta a sequência CIRCE (Homuth *et al.*, 1999).

Regulação da resposta ao choque térmico em *Escherichia coli*

Os principais genes induzidos em resposta ao choque térmico em *E. coli* são ativados por um fator sigma alternativo da RNA polimerase, a proteína σ^{32} , codificada pelo gene *rpoH*. A transcrição deste gene ocorre a partir de cinco promotores, pelo menos. Quatro deles, P1, P2, P4, e P5, são reconhecidos pela RNA polimerase contendo o fator sigma vegetativo, σ^{70} . A maior parte da transcrição em condições fisiologicamente normais ocorre a partir de P1 e P4. Já o promotor P3 é reconhecido pela holoenzima contendo o fator alternativo σ^{24} . A transcrição a partir deste promotor aumenta gradativamente com a elevação da temperatura dentro da faixa de crescimento de *E. coli*, sendo ele o único promotor dentre os cinco a permanecer ativo em temperaturas de choque térmico extremo (51 °C), assegurando assim a manutenção dos níveis de mRNA de *rpoH*, que tem uma curta meia vida (Georgopoulos *et al.*, 1994). Além dos dois fatores sigma (σ^{70} e σ^{24}), a

modulação da transcrição de *rpoH* é influenciada ainda pelo regulador global CRP (*catabolite repressor protein*), necessário para a transcrição a partir de P5; pela proteína DnaA, que reprime a transcrição via P3 e P4; e pelo anti-ativador CytR (Kallipolitis & Valentin-Hansen, 1988).

A maior parte do controle da concentração de σ^{32} na célula ocorre no nível de tradução ou pós-tradução. Quando a temperatura é elevada de 30 para 42 °C, há um grande aumento na concentração da proteína (cerca de 20 vezes), enquanto o mRNA para σ^{32} aumenta apenas duas vezes. Três regiões de controle, A, B e C, foram observadas dentro da região codificadora do mRNA. A região A assegura um alto nível de tradução, enquanto as regiões B e C a reprimem. Foi encontrada a função fisiológica ao menos da região C, que está envolvida com a ligação de σ^{32} ao cerne da RNA polimerase, conferindo a este fator sigma vantagem sobre os outros na competição pela ligação com a enzima (Arsène *et al.*, 1999).

A proteína σ^{32} madura é altamente instável em temperaturas normais de crescimento. Esta instabilidade é mediada pelo complexo DnaK/DnaJ/GrpE, que seqüestra σ^{32} , não permitindo sua ligação ao cerne da RNA polimerase, e media a sua degradação por proteases. Durante o choque térmico, o complexo DnaK/DnaJ/GrpE é requisitado para outras funções, liberando σ^{32} , que pode então se ligar ao cerne da RNA polimerase, ativando a transcrição dos genes de choque térmico. O excesso das proteínas de choque térmico (DnaK/DnaJ/GrpE) reprime novamente sua própria expressão seqüestrando σ^{32} de novo. Evidências sugerem que GroEL também possa estar envolvida no enovelamento correto de σ^{32} (Georgopoulos *et al.*, 1994).

Regulação da resposta ao choque térmico em *Bacillus subtilis*

A bactéria Gram-positiva *Bacillus subtilis* apresenta três classes de genes de choque térmico. Aqueles da classe I apresentam o elemento CIRCE na região 5' não codificadora. São representados pelos operons *groESL* e *hrcA-grpE-dnaK-dnaJ-orf25-orf38-orf50*. Estes operons são transcritos pela RNA polimerase contendo o fator sigma vegetativo σ^A (Hecker *et al.*, 1996; Homuth *et al.*, 1997). A regulação dos genes de choque térmico da classe I em *B. subtilis* está pormenorizada mais adiante, onde se discute o sistema CIRCE/HrcA, para fins de comparação com genes controlados pelo sistema CIRCE/HrcA em outros organismos.

Os genes de choque térmico de *B. subtilis* da classe II são regulados por um fator sigma alternativo, σ^B , que não é análogo ao fator σ^{32} de *E. coli*. A maioria dos genes de choque térmico estão incluídos neste regulon, que compreende também genes induzidos em fase estacionária e outras condições desfavoráveis à célula, como exposição a etanol, estresse salino e limitação de nutrientes ou oxigênio. Das cerca de 40 proteínas cuja expressão é regulada por σ^B , catalase, UDP-glicose pirofosforilase, e GsiB (*glucose starvation induced protein*), além do próprio fator σ^B , foram identificadas e são dispensáveis no crescimento vegetativo (Hecker *et al.*, 1996). Outros genes que são controlados por σ^B são *ctc* (Benson & Haldenwang, 1993), *gspA* (Antelmann *et al.*, 1995), *katB* (Engelmann *et al.*, 1995), e *trxA* (Scharf *et al.*, 1998).

A classe III inclui outros genes induzidos por choque térmico, entre eles os genes para proteases dependentes de ATP, como ClpP, ClpC, Lon, e FtsH (Hecker *et al.*, 1996), que em *E. coli* são todos regulados por σ^{32} . Foi demonstrado que a proteína codificada pelo primeiro gene do operon de *clpP* é capaz de se ligar a uma sequência heptamérica repetida

presente na região regulatória dos operons *clpC*, *clpP*, e *clpE*. Esta proteína foi denominada CtsR, de *class three stress gene repressor*, e funciona como um repressor de transcrição cuja forma ativa é um dímero, estando o sistema proteolítico ClpXP envolvido no controle da concentração de CtsR (Kruger & Hecker, 1998, Derré *et al.*, 1999, Derré *et al.*, 2000). Os promotores dos genes desta classe são reconhecidos pelo fator σ^B da RNA polimerase (Kruger & Hecker, 1998, Derré *et al.*, 1999), sendo que o promotor do operon de *ctsR* e *clpC* tem ainda uma sequência para reconhecimento pelo fator σ^A (Kruger & Hecker, 1998).

Regulação da resposta ao choque térmico em *Caulobacter crescentus*

Em *C. crescentus*, a síntese de aproximadamente 20 diferentes polipeptídeos é induzida durante a resposta ao choque térmico (Gomes *et al.*, 1986). Os mesmos polipeptídeos são induzidos em todas as fases do ciclo celular e sua expressão é regulada em nível da transcrição. Quatro dessas proteínas (Hsp70, Hsp62, Hsp24b, e Hsp23a) também estão presentes, porém em níveis mais baixos, em extratos de células cultivadas a temperaturas normais de crescimento, apresentando regulação temporal durante o ciclo. A Hsp mais abundante após o choque térmico, Hsp62, apresenta reação cruzada com o anticorpo contra a proteína GroEL de *E. coli*, sendo homóloga a esta. A Hsp70 de *Caulobacter* também apresentou reação cruzada com o anticorpo contra DnaK de *E. coli* (Gomes *et al.*, 1986).

O operon *dnaKJ* de *C. crescentus* foi isolado e caracterizado, apresentando dois promotores, P1 e P2 (Gomes *et al.*, 1990, Avedissian *et al.*, 1995). P1 é similar ao consenso

de promotores de choque térmico de *E. coli*, sendo o mRNA resultante de sua transcrição o mais abundante nesta condição. Já P2 é similar aos promotores regulados por σ^{70} de *E. coli*, sendo ativo apenas em condições normais de temperatura. A transcrição a partir de P2 apresenta controle temporal ao longo do ciclo celular da bactéria, ocorrendo durante a transição de célula móvel para célula talo (pontos a e b, respectivamente, da figura 1), imediatamente antes do início da replicação do DNA pela célula talo da progênie (Gomes *et al.*, 1990). A proteína DnaK sintetizada na célula prédivisional é segregada especificamente para a célula talo da progênie (Reuter & Shapiro, 1987), o que sugere seu envolvimento na replicação do DNA de *Caulobacter*. Foi demonstrada também a existência de uma sequência líder nos transcritos originados dos dois promotores, cuja deleção resulta no aumento da taxa de expressão das fusões de transcrição e tradução com um gene repórter, sem afetar a regulação temporal do operon. A formação de estruturas secundárias na sequência líder pode ser um mecanismo de modulação de *dnaK* e *dnaJ* em condições de choque térmico (Avedissian *et al.*, 1995).

O gene *rpoH*, que codifica para a proteína σ^{32} , foi isolado em *C. crescentus* por dois grupos, trabalhando independentemente (Reisenauer *et al.*, 1996, e Wu & Newton, 1996). A proteína apresenta 42% de identidade com σ^{32} de *E. coli*, 40% com *Pseudomonas aeruginosa* e 38% com *Citrobacter freundii*, além de ser reconhecida por anticorpos contra σ^{32} de *E. coli*. A proteína purificada permite a transcrição *in vitro* a partir do promotor P1 do operon *dnaKJ*, mas não a partir de P2, P1 mutado, ou promotores reconhecidos por σ^{70} , na presença do cerne da RNA polimerase (Wu & Newton, 1996).

Em *C. crescentus*, a regulação de *rpoH* parece ser diferente da estabelecida em *E. coli*. Foram encontrados dois inícios de transcrição para este gene, que definiram os

promotores P1 e P2. A sequência de P1 tem similaridade com o consenso de promotores σ^{70} , e a de P2 com aquele de σ^{32} , sugerindo uma auto regulação positiva. A 30 °C, os transcritos a partir de P2 são 8-10 vezes mais abundantes que os de P1, e após 5 minutos a 42 °C esta diferença aumenta para quase 100 vezes (Reisenauer *et al.*, 1996, Wu & Newton, 1996). Tanto a transcrição quanto os níveis da proteína não são regulados durante o ciclo celular.

Na Tabela 1 estão listados os operons de proteínas de choque térmico de *C. crescentus* estudados até o momento.

Tabela 1. Operons de choque térmico de *Caulobacter crescentus*

Operon	Função	Promotor	Regulação Temporal	Referência
<i>DnaKJ</i>	Chaperones moleculares	P1 $\rightarrow \sigma^{32}$ P2 $\rightarrow \sigma^{70}$	Maior expressão na célula talo	Gomes <i>et al.</i> , 1990; Avedissian <i>et al.</i> , 1995
<i>GroESL</i>	Chaperonina e co-chaperonina	σ^{32} CIRCE	Maior expressão na célula pré-divisonal	Avedissian & Gomes, 1996 Baldini <i>et al.</i> , 1998
<i>hrcA-grpE</i>	Repressor de groESL (HrcA) Chaperone molecular (GrpE)	σ^{32}	Constitutiva	Roberts <i>et al.</i> , 1996
<i>RpoH</i>	Fator σ^{32} da RNA polimerase	σ^{32}	Constitutiva	Reisenauer <i>et al.</i> , 1996
<i>FtsH</i>	Protease	?	?	Fischer <i>et al.</i> , 2002
<i>Lon</i>	Protease	?	Constitutiva(?)	Wright <i>et al.</i> , 1996
<i>ClpP</i>	Protease	σ^{32}	Maior expressão na célula pré-divisonal	Østerås <i>et al.</i> , 1999

O operon de *groESL* de *Caulobacter crescentus* apresenta um promotor regulado por σ^{32} e a sequência repetida invertida CIRCE, provável sítio de ligação da proteína repressora HrcA (Avedissian & Gomes, 1996; Roberts *et al.*, 1996). A expressão do operon *groESL*, além de ser induzida em condições de choque térmico, apresenta regulação temporal, sendo sua transcrição praticamente inexistente na célula talo e máxima na célula pré-divisional, pelo que foi sugerida a participação de GroEL na montagem do flagelo (Avedissian & Gomes, 1996). A regulação temporal é perdida em linhagens mutantes para *hrcA* e em fusões de transcrição de um gene repórter com a região regulatória de *groESL* contendo mutações no elemento CIRCE (Baldini *et al.*, 1998). Nos dois casos, ainda pode ser observada a indução por choque térmico, ao contrário do que ocorre em *Bacillus subtilis*, que não possui σ^{32} (Zuber & Shumman, 1994). Posteriormente demonstrou-se que o sistema CIRCE/HrcA em *Caulobacter crescentus* é responsável pela regulação da expressão de *groESL* no ciclo celular, sendo a indução por choque térmico controlada por σ^{32} (Baldini *et al.*, 1998), muito embora seja necessária a dissociação do complexo HrcA-CIRCE para a indução da expressão de *groESL* por σ^{32} (Fischer *et al.*, 2002). O controle no ciclo celular, entretanto, não se dá nem em nível da transcrição do gene *hrcA*, que é constitutiva (Roberts *et al.*, 1996), nem por variação da concentração celular da proteína HrcA ao longo do ciclo de vida da bactéria (Baldini, 1999). Isto indica que a variação na expressão do operon *groESL* em *C. crescentus* durante o ciclo celular deste organismo se dá, provavelmente, por uma modulação da afinidade da interação entre a proteína HrcA e a sequência operadora CIRCE presente na região regulatória daquele operon.

O sistema CIRCE/HrcA

A sequência repetida invertida encontrada na região regulatória do operon *groESL* de *Caulobacter crescentus* é homóloga a 51 outras encontradas em 41 genes e operons de 29 espécies diferentes de bactérias (listados em Segal & Ron, 1996a), sempre precedendo os operons *groE* e/ou *dnaK*, além dos genes *groEL* dos actinomicetos (Gram-positivas de alto G+C), *dnaJ* de *Mycoplasma genitalium* e *Lactococcus lactis*, e *dnaJ* de *Streptomyces albus* (Grandvalet *et al.*, 1998). Na verdade, essa é uma lista parcial bastante defasada e os números aumentaram muito nos últimos anos com o progressivo aumento do número de espécies bacterianas cujos genomas foram, e vem sendo, seqüenciados. Tal seqüência foi denominada CIRCE (*controlling inverted repeat of chaperonin expression*, Zuber & Schumman, 1994), e apresenta alto grau de conservação, sendo o seu consenso TTAGCACTC-N₉-GAGTGCTAA; as posições mais fortemente conservadas são as seis mais internas da repetição invertida (posições 4 a 9, e 19 a 24) e a segunda a partir de cada extremidade (posições 2 e 26) (Segal & Ron, 1996a).

Em *Bacillus subtilis*, onde não existe o fator σ^{32} , Zuber & Schumman (1994) observaram que mutações pontuais no elemento CIRCE elevaram a expressão de DnaK em temperatura normal de crescimento (37°C) ao mesmo nível da expressão a 48°C (temperatura de choque térmico). Estes autores sugeriram que o elemento CIRCE seria um sítio de ligação de uma proteína com atividade repressora. Mutantes de *Bacillus subtilis* para o gene *hrcA* (anteriormente denominado *orf39*) apresentaram altos níveis de transcrição dos operons *dnaK* e *groE*, que possuem o elemento CIRCE, em condições normais de crescimento (Schulz & Schumann, 1996), sendo que apenas *groE* ainda apresenta uma leve indução por choque térmico (Hecker *et al.*, 1996). Yuan & Wong

(1995) demonstraram que extratos protéicos de uma cepa de *E. coli* expressando HrcA de *B. subtilis* são capazes de retardar fragmentos de DNA contendo a sequência do elemento CIRCE, em géis de poliacrilamida não desnaturante. Esses dados sugerem que a proteína HrcA, codificada por um gene do operon *dnaK*, liga-se ao elemento CIRCE, reprimindo a transcrição, ou que está de alguma forma relacionada com outra proteína que se liga a este elemento. Estudos mais recentes (Mogk *et al.*, 1997, 1998) sugerem uma autorregulação da resposta ao choque térmico das proteínas de classe I de *B. subtilis* por GroEL, pela modulação da atividade de HrcA. A proteína HrcA seria ativada por mudanças conformacionais, de acordo com a quantidade de GroEL disponível na célula, de modo análogo ao sistema σ^{32} /DnaK de *E. coli* (Georgopoulos *et al.*, 1994). Os autores sugerem que HrcA não se liga fortemente ao operador, mas que se encontra em um estado de equilíbrio dinâmico, passando facilmente a uma forma agregada quando desligada e na ausência de GroEL. No choque térmico e em outras formas de estresse, o aumento da concentração de proteínas desnaturadas na célula aumenta a concentração de GroEL combinada com estas proteínas às custas de GroEL livre, diminuindo a disponibilidade desta última para interagir com HrcA, que formaria agregados e não seria capaz de reprimir a expressão dos operons *dnaK* e *groESL*. Muito recentemente foi descrito um alelo mutante de *hrcA* de *B. subtilis*, cuja proteína expressa tem a propriedade de agregar-se menos que aquela tipo selvagem, sendo portanto mais adequada à caracterização bioquímica: os primeiros resultados experimentais com a proteína mutante indicam que o modelo está essencialmente correto no caso de *B. subtilis* (Reischl *et al.*, 2002). Este modelo regulatório, porém, não parece ser absolutamente geral, pois em outros organismos situações diversas foram encontradas, como descrito a seguir.

Em *Lactococcus lactis*, deleções na região C-terminal de DnaK causaram o aumento da expressão dos operons contendo CIRCE (*hrcA-grpE-dnaK*, *groESL*, e *dnaJ*), mas não nos níveis do mRNA da protease HflB, envolvida na regulação do choque térmico, cujo gene *hflB*, não apresenta CIRCE (Koch *et al.*, 1998). Em *Clostridium acetobutylicum* foi demonstrada a prevenção da agregação de HrcA por DnaK, DnaJ, e pelo sistema DnaK-DnaJ-GrpE (Rüngeling *et al.*, 1999). Estes resultados sugerem que nestes organismos o sistema de regulação de CIRCE/HrcA é diferente daquele proposto para *B. subtilis*, onde apenas GroEL parece interferir na atividade de HrcA. Em *Bradyrhizobium japonicum*, tanto o sistema GroES/GroEL quanto o sistema DnaK/DnaJ não melhoraram a ligação de HrcA a CIRCE *in vitro* (Minder *et al.*, 2000). Já no caso de *Bacillus thermoglucosidasius* foi sugerido que a ligação entre HrcA e CIRCE é por si só termo-lábil (Watanabe *et al.*, 2001). Estudos em outras bactérias fazem-se necessários para elucidar o mecanismo da modulação da atividade de HrcA.

O gene *hrcA* de *C. crescentus* é o primeiro gene de um operon com *grpE*, que também funciona como “cicerone molecular” juntamente com DnaK e DnaJ (Roberts *et al.*, 1996). Foi identificado um promotor a 5' de *hrcA*, cujo transcrito compreende ambos os genes, e um outro mais fraco, a 5' de *grpE*. Ambos são levemente induzidos por choque térmico, transitoriamente. Na linhagem LS2293 de *C. crescentus*, na qual o gene *hrcA* foi interrompido, os níveis do transcrito de *groESL* estão aumentados em condições normais de temperatura, e fusões de transcrição da região regulatória de *groEL* com o gene repórter *lacZ* também apresentam atividade aumentada a temperaturas normais, sendo ainda observado um aumento com o choque térmico. Mutações sítio-dirigidas no elemento CIRCE também elevaram a atividade da β -galactosidase na linhagem selvagem contendo

fusões de transcrição da região regulatória de *groESL* com o gene *lacZ* (Roberts *et al.*, 1996, Baldini *et al.*, 1998), além de ocasionarem perda da regulação temporal do operon *groESL* e tornarem a resposta ao choque térmico mais transitória (Avedissian, 1996).

Os resultados dos estudos acima sugerem que em *C. crescentus* o elemento CIRCE está envolvido com a regulação da transcrição durante o ciclo celular, mais do que com a indução por choque térmico. Isso torna *C. crescentus* um modelo interessante do estudo da regulação do choque térmico, já que tanto a regulação de *groESL* quanto a de *rpoH* apresentam características marcadamente distintas das estabelecidas para *E. coli* ou *B. subtilis*, sendo que no caso do operon *groESL* em *C. crescentus* há a combinação simultânea de mecanismos de controle típicos tanto de *E. coli* quanto de *B. subtilis*.

HrcA em *Xylella fastidiosa*

Xylella fastidiosa é uma proteobactéria Gram-negativa do grupo γ , conhecida por causar uma série de patologias em vegetais, entre elas a clorose variegada dos cítricos, conhecida popularmente como “amarelinho”. Muito embora tenha sido encontrada anteriormente seqüências CIRCE na região regulatória dos operons *groESL* das γ proteobactérias *Chromatium vinosum* (Ferreyra *et al.*, 1993) e *Francisella tularensis* (Ericsson *et al.*, 1997), *X. fastidiosa* foi a primeira bactéria deste grupo na qual foi demonstrada a existência do gene *hrcA*, que aparece em muitas α e β proteobactérias, e em pelo menos duas proteobactérias do grupo ϵ , *Helicobacter pylori* (Tomb *et al.*, 1997) e *Campylobacter jejuni* (Thies *et al.*, 1999). O sequenciamento do genoma da linhagem de *X. fastidiosa* responsável pela citada fitopatologia foi concluído no ano de 2000 por um

conjunto de grupos de pesquisa brasileiros em trabalho coordenado e mereceu matéria de capa da revista Nature (Simpson *et al.*, 2000). *X. fastidiosa* contém um conjunto básico de proteínas de choque térmico e um gene codificando σ^{32} . Assim como *Caulobacter crescentus*, *X. fastidiosa* também apresenta o operon *groESL* contendo o elemento CIRCE em sua região regulatória e o gene *hrcA*, localizado no operon *dnaKJ*. Mais que isso, as HrcA's de *C. crescentus* e de *X. fastidiosa* apresentam elevada similaridade entre si (54,5 % de similaridade e 40,1 % de identidade usando-se o algoritmo de Needleman e Wunsch (alinhamento global) e parâmetros padrão do programa GCG do pacote de Winsconsin).

Estrutura das HrcA's: predições, oligomeria e domínios

Parece haver um certo consenso de que as HrcA's possuem um motivo *helix-turn-helix* (HTH) amino-terminal que seria responsável pela interação da proteína com a sequência do elemento CIRCE, como mostra a anotação disponível no Pfam-A para a família HrcA (Número de acesso - PF01628), obtida no endereço eletrônico do Pfam no Sanger Center (<http://www.sanger.ac.uk/Pfam>). Além disso, o ProDom (<http://protein.toulouse.inra.fr/prodom/doc/prodom.html>), outro sistema de predição de domínios e famílias de proteínas, aponta um único domínio presente em todas as HrcA's armazenadas em seu banco de dados (Número de acesso deste domínio no ProDom - PD003513), sendo tal domínio amino-terminal, presumivelmente responsável pela função comum de ligação ao DNA. Este domínio apresenta similaridade de sequência com o domínio de função desconhecida denominado pelo Pfam de DUF293. Este último

aparece em proteínas de pelo menos 9 espécies (7 gêneros) de arqueobactérias, sempre seguido de dois domínios regulatórios CBS (funcionalmente os domínios CBS são pouco conhecidos, embora sejam encontrados em uma ampla gama de proteínas, compreendendo proteínas de sinalização celular, transportadores/canais de membrana, enzimas e subunidades regulatórias das mesmas, e reguladores de transcrição tipo HTH: acredite-se que sejam capazes de ligar pequenas moléculas de função regulatória; para maiores informações, ver a página da internet <http://www.sanger.ac.uk/users/agb/cbs/cbs.html>). Presumivelmente as proteínas que possuem o domínio DUF293 também são reguladores de transcrição.

Motivos HTH estão presentes, em sua maioria, em reguladores de expressão procarióticos e de vírus, especialmente dos bacteriófagos. Nos casos clássicos estes reguladores ligam-se a seqüências repetidas invertidas por forças intermoleculares secundárias, na forma de dímeros simétricos onde os motivos HTH de cada monômero estão separados por cerca de 34 Å (o passo da hélice de DNA, que vem a ser a distância que separa o ponto médio de duas fendas maiores consecutivas, segundo o modelo de Watson e Crick). Cada monômero liga-se a uma das seqüências repetidas, formando um complexo ternário com o DNA, que é mantido pela interação cooperativa entre os monômeros da proteína e a cadeia de ácido nucléico, a qual é mais forte que a simples soma das interações entre os componentes individuais.

