

ADRIANA LUCKOW INVITTI

**DETERMINAÇÃO DE SEQUÊNCIAS DE DNA
RECONHECIDAS POR PROTEÍNAS REGULADORAS DE
TRANSCRIÇÃO DEPENDENTES DO FATOR SIGMA 54 DA
RNA-POLIMERASE DE *Herbaspirillum seropedicae***

**CURITIBA
2006**

ADRIANA LUCKOW INVITTI

**DETERMINAÇÃO DE SEQUÊNCIAS DE DNA
RECONHECIDAS POR PROTEÍNAS REGULADORAS DE
TRANSCRIÇÃO DEPENDENTES DO FATOR SIGMA 54 DA
RNA-POLIMERASE DE *Herbaspirillum seropedicae***

Dissertação apresentada ao curso de Pós-Graduação em Bioquímica da Universidade Federal do Paraná, como requisito parcial para obtenção do título de Mestre em Ciências - Bioquímica.

Orientadores: Prof. Emanuel Maltempo de Souza e Prof. Fábio de Oliveira Pedrosa

**CURITIBA
2006**

Aos meus melhores amigos e
ao meu irmão Henrique

AGRADECIMENTOS

Agradeço aos meus orientadores prof. Emanuel M. de Souza e ao prof. Fábio de O. Pedrosa pela orientação, paciência, estímulo, compreensão e amizade. Agradeço pelas críticas, sugestões e principalmente pela confiança depositada em mim para a realização deste projeto.

Aos professores Giseli Klassen e Jaime Paba pelas críticas e sugestões.

A profa. Rose Adele Monteiro pelas sugestões e auxílio.

A todos os professores do Núcleo de Fixação de Nitrogênio.

A D. Marilza e a coordenação do curso de pós-graduação em Bioquímica.

Aos colegas do laboratório pelo auxílio e momentos de descontração: Lílian, Luiza, Luciano, Ju Reis, André D., Gi, Cláudio, Roberta, Renata, Fernanda P., Vanessa, Ana C, Marcelo A., Juliana I, Gus, Geraldo, Cynthia, Magda, Carol, Michelle D., André Luis, Rafa, Ju Osaki, Marco Aurélio, Rose Adele, Marco Antônio, Michelle T., Lys, Helisson, Giovani, Giovana, Marcelo M., Stefan, Tuca, Pércio, Dani, Patilene, Karen, Luis, Thiago, Eliel e Leo.

Ao Augusto pela grande ajuda na parte de Bioinformática.

Ao Valter pelas sequências, seu empenho em sempre querer ajudar e por proporcionar a todos do laboratório ótimos momentos de descontração.

A Roseli Prado e D. Jú pelo carinho, apoio e atenção em todos os momentos.

Aos meus companheiros de laboratório e de “balada” pela troca de conhecimentos e diversão; Ana Claudia, Marcelão e Gus.

Em especial, à Fabi, que desde a iniciação científica foi minha “mãe científica” e acabou se tornando uma grande amiga, com a qual se pode sempre contar.

Ao Bruno por sua amizade e companhia nas madrugadas escrevendo.

Aos meus melhores amigos, Marco, Laura, Fer, Thay e Nick por terem me animado quando eu precisava, pelos momentos de descontração, pelas voltinhas para espairecer e principalmente por sempre estarem do meu lado me dizendo o que eu preciso ouvir.

Ao meu irmão Henrique pelo amor e apoio incondicional nos momentos que eu mais precisei, que me ajudaram a conseguir levar as coisas em frente.

Aos meus pais José Henrique e Elisabeth e à minha irmã Alessandra pelo carinho e apoio.

Ao CNPq e PRONEX (FINEP/MCT/CNPq) pelo apoio financeiro.

*Faça as coisas o mais simples que puder,
porém não as mais simples.*

Albert Einstein

SUMÁRIO

LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
LISTA DE ABREVIATURAS.....	xii
RESUMO.....	xiv
1 INTRODUÇÃO	1
1.1 <i>Herbaspirillum seropedicae</i>	1
1.2 Regulação da Expressão Gênica em Procariotos.....	1
1.2.1 Transdução de Sinal	2
1.3 Transcrição em Procariotos	6
1.4 Proteínas Reguladoras de Transcrição.....	9
1.4.1 Proteínas Repressoras de Transcrição.....	11
1.4.2 Proteínas Ativadoras de Transcrição dependentes de σ^{70}	11
1.4.3 Proteínas reguladoras de transcrição dependentes de σ^{54}	12
2 OBJETIVOS	18
2.1 Objetivos Específicos	18
3 MATERIAIS E MÉTODOS	19
3.1 Bactérias e Plasmídios.....	19
3.2 Meios de Cultura e Condições de Cultivo	20
3.3 Antibióticos.....	22
3.4 Preparo de Solução de RNase	23
3.5 Purificação de DNA.....	23
3.5.1 DNA Genômico de <i>H. seropedicae</i>	23
3.5.2 DNA Plasmidial	24
3.6 Separação Eletroforética.....	24
3.6.1 DNA	24
3.6.2 Proteínas em Condições Desnaturantes (SDS-PAGE).....	24
3.7 Ferramentas de Bioinformática	24
3.8 Planejamento dos Oligonucleotídeos Iniciadores	25
3.9 Amplificação de Fragmentos de DNA	26
3.10 Clivagem do DNA com Enzimas de Restrição	26
3.11 Reação de Ligação de DNA.....	27
3.12 Preparo de Células Eletrocompetentes.....	27
3.13 Transformação Bacteriana por Eletroporação	27

3.14	Expressão e Purificação de Proteínas	28
3.15	Dosagem de Proteínas	29
3.16	Reação de Interação DNA-Proteína e Isolamento de Aptâmeros	29
3.17	Análise de Plasmídios Recombinantes por PCR	30
3.18	Reações de Sequenciamento de DNA.....	30
4	RESULTADOS E DISCUSSÃO	31
4.1	Identificação dos Genes que Codificam para Proteínas Reguladoras de Transcrição no Genoma de <i>H. seropedicae</i>	31
4.1.1	Classificação e Identificação dos Domínios Estruturais das Proteínas Reguladoras de Transcrição de <i>H. seropedicae</i>	33
4.2	Análise das proteínas reguladoras de transcrição dependentes de sigma 54 localizadas no genoma de <i>H. seropedicae</i>	36
4.3	Amplificação dos Fragmentos de DNA que Codificam para o Domínio HTH das Proteínas 208.0173, 216.0019 e 287.0556.....	39
4.4	Clonagem dos Fragmentos de DNA que Codificam para o Domínio HTH das Proteínas 208.0173, 216.0019 e 287.0556 em Vetor de Expressão.	48
4.5	Expressão e Purificação dos Domínios 208.0173-His, 216.0019-His e 287.0556-His.	50
4.6	Isolamento dos Aptâmeros de DNA.....	52
4.6.1	Aptâmeros Correspondentes ao Domínio HTH da Proteína 216.0019	53
4.6.2	Aptâmeros Correspondentes ao Domínio HTH da Proteína 208.0173	58
4.6.3	Aptâmeros Correspondentes ao Domínio HTH da Proteína 287.0556	61
4.6.4	Comparação entre os Diversos Métodos de Isolamento de Aptâmeros de DNA	63
5	CONCLUSÕES	67
6	REFERÊNCIAS BIBLIOGRÁFICAS	68
	Apêndice I – Tabela da Análise das Proteínas Reguladoras de Transcrição de <i>H. seropedicae</i>	I
	Apêndice II – Grupos Funcionais Definidos pelo COG.....	X

Apêndice III – Alinhamento da Sequência de Aminoácidos das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de <i>H. seropedicae</i> e NtrC de <i>A. brasilense</i>	XI
Apêndice IV – Alinhamento da Sequência de Aminoácidos dos Domínios AAA das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de <i>H. seropedicae</i>	XIV
Apêndice V – Alinhamento da Sequência de Aminoácidos dos Domínios HTH das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de <i>H. seropedicae</i>	XV
Apêndice VI – Sequência do aptâmero 1 do domínio HTH da proteína 216.0019	XVI
Apêndice VII – Sequência dos aptâmeros 2 e 3 do domínio HTH da proteína 216.0019	XVII
Apêndice VIII – Sequência do aptâmero 4 do domínio HTH da proteína 216.0019	XVIII
Apêndice IX – Sequência dos aptâmeros 5 e 6 do domínio HTH da proteína 216.0019	XIX
Apêndice X – Sequência do aptâmero 1 do domínio HTH da proteína 208.0153	XX
Apêndice XI – Sequência do aptâmero 2 do domínio HTH da proteína 208.0153	XXI
Apêndice XII – Sequência do aptâmero 3 do domínio HTH da proteína 208.0153	XXII
Apêndice XIII – Sequência do aptâmero 4 do domínio HTH da proteína 208.0153	XXIII
Apêndice XIV – Sequência do aptâmero 5 do domínio HTH da proteína 208.0153	XXIV
Apêndice XV – Sequência do aptâmero 6 do domínio HTH da proteína 208.0153	XXV
Apêndice XVI – Sequência do aptâmero 7 do domínio HTH da proteína 208.0153	XXVI
Apêndice XVII – Sequência do aptâmero 8 do domínio HTH da proteína 208.0153	XXVII

Apêndice XVIII – Sequência do aptâmero 9 do domínio HTH da proteína	
208.0153	XXVIII
Apêndice XIX – Sequência do aptâmero 1 do domínio HTH da proteína	
287.0556	XXIX

LISTA DE FIGURAS

Figura 1 - Esquema do Funcionamento do Sistema de 2 Componentes em Procariotos	5
Figura 2 - Esquema da Transcrição em Procariotos	7
Figura 3 – Esquema de Ativação de Transcrição por Proteínas Reguladoras de Transcrição Dependentes de σ^{54}	14
Figura 4 – Domínios das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de <i>H. seropedicae</i>	37
Figura 5 - Dendrograma das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de <i>H. seropedicae</i> .	38
Figura 6 – Seqüência de Nucleotídeos e Aminoácidos das ORFs 208.0173, 216.0019 e 287.0556.	41
Figura 7 - Eletroforese dos Produtos de Amplificação de DNA Genômico de <i>H. seropedicae</i> com os Oligonucleotídeos Iniciadores 208N e 208C, 216N e 216C e 287N e 287C.	48
Figura 8 - Estratégia para Produção de Plasmídios para Expressão dos Domínios HTH com Cauda de Histidina N-terminal.	49
Figura 9 – Eletroforese Desnaturante das Frações Solúvel e Insolúvel de Extratos de <i>E. coli</i> BL21 λ (DE3) Expressando os Domínios HTH-His das Proteínas 208.0173, 216.0019 e 287.0556 .	51

Figura 10 – Eletroforese dos Domínios HTH-His das Proteínas 208.0173, 216.0019 e 287.0556 Purificados	52
Figura 11 - Eletroforese dos Produtos de PCR dos Plasmídios Recombinantes Contendo os Aptâmeros do Domínio HTH da Proteína 216.0019	54
Figura 12 – - Aptâmeros do Domínio HTH da Proteína 216.0019 Isolados e sua Localização no Genoma de <i>H. seropedicae</i>	55
Figura 13 – Provável Região Promotora do Gene <i>osmY</i>	57
Figura 14 – Provável Região Promotora do Gene <i>prhA</i>	57
Figura 15 - Aptâmeros do Domínio HTH da Proteína 208.0173 Isolados e sua Localização no Genoma de <i>H. seropedicae</i>	59
Figura 16 - Aptâmeros do Domínio HTH da Proteína 208.0173 Isolados e sua Localização no Genoma de <i>H. seropedicae</i>	62
Figura 17 – Esquema Comparativo dos Métodos SELEX e de Cromatografia de Afinidade para Isolamento de Aptâmeros de DNA	63
Figura 18 – Esquema do Isolamento de Aptâmeros por Imunoprecipitação de Cromatina – ChIP	65

LISTA DE TABELAS

Tabela 1 – Estirpes e Plasmídios	19
Tabela 2 - Distribuição das proteínas reguladoras de transcrição de <i>H. seropedicae</i> dentro das famílias de reguladores de transcrição	34
Tabela 3 – Fragmentos de DNA Obtidos por PCR e dos Peptídeos Expressos	40
Tabela 4 – Aptâmeros do Domínio HTH da Proteína 216.0019 que Foram Isolados	54
Tabela 5 – Apatâmeros do Domínio HTH da Proteína 208.0173 que Foram Isolados	59

LISTA DE ABREVIATURAS

A - Alanina

aa - aminoácido

ATP – 5´- trifosfato de adenosina

Amp^R – resistente à ampicilina

bis-acrilamida – N, N´-metileno-bis-acrilamida

°C – graus Celsius

COG – *Cluster of Orthologous Group*

Da – Dalton

dNTP - 5´didesoxinucleotídeo

DNA – ácido desoxirribonucléico

DTT - ditioneitol

EBP – *enhancer binding protein*

EDTA – ácido etilenodiamino-tetra-acético

EMSA – *Electrophoretic mobility shift assay*

F - Fenilalanina

G - Glicina

g – unidade de força centrífuga

µg – micrograma

His-tag – cauda de 6 resíduos de histidina

HK – histidina quinase

HTH – *hélix-turn-hélix* (hélice-volta-hélice)

IPTG - β-D-isopropil-tiogalactopiranosídeo

kb – kilobase, 1000 pares de base

Km^R – resistente à canamicina

L – litro

µL – microlitro

µmol – micromol

mg – miligrama

min – minuto

mL – mililitro

mmol – milimol

mRNA – RNA mensageiro

MW – *Molecular weight* (massa molecular)
ng – nanograma
NTA – ácido nitrilo-tri-acético
ORF – *open reading frame*
pb – pares de base
PEG – polietilenoglicol
Pfam – *Protein families database*
PMSF - Fenil-metil-sulfonil-fluoreto
pI – ponto isoelétrico
RNA – ácido ribonucleico
rpm – rotações por minuto
RR – regulador de resposta
s – segundo
SDS – dodecilsulfato de sódio
SDS-PAGE – eletroforese em gel de poliacrilamida contendo SDS
SELEX – *Systematic Evolution of Ligands by Exponential Enrichment*
Sm^R – resistente à estreptomicina
SMART – *Simple Modular Research Tool*
T - Treonina
Tris – tris(hidroximetil)-aminometano
U – unidade (enzimática)
UAS – *upstream activation sequence*
UV – luz ultra-violeta

RESUMO

A regulação da expressão gênica ocorre principalmente em nível transcricional. Proteínas reguladoras de transcrição, que podem funcionar como ativadores ou repressores, são os moduladores da atividade da RNA-polimerase e determinam os níveis de mRNA em respostas as condições fisiológicas prevalentes. Os ativadores de transcrição dependentes do fator sigma 54 da RNA-polimerase, também conhecidos por EBPs (*enhancer binding proteins*), constituem uma classe distinta de proteínas reguladoras de transcrição e exercem sua atividade de maneira semelhante àquela das proteínas reguladoras de transcrição de eucariotos. As EBPs ativam a transcrição a partir de promotores específicos e mediante a hidrólise de ATP catalisam a isomerização do complexo fechado a complexo aberto. Em *Herbaspirillum seropedicae* duas proteínas dessa família já foram caracterizadas, NifA e NtrC. Uma busca no banco de dados do Programa Genoma do Paraná (GENOPAR) identificou a presença de 11 prováveis proteínas reguladoras de transcrição dependentes de sigma 54 no genoma de *H. seropedicae*. Os domínios estruturais dessas proteínas foram identificados e três delas, denominadas 208.0173, 216.0019 e 287.0556, foram selecionadas para identificação dos seus sítios de ligação ao DNA. Fragmentos de DNA codificando o domínio de ligação ao DNA dessas três proteínas foram amplificados por PCR, clonados no vetor pGEM-T e subclonados no vetor de expressão pET28a.

Um novo método para isolamento de aptâmeros foi desenvolvido neste trabalho. Os domínios de ligação ao DNA das três proteínas foram então super-expressos, purificados, submetidos a reações de interação DNA-proteína e os aptâmeros foram isolados por cromatografia de afinidade em uma mini-coluna de agarose-NTA-Ni⁺². Os aptâmeros de DNA obtidos foram sequenciados e apresentaram homologia com regiões do genoma de *H. seropedicae* a montante de diversos genes. A proteína 208.0173 provavelmente regula a expressão dos genes *thdF*, *fhaC* e *xdhC*; 216.0019 dos genes *prhA* e *osmY* e 287.0556 do gene *fixG*.

1 INTRODUÇÃO

1.1 *Herbaspirillum seropedicae*

Herbaspirillum seropedicae são bactérias diazotróficas endofíticas, pertencentes a subclasse β das Proteobacterias (Young, 1992), isoladas dos órgãos vegetativos de gramíneas como cana de açúcar e arroz (Baldani et al., 1986). São bactérias gram-negativas, móveis, vibrióides, algumas vezes helicoidais e capazes de fixar nitrogênio em condições de microaerofilia (Baldani et al., 1986).

A capacidade de *Herbaspirillum* formar uma associação com gramíneas de interesse econômico como milho, sorgo e cana-de-açúcar (Döbereiner, 1992), sem provocar doença, tem despertado interesse para o seu estudo. Vários trabalhos demonstraram que a interação *Herbaspirillum*-planta pode trazer benefícios à planta. Baldani et al. (1995) observaram que a estirpe Z94 de *H. seropedicae* contribui com até 54% do nitrogênio total acumulado por plantas de arroz inoculadas. O mecanismo de colonização dos tecidos vegetais por essas bactérias ainda não está completamente esclarecido. Análises microscópicas de plantas colonizadas sugerem que ocorra a ligação à superfície da planta e proliferação das bactérias, seguida de penetração na planta e ocupação de espaços intercelulares e feixes vasculares com posterior colonização e estabelecimento nas partes aéreas e vasos do xilema (Olivares et al., 1995, Roncato-Maccari et al., 2003).

O potencial como biofertilizante de *H. seropedicae* torna importante o estudo dos mecanismos celulares deste microrganismo. O estudo dos reguladores de transcrição dependentes do fator σ^{54} desse organismo deve contribuir para o esclarecimento da regulação do metabolismo e da fixação biológica de nitrogênio.

1.2 Regulação da Expressão Gênica em Procaríotos

As bactérias vivem em ambientes onde os níveis de nutriente, toxinas, acidez, temperatura, osmolaridade, umidade e outras condições podem mudar rápida e inesperadamente. A sobrevivência das células depende do

monitoramento constante das condições intra e extracelulares e ao ajuste da estrutura, fisiologia e comportamento celular frente às alterações ambientais. Uma rede elaborada de sensores e reguladores de resposta suporta esse comportamento adaptativo (Parkinson & Kofoid, 1992).

As respostas adaptativas em bactérias envolvem em primeiro lugar mudanças rápidas na atividade das enzimas chaves das principais vias metabólicas (White, 2000). Estas mudanças têm como objetivo reorganizar o fluxo metabólico de acordo com as novas condições ambientais prevalentes. Em seguida ocorrem alterações mais duradouras em decorrência da reorganização do padrão de expressão gênica de acordo com os sinais que estão sendo recebidos (Stock, Ninfa & Stock, 1989) resultando na síntese *de novo* de proteínas necessárias para otimizar a exploração dos recursos nas novas condições ambientais.

A transcrição é o passo principal de controle dessa reorganização da expressão gênica e os fatores de transcrição são os instrumentos que interpretam os sinais ambientais e modulam precisamente a resposta àqueles sinais. De forma geral, estas proteínas exercem seu efeito alterando a eficiência de recrutamento da RNA-polimerase aos elementos promotores, alteração da conformação de promotores ou alteração da taxa de isomerização do complexo fechado a complexo aberto (McClure, 1985).

1.2.1 Transdução de Sinal

Transdução de sinal em bactérias é processo de sensoriamento de mudanças intra e extra-celulares, transferência e amplificação dos estímulos para que a célula elabore uma resposta. As bactérias possuem diversas redes de transdução de sinal que regulam a transcrição. Estas redes são constituídas por proteínas regulatórias que, em geral, contém domínios modulares capazes de interagir entre si para processar a informação e produzir a resposta apropriada aos estímulos ambientais (Stock, Robinson & Goudreau, 2000). Estes sistemas levam a uma mudança adaptativa da expressão gênica, que pode ser localizada (por exemplo, indução de um gene de resistência a um antibiótico específico) ou global (por exemplo, a indução de um conjunto de genes reparo de DNA e recombinação quando a célula é exposta à luz UV)

(White, 2000). Qualquer que seja a resposta são necessários processos que reconhecem os sinais ambientais e os transmitem para os sistemas responsáveis pela resposta. O mecanismo de sinalização mais simples e melhor caracterizado é conhecido como sistema de dois componentes (Hakenbeck & Stock, 1996; Stock, Robinson & Goudreau, 2000).

1.2.1.1 Sistemas de Dois Componentes

O sistema mais simples de sinalização em bactérias possui dois componentes: um sensor que monitora um parâmetro ambiental; e um regulador de resposta citoplasmático que altera a expressão gênica ou movimento celular em resposta às modificações ambientais (Stock, Ninfa & Stock, 1989). A maioria das proteínas sensoras contém domínios transmissores e estão ancoradas na membrana citoplasmática, com o domínio transmissor voltado para o interior da célula; já as proteínas reguladoras de resposta contém domínios receptores e são invariavelmente citoplasmáticas (Parkinson, 1993). Entretanto, é importante notar que a maioria dos sistemas de sinalização em procariotos não se comporta de modo tão simples quanto o paradigma de sistemas de dois componentes, pois podem possuir múltiplos elementos para amplificação de sinal, inter-conexões com outros circuitos regulatórios, alças de retro-alimentação e outras características complexas de sinalização (Parkinson & Kofoid, 1992).

O sistema de dois componentes normalmente consiste de membros de duas superfamílias de proteínas: uma proteína quinase (histidina quinase, HK), que se autofosforila num resíduo de histidina específico; e o regulador de resposta (RR) que recebe o grupo fosforil da proteína quinase ou de pequenas moléculas doadoras de fosfato em um resíduo de aspartato específico (Kofoid & Parkinson, 1988, Hakenbeck & Stock, 1996). A fosforilação do regulador de resposta modifica a sua atividade catalítica e capacidade de interação com outras proteínas e com DNA (Lukat et al., 1992).

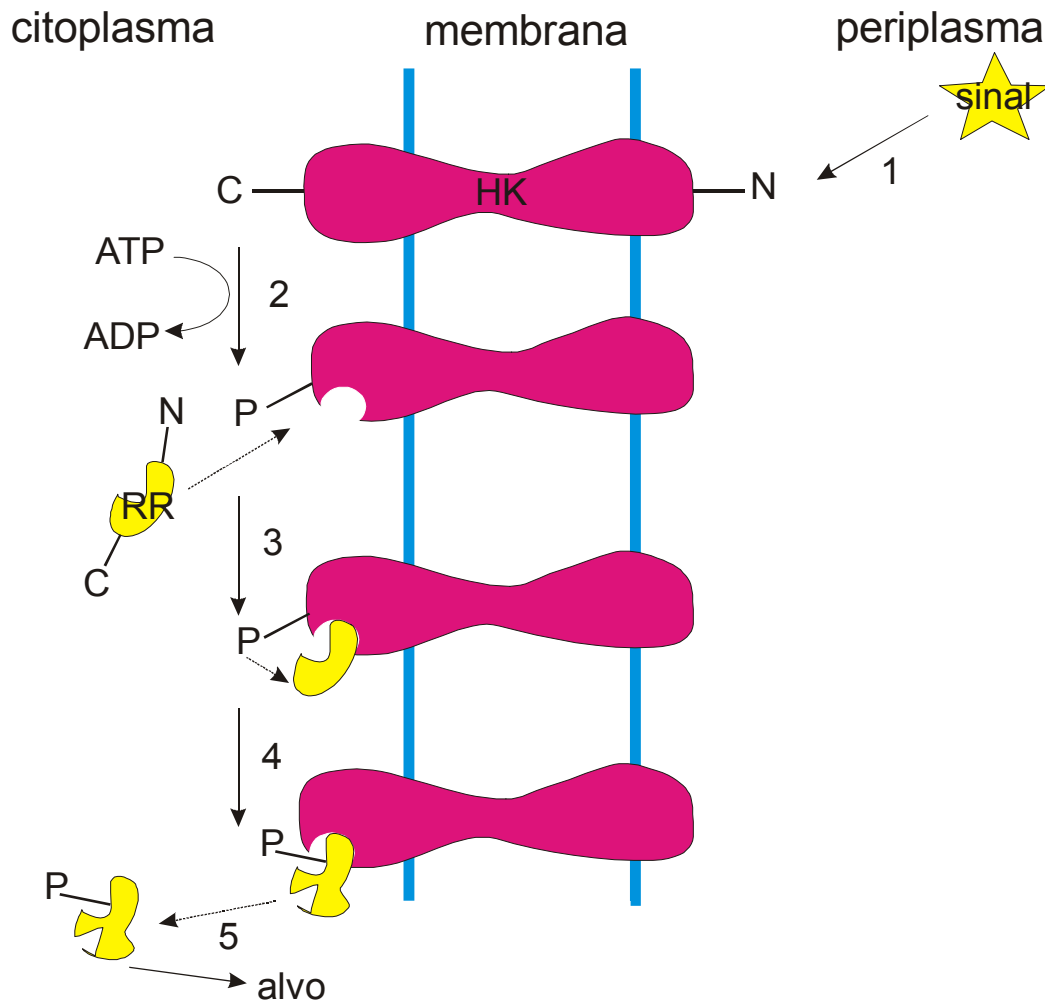
Os sistemas de dois componentes podem ser divididos em três classes. A primeira corresponde aos reguladores de resposta dependentes do fator σ^{70} da RNA-polimerase, onde a ativação da transcrição ocorre geralmente pelo recrutamento da RNA-polimerase ao promotor por interação do regulador de

resposta principalmente com a subunidade α da RNA-polimerase (Hakenbeck & Stock, 1996). A segunda classe compreende aos reguladores de resposta dependentes do fator σ^{54} da RNA-polimerase que possuem sequências de ligação ao DNA específicas e hidrolisam ATP para ativação da transcrição. O regulador de resposta (RR) desse sistema possui um domínio central (domínio AAA) responsável pela ligação, hidrólise de ATP e interação com a RNA-polimerase (Hakenbeck & Stock, 1996; Popham et al., 1989; Buck et al., 2000). As proteínas NtrB (HK) e NtrC (RR) constituem o sistema de dois componentes melhor caracterizado desta classe. (Hakenbeck & Stock, 1996). A terceira classe inclui os reguladores de resposta conhecidos como circuitos integrados, que são proteínas que possuem tanto o domínio transmissor quanto o receptor; são exemplos as proteínas ArcB (*E. coli*), que controla a expressão gênica em condições de baixas tensões de oxigênio; RscC (*E. coli*), que participa na regulação da síntese de polissacarídeos capsulares; BvgS (*Bordetella pertussi*), que controla o início da infecção virulenta em humanos (Parkinson & Kofoid, 1992).

Os domínios transmissores regulam de duas maneiras o estado de fosforilação dos seus pares receptores. Primeiro, os transmissores retiram um grupamento fosforil do ATP e o ligam a um resíduo de histidina conservado através de sua atividade de auto-quinase (Hess et al., 1988; Ninfa & Bennet, 1991). O produto, proteína-fosfohistidina, serve como um intermediário de alta energia para a transferência do grupamento fosforil para um resíduo de aspartato conservado no regulador de resposta (Sanders et al., 1992, 1989) (figura 1). Em alguns casos a histidina quinase pode ser capaz de fosforilar mais de um regulador de resposta e a competição destes reguladores por grupamentos fosforil definirá a resposta final (Stock, Robinson & Goudreau, 2000). Em segundo lugar, algumas proteínas transmissoras são bifuncionais e possuem também atividade de fosfatase, sendo capazes de remover o grupamento fosforil de seus receptores correlatos (Aiba et al., 1989; Igo et al., 1989; Keener & Kustu, 1988; Ninfa & Magasanik, 1986). O tipo de estímulo controla a inter-conversão entre as duas atividades (quinase ou fosfatase) dos transmissores, regulando assim o fluxo de grupamentos fosforil. Os reguladores de resposta podem ainda ser fosforilados e ativados na ausência

de proteína quinase por fosfo-metabólitos de baixa massa molecular (por exemplo, acetil fosfato), indicando que estes compostos podem ter um papel importante na regulação das vias de transdução de sinal (Lukat et al., 1992).

Figura 1 – Esquema do Funcionamento do Sistema de 2 Componentes em Procariotos



Legenda. Esquema de funcionamento do sistema de dois componentes. A proteína HK recebe um sinal e se auto-fosforila. O grupamento fosforil é então passado para a proteína reguladora de resposta, que sofre uma mudança conformacional que permite com que ela atue sobre seu alvo, que pode ser outra proteína ou DNA. Modificado de White, 2000.

As vias de transdução de sinal melhor caracterizadas em bactérias são a de controle da resposta à variação osmótica (EnvZ, sensor transmembrana, e OmpR, regulador de resposta citoplasmático), o sistema de regulação do metabolismo de nitrogênio (NtrB, sensor citoplasmático, e NtrC, regulador de

resposta citoplasmático) e o sistema de regulação de quimiotaxia (CheA (HK central), CheB (metilesterase), CheR (metiltransferase), CheW (HK), CheZ (defosforilação) e CheY (RR)) todas em *E. coli* (Stock, Ninfa e Stock, 1989; Parkinson, 1993; Hakenbeck & Stock, 1996).

1.3 Transcrição em Procariotos

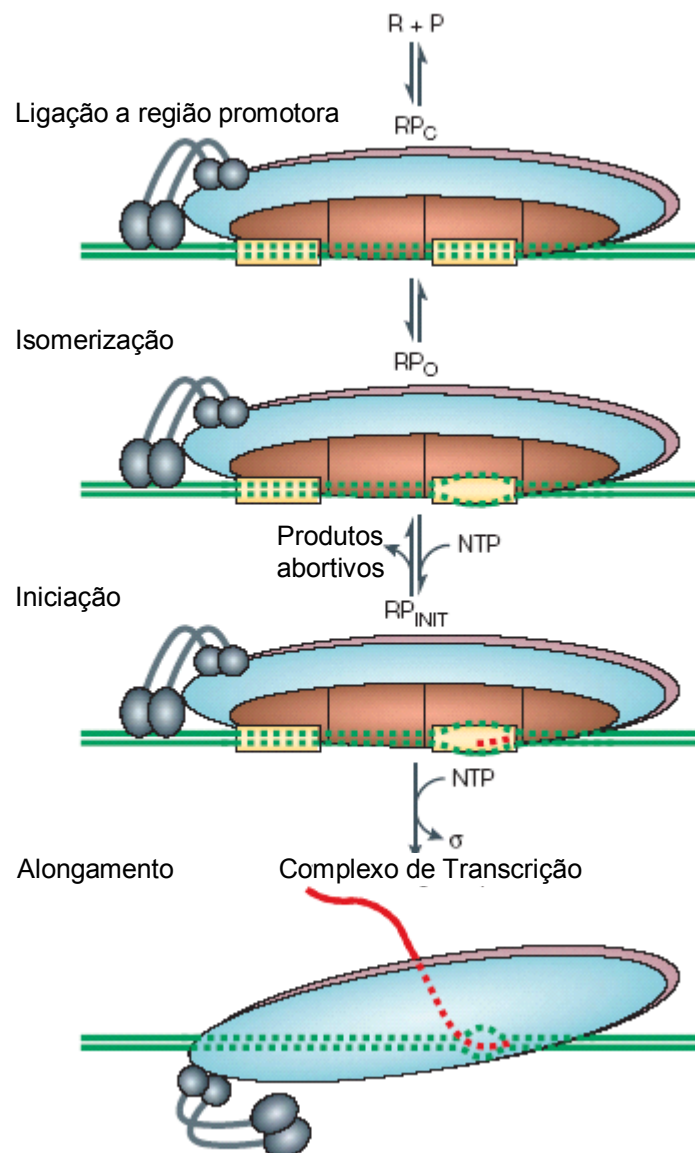
O componente central da transcrição em procariotos é a RNA-polimerase DNA-dependente. Essa enzima é composta por pelo menos cinco subunidades e pode ser isolada em duas formas: a) a holoenzima que possui as subunidades $\alpha_2\beta\beta'\sigma$, capaz de reconhecer promotores e iniciar a transcrição; e b) apoenzima ou núcleo da enzima, que possui todas as subunidades exceto a subunidade (ou fator) σ . Esta última forma não é capaz de iniciar a síntese de RNA a partir de moldes nativos (Kumar, 1981). A subunidade σ , ou fator sigma, é responsável pelo reconhecimento e ligação a seqüências específicas do DNA, denominadas promotores (Wösten, 1998).

O núcleo da RNA-polimerase precisa interagir com o fator sigma, formando a holoenzima, para que possa reconhecer a seqüência promotora e iniciar a transcrição. Somente a holoenzima é capaz de reconhecer, ligar a seqüência promotora e provocar as mudanças conformacionais no DNA necessárias para separar as duas fitas e iniciar a transcrição (Ishihama, 1990).

Os fatores σ de bactérias podem ser divididos em duas famílias: família σ^{70} e família σ^{54} . Os fatores pertencentes à família σ^{70} quando associados ao núcleo da RNA-polimerase, produzem uma forma de holoenzima capaz de iniciar a transcrição independente da presença de proteínas ativadoras (McClure, 1985). Aqueles pertencentes à família σ^{54} produzem uma holoenzima incapaz de formar o complexo aberto na ausência de proteínas ativadoras (Sasse-Dwight & Gralla, 1988; Morett & Buck, 1989 e Popham et al., 1989). Fatores sigma destas duas famílias possuem pouca ou nenhuma identidade de seqüência entre si. *E. coli* possui seis fatores sigma diferentes pertencentes à família σ^{70} : σ^{70} , σ^{38} , σ^{32} , σ^{28} , σ^E e σ^{FecI} e apenas um fator tipo σ^{54} (Wösten, 1998).

O processo de transcrição pode ser dividido em quatro etapas: reconhecimento da região promotora no DNA, iniciação, alongamento e terminação (Kumar, 1981; Von Hippel, 1984) (figura 2).

Figura 2 - Esquema da Transcrição em Procariotos



Legenda. Três primeiras fases da transcrição em procariotos. A RNA-polimerase holoenzima (R) interage com a região promotora no DNA (P), formando o complexo fechado (RP_C). As linhas tracejadas verdes indicam as regiões promotoras do DNA onde se liga a holoenzima e os dois quadrados amarelos as sequências conservadas localizadas nas posições -35 e -10 em relação ao início de transcrição. A dupla fita de DNA está desenrolada na região de

transcrição (bolha de transcrição) formando o complexo aberto (RP_0). O complexo de iniciação é formado (RP_{INT}) e a síntese de RNA (representado por linhas tracejadas vermelhas) é iniciada com formação da ligação fosfodiéster entre o nucleotídeo inicial e próximo NTP. O alongamento é a última fase mostrada, onde o tamanho da cadeia de RNA aumenta (linha vermelha contínua). Modificado de Browning & Busby, 2004.

O reconhecimento do promotor envolve a ligação da holoenzima associada ao fator σ^{70} ao DNA dupla fita dando origem ao complexo binário fechado. Em seguida, as fitas de DNA são separadas para permitir o pareamento de bases da fita molde com os ribonucleotídeos. A bolha de transcrição é formada quando ocorre a desnaturação da fita de DNA, que se inicia na região onde a RNA-polimerase está ligada, formando o complexo binário aberto (Von Hippel, 1984 e Lewin, 2000). A holoenzima associada a fatores sigma da família σ^{70} geralmente é capaz de iniciar a transcrição na ausência de proteínas ativadoras, embora existam exceções onde o complexo fechado RNA-polimerase-promotor não pode ser isomerizado a complexo aberto na ausência de ativadores de transcrição (Ptashne & Gann, 1997).

A iniciação corresponde à incorporação dos primeiros nucleotídeos do RNA utilizando uma das fitas do DNA como molde. A enzima permanece ligada ao promotor durante a incorporação dos primeiros nucleotídeos. A formação da primeira ligação fosfodiéster entre os dois primeiros ribonucleotídeos origina o complexo ternário (Lewin, 2000). A iniciação é prolongada pela ocorrência de inícios abortivos, onde a enzima produz transcritos curtos (até 8-9 nucleotídeos) que são liberados e então ela inicia novamente a síntese do RNA, desde a primeira base (Hsu et al., 2003).

A fase seguinte corresponde ao alongamento. Durante esta fase o núcleo da RNA-polimerase move-se pela fita de DNA e estende a cadeia de RNA nascente. A fita de DNA a jusante é desenrolada enquanto a enzima se move para expor um novo segmento do molde. Os nucleotídeos são covalentemente ligados à extremidade 3' do RNA nascente e formam um híbrido RNA-DNA na região desenrolada. A fita de DNA retoma a forma de dupla hélice na região a montante à enzima. O RNA fora da bolha de transcrição fica livre na forma de simples fita (Lewin, 2000). A estabilização do complexo ternário é dada principalmente pelas interações proteína-DNA e

proteína-RNA (Nudler et al., 1996).

