

EDUARDO LUIZ VOIGT

IDENTIFICAÇÃO E ANÁLISE ESTRUTURAL DE GENES A MONTANTE DO GENE *nifA* EM *Herbaspirillum seropedicae*

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

Orientadora:

Prof.^a Dr.^a Maria Berenice Reynaud Steffens

Co-orientador

Prof. Dr. Fábio de Oliveira Pedrosa

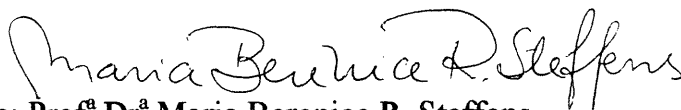
CURITIBA

2000

EDUARDO LUIZ VOIGT

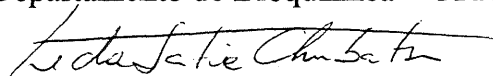
IDENTIFICAÇÃO E ANÁLISE ESTRUTURAL DE GENES A MONTANTE DO
GENE *nifA* EM *Herbaspirillum seropedicae*

Dissertação aprovada como requisito parcial para a obtenção de grau de Mestre no Curso de Pós-Graduação em Ciências - **Bioquímica** da Universidade Federal do Paraná, pela Comissão formada pelos professores:



Orientadora: Prof^a Dr^a Maria Berenice R. Steffens

Departamento de Bioquímica – UFPR



Prof^a Dr^a Leda Satie Chubatsu

Departamento de Bioquímica – UFPR



Prof^a Dr^a Liu Un Rigo

Departamento de Bioquímica – UFPR



Prof^a Dr^a Chirlei Glienke de Blanco

Departamento de Genética – UFPR

Curitiba, 28 de agosto de 2000.

Orientadora: Prof^a Dr^a Maria Berenice R. Steffens
Co-orientador: Prof. Dr. Fábio de Oliveira Pedrosa

Aos meus pais Rolf e Crista, com muito carinho.

AGRADECIMENTOS

Aos professores Maria Berenice R. Steffens e Fábio de Oliveira Pedrosa pelo apoio e pela orientação deste trabalho.

Aos professores Liu Un Rigo, Emanuel M. de Souza, Leda S. Chubatsu, Elaine M. Benelli, Marshall G. Yates, Iara Maria P. Machado e Hidevaldo Bueno Machado pela atenção e pelos auxílios prestados.

Às professoras Glaci T. Zancan, Maria Benigna M. de Oliveira e Maria Eliane Merlin pelo exemplo, pelo apoio e pelo incentivo.

Aos colegas de laboratório Adriana, Ana Cláudia, Andrea, Carolina, Cláudia, Cynthia, Darlene, Fabiane, Fernando, Giseli, Humberto, Igor, Josiane, Juliana Inaba, Juliana Ramos, Juliana Soares, Lauren, Leonardo, Lilian, Luciano, Luíza, Mariana, Patrícia, Raquel, Rose Adele e Roseli pelo companheirismo e pelos auxílios prestados.

Aos funcionários Julieta, Roseli e Válter pelo companheirismo e pelos serviços prestados.

Aos funcionários da Universidade Federal do Paraná pelos serviços prestados.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida e ao Conselho Nacional de Ensino e Pesquisa (CNPq) pelo financiamento deste trabalho.

Aos colegas de curso Ana Paula, Andrea, Ângela, Clarice, Daniele, Eliana, Fábio, Fabíola Iagher, Fabíola Stevan, Gisele, Jean, João, Júlio, Juan, Luciane e Sílvia pela amizade e pelo apoio.

Aos amigos Alcionete, Alexandre, Áurea, Carla, Carlos, Carolina, Daniela, Desirê, Dirceu, Édson, Josy, Leopoldo, Marcelo, Margherita, Noeli, Rodrigo e Sheila pelo companheirismo, pelo apoio e pelo incentivo.

Aos meus pais Rolf e Crista, aos meus irmãos Márcio e Ricardo e à minha cunhada Daniela pelo carinho, pelo apoio e pelo incentivo.

SUMÁRIO

LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	x
ABREVIATURAS.....	xi
RESUMO.....	xii
ABSTRACT.....	xiii
1 INTRODUÇÃO.....	1
1.1 ESTRUTURA DOS GENES DE PROCARIOTOS.....	1
1.1.1 Sequências estruturais.....	1
1.1.2 Promotores.....	2
1.1.3 Terminadores.....	6
1.2 <i>Herbaspirillum seropedicae</i>	8
1.3 ESTRUTURA DOS GENES IDENTIFICADOS EM <i>H. seropedicae</i>	9
1.4 OBJETIVOS.....	13
1.4.1 Objetivo geral.....	13
1.4.2 Objetivos específicos.....	13
2 MATERIAL E MÉTODOS.....	14
2.1 ESTIRPES BACTERIANAS E PLASMÍDEOS.....	14
2.2 MEIOS DE CULTURA.....	15
2.3 CONDIÇÕES DE CULTIVO E ESTOCAGEM DE BACTÉRIAS.....	16
2.4 ANTIBIÓTICOS.....	16
2.5 REAGENTES.....	17
2.6 SOLUÇÕES.....	18
2.6.1 Soluções de fenol tamponado, clorofórmio-álcool isoamílico e fenol- clorofórmio-álcool isoamílico.....	18

2.6.2 Solução de RNase.....	18
2.6.3 Outras soluções.....	19
2.7 MINIPREPARAÇÃO DE PLASMÍDEOS.....	19
2.8 MAXIPREPARAÇÃO DE PLASMÍDEOS.....	20
2.9 CLIVAGEM DE DNA COM ENDONUCLEASES DE RESTRIÇÃO.....	20
2.10 ELETROFORESE DE DNA EM GEL DE AGAROSE.....	21
2.11 SUBCLONAGEM DE DNA.....	22
2.12 TRANSFORMAÇÃO BACTERIANA.....	23
2.13 SEQÜENCIAMENTO DE DNA.....	24
2.13.1 Obtenção de deleções unidirecionais.....	24
2.13.2 Obtenção de DNA-molde.....	26
2.13.3 Reações de seqüenciamento.....	26
2.14 EDIÇÃO E ANÁLISE DE SEQÜÊNCIAS.....	27
3 RESULTADOS E DISCUSSÃO.....	29
3.1 OBTENÇÃO DE SUBCLONES PARA SEQÜENCIAMENTO DE DNA...	29
3.2 SEQÜENCIAMENTO E ANÁLISE ESTRUTURAL DA REGIÃO A MONTANTE DO GENE <i>nifA</i>	35
3.2.1 <i>orf4</i>	43
3.2.2 Gene <i>modE</i>	45
3.2.3 <i>orf5</i>	53
3.2.4 <i>orf6</i>	60
4 CONCLUSÕES.....	67
5 REFERÊNCIAS BIBLIOGRÁFICAS.....	69
6 ANEXOS.....	82

LISTA DE FIGURAS

Figura 1. Comparação estrutural entre promotores hipotéticos dependentes da subunidade sigma 70 (-35/-10) ou sigma 54 (-24/-12).....	4
Figura 2. Localização das seqüências regulatórias de um gene hipotético que apresenta um promotor dependente da subunidade sigma 54.....	5
Figura 3. Estrutura de um terminador intrínseco de <i>E. coli</i>	7
Figura 4. Mapa de restrição do plasmídeo pEMS1 e de plasmídeos derivados....	29
Figura 5. Mapa de restrição do plasmídeo pELV2B e de plasmídeos derivados..	30
Figura 6. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV2C	32
Figura 7. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV3B	33
Figura 8. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV3C	34
Figura 9. Seqüência de nucleotídeos da região seqüenciada a montante do gene <i>nifA</i> em <i>Herbaspirillum seropedicae</i>	36
Figura 10. Identificação das prováveis regiões codificadoras de proteína dos genes <i>orf4</i> , <i>modE</i> e <i>orf5</i> de <i>Herbaspirillum seropedicae</i> através do método de preferência de códons do programa Analyseq (STADEN, 1982).....	40
Figura 11. Identificação da provável região codificadora de proteína do gene <i>orf6</i> de <i>Herbaspirillum seropedicae</i> através do método de preferência de códons do programa Analyseq (STADEN, 1982).....	41
Figura 12. Representação das ORFs identificadas em <i>Herbaspirillum seropedicae</i> a montante do gene <i>nifA</i>	42
Figura 13. Seqüência de nucleotídeos da região distal da <i>orf4</i> e seqüência de aminoácidos deduzida da região C-terminal da proteína Orf4 de <i>Herbaspirillum seropedicae</i>	43