Ainda na década de 80, Brian Mathews chegou a uma série de restrições estereoquímicas para a seqüência de aminoácidos dos motivos HTH, por comparação das seqüências dos motivos HTH e das estruturas tridimensionais do repressor e da proteína Cro do fago lambda, e da proteína CRP de *E. coli*. Segundo suas conclusões um motivo HTH é composto por 20 aminoácidos, onde os sete resíduos amino-terminais formam uma

hélice estabilizadora, e os nove resíduos carboxi-terminais formam a hélice de reconhecimento do DNA, sendo que os quatro aminoácidos entre as duas hélices formam uma alça entre elas. Numerando os aminoácidos do amino para o carboxi-terminal, Mathews concluiu ainda que: (1) o resíduo 9, localizado na alça entre as hélices, tem que ser uma glicina, para suportar a conformação de hélice levógira naquela posição; (2) os resíduos 4, 8, 10, e 15 formam o cerne hidrofóbico do motivo, e devem ser hidrofóbicos ou possivelmente fracamente polares, sendo que os resíduos 4 e 15, em particular, não podem ser carregados; (3) é pouco provável que algum dos resíduos 3 a 7 e 15 a 20 seja de prolina, devido a impossibilidade desta formar ponte de hidrogênio pelo nitrogênio da cadeia principal, o que levaria a quebra da hélice naquelas posições; e (4) o resíduo 5 não deve ter uma cadeia lateral ramificada, já que esta deve se posicionar ao longo da interface entre as duas hélices, onde uma ramificação afetaria o posicionamento relativo entre elas. Ainda, segundo o modelo de Mathews, a hélice de reconhecimento do motivo HTH encaixava-se no sulco maior da repetição invertida e os resíduos mais prováveis de interagir especificamente com o DNA eram os de números 11, 12, 13, 16, 17, e 20, o que mais tarde foi verificado experimentalmente para muitas proteínas (Branden & Tooze, 1991). Arginina, asparagina, aspartato, glutamato e glutamina estão entre os aminoácidos que mais freqüentemente interagem especificamente com as bases do DNA alvo nos motivos HTH, devido à sua capacidade de participar de múltiplas pontes de hidrogênio simultaneamente (Darby & Creighton, 1993). Atualmente as restrições concluídas por Mathews são consideradas úteis, porém excessivamente estridentes. Tome-se o caso dos motivos HTH dos homeodomínios eucarióticos como exemplo; no lugar da glicina 9, estes tem uma cisteína, e se presume que o dobramento é suficientemente estável para que o aminoácido assuma a conformação energeticamente desfavorável (Branden & Tooze, 1991). Ainda

assim as restrições mais confiáveis, e portanto mais úteis, propostas por Mathews são aquelas com relação à prolina e, principalmente, aquela com relação à polaridade do aminoácido 15.

Voltando a HrcA, Huynen *et al.* (1998) sugeriram para a proteína de *Mycoplasma genitalium*, em um trabalho sobre genômica estrutural neste organismo, que o domínio amino terminal da proteína era similar ao domínio amino-terminal de LexA de *E. coli*, o qual contém um motivo HTH. Tal domínio de HrcA é um provável membro da família estrutural do domínio N-terminal de LexA, denominada *Winged Helix*. As análises da sequência de HrcA de *C. crescentus* corroboram este resultado (este trabalho), e já há ao menos um grupo de modelagem molecular trabalhando em modelos do amino-terminal de várias HrcA's a partir de alinhamentos com domínios protéicos de estrutura determinada, pertencentes a família estrutural *Winged Helix* (Ursula Pieper e Andrej Sali, Laboratories of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, The Rockefeller University - <http://pipe.rockefeller.edu/modbase>): o modelo mais bem sucedido elaborado pelo grupo, segundo seus próprios critérios, é do amino-terminal de HrcA de *Lactococcus lactis*, baseado no alinhamento com o amino terminal de LexA de *E. coli*.

Quanto à oligomerização, a estequiometria do complexo HrcA/CIRCE de *Bacillus thermoglucosidasius* foi calculada a partir dos dados de experimentos de migração retardada em eletroforese em gel, apresentando o resultado de 2,6 moles de HrcA por mol de DNA contendo CIRCE, o que indica que 2 ou 3 moléculas da proteína compõem o complexo com CIRCE (Watanabe *et al.*, 2001). Este achado é frontalmente oposto ao encontrado para HrcA de *Staphylococcus aureus* por microscopia de força atômica, que indicavam uma estequiometria 1:1 para o complexo de CIRCE com HrcA daquele organismo (Ohta *et al.*, 1996). Finalmente, pondo fim a esta controvérsia, foi recentemente

determinado, pela primeira vez, o peso molecular de uma HrcA na conformação nativa, por gel-filtração (Martirani *et al.*, 2001), sendo este aproximadamente igual ao do homodímero, que vem a ser oligomeria básica de uma proteína HTH. É de se esperar, portanto, que outras HrcA's, senão todas, sejam também dímeros em suas conformações nativas.

OBJETIVOS

A principal meta deste trabalho foi a de reunir informações que permitam propor um modelo estrutural da proteína HrcA de *Caulobacter crescentus* que explique as suas propriedades, contribuindo assim para o entendimento da modulação da atividade da proteína neste organismo, e testar o modelo através de estudos espectroscópicos e da determinação experimental da estrutura das proteínas de *C. crescentus* e de *Xylella fastidiosa*, a ser executada com colaboradores especializados nesta área. Para tanto, foi necessária a análise criteriosa das seqüências de aminoácidos de ambas as proteínas, tanto isoladamente quanto no conjunto de seqüências das proteínas da família HrcA. Tal análise permitiu a formulação de um modelo que aponta aminoácidos com funções prováveis, possíveis limites dos domínios e regiões funcionais na seqüência da proteína, e dobramentos específicos para cada domínio, o que possibilita explicar certos aspectos do comportamento da proteína nas análises espectroscópicas.

Mesmo que a produção de HrcA de *X. fastidiosa* tenha se mostrado inviável, a proteína recombinante de *C. crescentus* mostrou-se adequada às análises espectroscópicas, e avanços foram feitos no sentido de produzir amostras a partir das quais seja possível obter cristais adequados para experimentos de difração de raios X, muito embora a inesperada microagregação seja um obstáculo que ainda esteja por ser completamente superado. Neste sentido, os experimentos de espalhamento de luz dinâmico e de cromatografia de exclusão por tamanho tornaram-se ferramentas inestimáveis de avaliação rotineira das amostras. Já as análises de fluorescência e de dicroísmo circular possibilitaram o refinamento do modelo nas situações dúbias, e a verificação de que este pode realmente mostrar-se essencialmente correto quando testado pelo exame da proteína por técnicas de determinação estrutural fina.

MATERIAIS E MÉTODOS

Linhagens e plasmídeos

A Tabela 2 relaciona os oligonucleotídeos utilizados, e a Tabela 3 lista os plasmídeos usados. O fragmento contendo o gene *hrcA* de *X. fastidiosa*, da linhagem 9a5c, foi isolado utilizando-se amplificação com oligonucleotídeos específicos com sítios de restrição para *Bam*HI em uma extremidade e para *Hind*III na oposta. Para clonagem e replicação dos plasmídeos foi usada a linhagem DH5 α de *Escherichia coli*, ou eventualmente a linhagem DH10B da mesma bactéria. O vetor de expressão contendo a região codificadora do gene *hrcA* de *Xylella fastidiosa* foi introduzido na linhagem DH5 α . A mesma linhagem contendo o vetor de expressão pProEX-1 no qual foi inserida a região codificadora do gene *hrcA* de *Caulobacter crescentus*, encontrava-se previamente disponível no laboratório.

Tabela 2. Oligonucleotídeos utilizados

Oligonucleotídeo	Sequência (5' PARA 3')	Fornecedor
HRCAX3	GCATGAAGCTTCAAACGATGGCGAC	Gibco BRL
HRCAX5	TGGATCCTTGCCAACGGTCTCCCG	Promicro

Tabela 3. Plasmídios utilizados.

Plasmídeo	Características	Fonte ou referência
pProEX-1	vetor de expressão indutível, replicon pMB1, Ap ^r , lacI ^q , pTrc	Life Technologies
pGEM-T Easy	Ap ^r	Promega

Meios e condições de cultura

As linhagens de *E. coli* foram crescidas em meio LB (triptona 10 g/l; extrato de levedura 5 g/l; NaCl 10 g/l; pH 7,5), 2 x TY (triptona 16 g/l; extrato de levedura 10 g/l; NaCl 5 g/l; pH 7,5), TB (triptona 12 g/l; extrato de levedura 24 g/l; glicerol 0,4 %; KH₂PO₄ 2,31 g/l; K₂HPO₄ 12,54 g/l), ou Circle Growth (BIO 101 INC.) à temperatura de 30 ou 37 °C, a depender do experimento em execução. Os meios foram adicionados de ampicilina (100 µg/ml) quando necessário. Os meios sólidos continham 1,5 % de ágar.

Amplificação de DNA por reação em cadeia da polimerase

Para o isolamento do gene *hrcA* de *X. fastidiosa* contendo sítios de restrição a distâncias adequadas da região codificadora, foi necessário realizar experimentos de reação em cadeia da polimerase (PCR). Foram feitos experimentos de PCR utilizando como fonte da sequência desejada o cosmídio LAWRIST contendo o operon *dnaKJ* de *X. fastidiosa*, do qual faz parte o gene *hrcA* deste organismo, construído durante o processo de sequenciamento genômico da bactéria, ou ainda usando DNA genômico isolado de *X. fastidiosa* do estoque do laboratório. Foram usadas ainda *Taq* DNA polimerase (*Taq* 5U/µl - Amersham Pharmacia) ou Cloned *Pfu* DNA polimerase (*Pfu* 2,5 U/µl - Stratagene), e os oligonucleotídeos listados na tabela 2 como iniciadores.

Os PCR's executados usando-se o cosmídio contendo o operon *dnaKJ* de *X. fastidiosa* foram realizados com o uso da polimerase *Pfu*, e nos PCR's executados com DNA genômico de *X. fastidiosa* foi usada a polimerase *Taq*.

No caso dos experimentos de PCR usando o cosmídeo, foi seguido o procedimento otimizado previamente pela doutoranda Michelle Susin em nosso laboratório (comunicação pessoal), adaptado para o uso da polimerase *Pfu* quando necessário. Foram adicionados, para cada tubo:

1,00 µl de cosmídeo isolado

5,00 µl de solução 2 mM de cada dNTP

3,65 µl de tampão para polimerase *Pfu* (10x)

1,00 µl de primer HRCAX3 (25 pMoles)

1,00 µl de primer HRCAX5 (25 pMoles)

9,35 µl de água milliQ

0,50 µl DNA polimerase *Pfu* (adicionada após os primeiros 5 minutos a 95 °C)

O termociclador foi programado como se segue:

5 minutos a 95 °C (pausa após este tempo para adição da *Pfu* polimerase)

1 minuto a 95 °C

30 segundos a 55 °C

2 minutos e 17 segundos a 72 °C

7 minutos a 72 °C

35 ciclos

Já no caso dos experimentos utilizando DNA genômico de *X. fastidiosa* foi seguido o protocolo estabelecido para este fim, desenvolvido dentro do projeto CAGE-*Xylella* (Cooperação para Análise de Genes e sua Expressão; projeto temático FAPESP, sub-projeto *Xylella*) (dados não publicados).

Isolamento de DNA, digestão com enzimas de restrição, ligação, sequenciamento e transformações

Os plasmídeos foram extraídos pelo método de lise alcalina, como descrito em Sambrook *et al.* (1989), com adaptações para uso em processo de sequenciamento automático quando necessário. As enzimas de restrição utilizadas foram obtidas de New England Biolabs, Amersham e Böehringer-Mannheim, e usadas com os tampões fornecidos pelos fabricantes.

Os fragmentos resultantes de digestão com enzimas de restrição foram analisados por eletroforese em gel de agarose (0,8 a 1 %) em tampão TBE (Tris 89 mM; ácido bórico 89 mM; EDTA 2 mM; pH 8), com 100 Volts de potencial. Após eletroforese os géis foram corados com brometo de etídio e o DNA visualizado sob UV. Para o isolamento de fragmentos de DNA foram realizadas eletroforeses em gel de agarose e após detecção das bandas de interesse o gel foi cortado e o DNA eletroeluído, conforme descrito em Sambroock *et al.* (1989). As ligações dos fragmentos de DNA aos plasmídeos foram realizadas a temperatura ambiente, com T4 DNA ligase e tampão Gibco-BRL, ou com o kit do fabricante conforme as instruções contidas no manual no caso do plasmídeo pGEM-T Easy.

Os sequenciamentos de DNA necessários foram executados com o “BigDye terminator kit” em sequenciador automático ABI377.

As transformações foram realizadas segundo o procedimento de Nishimura *et al.* (1990). Nas transformações com o sistema de ligação do plamídeo pGEM-T Easy, foram adicionados IPTG (30 µl de uma solução 100 mM) e X-gal (30 µl de uma solução 2 % em

dimetilformamida) às placas de meio LB adicionadas de ampicilina (100 µg/ml), para seleção de possíveis transformantes.

Eletroforese de proteínas e “Western blots”

Extratos protéicos de *E. coli* expressando HrcA de *X. fastidiosa* ou de *C. crescentus* foram obtidos de 1 ml de cultura, após centrifugação e ressuspensão em solução O (Laemmli, 1970) e fervura por 3 minutos. As concentrações dos extratos foram igualadas levando-se em conta o número de células da cultura inicial, inferido pela leitura da absorbância das culturas a 600 nm (DO₆₀₀).

A separação das proteínas por eletroforese foi realizada com gel de poliacrilamida desnaturante em sistema de pH descontínuo (SDS-PAGE, Laemmli, 1970) e a transferência das proteínas para membranas de nitrocelulose foi realizada em sistema Novablot (Pharmacia). Após a transferência, as membranas foram coradas com 0,1 % de Ponceau S em ácido acético 10 %, e descoradas com água destilada, para verificar a eficiência da transferência. No procedimento de “Western blot” usando-se anticorpos contra poli-histidina, o filtro foi então bloqueado por 20 a 30 minutos em TBS (Tris-HCl 10mM, pH 8,0; NaCl 150mM) contendo 5 % de leite desnatado. A solução foi trocada e os anticorpos adicionados até o título de 1:3000, sendo a membrana incubada 1 a 2 horas à temperatura ambiente com agitação suave. No caso do ensaio com anticorpo anti-HrcA de *C. crescentus*, este foi adicionado à uma solução de TBS-TT (Tris-HCl 10mM, pH 8,0; NaCl 150mM; Tween 20 0,03 %; Triton X-100 0,02 %), sem adição de leite e incubação a temperatura ambiente durante seis horas. Após a incubação, as membranas foram lavadas quatro vezes por 5 minutos com TBS-T (TBS contendo Tween 20 0,05 %) e uma vez com

TBS por 10 minutos, procedendo-se então a incubação em TBS-Leite com o anticorpo secundário conjugado com fosfatase alcalina, contra anticorpos de coelho (título 1:5000), no caso de anticorpo primário anti-HrcA, ou contra anticorpos de camundongo (título 1:10000), no caso de anticorpo primário anti-poli-histidina, seguindo-se nova lavagem como descrito acima.

Para revelação da reação da fosfatase alcalina, foram usados BCIP (5-bromo-4-cloro-3-fosfato indol 0,3 mg/ml) e NBT (nitroblue tetrazolium 0,15 mg/ml) em tampão Tris-HCl 10mM pH 9,0; NaCl 100mM e MgCl₂ 5mM.

Expressão e purificação das proteínas recombinantes HrcA-His

O gene *hrcA* de *X. fastidiosa* foi clonado no vetor de expressão pProEx-1, que expressa a proteína de interesse com uma sequência de 6 histidinas ligada a sua extremidade amino terminal, o que possibilita uma etapa de purificação da proteína recombinante em colunas de afinidade de agarose contendo níquel imobilizado por complexação.

A indução da expressão da proteína recombinante foi testada por tempos de 3 a 6 horas, em meios LB, 2 x TY, TB, ou Circle Growth contendo ampicilina, cuja resistência é codificada pelo plasmídio, e adicionando-se IPTG até 0,6 ou 1,0 mM de concentração final quando a D.O. a 600nm da cultura alcançava um valor entre 0,65 e 0,85 para alguns experimentos a 37 °C, ou entre 0,45 a 0,65 para todos os demais. As temperaturas de expressão foram de 30 ou 37 °C. As células foram recolhidas por centrifugação (10 minutos a 7996 x g) e suspensas em 4 ml de tampão A (Tris-HCl 50 mM pH 8,0; EDTA 0,5 mM;

NaCl 50 mM; glicerol 5%) contendo 150 µg/ml de PMSF, para cada 100 ml de cultura centrifugada após o tempo de expressão, e as células foram lisadas por sonicação, em um sonicador Branson 450, com oito pulsos de 15 segundos (potência 3) intercalados com intervalos de repouso de 30 segundos, sempre no gelo e cuidando-se para não exceder o volume de 2 ml em cada tubo de sonicação.

Quando a expressão da proteína foi induzida com 0,6 mM de IPTG e a 30 °C, a maior parte do produto de expressão permaneceu na fração solúvel quando foi utilizado o seguinte tratamento: adicionou-se imediatamente desoxicolato (DOC) 20 % em tampão A (contendo 150 µg/ml de PMSF) ao lisado até concentração final de 2 % (p/v), seguindo-se repouso por 10 minutos a temperatura ambiente. Após este tempo centrifugou-se por 20 minutos a 9770 x g a 4 °C. Ao sobrenadante foi acrescentado igual volume de tampão I (Tris-HCl 10 mM pH 7,5; NaCl 100 mM; Imidazol 10 mM), e o sobrenadante assim diluído foi dialisado durante a noite a 4 °C contra tampão I, para eliminação do DOC, seguindo-se duas trocas de tampão, cada uma após mais 3 horas de diálise. O dialisado foi então incubado pelo menos 1 hora e meia com a resina de níquel-agarose a 4 °C em agitador rotatório, na proporção de 4 ml de dialisado para 1 ml de suspensão 50 % de resina Ni-NTA (Qiagen) em etanol 30 %, equilibrada em tampão I. A resina foi então separada por centrifugação, 5 minutos a 1086 x g, e lavada três vezes com 500 µl de tampão I para 1 ml de suspensão 50 % de resina. A eluição foi feita por incubação da resina por quinze minutos a 4 °C em agitador rotatório com concentrações crescentes de imidazol em tampão A, normalmente entre 25 e 300 mM de imidazol, na proporção de 750 µl de eluente para cada 1 ml de suspensão 50 % de resina. Aparentemente, entretanto, a proteína sofreu agregação após a diálise.

Quando a expressão é induzida com 1 mM de IPTG ou executada a 37 °C, a proteína passa a ser seqüestrada em corpos de inclusão. Neste caso, foi otimizado um método para extração seletiva da proteína. Após a lise das células por sonicação, triplica-se o volume do lisado com DOC 3 % em tampão A (contendo 150 µg/ml de PMSF), e deixa-se a suspensão em repouso por 1 hora. Após este tempo centrifuga-se a suspensão por 20 minutos a 9770 x g a 4 °C, separa-se o sobrenadante do precipitado e continua-se com este último. Este é ressuspensão com DOC 2 % em tampão A (10 % do volume de cultura centrifugada antes da lise) contendo 150 µg/ml de PMSF, deixando-se repousar 10 minutos a temperatura ambiente antes de centrifugar novamente a 4 °C, por 20 minutos a 9770 x g. A seguir separa-se o sobrenadante, e ressuspensão-se o precipitado novamente com o mesmo volume de DOC 2 % em tampão A, repetindo-se o procedimento ciclicamente mais seis vezes. Os sobrenadantes das seis últimas alíquotas são reunidas e seguem para diálise como anteriormente.

Já a indução da expressão da proteína recombinante de *C. crescentus* foi feita quando a cultura atingiu entre 0,45 e 0,65 de D.O. 600nm, em meio TB contendo ampicilina, cuja resistência é codificada pelo plasmídeo, com a adição de IPTG até 0,6 mM de concentração final, a temperatura de 30 °C. As células foram recolhidas por centrifugação a 4 °C e suspensas em tampão A (Tris-HCl 50 mM pH 8,0; EDTA 0,5 mM; NaCl 50 mM; glicerol 5%) contendo 150 µg/ml de PMSF e lisadas por sonicação, em um sonicador Branson 450, com oito pulsos de 15 segundos (potência 3) intercalados com intervalos de repouso de 30 segundos, sempre no gelo. Acrescentou-se então desoxicolato (DOC) 20 % até a concentração final de 2 %, deixando-se 10 minutos a temperatura ambiente, e centrifugando-se a 4 °C em seguida por 20 minutos a 9770 x g. O sobrenadante

foi então separado do precipitado, guardando-se este último e prosseguindo-se com o primeiro, o qual é dializado a 4 °C com 3 trocas de tampão I (Tris-HCl 10 mM pH 7,5; NaCl 100 mM; Imidazol 10 mM), sendo cada troca efetuada após 2 horas de diálise. Após diálise a proteína foi incubada por no mínimo uma hora e meia com a resina Ni-NTA (Qiagen) a 4 °C em agitador rotatório. A seguir, a resina foi centrifugada e o sobrenadante foi desprezado. A proteína foi então eluída da resina com concentrações crescentes de imidazol (25, 50, 75, 100, 150, e 300 mM) em tampão I. A eluição foi feita por incubação da resina por quinze minutos a 4 °C em agitador rotatório com a respectiva concentração de imidazol em tampão I, na proporção de 750 µl de eluente para cada 1 ml de suspensão 50 % de resina em etanol 30 %. As duas primeiras frações foram desprezadas e as demais reunidas para serem concentradas com o uso de centricon 30 ou centriprep plus-80 (Amicon), de acordo com o volume total, e substituído o diluente por tampão A (3 vezes diluído) com o uso do mesmo dispositivo, para eliminação do imidazol.

Para contornar o problema de microagregação foi avaliada uma metodologia baseada no procedimento de Bär *et al.* (2000) para fosfofrutoquinase-1 de *Saccharomyces cerevisiae*. Neste caso, após o tempo de indução, as células são recolhidas por centrifugação e suspensas em tampão A contendo, além de PMSF, 0,75 % de DOC. As células são então lisadas e deixadas descansar 10 minutos conforme descrito acima. Após isto centrifuga-se o lisado como já descrito, para separação de sobrenadante e precipitado. O sobrenadante é então aplicado diretamente à resina Ni-NTA, sem diálise prévia. A incubação com a resina e a eluição são idênticas ao descrito acima, exceto que os tampões de eluição tem, além da respectiva concentração de imidazol, 0,75 % de DOC. Esta concentração de DOC nos tampões foi estabelecida após vários experimentos com

variadas concentrações mais elevadas, por ser a concentração mais alta do detergente que causa mínimo prejuízo à ligação de HrcA recombinante à resina Ni-NTA. As frações recolhidas são então todas reunidas, concentradas por centrifugação em centricon, lavadas com tampão A (1/3 x) contendo 0,75 % de DOC para eliminação do imidazol, concentradas novamente, e diluídas cem vezes em tampão A 1/3 x contendo 40 mM de α ciclodextrina. A α ciclodextrina é um anel de seis unidades de glicose com o interior hidrofóbico e as hidroxilas das unidades de glicose voltadas para a superfície externa. Esta estrutura permite abrigar em seu interior moléculas e grupos hidrofóbicos, tornando a α ciclodextrina um capturador de detergentes e também uma molécula capaz de ocultar grupos hidrofóbicos de proteínas que podem causar a agregação das mesmas. Após diluição da proteína em tampão com ciclodextrina, a solução foi deixada descansando em câmara fria durante a noite. Em seguida, a proteína foi concentrada em centriprep plus-80 e lavada duas vezes com tampão A (1/3 x) contendo 40 mM de α ciclodextrina para eliminação do DOC, e concentrada uma última vez para poder ser finalmente usada.

Marcação de sondas e hibridação

O vetor pProEx não codifica β -galactosidase, sendo portanto incapaz de proporcionar uma seleção de colônias do tipo branca-e-azul. Por conseguinte, para a seleção de clones contendo insertos de interesse neste plasmídio, as colônias obtidas das transformações foram transferidas para membranas de nitrocelulose e seu DNA hibridizado com sondas apropriadas para verificar a presença do inserto de interesse. Os clones obtidos foram inoculados em membranas colocadas sobre placas contendo meio LB sólido com os

antibióticos necessários e incubados por 5-16 horas a 37 °C. As membranas foram tratadas com SDS 10 % por 3 minutos, solução de desnaturação (NaOH 0,5 M; NaCl 1,5 M) por 5 minutos, solução de neutralização (Tris-HCl 1M pH 7,5, NaCl 1,5 M) por 5 minutos, e SSC 2x (NaCl 0,3 M; citrato trissódico 0,03 M) por 2 minutos. O DNA foi então fixado na membrana por exposição à ultravioleta (UV) por 6 minutos.

A pré-hibridização e hibridização das membranas são realizadas em solução Rapid-Hyb (Amersham Pharmacia), conforme as especificações do fabricante.