A terminação é a última etapa e envolve o reconhecimento do ponto onde deve haver a parada da adição de bases à cadeia de RNA nascente. O fim da transcrição ocorre quando a formação de novas ligações fosfodiéster cessa e o complexo de transcrição é separado (Lewin, 2000). Uma seqüência denominada terminadora é necessária para que a bolha de transcrição se desmonte. Esta seqüência desestabiliza o híbrido de RNA-DNA permitindo que o DNA retorne a forma de dupla fita e a cadeia de RNA e a RNA-polimerase sejam liberadas (Platt & Richardson, 1992). Existem dois mecanismos de terminação conhecidos em *E. coli*: a formação de grampos no RNA nascente provocando a interrupção transitória da síntese processiva de RNA pela RNA-polimerase e liberação do RNA transcrito, e pela ação de um fator protéico conhecido como Rho (Richardson, 1990 e Platt & Richardson, 1992).

A RNA-polimerase pode ainda interagir com diversos fatores durante o processo de transcrição. Estas interações podem ocorrer em qualquer uma das fases de transcrição podendo provocar aumento ou diminuição da produção de transcrito. Proteínas acessórias como NusA, NusG, GreA e GreB podem atuar em diferentes momentos no processo de alongamento (Erie, 2002). Entretanto, a transcrição da maioria dos genes de bactérias é regulada na fase de iniciação. Essa regulação é resultado da ligação de proteínas, ou fatores de transcrição, aos promotores ou próximo a eles, ativando ou reprimindo a iniciação da transcrição de acordo com estímulos intra e extracelulares (Rhodius & Busby, 1998).

1.4 Proteínas Reguladoras de Transcrição

As proteínas reguladoras de transcrição podem provocar a ativação ou repressão da expressão gênica. Alguns fatores de transcrição possuem as duas atividades, funcionando como ativadores ou repressores dependendo do promotor em que se ligam ou do sinal recebido (Pérez-Rueda & Collado-Vides, 2000).

Baseados na presença de motivos estruturais de proteínas que ligam DNA, por análises *in silico*, Pérez-Rueda & Collado-Vides (2000) previram que

o genoma de *E. coli* K12 possui 314 genes que codificam para prováveis fatores de transcrição. Destes, 34,8% provavelmente funcionam como ativadores de transcrição, 42,8% como repressores e 22,3% que podem funcionar como ativadores ou repressores. Esse número corresponde a pouco menos de 10% do total de genes codificados pelo genoma de *E. coli* (Pérez-Rueda & Collado-Vides, 2000).

A maioria dos fatores de transcrição são proteínas que ligam DNA em seqüências específicas fazendo com que sua influência se dê em promotores específicos (Browning & Busby, 2004). Alguns desses fatores controlam a transcrição de um grande número de genes enquanto outros controlam apenas um ou dois operons. Estima-se que dentre os reguladores de transcrição de *E. coli* apenas sete (CRP, FNR, IHF, Fis, NarL e Lrp) controlem 50% dos genes com expressão regulada deste organismo, enquanto que aproximadamente sessenta proteínas regulatórias (~20%) controlam apenas um promotor (Martinez-Antonio & Collado-Vides, 2003). Existem várias famílias descritas de reguladores de transcrição, mas as famílias que melhor foram caracterizadas são NtrC, OmpR, LacI, CRP, AraC e LysR (Pérez-Rueda & Collado-Vides, 2000).

A atividade dos fatores de transcrição pode ser controlada de pelo menos quatro maneiras. A primeira delas é através da modificação da afinidade de ligação ao DNA modulada por pequenos ligantes; a concentração desses ligantes pode variar de acordo com a disponibilidade de nutrientes ou estresses (Browning & Busby, 2004). O segundo tipo de regulação de atividade é realizado por modificação covalente, principalmente fosforilação, da proteína regulatória (Stock, Robinson & Goudreau, 2000). Dessa forma ocorre o controle da atividade do regulador de resposta no sistema de dois componentes, como descrito anteriormente. A terceira forma corresponde à variação na concentração citoplasmática da proteína regulatória, que pode depender da regulação da sua expressão, degradação proteolítica, localização celular ou uma combinação destes (Browning & Busby, 2004). A última forma de regulação, e menos comum, é o seqüestro do fator de transcrição por outras proteínas regulatórias (Plumbridge, 2002).

1.4.1 Proteínas Repressoras de Transcrição

A transcrição pode ser inibida por diversos mecanismos; os repressores de transcrição não possuem somente a capacidade de ligar DNA, mas também de modificar a estrutura do DNA e interagir com a RNA-polimerase. Os mecanismos de repressão da expressão gênica são: inibição da ligação da RNA-polimerase ao promotor, inibição da desnaturação do promotor (atuando na fase de iniciação, impedindo a formação do complexo aberto) e desligamento do promotor e anti-ativação (Rojo, 2001; Browning & Busby, 2004). No caso da anti-ativação as proteínas ativadoras de transcrição podem ser seqüestradas por proteínas transmembrana, modulando assim a transcrição. Isso ocorre na regulação do metabolismo de glucose e maltose em *E. coli*; onde o fator de transcrição MalT pode ser seqüestrado pelo transportador de maltose transmembrana (Böhm & Boos, 2004).

Todos estes mecanismos são bastante eficientes e dependem do tipo de regulação do promotor, mas de forma geral o impedimento estérico da ligação da RNA-polimerase ao promotor é o mecanismo mais comum de repressão. Neste caso, a seqüência de ligação do repressor deve estar dentro ou muito próxima do promotor para que este tipo de repressão seja eficiente (Rojo, 2001 e Browning & Busby, 2004).

1.4.2 Proteínas Ativadoras de Transcrição dependentes de σ^{70}

O ativador de transcrição típico de bactérias atua ligando-se em seqüências específicas do DNA (regiões de ativação) localizadas nas proximidades das seqüências promotoras. Como consequência desta ligação a proteína ativadora se posiciona de forma a fazer contato com a RNA-polimerase holoenzima e afetar uma ou mais de suas atividades (Ptashne & Gann, 1997). Na maioria dos promotores a ativação da transcrição é simples e depende apenas de uma proteína ativadora. Três mecanismos gerais para a ativação simples foram descritos.

Ativadores da classe I ligam-se a seqüência alvo de DNA a montante do promotor recrutando a RNA-polimerase para o promotor por interação direta do ativador com o domínio C-terminal das subunidades α da RNA-polimerase. O

melhor exemplo desse tipo de ativação é a proteína receptora de AMP cíclico (CRP ou CAP) atuando no promotor *lac* (Ebright, 1993). Já na ativação de proteínas da classe II, o ativador liga-se a uma seqüência alvo que é sobreposta ao promotor e interage diretamente com a subunidade σ da RNA-polimerase, facilitando sua ligação ao promotor ou modificando a etapa de iniciação (Dove, Darst & Hoschschild, 2003). Em alguns promotores o ativador pode interagir com outras subunidades da RNA-polimerase, por exemplo, o domínio N-terminal da subunidade α (Busby & Ebright, 1997). Ativadores da classe III provocam mudanças conformacionais no promotor resultando em aumento da afinidade da RNA-polimerase. Essa forma de ativação requer a ligação da proteína ativadora na região promotora ou muito próximo a ela (Sheridan, Benham & Hatfield, 1998).

1.4.3 Proteínas reguladoras de transcrição dependentes de σ^{54}

O fator σ^{54} é uma subunidade da RNA-polimerase de procaríotos de características únicas e não tem homologia significativa com outro fator sigma conhecido (Wösten, 1998). O gene que codifica esta proteína foi descrito pela primeira vez em 1977 por Garcia e colaboradores e foi denominado *glnF*; seu produto era necessário para a expressão de *glnA* (gene que codifica para glutamina sintetase) em *Salmonella typhimurium*. Somente em 1985, Hirschman e colaboradores demonstraram que *glnF* (também chamado de *ntrA*, por estar envolvido na regulação do metabolismo de nitrogênio, e *rpoN*) codificava um fator sigma da RNA-polimerase.

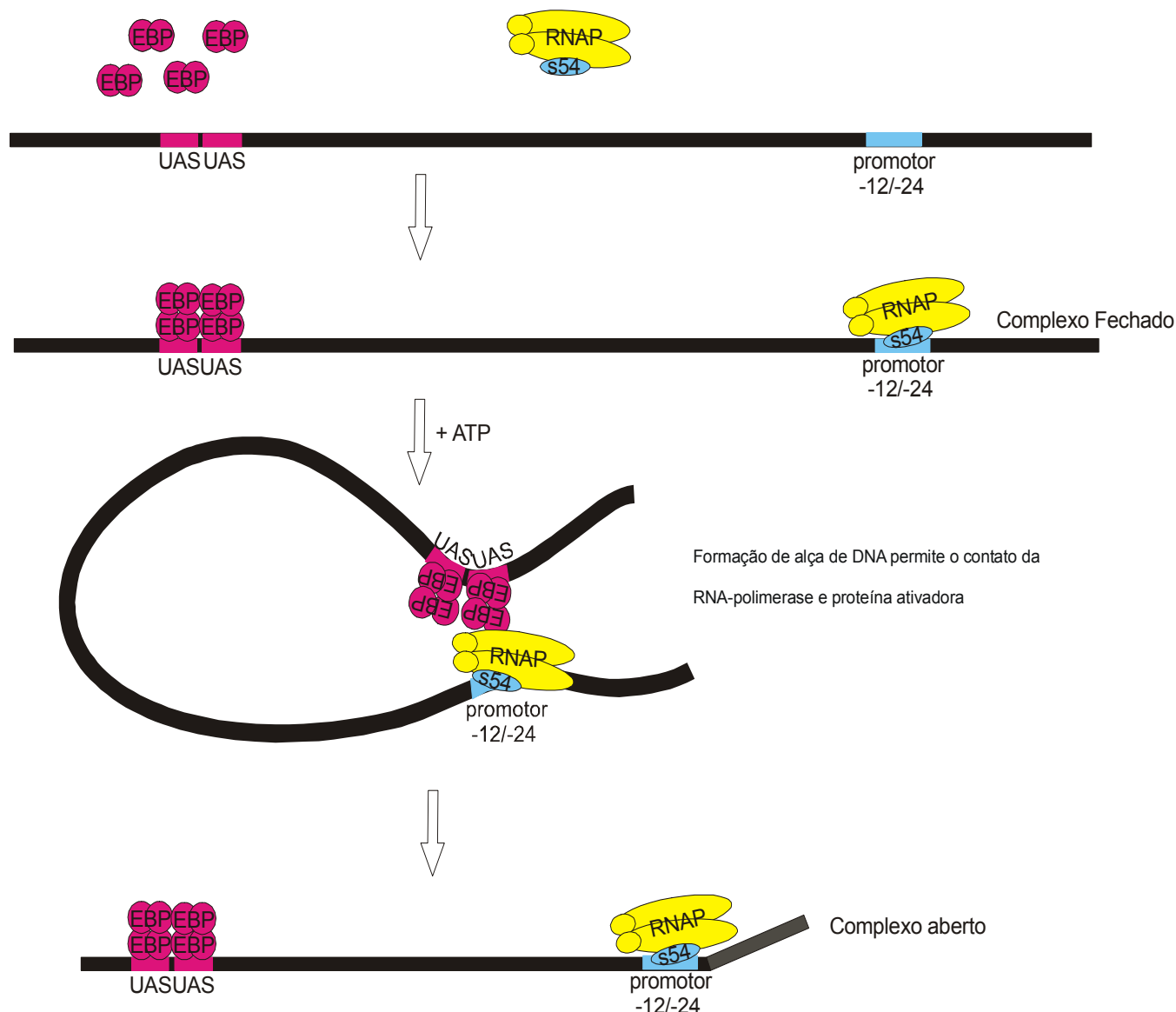
Até 1986 acreditava-se que o fator σ^{54} da RNA-polimerase estava envolvido exclusivamente no metabolismo de nitrogênio, porém Dixon (1986) mostrou que a expressão do operon *xyIABC*, que codifica para proteínas envolvidas no metabolismo de compostos aromáticos de *Pseudomonas putida*, era dependente do fator σ^{54} . No ano seguinte Ronson e colaboradores (1987) mostraram que em *Rhizobium meliloti* o transporte de dicarboxilatos era dependente do fator σ^{54} . Em seguida foi reportado que a expressão de algumas enzimas do metabolismo anaeróbico de *E. coli* também requeria o produto do gene *ntrA* (Birkmann, Sawers.& Bock, 1987). Kustu e colaboradores (1989) fizeram um levantamento de diversos genes que possuem sua expressão

dependente de σ^{54} , mas que não estão envolvidos no metabolismo de nitrogênio, tendo encontrado os genes *mela* (*Rhizobium leguminosarum*), metabolismo de melanina; *hax* (*Alcaligenes eutrophus*), hidrogenase; *flbG* (*Caulobacter crescentus*), movimentação flagelar; *dctA* (Rhizobia), metabolismo de dicarboxilato, *fdhF* (*E. coli*); selenopeptidase; *hyd-3* (*E. coli*), hidrogenase-3; *xylABCS* (*Pseudomonas putida*), metabolismo de tolueno; *pliA* (*Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Bacteroides nodosus*, *Moraxella bovis*), formação do pilli e *mbhA* (*Mixococcus xantus*), hemaglutinina.

Os promotores dependentes do fator σ^{54} possuem características que os diferenciam dos outros promotores de bactérias. A holoenzima RNA-polimerase/ σ^{54} reconhece uma sequência promotora distinta da sequência reconhecida pela RNA-polimerase/ σ^{70} . Esta sequência é caracterizada pelos dinucleotídeos “GG” e “GC” localizados nas posições -24 e -12 em relação ao início de transcrição (Kustu et al., 1989). A RNA-polimerase σ^{54} ao ligar-se ao seu promotor específico forma um complexo fechado inativo. Diferentemente da RNA-polimerase ligada a outros fatores sigma, a RNA-polimerase/ σ^{54} é incapaz de provocar a isomerização do complexo fechado a complexo aberto (Buck et al., 2000). A iniciação da transcrição só irá ocorrer na presença de um ativador de transcrição e da hidrólise de trinucleotídeo (Popham et al., 1989).

A proteína ativadora de transcrição geralmente se liga à cerca de 100 pares de base a montante do promotor. A ligação da proteína ativadora ao DNA induz sua oligomerização. Para que o complexo oligomérico ligado a montante interaja com a RNA-polimerase/ σ^{54} é necessário que o DNA se dobre ocorrendo a isomerização do complexo fechado à complexo aberto (Sasse-Dwight & Gralla, 1988 e Popham et al., 1989), em uma reação dependente da hidrólise de ATP (Su et al., 1990 e Rippe et al., 1997). O esquema desse mecanismo está mostrado na figura 3. Essa forma de ativação de transcrição é semelhante àquela das *enhancer-binding proteins* (EBPs) de eucariotos e, por esse motivo, os reguladores de transcrição dependentes de σ^{54} são também conhecidos como EBPs bacterianos (Kustu et al., 1985).

Figura 3 – Esquema de Ativação de Transcrição por Proteínas Reguladoras de Transcrição Dependentes de σ^{54}



Legenda: Modelo para ativação de transcrição dependente de σ^{54} . Um dímero ou vários da proteína ativadora de transcrição liga-se às seqüências UAS (*upstream activation sequence*) enquanto a RNA-polimerase liga-se à região promotora. A formação de uma alça no DNA permite que proteína ativadora de transcrição ligada a ATP interaja com a RNA-polimerase. A hidrólise de ATP provoca a isomerização do complexo fechado a complexo aberto. Modificado de Su et al., 1990 e Rippe et al., 1997.

As proteínas ativadoras de transcrição de promotores dependentes de σ^{54} possuem módulos estruturais bastante conservados. Na porção C-terminal esta presente um motivo de ligação a DNA do tipo hélice-volta-hélice (HTH), que permite a ligação do regulador de transcrição a seqüências ativadoras específicas a montante do promotor tipo σ^{54} (Morett & Segovia, 1993). A porção central possui um domínio AAA, capaz de hidrolisar nucleotídeos. Neste domínio encontra-se o motivo GAFTGA que é absolutamente conservado nos EBPs bacterianos. Acredita-se que esta seqüência de aminoácidos seja responsável pela interação com o fator σ^{54} e é essencial para a hidrólise de ATP e formação do complexo aberto (Zhang et al, 2002 e Rappas et al., 2005).

O domínio N-terminal tem função regulatória e controla a capacidade de ativar a transcrição destas proteínas em resposta a sinais ambientais através de modificação covalente (fosforilação), interação com pequenas moléculas efetoras ou proteínas sinalizadoras. Dependendo do mecanismo de recepção do sinal essa região do regulador de transcrição pode possuir vários módulos estruturais como o domínio sensor do regulador de resposta REC (Pfam 00072), domínio PAS (Pfam 00989), domínio GAF (Pfam 01590) e domínio V4R (Pfam 02830) (Studholme & Dixon, 2003).

A primeira proteína ativadora dependente do fator σ^{54} caracterizada foi à proteína NtrC (GlnG, NRI) (Gallardin & Magasanik, 1978 e Kustu et al., 1979). NtrC é necessária para a expressão do operon *glnAntrBC* em condições de limitação de nitrogênio (Hunt & Magasanik, 1985). Junto com proteína NtrB, a proteína NtrC constitui um sistema de dois componentes que controla a expressão de genes envolvidos no metabolismo de fontes alternativas de nitrogênio. Em condições de limitação de nitrogênio NtrB fosforila NtrC ativando-a enquanto em condições de excesso de nitrogênio NtrC é desfosforilada por NtrB e, conseqüentemente, inativada (Ninfa & Magasanik, 1986). A proteína NtrC fosforilada catalisa a formação do complexo aberto utilizando a hidrólise de ATP como fonte de energia (Popham et al., 1989). Em *E. coli* a indução da atividade de ATPase de NtrC requer a ligação cooperativa de NtrC-P ao DNA. NtrC desfosforilada compete com NtrC-P pela ligação ao DNA, podendo ligar-se fortemente a seqüência de ligação no DNA ou UAS (*upstream activation sequences*) inibindo a ligação cooperativa de NtrC-P

(Mettke et al., 1995). A Proteína NtrC liga na seguinte sequência consenso TGCACCAN3TGGTGCA (Reitzer et al., 1989).

A segunda proteína reguladora de transcrição dependente de σ^{54} caracterizada foi a proteína NifA de *Klebsiella pneumoniae*. Essa proteína é o ativador de transcrição dos genes relacionados à fixação de nitrogênio (*nif*) e sua atividade foi inicialmente descrita em *Klebsiella pneumoniae* e *Rhizobium meliloti* (Dixon et al., 1977; Ow & Ausubel, 1983; Sundaresan et al., 1983). Em 1986, Buck e colaboradores identificaram a sequência consenso de ligação ao DNA da proteína NifA nos promotores de genes relacionados à fixação de nitrogênio: TGT-N₁₀-ACA.

Uma das proteínas reguladoras de transcrição mais recentemente caracterizadas é a proteína YgaA de *E. coli*, também conhecida como NorR, ativadora da transcrição do sistema de detoxificação de óxido nítrico que atua nos mecanismos de proteção contra espécies reativas de nitrogênio (Gardener et al., 2002 e Hutchings et al., 2002). A proteína NorR possui um sítio para ligação de Fe²⁺ no domínio N-terminal GAF; o estado de oxidação do ferro determina a atividade da proteína (Autréaux et al., 2005).

A proteína XilR atua na regulação da degradação de tolueno em espécies de *Pseudomonas*, controlando a expressão de genes presentes no plasmídeo TOL (Dixon, 1984). XilR participa de um esquema de regulação complexo junto com a proteína reguladora de transcrição DmpR, onde a ligação de diferentes compostos aromáticos ao domínio N-terminal irá determinar a atividade dessas proteínas (Sze et al., 2002). A sequência consenso de ligação da proteína XilR é ATTTGN₂CAAAT (Abril et al., 1991).

Outros genes que codificam para proteínas reguladoras de transcrição dependentes de σ^{54} já estão caracterizados, como *hydH*, ativação da expressão da hidrogenase 3 em *E. coli* (Leonhartsberger et al., 2001); o gene *levR*, regulação da expressão do operon da levanase em *Bacillus subtilis* (Débarbouillé et al., 1991); *pspF*, regulação da expressão do operon *pspABCDE* em resposta à infecção pelo bacteriófago F1 em *E. coli* (Jovanovic et al., 1996). Mais recentemente foram caracterizados os genes *sfnR*, que está envolvido na regulação do operon *sfnFG* responsável pela utilização de dimetil-

sulfona em *Pseudomonas putida* (Endoh, 2005) e *hbpR*, que controla os genes responsáveis pelas etapas iniciais da degradação de 2-hidroxibifenol em *Pseudomonas azelaica* (Tropel & van der Meer, 2005).

Em *H. seropedicae* somente NtrC e NifA já foram caracterizadas. Em *H. seropedicae* NtrC é necessária para a ativação da transcrição do operon *glnAntrBC* em condições de limitação de nitrogênio e está envolvida tanto na regulação da síntese quanto da atividade da glutamina sintetase (GS) (Persuhn et al., 2000). NtrC é capaz de se ligar na região promotora de *glnA* mesmo desfosforilada (Twerdochlib et al., 2003).

O gene que codifica para a proteína NifA em *H. seropedicae* foi identificado por Souza e colaboradores (1991). Neste organismo a proteína NifA também atua na regulação da transcrição dos genes *nif*, tem sua atividade regulada por níveis de oxigênio e amônio e provavelmente interage com proteínas do tipo PII (Souza et al., 1999; Monteiro et al., 2001 e Wassem et al., 2002).

A sequência genômica completa de *H. seropedicae* determinada pelo consórcio GENOPAR (www.genopar.org) permitiu a localização *in silico* dos genes de reguladores de transcrição dependentes de σ^{54} por homologia de sequência. A identificação dos genes regulados por proteínas ativadoras de transcrição dependentes de σ^{54} permitirá a compreensão da função deste fator sigma alternativo na regulação da expressão gênica em *H. seropedicae* e em outras Proteobactérias.

2 OBJETIVOS

Identificar os genes que codificam para proteínas reguladoras de transcrição dependentes de σ^{54} de *Herbaspirillum seropedicae*, identificar as sequências de DNA reconhecidas por essas proteínas e os prováveis genes alvo desses reguladores.

2.1 Objetivos Específicos

- Identificar as proteínas reguladoras de transcrição presentes no genoma de *H. seropedicae* por análise *in silico*;

- Analisar as sequências das proteínas reguladoras de transcrição encontradas, identificando seus domínios estruturais e classificá-las nas subfamílias dos reguladores de transcrição;

- Planejar oligonucleotídeos iniciadores para amplificação dos domínios hélice-volta-hélice (HTH) das proteínas reguladoras de transcrição dependentes de σ^{54} identificadas;

- Clonar, super-expressar e purificar os domínios HTH das proteínas reguladoras de transcrição dependentes de σ^{54} de *H. seropedicae* selecionadas;

- Identificar as sequências de ligação ao DNA dos domínios HTH das proteínas reguladoras de transcrição dependentes de σ^{54} de *H. seropedicae*;

- Localizar no genoma de *H. seropedicae* as sequências de DNA reconhecidas pelas proteínas reguladoras de transcrição, identificando os genes que provavelmente são regulados pelos fatores de transcrição dependentes de σ^{54} de *H. seropedicae*.

3 MATERIAIS E MÉTODOS

3.1 Bactérias e Plasmídios

A estirpes de bactérias e plasmídios utilizados neste trabalho estão listados na tabela 1.

Tabela 1 – Estirpes e Plasmídios

Estirpe ou Plasmídio	Genótipo ou Características relevantes	Referências
<i>Herbaspirillum seropedicae</i> - SmR1	Derivada da estirpe Z78, Sm ^R , Nif ⁺	Souza et al., 1995
<i>Escherichia coli</i> – DH10B	Sm ^R , F'[proAB+ lacZΔM15	Gibco BRL
<i>Escherichia coli</i> - BL21λ (DE3)	B F ⁻ dcm ompT hsdS(rB- mB ⁻) gal λ(DE3)	New England Biolabs
pET-28a	vetor de expressão, gera proteína de fusão com segmento His-Tag, promotor T7, Km ^R	Novagen
pUC18	vetor para clonagem, p/lac, lacZDM15, Amp ^R	Yanisch-Perron et al., 1985
pGEM-T Easy	vetor para clonagem de produtos de PCR com adenina na ponta 3', pT7, Amp ^R	Promega
pALI 208	Possui a seqüência que codifica para o domínio HTH da proteína 208.0173 de <i>H. seropedicae</i> clonado em pGEM-T, Amp ^R	este trabalho
pALI 216	Possui a seqüência que codifica para o domínio HTH da proteína 216.0019 de <i>H. seropedicae</i> clonado em pGEM-T, Amp ^R	este trabalho
pALI 287	Possui a seqüência que codifica	este trabalho

	para o domínio HTH da proteína 287.0556 de <i>H. seropedicae</i> clonado em pGEM-T, Amp ^R	
pALIP 208	Possui a seqüência que codifica para o domínio HTH da proteína 208.0173 de <i>H. seropedicae</i> subclonado em pET28a, Km ^R	este trabalho
pALIP 216	Possui a seqüência que codifica para o domínio HTH da proteína 216.0019 de <i>H. seropedicae</i> subclonado em pET28a, Km ^R	este trabalho
pALIP 287	Possui a seqüência que codifica para o domínio HTH da proteína 287.0556 de <i>H. seropedicae</i> subclonado em pET28a, Km ^R	este trabalho
série pALiap 208	Possuem os aptâmeros do domínio HTH da proteína 208.0173 clonados em pUC18, Amp ^R	este trabalho
série pALiap 216	Possuem os aptâmeros do domínio HTH da proteína 216.0019 clonados em pUC18, Amp ^R	este trabalho
série pALiap 287	Possuem os aptâmeros do domínio HTH da proteína 287.0556 clonados em pUC18, Amp ^R	este trabalho

3.2 Meios de Cultura e Condições de Cultivo

H. seropedicae foi crescido em meio NFb-Malato (Klassen *et al.*, 1997), a 30°C.

A composição do meio de cultura Nfb é a seguinte:

	Concentração (g/L)
MgSO ₄ .7H ₂ O	2 x 10 ⁻¹
NaCl	1 x 10 ⁻¹

CaCl ₂	2×10^{-2}
Ácido nitrilo-triacético	$5,6 \times 10^{-2}$
FeSO ₄ . 7H ₂ O	2×10^{-2}
Malato de sódio	5
Biotina	1×10^{-4}
Na ₂ MoO ₄ .2 H ₂ O	2×10^{-3}
MnSO ₄ . H ₂ O	$2,35 \times 10^{-3}$
H ₃ BO ₃	$2,8 \times 10^{-3}$
CuSO ₄ .5 H ₂ O	8×10^{-5}
ZnSO ₄ .7 H ₂ O	$2,4 \times 10^{-4}$

O meio Nfb sólido foi obtido com a adição de 15 g/L de ágar a mistura de meio líquido. O meio completo é denominado NfbHPN (Nfb com adição de NH₄Cl 20 mmol/L e tampão fosfato 50 mmol/L – (7,97 g/L de KH₂PO₄ e 0,89 g/L de K₂HPO₄)). O pH final do meio foi 6,0.

As estirpes de *E. coli* foram cultivadas em meio LB (Luria – Bertani Broth), TB (Terrific Broth), SOB ou SOC (Sambrook et al., 1989) a 37°C. As culturas em meio líquido foram mantidas sob agitação de 130 rpm.

A composição do meio de cultura Luria-Bertani Broth (LB) (LA) é:

	Concentração (g/L)
Triptona	10
Extrato de levedura	5
NaCl	10

O pH foi corrigido para 7,0 com NaOH 1 mol/L. O meio sólido (LA) foi obtido com a adição de 15 g/L de ágar a mistura de meio líquido.

A composição do meio de cultura Terrific Broth é:

	Concentração (g/L)
Bacto-triptona	12
Extrato de levedura	24
KH ₂ PO ₄	2,3
K ₂ HPO ₄	12,54

Glicerol	4
----------	---

A composição do meio de cultura SOB é:

	Concentração (g/L)
Bacto-triptona	20
Extrato de levedura	5
NaCl	0,584
KCl	0,186

A composição do meio de cultura SOC é:

	Concentração (g/L)
Bacto-triptona	20
Extrato de levedura	5
NaCl	0,6
KCl	0,19
MgCl ₂	0,94
MgSO ₄	1,2
Glucose	3,6

Os meios de cultura foram esterilizados por autoclavação (120°C por 20 minutos).

3.3 Antibióticos

As soluções de antibióticos foram preparadas em água milli-Q estéril e esterilizadas por filtração em filtro Millipore HAWP 0,45 µm e estocadas a – 20°C. As concentrações de antibióticos no meio de cultura para *E. coli* foram: estreptomicina (Sm) 80 µg/mL, canamicina (Km) 50 µg/mL ou ampicilina (Amp) 250µg/mL; e para *H. seropedicae* foi: estreptomicina (Sm) 80 µg/mL.

3.4 Preparo de Solução de RNase

A solução estoque de RNase (10 mg/ml) foi preparada em Tris-HCl 10mmol/L pH 7,5 contendo 15 mmol/L de NaCl e fervida durante 20 minutos para inativação de desoxiribonucleases contaminantes.

3.5 Purificação de DNA

3.5.1 DNA Genômico de *H. seropedicae*

A purificação de DNA genômico foi realizada segundo um protocolo modificado de Souza (1990). Uma cultura de *H. seropedicae* (1500 μ L) foi centrifugada em microcentrífuga EppendORF minispin a 12000g por 1 minuto. O sobrenadante foi descartado e o precipitado de células foi ressuspenso em 500 μ L de tampão GET (glucose 50 mmol/L, Tris-HCl 25 mmol/L pH 8,0 e EDTA 10 mmol/L pH 8,0). A essa mistura adicionou-se 50 μ L de SDS 10% e a mistura foi incubada à 60°C por 20 minutos para que houvesse lise celular. Ao lisado celular foram adicionados 200 μ g/mL de Pronase E (solução estoque 20 mg/mL pré-tratada com incubação à 37°C por 1 hora) seguido de incubação a 37°C por 24 horas. A mistura foi tratada com 100 μ L de fenol:clorofórmio:álcool isoamílico (25:24:1), homogeneizada por inversão (cerca de 100 vezes) e posteriormente centrifugada a 12000g por 15 minutos. A fase aquosa foi transferida para outro tubo e o DNA precipitado com 2,5 volumes de etanol absoluto e recuperado por centrifugação por 20 minutos a 12000g. O precipitado de DNA foi lavado duas vezes com etanol 80%, seco à vácuo e dissolvido em 200 μ L de água milli-Q autoclavada. A solução de DNA foi conservada em geladeira.

Após a purificação, a concentração e pureza do DNA foram determinadas por espectrofotometria. Para isto, o DNA purificado foi diluído (50 vezes) em água milli-Q autoclavada. A relação absorvância $_{260\text{nm}}$ / absorvância $_{280\text{nm}}$, que indica o grau de pureza do DNA, variou de 1,86 a 1,93 nas diversas purificações que foram realizadas. Valores entre 1,8 e 2,0 indicam DNA com alto grau de pureza. A absorvância a 260 nm permite a determinação da concentração de DNA através da relação: absorvância 1,0 corresponde à concentração de 50 μ g/mL de DNA dupla fita.

3.5.2 DNA Plasmidial

A purificação de plasmídios foi realizada pelo método de lise alcalina conforme descrito em SAMBROOK *et al.* (1989). Quando foi utilizado como molde para sequenciamento, o DNA plasmidial sofreu um tratamento adicional com 2,5µg/mL de RNase por 2 horas.

3.6 Separação Eletroforética

3.6.1 DNA

As eletroforeses de DNA foram realizadas em gel de ágar ou agarose conforme descrito por SAMBROOK *et al.* (1989). A visualização do DNA foi realizada após tratamento com solução de azul de metileno (0,25%), no caso dos géis de agarose de baixo ponto de fusão ou sob luz UV (302 nm) após tratamento com brometo de etídeo (0,5 µg/mL).

3.6.2 Proteínas em Condições Desnaturantes (SDS-PAGE)

Foi utilizado o sistema descontínuo descrito por Laemmli (1970), com dodecil-sulfato de sódio (SDS) como agente desnaturante em géis com concentração de acrilamida de 20% (proporção de 29:1 acrilamida/bis-acrilamida).

3.7 Ferramentas de Bioinformática

Os genes que codificam para prováveis proteínas reguladoras de transcrição de *H. seropedicae* foram localizados no banco de dados do Programa Genopar (<http://www.genopar.org>) após categorização funcional de acordo com o banco de dados do COG (Cluster of Orthologous Group) (Tatusov, Koonin & Lipman, 1997) realizada pelo programa GAAT (Genomic Assembly and Analysis Tool) (Favetti *et al.*, 2003). Foram realizadas comparações através do programa Blast (Altschul, 1990) com o banco de dados nr do GeneBank (www.ncbi.nlm.nih.gov/blast).

Os domínios estruturais das proteínas reguladoras de transcrição de *H. seropedicae* foram identificados pelo programa SMART (Simple Modular Architecture Research Tool) (Schultz *et al.*, 1998 e Letunic *et al.*, 2002) e wise 2

que faz comparações com o banco de dados Pfam (Protein families database) (Bateman, 2004)(<http://www.sanger.ac.uk>).

O ponto isoelétrico e a massa molecular dos domínios HTH com cauda de resíduos de histidina foram calculados utilizando-se o programa Compute pI/MW Tool (<http://us.expasy.org>) (Bjellqvist et al., 1993; Bjellqvist et al., 1994 e Wilkins et al., 1998).

Os alinhamentos das seqüências de aminoácidos e de nucleotídeos foram realizados pelo programa ClustalW (Higgins, et al. 1994). Este programa realiza alinhamentos entre seqüências de proteínas, identificando similaridades.

3.8 Planejamento dos Oligonucleotídeos Iniciadores

Foram planejados 2 pares de oligonucleotídeos iniciadores para cada provável gene a partir da seqüência depositada no banco de dados do Programa Genoma do Paraná - GENOPAR. Os oligonucleotídeos 208N, 216N e 287N são complementares a seqüência de nucleotídeos que codifica para a região a montante ao início do domínio HTH de cada uma das proteínas e possuem o sítio de restrição para a enzima *NdeI*. Os oligonucleotídeos 208C, 216C e 287C são complementares a seqüência de nucleotídeos que codifica para a região a jusante ao final do domínio HTH das respectivas proteínas, possuindo o sítio de restrição para a enzima *HindIII*. Esses iniciadores encontram-se listados abaixo, os sítios de restrição estão destacados em vermelho.

208N – 5' TACACGAC**CATATG**CACACC 3'

208C – 5' GGCAA**AAGCTT**CCTGACG 3'

216N – 5' ACCCTC**CATATG**CTGCGT 3'

216C – 5' GTTGAA**AAGCTT**AAGTCCCT 3'

287N – 5' ATCTCTCG**CATATG**GAGC 3'

287C – 5' CAGGA**AAGCTT**GTCATCT 3'

A posição dos iniciadores e as alterações de seqüência inseridas estão mostradas na figura 6.

3.9 Amplificação de Fragmentos de DNA

A amplificação do DNA foi realizada pela técnica da reação em cadeia da polimerase (PCR) como descrito por Kocher e Wilson (1991).

O sistema de amplificação possuía 50 ng/ μ L de DNA genômico de *H. seropedicae*, tampão de PCR (Tris-HCl 20 mmol/L pH 8,4 e KCl 50 mmol/L), 1,5 mmol/L de $MgCl_2$, 0,2 mmol/L de dNTPs, 0,4 pmol/ μ L de cada oligonucleotídeo iniciador, 5 U de *Taq* DNA polimerase em um volume final de 20 μ L. Os parâmetros utilizados para amplificação da região que codifica pra o domínio HTH da proteína 208.0173 e da proteína 216.0019 foram: um ciclo de 95°C por 5 minutos, cinco ciclos de 95°C por 15 segundos, 45,4°C por 30 segundos e 72°C por 1 minuto, seguido por 35 ciclos de 95°C por 15 segundos, 52°C por 30 segundos e 72°C por 1 minuto e um ciclo de 72°C por 5 minutos. Para o domínio HTH da proteína 287.0556 foram: um ciclo de 95°C por 5 minutos, cinco ciclos de 95°C por 15 segundos, 47,4°C por 30 segundos e 72°C por 1 minuto, seguido por 35 ciclos de 95°C por 15 segundos, 54,2°C por 30 segundos e 72°C por 1 minuto e um ciclo de 72°C por 5 minutos

3.10 Clivagem do DNA com Enzimas de Restrição

O DNA plasmidial do vetor pET28a e dos plasmídios pALI 208, pALI 216 e pALI 287 foram clivados em tampão React 2 (Tris-HCl 50 mmol/L, $MgCl_2$ 10 mmol/L e NaCl 50 mmol/L ; pH 8,0; Invitrogen) com 10 U de cada enzima (*NdeI* e *HindIII*). O sistema de restrição do DNA plasmidial possuía ainda 2,5 μ g/mL de RNase, e foi incubado a 37° C por 4 horas.