Figura 14. Seqüência de nucleotídeos do gene <i>modE</i> e seqüência de aminoácidos deduzida da proteína ModE de <i>Herbaspirillum seropedicae</i>	46
Figura 15. Comparação da organização estrutural dos genes <i>mod</i> e/ou <i>mop</i> em <i>E. coli</i> , <i>H. influenzae</i> , <i>A. vinelandii</i> , <i>R. capsulatus</i> , e <i>H. seropedicae</i>	49
Figura 16. Alinhamento da seqüência de aminoácidos das proteínas ModE de <i>H. seropedicae</i> (Herse_ModE), de <i>A. vinelandii</i> (Azovi_ModE), de <i>H. influenzae</i> (Haein_ModE) e de <i>E. coli</i> (Ecoli_ModE) e das proteínas MopB (Rhoca_MopB) e MopA (Rhoca_MopA) de <i>R. capsulatus</i>	51
Figura 17. Seqüência de nucleotídeos da <i>orf5</i> e seqüência de aminoácidos deduzida da proteína Orf5 identificada em <i>Herbaspirillum seropedicae</i>	54
Figura 18. Alinhamento da seqüência de nucleotídeos da <i>orf5</i> identificada em <i>H. seropedicae</i> (Herse_orf5) e do gene <i>trs5</i> de <i>E. coli</i> (Ecoli_trs5).....	57
Figura 19. Seqüência de nucleotídeos da região distal da <i>orf6</i> e seqüência de aminoácidos deduzida da região C-terminal da proteína Orf6 de <i>Herbaspirillum seropedicae</i>	61
Figura 20. Alinhamento da seqüência de aminoácidos da região C-terminal da proteína Orf6 de <i>H. seropedicae</i> (Herse_Orf6) e das proteínas CyaA de <i>B. pertussis</i> (Borpe_CyaA), CyaA de <i>B. bronchiseptica</i> (Borbr_CyaA) e FrpC (Neime_FprC) e FprA (Neime_FprA) de <i>N. meningitidis</i>	64
Figura 21. Mapa de restrição do vetor pVK100.....	83
Figura 21. Mapa de restrição do vetor pTZ19R.....	84
Figura 21. Mapa de restrição do vetor pSport1.....	85

LISTA DE TABELAS

Tabela 1. Estirpes bacterianas.....	14
Tabela 2. Plasmídeos.....	15

ABREVIATURAS

ATP	-	adenosina trifosfato
bp	-	pares de bases nucleotídicas
DNA	-	ácido desoxirribonucléico
DNAse	-	desoxirribonuclease
dNTPs	-	desoxirribonucleosídeos trifosfato
DO	-	densidade óptica
EDTA	-	ácido etilenodiamino-tetracético
IHF	-	fator de integração de hospedeiro, do inglês <i>integration host factor</i>
kb	-	quilopares de bases nucleotídicas
ORF	-	seqüência codificadora de proteína, do inglês <i>open reading frame</i>
RBS	-	sítio de ligação de ribossomo, do inglês <i>ribosome-binding site</i>
RNA	-	ácido ribonucléico
RNAse	-	ribonuclease
rpm	-	rotações por minuto
RTX	-	repetições em toxina, do inglês <i>repeats in toxin</i>
SDS	-	dodecil sulfato de sódio
Tris	-	tris (hidroximetil)-aminometano
U	-	unidade de atividade enzimática
UAS	-	seqüência ativadora a montante, do inglês <i>upstream activator sequence</i>

RESUMO

Herbaspirillum seropedicae é uma bactéria diazotrófica endofítica encontrada nos tecidos de membros das Poaceae (BALDANI *et al.*, 1997), Ananaceae e Musaceae (WEBER *et al.*, 1999). O gene *nifA*, que codifica para um ativador de transcrição dos genes *nif*, foi isolado de uma biblioteca genômica de *H. seropedicae* Z78 e identificado no plasmídeo pEMS1 (SOUZA *et al.*, 1990). Com o objetivo de identificar genes localizados na região a montante de *nifA*, dois fragmentos contíguos *Bam*HI/*Sal*I de 3,5 Kb, correspondentes a esta região, foram subclonados. Deleções unidirecionais destes fragmentos foram seqüenciadas e alinhadas, resultando em uma seqüência de 4777 bp. Nesta seqüência, foram identificadas quatro prováveis ORFs denominadas *orf4*, *modE*, *orf5* e *orf6*. A região distal da *orf4* e o gene *modE* foram encontrados a cerca de 1,3 e 1,8 kb de *nifA*, respectivamente, fazendo parte de um possível operon. A *orf4* codifica para uma proteína com função desconhecida. O gene *modE* codifica para um potencial repressor de transcrição dos genes *modABC*, envolvidos com o transporte de molibdênio. A seqüência de aminoácidos deduzida da proteína ModE apresenta um possível motivo hélice-volta-hélice de ligação ao DNA e dois motivos característicos das proteínas ModE (SARNQ), localizados nos prováveis domínios MopE1 e MopE2, envolvidos com a ligação de íons molibdato. A *orf5* foi localizada a aproximadamente 3,7 kb de *nifA*. A sua seqüência deduzida de aminoácidos se mostrou idêntica à da transposase Trs5, codificada pela seqüência de inserção IS5 de *E. coli*. Esta seqüência pode ter sido adquirida por transferência horizontal ou a partir do DNA cromossômico da célula hospedeira do plasmídeo pEMS1, *E. coli* HB101. A região distal da *orf6* foi identificada a cerca de 5,3 kb de *nifA* e codifica para uma proteína semelhante às citotoxinas RTX, produzidas por patógenos de animais. Este gene pode ter sido herdado de um ancestral comum com espécies do gênero *Pseudomonas* ou adquirido por transferência horizontal de espécies filogeneticamente distantes.

ABSTRACT

Herbaspirillum seropedicae is an endophytic diazotrophic bacterium found in the tissues of several Poaceae (BALDANI *et al.*, 1997), Ananaceae and Musaceae (WEBER *et al.*, 1999). The *nifA* gene, which encodes for a transcriptional activator of the *nif* genes, was isolated from a genomic library of *H. seropedicae* and identified in the plasmid pEMS1 (SOUZA *et al.*, 1990). In order to identify genes located upstream from *nifA*, two contiguous 3.5 kb *Bam*HI/*Sal*I fragments corresponding to this region were subcloned. Unidirectional deletions from these fragments were sequenced and aligned, yielding a 4777 bp sequence. This sequence presented four putative ORFs, namely *orf4*, *modE*, *orf5* and *orf6*. The distal region of *orf4* and the *modE* gene were located about 1.3 and 1,8 kb from *nifA*, respectively, forming a putative gene cluster. The *orf4* encodes for a protein whose function was not determined and *modE* encodes for a potential transcriptional repressor of the *modABC* genes, involved in molybdenum uptake. The deduced amino acid sequence of the ModE protein has a putative helix-turn-helix DNA-binding motif and two ModE characteristic motifs (SARNQ), located in the putative MopE1 e MopE2 domains, which are probably involved with molybdate binding. The *orf5* was found approximately 3.7 kb from *nifA*. Its deduced amino acid sequence showed to be identical to the transposase Trs5 codified by the insertion sequence IS5 from *E. coli*. This sequence was probably acquired by horizontal transfer or from the chromosomal DNA of the host cell of the plasmid pEMS1, *E. coli* HB101. The distal region of the *orf6* was located about 5.3 kb from *nifA* and encodes for a protein similar to the RTX cytotoxins produced by animal pathogens. This gene could be acquired from a common ancestral with species of the genus *Pseudomonas* or by horizontal transfer.

1 INTRODUÇÃO

1.1 ESTRUTURA DOS GENES DE PROCARIOTOS

Os genes de procariotos estão distribuídos em seu genoma como unidades transcricionais independentes ou agrupadas, formando operons. Os operons são constituídos por genes transcritos e traduzidos conjuntamente, cujos produtos geralmente cumprem funções relacionadas (LEWIN, 1997).

Cada gene apresenta uma sequência estrutural, que codifica para um polipeptídeo (ORF), um RNA ribossômico (RNAr) ou um RNA transportador (RNAt), e sequências regulatórias, as quais são responsáveis pelo controle da expressão da sequência estrutural. A montante da sequência estrutural está localizado o promotor, envolvido com o início da transcrição, e a jusante, o terminador, relacionado com o término da mesma (HIPPEL, 1998).