Para a obtenção de sondas radioativas, cerca de 50 ng de DNA de fragmento de aproximadamente 1,1 kb, isolado do clone sequenciado contendo o gene *hrcA* de *X. fastidiosa*, foram marcados por “random priming” com o kit RediPrime (Amersham Pharmacia) e 25 µCi de [α -³²P] dCTP. Os nucleotídeos não incorporados foram retirados com o kit Nucleon QC (Amersham Pharmacia). Foi aplicado 1 µl da reação em papel 3MM, que foi então seco ao ar, para servir de amostra para determinar a incorporação dos nucleotídeos radioativos. Procedeu-se a contagem da radioatividade num contador de cintilação, pelo método Cerenkov.

Através deste método foram conseguidos 12 clones confirmados por digestão com *Bam*HI e *Hind*III, verificando-se a presença do fragmento de aproximadamente 1,1 kb e do fragmento do plasmídeo por eletroforese em gel de agarose.

Dicroísmo circular

Os espectros de dicroísmo circular foram obtidos em colaboração com a Dra. Shirley Schreier, com o auxílio e supervisão do Dr. Cláudio Shida. Inicialmente os

experimentos de dicroísmo circular foram executados em um espectropolarímetro de dicroísmo circular JASCO 720. A cela utilizada foi de 0,1 mm, e os espectros foram a média de cinco varreduras com resolução de 0,5 nm (step) e velocidade de 20 nm/min, com largura de banda de 1 nm. Nas amostras com ATP as condições foram idênticas, mas a cela usada foi de 0,2 mm. Os espectros usados para desconvolução foram executados em um espectropolarímetro de dicroísmo circular JOBIN-YVON CD6. A cela utilizada foi de 0,1 mm, e os espectros foram a média de quatro varreduras com resolução de 0,5 nm (step) e tempo de integração de 6 seg/step, com largura de banda de 1 nm.

Fluorescência

As medidas de fluorescência foram realizadas em colaboração com a Dra. Shirley Schreier. Os experimentos foram executados em um espectrofluorímetro Hitachi F4500, sob a supervisão do Dr. Cláudio Shida. As fendas de excitação e emissão foram ajustadas para 5,0 nm, e os espectros foram a média de duas varreduras a 60 nm por minuto.

Espalhamento de luz dinâmico com laser

As medidas de diâmetros de partículas por espalhamento de luz dinâmico foram realizadas em colaboração com o Dr. Maurício S. Baptista. Estas medidas foram obtidas em um analisador de potencial zeta ZETA PALS (Brookhaven Instruments Corp.), com o uso do programa BIC Particle Sizing Software, fornecido com o equipamento. As curvas de calibração foram obtidas com o uso de soluções 10 mg/ml das proteínas dos kits LMW e

HMW Gel filtration calibration kit (Pharmacia Biotech). As medidas foram analisadas com a ferramenta do programa para tratamento de amostras polidispersas e tomada a média ponderada do diâmetro da população predominante como medida do diâmetro medido das proteínas.

FPLC

Os cromatogramas de FPLC foram obtidos no Laboratório de Bioquímica dos Insetos do Instituto de Química da USP, com o auxílio e supervisão do doutorando Fernando A. Genta. Os experimentos de FPLC foram executados em um cromatógrafo ÄKTA FPLC (Amersham Biosciences) com uma coluna Superdex 200 HR 10/30 (Amersham Biosciences). A eluição foi isocrática (sem variação na composição do eluente), com tampão A como eluente, e vazão de 1 ml / minuto. A curva de calibração foi feita usando-se os mesmos padrões de peso molecular usados para a calibração do equipamento de espalhamento de luz, além de Blue Dextran para o *Void* (limite de exclusão da coluna). A grandeza imediatamente utilizável para determinação de massas moleculares por FPLC é o K_{av} , que é calculado pela fórmula $K_{av} = (\text{Volume eluição amostra} - \text{Volume eluição do } Void) / (\text{Volume total da coluna} - \text{Volume de eluição do } Void)$.

Métodos computacionais

As análises de sequência de proteína foram feitas, na sua maioria, com o uso dos programas do SeqWeb (GCG -Genetics Computer Group) versão 1.0, que é parte do pacote Winscosin versão 9.1. As matrizes de substituição de aminoácidos usadas foram sempre tipo Blosum, variando de Blosum30 a Blosum100, e as demais opções foram padrão. Foram usados ainda os seguintes programas e bases de dados disponíveis na internet, na grande maioria das vezes com as opções padrão: PROSITE (<http://www.expasy.ch/prosite/>); InterPro (<http://www.ebi.ac.uk/interpro/>); Pfam (<http://www.sanger.ac.uk/Pfam/>); SMART (<http://smart.embl-heidelberg.de/>); SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>); ClustalW (<http://www.ebi.ac.uk/clustalw/>); ProDom (<http://protein.toulouse.inra.fr/prodom/doc/prodom.html>); MEME/MAST (<http://meme.sdsc.edu/meme/website/>); e FUGUE/HOMSTRAD (<http://www-cryst.bioc.cam.ac.uk/~fugue/>). PROSITE, InterPro, Pfam, SMART, e ProDom são bases de dados de proteínas, sendo estas agrupadas por famílias ou domínios comuns, conforme o caso, por homologia. Estas bases foram usadas para encontrar similaridades entre a família HrcA e outras famílias de proteínas. Os demais programas terão suas utilidades explicadas na parte da discussão de seus resultados.

As seqüências usadas estão listadas na tabela 4, com seus devidos números de acesso, e foram obtidas no banco de dados do NCBI (National Center for Biotechnology Information - <http://www.ncbi.nlm.nih.gov/>), com uma única exceção (HrcA de *Enterococcus faecalis*). As seqüências de HrcA's formam um banco ligeiramente redundante, entretanto isto não invalida as conclusões deste trabalho.

Tabela 4. Lista das proteínas cujas seqüências foram usadas nas análises neste trabalho, com seus respectivos números de acesso, espécies e linhagens de origem.

Prot.	Organismo ((cepa/isol.) ou (Subsp./Sorvr./Patvr.-cepa/isol.))	Nº no GenBank
HrcA	<i>Acholeplasma laidlawii</i> (PG8)	21205839
HrcA	<i>Agrobacterium tumefaciens</i> (GV3101)	3170206
HrcA	<i>Agrobacterium tumefaciens</i> (C58)	15155234
HrcA	<i>Bacillus anthracis</i> (A2012)	21402363
HrcA	<i>Bacillus halodurans</i> (C-125)	15613907
HrcA	<i>Bacillus sphaericus</i> (1593)	7387777
HrcA	<i>Bacillus stearothermophilus</i> (NUB36)	3122229
HrcA	<i>Bacillus subtilis</i> (168/MB11)	140723
HrcA	<i>Bacillus thermoglucosidasius</i>	12082115
HrcA	<i>Bradyrhizobium japonicum</i> (110spc4)	6433782
HrcA	<i>Brucella melitensis</i> (16M)	17988059
HrcA	<i>Burkholderia fungorum</i>	**
HrcA	<i>Campylobacter jejuni</i> (NCTC11168)	10720039
HrcA	<i>Caulobacter crescentus</i> (CB15N/NA1000)	1708298
HrcA	<i>Chlamydia muridarum</i> (MoPn)	7404388
HrcA	<i>Chlamydia trachomatis</i> (D/UW-3/CX)	15605119
HrcA	<i>Chlamydophila pneumoniae</i> (AR39)	11361011
HrcA	<i>Chlamydophila pneumoniae</i> (CWL029)	15618412
HrcA	<i>Chlamydophila pneumoniae</i> (J138)	8978871
HrcA	<i>Chlorobium tepidum</i> (TLS)	21674306
HrcA	<i>Clostridium acetobutylicum</i> (ATCC 824)	15024208
HrcA	<i>Clostridium acetobutylicum</i> (DMS1731)	267490
HrcA	<i>Clostridium perfringens</i> (13)	18145699
HrcA	<i>Corynebacterium glutamicum</i> (ATCC 13032)	19553489
HrcA	<i>Enterococcus faecalis</i> (V583)	*TIGR: EF1306
HrcA	<i>Fusobacterium nucleatum</i> (Subs. nucleatum-ATCC 25586)	19713541
HrcA	<i>Helicobacter pylori</i> (26695/ATCC700392)	7387772
HrcA	<i>Helicobacter pylori</i> (J99)	4154610
HrcA	<i>Lactobacillus sakei</i> (LTH681)	7387779
HrcA	<i>Lactobacillus sanfranciscensis</i> (DSM20451)	21912937
HrcA	<i>Lactococcus lactis</i> (Subs. lactis-IL1403)	13878527
HrcA	<i>Lactococcus lactis</i> (Subs.cremoris-MG1363)	1176136
HrcA	<i>Leptospira interrogans</i> (Sorovar Copenhageni-Wijnberg)	7387774
HrcA	<i>Listeria innocua</i> (Clip11262)	16413985
HrcA	<i>Listeria monocytogenes</i> (EGD-e)	10720044
HrcA	<i>Listeria monocytogenes</i> (10403S)	5689037
HrcA	<i>Mesorhizobium loti</i>	13473879
HrcA	<i>Mycobacterium leprae</i> (TN)	13092803
HrcA	<i>Mycobacterium tuberculosis</i> (H37Rv)	15609511
HrcA	<i>Mycobacterium tuberculosis</i> (CDC1551)	13882153

Prot.	Organismo ((cepa/isol.) ou (Subsp./Sorvr./Patvr.-cepa/isol.))	Nº no GenBank
HrcA	<i>Mycoplasma capricolum</i> (ATCC25416)	3122220
HrcA	<i>Mycoplasma genitalium</i> (ATCC 33530/G37)	1351507
HrcA	<i>Mycoplasma pneumoniae</i> (M129)	13507863
HrcA	<i>Mycoplasma pulmonis</i> (UAB CTIP)	14089555
HrcA	<i>Nitrosomonas europaea</i>	**
HrcA	<i>Nostoc</i> sp. (PCC 7120)	17133179
HrcA	<i>Novosphingobium aromaticivorans</i>	**
HrcA	<i>Ralstonia metallidurans</i>	**
HrcA	<i>Ralstonia solanacearum</i> (GMI1000)	17547368
HrcA	<i>Rhodobacter sphaeroides</i>	**
HrcA	<i>Rhodopseudomonas palustris</i>	**
HrcA	<i>Sinorhizobium meliloti</i> (1021)	15073306
HrcA	<i>Staphylococcus aureus</i> (Subs. aureus-N315)	13701380
HrcA	<i>Staphylococcus aureus</i> (Subs. aureus-Mu50)	14247353
HrcA	<i>Staphylococcus aureus</i> (912)	487327
HrcA	<i>Streptococcus mutans</i> (GS-5)	10720393
HrcA	<i>Streptococcus pneumoniae</i> (CP1200)	7387785
HrcA	<i>Streptococcus pneumoniae</i> (R6)	15458021
HrcA	<i>Streptococcus pneumoniae</i> (TIGR4)	14971989
HrcA	<i>Streptococcus pyogenes</i> (Sorotipo M1-SF370)	15675608
HrcA	<i>Streptomyces albus</i> (J1074)	7387775
HrcA	<i>Streptomyces coelicolor</i> (A3 (2) / M145)	10720041
HrcA	<i>Synechocystis</i> sp. (PCC 6803)	1651883
HrcA	<i>Tetragenococcus halophilus</i>	15277197
HrcA	<i>Thermoanaerobacter tengcongensis</i> (MB4T)	20515958
HrcA	<i>Thermosynechococcus elongatus</i> (BP-1)	22294482
HrcA	<i>Thermotoga maritima</i> (MSB8)	7462402
HrcA	<i>Ureaplasma urealyticum</i> (Sorovar 3 Biovar 1)	13357983
HrcA	<i>Xanthomonas axonopodis</i> (Patovar citri-306)	21107698
HrcA	<i>Xanthomonas campestris</i> (Patovar campestris)	13879040
HrcA	<i>Xylella fastidiosa</i> (9A5C)	10720038
DeoR	<i>Escherichia coli</i> (O157:H7)	118448
LexA	<i>Escherichia coli</i> (O157:H7)	126224
ModE	<i>Escherichia coli</i> (K12)	973218
S2***	<i>Bordetella pertussis</i>	2392177
S3***	<i>Bordetella pertussis</i>	2392184
S4***	<i>Bordetella pertussis</i>	2392179
S5***	<i>Bordetella pertussis</i>	2392181

* = Sequência pré-publicada de projeto de sequenciamento genômico do The Institute for Genomic Research (TIGR - <http://www.tigr.org/>)

** = Sequências obtidas por BlastP contra genomas ainda não finalizados, com sequências parciais depositadas no banco de genomas microbianos do NCBI (<http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html>).

*** = Sequências das cadeias maduras de subunidades da toxina pertussica.

RESULTADOS E DISCUSSÃO

Análises de seqüências: análise geral, motivo HTH, e domínio amino-terminal

Conforme comentado na introdução, há um certo consenso de que o domínio de ligação ao DNA das HrcA's é amino terminal, e de que esta atividade é exercida através da interação do DNA com um motivo HTH presente naquele domínio. Portanto, a primeira meta de uma análise rigorosa das seqüências da família HrcA torna-se, obviamente, localizar quais regiões destas seqüências são suspeitas de abrigarem o motivo HTH.

Através da análise de um número significativo (cerca de 15) das seqüências de HrcA's listadas em Materiais e Métodos, buscando-se similaridades das mesmas com outras famílias de proteínas ligadoras de DNA existentes nos bancos de dados do InterPro, SMART, Pfam, e PROSITE, foi possível reconhecer similaridades fracas, porém sistemáticas com as proteínas das famílias LexA, e DeoR, especialmente esta última (não mostrado). Mais que isto, estas similaridades se concentram no terminal amino das HrcA's analisadas. As proteínas de ambas as famílias, LexA e DeoR, possuem elas próprias motivos HTH de ligação ao DNA em seus terminais amino: DeoR de *Escherichia coli* tem um motivo HTH típico do aminoácido 24 ao 43 (Valentin-Hansen *et al.* 1985), enquanto LexA do mesmo organismo apresenta um motivo HTH com pouco usuais 21 aminoácidos, do 28 ao 48, sendo o aminoácido extra localizado na alça que liga as duas hélices do motivo, entre a glicina da alça e a hélice de reconhecimento (Fogh *et al.* 1994).

Usando-se as seqüências de DeoR e LexA de *E. coli* como "protótipos" das seqüências de suas famílias, fez-se um alinhamento com estas seqüências e as seqüências das HrcA's disponíveis até então, utilizando-se o programa ClustalW, e apesar das grandes

diferenças entre os motivos HTH de LexA e de DeoR, inclusive de sequência, as hélices de reconhecimento dos motivos de ambas proteínas se alinharam com a mesma região das HrcA's, tanto quando era alinhada apenas uma daquelas proteínas com as HrcA's, como quando alinharam-se ambas as seqüências com as seqüências das HrcA's (figura 2 - alinhamentos apenas com LexA ou apenas com DeoR não mostrados).

A região de homologia das hélices de reconhecimento de DeoR e LexA com as HrcA's tem similaridade marcadamente maior, nestas últimas, com a hélice de DeoR que com a de LexA, embora os alinhamentos das HrcA's com uma ou outra daquelas proteínas isoladamente mostre uma similaridade geral significativamente maior entre os amino terminais de LexA e das HrcA's, que a existente entre os amino terminais de DeoR e das HrcA's (não mostrado). Isto provavelmente é reflexo de que como um todo, o domínio de ligação ao DNA das HrcA's é mais similar ao amino terminal de LexA que ao amino terminal de DeoR, embora o motivo HTH em si seja mais similar entre as HrcA's e DeoR que entre as HrcA's e LexA. Esta é uma possibilidade interessante, vez que LexA já possui estrutura determinada, enquanto que nenhuma DeoR teve ainda obtidas informações estruturais detalhadas, em parte por se tratar de uma família de proteínas cuja conformação nativa é constituída de oligômeros de ordem superior (tetrâmeros ou oligômeros maiores), ao invés de homodímeros como é o caso de LexA. As diferenças de oligomeria, aliás, são um indício de que esta hipótese pode estar correta, visto que recentemente uma HrcA foi descrita como sendo um dímero (Martirani *et al.*, 2001).

CLUSTAL W (1.81) multiple sequence alignment

```

HrcA_Mgen_ATCC33530/G-37_gi_1      -----MKN-----LTPRQAQILKAIINEYIAYAI 24
HrcA_Mpne_ATCC29342/M129_gi_1      -----MKN-----LTTRQAQILKAIINEYIAYPV 24
HrcA_Upar_Sorovr.-3_gi_107200      -----MKNTKTSQEQVIELKLSDRQKKVLKSIVDEYTI SAT 36
HrcA_Uure_Sorovr.-3_Biovr.-1       -----MKNTKTSQEQVIELKLSDRQKKVLKSIVDEYTI SAT 36
HrcA_Mpul_UAB_CTIP_gi_1408955      -----MNFENK-----EIKLETKKTILFKNIVELYLRTGK 30
HrcA_Hpyl_26695/ATCC700392_gi      -----MLRRIKVGSNLNKKESLLDAFVKTYLQILE 30
HrcA_Hpyl_J99_gi_7387788           -----MLKRIKVGSDLNKKESLLDAFVKTYLQILE 30
HrcA_Cpne_CWL029_gi_15618412       -----MLSVTIVLVGLEMARSKVSKRDSKILDILFATTELYLKTGQ 41
HrcA_Cpne_J138_gi_8978871          -----MLSVTIVLVGLEMARSKVSKRDSKILDILFATTELYLKTGQ 41
HrcA_Cpne_AR39_gi_11361011         MTFSGSFPIMLSVTIVLVGLEMARSKVSKRDSKILDILFATTELYLKTGQ 50
HrcA_Ctra_D/UW-3/CX_gi_156051      -----MENRIEMSQLRASKKDSKISYVLLMATKLYLEGGQ 35
HrcA_Cmur_MOPN_gi_7404388          -----MENRTEMSQLRASKKDSKISHVLLMAIKLYLESGQ 35
HrcA_Spne_R6_gi_15458021           -----MLIDQTERSDEMVTERRQDILNLIIDIFTKTHE 33
HrcA_Spne_CPL200_gi_7387785        -----MVTERQQDILNLIIDIFTKTHE 22
HrcA_Spne_TIGR4_gi_14971989        -----MVTERQQDILNLIIDIFTKTHE 22
HrcA_Spyo_Sorotp.-M1[SF370]_g      -----MITQRQNDILNLIVELFTQTHE 22
HrcA_Smut_GS-5_gi_10720393         -----MITQRQNDILNLIVELFTQTHE 22
HrcA_Llac_sub_lactis(il-1403)      -----MITERQRQILNLIVSLYAKEHT 22
HrcA_Llac_sub_cremoris[MG1363]     -----MITERQRQILNLIVSLYAKDHT 22
HrcA_Bste_NUB36_gi_3122229         -----MLTERQLLILQVIIDDFIRSGQ 22
HrcA_Bthe_gi_120082115             -----MLTDRQLLILQVIIDDFIRSGQ 22
HrcA_Bsub_168/MB11_gi_140723       -----MLTNRQLLILQVIIDDFIRSAQ 22
HrcA_Bsph_1593_gi_7387777          -----MLTNRQLQILQVIIDDFILSAQ 22
HrcA_Lmon_10403S_gi_10720044        -----MLTERQLLIFRAIIDHFTWTIQ 22
HrcA_BHal_C-125_gi_15613907        -----MLTDRQLLILQAIIVDDYIRSAE 22
HrcA_Lsak_LTH681_gi_7387779        -----MLTERQLMILKEIIRLFTESGQ 22
HrcA_Thal_gi_15277197              -----MLTQRQDNILHQIINHYNLNGK 22
HrcA_Saur_sub_aureus_N315_gi_       -----MITDRQLSILNAIVEDYVDFGQ 22
HrcA_Saur_sub_aureus_MU50_gi_       -----MITDRQLSILNAIVEDYVDFGQ 22
HrcA_Saur_912_gi_1176137           -----MITDRQLSILNAIVEDYVDFGQ 22
HrcA_Xfas_9A5C_gi_10720038         ---MPTVSRKTCIFPSSYQYSMCASFSP TLD SRSRQLRLTISCYIQNGE 47
HrcA_Xcam_Pv._campestris_gi_1       -----MRASQSPMLDPRARHLRLTIARYIRDGE 29
HrcA_Atum_C58_gi_15155234          -----MGFSAPL---SKDQAS-LLDERSREIFRRIVEGYLDTGE 35
HrcA_ATum_Gv3101_gi_7387776        -----MGFSAPL---SKDQAS-LLDERSREIFRRIVEGYLDTGE 35
HrcA_Smel_1021_gi_15073306         -----MVLRNPG---TKGIAS-ALDERSGEIFRRIVESYLESSE 35
HrcA_Mlot_gi_13473879              -----MTKAIDPG---SQPLALQSLDMRSRDI FRRIVDSYLRDGE 37
HrcA_Bjap_110spc4_gi_6433782       -----MAHHDPIHLIAPRAGLAQLNERSRDI FRQIVESYLATGE 39
HrcA_Ccres_CB15N/NA1000_gi_17       -----MTQLFPG-PIVRTPGLAELDARARDI FRRVYESYLETGE 38
HrcA_Mtub_H37Rv_gi_15609511        -----MGSADERRFEVLRAIVADFVATQE 24
HrcA_Mtub_CDC1551_gi_13882153      -----MGSADERRFEVLRAIVADFVATQE 24
HrcA_Mlep_TN_gi_13092803           -----MGNADERRFEVLRAIVADFVATKE 24
HrcA_Salb_J1074_gi_7387775         -----MLSERRLEVLRAIVQDYVGTEE 22
HrcA_Scoe_A3_2_gi_10720041         -----MLSERRLKVLRVQDYVGTEE 22
HrcA_Cace_ATCC824_gi_15024208       -----MEMEERKLKILQAIINDYINNGE 23
HrcA_Cace_DSM1731_gi_267490        -----MEMEERKLKILQAIINDYINNGE 23
HrcA_Mcap_ATCC25416_gi_312222      -----MLTKRQVKILQITVEEFIKTNQ 22
HrcA_Lint_Sorovr.-Copenhageni_     -----MDLTERHKRILKALVDKFIQENR 23
HrcA_Ssp_FCC6803_gi_1651883        -----MVKPLRLNDRHQILRATVQHYIATAE 27
HrcA_Cjej_NCTC11168_gi_107200      -----MMKSRDKKDLILESIIQTYLLDNA 24
HrcA_Tmar_MSB8_gi_7462402          -----MRRLNRKNNEALKKLNDQRKVLVCIVREYIENKK 35
LexA_Ecoli_O157_H7_gi_126224      -----MKALTARQQEVEFDLIRDHISQTGM 24
DeoR_Ecoli_K12_gi_118448          -----METRREERIGQLLQELKRSDK 21