O DNA genômico da estirpe SmR1 (235 μ g) foi clivado em tampão H (Tris-HCl 50 mmol/L, $MgCl_2$ 10 mmol/L, NaCl 100 mmol/L e DTT 100mmol/L; pH 7,5) com 10U da enzima *Sau3AI* e 2,5 μ g/mL de RNase; em um volume final de 300 μ L. O sistema foi incubado por 12 horas a 37° C.

3.11 Reação de Ligação de DNA

Os sistemas de ligação foram montados com uma proporção molar de vetor:inserto de 1:5. Foram utilizados tampão de T4 DNA ligase (50 mmol/L Tris-HCl pH 7,6; 10 mmol/L MgCl₂; 1 mmol/L DTT; 0,5 mmol/L ATP e 5% PEG 8000) e 0,5 U de T4 DNA ligase em um sistema de volume final 20 µL. A solução contendo apenas vetor e inserto foi aquecida a 65°C por 10 minutos e posteriormente mantidas em gelo por 5 minutos antes da adição do tampão e da enzima. A reação foi incubada a 18°C por 12 horas.

As ligações onde foi utilizado o vetor pGEM-T Easy foram realizadas com produtos de PCR sem nenhum tipo de tratamento, tampão de T4 DNA ligase (50 mmol/L Tris-HCl pH 7,6; 10 mmol/L MgCl₂; 1 mmol/L DTT; 0,5 mmol/L ATP e 5% PEG 8000), 1 U de T4 DNA ligase e 0,2µL da solução de vetor, em um sistema de volume final 20 µL. Esse sistema também foi incubado por 12 horas.

3.12 Preparo de Células Eletrocompetentes

Uma cultura com 40mL de *E. coli* crescida por 3 a 5 horas (DO 0,5) foi centrifugada a 2500g por 5 minutos, a 4°C. O sobrenadante foi descartado e o sedimento ressuspense em 40 mL de água milli-Q estéril gelada. Essa suspensão foi centrifugada novamente a 2500g por 5 minutos (4°C), e o sedimento ressuspense como anteriormente descrito. Após descartar o sobrenadante as células foram lavadas com solução de glicerol 15% estéril e gelado. Finalmente, após centrifugação de 2500g por 5 minutos, as células foram ressuspensas em 300 µL de glicerol 15% estéril e gelado, aliqüotadas (20 µL) e conservadas à -70°C.

3.13 Transformação Bacteriana por Eletroporação

Alíquotas de células eletrocompetentes de *E. coli* foram descongeladas em banho de gelo e foram adicionados 0,2-1 µL de solução de DNA ou reação de ligação. Essa mistura foi homogeneizada, transferida para câmaras de eletroporação e submetida a um pulso elétrico de 4 kV, 200 Ω, 330 µF. A seguir, as células foram incubadas com 1 mL de meio SOC, sem antibióticos, a

37°C e sob agitação (130 rpm por 40 minutos) para permitir sua recuperação. Alíquotas de 50-800 µL de células foram plaqueadas em meio LA com antibióticos adequados à seleção das células recombinantes e incubadas de 16-24 horas a 37°C.

3.14 Expressão e Purificação de Proteínas

Os plasmídios pALIP208, pALIP216 ou pALIP287 foram transformados em *E. coli* BL21λ (DE3) e os transformantes foram selecionados em meio LA enriquecido com 2% de glucose e 50µg/mL de canamicina. As colônias obtidas foram então inoculadas em 100mL de meio LB (50µg/mL de canamicina). A cultura foi incubada em agitador rotatório a 37°C até que atingisse a DO 0,5, quando foi adicionado 0,1mmol/L de IPTG, e a cultura permaneceu sob agitação por mais 4 horas a 37°C.

A cultura foi centrifugada a 12000g por 10 minutos a 4°C. O precipitado de células foi pesado e congelado à -70°C por 12 horas. O precipitado de células foi descongelado em banho de gelo e cada grama foi suspenso em 3 mL de tampão de sonicação (50mmol/L Tris-HCl pH 7,5; 500mmol/L NaCl e 10% de glicerol). A essa mistura foi adicionado 1mmol/L de PMSF (Fenil-metil-sulfonil-fluoreto) e 200µg/mL de lisozima, seguido de e incubação em banho de gelo por 30 minutos.

As células foram lisadas em sonicador em 3 pulsos de 20 segundos com 1 minuto de intervalo entre eles. O lisado celular foi centrifugado a 12000g por 10 minutos a 4°C. O sobrenadante foi aplicado em uma coluna de cromatografia de afinidade do tipo His-trap (GE Health Care, 1mL).

A coluna foi previamente carregada com 50mmol de cloreto de níquel e equilibrada com 5mL de tampão de sonicação. Após a injeção da amostra a coluna foi lavada com 5mL de tampão de sonicação, seguido de 10mL de tampão de lavagem 1 (50mmol/L Tris-HCl pH 7,5; 500mmol/L NaCl, 10% de glicerol e 50mmol/L de imidazol). A eluição foi realizada com 5mL de tampão I (50mmol/L Tris-HCl pH 7,5; 500mmol/L NaCl, 10% de glicerol e 100mmol/L de imidazol), 5mL de tampão II (50mmol/L Tris-HCl pH 7,5; 500mmol/L NaCl, 10% de glicerol e 300mmol/L de Imidazol) e 5 mL de tampão III (50mmol/L Tris-HCl pH 7,5; 500mmol/L NaCl, 10% de glicerol e 500mmol/L de imidazol). Na fase

de eluição as foram coletadas frações de 500 μ L. Todo o processo cromatográfico foi realizado em bomba peristáltica, com fluxo de 1mL/min e as frações obtidas foram analisadas por eletroforese SDS PAGE.

As frações que apresentaram a proteína de interesse foram reunidas, e dialisadas por 12 horas contra 50 volumes de tampão de estocagem (50mmol/L Tris-HCl pH 7,0; 500mmol/L NaCl e 50% de glicerol). A solução de proteínas dialisadas foi conservada à -70°C.

3.15 Dosagem de Proteínas

A quantificação de proteínas foi realizada através do método de Bradford (1976). Albumina bovina foi utilizada como padrão.

3.16 Reação de Interação DNA-Proteína e Isolamento de Aptâmeros

O ensaio de interação DNA-proteína para isolamento de aptâmeros continha 240 μ g de DNA genômico de *H. seropedicae* SmR1 clivado parcialmente com a enzima de restrição *Sau3AI* e 1 μ mol/L de proteína (HTH 208.0173-His, HTH 216.0019-His ou HTH 287.0556-His) em tampão TAP (50mmol/L Tris-acetato pH 8,0; 1mmol/L de acetato de potássio, 8mmol/L de acetato de magnésio e 27mmol/L de acetato de amônio) em um volume final de 600 μ L. Essa mistura foi mantida a temperatura ambiente por 30 minutos.

Os complexos DNA-proteína foram isolados em coluna Ni-NTA Spin (Qiagen). A coluna foi equilibrada com 600 μ L de TAP e posteriormente carregada com 600 μ L de reação de interação DNA-proteína. A coluna foi então lavada 2 vezes com 600 μ L tampão de lavagem (50mmol/L Tris-acetato pH 8,0; 1mmol/L de acetato de potássio, 8mmol/L de acetato de magnésio, 27mmol/L de acetato de amônio e 50mmol/L de imidazol). Os complexos DNA-proteína foram eluídos com 200 μ L de tampão de eluição (50mmol/L Tris-acetato pH 8,0; 1mmol/L de acetato de potássio, 8mmol/L de acetato de magnésio, 27mmol/L de acetato de amônio e 500mmol/L de imidazol), repetido 2 vezes. Todas as etapas de centrifugação foram realizadas em centrífuga mini spin à 700g por 2 minutos.

A fração eluída da coluna (400 μ L) foi tratada com 50 μ L de fenol:clorofórmio:álcool isoamílico (1:24:1), agitada em vórtex e centrifugada a 12.000g por 15 minutos. A fase aquosa foi coletada e o DNA precipitado com 3 volumes de etanol absoluto. O precipitado de DNA foi lavado 2 vezes com 500 μ L de etanol 70%, seco a vácuo e ressuspenso em 16 μ L de água milli-Q autoclavada. O DNA recuperado foi ligado a vetor pUC18 digerido com *Bam*HI e defosforilado com fosfatase alcalina de camarão.

3.17 Análise de Plasmídios Recombinantes por PCR

A presença de inserto nos plasmídios recombinantes foi determinada por PCR. As colônias foram coletas com uma ponteira estéril, inoculadas em 20 μ L (tubo) ou 10 μ L (placa) de reação de PCR contendo: tampão de PCR (Tris-HCl 20 mmol/L pH 8,4 e KCl 50 mmol/L), 1,5 mmol/L de MgCl₂, 0,2 mmol/L de dNTPs, 0,4 pmol/ μ L de cada oligonucleotídeo iniciador, 5 U de *Taq* DNA polimerase, e o sistema foi incubado em um termociclador. Os parâmetros de reação foram: um ciclo de 95°C por 5 minutos seguido de 30 ciclos de 95°C por 30 segundos, 60°C por 50 segundos, 72°C por 2 minutos e um ciclo de 72°C por 10 minutos. Os produtos de amplificação foram analisados por corrida eletroforese em gel de agarose.

3.18 Reações de Sequenciamento de DNA

As amostras de DNA foram seqüenciadas pelo método de terminação de cadeia por didesoxiribonucleotídeos (Sanger et al., 1977). O sequenciamento foi feito utilizando o Seqüenciador Automático de DNA Modelo ABI377 (Applied Biosystems).

As reações possuíam aproximadamente 100ng de DNA molde, 0,6pmol/ μ L de oligonucleotídeo iniciador reverso ou universal e 3 μ L de reativo de sequenciamento (ET mix GE Health Care); em um volume final de 8 μ L. Esta solução foi submetida à reação de amplificação linear utilizando o seguinte programa: um ciclo de 95°C por 5 minutos e 30 ciclos de 95°C por 30 segundos e 60°C por 1 minuto e 30 segundos.

4 RESULTADOS E DISCUSSÃO

4.1 Identificação dos Genes que Codificam para Proteínas Reguladoras de Transcrição no Genoma de *H. seropedicae*

A determinação do circuito de regulação de expressão gênica global de um organismo é necessária para a compreensão dos diversos processos celulares, como divisão celular e resposta a alterações ambientais. O sequenciamento completo de diversos genomas, principalmente de procariotos, disponibilizou uma grande quantidade de informações e permite a identificação de prováveis proteínas envolvidas na regulação da transcrição *in silico*. A compilação e organização das prováveis proteínas regulatórias de um genoma completo permite, algumas vezes, a postulação de circuitos regulatórios. Esse tipo de informação facilita a abordagem experimental, que é necessária e essencial para a confirmação das indicações obtidas pela abordagem computacional.

Existem diversas formas descritas de busca de classes de genes em um genoma (Pérez-Rueda & Collado-Vides, 2000). Entretanto, em todas abordagens automáticas 9 a 30% do total de genes detectados são falsos positivos (Pérez-Rueda & Collado-Vides, 2000). Outro problema relacionado com esse tipo de busca são falsos negativos, ou seja, genes que codificam proteínas reguladoras, mas que não são detectados. Para minimizar estes problemas, todas as prováveis ORFs do genoma de *H. seropedicae* (PROGRAMA GENOPAR) foram categorizadas por comparação de cada provável ORF com o banco de dados do COG (*Cluster of Orthologous Group*) e, então foi realizada uma procura manual por palavras chaves no banco de dados produzido pelo programa GAAT (*Genomic Assembly and Analysis Tool*) (Favetti et al., 2003) que consiste dos resultados de comparação obtidos utilizando o Programa BlastP (Altschul, 1990) e banco de dados GenBank (www.ncbi.nlm.nih.gov).

O banco de dados COG categoriza funcionalmente as proteínas identificadas em sequências genômicas completas de procariotos, organizando-as em famílias de proteínas ortólogas e, portanto, que devem

compartilhar funções comuns (Tatusov, Koonin & Lipman, 1997). Os grupos funcionais definidos pelo COG estão mostrados no apêndice II.

As proteínas classificadas no grupo K (Transcrição) estão de forma geral envolvidas no processo de transcrição e são, portanto, potencialmente proteínas regulatórias. Para identificar as proteínas de *H. seropedicae* pertencentes a este grupo foi realizada uma procura do banco de dados do programa GAAT utilizando as palavras “transcription” e/ou “response regulator”, que invariavelmente são incluídas na descrição dos membros do grupo K. Esta busca resultou em 444 ORFs potencialmente envolvidas no processo de transcrição. A análise dos resultados de comparação com o banco de dados revelou que nem todos os genes identificados correspondem a reguladores de transcrição: 62 destes genes codificam proteínas envolvidas em outras etapas do processo de transcrição. Foram localizadas 8 ORFs que provavelmente codificam helicases, 12 prováveis histona acetil-transferases, 2 exorribonucleases, 19 fatores sigma (5 semelhantes a *fecI*, 7 semelhantes a *rpoE*-like, 1 *prhI*, 1 *rpoS*, 1 *hrpL*, 1 *fliA*, 1 *rpoH*, 1 *rpoN*, e 1 *rpoD*), 4 subunidades da RNA-polimerase (genes *rpoA*, *rpoB*, *rpoC* e *rpoZ* que codificam as subunidades α , β , β' , e ω , respectivamente), 3 fatores de alongamento de transcrição (*greA*, *greB* e *nusA*), 2 fatores de terminação (*nusB* e *rho*), 3 fatores de anti-terminação (três genes semelhantes a *nusG*), 1 deacetilase NAD-dependente, 1 ribonuclease específica para ds-RNA, 2 proteínas envolvidas em resposta SOS (*lexA* e *umuD*), 1 guanosina polifosfato pirofosfotohidrolase/sintase, 3 fatores anti-sigma e 1 fator *fis*.

A sequência de aminoácidos de 248 das 444 ORFs identificadas foram então analisadas utilizando os programas SMART, Wise 2 (Pfam) e BlastX. Foram localizadas 169 prováveis proteínas que possuíam algum tipo de domínio de ligação ao DNA; sendo o domínio HTH o mais freqüente. Estas proteínas serão denominada de agora em diante reguladoras de transcrição. Apenas 24 ORFs (9,7% das analisadas) foram consideradas como falsos positivos, pois não apresentaram nenhum tipo de domínio conservado.

Os resultados obtidos são concordantes com aqueles observados em *E. coli* K-12 (Pérez-Rueda & Collado-Vides, 2000). Como o número de prováveis proteínas reguladoras que não foram analisadas foi de 162, e considerando que a proporção de falsos positivos seja mantida, então se pode dizer que *H.*

seropedicae possui aproximadamente 330 genes que codificam para proteínas reguladoras de transcrição. Esse número corresponde à cerca de 7% dos genes presentes nesse organismo.

Aproximadamente 1% das ORFs analisadas apresentavam mudança de fase de leitura devido à inserção ou deleção de bases. Embora uma parte destas seqüências representem a seqüência real e são, portanto pseudogenes inativos, é muito provável que a maioria das mudanças de fase de leitura foi decorrente de erro de sequenciamento e constitui genes ativos. Qualquer que seja o motivo, entretanto, a presença de mudança de fase de leitura produziu um superestimativa de número de prováveis reguladores já que a busca semi-automática utilizada não identifica a mudança de fase de leitura, levando a contabilizar 2 ou mais proteínas derivadas do mesmo gene. Esta superestimativa também ocorreu para as outras proteínas relacionadas à transcrição. Para este trabalho, quando foi detectada mudança de fase de leitura o gene foi contabilizado codificando apenas uma proteína e considerado um gene ativo.

4.1.1 Classificação e Identificação dos Domínios Estruturais das Proteínas Reguladoras de Transcrição de *H. seropedicae*

As famílias dos reguladores de transcrição geralmente possuem um domínio receptor de sinal que pode ligar pequenas moléculas, outras proteínas ou um substrato e um domínio de ligação ao DNA característicos.

As 169 ORFs analisadas foram classificadas de acordo com os domínios estruturais identificados pelos programas SMART e Wise 2 (Pfam) (<http://www.sanger.ac.uk/Software/Pfam/dnasearch.shtml>) (Apêndice I). A distribuição dos reguladores de transcrição de *H. seropedicae* em cada família está apresentada na Tabela 2 .

Tabela 2 - Distribuição das proteínas reguladoras de transcrição de *H. seropedicae* dentro das famílias de reguladores de transcrição

Família	Número de reguladores identificados
SIS	6
CheY	27
EBP	11
AraC	18
TetR	10
ArsR	2
AsnC	7
MerR	6
Xre	5
Lacl	7
GntR	18
IclR	11
MarR	7
LysR	34
Total	169

A maioria das proteínas reguladoras de transcrição identificadas em *H. seropedicae* são das famílias LysR (20,1%) e CheY (16%). Esse resultado é semelhante ao descrito para *E. coli* K-12 (Perez-Rueda & Collado-Vides, 2000). Os reguladores do tipo LysR são estruturalmente simples e possuem um domínio HTH N-terminal e um domínio receptor de sinal na porção C-terminal. Essa família de proteínas reguladoras atua sobre uma grande variedade de processos, regulando a expressão de genes envolvidos no metabolismo de carboidratos, compostos aromáticos, aminoácidos, toxinas, regulação osmótica, patogenicidade, e resistência a antibióticos (Schell, 1993).

Já os reguladores do tipo CheY geralmente são os reguladores de resposta de sistemas de dois componentes. A maioria dos reguladores do tipo CheY possuem o domínio HTH na porção C-terminal, entretanto alguns

possuem domínios do tipo *winged-helix*, característico de proteínas como OmpR (Martinez-Hackert & Stock, 1997). Os sistemas de dois componentes são comuns e regulam a expressão de genes necessários para diversas funções celulares, como metabolismo de nitrogênio e fosfato, colonização e infecção de hospedeiro, transportadores de íons, resistência a antibióticos, estresse oxidativo e adaptação a alterações de pH (Stock, Robinson & Goudreau, 2000).

As proteínas reguladoras de transcrição de *H. seropedicae* classificadas na família TetR, possuem apenas o domínio N-terminal conservado. A maioria das proteínas reguladoras de transcrição desta família funciona como repressoras de transcrição de operons envolvidos em transporte através da membrana e permeabilidade do envelope (Grkovic, Brown & Skurray, 2002). O domínio de ligação ao DNA é do tipo HTH e encontra-se na porção N-terminal da proteína. A porção C-terminal dessas proteínas contém diversas regiões que podem estar envolvidas na ligação de indutores, tais como antibióticos e drogas antimicrobianas. As proteínas reguladoras de transcrição tipo TetR incluem proteínas regulatórias bacterianas de sistemas de exportação de drogas, que protegem a bactéria contra antibióticos (por exemplo, tetraciclina), anti-sépticos, desinfetantes e outros compostos antimicrobianos (Grkovic, Brown & Skurray, 2002).

Todos os reguladores da família SIS identificados em *H. seropedicae* possuem na porção N-terminal o domínio característico desta família e na porção C-terminal um domínio de ligação de fosfo-açúcares. Esta família de proteínas regulatórias está envolvida no metabolismo de fosfo-açúcares (Bateman, 1999).

Os reguladores de transcrição classificados na família LacI em geral são repressores de transcrição. A maioria dos reguladores das famílias MarR, IclR, GntR, Xre, MerR, AsnC, ArsR e AraC de *H. seropedicae* analisados possuem os domínios de ligação ao DNA característicos de cada família, entretanto não possuem o domínio N-terminal receptor conservado. Não foi possível atribuir funções para a maioria destas proteínas regulatórias, pois apesar de possuírem domínios conservados não possuem similaridade com nenhum gene já caracterizado em outros organismos. Estes resultados sugerem que uma boa

parte destas proteínas estão presentes unicamente em *H. seropedicae* e devem participar de controle de expressão de genes específicos deste organismo. A presença de genes que codificam para proteínas reguladoras de transcrição com função desconhecida torna interessante o estudo mais aprofundado de cada um desses reguladores de *H. seropedicae*.

Foram identificadas 11 proteínas ativadoras de transcrição dependentes de σ^{54} (EBPs).

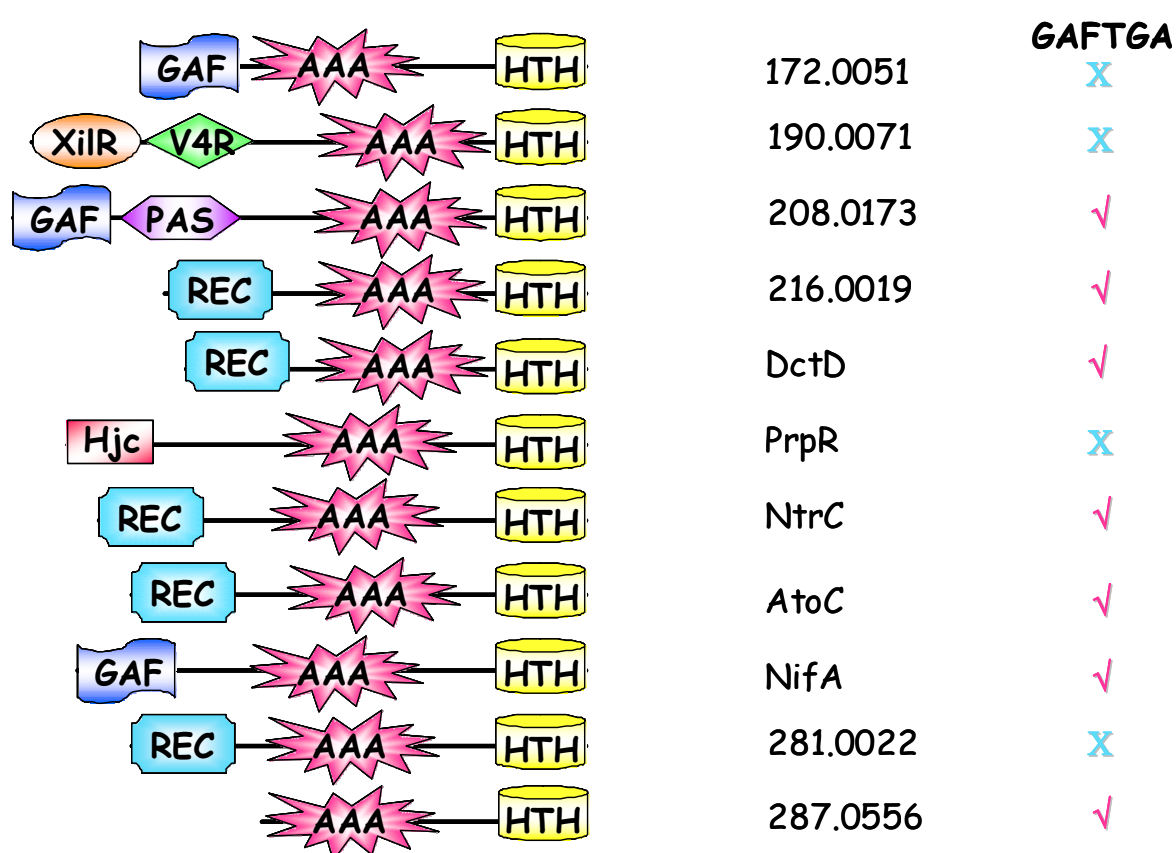
4.2 Análise das proteínas reguladoras de transcrição dependentes de sigma 54 localizadas no genoma de *H. seropedicae*

Todas as 11 proteínas reguladoras de transcrição dependentes de σ^{54} identificadas codificadas no genoma de *H. seropedicae* apresentaram o domínio para hidrólise de nucleotídeos (domínio AAA) e o motivo de ligação ao DNA do tipo HTH (hélice-volta-hélice), reunindo, portanto as características necessárias para serem caracterizados como EBPs (Dixon & Studholme, 2003). Neste grupo estão incluídas as proteínas NifA e NtrC de *H. seropedicae*, que já foram caracterizadas como proteínas ativadoras de transcrição dependentes de σ^{54} em outros estudos (Souza et al., 1999; Monteiro et al., 2001; Wassem et al., 2002; Persuhn et al., 2000 e Twerdochlib et al., 2003); a proteína reguladora de transcrição NorR (172.0051), que provavelmente está envolvida com o metabolismo de espécies reativas de nitrogênio; a proteína PoxR (190.0071), que provavelmente está envolvida no metabolismo de hidrocarbonetos aromáticos; a proteína DctD, provavelmente relacionada ao metabolismo de dicarboxilato; a AtoC, que provavelmente está envolvida no metabolismo de ácidos graxos de cadeia curta; a PrpR, que provavelmente está envolvida com o metabolismo de propionato; a proteína semelhante a HydG (281.0022), cuja função ainda não foi identificada; uma proteína que provavelmente está envolvida no metabolismo de acetoína (208.0173); uma proteína semelhante a NifA, porém sem o domínio N-terminal sensor (287.0556) e uma proteína com um domínio sensor tipo REC (216.0019) (Figura 4).

O motivo conservado “GAFTGA”, que provavelmente interage com o fator σ^{54} da RNA-polimerase (Zhang et al, 2002), não foi identificado em todas

estas proteínas reguladoras de transcrição: sete proteínas possuem esse motivo (NifA, NtrC, DctD, AtoC, 208.0173, 216.0019 e 287.0556) e as outras quatro (172.0051, 190.0071, PrpR, 281.0022) possuem alguma mudança de resíduo de aminoácido no motivo. A Figura 4 mostra esquematicamente os domínios estruturais de todas as prováveis proteínas reguladoras de transcrição dependentes de σ^{54} de *H. seropedicae* identificadas.

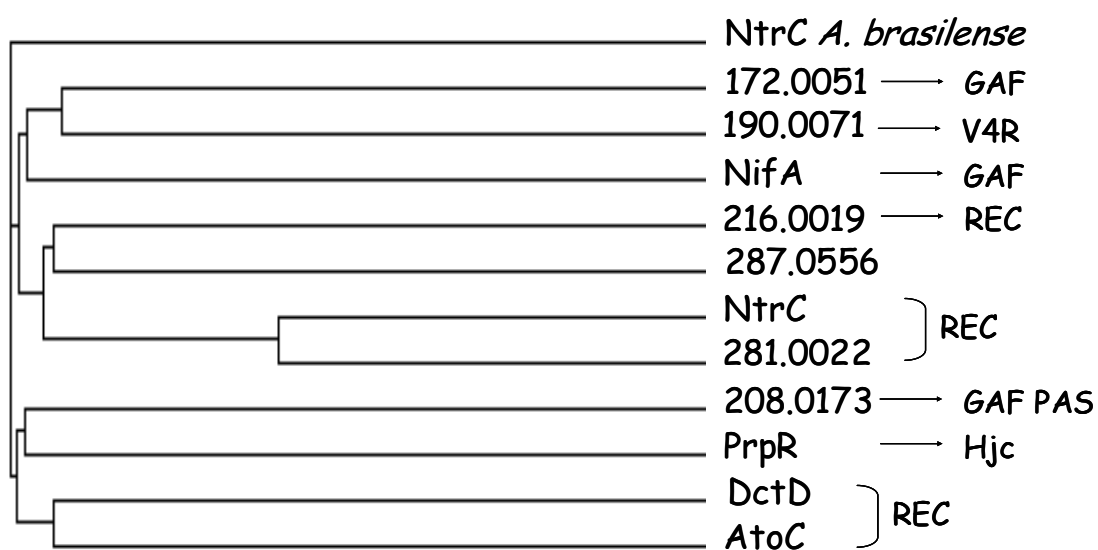
Figura 4 – Domínios das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de *H. seropedicae*



Legenda: As estrelas rosa correspondem ao domínio de hidrólise de nucleotídeos AAA; os quadrados azul claro ao domínio REC (domínio receptor de sinal); os paralelogramos azuis escuros ao domínio GAF (liga pequenas moléculas ou outras proteínas); o hexágono roxo, o domínio PAS; o retângulo vermelho, o domínio Hjc; a elipse marrom, o domínio XilR; o losango verde o domínio de ligação de hidrocarbonetos aromáticos, V4R; e os cilindros amarelos, o domínio de ligação ao DNA com motivo HTH. O símbolo “✓” indica a presença do motivo conservado GAFTGA e o símbolo “X” a ausência.

Para determinar a provável relação evolutiva existente entre as proteínas reguladoras de transcrição σ^{54} , as suas sequências de aminoácidos foram alinhadas pelo programa ClustalW, e este alinhamento foi então utilizado para construção de um dendrograma de similaridade (figura 5). A sequência de aminoácidos da proteína NtrC de *Azospirillum brasilense* foi utilizada como raiz do dendrograma. As proteínas que possuem domínio N-terminal do tipo GAF e V4R foram agrupadas no mesmo clado, enquanto as proteínas que possuem domínio REC encontraram-se distribuídas pelo dendrograma. Esse tipo de dendrograma coincide com o aquele apresentado por Studholme & Dixon (2003), quando mostraram a relação entre todos os reguladores de transcrição dependentes de σ^{54} descritos até então.

Figura 5- Dendrograma das Proteínas Reguladoras de Transcrição dependentes de σ^{54} de *H. seropedicae*.



O alinhamento das seqüências de aminoácidos de todas as proteínas reguladoras de transcrição dependentes de σ^{54} de *H. seropedicae* (apêndice III) permitiu o cálculo da identidade existente entre elas. O menor valor obtido foi de 17,8% (NifA e 208.0173) e o maior 35,5% (DctD e 216.0019). As sequências de aminoácidos dos domínios AAA (apêndice IV) e HTH (apêndice V) foram alinhadas isoladamente e o maior valor de identidade para o domínio AAA foi de 65,5% (281.0022 e 216.0019) e o menor 41,3% (PrpR e 208.0173); para o

domínio HTH os valores foram 48,7% (281.0022 e 190.0071) e 17% (NifA e NtrC), respectivamente. Esses valores mostram que o domínio mais conservado é o AAA. Este resultado justifica o fato de que todas estas proteínas compartilham a função de interagir com a RNA-polimerase/ σ^{54} , hidrolisar ATP e catalisar o início de transcrição (Buck et al., 2000). Por outro lado, o domínio HTH é mais variável, pois cada uma das proteínas reconhece uma sequência específica e diferente: alguns resíduos de aminoácidos são conservados e responsáveis interação com o esqueleto de ligações fosfato do DNA e pela conformação de hélice-volta-hélice, mas os outros são variáveis e fazem interações específicas com as bases nitrogenadas do seu sítio de ligação ao DNA (Brennan & Matthews, 1989 e Jones et al., 1999).

A presença do motivo GAFTGA, que provavelmente é necessário para interação com o fator σ^{54} da RNA-polimerase, sugere que a proteína reguladora é capaz de ativar a transcrição. A ausência desse motivo, por outro lado, indica que essa proteína provavelmente seja incapaz de ativar a transcrição e pode atuar como um repressor de transcrição.

Para este trabalho foram selecionadas as proteínas que possuíam o motivo GAFTGA sem alterações de aminoácidos e cuja sequência de ligação ao DNA em *H. seropedicae* ainda não foi caracterizada. As proteínas codificadas pelas ORFs 208.0173, 216.0019 e 287.0556 foram então escolhidas para purificação do domínio HTH e identificação do seu sítio de ligação ao DNA.

4.3 Amplificação dos Fragmentos de DNA que Codificam para o Domínio HTH das Proteínas 208.0173, 216.0019 e 287.0556.

Os pares de oligonucleotídeos iniciadores (208N e 208C, 216N e 216C e 287N e 287C) foram planejados para amplificar fragmentos de DNA contendo somente a região que codifica para o domínio HTH das respectivas proteínas. Nos oligonucleotídeos 5' foi incluída uma mutação para introduzir um sítio de restrição para a enzima *NdeI* e no oligonucleotídeos 3' foi introduzido sítio para a enzima *HindIII* (Material e Métodos). A temperatura de anelamento calculada, o provável peptídeo gerado pela expressão do fragmento de DNA amplificado

por cada par de iniciadores e seus respectivos pontos isoelétricos e massa molecular teóricos estão listados na Tabela 3.

Tabela 3 - Fragmentos de DNA obtidos por PCR e dos peptídeos expressos.

ORF	oligonucleotídeos iniciadores	T _m (°C)	Fragmento de DNA obtido (pb)	Peptídeo gerado após clonagem em pET28a (aa)	Massa Molecular (Da)	Ponto Isoelétrico
208.0173	208N	60	246	94	8392,41	11,23
	208C	56				
216.0019	216N	56	249	91	10547,86	10,09
	216C	56				
287.0556	287N	54	301	105	12010,45	7,11
	287C	52				

A análise dos domínios estruturais de cada proteína foi importante para determinar as prováveis regiões codificadoras de cada um deles. Os oligonucleotídeos correspondente à porção 5' hibridizam com a região entre os domínios AAA e HTH e os da região 3' após o códon de término do domínio HTH. Essa disposição dos iniciadores permitiu a amplificação de fragmentos de DNA pequenos que ao serem traduzidos produziram um peptídeo com conformação do domínio nativa do domínio HTH. Modelos estruturais dos peptídeos teóricos (incluindo os resíduos de aminoácidos da cauda de histidinas) foram obtidos utilizando os programas 3D-JIGSAW (Bates & Sternberg, 1999), CPH-models 2.0 (Lund, 2002) e ESyPred3D (Lambert, De Bolle & Depiereux, 2002) e, em todos os casos, foi prevista a formação do motivo volta-hélice-volta e não foi detectada evidência de que os resíduos de aminoácidos do domínio central ou da cauda de histidinas poderia causar impedir estericamente a ligação ao DNA.

As regiões correspondentes a cada um dos domínios, bem como a região que foi amplificada pelos pares de oligonucleotídeos iniciadores esta mostrada na Figura 6.

Figura 6 – Sequência de Nucleotídeos e Aminoácidos das ORFs 208.0173, 216.0019 e 287.0556.