1.1.1 Sequências estruturais

As sequências estruturais possuem tamanho variado e não contêm espaçadores. O início de transcrição corresponde, muitas vezes, à base A do motivo CAT e é considerada posição +1. As posições a montante são consideradas negativas e as posições a jusante, positivas (McCLURE, 1985). A jusante do início de transcrição, estão localizados o sítio de ligação de ribossomo (RBS), o códon de início de tradução e o códon de término de tradução. O sítio de ligação de ribossomo (sequência de Shine-Dalgarno) possui a sequência consenso AGGAGG, a qual é complementar a uma região do RNAr 16S da subunidade menor do ribossomo e auxilia no posicionamento correto do RNAm para o início da tradução. O códon de início de tradução corresponde à sequência ATG ou, alternativamente, GTG ou TTG, enquanto o códon de término de tradução corresponde às sequências TAG, TAA ou TGA (LEWIN, 1997).

As seqüências estruturais de um mesmo organismo ou de organismos diferentes apresentam variações na utilização de códons que codificam para um mesmo aminoácido (códons sinônimos), indicando preferência de códons (BULMER, 1987; PARKER, 1989; ANTEZANA e KREITMAN, 1999). A preferência de códons reflete a abundância dos RNAt correspondentes conforme as condições fisiológicas ou o organismo (SAIER, 1995). Genes expressos com elevada frequência geralmente apresentam maior número de códons comuns, conferindo maior eficiência à sua tradução (BULMER, 1987; PARKER, 1989; ANTEZANA e KREITMAN, 1999). Genes envolvidos com funções não essenciais para a célula, como a esporulação e a utilização de fontes alternativas de carbono e energia, podem apresentar maior número de códons raros. Assim sendo, a expressão destes genes ocorre em condições fisiológicas específicas, nas quais são expressos os RNAt correspondentes (SAIER, 1995).

1.1.2 Promotores

O promotor consiste na região de ligação da RNA polimerase, enzima responsável pelo processo de transcrição. Em procariotos, esta enzima apresenta um núcleo formado por duas subunidades alfa (α) e as subunidades beta (β) e beta' (β'). Este núcleo ($\alpha_2\beta\beta'$) contém o sítio catalítico da enzima e atua, portanto, na síntese de RNA. Ao núcleo da RNA polimerase se associa a subunidade sigma (σ), formando a holoenzima. A subunidade sigma é responsável pelo reconhecimento das seqüências específicas do promotor (McCLURE, 1985).

Durante o início do processo de transcrição, a holoenzima se liga inicialmente ao DNA dupla fita de modo reversível, formando o complexo fechado. Em seguida, a mesma separa as cadeias do DNA e inicia uma síntese abortiva de RNA, originando o complexo aberto. A síntese efetiva de RNA inicia com a liberação da subunidade sigma e a estabilização do complexo núcleo-DNA, no qual uma das fitas do DNA é utilizada como molde para a polimerização de nucleotídeos trifosfato (McCLURE, 1985).

A maioria dos promotores procarióticos são dependentes da subunidade sigma 70 (σ^{70}), apresentando as seqüências consenso TTGACA e TATAAT localizadas nas posições -35 e -10, respectivamente, separadas por 16-18 bp (TTGACA-N₁₆₋₁₈-TATAAT) (figura 1) (FASSLER e GUSSIN, 1996; HERTZ e STORMO, 1996). O motivo TTG da seqüência -35 e as bases TA---T da seqüência -10 são conservados, sugerindo que estas bases são essenciais para o reconhecimento pela subunidade sigma 70. A seqüência -35 está envolvida principalmente com a ligação da RNA polimerase e seqüência -10, com a separação das cadeias do DNA. Os genes que estão sob o controle destes promotores possuem funções diversificadas, incluindo aquelas essenciais para a sobrevivência da célula (McCLURE, 1985).

Além dos promotores dependentes da subunidade sigma 70, existem promotores dependentes de outras subunidades sigma, como sigma 32 (σ^{32}) (TAYLOR *et al.*, 1984), sigma 28 (σ^{28}) (HELMANN e CHAMBERLIN, 1987) e sigma 54 (σ^{54}) (SHINGLER, 1996). Assim como os promotores dependentes de sigma 70, os dependentes de sigma 32 ou 28 possuem seqüências consenso localizadas nas posições -35/-10 (TAYLOR *et al.*, 1984; HELMANN e CHAMBERLIN, 1987). Porém, nos promotores dependentes de sigma 54, estas seqüências estão localizadas nas posições -24/-12 (REITZER e MAGASANIK, 1985). Os genes que apresentam promotores dependentes de sigma 32, 28 ou 54 possuem funções não essenciais para a sobrevivência da célula, sendo transcritos em condições fisiológicas ou ambientais específicas (TAYLOR *et al.*, 1984; HELMANN e CHAMBERLIN, 1987; LORENZO e PÉREZ-MARTÍN, 1996).

Os promotores dependentes de sigma 32 possuem as seqüências consenso -35/-10 CCCTTGAA-N₁₃₋₁₅-CCCGATNT. Estes promotores estão presentes em genes expressos em condições de estresse, como choque térmico (TAYLOR *et al.*, 1984).

Os promotores dependentes de sigma 28 são caracterizados pela presença das seqüências consenso -35/-10 CTAAA-N₁₅-GCCGATAA. Os genes envolvidos com quimiotaxia e com a formação do flagelo estão sob o controle destes promotores (HELMANN e CHAMBERLIN, 1987).

Os promotores dependentes de sigma 54 apresentam as seqüências consenso -24/-12 TGGCAC-N₅-TTGC (figura 1) (REITZER e MAGASANIK, 1985). O motivo GG da seqüência -24 é invariável e o motivo GC da seqüência -12 é conservado, sugerindo que estes motivos são essenciais para o reconhecimento pela subunidade sigma 54 (SHINGLER, 1996). Estes promotores controlam a expressão de genes relacionados com diversos processos, como a fixação e o metabolismo de nitrogênio, a utilização de fontes alternativas de carbono e energia, o transporte de aminoácidos e ácidos dicarboxílicos e a produção de estruturas extracelulares ou determinantes de virulência (KUSTU *et al.*, 1991; LORENZO e PÉREZ-MARTÍN, 1996).

Outras seqüências regulatórias podem ser encontradas a montante das seqüências -35/-10 ou -24/-12, como os elementos UP (BUSBY e EBRIGHT, 1994) e os sítios para ligação de ativadores (KUSTU *et al.*, 1991) e repressores de transcrição (FASSLER e GUSSIN, 1996; GRALLA, 1996). Os elementos UP são seqüências ricas em A e T de aproximadamente 20 bp, localizadas imediatamente a montante da seqüência -35. O domínio C-terminal da subunidade α da RNA polimerase (α CTD) interage com esta seqüência, promovendo a transcrição. Os elementos UP estão presentes nos genes que codificam para os RNAr (BUSBY e EBRIGHT, 1994).

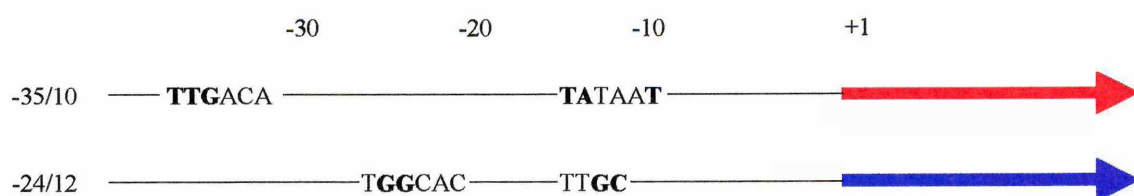


Figura 1. Comparação estrutural entre promotores hipotéticos dependentes da subunidade sigma 70 (-35/-10) ou sigma 54 (-24/-12). As seqüências consenso estão destacadas. As bases conservadas estão em negrito. As setas indicam a direção de transcrição das seqüências estruturais. Baseado em McCLURE (1985) e SHINGLER (1996).