```

```

HrcA_Mgen_ATCC33530/G-37__gi_1      PVGSKLLTKKYFKN-LSGGTLRNEMAALEKKGFLKKNHISSGRVPSQIGY 73
HrcA_Mpne_ATCC29342/M129__gi_1      PVGSKLLTKKYFKN-LSGGTLRNEMAVLEKEGYLKKNHISSGRIPSQLGY 73
HrcA_Upar_Sorovr.-3__gi_107200      PVSSKLLVQKEFKD-QSSATIRNEMMFLEKNGLIEKQHISGGRVPSLKG 85
HrcA_Uure_Sorovr.-3_Biovr.-1__      PVSSKLLVQKEFKD-QSSATIRNEMMFLEKNGLIEKQHISGGRVPSLKG 85
HrcA_Mpnl_UAB_CTIP__gi_1408955      PIGSKFLVEQDYGLGKSSATIRNMMHEFEEIGLLEKSHISSGRIPSTLGL 80
HrcA_Hpyl_26695/ATCC700392__gi      PISSKRLKELADLK-ISCATIRNYFQILSKEGMLYQAHSSGARLPTFKAF 79
HrcA_Hpyl_J99__gi_7387788           PISSKRLKELANLK-VSCATIRNYFQILSKEGMLYQAHSSGARLPTFKAF 79
HrcA_Cpne_CWL029__gi_15618412       PVGSKTLKESFCSD-LSTATIRNYFAELEAEGFLKKNHTSGGRIPTDLAL 90
HrcA_Cpne_J138__gi_8978871          PVGSKTLKESFCSD-LSTATIRNYFAELEAEGFLKKNHTSGGRIPTDLAL 90
HrcA_Cpne_AR39__gi_11361011         PVGSKTLKESFCSD-LSTATIRNYFAELEAEGFLKKNHTSGGRIPTDLAL 99
HrcA_Ctra_D/UW-3/CX__gi_156051      PVGSKLLKETCYSD-LSSATIRNYFAQLETDFLRKNHISGGRIPTDLAF 84
HrcA_Cmur_MOPN__gi_7404388          PVGSKTLKETCYSD-LSSATIRNYFAQLETDFLRKNHISGGRIPTDLAF 84
HrcA_Spne_R6__gi_15458021           PVGSK--ALQESIN-SSSATIRNDMAELEKQGLLEKAHTSSGRMPVAVGF 80
HrcA_Spne_CP1200__gi_7387785        PVGSK--ALQESIN-SSSATIRNDMAELEKQGLLEKAHTSSGRMPVAVGF 69
HrcA_Spne_TIGR4__gi_14971989        PVGSK--ALQESIN-SSSATIRNDMAELEKQGLLEKAHTSSGRMPVAVGF 69
HrcA_Spyo_Sorotp.-M1[SF370]__g      PVGSK--ALQRTID-SSSATIRNDMAELEKQGLLEKAHTSSGRMPSPAGF 69
HrcA_Smut_GS-5__gi_10720393         PIGSK--TLQNSIA-SSRATIRNDMAELEKGLLEKAHTSSGRLPVQEGY 69
HrcA_Llac_sub_lactis[i1-1403]__      PIGSK--SLLDISIQ-ASSATIRNDMKVLEKGLIQKEHTSSGRVPSVSGY 69
HrcA_Llac_sub_cremoris[MG1363]__     PIGSK--SLLDISIQ-ASSATIRNDMKALERLGLIQKEHTSSGRIPVSVGY 69
HrcA_Bste_NUB36__gi_312229          PVGSRRLSKKHEIA-LSSATIRNEMADLEELGYIEKTHVSSGRVPSEKGY 71
HrcA_Bthe_gi_120082115              PVGSRRLSKKHEIS-LSSATIRNEMADLEELGYIEKTHASSGRVPSEKGY 71
HrcA_Bsub_168/MB11__gi_140723       PVGSRRLSKKDEIT-FSSATIRNEMADLEELGFIEKTHSSSGRVPSEKGY 71
HrcA_Bsph_1593__gi_7387777          PVGSRQLSKKQGIT-FSPATIRNEMADLEELGYIEKTHSSSGRVPSEKGY 71
HrcA_Lmon_10403S__gi_10720044        PVGSKNLLKEKGLP-YSSATIRNEMGVLEEGYIEKTHSSSGRVPSEKGY 71
HrcA_BHal_C-125__gi_15613907         PVGSRSLSKREDIT-FSPATIRNEMADLEELGFLEKPHSSAGRIPVQKGY 71
HrcA_Lsak_LTH681__gi_7387779        PVGSKKLMSELPKH-VSSATIRNDMALEENGLIEKTHSSSGRVPSEKGY 71
HrcA_Thal_gi_15277197               PIGSKTLMEEG-IA-ASSATIRNEMKTLLEEGYLLVPHSSSGRIPVSLKGY 70
HrcA_Saur_sub_aureus_N315__gi_       PVGSKTLIERHNLN-VSPATIRNEMKQLEDLNYIEKTHSSSGRSPSQLGF 71
HrcA_Saur_sub_aureus_MU50__gi_       PVGSKTLIERHNLN-VSPATIRNEMKQLEDLNYIEKTHSSSGRSPSQLGF 71
HrcA_Saur_912__gi_1176137           PVGSKTLIERHNLN-VSPATIRNEMKQLEDLNYIEKTHSSSGRSPSQLGF 71
HrcA_Xfas_9A5C__gi_10720038         PVGSKTLAQAGLD-ISPATIRNILDLEELGLLNSPHTSAGRVPTAHGY 96
HrcA_Xcam_Pv._campestris__gi_1       PVGSKTLAQHAGLD-VSPATIRNILDLEDIGLLSSPHTSAGRVPTAHGY 78
HrcA_Atum_C58__gi_15155234           PLGSRSLRLLPMS-LSPASVRNVMSDLEELGLIYSPHISAGRLPTQTGL 84
HrcA_ATum_Gv3101__gi_7387776        PLGSRSLRLLPMS-LSPASVRNVMSDLEELGLIYSPHISAGRLPTQTGL 84
HrcA_Smel_1021__gi_15073306         PLGSRSLRLLPVS-LSPASVRNVMSDLEELGLIYSPHVSAGRLPTQLGL 84
HrcA_Mlot_gi_13473879               PVGSRSLRLLPSS-LSPATIRNVMSDLEHLGLIYAPHISAGRLPTQAGL 86
HrcA_Bjap_110spc4__gi_6433782       PVGSRNISRLIAMP-LSPASVRNVMADEQLGLIYAPHTSAGRLPTLGL 88
HrcA_Ccres_CB15N/NA1000__gi_17      PVGSRITISKGG-VA-LSPASIRNTMQDLAQLGLLDAPHTSAGRMPHTAGL 86
HrcA_Mtub_H37Rv__gi_15609511        PIGSKSLVERHNLG-VSSATVRNDMAVLEAEGYITQPHTSAGRVPTKGY 73
HrcA_Mtub_CDC1551__gi_13882153      PIGSKSLVERHNLG-VSSATVRNDMAVLEAEGYITQPHTSAGRVPTKGY 73
HrcA_Mlep_TN__gi_13092803           PIGSRALVERHNLG-VSSATIRNDMAVLEAEGYITQPHTSAGRVPTKGY 73
HrcA_Salb_J1074__gi_7387775         PVGSKALTERHRLG-VSPATVRNDMALEDEGYIAQPHTSAGRIPTDKGY 71
HrcA_Scoe_A3_2__gi_10720041         PVGSKALTERHRLG-VSPATVRNDMALEDEGYIAQPHTSAGRIPTDKGY 71
HrcA_Cace_ATCC824__gi_15024208      PVGSRITAKKYNLG-ISSATIRNEMADLEEMGYIEQLHTSSGRKPSDKGY 72
HrcA_Cace_DSM1731__gi_267490        PVGSRITAKKYNLG-ISSATIRNEMADLEEMGYIEQLHTSSGRKPSDKGY 72
HrcA_Mcap_ATCC25416__gi_312222      PVGSKRILELDDIK-ISSATIRNESAILHEGYIEQLHTSSGRTPSTKGY 71
HrcA_Lint_Sorovr.-Copenhageni__     PVGSKTLFDKHDIG-LSPASIRTVLKDLEDFGYLASKHTSGGRIPTKGY 72
HrcA_Ssp_PCC6803__gi_1651883        PVGSHTLAQEQYQFA-VSSATIRNALGQLEKAGLLYQPHVSAGRVPSDSGY 76
HrcA_Cjei_NCTC11168__gi_107200      PIGSNELNSNLCP---ASTIRVYLKRLSDEGLITQLHISSGRIPTILTM 71
HrcA_Tmar_MSB8__gi_7462402          PVSSQVRVLEVSNI-ESSATIRNDMKLEYLYGYIYQPHTSAGRIPTDKGL 84
LexA_Ecoli_O157_H7__gi_126224      PPTRAEIAQRLGFR--SPNAAEEHLKALARKGVIEIVSGASRGIR----- 67
DeoR_Ecoli_K12__gi_118448          ---LHLKDAALLGVSEMTIRRDNLNHSAPVLLG-----GY 55

```

Figura 2. Porção amino terminal do alinhamento das sequências de DeoR e LexA de *E. coli* com as sequências das HrcA's não repetidas encontradas no NCBI, utilizando-se o programa ClustalW e os parâmetros padrão. Os motivos HTH de LexA e DeoR sublinhados.

A possibilidade de que o amino-terminal das HrcA's tenha um dobramento semelhante ao do amino-terminal de LexA foi inicialmente sugerida por Huynen *et al.* (1998) para a proteína de *Mycoplasma genitalium*, no conjunto de um trabalho sobre genômica estrutural naquele organismo. Para descobrir se esta seria uma similaridade específica da proteína de *M. genitalium* ou se esta é de uma característica geral das proteínas HrcA, foi utilizado o programa FUGUE (<http://www-cryst.bioc.cam.ac.uk/~fugue/>, Shi *et al.*, 2001). Este programa utiliza matrizes de substituição ambiente-específicas para comparar seqüências fornecidas de proteínas com aquelas de proteínas de estrutura conhecida (matrizes ambiente-específicas levam em conta características físico-químicas dos aminoácidos substituídos nas estruturas, tais como acessibilidade do solvente, estrutura secundária e situação de pontes de hidrogênio). Em uma busca com parâmetros padrão, é fornecida uma seqüência a qual é inicialmente submetida a rodadas de PSI-Blast contra o banco do PIR (Protein Information Resource - <http://www-nbrf.georgetown.edu/>), a partir do qual é gerado um alinhamento múltiplo que é então submetido ao FUGUE. Esta busca foi feita com a seqüência de HrcA de *C. crescentus*, e o resultado apresenta a estrutura do domínio amino-terminal de LexA com o melhor escore (8,73). Para se ter noção do que significa este escore, um valor acima de 6,0 é dado como um alinhamento certo (99 % de confiança), valores entre 4,0 e 6,0 classificam o alinhamento como provável (95 % de confiança), valores entre 3,5 e 4,0 o alinhamento é classificado como marginal (90 % de confiança), e valores entre 2,0 e 3,5 classificam o alinhamento com "palpite" ("guess", 50 % de confiança). Entretanto, examinando o resultado da busca do PSI-Blast feita no banco do PIR, encontramos que foram incluídas no alinhamento muitas proteínas que não pertenciam a família HrcA. Foi feito então um alinhamento das seqüências de HrcA's, listadas em Materiais e Métodos,

com o uso do ClustalW, e submetido o alinhamento resultante ao FUGUE: o resultado foi um alinhamento com o domínio amino-terminal LexA de escore 12,55 em primeiro lugar. Mais ainda, os três alinhamentos de melhor escore são com estruturas do tipo "*Winged Helix*", indicando ser esta uma característica geral dos domínios amino-terminais das HrcA's. Adicionalmente, o escore do alinhamento com LexA é significativamente maior que os demais (mais que o dobro do segundo maior escore), indicando que deve existir uma similaridade estrutural entre os domínios amino-terminais de HrcA e LexA. O alinhamento de HrcA de *C. crescentus* com LexA gerado pelo FUGUE está mostrado na figura 3.

```

                                10          20          30          40          50
hsd1lea      (1)          mkaLtarQqeVFdlIrdhisqtgmpPtraeIAqrlg
HrcA C. cres      MTQLFPGPIVVRTPLAELDARARDIFRRVVESYLETGEPVGSRTISKGGV
                                aaaaaaaaaaaaaaaaaa      aaaaaaa

                                60          70          80          90          100
hsd1lea      (37) FrspnaAeehLkaLarkgVieivsgasrgIrllee
HrcA C. cres      ALSPASIRNTMQDLAQLGLLDAPHTSAGRMPHTAGLRMFVDGFLEVGDVA
                                aaaaaaaaaaaaaa      bbb      bbb

```

Figura 3. Alinhamento entre os domínios amino-terminais de HrcA de *C. crescentus* e LexA de *E. coli*, gerado pelo FUGUE. Abaixo do alinhamento as regiões de alfa hélice e cadeia beta são indicadas pelas letras "a" e "b" respectivamente, enquanto que na sequência de LexA as regiões com estas estruturas secundárias são indicadas pelas cores vermelha e azul, respectivamente.

Um ponto a ser observado na figura 3 é que segundo a estrutura do domínio amino-terminal de LexA publicada originalmente (Fogh *et al.*, 1994), as cadeias em conformação beta estendiam-se do resíduo 55 ao 59 e do 65 ao 69, sendo um resíduo em cada direção maiores que aquelas indicadas no alinhamento da figura 3, por conseguinte.

Um detalhe encorajador para o alinhamento dos domínios amino-terminais de LexA e de HrcA de *C. crescentus*, é que o dendrograma gerado pelo alinhamento feito com o uso do ClustalW entre LexA de *E. coli* e as HrcA's, agrupava LexA no ramo das HrcA's formado pelas proteínas de *Synechocystis sp.*, *Leptospira interrogans*, e proteobactérias dos grupos α , β , e γ , estando LexA especialmente próxima da HrcA de *Synechocystis* (não mostrado - com as seqüências de HrcA atualmente disponíveis, LexA agrupa-se com as proteínas do ramo dos micoplasmas, o que vem de encontro ao resultado de Huynen *et al.* (1998)). Acreditando que a proximidade fosse suficiente, foi tentado o alinhamento local de LexA de *E. coli* com o HrcA de *C. crescentus*, cujos resultados são mostrados na figura 4. Como resulta aparente dos alinhamentos, estes são coerentes dentro de uma ampla faixa de matrizes Blosum, e até o aminoácido 57 de LexA eles são idênticos àquele gerado pelo FUGUE (Figura 3). Mais que isso, como o valor da matriz Blosum usada é inversamente proporcional à distância filogenética presumida entre as seqüências analisadas, um alinhamento significativo usando-se uma matriz alta como Blosum85 sugere que as seqüências alinhadas são separadas por uma distância filogenética pequena (para efeito de comparação, as três matrizes Blosum mais elevadas permitidas pelo programa são Blosum100, Blosum90 e Blosum85).

A - Blosum40

HrcA *C. crescentus*

```

15 laeldarardifrrvvesyletgepvgsrtiskggvalspasirntmqdlaqlglld 71
   : | || :::| : : :|| | |.. || . :. ||: |:::
1  mkaltarqgevfdlirdhisqgtgmpptraeiaqrlgfrspnaaeehlkalarkgvie 57

```

LexA *E. coli*

B - Blosum85

HrcA *C. crescentus*

```

18 ldarardifrrvvesyletgepvgsrtiskggvalspasirntmqdlaqlglld 71
   | || :::| : : :|| | |.. || . :. ||: |:::
4  ltarqgevfdlirdhisqgtgmpptraeiaqrlgfrspnaaeehlkalarkgvie 57

```

LexA *E. coli*

Figura 4. Alinhamentos locais entre HrcA de *C. crescentus* e LexA de *E. coli*, obtidos com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se diferentes matrizes para o cálculo do alinhamento (todos os demais parâmetros são padrão): (A) Blosum 40, e (B) Blosum 85. A região sublinhada é o motivo hélice alça hélice de LexA: observe-se a fraca similaridade desta região em comparação com as regiões anterior e posterior do alinhamento.

A região imediatamente posterior ao alinhamento apresentado na figura 4 é de nosso interesse porque anteriormente foram construídos mutantes da HrcA de *C. crescentus* nas posições 81 e 87, cujas mutações são P81A e R87A respectivamente, e cujas proteínas se mostraram amplamente prejudicadas em sua capacidade de se ligar à sequência CIRCE *in vivo* e *in vitro* (Susin, M. F., Baldini, R. L., & Gomes, S. L., manuscrito em preparação). Estas posições estão incluídas na região de maior conservação da família HrcA (Tabela 5), como será visto adiante. No alinhamento da figura 3, a prolina 81 está alinhada com uma

das cadeias da folha beta de LexA, enquanto que a arginina 87 encontra-se pouco depois. Resolvemos então ver se conseguíamos alinhar esta região com LexA usando o programa do GCG para alinhamento local (BestFit), e uma matriz Blosum mais permissiva. O resultado é mostrado na figura 5.

LexA *E. coli*

```
1 mkaltarqqevfdlirdhisqtgmpptraeiaqrlgfrspnaaeehlkal 50
: | || :::| : : :|| | |. || . : |
15 laeldarardifrrvvesyletgepvgsrtiskggvalspasirntmqdl 64
```

HrcA *C. crescentus*

LexA *E. coli*

```
51 arkgvieiv.sgasrgi..rllqeeeeeglplvgrvaagepllaqqhie.g 96
|: |::: :|. | : :|| | | : :
65 aqlglldaphtsagrmpthaglrmfvdgflvvgdvaeqekraiearlavk 114
```

HrcA *C. crescentus*

Figura 5. Porção amino terminal do alinhamento local entre HrcA de *C. crescentus* e LexA de *E. coli*, obtidos com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), determinando-se a matriz Blosum35 para o cálculo do alinhamento (todos os demais parâmetros são padrão). A região sublinhada é o suposto motivo hélice-alça-hélice de HrcA, enquanto que a porção alinhada com a região de folha beta de LexA está duplamente sublinhada.

Segundo este novo alinhamento, a prolina 81 localiza-se na alça entre as cadeias da folha beta pregueada, e a arginina 87 alinha-se ao primeiro aminoácido após a folha beta do domínio amino-terminal de LexA, conforme descrito por Fogh *et al.* (1994). Com base na alta conservação da região onde estão aqueles aminoácidos (figura 6), na alta propensão da

alanina de formar alfa hélices, e nas informações de atividade prejudicada de ligação a CIRCE dos mutantes P81A e R87A, formulamos a hipótese de que na proteína tipo selvagem, a prolina 81 e a arginina 87 estão fazendo parte de uma folha beta, ou próxima a uma no caso da arginina, e que esta estrutura se encontra na interface entre os monômeros. Já no caso dos mutantes P81A e R87A, esta estrutura foi convertida em alfa hélice pela presença da alanina, causando alguma espécie de impedimento estérico para a oligomerização da proteína, que seria inativa na forma monomérica. Para testar esta hipótese foi usado o procedimento descrito por Zeng *et al.* (1996) para teste com mutantes de atividade comprovadamente reduzida, de proteínas multiméricas com atividade cooperativa. Este basicamente consiste em inserir um plasmídeo expressando a proteína mutante de atividade prejudicada na linhagem tipo selvagem do organismo, e verificar a atividade da proteína em questão através de um gene repórter, comparando-se com a atividade no tipo selvagem e no mutante nulo. Se o fenótipo for semelhante ao tipo selvagem significa que a mutação prejudicou a formação de hetero-oligômeros entre os monômeros mutante e tipo selvagem, e portanto os únicos oligômeros formados serão homo-oligômeros de proteína tipo selvagem, cuja atividade é normal. Se o fenótipo se assemelhar ao do mutante nulo, significa que a formação de hetero-oligômeros não foi prejudicada, por que os hetero-oligômeros de proteína tipo selvagem e mutante tem atividade significativamente menor que os oligômeros do tipo selvagem, o que resulta no fenótipo alterado. Isto significa que a mutação foi feita em um aminoácido realmente importante para o mecanismo de atividade da proteína.

Este trabalho foi feito pela doutoranda Michelle Susin em nosso laboratório, com os genes mutantes de HrcA de *C. crescentus* S56A, P81A, e R87A clonados no vetor pMR20, que replica-se em *Caulobacter*, sob o controle do promotor selvagem de *hrcA*, usando o

gene *groEL* da própria bactéria como repórter. A mutante na posição 56 substitui um aminoácido da hélice de reconhecimento do suposto motivo hélice alça hélice de HrcA, enquanto as demais foram comentadas anteriormente. Os resultados preliminares dos ensaios de Western blot indicam que a mutante S56A inserida em *C. crescentus* NA1000 faz a bactéria expressar bem mais GroEL que as outras duas mutantes, embora todas expressem mais GroEL que a linhagem NA 1000 na qual foi inserido apenas o plasmídeo sem inserto (comunicação pessoal). Este resultado indica que as mutações para alanina nas posições 81 e 87 de HrcA de *C. crescentus* devem afetar a oligomerização da proteína, em contraste com a mutação na posição 56, que aparentemente é parte de região que interage com o DNA.

Já havia sido relatado na literatura que a similaridade e identidade entre as seqüências das HrcA's varia amplamente ao longo dos alinhamentos, havendo regiões fortemente conservadas intercaladas a regiões grandemente variáveis (Roberts *et al.*, 1996). Portanto decidiu-se explorar as regiões conservadas dos alinhamentos para ter idéia de quais funções poderiam estar sendo desempenhadas por cada uma delas. Entre os programas do SeqWeb 1.0 (GCG), do pacote Winscosin, figura o PlotSimilarity, o qual produz um gráfico da similaridade ou identidade ao longo de um alinhamento do PileUP, que é o programa gerador de alinhamentos múltiplos do pacote. Os gráficos de similaridade e identidade para a família HrcA são mostrados na figura 6.

A

B

Figura 6. Gráficos de (A) Similaridade e (B) Identidade ao longo do alinhamento das seqüências de HrcA's listadas em Materiais e Métodos com o programa PileUP usando matriz Blosun45. HTH, região do suposto motivo HTH (posições 159 a 181 do alinhamento); * região conservada com estrutura presumível de folha beta; ** região conservada carboxi-terminal. Linha tracejada indica o valor da similaridade ou identidade média do alinhamento.

Dos gráficos gerados podemos tirar algumas conclusões: (1) notadamente a região carboxi-terminal do alinhamento é bem menos conservada que o amino-terminal; (2) a região do suposto motivo HTH é altamente conservada, abrangendo o segundo maior pico de similaridade e o mais alto pico de identidade, o que condiz com a alta conservação da sequência operadora CIRCE que ele deve reconhecer; (3) após o motivo HTH existe uma sequência altamente conservada que é o segundo maior pico de identidade e o maior pico de similaridade, região esta que presumivelmente forma uma folha beta envolvida na interação entre os monômeros no oligômero (este trabalho); (4) existe uma última sequência conservada no extremo carboxi-terminal do alinhamento.

Embora estes gráficos dêem uma idéia numérica da conservação relativa ao longo do alinhamento, eles têm a desvantagem de não apresentarem claramente os limites das regiões conservadas, e de não exibirem a sequência consenso dos motivos conservados. Para tanto recorreremos aos programas MEME (Multiple EM for Motif Elicitation - Bailey & Elkan, 1994) e MAST (Motif Alignment and Search Tool - Bailey & Gribskov, 1998), especialmente o primeiro. O motivo mais conservado da família HrcA começa no décimo primeiro aminoácido do suposto motivo HTH, que vem a ser o primeiro com possibilidades de interação com o DNA segundo Brian Mathews, e se estende numa sequência de 35 aminoácidos, estando presente em todas as 71 sequências de HrcA's processadas. O consenso deste motivo é **SSATIRNEMADLEELGLIEKPHTSSGRVPSEKGYR**, onde os aminoácidos em negrito são invariáveis: os dez primeiros correspondem ao último da alça e mais os nove da hélice de reconhecimento do suposto motivo HTH, e o motivo conservado ainda abrange, além desta parte do HTH, a região conservada imediatamente subsequente mostrada nos gráficos do PlotSimilarity. O E-value calculado pelo programa para este motivo é de 2×10^{-1555} , o que reflete bem a conservação desta sequência nas proteínas da

família HrcA. Já o segundo motivo mais conservado é amino-terminal ao anterior, se estende por 29 aminoácidos e termina dentro da hélice estabilizadora do suposto motivo HTH. O consenso é TERQLxILRAIVEDYIATGEPVGSKTLSE, e a sequência foi detectada pelo programa em todas as HrcA's processadas com uma única exceção (*Corynebacterium glutamicum*), a qual se deve a forma truncada do motivo apresentada neste caso, que não possui os seis aminoácidos amino-terminais do mesmo (não mostrado). Ainda assim o E-value calculado pelo programa para este motivo é de $1,3 \times 10^{-775}$, mostrando a elevada conservação do mesmo. Como dito acima, este segundo motivo conservado acaba no meio da hélice estabilizadora do suposto motivo HTH, em geral no quarto aminoácido desta hélice, com exceção das HrcA's de streptococcus (*S. mutans*, *S. pyogenes*, e as três linhagens de *S. pneumonie*), *Campylobacter jejuni* e *Lactococcus lactis* (ambas as subespécies), onde o motivo termina no sexto aminoácido do suposto HTH; de *Caulobacter crescentus*, *Tetragenococcus halophilus* e *Enterococcus faecalis*, onde o último aminoácido do motivo conservado é o quinto da suposta hélice estabilizadora; e de *Mycoplasma pulmonis*, em cuja HrcA o motivo conservado amino-terminal acaba no terceiro aminoácido do suposto motivo HTH. O terceiro motivo mais conservado encontrado pelo MEME corresponde ao pico de similaridade e identidade carboxi-terminal visto nos gráficos gerados pelo PlotSimilarity. Está presente em 66 das 71 seqüências e seu E-value calculado foi de $2,7 \times 10^{-670}$. O seu consenso é GGRxVGTIGVIGPTRMDYSRVIPLDxVA, tendo portanto 29 aminoácidos. As seqüências de HrcA's onde este motivo não aparece são as de *Campylobacter jejuni*, *Helicobacter pylori* (ambas as linhagens), *Thermotoga maritima*, e *Mycoplasma capricolum*. As HrcA's de *C. jejuni* e *H. pylori* são significativamente menores que as demais (aproximadamente 265 aminoácidos), o que indica que este último motivo pode ter

sido deletado de suas HrcA's. Em particular em *H. pylori* há motivos para suspeitar de que isto tenha realmente acontecido, já que nesta bactéria demonstrou-se que HrcA controla a expressão apenas do operon de seu próprio gene com *grpE* e *dnaK*. Neste organismo existe ainda um outro sistema de controle de transcrição de proteínas de choque térmico, HspR/Hair, que é o principal mecanismo de controle da resposta ao choque térmico nesta bactéria (Spohn & Scarlato, 1999; Homuth *et al.*, 2000; Spohn *et al.*, 2002), e que também controla o operon *hrcA-grpE-dnaK*. Assim, é possível que esta região conservada carboxi-terminal de HrcA tenha um papel na interação da proteína com os chaperones que modulam sua atividade, ou ainda que possa ter um papel na agregação que ocorre com a proteína no choque térmico. A distribuição dos motivos conservados nas HrcA's submetidas à análise está na Tabela 5, numerados por ordem de conservação, acompanhados do valor de probabilidade de todos os motivos terem surgido, com seus respectivos graus de conservação, em cada sequência em particular por simples acaso.