ORF 208.0173

	M Q Y Q A N R P S A G A L L G T S P L V	
81	AGAGGTCCCAAGGAGACAAGGATGCAATACCAGGCAAACCGGCCAGCGCCGGTGCATTGCTGGGCACCTCGCCGCTTGT	160
81	TCTCCAGGGTTCTCTGTTCTACGTTATGGTCCGTTTGGCCGGTTCGCGGCCACGTAACGACCCGTGGAGCGGCAACA	160
	Q A S T D G V P D L Q A I E R S H Q R S S S Y G I M	
161	GCAGGCCAGCACCGATGGCGTGCCGGATCTGCAGGCCATCGAGCGCTCTCACCAGCGCAGCTCGTCTACGGCATCATGC	240
161	CGTCCGGTCTGTGGCTACCGCACGGCCTAGACGTCCGGTAGCTCGCGAGAGTGGTTCGGTTCGAGCAGGATGCCGTAGTACG	240
	R G Q R P D L D S L S R A D L A Q T L E A N Q T L S G	
241	GCGGCCAGCGGCCCGACCTCGACTCGCTCTCGCGCGCCGACCTGGCGCAGACCCTCGAAGCCAACCAGACCCCTCTCCGGC	320
241	CGCCGGTTCGCGGGCTGGAGCTGAGCGAGAGCGCGGGCTGGACCGCTCTGGGAGCTTCGGTTGGTCTGGGAGAGGCCG	320
	H A R P V M E T L Y D Q I R N T H S M V I L S D A E G	
321	CACGCCCCCGGGTTCATGGAAACCCCTCTACGACCAGATCCGCAATACCCACAGCATGGTGATCCTGTCCGGATGCAGAAGG	400
321	GTGCGGGCGGGCCAGTACCTTTGGGAGATGCTGGTCTAGGCGTTATGGGTGTCGTACCCTAGGACAGCCTACGTCTTCC	400
	T I L H T L G D S D F L A K A D R V A L G P G G R W	
401	CACCATCCTGCATACCCCTCGGCGACAGCGATTTCTGGCCAAGGCCGATCGCGTCGCCCTTGGGCCGGGAGGGCGCTGGT	480
401	GTGGTAGGACGTATGGGAGCCGCTGTCGCTAAAGGACCGGTTCCGGCTAGCGCAGCGGAACCCGGCCCTCCCGCGACCA	480
	S E Q L R G T N A I G T A L F E Q K P T T V H A Q Q H	
481	CCGAGCAACTGCGCGGCACCAACGCCATCGGCACAGCGCTGTTTCGAGCAGAAGCCCACTACCGTACACGCCAGCAGCAT	560
481	GGCTCGTTGACGCGCCGTTGGTTGCGGTAGCCGTGTCGCGACAAGCTCGTCTTCGGGTGATGGCATGTGCGGGTCTGTCGTA	560
	Y L D A N R F L T C S A A P I F D H R G Q V L G V L D	
561	TACCTGGACGCCAATCGCTTCTGACCTGCTCGGCCGCCCATCTTCGACCATCGCGCCAGGTGCTGGGGGTGCTGGA	640
561	ATGGACCTGCGGTTAGCGAAGGACTGGACGAGCCGGCGGGGTTAGAAGCTGGTAGCGCCGGTCCACGACCCCCACGACCT	640
	V S G E C G S F H K H T M A L V R M S A Q M I E N H	
641	TGTCTCGGGTGAATGCGGCAGTTTCCACAAGCACACCATGGCGCTGGTGCATGTTCGGCGCAGATGATCGAGAATCACC	720
641	ACAGAGCCCACTTACGCCGTCAAAGGTGTTTCGTGTGGTACCGCGACCACGGTACAGCCGGTCTACTAGCTCTTAGTGG	720
	L L A G S F P D C I T L H F H S R A E F V G T L V E G	

721	TGCTGGCCGGCAGCTTTCCTGATTGCATCACCTGCATTTCCACAGCCGCGCCGAATTCGTCGGCAGCTGGTCTGAAGGC	800
721	ACGACCGCCGTCGAAAGGACTAACGTAGTGGGACGTAAAGGTGTCTGGCAGCCGTTAAGCAGCCGTGCGACCAGCTTCCG	800
	I V S F T P G G R F L A A N R S A Q F Q L G L T L A A	
801	ATCGTCAGCTTTCACGCCGGGCGGTCTGCTTTCCTGGCGGCCAATCGCAGCGCGCAGTTCCAGCTGGGCCTGACCCTGGCGGC	880
801	TAGCAGTCGAAGTGCGGCCCGCCAGCGAAGGACCGCGGTTAGCGTCGCGCGTCAAGGTGACCCGGACTGGGACCGCCG	880
	L Q S H T F A S L F G V P L S L L F E H H R K A A P	
881	GTTGCAATCGCATACTTCGCTTCACTGTTTCGGGGTGCCTGCTGTCGCTGCTGTTTCGAGCATCATCGCAAGGCCGCCCCGG	960
881	CAACGTTAGCGTATGGAAGCGAAGTGACAAGCCCCACGGCGACAGCGACGACAAGCTCGTAGTAGCGTTCCGGCGGGGCC	960
	G L L E L C L H N G I R I H G R A E L R L R D Q H F Q	
961	GCCTGCTGGAGCTGTGCCTGCACAATGGCATCCGCATCCATGGCCGCGCCGAATTGCGCCTGCGCGACCAGCATTTCAG	1040
961	CGGACGACCTCGACACGGACGTGTTACCGTAGGCGTAGGTACCGGCGCGGCTTAACGCGGACGCGCTGGTCTGTAAGGTC	1040
	A A R E A V I D A P G L L P P A P N T T T P P P S P P	
1041	GCCGCGCGGAGCCGTCATCGATGCGCCCGGCTGCTCCCGCTGCGCCCAATACCACCACCCCGCCCGCTCGCCGCC	1120
1041	CGGCGCGCTCCGGCAGTAGCTACGCGGGCCGGACGAGGGCGGACGCGGGTTATGGTGGTGGGGCGGCGGACGCGCGG	1120
	Q Q R A R H Q R R P D L Q A L D T G D V R M S A I I	
1121	CCAACAGCGCGCGCCACCAACGGCGACCGGACCTGCAGGCGCTCGATACCGGGGACGTACGCATGAGCGCCATCATTG	1200
1121	GGTTGTGCGCGCGCGGTGGTTGCCGCTGGCCTGGACGTCGCGAGCTATGGCCCTGCATGCGTACTCGGGTAGTAAC	1200
	A K L R R V I D R D I P I L I T G E T G T G K E W L A	
1201	CCAAGCTGCGCCGGGTGATCGACCGGACATCCCATCTCATCACCGGCGAGACCGGCACCGGCAAGGAATGGCTGGCG	1280
1201	GGTTCGACGCGGCCACTAGCTGGCGCTGTAGGGGTAGGAGTAGTGGCCGCTCTGGCCGTGGCCGTTCTTACCGACCGC	1280
	Q A I H G D S A R F G A R Q N R A G A A P F V A V N C	
1281	CAAGCCATCCATGGCGACAGTGCACGCTTTGGCGCCCGCAGAACCAGCGCAGGCGCGGCCCCCTTCGTCGCGAGTGAAGT	1360
1281	GTTGCGTAGGTACCGCTGTACGTCGAAACCGGGCGGCTTGGCGCGTCCGCGCCGGGGAAGCAGCGTCACTTGAC	1360
	S A I P E N L I E A E L F G Y E D G A F T G A R R K	
1361	CTCGCCATCCCGAAAACCTGATCGAAGCCGAAGTGTTCGGCTATGAAGACGGCGCTTTCAGTGGCGCACGGCGCAAGG	1440
1361	GAGCCGTTAGGGCCTTTGGACTAGCTTCGGCTTGACAAGCCGATACTTCTGCCGCGAAAGTGACCGCGTCCCGCTTCC	1440
	G A L G K I A Q A H G G T L F L D E I G D M P L A L Q	
1441	GTGCGCTGGGCAAGATCGCCAGGCGCACGGCGGCACCCTGTTCTGGATGAGATCGGCGACATGCCCTGGCCCTGCAA	1520
1441	CACGCGACCCGTTCTAGCGGGTCCGCGTCCGCGGTTGGGACAAGGACCTACTCTAGCCGCTGTACGGGGACCGGGACGTT	1520
	G R L L R V L Q E R A V T P L G S Q R V I P V D F A L	
1521	GGCCGGCTGCTGCGCGTGTGCAGGAACGCGCCGTGACCGCGTGGGAGCCAGCGCGTCATCCCGGTGGATTTGCACT	1600
1521	CCGGCCGACGACGCGCACGACGTCCTTGGCGGGCACTGCGGCGACCCGTCGGTCCGCGCAGTAGGGCCACCTAAAGCGTGA	1600
	V C A T N Q P L R E M A A R H A F R E D L Y Y R I N	

1601	GGTCTGCGGACCAATCAGCCGCTGCGGAGATGGCCGCGCCATGCCTTCCGCGAAGATCTGTACTACCGCATCAATG	1680
1601	CCAGACGCGCTGGTTAGTCGGCGACGCGCTCTACCGGCGCGGGTACGGAAGGCGCTTCTAGACATGATGGCGTAGTTAC	1680
	G L Q V K L P A L R D R S D L P Q L I E R I L E D E G	
1681	GCCTTGACAGGTGAAGCTGCCGGGCTACGGGACCGCAGCGACCTCCCCAGCTGATCGAACGCATCCTGGAAGACGAAGGC	1760
1681	CGAACGTCCACTTCGACGGCCGCGATGCCCTGGCGTCGCTGGAGGGGGTCTGACTAGCTTGCCTAGGACCTTCTGCTTCCG	1760
	G P H A P R F D A Q A Q A L L L A S H W P G N L R Q L	
1761	GGCCCCACGCGCCGCGCTTCGACGCGCAAGCACAGGCGCTGCTGCTGGCCAGCCACTGGCCCCGCAACCTGCGCCAGCT	1840
1761	CCGGGGGTGCGCGGCGCAAGCTGCGCGTTCGTGTCCGCGACGACGACCGGTTCGGTGACCGGGCCGTTGGACGCGGTGCA	1840
	A N L V R T S L A M A E G E E A I G P H H L P D D F	
1841	GGCCAACCTGGTCCGCACGTCGCTGGCCATGGCCGAAGGCGAAGAGGCCATCGGTCCGCATCACCTGCCCGATGATTTCC	1920
1841	CCGGTTGGACCAGGCGTGCAGCGACCGGTACCGGCTTCCGCTTCTCCGGTAGCCAGGCGTAGTGGACGGGCTACTAAAGG	1920
	L E E A G A P G A L H D G M H T M H A V R S D C L S I	
1921	TGGAAGAGGCAGGCGCACCCGGTGCCTACACGACGGCATGCACACCATGCACGCCGTGCGCAGCGATTGCTGTCCATC	2000
1921	ACCTTCTCCGTCCGCGTGGGCCACGCGATGTGCTGCCGTACGTGTGGTACGTGCGGCACGCGTCGCTAACGGACAGGTAG	2000
	208N →	
	Q D N E W A A I E R A L R T H A G N V S A A A K A L G	
2001	CAGGACAATGAATGGGCGCGATCGAACGGGCCCTGCGCACCCACGCCGCAATGTGACGCGCCGCCAAGGCCCTGGG	2080
2001	GTCCTGTACTTACCCGGCGCTAGCTTGCCCGGGACGCGTGGGTGCGGCCGTTACAGTCGCGGGCGGGTTCGGGACCC	2080
	V S R N T I Y R R L K S R G H *	
2081	CGTGTGCGGCAATACCATCTACCGGCGCCTCAAGTCACGCGGGCACTGAGCGGGGCCAGGGCGTCATACTCCTCATCGGT	2160
2081	GCACAGCGGTTATGGTAGATGGCCGCGGAGTTTACGTGCGCCCGTACTCGCCCCGGTCCCGCAGTATGAGGAGTAGCCA	2160
2161	AAACAGCCGCGAGCGGTCAGGAAGCGCTTGCCGCTGCCGTTCTCCAACGAGAACATGCCGCCATTGCCATTGACCACGT	2240
2161	TTTGTGCGGCGCTCGCGCAGTCTTCGCGAACGGCGACGCGCAAGAGGTTGCTCTTGTACGGCGGTAACGGTAACCTGGTGCA	2240
	← 208C	

ORF 216.0019

	M E L E S P M G K T A P L M L V D D D P D L L R L L T	
1	ATGGAAGTGGAGTCCCCGATGGGAAAGACTGCCCCCTGATGCTGGTTCGATGATGATCCCGACCTGCTTCGCCTGTTGAC	80
1	TACCTTGACCTCAGGGGCTACCCTTTCTGACGGGGGACTACGACCAGCTACTACTAGGGCTGGACGAAGCGGACAACTG	80
	L R L N A S G Y R I L A V E S A E A A L A Q L A I S	
81	GCTGCGCCTCAATGCCAGCGGCTACCGCATCCTGGCCGTGGAGAGCGCCGAGGCCGCGCTGGCGCAGCTGGCCATCTCGC	160
81	CGACGCGGAGTTACGGTCGCGGATGGCGTAGGACCGGCACCTCTCGGGCTCCGGCGGACCGCGTCGACCGGTAGAGCG	160
	L P A L V I T D V R L P G M D G L A L F A E I R S R Y	
161	TGCCGGCGCTGGTGCATCCCGATGTGCGCTTGCCCGCATGGATGGGCTGGCCCTGTTTCGCCGAGATCCCGACCGCTAT	240
161	ACGGCCGCGACCACTAGTGGCTACACGCGAACGGGCCGTACTACCCGACCGGGACAAGCGGCTCTAGGCGTCGGCGATA	240

	P A L P V I L L T A H G T I P D A V D A T T Q G A F A	
241	CCGGCCTTGCCGGTGATCCTTCTTACCGCGCATGGCACCATCCCTGATGCAGTGGACGCCACCACCAGGGCGCCTTCGC	320
241	GGCCGGAACGGCCACTAGGAAGAATGGCGCGTACCGTGGTAGGGACTACGTACCTGCGGTGGTGGGTCCCGCGGAAGCG	320
	Y L T K P F D G H V L L D K V A Q A L S L A A P A P	
321	CTATCTGACCAAGCCCTTCGACGGCCATGTGCTGCTCGACAAGGTCGCCAGGCCCTGAGCCTGGCGGCGCCGGCGCCGC	400
321	GATAGACTGGTTCGGGAAGCTGCCGGTACACGACGAGCTGTTCCAGCGGGTCCGGGACTCGGACCGCCGGCGCCGGCG	400
	L S P D Q S W R A D I L S R S Q R V A E L L A E A R L	
401	TGTCGCCGACAGTCTGTGGCGGGCCGACATCTCAGCCGACAGCGGGTGGCCGAGCTGCTGGCCGAGGCCCGGCTG	480
401	ACAGCGCCCTGGTCAGCACCGCCGGCTGTAGGAGTCGGCGTGGTTCGCCACCGGCTCGACGACCGGCTCCGGCGCGAC	480
	V A A S D A S I L I R G D S G T G K E L L A R A I H R	
481	GTGGCGGCCCTCCGACGCGAGCATCCTGATCCGGGGCGATAGCGGACCGGCAAGGAGCTGCTGGCGCGGCCATCCATCG	560
481	CACCGCCGGAGGCTGCGCTCGTAGGACTAGGCCCCGTATCGCCGTGGCCGTTCTTCGACGACCGCGCGGGTAGGTAGC	560
	A S A R A E Q P F I A V N C G A I P E Q L L E S E L	
561	CGCCAGTGCACGCGCCGAGCAGCCTTTCATCGCCGTCATTTGCGGCGCCATCCCCGAGCAGTTGCTGGAGTCCGAGCTGT	640
561	GCGGTACGTCGCGGCTCGTCGGAAGTAGCGGCAAGTAAACGCGCGGTAGGGGCTCGTCAACGACCTCAGGCTCGACA	640
	F G H V K G A F T G A V E H R E G L F Q A A H G G T L	
641	TCGGCCACGTCAAGGGCGCATTACAGGGCGCGTGCAGCATCGGAAGGGCTGTTCCAGGCCGCCATGGCGGTACCTTG	720
641	AGCCGGTGCAGTTCGCGGTAAGTGCAGCGCCAGCTCGTAGCGCTTCCCGACAAGGTCGCGGGGTACCGCCATGGAAC	720
	F L D E I G D M P L P L Q V K L L R V L Q E K T V R Q	
721	TTCTCGATGAGATCGGCGACATGCCCTGCCCTGCAGGTCAAGCTCTTGGCGGTGTTGCAGGAAAAACGGTGCGCCA	800
721	AAGGAGCTACTCTAGCCGCTGTACGGGGACGGGACGTCCAGTTCGAGAACGCCACAACGTCTTTTGGCACGCGGT	800
	V G A N Q A G E V D V R I I S A T H R D L E A A M R	
801	GGTCGGCGCAACCAGGCTGGCGAGGTGCATGTGCGCATCATCTCAGCCACCCATCGCGACCTCGAAGCGGCCATGCGCG	880
801	CCAGCCGCGGTTGGTCCGACCGCTCCAGCTACACGCGTAGTAGAGTCGGTGGGTAGCGCTGGAGCTTCGCCGGTACGCGC	880
	E G Q F R E D L Y Y R L N V V T L H L P A L E E R R E	
881	AAGGACAGTTCGCGAGGATCTGTATTACCGCCTCAATGTGGTACGCTGCACCTGCCGGCGCTGGAAGAGCGGCGGGAG	960
881	TTCTGTCAAGCGCTCCTAGACATAATGGCGGAGTTACACCACTGCGACGTGGACGGCCGCGACCTTCTCGCCGCCCTC	960
	D I A L L A N H F L K T I A A K Y P K R L T G F A P D	
961	GATATCGCGCTGCTGGCCAATCACTTCTCAAGACCATTGCGGCAAGTATCCCAAGCGCCTACCGGGTTTGCSCCCGA	1040
961	CTATAGCGCGACGACCGGTTAGTGAAGGAGTTCTGGTAACGCCGGTTCATAGGGTTCGCGGAGTGGCCAAACGGGGGCT	1040
	A L E I L S L A H W P G N V R Q L Y N V V E Q V C A	
1041	CGCGCTGGAGATATTGAGCCTGGCGCACTGGCCGGCAATGTGCGCCAGCTCTATAACGTGGTGGAGCAGGTCTGCGCCT	1120
1041	GCGCGACCTCTATAACTCGACCGGTGACCGGCCGTTACACGCGGTGAGATATTGCACCACCTCGTCCAGACGCGGA	1120

	L S T A P L I P A S L V Q R A L R T P T L D L L R Y A	
1121	TGTCCACCGCACCTTGTATCCCGGCGTCACTGGTGCAGCGCGCCCTGCGCACGCCG ACCCTCGATCTGCTGCGT TATGCC	1200
1121	ACAGGTGGCGTGGAACTAGGGCCGAGTGACCACGTGCGCGGGACGCGTGC GGCTGGGAGCTAGACGACGCAATACGG	1200
	216N →	
	E A K Q K F E R D Y L I R L L K L T D G N V S D A A R	
1201	GAGGCCAAG CAGAAGTTCGAGCGTGATTACCTGATCCGCTTGCTCAAGTTGACTGATGGCAACGTGTCTGACGCCGCCCG	1280
1201	CTCCGGTTC GTCTTCAAGCTCGCACTAATGGACTAGGCGAACGAGTTCAACTGACTACCGTTGCACAGACTGCGGCGGGC	1280
	L A D R N R T E F Y R L L Q K N G L D P R R F R G G	
1281	GCTGGCTGATCGCAACCGTACCAGATTCTATCGCCTGTTGCAGAAGAAT GGACTGGATCCCCGGCGTTTCCGCGGTGGGG	1360
1281	CGACCGACTAGCGTTGGCATGGCTCAAGATAGCGGACAACGTCTTCTT ACCTGACCTAGGGGCCGAAAGGCCACCCCC	1360
	E E A V V A G R Q E E R L *	
1361	AAGAGGCTGTCGTGCGGGGCGACAGGAGGAAAGGCTTTGAAATCAGGGACTTAATCTCTTCAACGGAGGAAATTTGTCC	1440
1361	TTCTCCGACAGCAGCGCCCCGTGTCTCTCTTCCGAAACTTTAG TCCCTGAATTAGAGAAGTTG CCTCTTTAAACAGG	1440
	← 216C	
1441	TCCCTCGGCGACAGAAAAGCGGGTTGGGGTGGGGCATTGCAGAGCCCCATTTCTTGG	1498
1441	AGGGAGCCGCTGTCTTTTCGCCAACCCACCCGTAACGTCTCGGGGTAAAGGAACC	1498

ORF 287.0556

	M H A S H R S A H L Y A L P S S P E G H G D L A A R A	
1	ATGCACGCCAGCCATCGCTCTGCCATCTGTATGCGCTGCCCTCCTCGCCGGAGGGTACGGCGATCTTGCAGCGCGCGC	80
1	TACGTGCGGTGCGTAGCGAGACGGGTAGACATACGCGACGGGAGGAGCGGCCTCCAGTGCCGCTAGAACGTGCGCGCGC	80
	S P P L L T L P H A R K L G T S I R A T A Q V F H D	
81	GTCGCCACCGCTGCTGACCTTGCCCATGCGCGCAAGCTGGGCACGTCCATCCGCGCCACTGCGCAGGTATTCCATGATC	160
81	CAGCGGTGGCGACGACTGGAACGGGGTACGCGGTTGACCCGTGACGGTAGGCGGGTACGCGTCCATAAGGTACTAG	160
	P V S L A L L E R V Q R V A P S D A N V V I I G A T G	
161	CGGTTTTCGCTGGCCTTGTGGAACGCGTGCAGCGGTGGCTCCC AGCGACGCCAACGTGGTCATCATCGGCGCTACCGGC	240
161	GCCAAAGCGACCGAACGACCTTGCACGCTGCGCACCGAGGG TCGCTGCGGTTGCCACAGTAGTAGCCGCGATGGCCG	240
	T G K E L I A R H V H E L S Q R R D A P F V A V N C G	
241	ACCGCAAGGAATTGATTGCGCGCCACGTGCACGAGCTGAGCCAGCGGGGATGCGCCTTTCGTGGCGGTGAACTGCGG	320
241	TGGCCGTTCTTAACTAACGCGCGGTGCACGTGCTGACTCGGTGCGCCCTACGCGGAAAGCACCGCCACTTGACGCC	320
	A F S E S L A E S E L F G H E K G A F T G A I T H K	
321	CGGTTTTTCGGAGAGCCTGGCCGAGAGTGAGTTGTTGCGCCATGAGAAAGGCGCATTACCGGCGCCATCACCCACAAGC	400
321	GCGCAAAGCCTCTCGGACCGGCTCTCACTCAACAAGCCGGTACTCTTCCGCGTAAGTGCCGCGGTAGTGGGTGTTTCG	400
	P G W F E A A N G G T L F L D E I G D L P L A I Q V K	
401	CGGGCTGGTTTCGAGGCTGCCAATGGCGGCACCCTGTTCTCTGATGAAATCGGCGACCTGCGGCTGGCCATCCAGGTCAAG	480

401	GCCCACCAAGCTCCGACGGTTACCGCCGTGGGACAAGGAGCTACTTTAGCCGCTGGACGGCGACCGGTAGGTCCAGTTC	480
	L L R V L Q E R E V V R L G S R R S I P I D V R I I A	
481	CTGTTGCGGGTGTGCAGGAACGCGAAGTGGTGC GGCTGGGCTCACGGCGCAGCATCCCCATCGATGTGCGCATCATTGC	560
481	GACAACGCCACGACGTCCTTGGCGTTACCACGCCGACCCGAGTGCCGCGTCGTAGGGGTAGCTACACGCGTAGTAACG	560
	A T N V R L Q D A V A A G H F R Q D L Y Y R L Q V V	
561	GGCCACCAACGTGCGCCTGCAGGATGCGGTGGCGGGCCGGCATTTCGCCAGGACCTGTACTATCGGCTGCAGGTGTTGC	640
561	CCGGTGGTTGCACGCGGACGTCCTACGCCACCGCCGGCCCGTAAAGGCGGTCTGGACATGATAGCCGACGTCACACAG	640
	Q L A I P R L R E R P L D I L P L A Y H F L H E Y R Q	
641	AGCTGGCCATTCCACGCTTGC GTGAGCGGCCTCTGGACATCCTGCCGCTGGCCTATCACTTCCTGCATGAATATCGCCAG	720
641	TCGACCGGTAAGGTGCGAACGCACCTCGCCGGAGACCTGTAGGACGGCGACCGGATAGTGAAGGACGTACTTATAGCGGTC	720
	R L G Y G A A R I D G E G E R K L R Q H D W P G N I R	
721	CGCCTGGGCTATGGCGCGGCCCGCATCGATGGCGAGGGCGAACGCAAGCTGCGCCAGCATGACTGGCCGGGCAATATCCG	800
721	GCGGACCCGATACCGCGCCGGCGTAGCTACCGCTCCCGCTTGC GTTCGACGCGGTCTACTGACCGGCCCGTTATAGGC	800
	E L E N A I H H A L L I C Q D D C L R A E H F H L S	
801	CGAACTGGAAAACGCCATCCACCATGCGCTGCTGATCTGCCAGGATGACTGCCTGCGGCCGAACACTTCCATCTCTCGG	880
801	GCTTGACCTTTTGC GGTTAGGTGGTACGCGACGACTAGACGGTCTACTGACGGACGCGCGGCTTGTGAAGGTAGAGAGCG	880
	287N →	
	H V E R R L D Y A V A A A D E R G N D P L H A L E Q A	
881	ATGTGGAGCGGGCGCTGGACTACGCAGTTGCCGCCGCCGACGAGCGCGGCAATGATCCCCTGCACGCCCTGGAACAAGCC	960
881	TACACCTCGCCCGGACCTGATGCGTCAACGGCGCGGCTGCTCGCGCGTTACTAGGGGACGTGCGGGACCTTGTTCGG	960
	L Q G L F N Q S Q E G L F E K I E D T V F R T A F E F	
961	TTGCAAGGTCTGTTCAACCAGTCGAGGAGGGCCTGTTTCGAGAAGATCGAGGACACCGTCTTTTCGCACCCGCTTCGAATT	1040
961	AACGTTCCAGACAAGTTGGTCAGCGTCTCCCGGACAAAGCTCTTCTAGCTCCTGTGGCAGAAAGCGTGGCGGAAGCTTAA	1040
	C E R N Q V Q A A R L L D I S R N V L R A R L I R N	
1041	CTGCGAACGCAACCAGGTGCAGGCCGCGGCCTGCTCGACATCAGCCGCAACGTGCTGCGGGCGCGGCTGATCCGTAACG	1120
1041	GACGCTTGC GTTGGTCCACGTCGGCGCGCGGACGAGCTGTAGTCGGCGTTGCACGACGCCCGCGCGGACTAGGCATTGC	1120
	G Q I S A L K *	
1121	GTCAGATCTCGGCGCTGAAATAGGCCGATTGCAAGATGACAAGTTTTCTGACGTTTTTTTACAAAAAAGTTACAATCA	1200
1121	CAGTCTAGAGCCGCGACTTTATCCGGCGTAACGTCTACTGTTCAAAGGACTGCAAAAAAGTGTTTTTTGAATGTTAGT	1200
	← 287C	

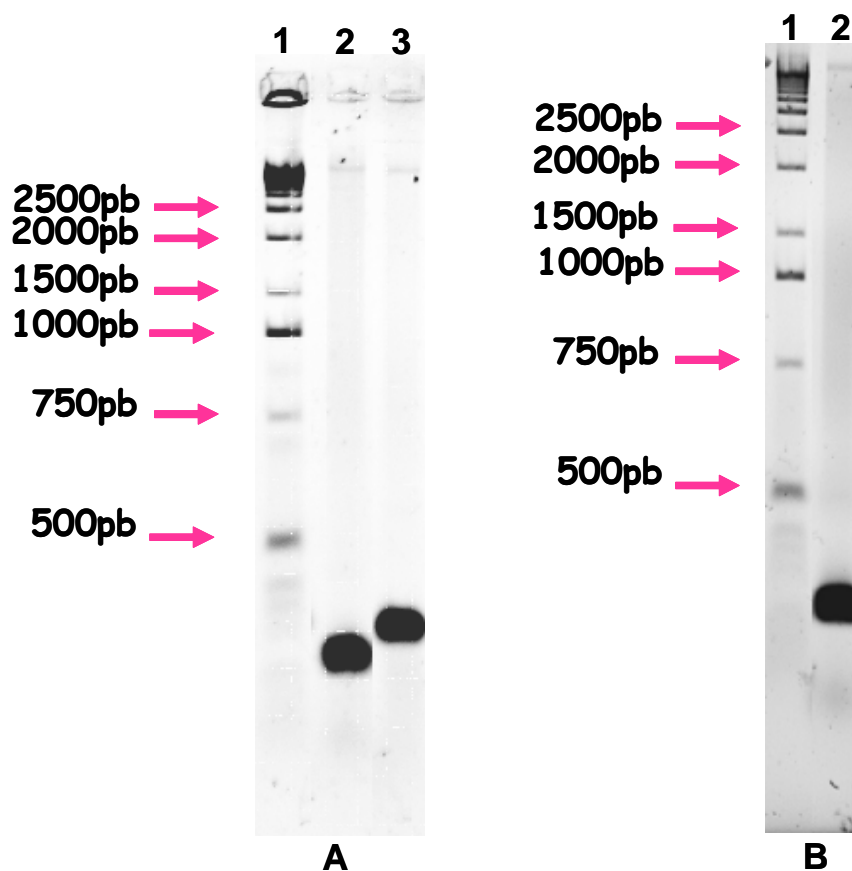
Legenda; Sequência de nucleotídeos e aminoácidos das ORFs 208.0173, 216.0019 e 287.0556. O domínio GAF está destacado em azul escuro, o domínio REC em azul claro, AAA em rosa e HTH em amarelo. O motivo GAFTGA de cada proteína está realçado em amarelo. As regiões de anelamento dos oligonucleotídeos iniciadores planejados para amplificar o

domínio HTH de cada uma das ORFs estão destacados em roxo, as setas indicam o sentido de cada iniciador.

O fragmento de DNA correspondente ao domínio HTH da proteína 208.0173 foi amplificado a partir de DNA genômico de *H. seropedicae* como molde e utilizando o par de oligonucleotídeos iniciadores 208N e 208C. A temperatura de anelamento utilizada foi de 45,4°C por 5 ciclos e 52°C por 35 ciclos. Inicialmente foi utilizada uma temperatura de anelamento menor para facilitar a amplificação, considerando que o oligonucleotídeos iniciadores possuem mudanças de nucleotídeos em relação ao molde nativo; após 5 ciclos de amplificação já existe quantidade suficiente de DNA sintetizado *de novo* para servir como molde para os próximos ciclos; permitindo assim o aumento da temperatura de anelamento. O fragmento de DNA obtido possui aproximadamente 250 pares de base.

O mesmo procedimento foi adotado para a amplificação da região que codifica para o domínio HTH da proteína 216.0019 e 287.0556. As temperaturas de anelamento utilizadas foram 45,4°C e 52°C para a 216.0019 e 47,4°C e 54,2°C para a 287.0556. Os fragmentos obtidos possuem aproximadamente 250 pares de base para 216.0019 e 300 pares de base para 287.0556. A Figura 7 mostra o eletroforetograma dos produtos de amplificação obtidos.

Figura 7 - Eletroforese dos Produtos de Amplificação de DNA Genômico de *H. seropedicae* com os Oligonucleotídeos Iniciadores 208N e 208C, 216N e 216C e 287N e 287C.



A. linha 1, marcador de massa molecular, linha 2, produto de amplificação dos iniciadores 208N e 208C, linha 3, produto de amplificação dos iniciadores 216N e 216C.

B. linha 1, marcador de massa molecular, linha 2, produto de amplificação dos iniciadores 287N e 287C. As flechas rosas indicam a banda do marcador de massa molecular equivalente à 500pb.

4.4 Clonagem dos Fragmentos de DNA que Codificam para o Domínio HTH das Proteínas 208.0173, 216.0019 e 287.0556 em Vetor de Expressão.

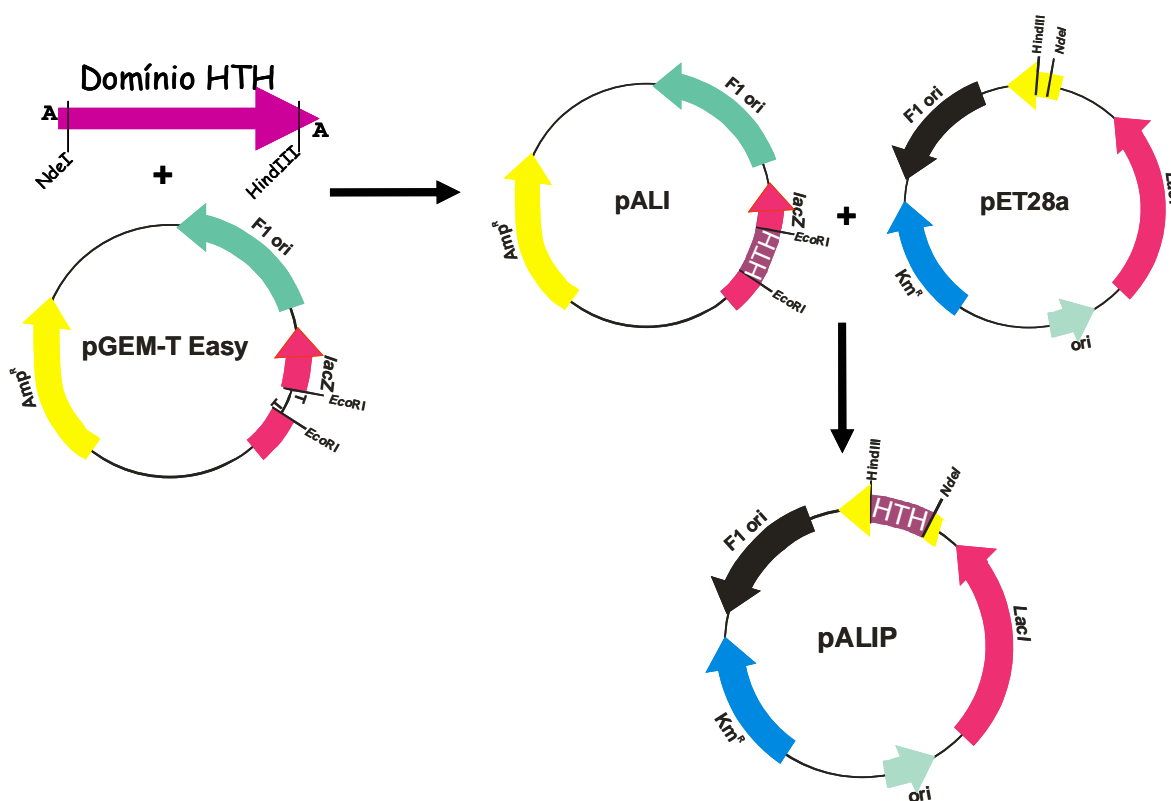
Os fragmentos de DNA obtidos por PCR foram amplificados com a enzima *Taq* DNA polimerase. Essa enzima possui a característica de adicionar uma adenina na ponta 3' do fragmento de DNA amplificado. Esta particularidade foi utilizada para a clonagem dos fragmentos no vetor pGEM-T, que possui timinas não pareadas nas pontas 5'. Essa primeira clonagem originou os plasmídios pALI 208, pALI 216 e pALI 287. Os insertos destes

plasmídios foram completamente sequenciados para confirmação da seqüência de nucleotídeos dos fragmentos clonados. A seqüência obtida foi 100% idêntica a do banco de dados GENOPAR.

Os plasmídios pALI 208, pALI 216 e pALI 287 foram clivados com as enzimas de restrição *NdeI* e *HindIII* para liberação do fragmento correspondente ao domínio HTH das proteínas 208.0173, 216.0019 e 287.0556, respectivamente. Estes fragmentos foram subclonados no vetor de expressão pET28a também tratado com as enzimas *NdeI* e *HindIII*. Os plasmídios pALIP 208, pALIP 216 e pALIP 287 obtidos foram sequenciados. O resultado obtido mostrou que todos plasmídios possuíam seqüência idêntica a do banco de dados e que os produtos de tradução teóricos possuíam uma cauda N-terminal de histidinas.

O esquema da estratégia das clonagens está mostrado na Figura 8.

Figura 8 - Estratégia para Produção de Plasmídios para Expressão dos Domínios HTH com Cauda de Histidina N-terminal.

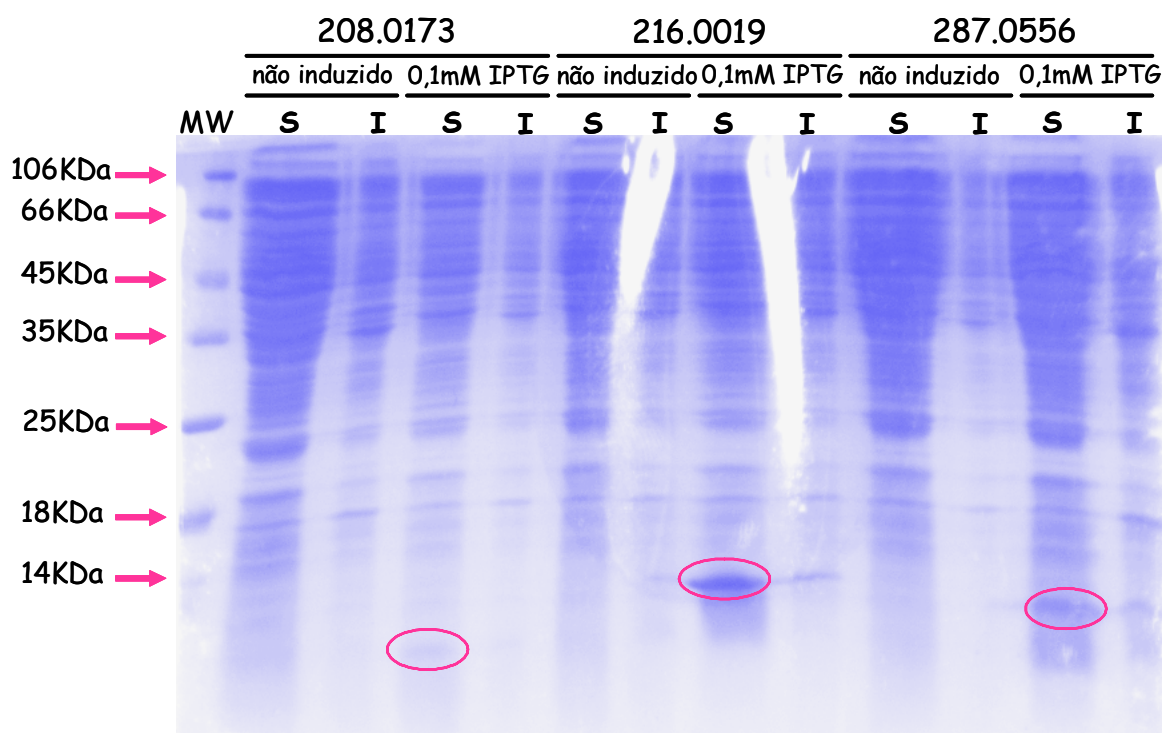


4.5 Expressão e Purificação dos Domínios 208.0173-His, 216.0019-His e 287.0556-His.

O domínio HTH-His das proteínas 208.0173, 216.0019 e 287.0556 foram super-expressos em *E. coli* BL21 λ (DE3) contendo os plasmídios pALIP 208, pALIP 216 e pALIP 287 após indução com 0,1mmol/L de IPTG por 4 horas a 37° C (Material e Métodos). As células foram coletadas por centrifugação e ressuspensas em tampão de sonicação (50mmol/l Tris-HCl pH 7,0; 500mmol/L NaCl e 10% de glicerol). A suspensão de células foi sonicada e posteriormente centrifugada como descrito em Materiais e Métodos. A eletroforese desnaturante (SDS-PAGE) revelou que os domínios HTH-His super-expressos estavam predominantemente na fração solúvel dos extratos (Figura 9).