Os ativadores de transcrição interagem com o α CTD ou a subunidade sigma da RNA polimerase. Os ativadores que se ligam a sítios próximos do promotor (entre as posições -60 e -40) interagem diretamente com a RNA polimerase, aumentando a sua afinidade pelo mesmo (GRALLA, 1996). Já os ativadores que se ligam a sítios mais distantes (entre as posições -200 e -100) interagem com esta enzima através do dobramento do DNA, favorecendo a transição de complexo fechado para complexo aberto (KUSTU *et al.*, 1991; SHINGLER, 1996). Estes sítios mais distantes são conhecidos como seqüências ativadoras a montante (UAS) (figura 2) e são constituídos por repetições invertidas (seqüências palindrômicas) (SHINGLER, 1996). A jusante da UAS, pode ser encontrado um sítio para ligação do fator de integração de hospedeiro (IHF) (figura 2), uma proteína que favorece o dobramento do DNA e, portanto, a interação entre o ativador e a RNA polimerase (HOOVER *et al.*, 1990; LORENZO e PÉREZ-MARTÍN, 1996).

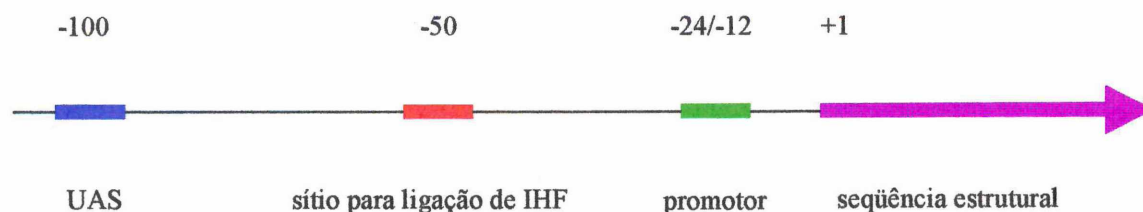


Figura 2. Localização das seqüências regulatórias de um gene hipotético que apresenta um promotor dependente da subunidade sigma 54. O promotor, um sítio para ligação de IHF e uma UAS estão destacados. A seta indica a direção de transcrição da seqüência estrutural. Adaptado de KUSTU *et al.* (1991).

Os repressores de transcrição se ligam a sítios denominados operadores, próximos ou sobrepostos aos promotores (entre as posições -60 e +12) (McCLURE, 1985). A ligação dos repressores impede a ligação da RNA polimerase ou a formação do complexo aberto (FASSLER e GUSSIN, 1996; GRALLA, 1996). Os repressores que possuem alta afinidade pelos seus respectivos operadores podem suprimir completamente a transcrição de certos genes conforme as condições fisiológicas. Entretanto, os repressores que apresentam baixa afinidade podem atuar apenas na modulação negativa da transcrição dos genes que controlam (McCLURE, 1985).

1.1.3 Terminadores

O terminador corresponde à região de término de transcrição. São conhecidos dois tipos de terminadores, os intrínsecos e os dependentes de Rho (PLATT, 1986; HENKIN, 1996; HIPPEL, 1998). Os terminadores intrínsecos, ou independentes de Rho, apresentam uma sequência palindrômica rica em G e C de 20-25 bp seguida por uma sequência de 6 a 8 resíduos de uridina (figura 3). A sequência palindrômica origina uma estrutura secundária em forma de grampo que interage com a RNA polimerase, causando uma pausa em sua atividade. Já a sequência de uridinas consiste na região de incorporação dos últimos nucleotídeos e forma um híbrido menos estável com o DNA-molde no sítio ativo da RNA polimerase (WILSON e HIPPEL, 1995). Deste modo, estas sequências contribuem para a desestabilização do complexo RNA polimerase-RNA-DNA, resultando na liberação do RNA nascente.

Os terminadores dependentes de Rho são sequências pouco conservadas de aproximadamente 40 bp, ricas em C e pobres em G, que interagem com o fator Rho. O fator Rho é uma proteína hexamérica em forma de anel que apresenta atividade de RNA-DNA helicase dependente da hidrólise de ATP. Cada monômero possui um domínio para ligação de RNA e um para ligação de ATP. Aparentemente, o fator Rho se liga ao RNA nascente em regiões carentes em estruturas secundárias a montante do terminador.

Através da sua atividade, o fator Rho se desloca ao longo do RNA nascente na direção 5'→3', alcançando o terminador, no qual promove a desestabilização do híbrido RNA-DNA e a liberação do RNA nascente. A ligação do ribossomo ao RNA nascente impede a ligação do fator Rho, permitindo a transcrição de genes localizados a jusante do terminador (RICHARDSON, 1990).

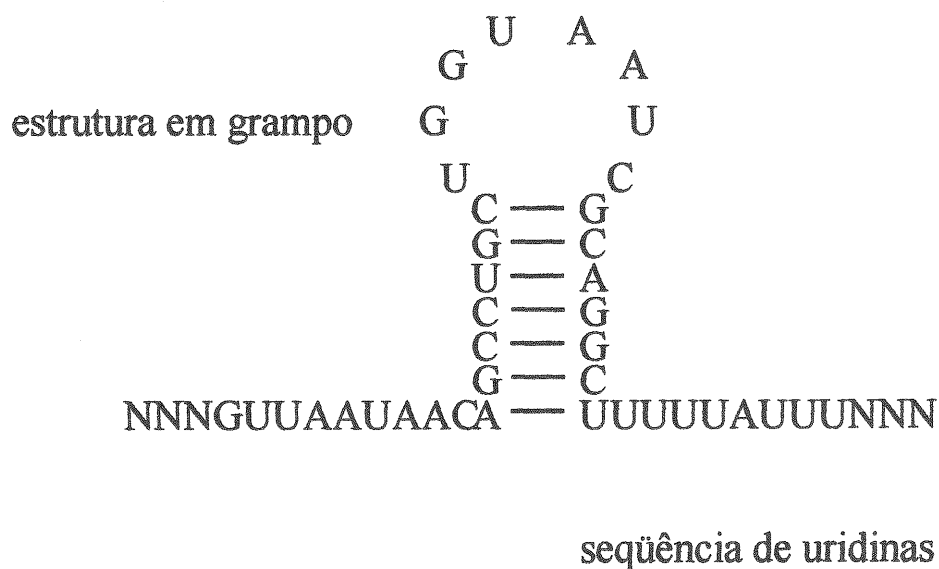


Figura 3. Estrutura de um terminador intrínseco de *E. coli*. Estão indicadas a estrutura em grampo e a seqüência de uridinas. Adaptado de WILSON e HIPPEL (1995).

1.2 *Herbaspirillum seropedicae*

Herbaspirillum seropedicae é uma bactéria Gram-negativa pertencente à divisão Proteobacteria, subdivisão Beta (YOUNG, 1992). Apresenta geralmente forma vibrióide, às vezes helicoidal, com um a três flagelos em um ou em ambos os polos. O diâmetro das células varia de 0,6 a 0,7 μm e o comprimento, entre 1,5 e 5,0 μm . Não fermenta açúcares, apresentando metabolismo respiratório típico. Ácidos orgânicos, como malato, fumarato, succinato, piruvato, citrato e trans-aconitato, são utilizados como fontes preferenciais de carbono. Manitol, sorbitol, glicerol e vários açúcares, como glucose, galactose e L-arabinose, também são oxidados. Fixa nitrogênio atmosférico em condições microaeróbias e assimila e desassimila nitrato e nitrito. Cresce entre 30 e 40°C e em pH variando de 5,3 a 8,0. O conteúdo de G + C do DNA é de $67 \pm 0,5$ mol%. É claramente diferenciado de *Azospirillum* spp. através de hibridização de RNA e DNA (BALDANI *et al.*, 1986). Pode ser encontrado em raízes, caules, folhas e sementes de membros das Poaceae (sorgo, arroz, cana-de-açúcar e plantas forrageiras) (BALDANI *et al.*, 1997), Ananaceae (abacaxizeiro) e Musaceae (bananeira) (WEBER *et al.*, 1999), sendo considerado um endófito obrigatório.