Com base nesses resultados mostrados, o domínio de ligação ao DNA de HrcA está, segundo o nosso modelo, localizado no amino-terminal com um motivo HTH e uma estrutura "*Winged Helix*" semelhante àquela do domínio amino-terminal de LexA. No caso de HrcA de *C. crescentus* o motivo HTH estende-se do aminoácido 43 ao 62, e a prolina 81 encontra-se em uma folha beta situada entre os aminoácidos 69 e 86. Já em HrcA de *Xyllera fastidiosa* o motivo HTH se estende do aminoácido 53 ao 72, e a folha beta situa-se entre os aminoácidos 79 ao 96.

Tabela 5. Distribuição dos motivos conservados nas seqüências de proteínas HrcA.

A última questão pendente em relação ao domínio amino-terminal das HrcA's é a que diz respeito aos limites deste domínio. Em uma proteína com uma distribuição de conservação tão desigual quanto em HrcA, uma primeira abordagem desta questão a partir das seqüências seria tomar as duas seqüências das HrcA's mais dissimilares (divergentes), e fazer um alinhamento local entre elas usando a matriz de substituição menos permissiva possível. Estas seqüências são as das HrcA's de *Staphylococcus aureus* e de *Streptococcus pneumoniae*, as quais foram alinhadas localmente com uso da matriz Blosum100 para o cálculo do alinhamento (figura 7). Como o nosso interesse está nas HrcA's de *C. crescentus*

e *X. fastidiosa*, as seqüências destas foram alinhadas localmente nas mesmas condições anteriores com a seqüência de *S. aureus* (figura 8), que é menos similar às das proteobactérias que aquela de *S. pneumonie*, na expectativa de um resultado coerente, o que efetivamente ocorreu. Não só estes alinhamentos terminavam na mesma posição da seqüência de *S. aureus* que aquele feito com a de *S. pneumonie*, como também terminam em posições equivalentes (alinhadas) nas HrcA's de *C. crescentus* e *X. fastidiosa*.

```
HrcA S. pneumonie
4 erqqdnlndiftkthepvgskal.qe.sinsssatirndmaelekqglllekahtssgrmpsvagfqyyv 73
  .|. | || : : | |||| | || ||||| : .|| || | . | |. || || .| : ||
9 drhqqilratvqhyiataepvgshlaqeyqfavssatirnalqglekaglllyqphvsagrpsdsgyriyv 80
HrcA S. aureus
```

Figura 7. Alinhamento local entre as HrcA's de *S. pneumonie* e de *S. aureus*, obtido com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote Winsconsin (GCG), usando-se a matriz Blosun100 para o cálculo do alinhamento (todos os demais parâmetros padrão).

Tudo isto, no entanto, seria apenas mera conjectura, se não houvesse sido publicada no começo deste ano a seqüência da HrcA da β proteobactéria *Ralstonia solanacearum*, no artigo sobre o sequenciamento do genoma deste organismo (Salanoubat *et al.*, 2002). Por sorte, a HrcA desta bactéria é altamente similar à de *X. fastidiosa* (43 % de identidade e 51 % de similaridade), ficando atrás em similaridade apenas daquelas das *Xanthomonas*. O mais interessante no alinhamento das HrcA's de *X. fastidiosa* e *R. solanacearum* (figura 9), porém, é que justamente na posição em que acaba o alinhamento local estridente entre as HrcA's de *X. fastidiosa* e *S. aureus*, começa uma região de grandes e freqüentes intervalos no alinhamento entre as HrcA's de *X. fastidiosa* e *R. solanacearum*.

Como se tratam de duas proteínas altamente similares, isto só seria possível se a posição em questão estiver próxima ao limite entre dois diferentes domínios.

(A)

HrcA *S. aureus*

```
9 ilnaivedyvdgqpvgsktlierhnlvspatirnemkqledlnyiekthsssgsrpsqlgfryyvnrllle 80
. | .: | : |.||||| | . |.:||||| | : ||. | . |. |. |. |. | : | : |. | .
34 llrtliscyiqngepvgsktlaqqagldispatirniladleelgllnsphsagrvtahgyrmfvdsllvq 105
```

HrcA *X. fastiosa*

(B)

HrcA *C. crescentus*

```
25 ifrrvvesyletgepvgsrt.iskggvalspasirntmqdlaqlglldaphtsagrmphaglrmlfvdgflle 95
| : || | . |.||||:| | : |||. || |. | | .. |. |. |. |. | : |. | |
9 ilnaivedyvdgqpvgsktlierhnlvspatirnemkqledlnyiekthsssgsrpsqlgfryyvnrllle 80
```

HrcA *S. aureus*

(C)

HrcA *C. crescentus*

```
18 ldarardifrrvvesyletgepvgsrtis.kggvalspasirntmqdlaqlglldaphtsagrmphaglrmlf 89
||. |. | . | : |. . |||||:|.. . | .|||. || : || .|||. . ||||| || | |||
27 ldsrsrqllrtliscyiqngepvgsktlaqqagldispatirniladleelgllnsphsagrvtahgyrmf 99
```

HrcA *X. fastiosa*

HrcA *C. crescentus*

```
90 vdgflle 95
|| .
100 vdsllvq 105
```

HrcA *X. fastiosa*

Figura 8. Alinhamento local entre as HrcA's de (A) *X. fastiosa* e *S. aureus*, (B) *C. crescentus* e *S. aureus*, e (C) *C. crescentus* e *X. fastidiosa* (apenas região amino terminal). Todos os alinhamentos obtidos com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se a matriz Blosum100 para o cálculo do alinhamento (todos os demais parâmetros padrão).

Já no caso da HrcA de *C. crescentus*, que também tem uma grande similaridade com aquela de *R. solanacearum* (41 % de identidade e 47 % de similaridade), o alinhamento com a sequência de *R. solanacearum* (figura 10) apresenta um único grande intervalo que começa apenas dois resíduos adiante da posição onde começam os intervalos no alinhamento da figura 9. Destes dados podemos supor com certa segurança, portanto, que o domínio amino terminal de HrcA de *C. crescentus* termina aproximadamente entre os aminoácidos 95 e 97.

Accession	Species	Position	Sequence	Length
HrcA	<i>R. solanacearum</i>	1	mdkratillktlieryiaegqpvg	24
HrcA	<i>X. fastidiosa</i>	1	mptvsrkticfpssyqysmcasfptldsrqrllrtliscyiqngepvg	50
HrcA	<i>R. solanacearum</i>	25	srtlskysgldlspatirnvmsdleemgfiasphtsagrvtprgyrlfv	74
HrcA	<i>X. fastidiosa</i>	51	sktlaqqagldispatirniladleelgllnsphtsagrvtahgyrmfv	100
HrcA	<i>R. solanacearum</i>	75	dtmltagpldildgmhdrmaqgiagtiaeqvraqlashgsessqrviaaaa	124
HrcA	<i>X. fastidiosa</i>	101	dslvq.....mqppseddi.....rrlrvmaggg...tqallgnas	134
HrcA	<i>R. solanacearum</i>	125	qtlsnlsqfagivmtprrth.afrqiefirlsetrillivvtpegdvqnr	173
HrcA	<i>X. fastidiosa</i>	135	eilsamthfvgvvsaprreqfvfrhidfvpldarqimailifadnevqnr	184
HrcA	<i>R. solanacearum</i>	174	iihtdtpytps qlveaanyinahyggmtfdavrerlrgelnalqadmtsl	223
HrcA	<i>X. fastidiosa</i>	185	vieprrvyepgelervsnylnahfigrtladirttvlcelrkakdemeql	234
HrcA	<i>R. solanacearum</i>	224	mqaaveagssav..sdedpvvisgerkllevedlssssmdklrrlfavfeq	271
HrcA	<i>X. fastidiosa</i>	235	lahsldlasqmlvpndsedivvtgqtrmlalqdl.sdmldrrelfeifas	283
HrcA	<i>R. solanacearum</i>	272	ktgllqllldvssraegvqifiggeesslvphedmavitapyevdgkivgtl	321
HrcA	<i>X. fastidiosa</i>	284	kreilqllertinapgvrrifigeetgmvsmedislvtapyaaahgqvlgl	333
HrcA	<i>R. solanacearum</i>	322	gvigpmmayervipivditakllsstlsqh...	352
HrcA	<i>X. fastidiosa</i>	334	gvigpkrmaydrviplvqvvaqvlgtalepptmp	367

Figura 9. Alinhamento global entre as HrcA's de *X. fastidiosa* e *R. solanacearum*, obtido com o uso do programa de alinhamento global (algoritmo de Needleman e Wunsh) do pacote de Wincosin (GCG) e matriz Blosum100 para o cálculo do alinhamento (todos os demais parâmetros são padrão). A região sublinhada em HrcA de *X. fastidiosa* corresponde à sequência de intervalos citada no texto.

```

HrcA C. crescentus      1 mtqlfpgpivrtpglaeldarardifrrvvesyletgepvgsrtisk.gg 49
                               :| || . : :| |. |.|||||.|| |
HrcA R. solanacearum    1 .....mdkratillktlieryiaegqpvgstlskysg 33

HrcA C. crescentus      50 valspasirntmqdlaqlglldaphtsagrmpthaglrnfvdgfvlg.. 97
                               ||||.|| | || :| . .|||||| | | | :|| | | |
HrcA R. solanacearum    34 ldlsparirnvmsdleemgfiasphtsagrvtprgyrlfvdtmltagpl 83

HrcA C. crescentus      98 .....dvaeeakraiearlavkg.rsfeealaeassilsglagg 135
                               :| : | || | | . .| |. || |.
HrcA R. solanacearum    84 dilgmhdrmagqiaqtiaeqvraqlashgssqrviaaaaqtlslsqf 133

HrcA C. crescentus      136 agivvtpvreggvkhvefvplgggqvlavmvfedggqvenrlmrqapgvtp 185
                               |||| || | :|| | :| : | .| |.||.. ||
HrcA R. solanacearum    134 agivmtprthafraqieflrlsetrillivvtpegdvqnriihtdtpytp 183

HrcA C. crescentus      186 salqearnflnarlrgrtlteartemggeldaarrqlnetaarlvdegla 235
                               | | ||. :|| | | : |||. | : || |.
HrcA R. solanacearum    184 sqlveaanyinahyggmtfdavrerlrgelnalqadmtslmqaaaveagss 233

HrcA C. crescentus      236 awsggegdaarslivrgqanlladararedidrvrqlfddleqkgqligll 285
                               | | | :| |. || .| : | || || |. ||
HrcA R. solanacearum    234 avs....dedpvvisgerkllevedlssmdklrrlfavfeqktgllql 279

HrcA C. crescentus      286 ddvrdaegvriyigaetrflslsgssviaapymtgrqkvlgavigvpar 335
                               | |||| | :|| |. | .|| || | : | .|||| |
HrcA R. solanacearum    280 dvssraegvqfiggesslvphedmavitapyevd.gkivgtlgvigpmr 328

HrcA C. crescentus      336 lnyarvipldyrtarvlgrmmdg. 358
                               : | ||||.|| || : | :
HrcA R. solanacearum    329 mayervipivditakllsstlsqh 352

```

Figura 10. Alinhamento global entre as HrcA's de *C. crescentus* e *R. solanacearum*, obtido com o uso do programa de alinhamento global (algoritmo de Needleman e Wunsch) do pacote de Wincosin (GCG) e matriz Blosum100 para o cálculo do alinhamento (todos os demais parâmetros padrão). A região sublinhada em HrcA de *C. crescentus* corresponde ao intervalo citado no texto.

Análises de seqüências : Domínio Carboxi-Terminal

Devido ao insuficiente percentual de identidade entre os amino-terminais de HrcA de *C. crescentus* e LexA de *E. coli*, procurou-se no banco de estruturas do Structural Classification of Proteins (SCOP - <http://scop.mrc-lmb.cam.ac.uk/scop/>) uma proteína da mesma família estrutural do amino-terminal de LexA ("*Winged Helix*") que apresentasse

uma maior similaridade com HrcA de *C. crescentus*, para tentativas de modelagem. Esta busca mostrou-se infrutífera em seu objetivo inicial, mas forneceu um subproduto igualmente interessante. O repressor do operon de utilização de molibdato, ModE, de *E. coli* apresentou um significativo alinhamento local com HrcA de *Xyllela fastidiosa*, entretanto este alinhamento surpreendentemente não ocorria entre os amino-terminais das proteínas (figura 11).

A - Blosum30

HrcA *X. fastidiosa*

```
88 grvptahgyrm..fvdslvqmqqpseddirrlrvemagggtqallgnase 135
   | | | | | : : . | | . | | .. | :      :      | :      |
75 gavltrygqrliqliydl laqi qqkaf.dvlsdddalplnsllaaisrfsl 123
```

ModE *E. coli*

HrcA *X. fastidiosa*

```
136 ilsamthfvgvvsaprreqfvfrhidfvpldarqimailifadnevqnrv 185
   || . . . | : : | |      | : : | . . | . : : | |      . : .
124 qtsarnqwfgtita.rdhddvqghvdvlladgktrlkvaitaqsgarlg 172
```

ModE *E. coli*

HrcA *X. fastidiosa*

```
186 ieprrvyepgeleri vsnylnahfigrtladirrttvlcelrkakdemegll 235
   | . | :      : . |      : : . .      : : : . . : | | :
173 degkevlillkapwvgi.tqdeavaqnadnqlpgiishiergaeqcevlm 221
```

ModE *E. coli*

HrcA *X. fastidiosa*

```
236 ahsldlasqmlvpndsdivvtggt 260
   |      . . || .      . ||.
222 alpdgqgtlcatvpvneatslqqgqn 246
```

ModE *E. coli*

B - Blosum45

HrcA *X. fastidiosa*

88 grvptahgyrm..fvdsfvqmqqpseddirrlrvemagggtqallgnase 135

| | | : | | : . | | | : | . | : : | : |

75 gavltrygqrliqliydliaqiqqkafdvlsddda.lplnsllaaisrfsl 123

ModE *E. coli*HrcA *X. fastidiosa*

136 ilsamthfvgvvsaprreqfvfrhidfvpldarqimailifadnevqnrvieprrv 191

| | . . | : : | | . | : | . | : : : | | . . . | : |

124 qtsarnqwfgtita.rdhddvqqhvdvlladgktrlkvaitaqsgarlglddegkev 178

ModE *E. coli*

C - Blosum80

HrcA *X. fastidiosa*

138 samthfvgvvsaprreqfvfrhidfvpldarqimailifa 177

| | . | : : | | . | : | . | : : : | |

126 sarnqwfgtita.rdhddvqqhvdvlladgktrlkvaita 164

ModE *E. coli*

Figura 11. Alinhamentos locais entre HrcA de *X. fastidiosa* e ModE de *E. coli*, obtidos com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se diferentes matrizes para o cálculo do alinhamento (todos os demais parâmetros padrão): (A) Blosum 30, (B) Blosum 45, e (C) Blosum 80.

Pode-se observar na figura 11 que o alinhamento na região entre os aminoácidos 138 a 177 da HrcA de *X. fastidiosa* mantém-se à medida que a permissividade da matriz usada diminui até a matriz Blosum80 (vale lembrar que o programa permite apenas três matrizes Blosum mais altas que esta). A região de ModE compreendida entre os aminoácidos 122 a

183 corresponde à maior região contígua de um subdomínio com 68 aminoácidos, chamado subdomínio 2a, responsável pela complexação de molibdato pela proteína (Hall *et al.*, 1999). O alinhamento calculado com Blosum80 corresponde a praticamente 63 % desta região, e a 57 % do subdomínio inteiro, enquanto o alinhamento com Blosum45 abrange praticamente 92 % daquela região, o que equivale a quase 84% de todo aquele subdomínio. Mas o que uma proteína que liga molibdato teria em comum com HrcA de *X. fastidiosa* ? Sabe-se que a estrutura do carboxi-terminal de ModE tem o seu melhor alinhamento estrutural com as subunidades S4 e S5 da toxina de *Bordetella pertussis* em complexo com ATP (Hall *et al.*, 1999). Tanto os subdomínios 2a e 2b de ModE, quanto as subunidades S4 e S5 são variantes de um dobramento conhecido como *OB Fold* ("Oligonucleotide/Oligosaccharide Binding Fold"), um dobramento baseado em um barril beta, com arquitetura de "Chave Grega" ("*Greek-key*") de cinco cadeias beta, encimado por uma hélice entre a terceira e a quarta cadeias beta (para topologia e figura no espaço, ver anexo 1).

Os ensaios de migração retardada feitos com HrcA de *C. crescentus* mostraram um pronunciado efeito da presença de ATP em concentrações tão baixas quanto 30 μ M (Susin, M. F., Baldini, R. L., & Gomes, S. L., manuscrito em preparação). Procurou-se então verificar quais as regiões de ModE que se alinhavam estruturalmente com as porções de S4 e S5 responsáveis por interações com o ATP, e se as seqüências das HrcA's que se alinhavam àquelas regiões de ModE tinham alguma similaridade com as respectivas regiões de S4 e S5. Primeiramente deve-se ressaltar que a toxina pertússica é um oligômero de seis subunidades, das quais uma subunidade S4 (o oligômero tem duas iguais) e a subunidade S5 possuem metade dos aminoácidos que fazem interações energeticamente favoráveis com

o ATP (Hazes *et al.*, 1996). Estas subunidades são ainda as únicas que interagem com outras porções da molécula de ATP que não os fosfatos (S5 com a porção básica, e S4 com a porção básica, a ribose, e os fosfatos α e β). Assim sendo estas subunidades são as mais importantes na ligação do ATP à toxina pertússica. Os aminoácidos de S4 que interagem com o ATP são S61, E65, e R69, onde o primeiro resíduo é o penúltimo aminoácido de uma alça que precede a hélice que contém os outros dois resíduos e se estende do aminoácido 63 ao 74. Já a subunidade S5 interage com a adenina do ATP através dos aminoácidos F59 e S62, sendo que o aminoácido 58 é obrigatoriamente glicina para evitar um impedimento estérico que decorreria da complexação do ATP se houvesse outro aminoácido naquela posição: estes resíduos fazem parte da hélice que se estende da posição 52 à 66. Até onde foi possível apurar, a região de ModE que se alinha estruturalmente com os aminoácidos de 61 a 74 de S4 está compreendida entre os aminoácidos 163 a 174, e a que se alinha com os aminoácidos de 52 a 66 de S5 situa-se entre os aminoácidos 235 a 243 de ModE. Este alinhamento estrutural com S5 cria o óbvio problema de como alinhar 15 aminoácidos com 9. A solução então foi tentar alinhar HrcA de *X. fastidiosa* diretamente com a subunidade S5 da toxina pertússica, o que não havia sido possível com a subunidade S4 para a região anterior, mesmo com uma matriz de substituição muito permissiva. O alinhamento obtido, ainda que com uma matriz amplamente permissiva, encontra-se na figura 12. Um detalhe importante é que ao contrário da região anterior, onde existe uma clara correspondência entre os resíduos que interagem com o ATP em S4 e seus supostos correspondentes nas HrcA's, só é possível identificar a glicina conservada nas HrcA's de α e β - proteobactérias, e o aminoácido hidrofóbico seguinte naquelas de α e γ - proteobactérias, não havendo nenhuma outra correspondência clara de aminoácidos

alinhados entre as HrcA's e a subunidade S5. A sequência de S4 dos aminoácidos 61 a 74 é (61) - SSPMEVMLRAVFMQ - (74) e a sequência de S5 dos aminoácidos 52 a 66 é (52) - WFTDMLGFAISAYAL - (66). A figura 13 mostra as comparações entre estas sequências e aquelas de HrcA de *X. fastidiosa* de 178 a 187 e de 243 a 263, mostrando ainda as sequências correspondentes nas HrcA's de outras proteobactérias, incluindo *C. crescentus*.

S5

```

23 ngefcltafmsgrslvracldsaghehdtwfdtmlgfaisayalksrial 72
   |. |   ... | : | .| | | : . :   | : |   ...   :
205 nahfigrtladirrttvlcelrkakdemeqllahsldlasqmlvpndsedi 254

```

HrcA *X. fastidiosa*

S5

```

73 tvedspypgtpgdlllel 89
   .|   .   || .:
255 vvtgqtrlmalqdlldm 271

```

HrcA *X. fastidiosa*

Figura 12. Alinhamento local entre a HrcA de *X. fastidiosa* e a subunidade S5 da toxina pertússica, obtido com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se a matriz Blosum30 para o cálculo do alinhamento (todos os demais parâmetros são padrão). A região sublinhada em HrcA de *X. fastidiosa* alinha-se com a hélice da subunidade S5 da toxina pertússica, que contém os aminoácidos que interagem com o ATP, marcados em negrito. A glicina 58 está sublinhada.

S4	<i>B. pertussis</i> (β)	61	-	SSPMEV MLRAVFMQ	-	74
HrcA	<i>X. fastidiosa</i> (γ)	176	-	FADNEV QNRVIEPR	-	189
HrcA	<i>C. crescentus</i> (α)	167	-	FEDGQ VENRLMRQA	-	180
HrcA	<i>X. campestris</i> (γ)	159	-	FADNEV QNRVIEPR	-	172
HrcA	<i>X. axonopodis</i> (γ)	152	-	FADNEV QNRVIEPR	-	165
HrcA	<i>R. solanacearum</i> (β)	165	-	TPEGDV QNR $\underline{\text{I}}$ IHTD	-	178
HrcA	<i>R. metallidurans</i> (β)	160	-	SPEGDV QNR $\underline{\text{I}}$ IQTD	-	173
HrcA	<i>B. fungorum</i> (β)	211	-	TPEGDV QNRIMATQ	-	224
HrcA	<i>N. europaea</i> (β)	152	-	TPEGDV QNR $\underline{\text{I}}$ ILTE	-	165
HrcA	<i>R. palustris</i> (α)	287	-	AEDGQ VENRVLTLP	-	300
HrcA	<i>R. sphaeroides</i> (α)	151	-	FADGHV ENRVFRPP	-	164
HrcA	<i>N. aromaticivorans</i> (α)	158	-	SEDGGI ENRVIDLP	-	171
HrcA	<i>B. japonicum</i> (α)	171	-	GEDGQ VENRVLTLP	-	184
HrcA	<i>M. loti</i> (α)	169	-	SQNGDV ENRVVEPT	-	182
HrcA	<i>B. melitensis</i> (α)	183	-	MQNGDV ENRVIDLP	-	194
HrcA	<i>S. meliloti</i> (α)	167	-	GEHDQ VENR $\underline{\text{I}}$ IELP	-	180
HrcA	<i>A. tumefaciens</i> (α)	168	-	GDHNQ VENR $\underline{\text{I}}$ IELP	-	179
S5	<i>B. pertussis</i> (β)	52	-	WFTDMLG FAI SAYAL	-	66
HrcA	<i>X. fastidiosa</i> (γ)	234	-	LLAHS L D L ASQMLVP	-	248
HrcA	<i>C. crescentus</i> (α)	227	-	AR L VED G LA A WSGG E	-	241
HrcA	<i>X. campestris</i> (γ)	217	-	LLAHS V D L ASEALVP	-	231
HrcA	<i>X. axonopodis</i> (γ)	210	-	LLAHS V E L ASEALVP	-	224
HrcA	<i>R. solanacearum</i> (β)	225	-	QA A VE A G S SAVSD E D	-	239
HrcA	<i>R. metallidurans</i> (β)	219	-	QA A VE A G S SAM A EDN	-	233
HrcA	<i>N. europaea</i> (β)	212	-	NA A IE I G N DALQ E SS	-	226
HrcA	<i>B. fungorum</i> (β)	269	-	MHA A VT A STDEADTG	-	283
HrcA	<i>R. palustris</i> (α)	347	-	QKV I A A G I ASWSGGD	-	361
HrcA	<i>R. sphaeroides</i> (α)	211	-	HDLVQ S G L AVWENPG	-	225
HrcA	<i>N. aromaticivorans</i> (α)	218	-	ADLVQ R G L VVWTQDA	-	232
HrcA	<i>B. japonicum</i> (α)	231	-	QKV I S A G I ASWSGG E	-	245
HrcA	<i>M. loti</i> (α)	229	-	QDLVE K G L AVWAG A E	-	243
HrcA	<i>B. melitensis</i> (α)	243	-	QELV A K G LAVW S GAG	-	257
HrcA	<i>S. meliloti</i> (α)	236	-	QDLVE R G L AIW S GSE	-	250
HrcA	<i>A. tumefaciens</i> (α)	228	-	QDLVE R G L AIWAGDN	-	242

Figura 13. Comparação das seqüências das regiões que ligam ATP nas subunidades S4 e S5 da toxina pertússica e as supostas regiões análogas nas HrcA's de α , β , e γ proteobactérias. Os resíduos que se ligam ao ATP e seus possíveis análogos nas HrcA's estão em negrito. G58 de S5 e seus prováveis substitutos estão grifados.