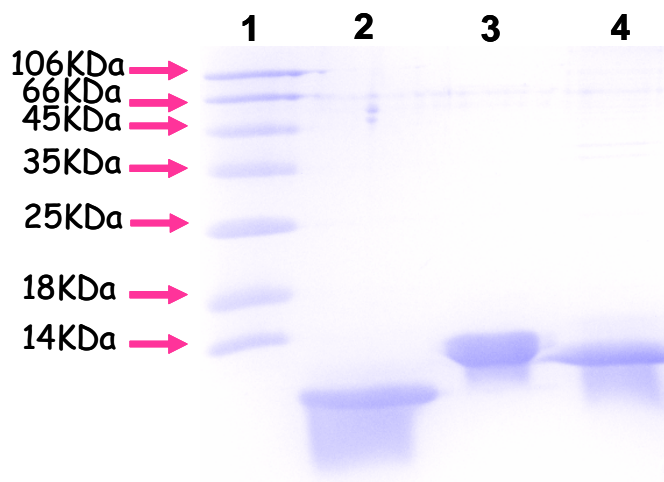
Considerando que todas as proteínas se apresentaram na fração solúvel foi então realizada a purificação das três proteínas, em apenas uma etapa cromatográfica utilizando uma coluna His-Trap (GE Health, 1mL) como descrito em Materiais e Métodos. Todas as proteínas começaram a ser eluídas com 300 mmol/L de imidazol (dados não mostrados). O eletroforetograma das proteínas purificadas está mostrado na figura 10. A pureza das proteínas foi determinada por densitometria dos eletroforetogramas corados com Coomassie Blue e digitalizados com uma câmera CCD. Os valores obtidos foram de 91% para o domínio HTH-His da proteína 208.0173, 90% para o domínio HTH-His da proteína 216.0019 e 90% para o domínio HTH-His da proteína 287.0556.

Figura 9 – Eletroforese Desnaturante das Frações Solúvel e Insolúvel de Extratos de *E. coli* BL21λ (DE3) Expressando os Domínios HTH-His das Proteínas 208.0173, 216.0019 e 287.0556 .



Legenda: Extratos de *E. coli* estirpe BL21λ(DE3), “MW”, indica o marcador de massa molecular, as linhas indicadas por “S”, representam a fração solúvel dos extratos e por “I” as frações insolúveis, “não induzido” representam os extratos sem adição de IPTG e “0,1mM IPTG” representa os extratos induzidos com 0,1mmol/L de IPTG de acordo com o descrito em Materiais e Métodos. O número 208.0173 indica os extratos celulares de *E. coli* contendo o plasmídeo pALIP 287, 216.0019, contendo o plasmídeo pALIP 216 e 287.0556, contendo o plasmídeo pALIP 287. As elipses rosas indicam as bandas correspondentes às proteínas de interesse super-expressas.

Figura 10 – Eletroforese dos Domínios HTH-His das Proteínas 208.0173, 216.0019 e 287.0556 Purificados



Legenda: Linha 1, Marcador de massa molecular; Linha 2, Domínio HTH-His da proteína 208.0173 (5µg); Linha 3, Domínio HTH-His da proteína 216.0019(5µg); Linha 4, Domínio HTH-His da proteína 287.0556(5µg).

4.6 Isolamento dos Aptâmeros de DNA

Aptâmeros são pequenas moléculas (fragmentos de DNA ou RNA, ou partes de proteínas) que possuem afinidade de ligação por outras moléculas (<http://en.wikipedia.org/wiki/Aptamer>). A determinação dos aptâmeros de DNA para cada um dos domínios HTH purificados foi realizada como descrito em Materiais e Métodos. Resumidamente, as proteínas purificadas foram incubadas com DNA genômico da estirpe SmR1 de *H. seropedicae* parcialmente digerido com a enzima de restrição *Sau3A1*. Em seguida, o complexo de DNA-proteína-His foi posteriormente aplicado em uma coluna Ni-NTA Spin (Qiagen).

Os complexos DNA-proteína-His devem permanecer ligados à resina, enquanto que o DNA não ligado à proteína foi eluído no tampão de lavagem.

Os complexos DNA-proteína-His foram então eluídos com 500mmol/L de imidazol. Os fragmentos de DNA ligados à proteína-His correspondem aos aptâmeros do domínio HTH da respectiva proteína.

A mistura de aptâmeros de DNA foi então ligada ao vetor pUC18 digerido com *Bam*HI. A reação de ligação foi eletrotransformada em *E. coli* DH10B e as colônias brancas obtidas foram analisadas como descrito em Materiais e Métodos.

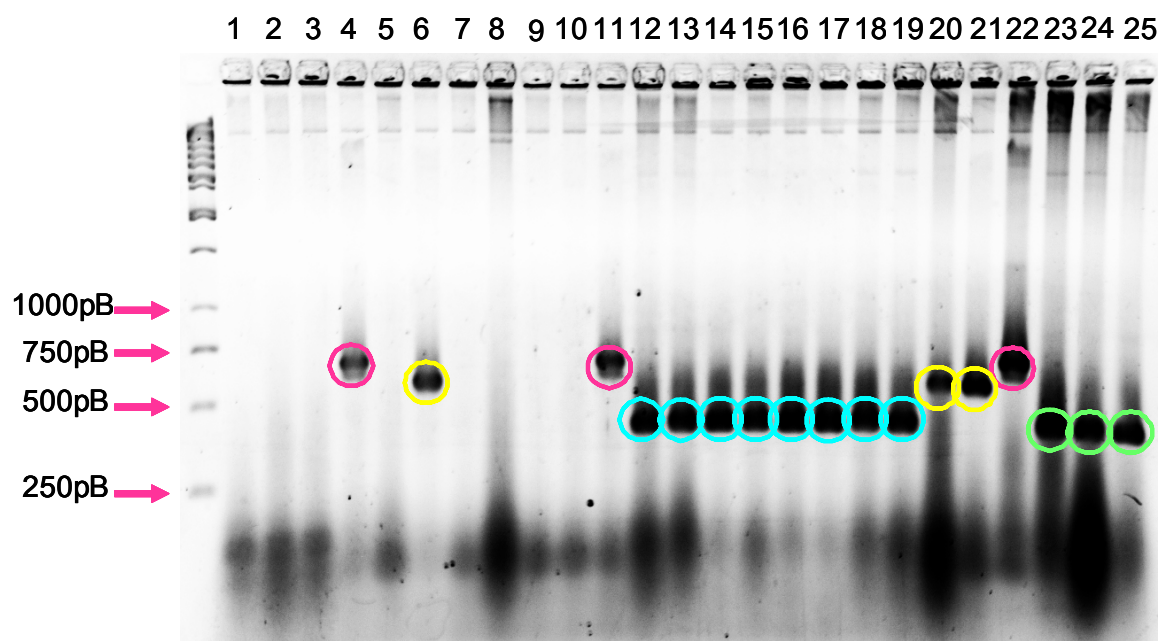
Um ensaio controle contendo todos os componentes do sistema com exceção das proteínas foi realizado inicialmente. Nenhuma colônia transformante contendo plasmídeo com inserto foi recuperada; indicando que o DNA de *H. seropedicae* parcialmente digerido com *Sau*3AI não se liga à coluna de NTA-Ni⁺² na ausência de proteína.

4.6.1 Aptâmeros Correspondentes ao Domínio HTH da Proteína 216.0019

Dois experimentos para isolamentos dos aptâmeros para o domínio HTH da proteína 216.0019 foram realizados, produzindo 144 colônias que provavelmente continham plasmídios com inserto. Essas colônias foram analisadas por PCR (Figura 11) e os insertos de 11 plasmídios foram seqüenciados parcialmente, como o descrito em Materiais e Métodos.

O sequenciamento parcial de pelo menos dois clones para cada tamanho de inserto revelou a presença do mesmo fragmento de DNA ligado ao vetor nos dois sentidos, indicando fortemente que os insertos foram selecionados pela ligação ao domínio HTH da proteína 216.0019 e não foram ligados ao vetor aleatoriamente. Os insertos obtidos possuem de 400 a 750 pares de base (tabela 4) e correspondem a quatro regiões genômicas diferentes (figura 12).

Figura 11 – Eletroforese dos Produtos de PCR dos Plasmídios Recombinantes Contendo os Aptâmeros do Domínio HTH da Proteína 216.0019

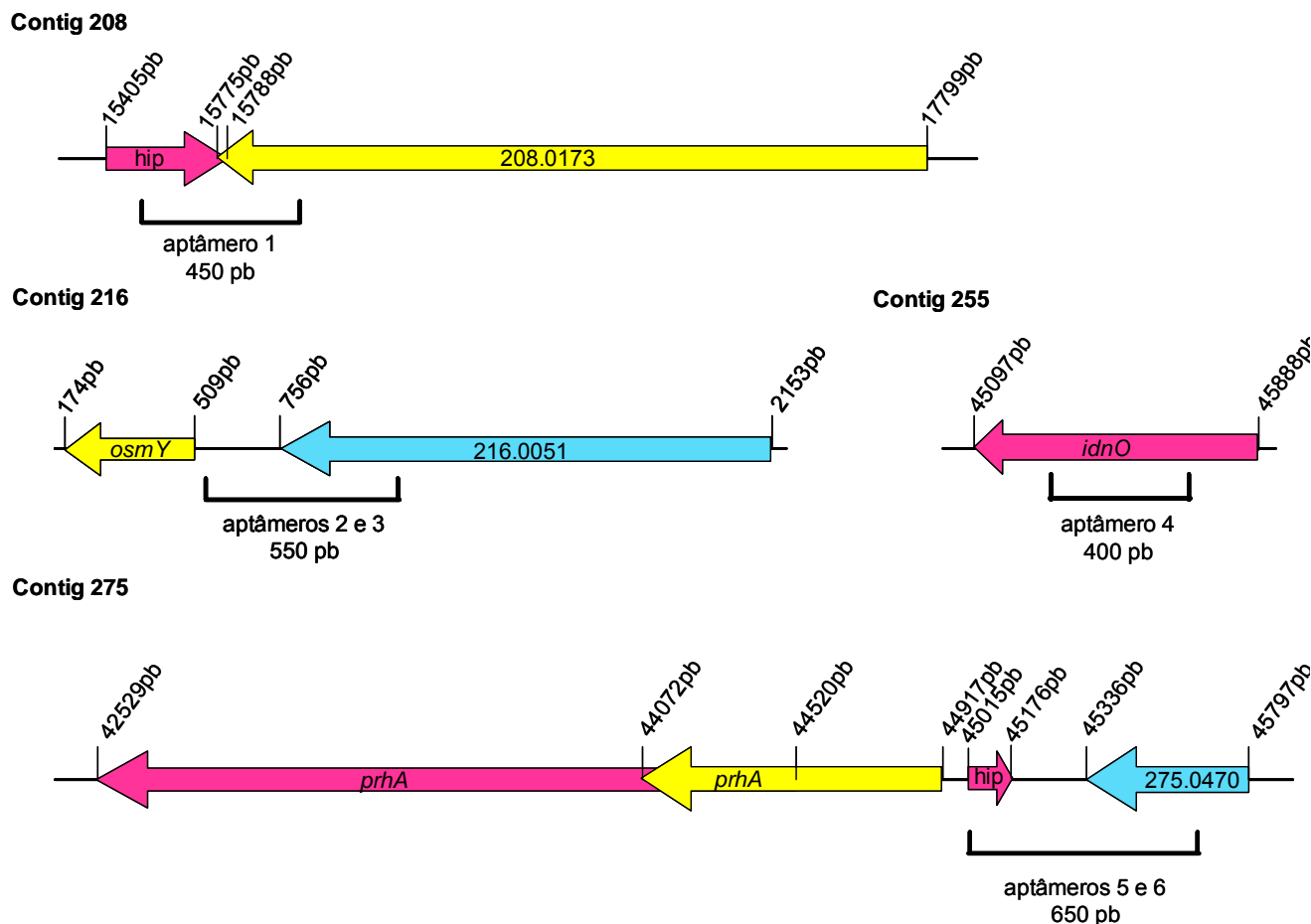


Legenda: Linha 1, Marcador de massa molecular; linhas 2, 3 e 4, produto de PCR das colônias controle (que continham o plasmídio vetor sem inserto); linhas 5 a 25, produto de PCR dos transformantes. Os círculos indicam os produtos de amplificação obtidos, o rosas indicam produtos de aproximadamente 650pb, os amarelos de aproximadamente 550pb, os azuis de aproximadamente 450pb e os verdes de 400pb.

Tabela 4 – Aptâmeros do Domínio HTH da Proteína 216.0019 que Foram Isolados

Aptâmero	Tamanho(pb)	Contig	Valor de e da melhor sequência
216 - 1	450	208	e -50
216 - 2	550	216	e -71
216 - 3	550	216	e -123
216 - 4	400	255	e -134
216 - 5	650	275	0.0
216 - 6	650	275	5e -92

Figura 12 – Aptâmeros do Domínio HTH da Proteína 216.0019 Isolados e sua Localização no Genoma de *H. seropedicae*



Legenda: As direções das setas indicam a direção de transcrição dos genes. As regiões dos aptâmeros estão delimitadas pelas chaves.

O alinhamento das seqüências dos aptâmeros não permitiu a determinação de um sítio de ligação para o domínio HTH da proteína 216.0019, entretanto as regiões genômicas puderam ser localizadas. Como pode ser observado na Figura 12, dois aptâmeros correspondem a regiões codificadoras de proteínas, o que torna improvável a interação funcional da proteína 216.0019 com essas regiões. É possível que tenha ocorrido ligação inespecífica durante a reação de interação DNA-proteína. Novamente, a presença da cauda de histidinas ou a falta dos domínios central e N-terminal da

proteína 216.0019 pode ter provocado uma alteração conformacional do domínio HTH permitindo a ligação não-específica ao DNA. Por outro lado, estas prováveis interações não-específicas podem ter sido favorecidas pelas condições (pH, força e composição iônica, concentração de DNA e proteína) da reação que não coincidem com a condição intracelular.

Quatro aptâmeros isolados foram correspondentes a regiões intergênicas. Dois deles (2 e 3) correspondem à região intergênica entre a ORF 216.0019 e o gene *osmY*, que codifica para uma proteína periplasmática sensora envolvida na regulação osmótica (Yim & Villarejo, 1992). Em *E. coli* a transcrição de *osmY* é dependente do fator sigma RpoS (Yim, Brems & Villarejo, 1994), em contraste com o que aparentemente ocorre em *H. seropedicae*, onde este gene deve ser dependente do fator sigma 54. É interessante notar que a região de DNA isolada corresponde à região a jusante do gene 216.0019.

Os aptâmeros 5 e 6 da proteína 216.0019 correspondem à região intergênica entre a ORF hip (275.0459) e o gene *prhA*. O gene *prhA* codifica para uma proteína transmembrana que provavelmente liga sideróforos. Em *Ralstonia solanacearum* PrhA funciona como um sensor de sinais específicos da planta hospedeira e induz a expressão de genes *hrp* (codificam para o sistema de secreção do tipo III) (Marenda, 1998). Como o sentido da ORF 275.0459 é inverso ao do gene *prhA* é possível que essa ORF hipotética é que seja regulada pela proteína 216.0019 e não *prhA*; ou ambos os genes.

As regiões amontante aos genes *osmY* e *prhA* foram analisadas quanto a presença de sequências consenso para promotores σ^{54} . Foi localizado um provável promotor na região amontante a *osmY* (figura 13), o que indica que provavelmente esse gene seja transcrito pela RNA-polimerase/ σ^{54} . Já amontante ao gene *prhA* foram encontrados dois prováveis promotores do tipo σ^{54} (figura 14), porém não tão conservados quanto o amontante a *osmY*

Figura 13 – Provável Região Promotora do Gene *osmY*

	G G E E A V V A G R Q E E R L *	
1	GGTGGGAAGAGGCTGTCGTCGCGGGGCGACAGGAGGAAAGGCTTTGA	80
1	CCACCCCTTCTCCGACAGCAGCGCCCCGCTGTCCTCCTTTCCGAAACT	80
81	TTTGTCTCCCTCGGCGACAGAAAAGCGGGTTGGGGTGGGGCATTGCAGAGCCCCATTTCTTGGCGAGATCGCCAAGTC	160
81	AAACAGGAGGGAGCCGCTGTCTTTTCGCCAACCCACCCCGTAACGTCTCGGGGTAAAGGAACCGCTCTAGCGGTTTCAG	160
161	ATTGAATTCATTGAGGAAAATTCCTGCTGGCATGGGCTTTGCG	240
161	TAACCTAAGTAACTCCTTTTAAGGACGACCGTACCCGAAACGC	240
	M K K	
241	TGTCCCTATATAGAGCGCTGCCACGATGACTGATCAATAAGGAGATAGACC	300
241	ACAGGGATATATCTCGCGACGGTGCTACTGACTAGTTATTCCTCTATCTGG	300

Legenda: Sequência de nucleotídeos e aminoácidos da região intergênica de *osmY* e a ORF 216.0019. O final da ORF 216.0019 está destacado em rosa, o início de *osmY* destacado em azul e o provável promotor σ^{54} está destacado em roxo.

Figura 14 – Provável Região Promotora do Gene *prhA*

	L M Y A D *	
1	CTGATGTACCGGATTAG	80
1	GACTACATGCGCCTAATC	80
81	GAAAATTTAAATTCACCTGGAGTTGCATGTGAAGTTTGAAAAATAAATTCGATTCCTCTGGATAAACTTTTATATCAAT	160
81	CTTTTAAATTTAAGTGGACCTCAACGTACACTTCAAACTTTTTATTTAAGCTAACGGAGACCTATTTGAAAATATAGTTA	160
161	TCGACACGCTATCAGGCGACTTTCGAGTGGCAGGGAAGTATCA	240
161	AGCTGTGCGATAGTCCGCTGAAAGCTCACCGTCCCTTCATAGT	240
241	ACAGCGTGGAAGTTTCGGTTTTTCGGTGACGGCAAAGATGTGCA	320
241	TGTGCGACCTTCAAGCCAAAAGCCACTGCCGTTTTCTACACGTT	320
321	CATCGGCAAAACGCAAAGAGAAAATCGTTTTTCAGAACTCTGCGATCCGCCTGTTTCAGGCTGCTTTCAATTTGTTTCAGC	400
321	GTAGCCGTTTTGCGTTTTCTCTTTTAGCAAAAGTCTTGAGACGCTAGGCGACAAAGTCCGACGAAAGTTAAACAAGTCG	400
	M K L I S H	
401	AGAGCCAATCTAACATCATTATTAGAAATCAAC	451
401	TCTCGGTTAGATTGTAGTAATAATCTTTAGTTG	451

Legenda: Sequência de nucleotídeos e aminoácidos da região intergênica de *prhA* e a ORF 275.0470. O final da ORF 245.0470 está destacado em rosa, o início de *prhA* destacado em azul e os prováveis promotores σ^{54} estão destacados em roxo.

A identificação dos aptâmeros da proteína ativadora de transcrição a montante de *osmY* e *prhA* sugere uma relação funcional entre os produtos destes genes. Em muitos organismos a expressão dos genes *hrp* é dependente de fatores ambientais como pH, osmolaridade e disponibilidade de carbono e nitrogênio (Arlat, 1992). Considerando que PrhA participe da expressão dos genes *hrp* do sistema de secreção do tipo III de *H. seropedicae*, como ocorre em *R. solanacearum*, é possível especular que as proteínas OsmY e PrhA provavelmente participam de um sistema de sensoriamento envolvido na regulação da transcrição de genes responsáveis pela regulação da expressão do sistema de secreção do tipo III em *H. seropedicae*. Entretanto ensaios de interação DNA-proteína, análise de mutantes e de expressão gênica são necessários para confirmar a funcionalidade dos prováveis regiões regulatórias identificadas.

4.6.2 Aptâmeros Correspondentes ao Domínio HTH da Proteína 208.0173

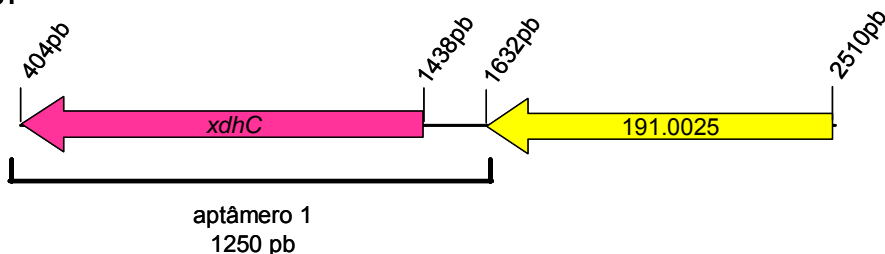
Dois experimentos de interação DNA-proteína seguidos de cromatografia foram realizados para isolar os aptâmeros de DNA do domínio HTH da proteína 208.0173 e nesses experimentos foram isolados aptâmeros de 9 tamanhos diferentes (tabela 5), variando de 400 e 1500 pares de base. Dois clones de cada um dos aptâmeros foram parcialmente seqüenciados e as regiões do genoma de *H. seropedicae* que correspondem a cada aptâmero podem ser visualizadas na figura 15.

Tabela 5 – Aptâmeros do Domínio HTH da Proteína 208.0173 que Foram Isolados

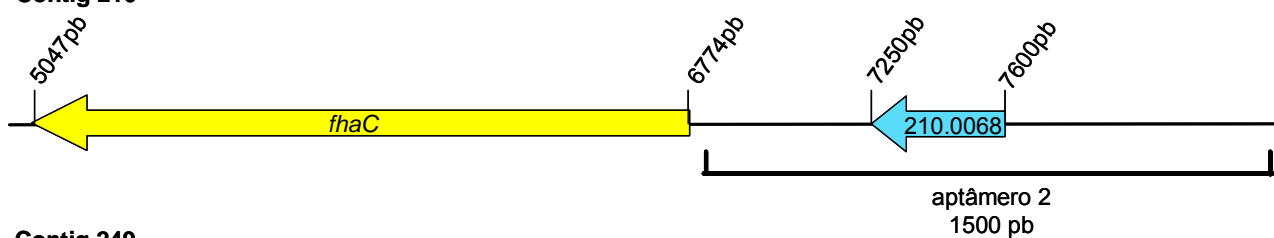
Aptâmero	Tamanho(pb)	Contig	Valor de e da melhor sequência
208 - 1	1250	191	e -146
208- 2	1500	210	0.0
208 - 3	600	249	0.0
208 - 4	400	254	e -108
208 - 5	400	257	e - 70
208 - 6	750	268	e - 149
208 - 7	700	274	e - 152
208 - 8	600	283	e - 101
208 - 9	650	284	e - 107

Figura 15 - Aptâmeros do Domínio HTH da Proteína 208.0173 Isolados e sua Localização no Genoma de *H. seropedicae*

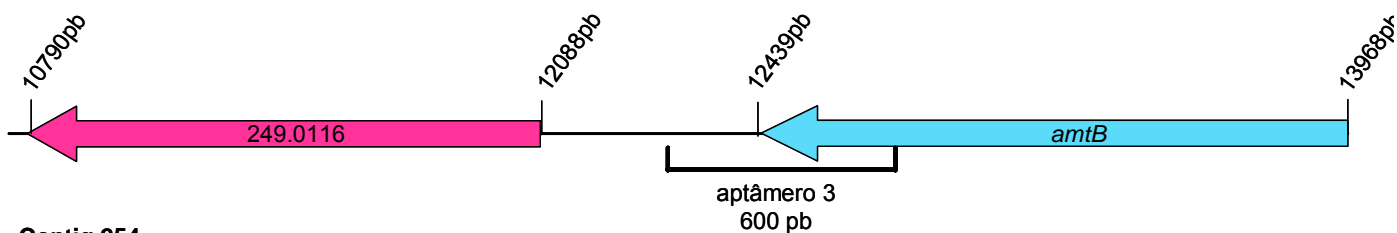
Contig 191



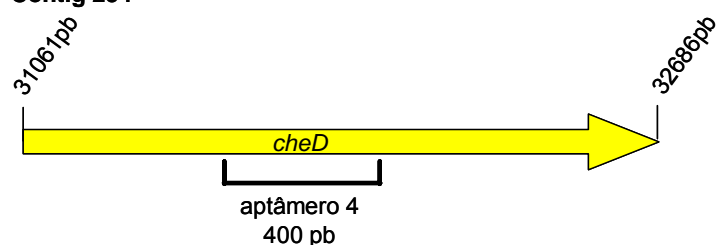
Contig 210

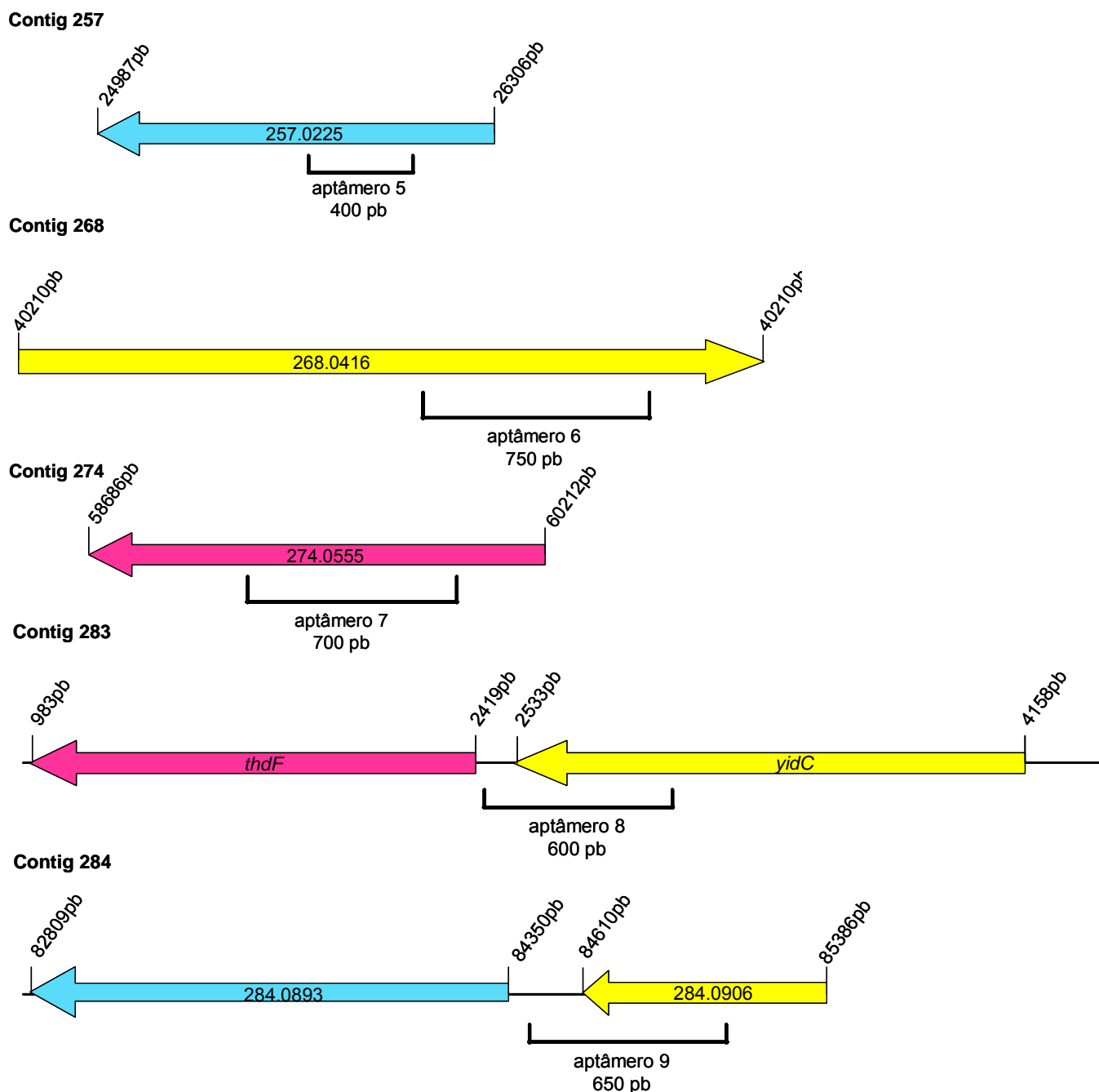


Contig 249



Contig 254





Legenda: As direções das setas indicam a direção de transcrição dos genes. As regiões dos aptâmeros estão delimitadas pelas chaves.

Quatro dos nove aptâmeros obtidos para o domínio HTH da proteína 208.0173 estão localizados em prováveis regiões codificadoras do genoma de *H. seropedicae*. Esse resultado não foi esperado, pois como se trata de uma proteína reguladora de transcrição esperava-se que as seqüências correspondessem a regiões intergênicas, mais especificamente próximo à

região promotora de um gene ou operon. É provável que a recuperação destes aptâmeros tenha sido consequência de ligação inespecífica do domínio HTH da proteína 208.0173-His ao DNA, que pode ocorrer pelo dobramento incorreto do domínio na ausência do restante da proteína e os resíduos de histidina inseridos na porção N-terminal do domínio também podem ter afetado a conformação do domínio ou ainda a alta concentração de DNA e proteína utilizada no ensaio.

Dois dos aptâmeros estão localizados a montante de genes que codificam para proteínas hipotéticas conservadas, mas de função desconhecida. Os experimento de EMSA com essas regiões confirmariam se a ligação foi inespecífica, entretanto para elucidar a função da proteína reguladora de transcrição seria necessário identificar a função de cada um desses genes que codifica para uma proteína hipotética conservada.

Somente três aptâmeros corresponderam a regiões intergênicas a montante de genes já caracterizados. Um deles corresponde à região promotora do gene *thdF* que, em *E. coli*, codifica para uma proteína responsável pela oxidação de tiofeno (Alam & Clark, 1991). O segundo está localizado na região a montante ao gene *xdhC*, que em *Rhodobacter capsulatus* codifica para uma provável chaperona da xantina desidrogenase (Leimkühler & Klipp, 1999). O último aptâmero da proteína 208.0173 correspondente à região a montante ao gene *fhaC*, que em *Bordetella pertussis* codifica para uma subunidade da proteína FHA que é uma hemaglutinina filamentosa envolvida na interação com células do hospedeiro (Willems, 1994).

Esses três genes possivelmente regulados pela proteína 208.0173 não parecem possuir relação funcional entre si. Novos experimentos para caracterização da interação entre os fragmentos de DNA identificados e a proteína 208.0173 e análise de expressão gênica são necessários para confirmação da funcionalidade destas regiões e dos genes regulados e sua função no metabolismo de *H. seropedicae*.

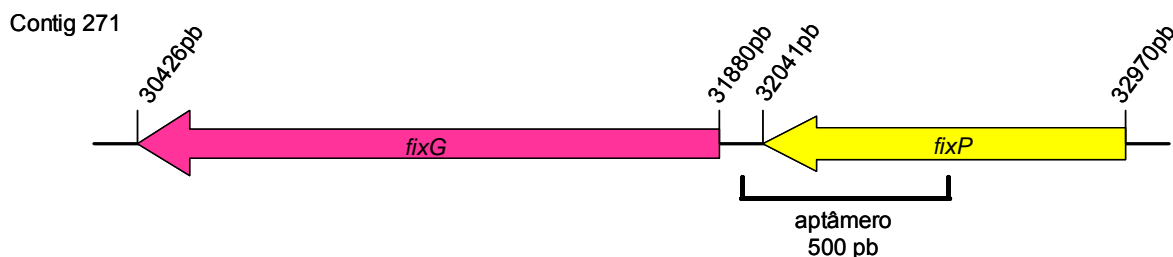
4.6.3 Aptâmeros Correspondentes ao Domínio HTH da Proteína 287.0556

Três diferentes aptâmeros foram identificados para o domínio HTH da proteína 287.0556 em 120 colônias obtidas. O tamanho desses aptâmeros variou de 300 a 600 pares de base. Os insertos de pelo menos dois clones

para cada tamanho de aptâmero foi sequenciado, mas só foram obtidas seqüências com qualidade necessária para análise para o aptâmero de 500 pares de base.

O aptâmero de 500 pares de base corresponde à região do genoma de *H. seropedicae* entre os genes *fixP* e *fixG*, no contig 271 (montagem de 18/08/2004) (Figura 16).

Figura 16 - Aptâmeros do Domínio HTH da Proteína 287.0556 Isolados e sua Localização no Genoma de *H. seropedicae*



Legenda: As direções das setas indicam a direção de transcrição dos genes. As regiões dos aptâmeros estão delimitadas pelas chaves.

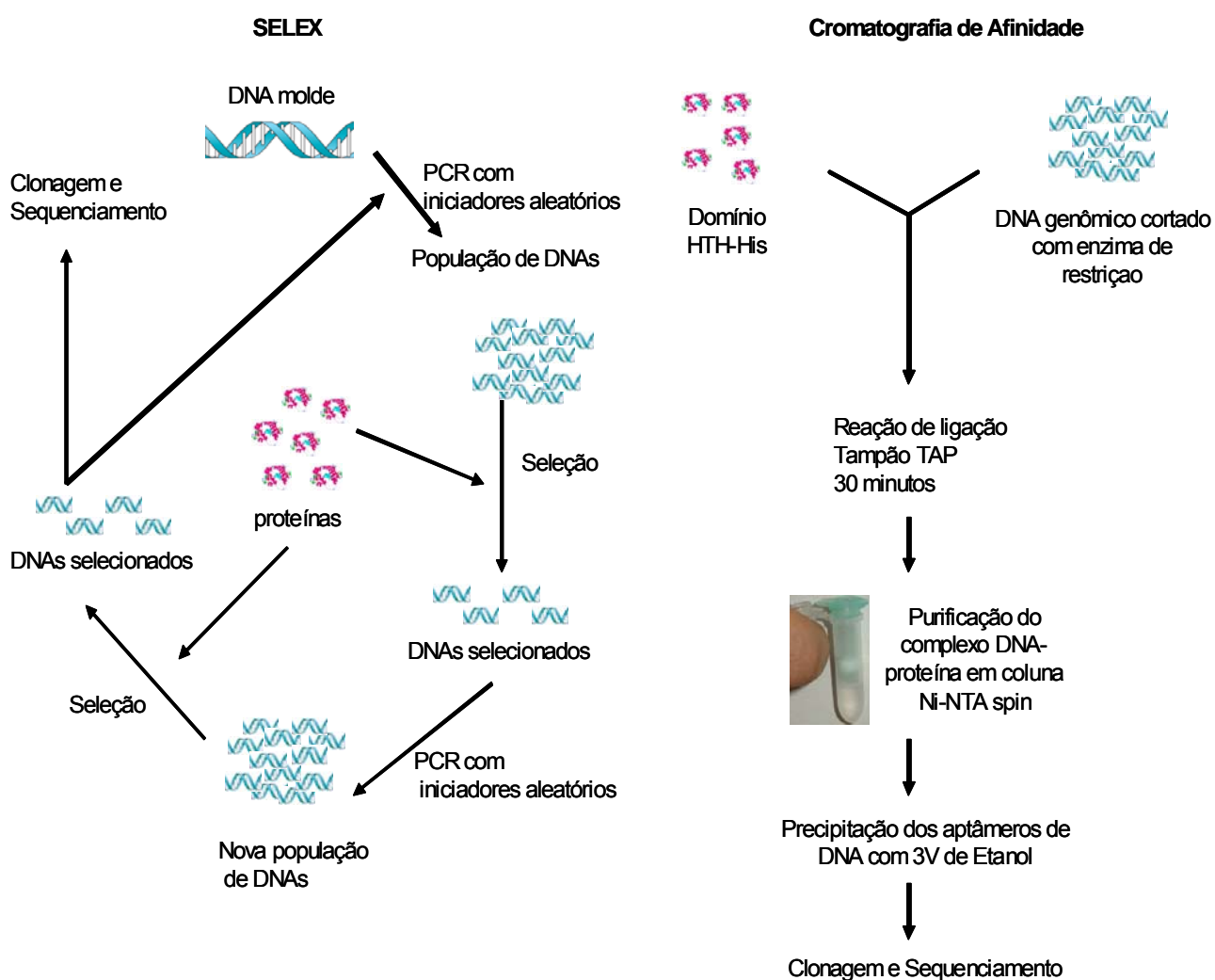
Os genes *fix* estão envolvidos com a fixação biológica de nitrogênio. O gene *fixP* codifica para a subunidade III da citocromo oxidase *cbb3*. FixP é uma proteína transmembrana indispensável para a fixação de nitrogênio em *Rhizobium etli* (Lopes et al., 2001). Já *fixG* foi descrito como gene que codifica para uma ferredoxina transmembrana que contém um grupo 4Fe-4S e provavelmente é parte de uma bomba de cátions que influencia a fixação simbiótica de nitrogênio em *Rhizobium meliloti* (Kahn et al., 1989).