A potencial aplicação de *H. seropedicae* como fertilizante biológico está sendo estudada. PEREIRA *et al.* (1989) inocularam sorgo e arroz com as estirpes H25, Z89 e Z95 de *H. seropedicae* em condições de campo e verificaram um aumento significativo da germinação de sementes de arroz na presença das estirpes Z89 e Z95. Entretanto, não houve incremento significativo do peso seco de plantas de sorgo e arroz após 60 e 90 dias de inoculação, assim como nenhuma das três estirpes foi reisolada a partir das raízes das mesmas. Desta forma, não foi possível verificar a colonização efetiva das plantas inoculadas, nem estabelecer relações entre a inoculação e o desenvolvimento destas plantas. PIMENTEL *et al.* (1991) e JAMES *et al.* (1997) verificaram o estabelecimento da associação em folhas de sorgo, das quais *H. seropedicae* pôde ser reisolado, mas não avaliaram a eficiência da inoculação na promoção do crescimento vegetal. PEREIRA e

BALDANI (1995) observaram um aumento significativo na produção de grãos em plantas de arroz inoculadas com as estirpes Z89 e Z95 de *H. seropedicae*. Experimentos de inoculação de diferentes plantas hospedeiras com várias estirpes de *H. seropedicae* devem ainda ser feitos para uma avaliação mais aprofundada do seu efeito em relação ao crescimento vegetal.

O estudo dos genes relacionados com a fixação e o metabolismo de nitrogênio e a sua regulação (genes *nif*, *fix*, *gln* e *ntr*) (MERRICK e EDWARDS, 1995) pode auxiliar não apenas no esclarecimento destes processos em *H. seropedicae*, mas também na obtenção de estirpes geneticamente modificadas que sejam mais eficientes na promoção do crescimento vegetal. Estas estirpes poderiam ser utilizadas nas lavouras como inoculantes em substituição à adubação nitrogenada, proporcionando uma economia substancial nos gastos com esta adubação, já que *H. seropedicae* se associa a gramíneas cultivadas em larga escala.

1.3 ESTRUTURA DOS GENES IDENTIFICADOS EM *Herbaspirillum seropedicae*

Herbaspirillum seropedicae foi isolado recentemente e poucos grupos de pesquisa se dedicam a seu estudo, de modo que seu genoma é pouco conhecido. A maioria dos genes identificados em *H. seropedicae* estão envolvidos com a fixação e o metabolismo de nitrogênio e a sua regulação. Até o momento, foram identificados os genes *nifA* (SOUZA *et al.*, 1990), *nifB* (SOUZA *et al.*, 1991), *glnAntrBC* (TEIXEIRA, 1991), *recA* (STEFFENS *et al.*, 1993), *nifHDK* (MACHADO *et al.*, 1996), *glnB* (BENELLI *et al.*, 1997), *recX* (GALVÃO *et al.*, 1998), *amtB* (NOINDORF *et al.*, 1999), *nifENXorf1orf2* (KLASSEN *et al.*, 1999), *nifQmodABC* e *fixXC* (KLASSEN, 2000).

O gene *nifA* consiste em uma unidade transcricional independente que codifica para a proteína NifA, um ativador de transcrição dos genes de fixação de nitrogênio (genes *nif*) (SOUZA *et al.*, 1991). Este gene possui um promotor dependente de sigma 54, a montante do qual foram encontrados dois sítios para ligação da proteína NtrC (GCAC-

N₅-TGGTGCA), dois sítios e um hemi-sítio para ligação da proteína NifA (TGT-N₁₀-ACA) e um sítio para ligação de IHF (WATCAA-N₄-TTR, onde W = A/T e R = A/G) (SOUZA *et al.*, 1991; WASSEM *et al.*, 2000).

O operon *nifHDKENXorf1orf2* contém os genes estruturais da nitrogenase, *nifHDK*, os quais codificam para a subunidade γ da proteína Fe e as subunidades α e β da proteína MoFe desta enzima, respectivamente (MACHADO *et al.*, 1996). Este operon apresenta, ainda, os genes *nifENX*, relacionados com a fixação de nitrogênio, e os genes *orf1orf2*, cujas funções são desconhecidas (KLASSEN *et al.*, 1999). As proteínas NifE, NifN e NifX devem atuar na biossíntese do cofator ferro-molibdênio (FeMoco) da nitrogenase. A *orf1* codifica para uma proteína semelhante à ferredoxina encontrada em *Rhodobacter capsulatus* e *Rhizobium* sp. O produto da *orf2* não apresenta similaridade com nenhuma proteína conhecida (KLASSEN *et al.*, 1999; KLASSEN, 2000). A montante do gene *nifH* foram identificados um promotor dependente de sigma 54, dois sítios para ligação de NifA e um para ligação de IHF (MACHADO *et al.*, 1996).

O gene *nifB* está parcialmente sequenciado. Está localizado a jusante do gene *nifA* e é transcrito na mesma direção (SOUZA *et al.*, 1991). Codifica para a proteína NifB, possivelmente envolvida no processamento do cofator ferro-molibdênio da nitrogenase. Este gene está sob controle de um promotor dependente de sigma 54, a montante do qual foram identificados dois sítios funcionais para ligação de NifA e um para ligação de IHF (SOUZA *et al.*, 1991; REGO, 1997).

O operon *nifQmodABC* está localizado a jusante do operon *nifHDKENXorf1orf2* e está orientado na mesma direção. A proteína NifQ deve estar relacionada com a incorporação de molibdênio no cofator ferro-molibdênio da nitrogenase. As proteínas ModA, ModB e ModC formam um sistema de transporte de molibdênio de alta afinidade, pertencente à família de transportadores ABC (KLASSEN, 2000).

O operon *fixXC* está parcialmente sequenciado. Está disposto a jusante do operon *nifQmodABC* e é transcrito na direção oposta. Codifica para as proteínas FixX e FixC, as

quais provavelmente atuam na transferência de elétrons para a nitrogenase (KLASSEN, 2000).

O gene *glnB* é uma unidade transcricional independente que codifica para a proteína PII, relacionada com a regulação do metabolismo de nitrogênio conforme a disponibilidade de amônio no ambiente (BENELLI *et al.*, 1997). O promotor deste gene é aparentemente dependente de sigma 70. A jusante da sequência estrutural, foi encontrado um provável terminador intrínseco.

O operon *glnAntrBC* codifica para glutamina sintetase (GS) e as proteínas NtrB e NtrC, respectivamente (RIGO *et al.*, 1996; STEFFENS *et al.*, 1997). A enzima glutamina sintetase catalisa a síntese de glutamina a partir de amônio e glutamato. As proteínas NtrB e NtrC constituem um sistema regulatório de dois componentes envolvido com a fixação e o metabolismo de nitrogênio. NtrB deve atuar como uma histidina quinase de membrana, fosforilando NtrC em baixas concentrações de amônio, enquanto NtrC fosforilada deve atuar como um ativador de transcrição dos genes de fixação e metabolismo de nitrogênio. Este operon está aparentemente sob controle de dois promotores, um dependente de sigma 70 e outro dependente de sigma 54. A montante do promotor dependente de sigma 54, foram encontrados dois sítios para ligação de NtrC, um dos quais está sobreposto ao promotor dependente de sigma 70 (RIGO *et al.*, 1996; STEFFENS *et al.*, 1997; PERSUHN *et al.*, no prelo). A jusante do gene *glnA*, foi identificado um provável atenuador de transcrição (STEFFENS *et al.*, 1997) e, a jusante do gene *ntrC*, um possível terminador intrínseco (TWERDOCHLIB, 1999).

O gene *amtB* está parcialmente sequenciado. Codifica para a proteína AmtB, um provável transportador de íons amônio (NOINDORF *et al.*, 1999).

O operon *recAX* codifica para a proteína RecA, que atua na recombinação homóloga e no reparo do DNA (STEFFENS *et al.*, 1993), e para a proteína RecX, que também deve atuar nestes processos (GALVÃO *et al.*, 1998). A montante do gene *recA*, foi encontrado um possível promotor dependente de sigma 70 e, a jusante do gene *recX*, um provável terminador intrínseco (GALVÃO *et al.*, 1998).

Os genes identificados em *H. seropedicae* formam unidades transcricionais independentes ou agrupadas em operons. Podem estar sob o controle de promotores dependentes de sigma 70 ou sigma 54, conforme a função que desempenham. Os genes de fixação e metabolismo de nitrogênio geralmente apresentam promotores dependentes de sigma 54, ativados por NifA ou NtrC. Deste modo, a sua transcrição pode ser regulada de acordo com a disponibilidade de amônio no ambiente.

A identificação de outros genes pode ampliar o conhecimento do genoma de *H. seropedicae* e pode contribuir para o esclarecimento dos processos de fixação e metabolismo de nitrogênio. Além disso, a análise estrutural destes genes pode fornecer informações que permitam inferir acerca dos processos de regulação da expressão gênica nesta espécie.