Observe-se que com relação a S4 é possível estabelecer uma relação nítida, até mesmo surpreendente se consideramos apenas os nove primeiros aminoácidos, entre as seqüências da região em questão com suas correspondentes nas HrcA's, ao contrário do que acontece com S5. No caso dos aminoácidos que se alinham com a serina 62 da subunidade S5, há uma conservação nas HrcA's de α -proteobactérias sendo constante um triptofano naquela posição, cuja cadeia lateral indol possui um grupo N-H que pode fazer o mesmo tipo de ponte de hidrogênio que a hidroxila da serina 62 de S5 faz com a adenina. A glutamina de HrcA de *X. fastidiosa* também pode igualmente substituir a serina 62 nesta interação. É interessante ainda observar que nas HrcA's apresentadas na figura 13, o aminoácido supostamente correspondente à alanina 63 de S5 é serina ou treonina em mais da metade das seqüências, levantando a suspeita de que possa ter havido um evento de inserção do aminoácido anterior à serina/treonina que corresponderia à posição 62 de S5, ao longo da história evolutiva das HrcA's de α , β , e γ -proteobactérias.

Seguindo a linha de raciocínio que estes fatos sugerem, foram procuradas similaridades entre as demais subunidades da toxina pertússica que interagem com ATP, e HrcA de *C. crescentus* ou *X. fastidiosa*. As subunidades citadas, S2 e S3, são produtos de genes parálogos e tem seqüências muito semelhantes (70 % de identidade, e 74 % de similaridade de seqüências de aminoácidos calculados com o programa de alinhamento global do pacote de Wincosin (GCG) e a matriz Blosum100). São proteínas de 199 aminoácidos que possuem dois domínios, sendo o domínio carboxi terminal, de 102 aminoácidos, o responsável pelas interações com o ATP, e o mais conservado entre as duas proteínas (76 % de identidade, e 80 % de similaridade de seqüências de aminoácidos calculados com a ferramenta e parâmetros supracitados). Já o domínio amino-terminal é

responsável pela ligação a oligossacarídeos, especialmente aqueles contendo ácido siálico, da membrana das células alvo, o que leva, em última instância, à internalização da toxina.

Enquanto não se mostrou produtivo alinhar a subunidades S2 e S3 com HrcA de *X. fastidiosa*, realmente foi possível alinhar localmente a região da subunidade S3 onde estão situados os aminoácidos que interagem com ATP, com uma região carboxi terminal, de função até então insuspeita, de HrcA de *C. crescentus* (figura 14), coisa que não foi possível fazer com as subunidades S4 e S5. Os alinhamentos feitos com o ClustalW entre HrcA de *C. crescentus* e S3 ou S2 corroboram este resultado (não mostrado). Melhor que isto, a região apresenta marcada conservação principalmente ao redor dos supostos aminoácidos funcionais (figura 17).

S3

```
139 egryrdmydalrrlllymiysglavrv..hvskeeqyydyedatfqtyaltgisl 191
    : | | : | . | . : | : . | | | . : | : | |
258 dararedidrvrqlfddlegkgqligllddvrdaegvriyigaetrllfslsgssv 312
```

HrcA *C. crescentus*

Figura 14. Alinhamento local entre HrcA de *C. crescentus* e a subunidade S3 da toxina pertússica, obtido com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se a matriz Blosun45 para o cálculo do alinhamento (todos os demais parâmetros padrão). A região sublinhada em HrcA de *C. crescentus* alinha-se com a hélice da subunidade S3 da toxina pertússica, que contém os aminoácidos que interagem com o ATP, marcados em negrito.

Como as regiões correspondentes aos alinhamentos citados até aqui abrangem apenas os primeiros 312 aminoácidos, restando ainda 56 aminoácidos carboxi terminais de

HrcA de *C. crescentus* com função ainda indefinida, foi tentado um alinhamento da subunidade S2 da toxina pertússica com esta região truncada, e somente com ela devido ao alto grau de identidade entre S2 e S3 que levaria a um resultado coincidente ao anterior caso fosse usada a sequência inteira de HrcA de *C. crescentus*. O resultado desta tentativa foi um alinhamento de 31 aminoácidos incluindo toda a hélice de S2 na qual se encontram os aminoácidos que interagem com o ATP (figura 15). Entretanto salta aos olhos que este alinhamento pode ser significativamente melhorado com a introdução de um único intervalo, que resulta em um ganho de 3 identidades no alinhamento (figura 15). A maior vantagem desta correção é a maior conservação aparente dos supostos aminoácidos funcionais, em especial o segundo (lisina - figura 17). Para ter certeza que este alinhamento não estava excessivamente manipulado pela truncagem da sequência de HrcA, submetemos ao programa uma sequência de HrcA de *C. crescentus* truncada apenas na região que alinha-se com S3, para ser alinhada contra a sequência de 54 aminoácidos do extremo carboxi-terminal de S2, e submetemos ainda a sequência de 31 aminoácidos de S2 da figura 15 a um alinhamento global com a sequência inteira de HrcA de *C. crescentus* (figura 16). Os alinhamentos obtidos são todos coerentes, indicando que o alinhamento inicial não se trata de mero resultado da manipulação das seqüências.

Neste ponto deve ser feita uma ressalva quanto a esta última região de HrcA: os últimos onze aminoácidos da região alinhada com a hélice de S2 que interage com o ATP (figura 17) são parte de uma região conservada em quase todas as HrcA's (motivo 3 - exceção feita as HrcA's de *Helicobacter pylori*, *Campylobacter jejuni*, *Mycoplasma capricolum*, e *Thermotoga maritima*), sendo os seis últimos especialmente conservados, conforme encontrado com o uso do programa MEME (Bailey & Elkan, 1994) para busca de motivos conservados, cujos resultados encontram-se sumarizados na tabela 5.

A - Original

HrcA *C. crescentus* (aa de 313 a 358)
 317 ymtgrqkvlgaigvigparlnyarviplvdy 347
 | | :.: . .|. ... :.||
 146 ysrl**rk**mlyliyvagisvrvhvskeeqyydy 176
 S2

B - Corrigido

HrcA *C. crescentus* (aa de 313 a 358)
 318 mtgrqkvlgaigvig.parlnyarviplvdy 347
 : :| | | | .|. ... :.||
 146 ysrl**rk**mlyliyvagisvrvhvskeeqyydy 176
 S2

Figura 15. Alinhamento local entre HrcA de *C. crescentus* e a subunidade S2 da toxina pertússica, obtido com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se a matriz Blosum30 para o cálculo do alinhamento (todos os demais parâmetros padrão): (A) alinhamento original computadorizado; (B) alinhamento corrigido com a introdução manual de um intervalo. A região sublinhada em HrcA de *C. crescentus* alinha-se com a hélice da subunidade S2 da toxina pertússica, que contém os aminoácidos que interagem com o ATP, marcados em negrito.

A - Local, Blosum30

HrcA *C. crescentus* (truncada entre os aa 258 a 312)

317 ymtgrqkvlgaigvigparlnyarviplvdy 347

| | :.: . .|. ... :.||

146 ysrl**rk**mlyliyvagisvrvhvskeeqqyydy 176

S2 (apenas aa de 146 a 199)

B - Global, Blosum65

S2 (apenas aa de 146 a 176)

146ysrl**rk**mlyliyvagisvrvhvskeeqqyydy.. 176

. .|. | | | |.. :. ||

301 etrlfslsgssviaapymtgrqkvlgaigvig.parlnyarviplvdyta 349

HrcA *C. crescentus*

Figura 16. (A) Alinhamento local entre HrcA de *C. crescentus* truncada do aminoácido 258 ao 312, e os 54 aminoácidos carboxi terminais da subunidade S2 da toxina pertússica, obtido com o uso do programa de alinhamento local (algoritmo de Smith e Waterman) do pacote de Winsconsin (GCG), usando-se a matriz Blosum30 para o cálculo do alinhamento (todos os demais parâmetros padrão). (B) alinhamento global entre a sequência de aminoácidos de 146 a 176 de S2, e HrcA de *C. crescentus*, obtido com o uso do programa de alinhamento global (algoritmo de Needleman e Wunsch) do mesmo pacote e matriz Blosum65 para o cálculo do alinhamento (todos os demais parâmetros padrão). As regiões sublinhadas em HrcA de *C. crescentus* alinham-se com a hélice da subunidade S2 da toxina pertússica, que contém os aminoácidos que interagem com o ATP, marcados em negrito.

S3	<i>B. pertussis</i> (β)	146	-	YDAL R RLLYMIYM	-	158
HrcA	<i>C. crescentus</i> (α)	265	-	IDRV R QLFDDLEQ	-	277
HrcA	<i>B. melitensis</i> (α)	282	-	IERL R HLFDDLET	-	294
HrcA	<i>M. loti</i> (α)	268	-	IELL R HLFEDMET	-	280
HrcA	<i>A. tumefaciens</i> (α)	268	-	IDRV R MLFDDLER	-	280
HrcA	<i>S. meliloti</i> (α)	267	-	IERL R MLFDDLEK	-	279
HrcA	<i>B. japonicum</i> (α)	269	-	LERV R LLFDDLET	-	281
HrcA	<i>N. aromaticivorans</i> (α)	255	-	IERV R QLLEELEG	-	267
HrcA	<i>R. sphaeroides</i> (α)	249	-	FERI R TLFDDLER	-	261
HrcA	<i>R. palustris</i> (α)	385	-	LERV R LLFDDLET	-	397
HrcA	<i>X. fastidiosa</i> (γ)	271	-	MDRL R ELFEIFAS	-	283
HrcA	<i>X. campestris</i> (γ)	254	-	LDRL R ELFEAFAS	-	266
HrcA	<i>X. axonopodis</i> (γ)	247	-	LDRL R ELFEAFAS	-	259
HrcA	<i>R. solanacearum</i> (β)	259	-	MDKL R RLFAVFEQ	-	271
HrcA	<i>R. metallidurans</i> (β)	255	-	MDKL R RLFDVFEH	-	267
HrcA	<i>B. fungorum</i> (β)	303	-	MARL R KLFDVFDQ	-	315
HrcA	<i>N. europaea</i> (β)	247	-	LSSL K SLFEMFER	-	259
S2	<i>B. pertussis</i> (β)	146	-	YSRL R KMLYLIYVA	-	159
HrcA	<i>C. crescentus</i> (α)	318	-	MTGR Q KVLGAIGVI	-	331
HrcA	<i>B. melitensis</i> (α)	335	-	RDSE Q RVIGALGVI	-	348
HrcA	<i>M. loti</i> (α)	321	-	RDKD A RVVGALGVI	-	334
HrcA	<i>A. tumefaciens</i> (α)	321	-	RDEEN R VVGAVGVI	-	334
HrcA	<i>S. meliloti</i> (α)	320	-	RDSDD R IVGAVGVI	-	333
HrcA	<i>B. japonicum</i> (α)	322	-	RDAAG H IVGVLGVI	-	335
HrcA	<i>N. aromaticivorans</i> (α)	308	-	RDGD G RVVGAVGVI	-	321
HrcA	<i>R. sphaeroides</i> (α)	302	-	MNADR K IIGAIGVI	-	315
HrcA	<i>R. palustris</i> (α)	438	-	SDGAG H IVGVLGVI	-	451
HrcA	<i>X. fastidiosa</i> (γ)	323	-	YAAH G QVLGVLGVI	-	336
HrcA	<i>X. campestris</i> (γ)	306	-	YTAG G QVLGVLGVI	-	319
HrcA	<i>X. axonopodis</i> (γ)	299	-	YAAN G QVLGVLGVI	-	312
HrcA	<i>R. solanacearum</i> (β)	311	-	YEV D GKIVGTLGVI	-	324
HrcA	<i>R. metallidurans</i> (β)	307	-	YGV D GQIVGTLGVI	-	320
HrcA	<i>B. fungorum</i> (β)	355	-	YEV N GKIVGTLGVI	-	368
HrcA	<i>N. europaea</i> (β)	299	-	YEME G KIVGTVGVI	-	312

Figura 17. Comparação entre as seqüências das regiões que interagem com o ATP nas subunidades S3 e S2 da toxina pertússica e as regiões supostamente correspondentes nas HrcA's de α , β , e γ proteobactérias. Em negrito estão os aminoácidos envolvidos nesta interação e seus possíveis substitutos nas HrcA's.

Segundo o nosso modelo, portanto, o domínio carboxi-terminal de HrcA de *Caulobacter crescentus* começa entre os resíduos 96 e 98, e possui 3 ou 4 dobramentos *OB* ("Oligonucleotide/Oligosaccharide Binding Fold") consecutivos, cada um deles semelhante ao mesmo tipo de dobramento existente nas subunidades S2, S3, S4 e S5 da toxina pertússica. O domínio carboxi-terminal de HrcA de *Xyllela fastidiosa* deve começar por volta do resíduo 106, e tem a mesma arquitetura daquele da proteína de *C. crescentus*. O provável sítio de ligação do ATP é constituído por quatro hélices entre as quais se acomoda o nucleotídeo, a semelhança da toxina pertússica, as quais se estendem, no mínimo, do resíduo 169 ao 176, do 224 ao 237, do 265 ao 278, e do 318 ao 328, levando-se em conta as previsões de estrutura secundária pelos métodos de Chou-Fasman e de GOR (não mostradas). Os resíduos que provavelmente interagem com ATP em HrcA de *C. crescentus* são glutamina 171, arginina 175, leucina 234, arginina 269, glutamina 270, glutamina 322 e lisina 323. Outros resíduos possivelmente envolvidos nesta interação são triptofano 237, serina 238 e glutamina 322.

O nosso modelo estrutural para HrcA de α , β , e γ - proteobactérias é, a grosso modo, semelhante àquele obtido por cristalografia para ModE de *E. coli*, consistindo de um domínio amino-terminal "*Winged Helix*", e de um domínio carboxi-terminal formado por alguns dobramentos *OB* consecutivos. As maiores diferenças são o domínio amino-terminal mais curto e mais parecido com o de LexA que com o de ModE, e o número de dobramentos *OB* no carboxi-terminal, maior no caso das HrcA's, que ligam a volumosa molécula de ATP ao invés do pequeno íon molibdato.

Clonagem do gene *hrcA* de *Xyllela fastidiosa* em vetor de expressão de *E. coli*

Para amplificar o gene *hrcA* de *Xyllela fastidiosa* foi tentada a amplificação e o isolamento do mesmo por reação em cadeia da polimerase (PCR) a partir do cosmídeo LAWRIST contendo o operon *DnaK* da bactéria, porém não obtivemos sucesso, aparentemente devido a uma contaminação dos nossos estoques do cosmídeo. Foram então feitas reações de PCR com DNA genômico de *X. fastidiosa* e o par de iniciadores HRCAX3/HRCAX5 (Materiais e Métodos, Tabela 2). Apesar de terem ocorrido amplificações inespecíficas, o produto principal da reação sempre tinha o tamanho esperado, enquanto as outras bandas sempre tinham tamanhos muito diferentes (figura 18). Um problema da distância entre a extremidade dos iniciadores e seu sítios de restrição obrigou a que o produto das amplificações fossem clonados em pGEM-T Easy, após corte do gel e eletroeluição da banda de tamanho correto. Foram obtidos 20 clones positivos da transformação de *E. coli* DH5 α com a ligação do pGEM-T Easy com o produto do PCR, e dois deles tiveram os insertos de seus plasmídeos seqüenciados com iniciadores T7 e SP6. As seqüências destes insertos confirmaram tratar-se do gene *hrcA* de *X. fastidiosa*, conforme verificação feita usando-se o Blast do NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). A presença dos iniciadores nas posições corretas dos insertos também foi verificada, por inspeção visual das seqüências.

Minipreps dos plasmídeos seqüenciados foram digeridas com *Bam*HI e *Hind*III para obtenção do fragmento contendo o gene *hrcA* de *X. fastidiosa*. As digestões foram submetidas a eletroforese em gel de agarose e a banda daquele fragmento foi cortada do gel e eletroeluída. O fragmento assim purificado foi ligado ao vetor pProEx-1 digerido com as mesmas enzimas. Este vetor produz proteínas recombinantes contendo uma cauda de poli-

histidina que permite a purificação da proteína com o uso de colunas de afinidade de agarose contendo níquel imobilizado por complexação. Parte do fragmento contendo o gene *hrcA* foi usada na marcação de sonda radioativa, já que pProEx-1 não possui seleção branca-e-azul, sendo então necessária varredura por hibridação de colônia. Com este procedimento foram selecionados 12 clones, todos confirmados como positivos para a presença do inserto do gene *hrcA* por digestão com *Bam*HI e *Hind*III. Destes foram selecionados dois que expressaram a proteína HrcA em quantidade ligeiramente maior em um experimento preliminar de expressão, tendo sido denominados clones A e B.

Figura 18. Análise por eletroforese em gel de agarose (1 %) do produto de PCR da região codificadora do gene *hrcA* a partir do DNA genômico de *X. fastidiosa*.

Expressão e purificação de HrcA de *Xyllela fastidiosa*

Como já foi descrito anteriormente, dois clones contendo o gene *hrcA* de *X. fastidiosa* adequadamente clonado no vetor pProEx-1, entre os sítios de *Bam*HI e *Hind*III, foram selecionados em um experimento de expressão preliminar. Nenhum dos clones expressa uma quantidade muito grande de HrcA, provavelmente devido a um problema de degradação da proteína *in vivo*, em virtude de uma instabilidade intrínseca da sequência da proteína. Analisando as sequências das HrcA's selvagens e recombinantes de *C. crescentus* e *X. fastidiosa* com o ProtParam Tool do Expert Protein Analysis System (ExPASy - <http://us.expasy.org/tools/protparam.html>), um dos parâmetros calculados é o "Instability index" (Guruprasad *et al.*, 1990). Este índice é derivado de um estudo estatístico de proteínas estáveis e instáveis *in vivo*. São classificadas como estáveis aquelas cujo índice calculado é inferior a 40, enquanto que aquelas cujo índice é superior a este limite são classificadas como instáveis. A proteína de *C. crescentus* apresentou índice de instabilidade de 35,18 para a selvagem e de 34,20 para a recombinante, sendo predita como estável em ambos os casos. Já a proteína de *X. fastidiosa* apresentou índices de 49,24 para a recombinante e de 51,10 para a selvagem, como seria de se esperar para uma proteína instável *in vivo*. Como *X. fastidiosa* e *E. coli* são ambas γ proteobactérias, é provável que seus sistemas proteolíticos sejam muito semelhantes, e possivelmente uma proteína de *X. fastidiosa* expressa em *E. coli* terá estabilidade semelhante à apresentada no organismo original.

Mesmo assim a expressão é visível em gel de poliacrilamida corado com azul de comassie (figura 19) e em western blot (figura 20), tanto utilizando anticorpo monoclonal anti poli-histidina como usando antisoro anti-HrcA de *C. crescentus*. Inicialmente foi

tentado o protocolo de expressão e purificação usado em nosso laboratório para HrcA de *C. crescentus*. Entretanto HrcA de *X. fastidiosa* aparentemente agrega-se, isoladamente ou co-agregando-se com outras proteínas no sobrenadante, durante a diálise a 4 °C, talvez devido a hidrofobicidade e massa da proteína de *X. fastidiosa*, que são maiores que aquelas de *C. crescentus*. Foram testadas várias condições de expressão alterando-se o meio de cultura (LB, 2xTY, TB, e Circle Growth), temperatura (30 ou 37 °C), concentração do indutor (0,6 ou 1 mM), tempo de expressão (3, 4, ou 6 horas), e momento de indução (fase exponencial precoce - D.O._{600nm} entre 0,45 a 0,65 - ou tardia - D.O._{600nm} entre 0,65 a 0,85). As condições em que se consegue uma maior expressão são exatamente aquelas em que a proteína forma corpos de inclusão, o que é coerente com a hipótese de degradação da fração solúvel da proteína. Foi decidido, em virtude da agregação da proteína durante a diálise, tentar uma condição de expressão em que se formam corpos de inclusão, utilizando-se 4 horas de expressão a 37 °C de temperatura em meio TB, e um protocolo de desnaturação das proteínas incluídas nestes corpos. Foi tentada uréia 8 M como agente desnaturante, mas pouquíssima proteína foi solubilizada, indicando uma elevada hidrofobicidade desta no precipitado (não mostrado). Entretanto foi observado que um tempo maior (1 hora) de contato do lisado com DOC 2 %, a temperatura ambiente, não só aumenta a solubilização da proteína no sobrenadante como também aumenta a sua solubilidade nas lavagens posteriores com DOC 2 %, sem que se precise aumentar o tempo de permanência do ressuspenso nestas lavagens. Após a primeira lavagem a proteína sai quase pura nas lavagens seguintes. Observou-se ainda que é obtido um maior rendimento da proteína nas lavagens quando a cultura está em fase exponencial mais tardia, isto é, D.O._{600nm} entre 0,65 e 0,85, no momento da indução da expressão. Contudo, apesar dos vários artifícios tentados

não foi possível, infelizmente, obter a proteína em uma forma solúvel, devido a sua forte tendência a agregar-se. Por tudo isso ficou decidido encerrar os trabalhos com a proteína de *Xyllela fastidiosa* e prosseguir os estudos apenas com a HrcA de *C. crescentus*.

Figura 19. Análise por eletroforese em gel de poliacrilamida (SDS-PAGE) dos extratos protéicos dos clones A e B em diferentes tempos de indução expressão de HrcA recombinante de *X. fastidiosa* (T = 30 °C).

Figura 20. Experimentos de "Western blot" de extratos protéicos de *E. coli* contendo o plasmídeo pProEx/hrcAXf. As células foram coletadas após diferentes tempos de indução da expressão de HrcA recombinante de *X. fastidiosa* ($T = 30\text{ }^{\circ}\text{C}$). (A) experimento feito com anticorpo anti-poli-histidina. Controle +: proteína com cauda de poli-histidina. (B) experimento feito com antisoro anti-HrcA de *C. crescentus*. M: marcador de peso molecular.

Expressão e purificação de HrcA de *Caulobacter crescentus*

HrcA (recombinante) de *C. crescentus* é obtida na fração solúvel de extratos de *E. coli* com o protocolo estabelecido pela doutoranda Michelle Susin em nosso laboratório (M. F. Susin, R. Baldini, e S. L. Gomes, manuscrito em preparação; para maiores detalhes ver Materiais e Métodos), e foi produzida em escala preparativa para concentração e estudos estruturais sem maiores dificuldades. Uma modificação introduzida posteriormente no protocolo original foi a adição de DOC imediatamente antes da sonicação das células, pois a proteína não apresenta atividade sem a adição e posterior diálise deste detergente. Este possivelmente age em HrcA semelhantemente ao descrito por Doñate *et al.* (1998) para Triton X-100 e Lubrol em aspartato aminotransferase mitocondrial e citossólica. Esta pequena modificação melhora ligeiramente o rendimento da purificação por prevenir parcialmente a agregação da proteína imediatamente após a lise celular.