A identificação de um aptâmero da proteína 287.0556 localizado a montante de *fixG* sugere que essa proteína A expressão do gene *fixG*. A proteína 287.0556 é semelhante à proteína NifA (62% de identidade com NifA de *Frankia*), mas não possui o domínio N-terminal receptor de sinal. Os resultados obtidos sugerem que essa proteína pode estar envolvida na regulação da expressão de oxidase terminal e da fixação de nitrogênio em *H. seropedicae*.

4.6.4 Comparação entre os Diversos Métodos de Isolamento de Aptâmeros de DNA

O método para isolamento de aptâmeros de DNA desenvolvido neste trabalho é inédito e os resultados apresentados sugerem que é uma opção eficiente para a identificação de prováveis seqüências regulatórias no DNA de procariotos. Entretanto, é importante frisar que é um método de triagem, assim como a SELEX (*Systematic Evolution of Ligands by Exponential Enrichment*) (Tuerk & Gold, 1990 e Ellington & Szostak, 1990), requerendo experimentos de modificação de mobilidade eletroforética de DNA (*electrophoretic mobility shift assay* ou EMSA) e *footprinting* de DNA para confirmação da interação DNA-proteína e determinação do sítio de ligação da proteína no fragmento de DNA. Este novo método de isolamento de aptâmeros de DNA requer menos etapas e é mais simples do que a SELEX (Figura 17).

Figura 17 – Esquema Comparativo dos Métodos SELEX e de Cromatografia de Afinidade para Isolamento de Aptâmeros de DNA



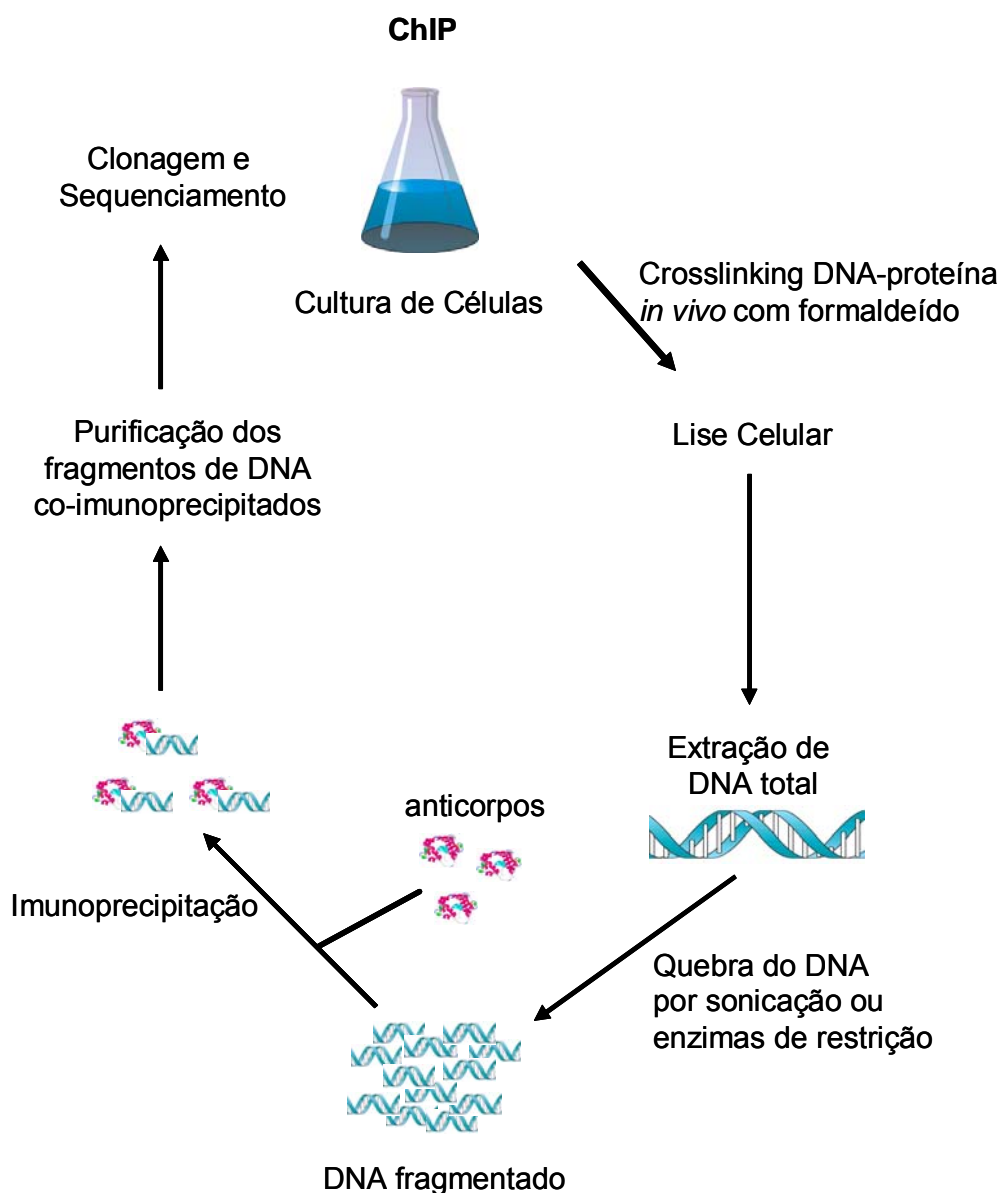
O isolamento de aptâmeros por ambos métodos requer a proteína, ou o seu domínio de ligação ao DNA, purificada. Entretanto, para SELEX a proteína pode ser nativa, enquanto que o método descrito neste trabalho requer que a proteína tenha capacidade de se ligar a uma matriz sólida. Uma grande diferença entre as duas técnicas é que na SELEX são utilizados oligonucleotídeos sintéticos e no método por cromatografia de afinidade é utilizado o DNA genômico parcialmente cortado com uma enzima de restrição que será utilizado para a formação dos complexos DNA-proteína; isso faz com que na SELEX sejam obtidos sempre aptâmeros do mesmo tamanho (cerca de 100pb) enquanto no método descrito neste trabalho os aptâmeros obtidos são de tamanhos variados (200 – 1500pb). A variabilidade do tamanho dos aptâmeros isolados dificulta o alinhamento das seqüências para determinar o sítio de ligação da proteína. Apesar disto, em ambos métodos outros experimentos devem ser realizados para a determinação inequívoca dos sítios. A técnica SELEX foi utilizada por Freede & Brantl (2004) para identificação de aptâmeros da proteína CopR, que foram confirmados por ensaios de modificação de mobilidade eletroforética de DNA, e por Ferrières (2004) que confirmou o sítio de ligação da proteína FixJ de *Sinorhizobium meliloti* utilizando mutantes sítio dirigidos.

Schultzaberger & Schneider (1998) fizeram uma comparação o método SELEX de identificação de aptâmeros e métodos *in vivo* (*footprinting in vivo* e mutações sítio específicas no DNA) e concluíram que os resultados obtidos por SELEX não corroboravam os resultados obtidos *in vivo*. Os autores discutem que esse resultado deve ter sido devido a condições inadequadas de interação entre DNA-proteína *in vitro* (alta concentração de sal, pH incorreto) que não mimetizam fielmente a situação *in vivo* e, portanto, ocasionam a produção de complexo DNA-proteína não nativo. Foi aventado também que as etapas de amplificação por PCR para aumentar o número e variabilidade dos aptâmeros podem gerar fragmentos de DNA com seqüências em que a proteína pode ligar, mas que não correspondem aos reais sítios de ligação *in vivo*.

Outro método de isolamento de aptâmeros de DNA largamente difundido é a imunoprecipitação de cromatina (ChIP) (Boyd et al, 1998 e Boyd & Farnham, 1997). É um método comumente utilizado para eucariotos e muito pouco utilizado para procariotos. Nesta metodologia a proteína de interesse não

precisa estar purificada, entretanto é necessário que se tenha um anticorpo contra esta proteína. A proteína é ligada covalentemente ao DNA *in vivo*, o complexo DNA-proteína é isolado por imunoprecipitação e o DNA ligado é finalmente isolado e seqüenciado (Figura 18) (Boyd et al, 1998 e Boyd & Farnham, 1997).

Figura 18 – Esquema do Isolamento de Aptâmeros por Imunoprecipitação de Cromatina - CHIP



O método de imunoprecipitação gera aptâmeros de diversos tamanhos e muitas vezes grandes, assim como a técnica de cromatografia de afinidade descrita neste trabalho. Aptâmeros muito grandes dificultam a determinação de

sítios de ligação, tornando necessários experimentos complementares (EMSA, *footprinting*) para determinação da localização exata dos sítios de ligação. Além disso, a CHIP é uma metodologia com muitas etapas, o que pode levar a uma menor eficiência. Os anticorpos devem ser bastante específicos para que não ocorra a precipitação de fragmentos de DNA não relacionados à proteína de interesse.

Existe ainda uma metodologia que, diferente dos métodos anteriores, não é de triagem e possibilita a determinação dos sítios de ligação da proteína mais facilmente. Neste método a proteína que liga DNA é fundida ao domínio hidrolítico de uma enzima de restrição (*FokI*). Desta forma o DNA será cortado nas regiões em que a proteína de interesse se liga e os fragmentos obtidos podem ser clonados e sequenciados (Ruminy et al, 2001). Entretanto a grande dificuldade dessa técnica é ter acesso aos vetores que possibilitam a formação dessa proteína quimera, bem como a metodologia utilizada para a construção desses vetores.

A grande vantagem do método desenvolvido aqui em relação a SELEX e CHIP é a facilidade, pois os aptâmeros são isolados após a reação de interação DNA-proteína com apenas um passo cromatográfico e podem ser clonados e seqüenciados. A SELEX requer várias reações de amplificação por PCR; que podem levar a formação de aptâmeros diferentes daqueles obtidos *in vivo* (Schultzaberger & Schneider, 1998) e a CHIP requer anticorpos muito específicos e eficientes para que não ocorra a precipitação de fragmentos de DNA indesejados.

5 CONCLUSÕES

A análise das proteínas reguladoras de transcrição de *H. seropedicae* identificou 444 ORFs que provavelmente estão envolvidas em transcrição. Destas, 169 são proteínas reguladoras de transcrição e possuem algum tipo de domínio de interação com DNA. Dentre estas proteínas reguladoras de transcrição foram identificadas 11 que possuem domínios característicos de proteínas ativadoras de transcrição dependentes do fator sigma 54 da RNA-polimerase. Três dessas proteínas (208.0173, 216.0019 e 287.0556) foram selecionadas para identificar as suas respectivas seqüências de reconhecimento e seus genes alvos.

Um novo método para triagem de aptâmeros de DNA foi desenvolvido que permitiu o isolamento e identificação de aptâmeros para as três proteínas purificadas.

Os aptâmeros obtidos para o domínio HTH da proteína 208.0173 de *H. seropedicae* sugerem que esta proteína controla a expressão dos genes *thdF*, *fhaC* e *xdhC*.

Os aptâmeros obtidos para o domínio HTH da proteína 216.0019 de *H. seropedicae* sugerem que provavelmente essa proteína esteja envolvida na regulação de genes *osmY* e o gene *prhA*.

Três diferentes aptâmeros foram isolados para o domínio HTH da proteína 287.0556 de *H. seropedicae*, mas apenas um destes foi caracterizado. A localização deste aptâmero sugere que a proteína 287.0556 participa da regulação da expressão do gene *fixG* neste organismo.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABRIL M.A.; BUCK M.; RAMOS J.L. Activation of the *Pseudomonas* TOL plasmid upper pathway operon. Identification of binding sites for the positive regulator XylR and for integration host factor protein. **J Biol Chem.** v.266 (24), p.15832-8, 1991.

AIBA, H.; NAKASAI, F.; MIZUSHIMA, S. e MIZUNO, T. Evidence for the physiological importance of the phosphotransfer between the two regulatory components, EnvZ and OmpR, in osmoregulation in *Escherichia coli*. **J. Biol. Chem.** v.264, p.14090-14094, 1989.

ALAM KY, CLARK DP. Molecular cloning and sequence of the *thdF* gene, which is involved in thiophene and furan oxidation by *Escherichia coli*. **J Bacteriol.** v.173 (19), p.6018-24, 1991.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. Basic local alignment search tool. **J. Mol. Biol.** 215:403-410, 1990.

ARLAT, M., GOUGH, C.L., ZISCHEK, C., BARBERIS, P.A., TRIGALET, A., AND BOUCHER, C.A. Transcriptional organization and expression of the large *hrp* gene cluster of *Pseudomonas solanacearum*. **Mol Plant–Microbe Interact** v.5, p.187–193, 1992.

D'AUTREAUX B, TUCKER NP, DIXON R, SPIRO S. A non-haem iron centre in the transcription factor NorR senses nitric oxide. **Nature** v.437(7059), p.769-72, 2005.

BALDANI, J.I.; BALDANI, V.L.D.; SELDIN, L.; DOBEREINER, J. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen fixing bacterium. **Int. J. Syst. Bacteriol.**, Washington, v. 36, p. 86-93, 1986.

BALDANI; V. L. D.; OLIVARES, F; DÖBEREINER, J. Selection of *Herbaspirillum spp.* strains associated with rice seedlings amended with ¹⁵N-labelled fertilizer. In: **International Symposium on Sustainable Agriculture for the Tropics – The Role of Biological Nitrogen Fixation**, Angra dos Reis, p.202-203, 1995.

BATES, P.A. AND STERNBERG, M.J.E. Model Building by Comparison at CASP3: Using Expert Knowledge and Computer Automation. **Proteins: Structure, Function and Genetics**, v. 3:47-54, 1999.

BATEMAN, A. The SIS domain: a phosphosugar-binding domain. **Trends Biochem Sci** v.24, p.94-95, 1999.

BATEMAN, A.; COIN, L. DURBIN, R. FINN, R. D.; HOLLICH, V. GRIFFITHS-JONES, S.; KHANNA, A.; MARSHALL, M. MOXON, S. SONNHAMMER, E. L. L.; STUDHOLME, D.J.; YEATS, C. e EDDY, S.R. The Pfam Protein Families Database. **Nuc. Acid. Res.** v.32, database issue, 2004.

BIRKMANN, A.; SAWERS, R.G. & BOCK, A. Involvement of the *ntrA* gene product in the anaerobic metabolism of *Escherichia coli*. **Mol. Gen. Genet.** v.210, p. 535-542, 1987.

BJELLQVIST, B.; HUGHES, G.J.; PASQUALI, CH.; PAQUET, N.; RAVIER, F.; SANCHEZ, J.C.H.; FRUTIGER, S.; HOCHSTRASSER, D.F. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. **Electrophoresis**, v.14, p.1023-1031, 1993.

BJELLQVIST, B.; BASSE, B.; OLSEN, E.; CELIS, J.E. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. **Electrophoresis**, v.15, p.529-539, 1994.

BÖHM, A. & BOOS, W. Gene regulation in procariotes by subcellular relocalization of transcription factors. **Curr. Op. Microbiol.** v.7, p.151-156, 2004.

BOYD, K.E. & FARNHAM, P.J. Myc versus USF: Discrimination at the *cad* gene is determined by core promoter elements. **Mol. Cell Biol.** V.17(5), p. 2529-2537, 1997.

BOYD, K.E.; WELLS, J.; GUTMAN, J. BARTLEY, S.M. e FARNHAM, P.J. c-Myc target gene specificity is determined by a post-DNA-binding mechanism. **Proc. Natl. Acad. Sci. USA**, v.95, p. 13887-13892, 1998.

BRADFORD, M.M. A rapid and sensitive method for the quantification of micrograms quantities of protein utilization the principle of protein-dye binding. **Anal. Biochem.**,v. 72, 248-254, 1976.

BRENNAN RG, MATTHEWS BW. Structural basis of DNA-protein recognition. **Trends Biochem Sci.** v.14(7), p.286-90, 1989.

BROWNING, D.F. & BUSBY, S.J.W. The regulation of bacterial transcription initiation. **Nature Rev. Microbiol.** v.2, p.1-9, 2004.

BUCK, M.; MILLER, S.; DRUMMOND, M. e DIXON, R. Upstream activator sequences are present in the promoters of nitrogen fixation genes. **Nature**, v.320, p. 374-378, 1986.

BUCK, M.; GALLEGOS, M.T.; STUDHOLME D.J.; GUO, Y. e GRALLA, J.D. The bacterial enhancer-dependent σ^{54} (σ^N) transcription factor. **J. Bacteriol.** v.182, p. 4129-4136, 2000.

BUSBY, S. & EBRIGHT, R.H. Transcription activation at class II CAP-dependent promoters. **Mol. Microbiol.** v.23, p. 853-859, 1997.

DEBARBOUILLE M, MARTIN-VERSTRAETE I, KLIER A, RAPOPORT G. The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to

both sigma 54- and phosphotransferase system-dependent regulators. **Proc Natl Acad Sci U S A.** v.88(6), p.2212-6, 1991.

DIXON R, KENNEDY C, KONDOROSI A, KRISHNAPILLAI V, MERRICK M. Complementation analysis of *Klebsiella pneumoniae* mutants defective in nitrogen fixation. **Mol Gen Genet.** v.157(2), p.189-981977.

DIXON, R.A. The genetic complexity of nitrogen fixation. **J. Gen. Microbiol.** v.130, p. 2745-2755, 1984.

DIXON, R. The *xyIABC* promoter from the *Pseudomonas putida* TOL plasmid is activated by nitrogen regulatory genes in *Escherichia coli*. **Mol. Gen. Genet.** v. 203, p. 129-136, 1986.

DOBEREINER, J. History and new perspectives of diazotrophs in association with non leguminous plants. **Symbiosis**, Rehovot, v. 13, p. 1-13, 1992.

DOVE, S.L.; DARST, S.A. & HOCHSCHILD, A. Region 4 of σ as a target of transcription regulation. **Mol. Microbiol.** v.48, p. 863-874, 2003.

EBRIGHT, R.H. Transcription activation at class I CAP-dependent promoters. **Mol. Microbiol.** v.8, p. 797-802, 1993.

ELLINGTON, A.D. & SZOSTAK, J.W. *In vitro* selection of RNA molecules that bind specific ligands. **Nature** v. 346, p.818-822, 1990.

ENDO H, HABE H, NOJIRI H, YAMANE H, OMORI T. The sigma54-dependent transcriptional activator SfnR regulates the expression of the *Pseudomonas putida* *sfnFG* operon responsible for dimethyl sulphone utilization. **Mol Microbiol.** v.55(3), p.897-911, 2005.

ERIE, D. A. The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. **Biochim. Biophys. Acta** v.1577, p.224-39, 2002.

FAVETTI, A.F.J.; FRIEDERMAN, G.R.; CRUZ, L.M.; SOUZA, E.M.; MONTEIRO, R.A.; WASSEM, R.; RAMOS, H.J.O.; CHUBATSU, L.S.; RIGO, L.U.; STEFFENS, M.B.R.; PEDROSA, F.O. GAAT - a bioinformatic tool for automated genomic assembly. **in: XXXII Reunião Anual da Sociedade Brasileira de Bioquímica**, Caxambu, p.262, 2003.

FERRIÈRES, L.; FRANCEZ-CHARLOT, A.; GOUZY, J.; ROUILLÉ, S.; KAHN, D. FixJ-regulated genes evolved through promoter duplication in *Sinorhizobium meliloti* **Microbiol.** v.150, p.2335-45, 2004.

FREEDE P. & BRANTL S. Transcriptional Repressor CopR: use of SELEX to study the *copR* operator indicates that evolution was directed at maximal binding affinity. **J Bacteriol.** v.186(18), p.6254-64, 2004.

GAILLARDIN, C.M. & MAGASANIK, B. Involvement of the product of the *glnF* gene in the autogenous regulation of glutamine synthetase formation in *Klebsiella aerogenes*. **J. Bacteriol.** v. 1333, p. 1329-1338, 1978.

GARCIA, E.; BANCROFT, S.; RHEE, S.G. e KUSTU, S. The product of a newly identified gene, *glnF*, is required for synthesis of glutamine synthetase in *Salmonella* **Proc. Natl. Acad. Sci. USA** v. 74(4) p. 1662-1666, 1977.

GARDNER, A.M.; HELMICK, R.A. & GARDNER, P.R. Flavorubredoxin, an inducible catalyst for nitric oxide reduction and detoxification in *Escherichia coli* **J. Biol. Chem.** v.277, p.8172-8177, 2002.

GRKOVIC S., BROWN M.H., SKURRAY R.A. Regulation of bacterial drug export systems. **Microbiol. Mol. Biol. Rev.** v.66, p.671-701, 2002.

HAKENBECK, R. & STOCK, J.B. Analysis of two-component signal transduction systems involved in transcription regulation. **Met. Enzym.** v.273, p.281-301, 1996.

HESS, J.F.; BOURRET, R.B. e SIMON, M.I. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. **Nature** v.336, p. 139-143, 1988.

HIGGINS,D.; THOMPSON, J.; GIBSON,T.; THOMPSON,J.D.; HIGGINS,D.G.; GIBSON T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nuc. Ac. Res.** v.22, p.4673-4680, 1994.

HIRSHMAN, J.; WONG, P.K.; SEI, K.; KEENER, J. e KUSTU, S. Products of nitrogen regulatory genes *ntxA* and *ntxC* of enteric bacteria activate *glnA* transcription *in vitro*: Evidence that the *ntxA* product is a σ factor. **Proc. Natl. Acad. Sci. USA**, v. 82, p. 7525-7529, 1985.

HSU, L.; VO, N.V.; KANE, C.M.; CHAMBERLIN, M.J. In Vitro Studies of Transcripts Initiation by *Escherichia coli* RNA Polymerase. 1.RNA Chain Initiation, Abortive Initiation, and Promoter Escape at Three Bacteriophage Promoters. **Biochemistry**, v.42, p.3777-3786, 2003.

HUNT, T.P. & MAGASANIK, B. Transcription of *glnA* by purified *Escherichia coli* components: Core RNA polymerase and the products of *glnF*, *glnG* and *glnL*. **Proc. Natl. Acad. Sci. USA** v. 82, p. 8453-8457, 1985.

HUTCHINGS, M.I.; MANDHANA, N. & SPIRO, S. The NorR protein of *Escherichia coli* activates expression of the flavin monooxygenase gene *norV* in response to reactive nitrogen species **J. Bacteriol.** v.184, p.4640-4643,2002.

IGO, M.M.; NINFA, A.J.; STOCK, J.B. e SILHAVY, T.J. Phosphorylation and dephosphorylation of a bacterial transcription activator by transmembrane receptor. **Genes Dev.** v.3, p. 1725-1734, 1989.

ISHIHAMA, A. Molecular Assembly and functional modulation of *Escherichia coli* RNA polymerase. **Adv. Biophys.** V.26, p. 19-31, 1990.

JONES S, VAN HEYNINGEN P, BERMAN HM, THORNTON JM. Protein-DNA interactions: A structural analysis. **J Mol Biol.** v.287(5), p.877-96, 1999.

JOVANOVIC G, WEINER L, MODEL P. Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the *Escherichia coli* stress-induced psp operon. **J Bacteriol.** v.178(7), p.1936-45, 1996.

KAHN, D., DAVID, M., DOMERGUE O., DAVERAN, M-L, GHAI, J.,HIRSH, P. R., BATUT, J. Rhizobium meliloti fixGHI Sequence Predicts Involvement of a Specific Cation Pump in Symbiotic Nitrogen Fixation **J Bacteriol**, v.171(2), p. 929-939, 1989.

KEENER, J. & KUSTU, S. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NtrB and NtrC of enteric bacteria: roles of the conserved amino-terminal domain of NtrC. **Proc. Natl. Acad. Sci. USA** v.85, p.4976-4980, 1988.

KLASSEN,G.; PEDROSA, F.O.; SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U. Effect of nitrogen compounds on nitrogenase activity in *Herbaspirillum seropedicae* **Can. J. Microbiol.** V.43, p887-891, 1997.

KOFOID, E.C. & PARKINSON, J.S. Transmitter and receiver modules in bacterial signaling proteins. **Proc. Natl. Acad. Sci. USA** v.85, p.4981-4985, 1988.

KOCHER, T. D.; WILSON, A. C. DNA amplification by the polymerase chain reaction. In BROWN, T. A., **Essential Molecular Biology: A Practical Approach**, IRL Press, Oxford University Press, Oxford, p.185-208, 1991.

KUMAR, S. A. The Structure and Mechanism of Action of Bacterial DNA-Dependent RNA polymerase. **Prog. Biophys. Molec. Biol.** V.38, p.163-210, 1981.

KUSTU, S.G.; McFARLAND, N.C.; HUI, S.P.; ESMON, B. e AMES, G.F. Nitrogen control in *Salmonella typhimurium*: co-regulation of synthesis of glutamine synthetase and amino acid transport systems **J. Bacteriol.** v.138(1), p. 218-234,1979.

KUSTU, S.; NORTH,A.K. & WEISS, D.S. Prokaryotic transcriptional enhancers and enhancer binding proteins. **Trends Biochem. Sci.** v.16, p.397-402, 1985.

KUSTU, S.; SANTERO, E.; KEENER, J.; POPHAM, D. e WEISS, D. Expression of σ^{54} (ntrA)-dependent genes is probably united by a common mechanism. **Microbiol. Rev.** V.53(3), p. 367-376, 1989.

LAEMMLI, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T7. **Nature**, 277, 680-685, 1970.

LAMBERT C, LEONARD N, DE BOLLE X, DEPIEREUX E. ESyPred3D: Prediction of proteins 3D structures. **Bioinformatics.** v.18(9), p.1250-1256, 2002.

LEIMKÜHLER, S. & KLIPP, W. Role of XDHC in Molybdenum Cofactor Insertion into Xanthine Dehydrogenase of *Rhodobacter capsulatus* **J Bacteriol.** v.181(9), p.2745–2751, 1999.

LEONHARTSBERGER S, HUBER A, LOTTSPREICH F, BOCK A. The *hydH/G* Genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. **J Mol Biol.** v.307(1), p.93-105, 2001.

LETUNIC, I.; GOODSTADT, L.; DICKENS, N.J.; DOERKS, T.; SCHULTZ, J.; MOTT, R.; CICCARELLI, F.; COPLEY, R.R.; PONTING, C.P.; BORK, P. Recent improvements to the SMART domain-based sequence annotation resource. **Nucleic Acids Res.** v.30, p.242-244, 2002.

LEWIN, B. **Genes VII**, Cambridge, Oxford University Press, 2000.

LOPEZ O, MORERA C, MIRANDA-RIOS J, GIRARD L, ROMERO D, SOBERON M. Regulation of gene expression in response to oxygen in *Rhizobium etli*: role of FnrN in fixNOQP expression and in symbiotic nitrogen fixation. **J Bacteriol.**, v.183(24), p.6999-7006, 2001.

LUKAT, G.S.; McCLEARY, W.R.; STOCK, A.M. & STOCK, J.B. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors **Proc. Natl. Acad. Sci. USA** v.89, p.718-722, 1992.

LUND O., NIELSEN M., LUNDEGAARD C., WORNING P. CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. **resumo na conferência CASP5** , A102, 2002.

MARENDA M, BRITO B, CALLARD D, GENIN S, BARBERIS P, BOUCHER C, ARLAT M. PrhA controls a novel regulatory pathway required for the specific induction of *Ralstonia solanacearum* *hrp* genes in the presence of plant cells. **Mol. Microbiol.** v.27(2), p.437-53, 1998.

MARTINEZ-ANTONIO,A. & COLLADO-VIDES, J. Identifying global regulators in transcriptional regulatory networks in bacteria. **Curr. Op. Microbiol.** v.6, p.482-489, 2003.

MARTINEZ-HACKERT E, STOCK AM. The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. **Structure**, v.5(1) p.109-24, 1997.

METTKE, I.; FIEDLER, U. & WEISS, V. Mechanism of activation of a response regulator: interaction of NtrC-P dimers induces ATPase activity. **J. Microbiol.** v.177, p.5056-5061, 1995.

MC CLURE, W. R. Mechanism and Control of Transcription Initiation in Prokaryotes. **Ann. Rev. Biochem.** V.54, p.171-204, 1985.

MONTEIRO, R.A.; SOUZA, E.M.; WASSEM, R.; YATES, M.G.; PEDROSA, F.O. & CHUBATSU, L.S. Inter-domain cross-talk controls the NifA protein activity of *Herbaspirillum seropedicae*. **FEBS Let.** v.508, p.1-4, 2001.

MORETT, E. & BUCK, M. *In vivo* studies on the interaction of RNA polymerase- σ^{54} with the *Klebsiella pneumoniae* and *Rhizobium meliloti* *nifH* promoters. **J. Mol. Biol.** v.210, p.65-77, 1989.

MORETT, E. & SEGOVIA, L. The sigma 54 bacterial enhancer-binding protein family: mechanisms of action and phylogenetic relationship of their functional domains. **J. Bacteriol** v.175, p.6067-6074, 1993.

NINFA, A.J. & BENNETT, R.L. Identification of the site of autophosphorylation of the bacterial protein kinase/phosphatase NRII. **J. Biol. Chem.** v.266, p. 6888-6893, 1991.

NINFA, A.J. & MAGANSANIK, B. Covalent modification *glnG* product, NRI by *glnL* product NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli* **Proc. Natl. Acad. Sci. USA** , 1986.

NUDLER, E.; AVETISSOVA, E.; MARKOVTSOV, V.; GOLDFARB, A. Transcription Processivity: Protein-DNA Interactions Holding Together the Elongation Complex. **Science**, v.273, p.211-217, 1996.

OLIVARES, F.L; DOS REIS JR, F. B.; REIS, V. M.; BALDANI, J. I.; DÖBEREINER, J. Infection of sugarcane roots by the endophytic diazotrophs *Herbaspirillum seropedicae* and *H. rubrisubalbicans*. In: . In: **International Symposium on Sustainable Agriculture for the Tropics – The Role of Biological Nitrogen Fixation**, Angra dos Reis, p. 65-66,1995.

OW, D. W. e AUSUBEL, F.M. Regulation of nitrogen metabolism genes by *nifA* gene product in *Klebsiella pneumoniae* **Nature** v. 301, p.307-313, 1983.

PARKINSON, J.S. & KOFOID, E.C. Communication modules in bacterial signaling proteins **Annu. Rev. Genet.** v.26, p.71-112, 1992.

PARKINSON, J.S. Signal transduction schemes of bacteria **Cell** v.73, p.857-871, 1993.

PÉREZ-RUEDA, E. & COLLADO-VIDES, J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K12. **Nuc. Ac. Res.** v.28(8), p.1838-1847, 2000.

PERSUHN, D.C.; SOUZA, E.M.; STEFFENS, M.B.; PEDROSA, F.O.; YATES, M.G. & RIGO, L.U. The transcriptional activator NtrC controls the expression activity of glutamine synthetase in *Herbaspirillum seropedicae*. **FEMS Microbiol.** v. 192, p.217-221, 2000.

PLATT, T. & RICHARDSON, J.P. Escherichia coli Rho Factor: Protein and Enzyme of Transcription Termination in MCKNIGHT, S.L. & YAMAMOTO, K.R. **Transcriptional Regulation**, v.1, p.365-388, 1992.

PLUMBRIDGE, J. Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. **Curr. Opin. Microbiol.**v.5, p.187-193,2002.

POPHAM, D.L.; SZETO, D.; KEENER, J. e KUSTU, S. Function of a bacterial activator protein that binds to transcriptional enhancers. **Science** v.243, p.629-635, 1989.

PTASHNE, M. & GANN, A. Transcription activation by recruitment. **Nature**, v.386, p.569-577, 1997.

RAPPAS M, SCHUMACHER J, BEURON F, NIWA H, BORDES P, WIGNESHWERARAJ S, KEETCH CA, ROBINSON CV, BUCK M, ZHANG X. Structural insights into the activity of enhancer-binding proteins. **Science** v. 307(5717), p.1972-5, 2005.

REITZER LJ, MOVASAS B, MAGASANIK B. Activation of *glnA* transcription by nitrogen regulator I (NRI)-phosphate in *Escherichia coli*: evidence for a long-range physical interaction between NRI-phosphate and RNA polymerase. **J Bacteriol.** v.171(10), p.5512-22, 1989.

RICHARDSON, J.P. Rho dependent transcription termination **Biochim. Biophys. Acta**, v. 1048, p. 127-138, 1990.

RHODIUS, V.A. & BUSBY, S.J.W. Positive activation of gene expression **Cur. Op. Microbiol.** v.1, p. 152-159, 1998.

RIPPE K, GUTHOLD M, VON HIPPEL PH, BUSTAMANTE C. Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of *E. coli* RNA polymerase \times sigma 54 holoenzyme by scanning force microscopy. **J Mol Biol.** v.270(2), p.125-38, 1997.

ROJO, F. Mechanisms of transcriptional repression. **Curr. Op. Microbiol.** V.4, p. 145-151, 2001.

RONCATO-MACCARI, L.D.; RAMOS, R.J.O.; PEDROSA, F.O.; ALQUINI, Y.; CHUBATSU, L.S.; YATES, M.G.; RIGO, L.U.; STEFFENS, M.B.R. e SOUZA, E. M. Endophytic *Herbaspirillum seropedicae* Expresses *nif* Genes in Gramineous Plants. **FEMS Microbiol. Eco.**, V. 45(1), p. 39-47, 2003.

RONSON, C.W. NIXON, B.T.; ALBRIGHT, L.M. e AUSUBEL, F.M. Rhizobium meliloti *ntrA* (*rpoN*) Gene Is Required for Diverse Metabolic Functions **J. Bacteriol.** v.169(6), p. 2424-2431, 1987.

RUMINY, P.; DERAMBURE, C.; CHANDRASEGARAN, S. e SALIER, J.P. Long-range identification of hepatocyte nuclear factor-3 (FoxA) high and low-affinity binding sites with a chimeric nuclease. **J. Mol. Biol.** V.310, p. 523-535, 2001.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning a laboratory manual**. 2 ed., New York: Cold Spring Harbour Laboratory Press, 1989.

SANDERS, D.A.; GILLECE-CASTRO, B.L.; STOCK, A.M.; BURLINGAME, A.L. e KOSHLAND, D.E.Jr Identification of the site of phosphorylation of the chemotaxis response regulator protein CheY. **J. Biol. Chem.** v.264, p. 217770-217778, 1989.

SANDERS, D.A.; GILLECE, C.B.; BURLINGAME, A.L. e KOSHLAND, D.E.Jr. Phosphorylation site of NtrC, a protein phosphatase whose covalent intermediate activates transcription. **J. Bacteriol.** v.174, p. 5117-5122, 1992.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-termination inhibitors. **Proc. Natl. Acad. Sci. USA.**, v. 74, p. 5463-5467, 1977.

SASSE-DWIGHT, S. & GRALLA, J.D. Probing the *Escherichia coli glnALG* upstream activation mechanism *in vivo*. **Proc. Natl. Acad. Sci. USA** v.85, p. 8934-8938, 1988.

SCHELL, M.A. Molecular Biology of the LysR family of transcription regulators. **Annu. Rev. Microbiol.** v47, p 597-626, 1993.

SCHULTZ J.; MILPETZ F.; BORK P.; PONTING C.P. SMART, a simple modular architecture research tool: identification of signaling domains. **Proc. Natl. Acad. Sci. USA** v.95, p.5857-5864, 1998.

SCHULTZABERGER, R.K. & SCHENEIDER, T.D. Using sequence logos and information analysis of Lrp DNA binding sites to investigate discrepancies between natural selection and SELEX. **Nuc. Ac. Res.** v.27(3), p. 882-87, 1999.

SHERIDAN, S.D.; BENHAM, C.J. & HATFIELD, G.W. Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activation sequence. **J. Biol. Chem.** v.273, p.21298-21308, 1998.

SOUZA, E.M. Clonagem, caracterização e sequenciamento dos genes *nifA* e *nifB* de *Herbaspirillum seropedicae*. Curitiba, 1990. 264f. Tese (Doutorado em Ciências- Bioquímica) – Setor de Ciências Biológicas, Universidade Federal do Paraná.

SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U.; YATES, M.G. e PEDROSA, F. Sequence and structural organization of *nifA*-like gene and part of a *nifB*-like gene of *Herbaspirillum seropedicae* strain Z78. **J. Gen. Microbiol.** v.137, p.1511-1522, 1991.

SOUZA, E. M.; MACHADO, H.B.; YATES, G.M. Deletions analysis of the promoter region of the *nifA* gene from *Herbaspirillum seropedicae*. In TIKHONOVICH, I. A.; NIKOLAI, A. P.; VASSILY, I. R.; WILLIAM, E. N., **Nitrogen fixation fundamentals and applications**. X International Congress in Nitrogen Fixation. St. Petersburg, p.260, 1995.

SOUZA, E.M.; PEDROSA, F.O.; DRUMMOND, M.; RIGO, L.U. & YATES, M.G. Control of *Herbaspirillum seropedicae* NifA activity by ammonium ions and oxygen. **J. Bacteriol.** V.181, p. 681-684, 1999.

STOCK, J.B.; NINFA, A.J. & STOCK, A.M. Protein phosphorylation and regulation of adaptive responses in bacteria. **Microbiol. Rev.** v.53, p.450-490, 1989.

STOCK, A. M.; ROBINSON, V.L. & GOUDREAU, P.N. Two-Component signal transduction **Annu. Rev. Biochem.** v.69, p.183-215, 2000.

STUDHOLME, D. J. & DIXON, R. Domain architecture of σ^{54} - dependent transcriptional activators **J. Bacteriol.** v.185, p.1757-1767, 2003.

SU W, PORTER S, KUSTU S, ECHOLS H. DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. **Proc Natl Acad Sci U S A**, v.87(14), p.5504-8, 1990.

SUNDERSAN, V.; JONES, J.D.G.; OW, D.W. & AUSUBEL, F. M. *Klebsiella pneumoniae nifA* product activates the *Rhizobium meliloti* nitrogenase promoter. **Nature**, v.301, p. 728-732, 1983.

SZE CC, BERNARDO LM, SHINGLER V. Integration of global regulation of two aromatic-responsive sigma(54)-dependent systems: a common phenotype by different mechanisms. **J Bacteriol.** v.184(3), p. 760-70, 2002.

TATUSOV, R.L.; KOONIN, E.V. & LIPMAN, D. J. A genomic perspective in protein families. **Science** v. 278, p. 631-637, 1997.

TROPEL D, VAN DER MEER JR. Characterization of HbpR binding by site-directed mutagenesis of its DNA-binding site and by deletion of the effector domain. **FEBS J.** v.272(7), p.1756-66, 2005.

TUERK, C. & GOLD, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. **Science** v. 249, p.505-510, 1990.

TWERDOCHLIB, A.L.; CHUBATSU, L.S.; SOUZA, E.M.; PEDROSA, F.O.; STEFFENS, M.B.; YATES, M.G. e RIGO, L.U. Expression, purification and DNA-binding activity of the solubilized NtrC protein of *Herbaspirillum seropedicae*. **Prot. Exp. Pur.** v. 30, p. 117-123, 2003.

VON HIPPEL, P. H.; BEAR, D.G.; MORGAN, W.D. e MC SWIGGEN Protein-Nucleic Acid Interactions in Transcription: A Molecular Analysis. **Ann. Rev. Biochem.** V.53, p. 389-446, 1984.

WASSEM, R.; PEDROSA, F.O.; YATES, M.G.; REGO, F>G.M.; CHUBATSU, L.S.; RIGO, L.U. & SOUZA, E. M. Control of autogenous activation of *Herbaspirillum seropedicae nifA* promoter by IHF protein **FEMS Microbiol.** v.212, p.177-182, 2002.

WHITE, D. **The Physiology and Biochemistry of Prokaryotes Oxford University Press**, segunda edição, 2000.

WILLEMS RJ, GEUIJEN C, VAN DER HEIDE HG, RENAULD G, BERTIN P, VAN DEN AKKER WM, LOCHT C, MOOI FR. Mutational analysis of the *Bordetella pertussis fim/fha* gene cluster: identification of a gene with sequence similarities to haemolysin accessory genes involved in export of FHA. **Mol Microbiol.** v.11(2), p.337-47, 1994.

WILKINS M.R.; GASTEIGER E.; BAIROCH A.; SANCHEZ J.C.; WILLIAMS K.L.; APPEL R.D.; HOCHSTRASSER D.F. Protein Identification and Analysis Tools in the ExPASy Server in: **2-D Proteome Analysis Protocols**, 1998.

WÖSTEN, M.M.S.M. Eubacterial Sigma-factors. **FEMS Microbiol. Rev.** v.22, p. 127-150, 1998.

YANISCH-PERRON, C., VIEIRA, J. AND MESSING, J., Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors **Gene** v 33, p 103-119, 1985.

YIM HH, VILLAREJO M. *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. **J Bacteriol.** v.174(11) p.3637-44, 1992.

YIM HH, BREMS RL, VILLAREJO M. Molecular characterization of the promoter of *osmY*, an *rpoS*-dependent gene. **J Bacteriol.** v.176(1), p.100-7, 1994.

YOUNG, J.P.W. Phylogenetic classification of nitrogen-fixing organisms. In: STACEY,G.; BURRIS, R. H.; EVANS, H. J. ed. **Biological Nitrogen Fixation**. London: Chapman & Hall, New York, p.43-86. 1992.

ZHANG, X.; CHANEY, M.; WIGNESHWERARAJ, S.R.; SCHUMACHER,J.; BORDES, P.; CANNON, W. & BUCK, M. Mechanochemical ATPases and transcriptional activation. **Mol. Microbiol.** v.45, p. 895-903, 2002.

Apêndices

Apêndice I – Tabela da Análise das Proteínas Reguladoras de Transcrição de *H. seropedicae*

	ORF	motivos estruturais	COG	Provável função	aa	MW (Da)	pI
LysR	186.0089	HTH_1 LysR_substrate	0583	Associação simbiótica, <i>nodD</i>	309	34752,2	6,52
	202.0060	HTH_1 LysR_substrate	0583	Resistência a antibióticos	332	36170,12	10,7
	205.0079	HTH_1 LysR_substrate	0583	Via de clivagem de glicina	336	37530,19	7,25
	214.0040	HTH_1 LysR_substrate	0583		303	33186,89	8,75
	195.0139	HTH_1 LysR_substrate	0583	Regulação osmótica e a pH alcalino	303	33537,77	8,83
	236.0256	HTH_1 LysR_substrate	0583		309	33572,94	7,89
	242.0355	HTH_1 LysR_substrate	0583	Via da D-serina deaminase	301	33271,60	9,87
	250.0188	HTH_1 LysR_substrate ou DUF323	0583	Via de clivagem de glicina	306	34047,80	6,61
	256.0386	HTH_1 LysR_substrate	0583	Patogenicidade ou síntese de aminoácidos	331	35101,89	9,56
	258.0050	HTH_1 LysR_substrate	0583	Via de clivagem de glicina	300	33376,47	9,81
	260.0285	HTH_1 LysR_substrate	0583		299	32362,61	6,07
	268.0471	HTH_1 LysR_substrate	0583	Metabolismo de alose	300	33662,84	6,76
	272.0045	HTH_1 LysR_substrate	0583		291	31733,76	8,78
	274.0154	HTH_1 LysR_substrate	0583	Metabolismo de alose	322	34871,37	8,85
	275.0162	HTH_1 LysR_substrate	0583		305	32799,16	11,2
	281.0078	HTH_1 LysR_substrate	0583		308	33898,29	9,23
	283.0147	HTH_1 LysR_substrate	0583		310	34382,89	9,42
	283.1131	HTH_1 LysR_substrate	0583	Glutaciona-S-transferase/toxinas	307	34274,43	7,21
	287.0615	HTH_1 LysR_substrate	0583		313	34648,68	7,26
	181.0111	HTH_1 LysR_substrate	0583	Metabolismo de compostos aromáticos	426	47357,27	10,3
	163.0009	HTH_1 LysR_substrate	0583	Controle de adesão ao hospedeiro	291	32670,47	6,29
	137.0036	HTH_1 LysR_substrate	0583	Sensoriamento de naftaleno/nitrotolueno	319	35646,41	9,22
	153.0065	HTH_1 LysR_substrate	0583		308	34192,05	6,01
	169.0042	HTH_1 LysR_substrate	0583		337	38111,75	5,81
195.0154	HTH_1 LysR_substrate	0583	Metabolismo de glutamato	287	31598,54	9,54	
197.0047	HTH_1 LysR_substrate	0583		331	37176,39	8,35	

	201.0171	HTH_1 LysR_substrate	0583	Proteína ligadora de galactose	321	34515,21	8,02
	203.0166	HTH_1 LysR_substrate	0583		299	33228,18	9,46
	220.0138	HTH_1 LysR_substrate	0583	Sensoriamento de naftaleno/nitrotolueno	311	35082,63	5,9
	210.0134	HTH_1 LysR_substrate	5836		314	36005,77	9,31
	215.0158	HTH_1 LysR_substrate	0583	Metabolismo de nopalina	312	33618,66	6,42
	233.0235	HTH_1 LysR_substrate	0583		201	38363,95	10,8
	233.0237	FRAMESHIFT	0583		156		
	237.0313	HTH_1 LysR_substrate	0583		297	33645,79	8,57
	238.0250	HTH_1 LysR_substrate	0583		297	32301,33	8,93
MarR	228.0028	PFAM:MarR	5631		176	19618,77	9,46
	191.0064	HTH:MarR	1846		146	16245,74	8,64
	125.0012	HTH MarR	1846		153	17207,96	6,63
	169.0085	HTH MarR	1846		208	23485,01	10,1
	164.0046	HTH MarR	0583		315	34668,93	9,53
	219.0259	HTH MarR	1846		141	16258,46	8,93
	232.0105	HTH MarR	1846		174	19635,66	7,94
IcIR	182.0010	HTH:ICLR	1414		282	28634,03	6,22
	188.0158	HTH:ICLR	1414		176	32986,56	7,3
	199.0095	HTH:ICLR	1414		261	28836,93	7,07
	243.0053	HTH_ICLR	1414	Quimiotaxia via beta-cetoadipato	273	29786,29	7,81
	243.0061	HTH_ICLR	1414	Quimiotaxia via beta-cetoadipato	260	28705,09	6,85
	256.0289	HTH:ICLR	1414		263	29001,08	5,69
	273.0226	IcIR	1414		192	21036,13	7,88
	273.0571	IcIR	1522		159	21036,13	7,88
	273.0632	IcIR	1414	Quimiotaxia via beta-cetoadipato	216	23544,63	9,23
	284.1389	HTH IcIR	1414	Quimiotaxia via beta-cetoadipato	260	28047,43	9,35
	264.0610	IcIR	1414		289	31408,28	9,01
GntR	239.0045	HTH:GNTR PFAM:UTRA	2188	Metabolismo de fosfato	244	27318,54	9,27
	239.0094	HTH:GNTR	2186		240	26360,81	6,46
	246.0015	HTH_GNTR	1802		228	25199,88	5,58

	250.0063	HTH_GNTR	1846		234	25849,69	7,88
	252.0462	HTH_GNTR	2186	Sistema do hexuronato	256	29035,21	6,68
	262.0695	HTH_GNTR	1802		243	27123,92	5,7
	263.0113	HTH_GNTR	2186		233	25326,57	5,69
	264.0232	HTH_GNTR	2186		65	10204,62	6,58
	264.0604	HTH_GNTR	1802		243	27115,68	6,81
	254.0501	HTH_GNTR	2186		255	27961,47	6,11
	270.0610	HTH_GNTR	2186		249	27827,77	6,61
	276.0253	HTH_GNTR	1802		213	23661,19	9,13
	283.1351	HTH_GntR	1802	Resistência a antibióticos	222	25459,2	6,84
	285.0009	HTH_GNTR	2188		30	60589,05	8,86
	285.0011	FRAMESHIFT	2188		212		
	253.0410	HTH_GntR	1802		245	27528,46	6,6
	253.0472	HTH_GntR	1802	Metabolismo do malonato	249	27055,78	6,26
	267.0688	HTH_GntR	1802	Metabolismo do malonato	234	26908,98	6,55
	201.0179	HTH_GntR	1846	Metabolismo de ác siálico	235	26545,29	6,09
Laci	219.0270	HTH:LACI	1609	Repressão da síntese de gluconato	340	36520,86	6,13
	220.0019	HTH:LACI	1609	Produção de toxinas	366	39489,32	8,92
	240.0050	HTH:LACI	1609				
	240.0053	FRAMESHIFT	1609				
	255.0457	HTH:LACI	1609	Repressão da síntese de gluconato	352	38322,83	8,96
	256.0343	HTH:LACI	1609	Repressão da síntese de gluconato	365	38943,57	8,97
	260.0189	HTH_LACI	1609	Proteína de controle catabolico	345	36733,15	5,84
	270.0189	HTH_LACI	1609		338	36509,95	8,63
Xre	283.0732	HTH:XRE	5499		132	149444,3	6,28
	199.0062	HTH:XRE PAS	1846		285	32302,92	9,05
	278.0446	HTH Xre	1396	Regulação de enzima de restrição	89	10307,73	5,44
	285.0693	HTH Xre	1396	Repressor de profago	143	15781,74	5,12
	287.1262	HTH Xre	1396	Repressor de profago	145	15357,69	9,52
M er R	201.0085	HTH:MERR	0789		330	35826,73	5,31

	234.0092	HTH:MERR	0789		135	15594	7,87
	241.0206	HTH:MERR	0789		144	16202,47	9,2
	258.0274	HTH_MERR	0789	Resistência a metais pesados	150	16748,09	6,06
	258.0298	HTH_MERR	0789	Resistência a metais pesados	141	15551,04	7,01
	281.0595	HTH MerR	0789		142	16477,89	9,33
AsnC	271.0224	HTH_AsnC	1522	Metabolismo de leucina	153	17390,37	8,93
	274.0004	HTH_AsnC	1522		170	19632,69	8,55
	275.0470	HTH_AsnC	1522		244	27686,94	9,43
	158.0067	HTH_AsnC	1522	Metabolismo de leucina	162	18516,51	6,85
	282.0406	HTH_AsnC	1522	Metabolismo de leucina	154	17310,25	7,74
	283.0949	HTH_AsnC	1522		163	18106,61	6,1
	218.0074	HTH_AsnC	1522		172	19376,32	6,98
ArsR	240.0237	HTH:ARSR	0640	Resposta a sinais simbióticos	110	12475,55	6,33
	246.0322	HTH_ARSR	0640	Associação simbiótica (N ₂ , nodulação)	110	12317,15	9,62
TetR	227.0272	TetR_N	1309		209	22398,87	9,99
	261.0507	TetR_N	1309		211	27730,36	8,92
	270.0486	TetR_N	1309		233	25989,89	8,85
	276.0600	TetR_N	1309	Metabólito antifúngico	239	26759,94	8,32
	135.0020	TetR_N	1309		146	20587,7	6,44
	178.0046	TetR_N	1309	Resistência a drogas	213	23457,03	9,59
	188.0079	TetR_N	1309		206	23024,83	6,84
	201.0047	TetR_N	1309	Resistência a drogas	204	22562,17	5,7
	221.0015	TetR_N	1309		201	22078,65	8,77
	238.0317	TetR_N	1309		203	22187,26	6,41
AraC	181.0100	HTH_AraC	2207		331	38846,53	9,01
	236.0242	DJ1_Pfpl HTH_AraC	4977		190	46617,14	10,3
	236.0248	FRAMESHIFT	4977		247		
	239.0384	DJ1_Pfpl HTH_AraC	4977		313	34448,33	9,37
	245.0069	HTH_AraC	2207		337	36647,23	7,85
	265.0304	AraC_binding HTH_AraC	2207		255	27907,17	9,75

	268.0678	DJ1_Pfpl HTH_AraC	4977	Metabolismo de glucose	330	35912,97	9,83
	277.0458	DJ1_Pfpl HTH_AraC	4977		375	41688,47	7,73
	280.0613	HTH_AraC	2207		299	33393,14	9,37
	282.0399	HTH_AraC	2207	<i>Multiple transferable resistance</i>	323	34145,28	8,78
	286.1343	HTH_AraC	2207	Metabolismo etanolamina	351	39485,77	7,02
	286.1596	DJ1_Pfpl HTH_AraC	4977		335	36630,01	6,84
	287.0681	AraC_binding HTH_AraC	2207		271	30038,83	7,09
	287.0725	HTH_AraC	2207	Metabolismo de nitrogênio	405	39101,41	8,84
	287.0874	HTH_AraC	2207		307	34218,83	6,46
	287.0929	HTH_AraC	2207		349	39089,17	8,51
	169.0055	AraC_N HTH AraC	2207		348	39024,87	5,52
	218.0069	Ada_ZN_binding HTH AraC Metil_transF_1N Pfam:DNA binding	2207		369	39633,34	8,96
	218.0031	HTH AraC	2207		273	30237,01	9,54
sigma54	216.0019	REC AAA HTH_3	2204		466	51157,96	6,12
	262.0268	REC AAA HTH_3	2204	Metabolismo de nitrogênio/ ác. graxos de cadeia curta	450	49544,13	6,31
	281.0022	REC AAA HTH_3	2204		448	47934,71	5,85
	172.0051	GAF AAA HTH_3	3604	Metabolismo de espécies reativas de nitrogênio	526	57216,21	6,72
	241.0256	REC AAA HTH_3	2204	metabolismo de nitrogênio/ C4 decarboxilato	445	48204,22	5,26
	190.0071	Xil_R_N V4R AAA HTH_8	3604	Metabolismo de compostos fenólicos	611	66711,69	5,74
	254.0323	REC AAA HTH_3	2204	Metabolismo de nitrogênio	500	54458,02	5,68
	275.0911	GAF AAA HTH_3	3604	Metabolismo de nitrogênio	542	60807,21	8,99
	287.0556	AAA HTH_3	1221		380	42480,4	7,37
	254.0269	Hjc AAA HTH_3	2204	Metabolismo de propionato	328	62645,15	9,02
	254.0276	FRAMESHIFT			243		
	208.0173	GAF PAS AAA HTH_8 ou IciR PAS AAA HTH_8	3284	Metabolismo de acetoína/glicerol	675	73442,43	7,04
CheY	281.0067	PAS PAC HTH	2771		210	23689,32	8,51
	127.0031	ZnF NFX	0784		68	7291,26	6,38
	187.0121	REC HTH_LuxR	4566	Metabolismo de nitrogênio	196	21375,76	5,55
	196.0145	REC HTH_LuxR	2197	Controle de infecção a planta	146	23385,91	5,56

	213.0084	REC HTH_LuxR	2197	Controle da nitrato redutase	216	24113,96	5,87
	221.0220	REC HTH_LuxR	4566	Metabolismo de nitrogênio	252	28165,77	9,45
	243.0210	REC trans_reg_C	0745		236	26415,17	6,53
	243.0335	Response_reg LuxR	4566	Metabolismo de nitrogênio	195	21317,43	9,09
	245.0017	REC HTH_LuxR	0784		307	33237,32	6,32
	248.0100	REC HTH_3	4566	Metabolismo do heme / virulência	182	19901,54	8,14
	276.0373	REC HTH_3	4566	Metabolismo do heme / virulência	178	19616,46	6,16
	280.0468	REC HTH_LuxR	4566	Metabolismo de nitrogênio	242	30335,05	8,62
	176.0021	REC Trans_reg_C	0745	Sistema de transporte de K ⁺	236	26255,9	6,56
	184.0005	HATPase	0642		122	13031,68	5,64
	184.0011	REC HTH_LuxR	2197	Mecanismos de colonização	217	23641,16	6,07
	184.0098	REC trans_reg_C	0745	Resistência a antibióticos	234	25825,7	6,82
	192.0159	REC trans_reg_C	0745		262	28975,51	7,87
	198.0006	REC trans_reg_C	0745	Stress oxidativo/patogenicidade	185	30695,26	6,35
	198.0007		0745		91		
	200.0077	REC trans_reg_C	0745	Metabolismo de fosfato	220	24192,83	6,1
	249.0302	REC HTH_8	2204	Metabolismo de nitrogênio	249	27083,87	5,8
	201.0144	REC HTH_AraC	3706	Metabolismo de fosfato	262	28979,27	9,24
	201.0154	Trans_reg_c	0745		224	24956,26	9,69
	204.0151	REC HTH_LuxR	2197	Metabolismo de nitrito/nitrato	219	23708,44	5,76
	214.0135	REC trans_reg_C	0745	Patogenicidade/virulência	253	28627,79	5,52
	217.0047	REC trans_reg_C	0745	Metabolismo de fosfato	230	26036,19	6,67
	228.0222	HTH_GntR Amintran_1_2 ou HTH_GntR DegT_DnrJ_EryC1	1167		516	56598,02	8,86
	236.0325	REC trans_reg_C	0745		239	26587,34	6,14
	254.0492	ZnF_NFX	1959	Metabolismo de nitrogênio	179	19620,44	6,57
SIS	286.1385	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	293	31826,62	8,67
	255.0359	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	286	30885,73	8,73
	266.0033	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	327	39928,04	9,6
	277.0009	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	280	31325,81	7,8
	228.0174	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	329	35796,77	7,22
	286.1509	HTH_6 PFAM:SIS	1737	Metabolismo de fosfoaçúcares	319	34380,2	8,79

Descrição dos Domínios estruturais citados na tabela anterior.

- > **MarR – Pfam** - Resistência múltipla a antibióticos, sistema de resistência não específico. Expressão dos operons Mar é regulada por um repressor, MarR; inúmeros compostos se ligam a MarR fazendo com que ele não ligue mais o DNA e a transcrição do operon ocorra.
- > **PAS – SMART** - Domínio sensor de sinal, geralmente dependente da ligação de um cofator. Aparece em canais de K⁺.
- > **PAC – SMART** - Domínio C-terminal ligado a PAS. Contribui para a montagem do domínio PAS.
- > **HTH-LUXR – SMART** - Domínio localizado na região C-terminal da sequência que possui um motivo de ligação ao DNA do tipo HTH e é ativado por moléculas auto-indutoras como N-(3-oxohexanoyl)-L-homoserina-lactona (OHHL).
- > **HTH-XRE – SMART** - Grande família de proteínas com domínio HTH que incluem: proteínas controladoras de cópias de plasmídios, metilase de bactérias; proteínas de controle de transcrição de bacteriófagos e proteínas vegetativas específicas de “slime mold”.
- > **DUF387 – Pfam** - Provável domínio HTH de regulador de transcrição. Bastante conservado em bactérias.
- > **HTH –ICLR – SMART** - Motivo HTH que ocorre na região N-terminal, onde a porção C-terminal da proteína provavelmente deva interagir com substrato regulatório.
- > **HTH – MARR – SMART** - Resistência múltipla a antibióticos, sistema de resistência não específico.
- > **HTH_1 – Pfam** - Domínio relacionado às proteínas da família LysR, é do tipo hélice-volta-hélice e faz interação com de DNA, geralmente presente na porção na porção N terminal.
- > **LysR_substrate – Pfam** - Estrutura semelhante ao domínio de ligação de proteínas de periplasma. Domínio característico dos reguladores de transcrição da família LysR, geralmente se relaciona com um domínio HTH N-terminal.
- > **HTH-LACI – SMART** - Domínio HTH em direção a região N-terminal, presente em várias proteínas reguladoras de transcrição de procaríotos.
- > **TetR_N – Pfam** - Subfamília de domínios HTH. Muitas das proteínas que possuem este domínio funcionam como repressores que controlam o nível de suscetibilidade a antibióticos hidrofóbicos e detergentes.
- > **HTH_6 – Pfam** - Domínio característico de genes regulatórios do metabolismo de fosfoaçúcares. (sem estrutura resolvida) Domínio HTH N-terminal relacionado a domínios SIS.
- > **SIS – Pfam** - Sugar Isomerase. Encontrado em fosfoaçúcar isomerase e proteínas ligadoras de fosfoaçúcar. Encontrado também em proteínas que regulam a expressão de genes da síntese de fosfoaçúcares. provavelmente o domínio SIS se ligue ao produto final da via.
- > **HTH-MERR – SMART** - Subfamília de domínios HTH. Caracterizado na proteína responsável pela indução dependente de mercúrio no operon de resistência a mercúrio.
- > **HTH_GNTR – SMART** - Subfamília de domínios HTH. Presente no repressor do operon do gluconato. Nesta subfamília o domínio HTH encontra-se em direção ao N-terminal da proteína. O C-terminal destas proteínas é geralmente um domínio de oligomerização efetor-ligante.
- > **UTRA – Pfam** - UbiC transcription regulator-associated. Acredita-se que module a atividade de fatores de transcrição de bactérias em resposta a pequenas moléculas ligantes.
- > **HTH_ARSR – SMART** - Família de domínios HTH em que a proteína regulatória se desliga do DNA na presença de íons metálicos. ArsR atua como um repressor do operon de resistência a arsênio.
- > **ZNF/NFX – SMART** - Repressor de transcrição. Domínio de coordenação de zinco, dedo de zinco C-X(1-6)-H-X-C-X3-C(H/C)-X(3-4)-(H/C)-X(1-10)-C, onde X representa qualquer aminoácido e os números entre parênteses representam o número de resíduos.
- > **ICLR – Pfam** - Domínio presente em fatores de transcrição de bactérias, incluindo a proteína regulatória do operon do glicerol e o repressor do operon do acetato. Possui um motivo HTH na porção N-terminal e a região C-terminal provavelmente liga substratos regulatórios.

- > **AsnC** - Regulador ativador autógeno de asparagina sintetase A. Subfamília de domínios HTH onde o domínio de ligação ao DNA encontra-se na região N-terminal.
- > **Neg_Reg – Pfam** - Regulação negativa da síntese de proteases e esporulação.
- > **HTH-ARAC – SMART** - Domínio HTH semelhante ao da proteína AraC, a maioria se caracteriza por ser um regulador de transcrição positivo. O domínio HTH geralmente localiza-se no ¾ final das proteínas, próximo ao C-terminal.
- > **DJ-1_Pfpl – Pfam** - Domínio característico de proteases presente em alguns reguladores de transcrição.
- > **AraC_Binding – Pfam** - Domínio de ligação de arabinose e dimerização. Encontrado em conjunto com HTH_AraC.
- > **AAA – SMART** - Domínio ATPase associado a diversos processos celulares.
- > **REC – SMART** - Presente em proteínas da família CheY; que regula o movimento rotatório de flagelos. Esse domínio contém um sítio aceptor de fosfato que é fosforilado por um homólogo de histidina quinase. Esse domínio recebe o sinal de um sensor nos sistemas de dois componentes de bactérias. Usualmente é encontrado na porção N-terminal de um domínio de ligação de DNA.
- > **HTH_3 – Pfam** - Domínio de ligação ao DNA HTH, que possui grande similaridade com domínios HTH de reguladores de transcrição dependentes de sigma54 de bactérias; como a proteína NtrC.
- > **Trans_reg_C – Pfam** - Domínio de ligação de DNA geralmente encontrado em associação com um domínio sensor. Encontrado principalmente na porção C-terminal de reguladores de transcrição.
- > **Response_reg – Pfam** - Domínio que recebe um sinal de um parceiro em um sistema de dois componentes. Geralmente encontrado no N-terminal de um domínio efetor de ligação de DNA.
- > **AraC_N – Pfam** - Domínio N-terminal característico dos reguladores de transcrição do tipo AraC. Esta família pode representar um tipo de braço N-terminal que liga o domínio C-terminal de ligação ao DNA para manter a proteína em uma conformação inativa.
- > **XiIR_N – Pfam** - N-terminal de reguladores de transcrição dependentes de sigma-54 em proteobactérias; incluindo ativadores da degradação de fenol.
- > **V4R – Pfam** - Vinil 4 redutase. Domínio ligador de pequenas moléculas, provavelmente liga hidrocarbonetos.
- > **GAF – SMART** - Domínio presente em fitocromos e fosfodiesterases cGMP específicas. Fosfodiesterases cGMP específicas catalisam a conversão de 3'5'guanossina fosfato cíclica para 5'guanossina fosfato. O fitocromo é um fotorreceptor que existe em duas formas que são reversivelmente interconversíveis por luz. Esse domínio é encontrado em NifA.

> **Ada_Zn_binding – Pfam** - Proteínas Ada de *E. coli* reparam resíduos de 6O – metilguanina e metilfosfotriesters no DNA por transferência do grupo metil para um resíduo de cisteína. Esse domínio contém 4 cisteínas conservadas que fazem coordenação com zinco. Uma dessas cisteínas é o aceptor do metil. Um domínio metilado pode então especificamente ligar ao ada box no DNA duplex.

> **Methyltransf_1N – Pfam** - 6-O-methylguanine DNA methyltransferase, ribonuclease-like domain.

> **Methyltransf_1 – Pfam** - 6-O-methylguanine DNA methyltransferase, DNA binding domain

Ref. <http://smart.embl-heidelberg.de/> e <http://www.sanger.ac.uk/Software/Pfam/>

Apêndice II – Grupos Funcionais Definidos pelo COG

Grupo	Descrição
Armazenamento e Processamento de Informação	
J	Tradução, Estrutura de Ribossomo e Biogênese
K	Transcrição
L	Replicação de DNA, Recombinação e Reparo
Processos Celulares	
D	Divisão Celular e Partição Cromossomal
O	Modificações Pós-traducionais, Turnover protéico e Chaperonas
M	Biogênese do Envelope Celular, Membrana Externa
N	Mobilidade e Secreção celular
P	Transporte e Metabolismo de Íons Inorgânicos
T	Mecanismos de Transdução de Sinal
Metabolismo	
C	Produção e Conversão de Energia
G	Metabolismo e Transporte de Carboidratos
E	Metabolismo e Transporte de Aminoácidos
F	Metabolismo e Transporte de Nucleotídeos
H	Metabolismo de Coenzimas
I	Metabolismo de Lipídios
Q	Biossíntese, transporte e catabolismo de metabólitos secundários
Pouco caracterizados	
R	Somente Predição da Função Geral
S	Função Desconhecida

Apêndice III – Alinhamento da Sequência de Aminoácidos das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de *H. seropedicae* e NtrC de *A. brasilense*.

172.0051 -----
 281.0022 -----
 A.brasilense -----
 NtrC -----
 DctD -----
 216.0019 -----
 AtoC -----
 190.0071 -----MPSSPRISLAEMARHQGGGLSSAGTDRWSVDAHPTLADLSECLFFSPG
 287.0556 -----
 NifA -----MATILDDRNVNLELVTIYEISKILGSSLD
 PrpR -----MKRPAPVDRHADKPVIWTVSVSRLSE
 208.0173 MQYQANRPSAGALLGTSPLVQASTDGVPDLQAIERSHQRSSSYGIMRGQRPDLDSLRSAD

172.0051 -----MTSRTLLDALVPLVADLSR-----
 281.0022 -----MARILIVDDDAAFRGLAE-----
 A.brasilense -----MSAATILVADDDR-----
 NtrC -----MKPWIIVDDDE-----
 DctD -----MDKPGGAASALLIEDEQ-----
 216.0019 -----MELESPMGKTAPLMLVDDDP-----
 AtoC -----VSENKPKLLIVEDDP-----
 190.0071 DGR IWLNDQRM LLIHSRSMGTLRRELIDQLGIDQARGLL-----
 287.0556 -----
 NifA LSKTLREVLNVLSAHLETKRVL LSLMQDSGELQLVSAIG-----
 PrpR LFRDITLFDALADIEPIHLGFDDAVREIRERLATERCD-----
 208.0173 LAQTLEANQTL SGHARPVMETLYDQIRNTHSMVILSDAEGTILHTLGSDSDFLAKADRVAL

172.0051 -----EIDDRERYRRLQLTRLTLFPCDAAALLRLEGDVLVPLAIDGLSSDTLGR
 281.0022 -----TVSDLG-----HDVLEAGSTAAGLKLLAAHAVDAAIVDLRMPDED-
 A.brasilense -----AIRTVLTLQALARLGHVVRTTGNASTLWRWVADGGDLIITDVMVP-----
 NtrC -----SIRWVLEKALARENLATQSFSSARDAIAALQNGTPQVLVSDIRMP-----
 DctD -----TVRRATAQALELGGFAVTACGSAAEALPLITRDFAGVVVSDVRLP-----
 216.0019 -----DLLRLLTLRLNASGYRILAVESAEAAALQLAISLPALVITDVRLP-----
 AtoC -----GLQKQLRWSL DAY--EVLVAGDRESALAMVRRHEPAVVMTDGLGPPDPD
 190.0071 -----TRAGYISGARDAQLVRERWPDADPSTILMAGTRLHTLEGLVKVVPVVSFS
 287.0556 -----
 NifA -----LSYEEFQSGRYRVGEGITGKIFQTETPIVVRDLAQEPLFLARTSPRQSQ
 PrpR -----VVISAGSNGAYLKSRLPVPVIVARASGFDMQALARARQLTRRIALVTY
 208.0173 GPGGRWSEQLRGTNAIGTALFEQKPTTVHAQQHYLDANRFLTCSAAPIFDHRGQVLGVLD

172.0051 RFRITEHPRFGILLGNAAPTRFAADSPLPDPYDGLVDGLNDQHSGELEVHDCMGCALTVN
 281.0022 -----GLAFLQRAPAIAAIP-----CIMLT
 A.brasilense -----DENGLDLIPRIKKIRPDLR-----IIVMS
 NtrC -----GASGLELLQTVKAKHPGIP-----VIIT
 DctD -----GLSGLDVLARVVAIDHDL P-----VIVVT
 216.0019 -----GMDGLALFAEIRSRYPALP-----VILLT
 AtoC -----GASEGFATLQQVLALAPDTK-----VIMLT
 190.0071 -----YDPETGRYEGEFLWHNSCEA-----DEHLA
 287.0556 -----
 NifA -----DGEVISFVGVPIKAAREMLG-----VLCVF
 PrpR QEAMPALAEFGQRFGL EIVQRTYATEEDAR-----AVV
 208.0173 VSGEGCSFHKHTMALVRMSAQMIENHLLAGSFPDCITLHFHSRAEFVGTLVEGIVSFTPG

172.0051 GRAWGLMTLDALGRGRFDRVDMDSLQAF AALAAATVNVAERISDLAQAGQQAQRAEAYR

281.0022 AYASGDNTIEAIRLGFADHLTKP-----VPRAALVETLERALRQGASTPAAAS
A.brasilense AQNTLITAVKAAERGAFFEYLKPK-----FDLKELVSVVERALNSN---TPPAA
NtrC AFSDLDASVSAFQGGAFEYLAKP-----FDVDKAVELIRRALDES---LRETD
DctD GHGDVGMAVEAMRAGAYDFVEKP-----FSSDTLLEIALRAQEKRNVLNRR
216.0019 AHGTIPDAVDATTQGAFAFYLTKP-----FDGHVLLDKVAQALSLA----APAP
AtoC GNQDHAHAVKAISMAYDFHQKP-----CIDETLRLVVQRAFYLHALQRENQR
190.0071 AYGVATAPSCWQQVGYAIGYVST-----LLGHLVIFREVQCGRAMGTAHCRVIG
287.0556 -----MHASHRSAHLYALPSS-----PEGHGDLAARASPLLLTLPHARKLG
Nifa RDGQSPSRSDHEVRLTLMVANL-----IGQTVRLYRSVAAERQQLQEERQQL
PrpR NELKAAGIEAVAGAGLITDLAEAEAGMRGIFLYSAATVRQAFEDALELARLTQLESTRSPI
208.0173 GRFLAANRSAQFQLGLTLAALQSHTFASLFGVPLSLLFEHHRKAAAPGLLELCLHNGIRIH

172.0051 LG-----LEQSQGQRELIQOSP
281.0022 PA-----TEE---RNQLVGESE
A.brasilense LP-----ADAGEADEQLPLIGRSP
NtrC IE-----QGAAQTPEIL---GQAP
DctD LR-----EAWAHDPELPLVGGQSP
216.0019 LS-----PDQSWRADILS---RSQ
AtoC MQ-----QAGGHTALGG-LITRDA
190.0071 RSAELWSDAEEDLRYLNAQDFVSGMLASEADASAEDGAQGAGEQDDDTDLVTALVGTSS
287.0556 -----TSIRATAQVFHDP
Nifa SR-----QLQGKYKLDNVIIGISK
PrpR LP-----VADSLRARHHLKDLRGRDSA
208.0173 GRAELRLRDQHFQAAREAVIDAPGLLPPAPNTTTPPPSPPQQRARHQRRPDLQALDGTGDV