1.4 OBJETIVOS

1.4.1 Objetivo geral

Contribuir para o conhecimento do genoma de *H. seropedicae* através da identificação e da análise estrutural de genes presentes a montante do gene *nifA*.

1.4.2 Objetivos específicos

Subclonar e seqüenciar a região a montante do gene *nifA*.

Identificar seqüências codificadoras de proteínas, assim como promotores, seqüências para ligação de proteínas, sítios de ligação de ribossomo e terminadores.

Determinar o grau de similaridade entre as proteínas codificadas pelos genes identificados e as descritas para outras bactérias.

2 MATERIAL E MÉTODOS

2.1 ESTIRPES BACTERIANAS E PLASMÍDEOS

As estirpes bacterianas e os plasmídeos empregados no desenvolvimento deste trabalho estão listados nas tabelas 1 e 2, respectivamente.

Tabela 1. Estirpes bacterianas.

Estirpe	Genótipo / fenótipo	Referência
<i>Escherichia coli</i>		
HB101	F ⁻ <i>mcrB mrr hsdS20</i> (r _B ⁻ , m _B ⁻) <i>recA13 supE44 ara14 galK2 lacY1 proA2 rpsL20</i> (Sm ^R) <i>xyl5 λ⁻ leu mtl1</i>	BOYER e ROULLAND-DUSSOIX (1969)
DH10B	F ⁻ <i>mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL nupG</i>	GRANT <i>et al.</i> (1990)

Tabela 2. Plasmídeos.

Plasmídeo	Genótipo/fenótipo	Referência/origem
pEMS1	pVK102 :: 31 kb <i>SalI/SalI</i> <i>H. seropedicae</i> contendo o gene <i>nifA</i>	SOUZA <i>et al.</i> (1990)
pTZ19R	Ap ^R <i>lacZ</i> fl	MEAD <i>et al.</i> (1986)
pSport1	Ap ^R <i>lacI</i> fl	Life Technologies (Gaithersburg, EUA)
pELV2B	pTZ19R :: 3,5 kb <i>BamHI/SalI</i> <i>H. seropedicae</i> contendo o gene <i>nifA</i> (parcial)	Este trabalho
pELV2C	pSport1 :: 3,5 kb <i>BamHI/SalI</i> <i>H. seropedicae</i> contendo o gene <i>nifA</i> (parcial)	Este trabalho
pELV21B	pTZ19R :: 0,9 kb <i>BamHI/PstI</i> <i>H. seropedicae</i>	Este trabalho
pELV22B	pTZ19R :: 0,9 kb <i>EcoRI/EcoRI</i> <i>H. seropedicae</i>	Este trabalho
pELV3B	pTZ19R :: 3,5 kb <i>BamHI/SalI</i> <i>H. seropedicae</i>	Este trabalho
pELV3C	pSport1 :: 3,5 kb <i>BamHI/SalI</i> <i>H. seropedicae</i>	Este trabalho

2.2 MEIOS DE CULTURA

As estirpes de *E. coli* foram cultivadas nos meios Luria Broth (LB), Terrific Broth (TB) (SAMBROOK *et al.*, 1989) e MacConkey (MILLER, 1992).

O meio LB apresenta a seguinte composição:

Extrato de levedura 5,0 g/L

NaCl 10,0 g/L

Triptona 10,0 g/L

O pH foi ajustado para 7,0 com solução de NaOH 1 mol/L. O meio sólido, denominado LA, foi obtido pela adição de ágar na concentração de 12,0 g/L ao meio líquido.

O meio TB possui a seguinte composição:

Extrato de levedura	24,0 g/L
Glicerol	4,0 mL/L
Triptona	12,0 g/L
KH ₂ PO ₄	2,31 g/L
K ₂ HPO ₄	12,54 g/L

O meio sólido MacConkey apresenta a seguinte composição:

Peptona	17,0 g/L
Polipeptona	3,0 g/L
Lactose	10,0 g/L
Sais biliares	1,5 g/L
NaCl	5,0 g/L
Vermelho neutro	0,03 g/L
Cristal violeta	0,001 g/L
Ágar	13,5 g/L

2.3 CONDIÇÕES DE CULTIVO E ESTOCAGEM DE BACTÉRIAS

As estirpes bacterianas foram cultivadas em meio LB ou TB contendo antibióticos apropriados, a 37°C, sob agitação a 180 rpm, por 12-15 h. As células foram estocadas em solução de glicerol a 50% (v/v) a -20°C (SAMBROOK *et al.*, 1989).

2.4 ANTIBIÓTICOS

Os antibióticos ácido nalidíxico, ampicilina e canamicina foram adicionados aos meios de cultura nas concentrações finais de 10, 250 e 50 µg/mL, respectivamente. As soluções estoque de ampicilina (250 mg/mL) e canamicina (100 mg/mL) foram preparadas em água ultrapura (água destilada e filtrada através de uma matriz de carvão

ativo e de resinas de troca iônica no aparelho Milli-Q Plus, Millipore Corporation, Bedford, EUA). Estas soluções foram esterilizadas por filtração através de uma membrana com poros de 0,45 µm de diâmetro (Millipore Corporation, Bedford, EUA). A solução estoque de ácido nalidíxico (20 mg/mL) foi preparada em água ultrapura, neutralizada com solução de NaOH 1 mol/L para completa dissolução e esterilizada por filtração. Estas soluções foram mantidas a 0°C e adicionadas ao meio de cultura imediatamente antes do uso (SAMBROOK *et al.*, 1989).

2.5 REAGENTES

Ácido fórmico, álcool isoamílico, azul de bromofenol, clorofórmio, D-glucose, etanol, fosfato dibásico de potássio, fostato monobásico de potássio, glicerol, hidróxido de sódio, isopropanol e sulfato de zinco foram adquiridos de E. Merck KGaA (Darmstadt, Alemanha). Ácido bórico, agarose, cloreto de sódio, extrato de levedura, fenol, SDS, triptona e Tris-base foram adquiridos de Life Technologies (Gaithersburg, EUA). Ácido nalidíxico, brometo de etídeo, cloreto de magnésio, 8-hidroxiquinoleína, RNase e xileno cianol foram adquiridos de Sigma Chemical Company (St. Louis, EUA). Ampicilina, canamicina, EDTA, Ficoll e dNTPs foram adquiridos de Amersham Pharmacia Biotech Europe GmbH (Uppsala, Suécia). Acetato de potássio foi adquirido de Sigma - Aldrich Chemie GmbH (Steinheim, Suécia). A mistura para meio sólido MacConkey foi adquirida de Oxoid Ltd. (Basingstone, Inglaterra).

2.6 SOLUÇÕES

2.6.1 Soluções de fenol tamponado, clorofórmio-álcool isoamílico e fenol-clorofórmio-álcool isoamílico

Para a preparação da solução de fenol tamponado, fenol cristalino foi fundido a 68°C. Ao fenol fundido, foram adicionados 8-hidroxiquinoleína a 0,1% (p/v) e um volume de solução de Tris-base 0,1 mol/L. Esta mistura foi agitada em agitador eletromagnético por 15 min e a emulsão foi mantida em repouso para a separação das fases aquosa e orgânica. A fase aquosa foi removida e a fase orgânica foi novamente equilibrada uma ou duas vezes com um volume de solução de Tris-HCl 0,1 mol/L pH 8,0 para atingir pH igual ou superior a 7,5. Na última lavagem, foi acrescentado β -mercaptoetanol a 0,2% (p/v). A solução de fenol tamponado foi conservada em frasco escuro a 4°C (SAMBROOK *et al.*, 1989).

A solução de clorofórmio-álcool isoamílico foi preparada misturando-se 24 partes de clorofórmio para 1 parte de álcool isoamílico. A solução de fenol-clorofórmio-álcool isoamílico foi preparada misturando-se 1 parte de solução de fenol tamponado para 1 parte de solução de clorofórmio-álcool isoamílico (24:1). Ambas as soluções foram mantidas em frasco escuro a 4°C (SAMBROOK *et al.*, 1989).

2.6.2 Solução de RNase

A solução estoque de RNase (10 mg/mL) foi preparada em tampão Tris-HCl 10 mmol/L pH 7,5 contendo NaCl 15 mmol/L. Esta solução foi fervida por 30 min, resfriada à temperatura ambiente para inativação de DNases contaminantes e conservada a -20°C (SAMBROOK *et al.*, 1989). Esta solução foi adicionada aos sistemas de reação adequados ajustando-se a concentração final para 0,1 mg/mL.