Apesar de ativa, a proteína de *C. crescentus* apresentou um problema de microagregação após a sua purificação, verificado por espalhamento de luz dinâmico, e em uma tentativa inicial de contornar tal imprevisto optou-se pelo uso de detergentes. Um detergente ideal para este fim tem que responder a alguns critérios, por exemplo: não pode conter grupos aromáticos que viriam a interferir nos estudos espectroscópicos na região do ultra violeta; deve ter efeito abaixo da concentração micelar crítica, para evitar a formação de micelas exclusivamente do detergente, que interferem na caracterização de tamanhos de partículas; e por fim deve ser, preferencialmente não iônico, para evitar interações desta natureza entre o detergente e a proteína. A dificuldade de chegar a uma concentração adequada de um detergente que respondesse a todos estes critérios forçou a uma mudança de abordagem. Por sugestão do Dr. Maurício Baptista, que colabora conosco

nas medidas de espalhamento de luz, foi estudado o uso de ciclodextrinas para este fim. As ciclodextrinas são oligossacarídeos circulares contendo 6, 7, ou 8 resíduos de glicose (α , β , e γ ciclodextrina respectivamente), cuja cavidade interior é hidrofóbica, e os grupos hidroxila hidrofílicos ficam voltados para o exterior. Isto torna estas substâncias ideais para carrear pequenas moléculas apolares em meio aquoso, ou para proteger grupos hidrofóbicos de moléculas maiores, como proteínas, da exposição a moléculas de água. Após um levantamento bibliográfico criterioso, a escolha recaiu em uma adaptação do método usado por Bär *et al.* (2000) para fosfofrutoquinase-1 de *S. cerevisiae*, conforme detalhado em Materiais e Métodos, em virtude da demonstrada interação da α ciclodextrina com a proteína por este método. Em nosso protocolo acreditamos que a α ciclodextrina cumpre a dupla função de capturar as moléculas de detergente, extraíndo-o das moléculas de proteína com as quais estaria interagindo, e de proteger as regiões hidrofóbicas da proteína, expostas após a saída do detergente e que estariam levando a proteína a agregar-se. Este protocolo, após otimização, pode inclusive possibilitar a purificação da proteína de *X. fastidiosa* na fração solúvel do extrato celular, devido ao fato de que o detergente em interação com a proteína ser substituído *in situ* pela ciclodextrina, e ainda de não haver, neste caso, nenhuma fase de diálise na qual a proteína possa vir a se agregar.

Dicroísmo Circular

A figura 21 mostra um típico espectro de dicroísmo circular (CD) no ultra-violeta distante, não tratado, obtido para HrcA de *C. crescentus*. Esta análise reflete o conteúdo de cada tipo de estrutura secundária da proteína. Observa-se um máximo no comprimento de onda de aproximadamente 190 nm, e dois mínimos a cerca de 210 e 220 nm, o que é consistente com uma proteína com uma significativa porção helicoidal, de acordo com a teoria de DeVoe na forma estendida (Fleischhauer *et al.*, 1993: dentre as teorias de dicroísmo circular esta tem três vantagens; ela não assume nada *a priori* sobre as bandas, os cálculos fornecem diretamente os valores de dicroísmo e absorbância em função do comprimento de onda, e na forma original o tratamento é puramente por física clássica (DeVoe, 1964, 1965), o que a torna mais facilmente compreensível). Mais que isso, o mínimo em 210 nm é sutilmente mais intenso do que aquele em 220 nm, característica indicativa de proteínas que possuem regiões de α -hélice e estrutura β segregadas em diferentes domínios (Manalavan & Johson, 1983), o que está de acordo com a hipótese de trabalho proposta.

Uma observação dos espectros de dicroísmo circular que indica a boa representação do modelo em relação a real estrutura de HrcA de *C. crescentus* é a semelhança surpreendentemente elevada entre os espectros normalizados desta proteína e de ModE de *E. coli* (Anderson *et al.*, 1997), em especial quanto à forma das curvas (figura 22). Dos espectros superpostos em escala, a maior diferença observável é uma maior magnitude do espectro de HrcA, na região negativa de θ a partir de 207 nm, aproximadamente, até 260 nm (ver anexo 2 para sobreposição visual em escala), o que pode ser interpretado como resultado de uma maior fração de cadeias beta na estrutura de HrcA de *C. crescentus* em relação àquela de ModE de *E. coli*, observação facilmente explicável pelo nosso modelo

devido ao maior número de dobramentos OB ("Oligonucleotide/Oligosaccharide Binding Fold") existentes no carboxi-terminal de HrcA de *C. crescentus*. Outra possibilidade é que HrcA de *C. crescentus* tenha uma maior fração helicoidal em relação a ModE, associada a uma significativa fração desordenada, a qual dá um sinal negativo no espectro da proteína na mesma região do pico positivo da fração helicoidal (190-200 nm). Isto resultaria em um menor pico positivo a 195 nm, sem tanta influência na região negativa do espectro (acima de 200 nm).

Figura 21. Espectro de dicroísmo circular medido, não tratado, de HrcA de *C. crescentus*, obtido com cubeta cilíndrica de 0,1 mm de caminho ótico ($[HrcA] = 25,2 \mu M$, temperatura ambiente).

Figura 22. Comparação entre os espectros normalizados de dicroísmo circular, no UV distante, de HrcA de *C. crescentus* (A) e ModE de *E. coli* (B, Anderson *et al.*, 1997 - linha sólida, proteína pura; tracejada, com MoO_4^{2-} 1 mM; pontilhada, com WO_4^{2-} 1 mM). Espectro (A) é resultado da normalização do espectro da figura 21.

A desconvolução do espectro para a estimativa dos percentuais de alfa hélice e folha beta foi feita com o uso do pacote de programas CDPro (Sreerama & Woody, 2000), que inclui um programa para determinação da classe de estrutura terciária, o "Cluster", que analisa as características do espectro importantes para esta classificação. Os resultados obtidos com o uso deste programa são mostrados na tabela 6. Note-se que na tabela são apresentados tanto os resultados da análise do espectro medido quanto aqueles do espectro suavizado com o algoritmo de Savitzky-Golay.

Tabela 6. Análise do espectro de CD de HrcA de *C. crescentus* usando o programa Cluster.

Espectro medido (normalizado)		Espectro suavizado	
Class	Score	Class	Score
-----		-----	
All Alpha	3	All Alpha	3
Alfa+Beta	3	Alfa+Beta	4
Alfa/Beta	3	Alfa/Beta	2
All Beta	1	All Beta	1
Denatured	0	Denatured	0
Maximum Score Obtained: 3		Maximum Score Obtained: 4	
PREDICTED Tertiary Class: All Alpha		PREDICTED Tertiary Class: Alfa+Beta	
PREDICTED Tertiary Class: Alfa+Beta			
PREDICTED Tertiary Class: Alfa/Beta			

Embora o programa não tenha sido capaz de determinar inequivocamente a classe estrutural terciária da proteína a partir do espectro medido normalizado, possivelmente devido a um efeito de ruído (figura 23), quando o espectro é suavizado a classe é estabelecida sem ambiguidades, confirmando o que era evidente na inspeção visual. Em termos numéricos, as percentagens de cada estrutura secundária são estimadas por três

diferentes programas do pacote CDPro, cabendo ao pesquisador ter o bom senso de definir qual delas, ou ainda se a média das três, é a que mais se aproxima da realidade, levando em consideração dados de literatura sobre a aplicabilidade de cada método (Sreerama & Woody, 2000). Na tabela 7, temos os resultados dos três programas, o CONTIN/LL, CDSSTR, e SELCON3. Novamente temos primeiramente os resultados do espectro medido, e em seguida aparecem os do espectro suavizado.

Tabela 7. Desconvolução, pelos programas do pacote CDPRO, do espectro de CD de HrcA recombinante de *C. crescentus*.

A - Espectro medido (normalizado)

SAMPLE: HrcAoriginal
 PROGRAM: CDSSTR
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Trn Unrd
 Fractions: .193 .145 .113 .084 .166 .287
 RMSD(Exp-Calc): .171

SAMPLE: HrcAoriginal
 PROGRAM: CONTINLL
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Turn Unrd
 Fractions: .172 .134 .106 .084 .196 .307
 RMSD(Exp-Calc): .213

SAMPLE: HrcAoriginal
 PROGRAM: SELCON3
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Trn Unrd
 Fractions: .171 .137 .088 .076 .219 .304
 RMSD(Exp-Calc): .352

Tabela 7. Continuação**B - Espectro suavizado**

SAMPLE: HrcA suavizado
 PROGRAM: CDSSTR
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Trn Unrd
 Fractions: .176 .144 .127 .093 .187 .268
 RMSD(Exp-Calc): .084

SAMPLE: HrcA suavizado
 PROGRAM: CONTINLL
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Turn Unrd
 Fractions: .168 .144 .106 .085 .214 .284
 RMSD(Exp-Calc): .036

SAMPLE: HrcA suavizado
 PROGRAM: SELCON3
 Protein Set: CA Reference Set
 Structure: H(r) H(d) S(r) S(d) Trn Unrd
 Fractions: .159 .132 .096 .083 .228 .289
 RMSD(Exp-Calc): .193

As colunas de estrutura secundária H, S, Trn, e Unrd, significam Hélice, Cadeia beta (Strand), Alça (Turn), e Desordenada (Unordered), respectivamente. Os índices r e d entre parênteses significam regular e distorcida. A linha RMSD dá o desvio quadrático médio, em fração, do espectro experimental em relação a um espectro teórico com as mesmas frações de estrutura secundária, calculado com cada um dos três programas. No caso da desconvolução do espectro de HrcA de *C. crescentus* recombinante, em fusão com a cauda de poli-histidina, é aparente que devem ser considerados os valores do espectro suavizado,

dada a redução do RMSD, devido à diminuição de efeito do ruído. Parece mais próximo da realidade, em nosso entender, a estimativa do CONTIN/LL a partir do espectro suavizado, não só pelo baixo RMSD (3,6 %) como também pelo fato de que as frações de proteína em cada estrutura secundária, estimadas por este método resultam, ou aproximadamente coincidentes com um dos outros dois métodos, ou próxima à média entre os três métodos. Portanto, a estimativa de percentagens de estrutura secundária aqui apresentada para HrcA recombinante de *C. crescentus* é de 31,2 % de α hélice (16,8 % regular e 14,4 % distorcida), 19,1 % de folhas β (10,6 % regular e 8,5 % distorcida), 21,4 % de alças, e 28,4 % desordenada. Esta apreciável fração desordenada provavelmente é parcialmente devida à cauda amino-terminal de poli-histidina contendo o sítio da protease TEV, cauda esta que tem 24 aminoácidos, o que corresponde a quase 7 % da proteína recombinante.

Com relação ao espectro de dicroísmo da preparação com α ciclodextrina, este não apresentou diferenças significativas com relação àquele da proteína purificada sem o uso do oligossacarídeo (não mostrado).

A figura 24 mostra a sobreposição dos espectros não tratados obtidos para a proteína pura e na presença de 1 mM de ATP. Não há diferença significativa entre eles, o que indica que a presença de ATP não altera a estrutura secundária da proteína, tal qual nas subunidades da toxina pertússica.

Figura 23. Espectros de dicroísmo circular normalizados (medido e suavizado) de HrcA de *C. crescentus* entre 190 e 260 nm, usados na desconvolução.

Figura 24. Espectros de dicroísmo circular medidos (não tratados) de HrcA de *C. crescentus*, em presença (preto) e ausência (vermelho) de 1 mM de ATP, obtidos com cela de 0,1 mm.

Espectros de Fluorescência

Foram obtidos espectros de fluorescência intrínseca da proteína HrcA recombinante de *C. crescentus* na presença de concentrações crescentes de ATP, tanto utilizando o comprimento de onda de excitação ajustado na faixa onde tirosina e triptofano apresentam absorção, quanto na faixa de absorção exclusiva do triptofano (figura 25). Como HrcA de *C. crescentus* tem apenas um triptofano, qualquer ensaio específico para triptofanos será

específico para este único resíduo da proteína. A figura 26 mostra a variação com o aumento da concentração de ATP, da intensidade de fluorescência em $\lambda_{em} = 335$ nm com excitação nos comprimentos de onda $\lambda_{ex} = 280$ nm e $\lambda_{ex} = 297$ nm, em termos absolutos e da relação entre a fluorescência na ausência de ATP e na presença da respectiva concentração do nucleotídeo. A fluorescência sofre uma supressão significativa na presença de concentrações muito baixas de ATP. O fato de praticamente não haver diferença entre as curvas, guardadas as proporções, com excitação nos diferentes comprimentos de onda indica a participação predominante do triptofano no processo que está ocorrendo. Como as curvas de F_0/F mostram um comportamento evidentemente saturante, fugindo portanto do predito pela equação de Stern-Volmer, o ATP age como uma espécie ligada à proteína (Lehrer, 1976). Não foi observada nenhuma alteração significativa nos λ máximos de emissão de fluorescência com o aumento da concentração de ATP, nos espectros com excitação tanto a 280 quanto a 297 nm (figura 25). O λ dos máximos de emissão geralmente se situava entre 335 e 336 nm. Com relação ao micro-ambiente do triptofano isto é indicativo, juntamente com a supressão observada, não de uma interação hidrofóbica com o anel de adenina do ATP, mas de um processo de transferência de energia para um grupo eletricamente carregado ou insaturado quimicamente, o que pode indicar a proximidade entre o triptofano 237 de HrcA e um fosfato do ATP quando este está complexado pela proteína, o que é explicável pelo nosso modelo estrutural.

Não foi possível obter as curvas de fluorescência em função da concentração de ATP para a preparação com α ciclodextrina pois esta forma um complexo de inclusão com o ATP, cuja constante de formação a pH = 6 é de $2,51 \text{ mM}^{-1}$ (Seno *et al.*, 1990). Uma das

metas da otimização do protocolo de purificação é minimizar a concentração de ciclodextrina, a fim de que se possam realizar ensaios de fluorescência com ATP.

Figura 25. Espectros de fluorescência intrínseca, subtraídos dos respectivos brancos e corrigidos quanto a concentração, de HrcA recombinante de *C. crescentus* em presença de várias concentrações de ATP, com comprimento de onda de excitação ajustado a 280 nm (a esquerda) e a 297 nm (a direita).

Figura 26. Emissão de fluorescência intrínseca em 335 nm de HrcA de *C. crescentus* com $\lambda_{\text{ex}} = 280\text{ nm}$ (preto) e $\lambda_{\text{ex}} = 297\text{ nm}$ (azul) em função da concentração de ATP (0, 2.5, 5, 10, 25, e 125 mM), em medida absoluta (a esquerda, escala na cor da curva, unidades arbitrárias) e em relação F_o/F (a direita, na mesma escala para ambas curvas).

Espalhamento de Luz Dinâmico

Os experimentos de espalhamento de luz com laser foram feitos para determinação da ordem de oligomerização de HrcA. Em *Streptococcus thermophilus*, a forma ativa de HrcA foi identificada como sendo o dímero (Martirani *et al.*, 2001), e esperávamos portanto obter um resultado semelhante. Para tanto o aparelho foi calibrado com um kit de proteínas de baixo peso molecular (12,5 a 67 kDa) de raios de Stokes conhecidos, tomando-se as médias ponderadas da primeira população do tratamento para amostras polidispersas e verificando as medidas dos diâmetros com aqueles declarados no manual do kit (figura 27). Para nossa surpresa o diâmetro medido para a população dominante (primeira população, 99,62 % do total) da amostra de HrcA tinha um diâmetro médio (média ponderada) de aproximadamente 13,15 nm (Tabela 8), enquanto que o esperado para o dímero era algo em torno de 8,2 nm. O diâmetro medido resultava em um peso molecular calculado de aproximadamente 140,86 kDa, muito além da faixa de linearidade obtida. Foi decidido então adicionar à curva de calibração as medidas de proteínas de maior peso molecular, o que foi conseguido com o uso de um kit de proteínas de alto peso molecular (158 a 669 kDa) semelhante ao primeiro. As novas medidas esclareceram a situação: parece haver duas faixas de linearidade, uma para baixos pesos moleculares e outra para pesos mais elevados (figura 27), com o diâmetro da BSA (67 kDa) como ponto comum. Entretanto a nova curva de calibração apontou um peso molecular calculado ainda maior para a amostra de HrcA, cerca de 418,6 kDa, o que significa que cada agregado de proteína possui em média cerca de dez cadeias de HrcA (tabela 8).

Figura 27. Curvas de calibração para experimentos de espalhamento de luz com laser, utilizando kit de proteínas de baixo e alto peso molecular: em rosa, dados de experimento de calibração com proteínas de baixo peso molecular; em roxo, dados de experimento de calibração com proteínas de alto peso molecular.

O diâmetro medido para HrcA explica porque em experimentos de eletroforese de proteína em gel não desnaturante executados anteriormente ao início deste estudo, uma

fração muito pequena da proteína penetrava no gel, sendo necessária a desnaturação da proteína com uréia em concentração a partir de 1M, para que uma fração significativa da proteína entrasse no gel (Michelle Susin, comunicação pessoal). Explica ainda, ao menos parcialmente, porque é necessária uma quantidade relativamente grande de proteína para realizar os ensaios de migração retardada em gel (Susin M. F., Baldini R. L., Gomes S. L., manuscrito em preparação). A agregação da proteína cria um problema adicional para os estudos por dicroísmo circular de alteração de estrutura terciária em presença de ATP. Está atualmente sendo avaliada a diminuição de tamanho destes agregados com o uso de um protocolo alternativo de purificação, cujos resultados são discutidos adiante.

Tabela 8. Distribuição da primeira população de sinais de HrcA em espalhamento de luz.

Diâmetro (nm)	% numérico (do total de sinais)
11,48	29,46
11,95	33,01
14,60	23,83
16,47	9,70
18,57	2,76
20,95	0,86
Média ponderada ~ 13,146	Total da 1ª população = 99,62

As medidas de espalhamento de luz feitas com amostras preparadas pelo protocolo alternativo, foram replicadas com o filtrado da etapa de concentração final ("água-mãe" recolhida após a etapa de concentração centrífuga final da proteína - vide Materiais e Métodos) das mesmas como branco, já que o número de sinais em regiões de diferentes tamanhos obtidos na análise da amostra impedia a atribuição segura dos mesmos. A grande diferença entre os sinais da amostra e do filtrado aparece na região de 4,47 a 7,85 nm, e em 11,41 nm de diâmetro, onde aparecem dois conjuntos de sinais nas amostras, sinais estes

inexistentes no filtrado. A presença destes sinais só pode ser atribuída a proteína existente na amostra (não mostrado). O conjunto de sinais entre 4,4 e 7,9 nm tem tamanhos variando aproximadamente daquele do monômero ao do trímero ou tetrâmero (oligomeria calculada igual a 3,56), enquanto que o sinal a 11,41 nm tem massa calculada entre octâmero e nonâmero (oligomeria calculada igual a 8,49). Entretanto a presença de 40 mM de α ciclodextrina impediu a quantificação relativa dos sinais da proteína, já que o programa toma como referência o sinal mais elevado, e a molécula de ciclodextrina, com mais de 1 nm de diâmetro espalha luz significativamente. Ainda assim foi possível observar que os sinais de monômero predominam entre aqueles da proteína, nesta preparação (não mostrado). Para uma quantificação mais detalhada, foram executados experimentos com FPLC, cujos resultados são discutidos a seguir.

FPLC

Foram feitas corridas em FPLC, de amostras contendo a HrcA recombinante de *C. crescentus*, para confirmar os resultados de espalhamento de luz. A curva de calibração da coluna está mostrada na figura 28. A figura 29A mostra os picos referentes a HrcA, quando obtida com a purificação sem ciclodextrina. Observa-se o enorme pico de agregado, de tamanho superior ao limite de exclusão da coluna (1300 kDa). O trio de pequenos picos ao redor de 14,76 ml inclui as formas oligoméricas de ordem inferior (monômero a trímero) de HrcA. Na figura 29B temos o cromatograma completo da HrcA purificada usando α ciclodextrina. Os picos entre 8 e 15 ml tem massa compatível com formas diversas de HrcA. Os picos acima de 19 ml correspondem à ciclodextrina e outras

substâncias de baixo peso molecular. O pico de 8,51 ml (913,36 kDa) é atribuído ao agregado, que tem massa média correspondente 16,5 moléculas de proteína com massa do pico a 14,77 ml (55,37 kDa), atribuído ao monômero. A diferença em relação ao peso teórico do monômero pode ser atribuída ao erro inerente ao método, ou a moléculas de ciclodextrina associadas à proteína (MW = 0,972 kDa), ou a ambos. O pico em 12,35 ml (massa de 163,65 kDa) tem massa do trímero do pico em 14,77 ml, ou do tetrâmero teórico, o que pode ser atribuído à exclusão das moléculas de ciclodextrina durante a oligomerização. A proporção das diferentes formas da proteína é de 47 % no caso do monômero, 28 % no caso de oligômero, e 25 % no caso de agregado, o que indica que o processo de purificação ainda precisa ser otimizado para eliminação da agregação. Estes resultados são semelhantes aos encontrados por espalhamento de luz, e indicam a necessidade de uma otimização adicional do protocolo de purificação.

Figura 28. Curva de calibração para separação por exclusão em gel com coluna Superdex 200 e fluxo de 1 ml / min de tampão A.

A **B**

Figura 29. Cromatogramas de amostras de HrcA de *C. crescentus* (eluição em coluna Superdex 200 com fluxo de 1 ml / min de tampão A): (A) purificada sem ciclodextrina (apenas picos referentes à proteína); (B) purificada com α ciclodextrina.

CONSIDERAÇÕES FINAIS E CONCLUSÃO

O objetivo principal deste trabalho foi realizado a contento: o nosso modelo estrutural de HrcA de *Caulobacter crescentus* permite explicar e prever aquelas propriedades da proteína testadas até o momento. Este modelo indica que a proteína é composta de dois domínios: um domínio amino-terminal, com 95 a 97 aminoácidos, cuja estrutura pertence ao grupo "*Winged Helix*", sendo muito semelhante àquela do domínio amino-terminal de LexA, e um domínio carboxi-terminal, composto de pelo menos três dobramentos OB ("*Oligonucleotide/Oligosaccharide Binding Fold*"), similares aos das subunidades S2 a S5 da toxina pertússica. A combinação de domínios é parecida com aquela do repressor ModE de *E. coli*. O motivo HTH do domínio amino-terminal foi localizado, estendendo-se, no mínimo, do resíduo 43 ao 62: a hélice de reconhecimento do DNA deste motivo compreende os resíduos 54 a 62, pelo menos. Por sua vez, o domínio carboxi-terminal é composto de 261 a 263 resíduos, com uma região de 29 aminoácidos conservados no extremo carboxi-terminal de todas as HrcA's, que na proteína de *C. crescentus* se situa entre os resíduos 323 e 351. Os dobramentos OB teriam como finalidade a complexação de ATP, que se acomodaria entre quatro supostas hélices que se estendem, provavelmente, do resíduo 169 ao 176, do 224 ao 237, do 265 ao 278, e do 318 ao 328, respectivamente. As similaridades com as subunidades da toxina pertússica apontam como prováveis resíduos responsáveis pela interação com o ATP a glutamina 171, a arginina 175, a leucina 234, a arginina 269, a glutamina 270, e a lisina 323, embora esta última já seja parte da região conservada supracitada. Outros aminoácidos que poderiam interagir com ATP são o triptofano 237, serina 238, e a glutamina 322, no entanto não dispomos ainda de evidências que abonem ou refutem esta suposição.

Foi ainda tentada a expressão em *E. coli* e purificação da HrcA de *Xylella fastidiosa*. Com este intuito foi isolada e clonada a região codificadora do gene *hrcA* deste organismo, mas foi impossível no tempo deste projeto expressar e purificar a proteína em quantidade e concentração razoável para testar a sua atividade, e fazer ensaios estruturais.

Com relação à estrutura de HrcA de *Caulobacter crescentus*, as análises de sequência de aminoácidos da proteína tem sido corroboradas até o momento por ensaios com mutantes, realizados em nosso laboratório, e pelas análises espectroscópicas, principalmente de dicroísmo circular e fluorescência, cujos resultados mostram concordância, respectivamente, com as previsões de dobramentos e com o possível sítio de ligação do ATP ao domínio carboxi-terminal, indicados pelos alinhamentos. A sequência dos experimentos de fluorescência deve ainda colaborar para uma melhor definição das características da ligação entre HrcA e ATP, como sua estequiometria, por exemplo.