172.0051 AHRQLMQEIGMAAPSDMTVLVAGETGTGKELVAHALHAAS-----PRAHKPMISLNC
281.0022 AMRQVFKQIGRAATSDATVVLVGETGTGKELVARALHRNG-----ARAARPFVAVNC
A.brasilense AMQEIYRVLARLMTDLTTLITGESGTGKELVARALHDYG-----KRRNGPFVAINM
NtrC AMQDVFRAIGRLSQSNVTVLITGESGSGKELVARALHKHS-----PRASQPFVALNT
DctD AIERVRTLIRSLGPADVVDVLLINGQGTGTGKEVVARQLHAAS-----GRK-GPFVALNC
216.0019 RVAELLAEARLVAASDASILIRGDSGTGKELLARAIHRAS-----ARAEQPFVAVNC
AtoC GMLAVCRSIEKVAPSSASVMLLGSSTGKEVLARALHQLS-----PRSGRRFMAINC
190.0071 AFNAACHLLNRVAATDATVLFVTGESVGGKERFARMLHQIG-----KRAKRACIAINC
287.0556 VSLALLERVQRVAPSDANVVIIGATGTGKELIARHVHELSE-----QRRDAPFVAVNC
Nifa AMQEVFAQVHQSAAPSSTMLRGESGTGKEVIARAIHYLS-----PRKDGPFVAVNC
PrpR EMETVRQSITLFAKSPANVLIQGETGTGKELAAQAIHRAHPLM--QKGTQAAHPFVAVNC
208.0173 RMSAIIAKLRRVIDRDIPILITGETGTGKEWLAQAIHGDSARFGARQNRAGAAPPFVAVNC
: . * : * * * * : : * : : *

172.0051 AALPDTLVESELFGHVRFAGFSGAVAD-RRGKFEMADGG-----
281.0022 AAIPAEELIESELFGHVKGAFSGATSP-RLGRFQEADGG-----
A.brasilense AAIPRELIIESELFGHEKGAFTGATNR-STGRFQEAGGG-----
NtrC AAIPKDLLESELFGHERGAFTGAQAM-RRGRFQEAGGG-----
DctD GALPETVFESEIFGHEPGAFTGAQKR-RIGKLEYANGG-----
216.0019 GAIPEQLLESELFGHVKGAFSTGAVEH-REGLFQAAGGG-----
AtoC AAIPENLLESELFGEYKGAFTGAQKQ-TIGKVELAHGG-----
190.0071 AAIPETLIEAELFVVERGAYTGATQS-RAGRFEADGG-----
287.0556 GAFSESLAESELFGHEKGAFTGATH-KPGWFEEANGG-----
Nifa AALSETLLESELFGHEKGAFTGAQGE-RKGRFELAHGG-----
PrpR GALAESLLESELFGEYEGAFSTGRRGGRAGVFEAAHRGTCARMRKAPLVRAGAGVPACS
208.0173 SAIPENLIEAELFGEYEDGAFSTGARRKGAALGKIAQAHGG-----
.*:. : *:*:* *:*:* * . * . *

172.0051 -----TLFLDEVGELPLAVQAKLLRVLQSGQLQRVGSDREHRVDVRLIAATNRDLADEVK
281.0022 -----TLFLDEIGDMELPTQAKILRVLQERELTPVGGNQSVRVDVRLVIAATHRDLQHEIT
A.brasilense -----TLFLDEIGDMPLAQTRLLRVLQEGEYTTVGGRTPIKTDVRLVIAATHRDLRDLIR
NtrC -----TLFLDEIGDMPLDLQTRLLRVLSDGHFYRVGGHQLKANVRVIAATHQNLEQRVR
DctD -----TLFLDEIESMPLALQVKLLRVLQERLERLGGNESVAIDCRVVAASKADLLKLA
216.0019 -----TLFLDEIGDMPLPLQVKLLRVLQEKTVRQVQGANQAGEVDVRIISATHRDLAAMR
AtoC -----TFFLDEVGDLPMPLQAKLLRVLQERVIERVGGHTEIPVDVRLVICATHQNKLKELTA
190.0071 -----TLFLDEVGTLSLVAQKLLRVLQEGEFERVGSARPVKVDVRLVIAATNEDLAQAVR
287.0556 -----TLFLDEIGDLPLAIQVKLLRVLQEREVVRLGSRRSIPIDVRIIAATNRLQDVA
Nifa -----TLFLDEIGEISPAFQAKLLRVLQEREFERVGGSRSIKVDVRLVLTATNRDLEKAVA
PrpR RQRIVAPVLDEIGEMPLPLQTRLLRVLQEREVVRLGSRRSIPIDVRIISATHCDLQARVR
208.0173 -----TLFLDEIGDMPLALQGRLLRVLQERAVTPLGSQRVIPVDFALVCATNQPLREMAA
: .***: : * : **.*.. : * . : : * : *

```

172.0051      RGHFRAFDFYHRLSVFPPIRVPPPLRERGRDVLVLLAGYFLEQNRARLGLSSLRLSADAQR---
281.0022      QGRFREDLWYRLQVLTIMVPALRERPGDILTIAEHFLRQLGG---DAPKRISPAAR---
A.brasilense  QGLFREDLFYRLCVVPIRLPPLRERTEDVPLLVRHFLNQCASAQG-LPVKSIDQPAMD---
NtrC          EGLFREDLYHRLNVIRLRLPSPLEERSEDIPILARYFLAQSARQLGVEAKRLSPQAMQ---
DctD          EGQFREDLYYRIGVVAIDL PALNQRREDIALLLAHFCQAAAVRYRRESPAWSAAQMQ---
216.0019      EGQFREDLYYRLNVVTLHLPALEERREDIALLANHFLKTTIAAKYPKRLTGFAPDALE---
AtoC          SGGFREDLYYRLCEII IKIPALKERHGDAVLLARHFVQKFCAKEGRATLHLSQDAVA---
190.0071      EGRFRQDLFFRLNVFPIHLPPLRERRDDIPLLMNHFLKLYGQRHGIAVRGFTQRAIQ---
287.0556      AGHFRQDLYYRLQVVQLAIPRLRERPLDILPLAYHFLHEYRQLRGYGAARIDGEGER---
Nifa          KGEFRADLYYRINVSIFIPPLRERREDIPYLVEHFLEKFRVENQRAMVAMSPQAMK---
PrpR          EGRFRADLYYRLSVLCLQLPPLRERRDDIAALAEWHLKHALAALDVRPHANLGAETVRCR
208.0173      RHAFFREDLYYRINGLQVKLPALRDRSDLPQLIERILEDEGGPHAPRFDAQAQALLLAS--
                ** *::.*: . : :* *.:* :

```

```

172.0051      -ALLDYAWPGNIRELEHLIGRSALRSLATHRTRPTILTLDAAALELPAP-----
281.0022      -LLLQHAWPGNVRELNRNTMERAAMVMSHG-----LCIEAEDIGLSARPP-----
A.brasilense  -RLKRYRWPGNVRELENLVRRLAALYSQ-----EVIGLDVVEAELADT-----
NtrC          -FLSQLEFPGNVRQLENLNCNWTVMAPG-----QTVEIKDLPQDLVEERVHAPQPAQTV
DctD          -QWQQRDWPNGNVRELNRNFADRWVLGVEQ-----MAASAVSSAVPVLSS-----
216.0019      -ILSLAHWPNGNVRLYNVVEQVCAALSTA-----PLIPASLVQRALRTP-----
AtoC          -AIERYEWPNGNVREMENCVMRAVIMADG-----QQITAEDLGLSVEGR-----
190.0071      -ALLNYSFPGNIRELQNLIERGVILAQG-----HPLDLPHPFISGERLNDEV-----
287.0556      -KLRQHDWPGNIRELENAIHHALLICQD-----DCLRAEHFHL-----
Nifa          -VMMNCYWPGNVRELENCVERTATMMRG-----DLITEVHFSCQQNK-----
PrpR          PLLCRYDWPNGNVRELNRNLMERLALFLAAEPLQALTPALIQRVAPELGQAP-----
208.0173      -----HWPGNLRQLANLVRTSLAMAEGEEAIGPHHLRDDDFLEEAGAPG-----
                :***:*::: :

```

```

172.0051      -----SLATVAAAEPSLPAESTPAQASGSLRDALERY
281.0022      -----AGAASAASAPAIDWD-----GPLAVALAEV
A.brasilense  -----TPAAQPVEEPQGEGLSAAVERHLKDYFAAHKDGMPNSGLYDRVLREV
NtrC          AAPASAAALPTAYTEPAPGVASASADGAAGASWITLLETEAAQMLASEQPEVMDILGRQF
DctD          -----LPEQVEQF
216.0019      -----TLDLLRYAEAKQKF
AtoC          -----EAEPLNLRQVRDEA
190.0071      --LSLALQGETGMLAGKSVEKSVEKVAHPTAQG--LLDVLQRWRQEAAGEAAALTLPEM
287.0556      -----HVERRLDYAVAAADERGNDPLHALEQALQGLFNQSQEGLFEKI
Nifa          -----CLTKVLHEPQQQPVVVVPLERISAHTAPSSPSGMAKDKPLAAAPPTS
PrpR          -----PSPAAPDEAQENITP
208.0173      -----ALHDGMHTMHA VRSDCLLSIQDN

```

```

172.0051      ERQLVSAALQRHGGNLAATARELGVDRANLNRLAKRLGLR-----
281.0022      ERVMIERALQGAAGNRAEAARRLGISRQQLYRRLEEAGKSHD-----
A.brasilense  ERPLISLSLSATRGNQIKAAQLGLNNTLRKKIRDLDIQVVRGLK-----
NtrC          EAALIKVALKHTGRKNDAAVRLGIGRNTITRKIQELGIGGAEDD-----
DctD          EAGLIAAALKDSEGSVAVAAERLGI PKKTLYDKIRKYQLAGS-----
216.0019      ERDYLIRLLKLTGDNVSDAARLADRNRTEFYRLLQKNGLDPRFRFGGEEAVVAGRQEERL
AtoC          ERKAVVKALTRMDGNILKAAELLGVSRTLYDLMSRHEIK-----
190.0071      ERRLIAEAI ERAEGNLSAAARLLGISRAQLAYRQKQGQ-----
287.0556      EDTVFRTAFAFCERNQVQAARLLDISRNVLRARLIRNGQISALK-----
Nifa          ERERLIWAMEQCQGWVQAKAARALNISPRQMGYALQKFNIEVKKF-----
PrpR          DNASIAAALRRFRGNRAAAARYLGMSRTTLWRKLESGTT-----
208.0173      EWAAIERALRTHAGNVSAAKALGVSRRNTIYRRLKSRGH-----
                : . : :* . :

```

Legenda: Hífens representam gaps, asteriscos aminoácidos conservados e .ponto e dois pontos aminoácidos com substituições conservadas.

Apêndice IV – Alinhamento da Sequência de Aminoácidos dos Domínios AAA das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de *H. seropedicae*

172.0051	SDMTVLVAGETGTGKELVAHALHAAS-----PRAHKPMISLNCAALPDTLVESELFG
190.0072	TDATVLFVTGESGVGKERFARMLHQIG-----KRAKRACIAINCAAIPETLIEAELFG
NtrC	SNVTVLITGESGSGKELVARALHKHS-----PRASQPFVALNTAAIPKDLLESELFG
216.0019	SDASILIRGDSGTGKELLARAIHRAS-----ARAEQPFIAVNCGAIPQLLESELFG
281.0022	SDATVVLVLETGTGKELVARALHRNG-----ARAARPFVAVNCAAI PAELIESELFG
AtoC	SSASVMLLGSSGTGKEVLARALHQLS-----PRSGRRFMAINCAAIPENLLESELFG
Nifa	SRSTMLLRGESGTGKEVIARAIHYLS-----PRKDGPFKIVNCAALSETLLESELFG
DctD	ADVDVLINGQTGTGKEVVARQLHAAS-----GRKG-PFVALNCGALPETVFESEIFG
287.0556	SDANVVIIGATGTGKELIARHVHELSE-----QRRDAPFVAVNCGAFSES LAESELFG
PrpR	SPANVLIQGETGTGKELAAQAIHRAHPLM--QGKTQAAHPFVAINCGALAESLLESELFG
208.0173	RDIPILITGETGTGKEWLAQAIHGDSARFGARQNRAGAAPFVAVNCSAIPENLIEAELFG
	: : . * : * * * * * : : * : * . * : : * : * : * * *
172.0051	HVRGAFSGAVAD-RRGKFEMADGG-----TLFLDEVGE
190.0072	VERGAYTGATQS-RAGRFEADGG-----TLFLDEVGT
NtrC	HERGAFTGAQAM-RRGRFEQAEGG-----TLFLDEIGD
216.0019	HVKGAFTGAVEH-REGLFQAHHGG-----TLFLDEIGD
281.0022	HVKGAFTGATSP-RLGRFQEADGG-----TLFLDEIGD
AtoC	YEKGAFTGAAKQ-TIGKVELAHGG-----TFFLDEVGD
Nifa	HEKGAFTGAQGE-RKGRFELAHGG-----TLFLDEIGE
DctD	HEPGAFTGAQKR-RIGKLEYANGG-----TLFLDEIES
287.0556	HEKGAFTGAITH-KPGWFEEANGG-----TLFLDEIGD
PrpR	YEEGAFTGSRRGGRAGVFEEAHRGT CARMRKAPLPVRAGAGVPACSRQRIVAPVLDEIGE
208.0173	YEDGAFTGARRK GALGKIAQAHHGG-----TLFLDEIGD
	* * : * : * * * * * : : . * * * * :
172.0051	LPLAVQAKLLRVLQSGQLQRVGSDREHRVDVRLIAATNRDLADEVKRGHFRADFYHRLSV
190.0072	LSLVAQGKLLRVLQEGEFERVGSARFPVKVDVVRVVAATNEDLAQAVREGRFRQDLFFRLNV
NtrC	MPLDLQTRLLRVLSDGHFYRVGGHQSLKANVRVIAATHQNLEQRVREGLFREDLYHRLNV
216.0019	MPLPLQVKLLRVLQEKTVRQVGANQAGEVDVRIISATHRDLEAMREGQFREDLYYRLNV
281.0022	MELPTQAKILRVLQERELTPVGGNQSVRVDVVRVIAATHRDQHEITQGRFREDLWYRLQV
AtoC	LPMP LQAKLLRFLQERVI ERVGGHTEIPVDVVRVICATHQNLKELTASGGFREDLYYRLCE
Nifa	ISPAFQAKLLRVLQEREFERVGGSRSIKVDVRLVTATNRDLEKAVAKGEFRADLYYRINV
DctD	MPLALQVKLLRVLQERRLERLGGNESVAIDCRVVAASKADLLKLAEGQFREDLYYRIGV
287.0556	LPLAIQVKLLRVLQEREVVRVLRGSRRSIPIDVRIIAATNVRLQDAVAAGHFRQDLYYRLQV
PrpR	MPLPLQTRLLRVL EEREVVRVGGTRPIPVQVRIISATHCDLQARVREGRFRADLYYRLSV
208.0173	MPLALQGRLLRVLQERAVTPLGSRVIVPVDFAVLCATNQPLREMAARHAFREDLYYRLING
	: * : * * * . * . . : * . : : * : * * * * * * * * * : * : * :
172.0051	-
190.0072	F
NtrC	-
216.0019	-
281.0022	L
AtoC	-
Nifa	-
DctD	-
287.0556	-
PrpR	-
208.0173	-

Legenda: Hífens representam gaps, asteriscos aminoácidos conservados e .ponto e dois pontos aminoácidos com substituições conservadas.

Apêndice V – Alinhamento da Sequência de Aminoácidos dos Domínios HTH das Proteínas Reguladoras de Transcrição Dependentes de σ^{54} de *H. seropedicae*

172.0051	LERYERQLVSAALQRHGGNLAATARELGVDRANLNRLAKR-
AtoC	RDEAERKAVVKALTRMDGNILKAAELLGVSRTLYDLMSRH
208.0173	-QDNEWAAIERALRTHAGNVSAAAKALGVSRTIYRRLKSR
PrpR	--TPDNASIAAALRRFRGNRAAAARYLGMSRTTLWRKLKES
281.0022	LAEVERVMIERALQGAAGNRAEAARRLGISRQQLYRRLEEA
DctD	VEQFEAGLIAAALKDSEGSVAVAAERLGIKPKTLYDKIRKY
NtrC	GRQFEAALIKVALKHTHGRKNDAAVRLGIGRNTITRKIQEL
216.0019	-QKFERDYLIIRLLKLTGDNVSDAARLADRNRTEFYRLLQKN
191.0072	LPEMERRLIAEAIERAEGNLSAAARLLGISRAQLAYRQQKG
NifA	PPTSERERLIWAMEQCGWVQAKAARALNISPRQMGYALQKF
287.0556	FEKIEDTVFRTAFEFCERNQVQAARLLDISRNVLRARLIRN
	: . : :* . :

Legenda: Hífens representam gaps, asteriscos aminoácidos conservados e .ponto e dois pontos aminoácidos com substituições conservadas.

ApêndiceVI – Sequência do aptâmero 1 do domínio HTH da proteína**216.0019**

```
1  AAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAG  80
1  TFCGATAAATCCACTGTGATATCTTATGAGTTCGATACGTAGGTTGCGCAACCCTCGAGAGGGTATACCAGCTGGACGTC  80

81  GCGGCCGCGAATTCAGTACTAGTATTGGCAAAGCTTCCTGACGCGCTCGCGGCTGTTTACCGATGAGGAGTATGACGCCCT  160
81  CGCCGGCGCTTAAGTGATCACTAACCGTTTTTCGAAGGACTGCGCGAGCGCCGACAAATGGCTACTCCTCATACTGCGGGA  160

161  GGCCCCGCTCAGTGCCCGCGTGACTTGAGGCGCCGGTAGATGGTATTGCATATGTGTCTCAATCGAATTCGCGGGCCGC  240
161  CCGGGGCGAGTCACGGGCGCACTGAACTCCGCGCCATCTACCATAACGTATACACAGAGTTAGCTTAAGGGCGCCGGCG  240

241  CATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAG  290
241  GTACCGCCGGCCCTCGTACGCTGCAGCCCGGTTAAGCGGGATATCACTC  290
```

Apêndice VII – Sequência dos aptâmeros 2 e 3 do domínio HTH da proteína 216.0019

```
1 CTTTTTNGTNTATTACGCGGCAGNTTGGCTGNAAAGGGGGATGTGCTGGGCNAAGGCGATTAATGTTGGTTAACGGCC 80
1 GAAAAAANCANATAATGCGCCGTCNAACCGACNTTTCCTCCCTACACGACCCGNTTCCGCTAATTACAACCAATTGCGCG 80

81 AGGGTTTTTNCCATTTAGCGAAGTTTGGTAANAACGAGCGGNCAGTTNGAATTGTAATAGGACTTCACTATAGGGC 160
81 TCCCAAAAANGGGTAAAGTCGCTTCAAACCATNTTGTCTCGCCNGGTCAANCTTAACATTATCCTGAAGTGATATCCCG 160

161 GAATTGGGGCNCGGACGGTCGTCATGTTCCCGCCCCCATTTGGCGCCCGGGGAATTCGATTACCCCTCCATATGCGTG 240
161 CTTAACCCCGNGCCTGCCAGCAGTACAAGGGCGGGGGTAACCGCGGGCGCCCTTAAGCTAATGGGGAGGTATACGCAC 240

241 CGTTATGCCGAGGGCCAAGCAGAAGTTCGAGCGTGATTACCTGATCCCGCTTGCTCAAGTTGACTGATGGCAACGTGTC 320
241 GCAATACGGCCTCCCGGTTCTTCAAGCTCGCACTAATGGACTAGGGCGAACGAGTTCAACTGACTACCGTTGCACAG 320

321 TGACGCGCCCGGCTGGCTGATCGCAACCGTACCGAGTTCTATCGCCTGTTGCAGAAGAATGGACTGGATC 391
321 ACTGCGCGGGCCGACCGACTAGCGTTGGCATGGCTCAAGATAGCGGACAACGTCTTCTTACCTGACCTAG 391
```

Apêndice VIII – Sequência do aptâmero 4 do domínio HTH da proteína

216.0019

```

1  TATGNGCCAATATTGGGGACCATCANNNTTGAGNGNTGTAGTCCATGTGTCCANTCATNGATNCACCTCCACCACCTCGG  80
1  ATACNCGGTTATAACCCTGGTAGTNNAACTCNCNACATCAGGTACACAGGTNAGTACNCTANGTGGAGGTGGTGGAGCC  80

81  GAATGCAGGTGACNACATCGCGGAACTTCTTCANCCAGGAGGCGCTGTGCTGGCGCGTCTTGATCTCCACCAGAACCCTC  160
81  CTTACGTCCACTGNTGTAGCGCCTTGAAGAAGTNGGTCCTCCGCGACACGACCGCGCAGAACTAGAGGTGGTCTTGGCAG  160

161  ACCGAGACATTTAGTTTTTCGGCATCCANACACCACGCGGNGCTTGATGACACCTTCGGCCTCCAACCTCTGGATGCNA  240
161  TGGCTCTGTAAATCAAAAAGCCGTAGGTNGTGTGGTGCGCCNCGAACTACTGTGGAAGCCGGAGGTTGGAGACCTACGNT  240

241  NGCCAGCATGGGGTTGGTCGACAAGTTGACGCACTCGGCGATTTCCGGCCACAGNGCGCGTGNITGTCTTCTTGTAGGATCT  320
241  NCGGTCGTACCCCAACCAGCTGTTCAACTGCGTGAGCCGCTAAAGCCGGTGTTCNCGCGCACNACAGAAGAACATCCTAGA  320

321  CCAGGN      326
321  GGTCN      326

```

Apêndice IX – Sequência dos aptâmeros 5 e 6 do domínio HTH da proteína 216.0019

```
1 CCATATGNGCCAATATTGGGCACCATCACCTTGAGCATGTANTCGATGTGTCCACTCATGCGATGCACCTCCACCACCTC 80
1 GGTATACNCGGTATATAACCCGTGGTAGTGGAACCTCGTACATNAGCTACACAGGTGAGTACGCTACGTGGAGGTGGTGGAG 80

81 GGGAAATGCAGGTGACGACATCGCGGAACTTCTTCAGCCAGGAGGCGCTGTGCTGGCGCGTCTTGATCTCCACCAGAACCG 160
81 CCCTTACGTCCACTGCTGTAGCGCCTTGAAGAAGTCGGTCCTCCGCGACACGACCGCGCAGAACTAGAGGTGGTCTTGGC 160

161 TCACCGAGACATTGAGTTTTTTCGGCATCCAGCAGCACCACGCGGCGCTTGATGACACCTTCGGCCTCCAGCCTCTGGATG 240
161 AGTGGCTCTGTAACTCAAAAAGCCGTAGGTCGTCTGTTGGTGCGCCGGAACACTGTGGAAGCCGGAGGTCGGAGACCTAC 240

241 CGACGCCAGCATGGGGTGGTCGACAAGTTGACGCACTCGGCGATTTTCGGCCACAGCGCGCTGCTGTCTTCTTGTAGGAT 320
241 GCTGCGGTCGTACCCACAGCTGTTCAACTGCGTGAGCCGCTAAAGCCGGTGTGCGCGCACGACAGAAGAACATCCTA 320

321 CTCCAGGATCCAA 333
321 GAGGTCTAGGTT 333
```

**Apêndice X – Sequência do aptâmero 1 do domínio HTH da proteína
208.0153**

```
1 CAAGAATGCGCCGGTCTCGGGTGGCACGGTCAATGGGGTTCAGGTGGGTATCCGCCATCTGTTCTGAGTTTGTGATGTTT 80
1 GTTCTTACGCGGCCAGAGCCCACCGTGCCAGTTACCCCAAGTCCACCCATAGGCGGTAGACAAGACTCAAACACTACAAG 80

81 TGAAAAGGAGTCAAGCAATGGACAGCATCGACCTCGATGTACTIONCAGTGGCTTGAGGCTGGCCGTGGC 160
81 ACTTTTCCTCAGTTCGTTACCTGTCGTAGCTGGAGCTACATGAGTTNGCGTAAGTNGTCACCGAACTCCGACCCGCCACCG 160

161 GTNTTGCTGGCCACAGTGGTGCACCTGNGGTTTCNGCGCCACCGCCGGTCGGGGCNATACTTNCCTTNCGTG 233
161 CANAACGACCGGTGTCACCACGCGTGGACNCCAAGNCGCGGTGGCGGCCAGCCCCGNTATGAANGGAANGCAC 233
```

Apêndice XI – Sequência do aptâmero 2 do domínio HTH da proteína**208.0153**

```
1 GGGGATCCTGGCTCATTGTGGTCTTGTGGNGCTGGTGAACGAAAGTGGGTGCGGGGCGTCGTTTGATGGCCCCACTGCC 80
1 CCCCTAGGACCGAGTAACACCAGAACACCNCGACCACCTTTGACTTCACCCACGCCCCGAGCAAACCTACCGGGGTGACGG 80

81 TGGCTGCGGGGCAATTTCCACGCGGAAATGGTGCCGCACAGCAATTTTACCTAAAATTACAATCTCATGAAACAGGAACG 160
81 ACCGACGCCCCGTTAAAGGTGCGCCTTTACCACGGCGTGTTCGTTAAAATGGATTTTAAATGTTAGAGTAACTTGCCTTGC 160

161 TAACAGGTTGCGCTCGATCAGCGGCGATATGGCCAACCTTCGTTGCTCTGTTACACCCTCATGGGGAAGTTACACCAAAGA 240
161 ATTGTCCAACGCGAGCTAGTCGCCGCTATACCGGTTGAAGCAACGAGACAATGTGGGAGTACCCCTTCAATGTGGTTTCT 240

241 AATAAAAAGTCAATATCTGTTAAATATAAAAAGTCATCAGGGAACAAAAACAATCAACCGCGCAAAAGTATTAATATAA 320
241 TTATTTTTTCAGTTATAGACAATTTATATTTTCAGTAGTCCCTTGTTTTTGTAGTTGGCGCGTTTTTCATAATTTATATT 320

321 AACCAATAATTTCCCATCTAACAACCTTGGGGGAGCTTGCGTGGCCAACGACGACGCCGTTGTTGATCG 388
321 TTGGTTATTAAGGGTAGATTGTTGAACCCCTCGAACGCACCGGTTGCTGCTGCGGCAACAACACTAGC 388
```

Apêndice XII – Sequência do aptâmero 3 do domínio HTH da proteína 208.0153

```
1 GATCGATTTTAGCGGATTCCTGACATCGTGTGCGACGCGACACAAGCCACTGTTGCCATTCCGGCAAAGGCCGGCCCTG 80
1 CTAGCTAAAATCGCCTAAGGACTGTAGCACAAACGCTGCGCTGTGTTTCGGTGACAACGGTAAGGCCGTTTCCGGCCGGGAC 80

81 GCGATGCCCGCTGCGAAAAGCGTCGCTGGGAGTGTTTCAGATGGCGATGATGGAGGGATTTTAAAGAGTCCGACTCAAATAG 160
81 CGCTACGGGCGACGCTTTCGACGCGACCCTCACAAAGTCTACCGCTACTACCTCCCTAAAATTTCTCAGGCTGAGTTTATC 160

161 TCGGGAATTATCCGGGGCGACGCGCGAATGCCACAGAAAAAAGCAGCTTGCATCCAGGCAATGAAAAAAGGCCGCCGAAG 240
161 AGCCCTTAATAGGCCCCGCTGCGCGCTTACGGTGTCTTTTTTCGTCGAACGTAGGTCGGTTACTTTTTTCCGCGGGCTTC 240

241 GCGCCTTTGAAAGAGGCCGTCGCCGGCCCCGAAGGAATACTTGCAGACTCAGTCGATCAGTACGGCAGANGCTTTAGTGG 320
241 CGCGGAAACTTTCTCCGGCAGCGGCCGGGGCTTCTTATGAACGCTGAGTCAGCTAGTCATGCCGTCTNCGAAATCACC 320

321 TTAAGCCGATTCGCCGTNGGCTTGAAAATATCCAGGCCTTCGCGTTCCTTCTCCTCGGTCACACGCAGGCCGAT 394
321 AATTCGGCTAAGCGGCANCCGAACCTTATAGGTCCGGAAGCGCAAGAAGAAGGAGCCAGTGTGCGTCCGGCTA 394
```

Apêndice XIII – Sequência do aptâmero 4 do domínio HTH da proteína**208.0153**

```
1 CAGCATGCCAGCACGTCCATCCACGTGGCTGCCAGGAGATCGCGCAGGGCAACATCGATCTCTCGCAGCGCTTCGAAGAA 80
1 GTCGTACGGTCGTGCAGGTAGGTGCACCGACGGTCCTCTAGCGCGTCCCGTTGTAGCTAGAGAGCGTCGCGAAGCTTCTT 80

81 CAGGCCGCCTCGCTGGAAGAAACCGCCTCCAGCATGGAAGAAGTACCAGCACTGTGCAACAGAATGCCGGTNACGCCAA 160
81 GTCCGGCGGAGCGACCTTCTTTGGCGGAGGTCTACCTTCTTACTGGTCTGTGACACGTTGTCTTACGGCCANTGCGGTT 160

161 CAGGCCAGCAGCATGGCCAGCGCCGGTGCANAAGTGGCGCGCCGGGGCGAGGCCGAA 218
161 CGTCCGGTCGTCTACCGGTCTCGCGGCCACGTNTTACCGCGCGGCCCGCTCCGGCTT 218
```

**Apêndice XIV – Sequência do aptâmero 5 do domínio HTH da proteína
208.0153**

```
1 GATCTGGGTGCTCAACGCCAGCTCCAACCTGGCTCTTGCGCGTGGCAGGCCTGGGTGCCGGCCATGGTCATGACGCCNNCT 80
1 CTAGACCCACGAGTTGCGGTCGAGGTTGACCGAGAACGCGCACCGTCCGGACCCACGCCCGGTACCAGTACTGCGGNNGA 80

81 ATTCGTCCGACGAACTCAAGCTCATCCTGNGCGCCGGCAACAAGAGCGGCAANAACGGCAAGTTCACCCGCGACGAATGG 160
81 TAAGCAGGCTGCTTGAGTTCGAGTAGGACNCGCGCCGTTGTTCTCGCCGTTNTTGCCGTTCAAGTGGGCGCTGCTTACC 160

161 AATGTGCTGACGCAAAGCCTGAACTTCGCCGAGCTGGACGTGGCCGACATCATGCGTCCGGCCAGCGAGATC 232
161 TTACACGACTGCGTTTCGGACTTGAAGCGGCTCGACCTGCACCGGCTGTAGTACGCAGGCCGCGTCTAG 232
```

**Apêndice XV – Sequência do aptâmero 6 do domínio HTH da proteína
208.0153**

```
1 TACCTGATCCGTATTTTTCTCCAAGAGCAAATTCGCAACGGGATATCCAAGTTTTTAGCGCACTGTCGTTACAGAGCGCT 80
1 ATGGACTAGGCATAAAAAGAGGTTCTCGTTTAAGCGTTGCCCTATAGGTTCAAAAATCGCGTGACAGCAAGTGCTCGCGA 80

81 ATCGCTATCGAAACCAAAAAGACGAAACAAGAAATGGCCAGAAAATACCGCCGATAGTGCTGTAGGCAGCGATGGAAAT 160
81 TAGCGATAGCTTTGGTTTTTTCTGCTTTGTTCTTTACCGGCTTTTATGGCGGCTATCACGACATCCGTCGCTACCTTTA 160

161 TCCAGCAGAAAGCGTTCATTTCGCTAGATACAACCAACCAAGGAGCTGTTTTTTTTCTCAAATGCCACGCGAGCGCTAA 240
161 AGGTCGTCTTTCGCAAGGTAAGCGATCTATGTTGGTTGGTTCCTCGACAAAAAAGAGGTTTACGGTGCCTCGCGATT 240

241 CGGCAATATCGAAAGCGCCCGCCAGGATGCTGCCCCATACCTGCCAGACGGGCCCCCTGGAT 303
241 GCCGTTATAGCTTTCGCGGGCGGTCCTACGACGGGGTATGGACGGTCTGCCCGGGGACCTA 303
```

Apêndice XVI – Sequência do aptâmero 7 do domínio HTH da proteína**208.0153**

```
1 CGTTGATCGACTTCAATACACTGTCGGCGCAGATGCCCAACGAGACACGTAACCTACGTGCCCAAGCTGCAGGCGGTCAAG 80
1 GCAACTAGCTGAAGTTATGTGACAGCCGCTCTACGGGTTGCTCTGTGCATTGATGCACGGGTTTCGACGTCCGCCAGTTC 80

81 AACATCATCGCCAACCCGGCTGCCTTCGACATCGCCCTGCCCAAGGTGGACAACCAGCCCTACTTCGTCAACATCGGCAA 160
81 TTGTAGTAGCGGTTGGGCCGACGGAAGCTGTAGCGGGACGGGTTCCACCTGTTGGTCGGGATGAAGCAGTGGTAGCCGTT 160

161 GACCCGCAACATCGACGTGAAGGTGCGGGCGCAACTGGCCGAGCTGTCGCTGGACGAGTTCAAGGNCTTGAACCCGCAAT 240
161 CTGGGCGTTGTAGCTGCACTTCCAGCGCCGCTTGACCGGCTCGACAGCGACCTGCTCAAGTTCNGAACTTGGGCGTTA 240

241 TCAACCGCCCGGTTCATCATCGGCGGNGAANACACCAAGATCC 282
241 AGTTGGCGGGCCAGTAGTAGCCGCCNCTTNTGTGGTTCTAGG 282
```

**Apêndice XVII – Sequência do aptâmero 8 do domínio HTH da proteína
208.0153**

```
1 ATCGGCGGGTCTGGGTTGAGCTTGGNCTGCAGGAACATCGAAGCAGCCATGATGATGGGCAACAGACCGATGGGNNTGTT 80
1 TAGCCGCCCCAGCCCAACTCGAACCNGACGTCCTTGTAGCTTCGTTCGGTACTACTACCCGTTGTCTGGCTACCCNACAA 80

81 GAATTACGGGATGGTGCCAAACAGGCGGTTCAGGCGCGGTTCAGGTCATGCACCCAGCCAGCCAGTGCAGCGCCGCGCATT 160
81 CTTAATGCCCTACCACGGTTTGTCCGCCAGTCCGCGCCAGTCCAGTACGTGGGTCGGGTCAGCGCGCCGCGTAAA 160

161 CCACCGACGCCAGCAGCACCGAGTACAGCGCGATGAACACCGGGATCCT 209
161 GGTGGCTGCGGTCGTCTGGCTCATGTCGCGCTACTTGTGGCCCTAGGA 209
```

Apêndice XVIII – Sequência do aptâmero 9 do domínio HTH da proteína**208.0153**

```
1 CCTTGATCATCTACATCGATAGTCAGAAGGAACTGATGTGGGATTAACGCGCTACCCGGGAGGNTGACGCTGCGCGCTTC 80
1 GGAAC TAGTAGATGTAGCTATCAGTCTTCCTTGACTACACCCTAATTGCGCGATGGGCCCTCCNACTGCGACGCGGAAG 80

81 GCCCGTGCTCAGGGTAAAGTAATGGTCAGATTGGCAGTGC GCGGTGCCAATTTGACGATATCTGAATAGCCGGCCAGCCC 160
81 CGGGCACGAGTCCCATTTCATTACCAGTCTAACCGTCACGCGCCACGGTTAAACTGCTATAGACTTATCGGCCGGTCGGG 160

161 ATTTTCGGTGTGCTGATAAGCATCCAGCGCACTCCAGATAGGCGAACAAGGTGATCCT 220
161 TAAAAGCCACAGCGACTATTCGTAGGTCGCGTGAGGGTCTATCCGCTTGTTCCACTAGGA 220
```

**Apêndice XIX – Sequência do aptâmero 1 do domínio HTH da proteína
287.0556**

```
1 ATCTGTGTCAGCGTCGGCCCCGGCGGGGAACTAGCACCTCATGCGCGGCATGAGGTCTTACTTGGCTGCCGCGTTGCTG 80
1 TAGACACAGTCGCAGCCGGGGCCGCCCTTGATCGTGGAGTACGCGCCGTACTCCAGAATGAACCGACGGCGCAACGAC 80

81 GGCTTGTGGAAAGACTCCACACATAGGCAGCCAGCACATGCACCTTGGGTTACCCGAGGAACTCGCCAAACGCCGGCAT 160
81 CCGAACAACTTTCTGAGGTGTGTATCCGTCGGTCGTGTACGTGGAACCAAGTGGCTCCTTGAGCGGTTTGCGGCCGTA 160

161 GGTATTGTCGCGACCCTTGTTGATGGTTTCCTTGATGGTGTGATACTGCCGCCATACAGCCAGATCT 228
161 CCATAACAGCGCTGGGAACAACCTACCAAAGGAACTACCACAGCTATGACGGCGGTATGTCGGTCTAGA 228
```