2.6.3 Outras soluções

As soluções empregadas na manipulação de DNA, assim como as soluções de glicerol a 10 e 50 % (v/v), foram preparadas em água ultrapura e esterilizadas por autoclavação a 127°C e 1,5 atm ou por filtração (item 2.4), se necessário. Tampão TBE 1x concentrado (Tris-base 89 mmol/L, ácido bórico 89 mmol/L e EDTA 2 mmol/L pH 8,0), solução corante FSUDS-xileno cianol (azul de bromofenol a 0,25% (p/v), xileno cianol a 0,25% (p/v), Ficoll a 25% (p/v), SDS a 0,5% (p/v), Tris-HCl 10 mmol/L pH 8,0 e EDTA 1 mmol/L pH 8,0) e solução de brometo de etídeo a 0,5 µg/mL foram preparadas em água destilada (SAMBROOK *et al.*, 1989).

2.7 MINIPREPARAÇÃO DE PLASMÍDEOS

A purificação de plasmídeos em pequena escala foi realizada por meio de uma adaptação do método de lise alcalina (SAMBROOK *et al.*, 1989). Uma alíquota de 1,5 mL de uma cultura bacteriana ($DO_{600} \cong 2,0$) foi transferida para um tubo de microcentrífuga e centrifugada a 12.800 xg por 30 segundos. O meio de cultura foi descartado e as células foram ressuspensas em 100 µL de solução GET (D-glucose 50 mmol/L, EDTA 10 mmol/L pH 8,0 e Tris-HCl 25 mmol/L pH 8,0). As mesmas foram rompidas pela adição de 200 µL de solução de lise (SDS a 1% (p/v) e NaOH 0,2 mol/L), seguida de homogeneização suave e incubação à temperatura ambiente por 5 min. DNA cromossômico, proteínas, SDS e fragmentos celulares foram precipitados pelo acréscimo de 200 µL de solução KAcF (acetato de potássio 3 mol/L e ácido fórmico 1,8 mol/L pH 4,8) gelada, seguido de agitação suave e incubação em banho de gelo por 5 min. A mistura foi centrifugada a 12.800 xg por 5 min, o sobrenadante foi coletado em um tubo novo e foram adicionados 100 µL de solução de fenol-clorofórmio-álcool isoamílico (25:24:1). A mistura foi homogeneizada vigorosamente e centrifugada a 12.800 xg por 5 min. A fase aquosa foi transferida para um tubo novo e o DNA plasmidial foi precipitado

pelo acréscimo de 2 volumes de etanol a 95% (v/v), seguido de agitação vigorosa e centrifugação a 12.800 xg por 20 min. O sobrenadante foi descartado e o precipitado foi lavado duas vezes com 1 mL de etanol a 70% (v/v), seco em câmara de vácuo e dissolvido em 30-50 µL de água ultrapura.

2.8 MAXIPREPARAÇÃO DE PLASMÍDEOS

A purificação de plasmídeos em larga escala foi feita por meio de uma adaptação do método de lise alcalina (SAMBROOK *et al.*, 1989), adequando-se os volumes. Um volume de 40 mL de uma cultura bacteriana ($DO_{600} \cong 2,0$) foi transferido para um tubo de centrífuga e centrifugado a 4.300 xg, a 5°C, por 5 min. O meio de cultura foi retirado e as células foram ressuspendidas em 4 mL de solução GET. A lise foi feita por meio da adição de 8 mL de solução de lise, seguida de homogeneização suave e incubação à temperatura ambiente por 5 min. Foram acrescentados 6 mL de solução KAcF gelada e a mistura foi agitada suavemente. Foram, ainda, adicionados 3 mL de solução de fenol-clorofórmio-álcool isoamílico (25:24:1) e a mistura foi homogeneizada vigorosamente, incubada em banho de gelo por 10 min e centrifugada a 17.200 xg, a 5°C, por 10 min. O sobrenadante foi transferido para outro tubo, ao qual foi acrescentado 0,6 volume de isopropanol. A mistura foi agitada, incubada à temperatura ambiente por 20 min e centrifugada a 17.200 xg, a 15°C, por 15 min. O sobrenadante foi descartado e o precipitado foi lavado duas vezes com 4 mL de etanol a 70% (v/v), seco em câmara de vácuo e ressuspenso em 1 mL de água ultrapura.

2.9 CLIVAGEM DE DNA COM ENDONUCLEASES DE RESTRIÇÃO

Os plasmídeos purificados foram clivados por endonucleases de restrição apropriadas conforme instruções do fabricante (Life Technologies, Gaithersburg, EUA; Amersham Pharmacia Biotech Europe GmbH, Uppsala, Suécia). Em um sistema de

reação de 20 μ L foram utilizados 1-2 μ L de solução de DNA (200-400 ng), 2 μ L de tampão de restrição 10x concentrado, 0,2 μ L de solução de RNase a 10 mg/mL, 1 U de endonuclease de restrição e água ultrapura em quantidade suficiente para 20 μ L. Este sistema foi incubado a 37°C por 3-5 h.

2.10 ELETROFORESE DE DNA EM GEL DE AGAROSE

Os plasmídeos íntegros ou clivados por endonucleases de restrição foram submetidos à eletroforese horizontal em gel de agarose (SAMBROOK *et al.*, 1989). Foram empregados géis de agarose a 0,7% (p/v) em tampão TBE 1x concentrado. A agarose foi fundida neste tampão a 100°C e resfriada a cerca de 50°C, sendo então transferida para placa suporte contendo um ou dois pentes e mantida em repouso à temperatura ambiente até completa geleificação. O pente foi removido e o gel foi submergido em tampão TBE 1x concentrado. As amostras foram misturadas ao corante FSUDS em uma proporção de 3:1 e aplicadas nos poços do gel. A corrida eletroforética foi conduzida a 2 V/cm por 3-4 h ou a 0,5 V/cm por 12-15 h. O gel foi corado com solução de brometo de etídeo (0,5 μ g/mL) por 15-30 min e lavado em água destilada para remoção do excesso de corante. O DNA foi visualizado sob luz ultravioleta (312 nm) em transiluminador. O perfil eletroforético foi fotografado (filme ISO/ASA 100) ou registrado em sistema de vídeo acoplado a equipamento de impressão a seco GDS 5000 (UVP Ltd., Cambridge, Inglaterra). O tamanho molecular dos fragmentos analisados foi estimado através de comparação com marcador de tamanho molecular “1 kb DNA ladder” (Life Technologies, Gaithersburg, EUA).

2.11 SUBCLONAGEM DE DNA

Fragmentos de DNA obtidos do plasmídeo pEMS1 foram subclonados nos vetores pTZ19R ou pSport1 (SAMBROOK *et al.*, 1989). O plasmídeo pEMS1 foi purificado em larga escala (item 2.8) e os vetores, em pequena escala (item 2.7). Os mesmos foram clivados com as endonucleases de restrição *Bam*HI e *Sal*I e a mistura de fragmentos foi desproteinizada pela adição de 0,5 volume de solução de fenol-clorofórmio-álcool isoamílico (25:24:1), seguida de agitação vigorosa e centrifugação a 12.800 xg por 5 min. O sobrenadante foi coletado e o DNA foi precipitado pelo acréscimo de 3 volumes de etanol a 95% (v/v), seguido de homogeneização, incubação à temperatura ambiente por 20 min e centrifugação a 12.800 xg por 20 min. O sobrenadante foi descartado e o precipitado foi lavado duas vezes com 1 mL de etanol a 70% (v/v), seco em câmara de vácuo e dissolvido em 20 µL de água ultrapura.

Os fragmentos *Bam*HI/*Sal*I do plasmídeo pEMS1 foram empregados como DNA-inserto em reações de ligação com cada um dos vetores. Em um sistema de 15 µL, foram utilizadas soluções de DNA-inserto e vetor na proporção de 20:1, 3 µL de tampão de ligação 5x concentrado (polietilenoglicol-8000 a 25% (p/v), Tris-HCl 250 mmol/L pH 7,6, MgCl₂ 50 mmol/L, ATP 5 mmol/L e ditiotretitol 5 mmol/L) e 1 U de T₄ DNA ligase (Life Technologies, Gaithersburg, EUA). Inicialmente, DNA-inserto e vetor foram aquecidos a 65°C por 5 min e resfriados em banho de gelo e, em seguida, foram adicionados os demais reagentes. O sistema foi incubado a 16°C por 12-15 h.