A análise de seqüências desenvolvida nesta dissertação deve guiar futuras mutações sítio-dirigidas no gene *hrcA* de *C. crescentus*, o que de fato já está acontecendo. Embora seja prematuro falar detalhadamente sobre isto, já nos é possível esboçar um mecanismo de regulação por ATP para a HrcA de *C. crescentus*, entretanto ainda são necessários alguns dados experimentais para confirmar as nossas hipóteses. Já a delimitação do domínio amino-terminal abre as portas para a produção de proteínas compreendendo apenas um dos domínios de HrcA, o que geraria peptídeos de tamanho adequado a estudos por ressonância magnética nuclear. Este estudo de seqüências rendeu ainda a localização de um motivo carboxi-terminal conservado, que pode estar relacionado com a interação da proteína com chaperones de papel regulatório, o qual pode levar à produção de uma proteína truncada cuja regulação tenha sido perdida em função da ausência de interação com aqueles chaperones.

O estudo da ordem de oligomerização da proteína está em vias de ser concluído, e do ponto de vista prático um protocolo de purificação que elimine a agregação da proteína, que está em desenvolvimento, pode levar a uma preparação passível de se obter cristais com os quais se possa chegar, esperamos, a uma estrutura cristalográfica de HrcA de *C. crescentus*, por difração de raios X. O protocolo em estudo pode ainda possibilitar a purificação da proteína de *Xylella fastidiosa* com rendimento aceitável.

Este mestrado teve assim por finalidade, a coleta de dados e preparação para um projeto mais avançado, que terá como objetivo principal definir inequivocamente a estrutura de HrcA de *Caulobacter crescentus*, tanto da proteína tipo-selvagem quanto de suas mutantes pontuais e trucadas já existentes ou que venham a ser produzidas, e estudar espectroscopicamente sua interação com ATP e moléculas análogas, além de estudar a interação da proteína com diversos chaperones, na tentativa de estabelecer quais os mecanismos que ditam a sua atividade como repressor, tanto no choque térmico quanto nos diferentes estágios do ciclo celular de *Caulobacter crescentus*.

BIBLIOGRAFIA

- ANDERSON, L. A., PALMER, T., PRICE, N. C., BORNEMANN, S., BOXER, D. H., & PAU, R. N. 1997. Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the *modABCD* (molybdenum transport) operon of *Escherichia coli*. *Eur. J. Biochem.* 246: 119-126.
- ANTELMANN, H.; BERNHARDT, J.; SCHMID, R. & HECKER, M. 1995. A gene at 333 degrees on the *Bacillus subtilis* chromosome encodes the newly identified sigma B-dependent general stress protein GspA. *J. Bacteriol.* 177: 3540-3545.
- ARSÈNE, F.; TOMOYASU, T.; MOGK, A.; SCHIRRA, C.; SCHULZE-SPECKING, A. & BUKAU, B. 1999. Role of region C in regulation of the heat shock gene-specific sigma factor of *Escherichia coli*, σ^{32} . *J. Bacteriol.* 181: 3552-3561.
- AVEDISSIAN, M. 1996. *Regulação dos genes groES e groEL em Caulobacter crescentus*. Tese de Doutorado, Instituto de Química, Universidade de São Paulo.
- AVEDISSIAN, M. & GOMES, S. L. 1996. Expression of the *groESL* operon is cell cycle controlled in *Caulobacter crescentus*. *Mol. Microbiol.*, 19: 79-89.
- AVEDISSIAN, M.; LESSING, D.; GOBER, J. W.; SHAPIRO, L. & GOMES, S. L. 1995. Regulation of the *Caulobacter crescentus* *dnaKJ* operon. *J. Bacteriol.* 177: 3479-3484.
- BAILEY, T. L. & ELKAN, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California.
- BAILEY, T. L. & GRIBSKOV, M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, 14: 48-54.
- BALDINI, R. L. 1999. Expressão do operon de choque térmico *groESL* durante o ciclo celular de *Caulobacter crescentus*. Tese de Doutorado, Instituto de Química, Universidade de São Paulo.

- BALDINI, R. L.; AVEDISSIAN, M. & GOMES, S. L. 1998. The CIRCE element and its putative repressor control cell cycle expression of the *Caulobacter crescentus* *groESL* operon. *J. Bacteriol.* 180: 1632- 1641.
- BÄR, J.; GOLBIK, R.; HÜBNER, G. & KOPPERSHLÄGER, G. 2000. Denaturation of phosphofructokinase-1 from *Saccharomyces cerevisiae* by guanidinium chloride and reconstitution of the unfolded subunits to their catalytic active form. *Biochemistry*, 39: 6960-6968.
- BENSON, A. K. & HALDENWANG, W. G. 1993. The sigma B-dependent promoter of the *Bacillus subtilis* *sigB* operon is induced by heat shock. *J. Bacteriol.*, 175: 1929-1935.
- BRAIG, K. 1998. Chaperonins. *Curr. Opinion Structural Biol.*, 8: 159-165.
- BRANDEN, C. & TOOZE, J. 1991. *Introduction to protein structure*. Garland Publishing Inc.
- BRUN, Y. V.; MARCZYNSKI, G. & SHAPIRO, L. 1994. The expression of asymmetry during *Caulobacter* cell differentiation. *Annu. Rev. Biochem.*, 63: 419-450.
- BUKAU, B & HORWICH, A. L. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell*, 92: 351-366.
- CRAIG, E. A.; GAMBILL, B. D. & NELSON, R. J. 1993. Heat shock proteins: Molecular chaperones of protein biogenesis. *Microbiol. Rev.*, 57: 402-414.
- DARBY, N.J. & CREIGHTON, T. E. 1993. *Protein structure*. Oxford University Press Inc.
- DERRÉ, I.; RAPOPORT, G.; MSADEK, T. 2000. The CtsR regulator of stress response is active as a dimer and specifically degraded in vivo at 37 °C. *Mol. Microbiol.*, 38: 335-347.
- DERRÉ, I.; RAPOPORT, G.; MSADEK, T. 1999. CtsR, a novel regulator of stress and heat shock response, controls *clp* and molecular chaperone gene expression in gram-positive bacteria. *Mol. Microbiol.*, 31: 117-131.

- DEVOE, H. 1965. Optical properties of molecular aggregates. II. Classical theory of the refraction, absorption, and optical activity of solutions and crystals. *J. Chem. Phys.*, **43**: 3199-3208
- DEVOE, H. 1964. Optical properties of molecular aggregates. I. Classical model of electronic absorption and refraction. *J. Chem. Phys.*, **41**: 393-400
- DINGWALL, A.; GOBER, J. W. & SHAPIRO, L. 1990. Identification of a *Caulobacter* basal body structural gene and a *cis*-acting site required for activation of transcription. *J. Bacteriol.*, **172**: 6066-6076.
- DOÑATE, F.; ARTIGUES, A.; IRIARTE, A. & MARTINEZ-CARRION, M. 1998. Opposite behavior of two isozymes when refolding in the presence of non-ionic detergents. *Prot. Sci.*, **7**: 1811-1820.
- ENGELMANN, S.; LINDNER, C. & HECKER, M. 1995. Cloning, nucleotide sequence, and regulation of *katE* encoding a sigma B-dependent catalase in *Bacillus subtilis*. *J. Bacteriol.*, **177**: 5598-5605.
- ERICSSON, M.; GOLOVLIOV, I.; SANDSTRÖM, G.; TÄRNVIK, A. & SJÖSTEDT, A. 1997. Characterization of the nucleotide sequence of the *groE* operon encoding heat shock proteins chaperone-60 and -10 of *Francisella tularensis* and determination of the T-Cell response to the proteins in individuals vaccinated with *F. tularensis*. *Infect. Immun.*, **65**: 1824-1829.
- FERNANDEZ, R. C. & WEISS, A. A. 1995. Cloning and sequencing of the *Bordetella pertussis* *cpn10/cpn60* (*groESL*) homolog. *Gene*, **158**: 151-152.
- FERREYRA, R. G.; SONCINI, F. C. & VIALE, A. M. 1993. Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the phototrophic sulfur bacterium *Chromatium vinosum*. *J. Bacteriol.*, **175**: 1514-1523.
- FISCHER, B.; RUMMEL, G.; ALDRIDGE, P. & JENAL, U. 2002. The FtsH protease is involved in development, stress response and heat shock control in *Caulobacter crescentus*. *Mol. Microbiol.*, **44**: 461-478.
- FISCHER, H. M.; BABST, M.; KASPAR, T.; ACUÑA, G.; ARIGONI, F. & HENNECKE, H. 1993. One member of a *groESL*-like chaperonin multigene family in

Bradyrhizobium japonicum is co-regulated with symbiotic nitrogen fixation genes. *EMBO J.*, 12: 2901-2912.

FLEICHHAUER, J., KRAMER, B., LÖHKAMPER, R., ZOBEL, E., GRÖTZINGER, J., KRÜGER, P. & WOLLMER, A. 1993. Calculations of the circular dichroism of polypeptide helices with the matrix method and the theory of De Voe, *in: Proceedings of the 5th International Conference on Circular Dichroism*, Pingree Park, Colorado, EUA.

FOGH, R.H., OTTLEBEN, G., RUTERJANS, H., SCHNARR, M., BOELEN, R. & KAPTEIN, R. 1994. Solution structure of the LexA repressor DNA binding domain determined by ¹H NMR spectroscopy. *EMBO J.*, 13: 3936-3944.

GEORGOPOULOS, C.; LIBEREK, K.; ZYLICZ, M. & ANG, D. 1994. Properties of the heat shock proteins of *Escherichia coli* and the auto-regulation of the heat shock response. *In* Morimoto, R. I.; Tissières, A. & Georgopoulos, C. 1994. *The biology of Heat Shock Proteins and Molecular Chaperones*. Cold Spring Harbour Laboratory Press.

GOBER, J. W. & MARQUES, M. V. 1995. Regulation of cellular differentiation in *Caulobacter crescentus*. *Microbiol. Rev.*, 59: 31-47.

GOMES, S. L.; GOBER, J.W. & SHAPIRO, L. 1990. Expression of the *Caulobacter* heat shock gene *dnaK* is developmentally controlled during growth at normal temperatures. *J. Bacteriol.*, 172: 3051-3059.

GOMES, S. L.; JULIANI, M. H.; MAIA, J. C. & SILVA, A. M. 1986. Heat shock protein synthesis during development in *Caulobacter crescentus*. *J. Bacteriol.*, 168: 923-930.

GRANDVALET, C.; RAPOPORT, G. & MAZODIER, P. 1998. *hrcA*, encoding the repressor of the *groEL* genes in *Streptomyces albus* G, is associated with a second *dnaJ* gene. *J. Bacteriol.*, 180: 5129-5134.

GURUPRASAD, K.; REDDY, B. V. & PANDIT, M. W. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng.*, 4: 155-161.

- HALL, D. R.; GOURLEY, D.G.; LEONARD, G.A.; DUKE, E. M. H.; ANDERSON, L. A.; BOXER, D. H. & HUNTER, W. N, 1999. The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds. *EMBO J.*, 18: 1435-1446.
- HAZES, B.; BOODHOO, A.; COCKLE, S. A. & READ, R. J. 1996. Crystal structure of the pertussis toxin-ATP complex: a molecular sensor. *J. Mol. Biol.*, 258: 661-671.
- HECKER, M.; SCHUMANN, W. & VÖLKER, U. 1996. Heat-shock and general stress response in *Bacillus subtilis*. *Mol. Microbiol.*, 19: 417-428.
- HOMUTH, G.; MOGK, A. & SCHUMANN, W. 1999. Post-transcriptional regulation of the *Bacillus subtilis dnaK* operon. *Mol. Microbiol.*, 32: 1183-1197.
- HOMUTH, G.; DOMM, S.; KLEINER, D. & SCHUMANN, W. 2000. Transcriptional analysis of major heat shock genes of *Helicobacter pylori*. *J. Bacteriol.*, 182: 4257-4263.
- HUYNEN, M.; DOERKS, T.; EISENHABER, F.; ORENGO, C.; SUNYAEV, S.; YUAN, Y.; BORK, P. 1998. Homology-based fold predictions for *Mycoplasma genitalium* proteins. *J. Mol. Biol.*, 280: 323-326.
- JENAL, U. & FUCHS, T. 1998. An essential protease involved in bacterial cell-cycle control. *EMBO J.*, 19: 5658-5669.
- KALLIPOLITIS, B. H. & VALENTIN-HANSEN, P. 1998. Transcription of *rpoH*, encoding the *Escherichia coli* heat-shock regulator sigma32, is negatively controlled by the cAMP-CRP/CytR nucleoprotein complex. *Mol. Microbiol.*, 29: 1091-1099.
- KARLS, R. K.; BROOKS, J.; ROSSMEISL, P.; LUEDKE, J. & DONOHUE, T. J. 1998. Metabolic roles of a *Rhodobacter sphaeroides* member of the σ^{32} family. *J. Bacteriol.*, 180: 10-19.
- KELLY, A. J.; SACKETT, M.J.; DIN, N.; QUARDOKUS, E. & BRUN, Y. V. 1998. Cell cycle-dependent transcriptional and proteolytic regulation of FtsZ in *Caulobacter*. *Genes Dev.*, 15: 880-893.

- KOCH, B.; KILSTRUP, M.; VOGENSEN, F. K. & HAMMER, K. 1998. Induced levels of heat shock proteins in a *dnaK* mutant of *Lactococcus lactis*. *J. Bacteriol.*, 180: 3873-3881.
- KRUGER, E. & HECKER, M. 1998. The first gene of the *Bacillus subtilis* *clpC* operon, *ctsR*, encodes a negative regulator of its own operon and other class III heat shock genes. *J. Bacteriol.*, 180: 6681-6688.
- LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of head of bacteriophage T4. *Nature*, 227: 676-685.
- LEE, W. T.; TERSLESKY, K. C. & TABITA, F. R. 1997. Cloning and characterization of two *groESL* operons of *Rhodobacter sphaeroides*: transcriptional regulation of the heat-induced *groESL* operon. *J. Bacteriol.*, 179: 487-495.
- LEHRER, S. S. 1976. Perturbation of intrinsic protein fluorescence. In CHEN, R. F. & EDELHOCH, H. 1976. Biochemical fluorescence: concepts, V. 2. Marcel Dekker Inc.
- LINDQUIST, S. & CRAIG, E. A. 1988. The heat shock proteins. *Annu. Rev. Genet.*, 22: 631-677.
- MANALAVAN, P. & JOHNSON, W. C., Jr. 1983. Sensitivity of circular dichroism to protein tertiary structure class. *Nature*, 305: 831-832.
- MARTIRANI, L.; RANIELLO, R.; NACLERIO, G.; RICCA, E.; FELICE, M. 2001. Identification of the DNA-binding protein, HrcA, of *Streptococcus thermophilus*. *FEMS Microbiol. Lett.*, 198: 177-182.
- MINDER, A. C.; FISCHER, H.; HENNECKE, H. & NARBERHAUS, F. 2000. Role of HrcA and CIRCE in the heat shock regulatory network of *Bradyrhizobium japonicum*. *J. Bacteriol.*, 182: 14-22.
- MOGK, A.; HOMUTH, G.; SCHOLZ, C.; KIM, L.; SCHID, F. X. & SCHUMANN, W. 1997. The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *EMBO J.*, 16: 4579-4590.

- MOGK, A.; VÖLKER, A.; ENGELMANN, S.; HECKER, M.; SCHUMANN, W. & VÖLKER, U. 1998. Nonnative proteins induce expression of the *Bacillus subtilis* CIRCE regulon. *J. Bacteriol.*, 178: 2895-2900.
- NAKAHIGASHI, K.; RON, E. Z.; YANAGI, H. & YURA, T. 1999. Differential and independent roles of HrcA repressor in the heat shock response of *Agrobacterium tumefaciens*. *J. Bacteriol.*, 181: 7509-7515.
- NISHIMURA, A.; MORITA, M.; NISHIMURA, Y. & SUGINO, Y. 1990. A rapid and highly efficient method for preparation of competent *Escherichia coli* cells. *Nucleic Acids Res.*, 18: 6169.
- OHTA, T.; NETTIKADAN, S.; TOKUMASU, F.; IDENO, H.; ABE, Y.; KURODA, M.; HAYASHI, H. & TAKEYASU, K. 1996. Atomic Force Microscopy Proposes a Novel Model for Stem-Loop Structure That Binds a Heat Shock Protein in the *Staphylococcus aureus* HSP70 Operon. *Biochem. Biophys. Res. Commun.*, 226: 730-734.
- ØSTERÅS, M.; STOTZ, A.; SCHMID-NUOFFER, S. & JENAL, U. 1999. Identification and transcriptional control of the genes encoding the *Caulobacter crescentus* ClpXP protease. *J. Bacteriol.*, 181: 3039-3050.
- POINDEXTER, J. S. 1981. The caulobacters: ubiquitous unusual bacteria. *Microbiol. Rev.*, 45: 123-179.
- QUON, K. C.; MARCZINSKY, G. T. & SHAPIRO, L. 1996. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell*, 84: 83-93.
- REISCHL, S.; WIEGERT, T.; SCHUMANN, W. 2002. Isolation and analysis of mutant alleles of the *Bacillus subtilis* HrcA repressor with reduced dependency on GroE function. *J. Biol.Chem.*, 277: 32659 - 32667
- REISENAUER, A.; MOHR, C. D. & SHAPIRO, L. 1996. Regulation of a heat shock σ^{32} homolog in *Caulobacter crescentus*. *J. Bacteriol.*, 178: 1919-1927.
- REUTER, S. H. & SHAPIRO, L. 1987. Assymetric segregation of heat-shock proteins upon cell division in *Caulobacter crescentus*. *J. Mol. Biol.*, 194: 653-622.

- ROBERTS, R.; TOOCHINDA, C.; AVEDISSIAN, M.; BALDINI, R. L.; GOMES, S. L. & SHAPIRO, L. 1996. Identification of a *Caulobacter crescentus* operon encoding *hrcA*, involved in negatively regulating heat-inducible transcription, and the chaperone gene *grpE*. *J. Bacteriol.*, 178: 1829-1841.
- RÜNGELING, E.; LAUFEN, T.; BAHL, H. 1999. Functional characterisation of the chaperones DnaK, DnaJ, and GrpE from *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.*, 170: 119-123.
- SALANOUBAT, M.; GENIN, S.; ARTIGUENAVE, F. *et al.* 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, 415 (6871): 497-502.
- SAMBROOK, J.; FRITCH, E. F. & MANIATIS, T. 1989. *Molecular cloning. A Laboratory Manual*. Segunda edição, Cold Spring Harbour Laboratory press.
- SCHARF, C.; RIETHDORF, S.; ERNST, H.; ENGELMANN, S.; VOLKER, U. & HECKER, M. 1998. Thioredoxin is an essential protein induced by multiple stresses in *Bacillus subtilis*. *J. Bacteriol.* 180: 1869–1877.
- SCHULZ, A. & SCHUMANN, W. 1996. *hrcA*, the first gene of *Bacillus subtilis dnaK* operon encodes a negative regulator of class I heat shock genes. *J. Bacteriol.*, 178: 1088-1093.
- SEGAL, G. & RON, E. Z. 1996a. Regulation and organization of the *groE* and *dnaK* operons in Eubacteria. *FEMS Microbiol. Lett.*, 138: 1-10.
- SEGAL, G. & RON, E. Z. 1996b. Heat shock activation of the *groESL* operon of *Agrobacterium tumefaciens* and the regulatory roles of the inverted repeat. *J. Bacteriol.*, 178: 3634-3640.
- SEGAL, G. & RON, E. Z. 1995. The *groESL* operon of *Agrobacterium tumefaciens*: evidence for heat shock-dependent mRNA cleavage. *J. Bacteriol.*, 177: 750-757.
- SENO, M.; LIN, M. L.; IWAMOTO, K. 1990. Chromatographic behaviour of cyclodextrin complexes of nucleotides, nucleosides and their bases. *J. Chromatogr.*, 523: 293-299.
- SHAPIRO, L. 1994. The control of timing and spatial organization during *Caulobacter* cell differentiation. *The Harvey Lectures Series*, 88: 23-48.

- SHI, J.; BLUNDELL, T. L. & MIZUGUCHI, K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.*, 310: 243-257.
- SIMPSON, A. J. G.; REINACH, F. C.; ARRUDA, P. *et al.* 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, 406 (6792): 151-157.
- SPOHN, G.; DELANY, I.; RAPPUOLI, R.; SCARLATO, V. 2002. Characterization of the HspR-mediated stress response in *Helicobacter pylori*. *J. Bacteriol.*, 184: 2925-2930.
- SPOHN, G.; SCARLATO, V. 1999. The autoregulatory HspR repressor protein governs chaperone gene transcription in *Helicobacter pylori*. *Mol. Microbiol.*, 34: 663-674.
- SREERAMA, N.; WOODY, R. W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, 287: 252-260.
- STAHL, D. A.; KEY, R.; FLESHER, B. & SMIT, J. 1992. The phylogeny of marine and freshwater caulobacters reflects their habitat. *J. Bacteriol.*, 174: 2193-2198.
- THIES, F.L.; KARCH, H.; HARTUNG, H.P. & GIEGERICH, G. 1999. Cloning and expression of the *dnaK* gene of *Campylobacter jejuni* and antigenicity of heat shock protein 70. *Infect. Immun.*, 67: 1194-1200.
- TOMB, J. F.; WHITE, O.; KERLAVAGE, A. R. *et al.* 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388 (6642): 539-547.
- VALENTIN-HANSEN, P., HOJRUP, P. & SHORT, S. 1985. The primary structure of the DeoR repressor from *Escherichia coli* K-12. *Nucleic acids res.*, 13: 5927-5936.
- WATANABE, K., YAMAMOTO, T. & SUZUKI, Y. 2001. Renaturation of *Bacillus thermoglucosidasius* HrcA Repressor by DNA and Thermostability of the HrcA-DNA Complex In Vitro. *J. Bacteriol.*, 183: 155-161.
- WRIGHT, R.; STEPHENS, C.; ZWEIGER, G.; SHAPIRO, L. & ALLEY, M. R. 1996. *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev.*, 10: 1532-1532.

- WU, J. & NEWTON, A. 1996. Isolation, identification, and transcriptional specificity of the heat shock sigma factor σ^{32} from *Caulobacter crescentus*. *J. Bacteriol.*, 178: 2094-2101.
- WU, J.; OHTA, N. & NEWTON, A. 1998. An essential, multicomponent signal transduction pathway required for cell cycle regulation in *Caulobacter*. *Proc. Natl. Acad. Sci. USA*, 95: 1443-1448.
- YUAN, G. & WONG, S. 1995. Isolation and characterization of *Bacillus subtilis* *groE* regulatory mutants: evidence for *orf39* in the *dnaK* operon as a repressor gene in regulating the expression of *groE* and *dnaK*. *J. Bacteriol.*, 177: 6462-6468.
- ZENG, G.; YE, S. & LARSON, T. J. 1996. Repressor for the sn-Glycerol 3-Phosphate Regulon of *Escherichia coli* K-12: Primary Structure and Identification of the DNA-Binding Domain. *J. Bacteriol.*, 178: 7080-7089.
- ZHANG, R. G.; KIM, Y.; SKARINA, T.; BEASLEY, S.; LASKOWSKI, R.; ARROWSMITH, C.; EDWARDS, A.; JOACHIMIAK, A. & SAVCHENKO, A. 2002. Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. *J. Biol.Chem.* 277:19183-19190.
- ZUBER, U. & SCHUMANN, W. 1994. CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *Bacillus subtilis*. *J. Bacteriol.*, 176: 1359-1363.

ANEXOS

Anexo 1. À esquerda, visão no espaço de um dobramento OB ("Oligonucleotide/Oligosaccharide Binding Fold") idealizado (extraído e adaptado de AGRAWAL, V. & KISHAM, R. K. V. 2001. Functional evolution of two subtly different (similar) folds. *BMC Structural Biology*, 1:5 (publicação eletrônica)). À direita, diagrama topológico correspondente, onde α indica a hélice (cilindro), e β indica as cadeias em conformação β (setas), numeradas do terminal amino para o carboxi, indicados por N e C, respectivamente. Correspondência entre as cadeias β das duas figuras: B2 = $\beta 1$, B3 = $\beta 2$, B4 = $\beta 3$, B1 = $\beta 4$, e B5 = $\beta 5$

Anexo 2. Sobreposição visual dos espectros normalizados de dicroísmo circular no UV distante, de HrcA recombinante de *Caulobacter crescentus* (linha fina) e de ModE de *Escherichia coli* (linha espessa, ANDERSON, L. A., PALMER, T., PRICE, N. C., BORNEMANN, S., BOXER, D. H., & PAU, R. N. 1997. Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the *modABCD* (molybdenum transport) operon of *Escherichia coli*. *Eur. J. Biochem.* 246: 119-126.).