Por meio destes procedimentos, foram obtidos quatro subclones denominados pELV2B, pELV2C, pELV3B e pELV3C (item 3.1, figura 4). A partir do subclone pELV2B, carregado pelo vetor pTZ19R, foram obtidos fragmentos *Bam*HI/*Pst*I e *Eco*RI/*Eco*RI, os quais foram subclonados neste mesmo vetor. Para tanto, o plasmídeo pELV2B foi purificado em pequena escala e clivado com as endonucleases de restrição *Bam*HI e *Pst*I ou *Eco*RI. As misturas de fragmentos foram desproteinizadas, precipitadas, lavadas, secas e ressuspensas separadamente em 20 µL de água ultrapura. Estas

soluções foram empregadas em reações de ligação conforme os procedimentos descritos acima, através dos quais foram obtidos os subclones pELV21B e pELV22B (item 3.1, figura 5).

2.12 TRANSFORMAÇÃO BACTERIANA

As misturas de ligação foram empregadas na transformação de células de *E. coli* DH10B por meio de eletroporação, de acordo com as instruções do manual do eletroporador “Cell-Porator[®] Electroporation System” (Life Technologies, Gaithersburg, EUA). Células eletrocompetentes foram preparadas a partir de culturas bacterianas recentes ($DO_{600} = 0,3-0,7$) em meio LB, acrescido de 10 $\mu\text{g/mL}$ de ácido nalidíxico. Cerca de 100 mL de cultura foram transferidos para um tubo de centrífuga e submetidos à centrifugação a 4.300 xg, a 5°C, por 5 min. O meio de cultura foi removido e as células foram lavadas duas vezes com 30-40 mL de água ultrapura estéril e uma vez com 30-40 mL de solução de glicerol a 10% (v/v) estéril, sendo finalmente ressuspensas em 1 mL desta solução. Aliquotas de 40-80 μL da suspensão celular foram distribuídas em tubos de microcentrífuga e conservadas a -70°C.

Para a transformação, foram reunidos 20-40 μL de suspensão de células eletrocompetentes e 1-2 μL de mistura de ligação, os quais foram transferidos para uma câmara de eletroporação mantida em gelo. Esta mistura foi submetida a um pulso elétrico (4 kV, 200 Ω , 300 μF), coletada e diluída em 800-1000 μL de meio LB e incubada a 37°C, sob agitação a 180 rpm, por 1 h. Em seguida, 200-300 μL desta cultura foram plaqueados em meio MacConkey contendo 250 $\mu\text{g/mL}$ de ampicilina e 10 $\mu\text{g/mL}$ de ácido nalidíxico. As placas foram incubadas a 37°C por 12-24 h e as colônias foram triadas por α -complementação em relação à atividade de β -galactosidase (MILLER, 1992).

2.13 SEQÜENCIAMENTO DE DNA

2.13.1 Obtenção de deleções unidirecionais

O seqüenciamento dos fragmentos de DNA subclonados envolveu a obtenção de deleções unidirecionais (HENIKOFF, 1984). Para tanto, cerca de 3,0 a 5,0 µg de DNA plasmidial foram clivados com as endonucleases de restrição *Bam*HI e *Kpn*I ou *Sal*I e *Kpn*I. Este material foi desproteinizado, precipitado, lavado e ressuspendido em 50 µL de água ultrapura (item 2.11). Tal tratamento permitiu a obtenção de plasmídeos linearizados com uma ponta protudente 5' e uma ponta protudente 3'.

Os plasmídeos linearizados foram tratados com a enzima exonuclease III (ExoIII, Life Technologies, Gaithersburg, EUA). Em um sistema de deleção de 60 µL, foram empregados 25 µL da solução de plasmídeos linearizados (1,5-3,0 µg de DNA), 6 µL de tampão "ExoIII" 10x concentrado (Tris-HCl 660 mmol/L pH 8,0 e MgCl₂ 6,6 mmol/L), 400 U de enzima ExoIII e água ultrapura em quantidade suficiente para 60 µL. DNA, tampão e água ultrapura foram reunidos e aquecidos em banho-maria a 37°C por 5 min. A enzima foi adicionada em seguida e o sistema foi incubado por mais 12 min na mesma condição. Uma amostra de 2,5 µL foi coletada a cada 30 segundos e transferida para um tubo contendo 30 µL de tampão "ExoIII stop" (NaCl 200 mmol/L e EDTA 5 mmol/L pH 8,0) mantido em banho de gelo. Foram coletadas 24 amostras, as quais foram reunidas 4 a 4 em um mesmo tubo, perfazendo um total de 6 tubos. Neste tratamento, as pontas protudentes 5' dos plasmídeos linearizados serviram como regiões de ligação da exonuclease III, a qual hidrolisou a cadeia complementar na direção 3'→5'. Os diferentes tempos de incubação das amostras permitiram a redução progressiva do tamanho do DNA-inserto.

As amostras coletadas foram incubadas a 70°C por 10 min para a inativação da exonuclease III. Os plasmídeos deletados foram precipitados pela adição de 3 volumes de etanol a 95% (v/v), seguida de homogeneização, incubação a -70°C por 20 min e

centrifugação a 12.800 xg por 10 min. O sobrenadante foi retirado e o precipitado foi lavado uma vez com 100 μ L de etanol a 70% (v/v), seco em câmara de vácuo e ressuscitado em 50 μ L de tampão “S1” 1x concentrado (glicerol a 5% (v/v), acetato de potássio 30 mmol/L pH 4,6, NaCl 250 mmol/L e ZnSO₄ 1 mmol/L). Em seguida, foram adicionadas 3,5 U da enzima nuclease S1 (Life Technologies, Gaithersburg, EUA). Os sistemas foram incubados a 25°C por 30 min e a reação foi interrompida pela adição de 6 μ L de tampão “S1 stop” (Tris-HCl 500mmol/L pH 8,0 e EDTA 500 mmol/L pH 8,0). Nesta reação, as pontas protudentes dos plasmídeos deletados foram aparadas pela nuclease S1.

Os sistemas foram desproteinizados e o DNA foi precipitado, lavado e ressuscitado em 8,75 μ L de água ultrapura (item 2.11). Este material foi submetido a tratamento com o fragmento Klenow da DNA polimerase I (Life Technologies, Gaithersburg, EUA). Para tanto, foram adicionados 1 μ L de tampão 10x concentrado (Tris-HCl 500 mmol/L pH 8,0, MgCl₂ 100 mmol/L e NaCl 500 mmol/L) e 2 U de enzima. Os sistemas foram incubados a 37°C por 5 min, foram acrescentados 0,4 μ L de solução de dNTPs a 5 mmol/L e os mesmos foram mantidos nesta temperatura por mais 10 min. Durante este tratamento, as pontas dos plasmídeos deletados foram reparadas, originando pontas cegas.

Os plasmídeos deletados foram circularizados através do tratamento com a T₄ DNA ligase (Life Technologies, Gaithersburg, EUA). A cada sistema, foram acrescentados 8 μ L de tampão de ligação 5x concentrado, 1 U de enzima e 31 μ L de água ultrapura, sendo incubados a 16°C por 12-15 h. As misturas de ligação foram utilizadas separadamente para a transformação de células de *E. coli* DH10B por eletroporação (item 2.12). As células foram plaqueadas em meio LA contendo de 250 μ g/mL de ampicilina e 10 μ g/mL de ácido nalidíxico e cultivadas a 37°C por 12-24 h. As colônias foram coletadas e cultivadas em meio LB (item 2.3). Os plasmídeos foram purificados (item 2.7), tratados com endonucleases de restrição (item 2.9) e analisados quanto ao tamanho

molecular por eletroforese (item 2.10). Plasmídeos com tamanho molecular intermediário ao original e o vetor foram selecionados e organizados em ordem decrescente de tamanho molecular.

2.13.2 Obtenção de DNA-molde

Os plasmídeos originais e seus derivados foram utilizados como DNA-molde em reações de seqüenciamento. Para tanto, os mesmos foram purificados em pequena escala manualmente (item 2.7) ou por meio de “kits” de purificação (“Plasmid Mini Kit” e “QIAprep Spin Miniprep Kit”, Qiagen GmbH, Hilden, Alemanha). Quando purificado manualmente, o DNA plasmidial precipitado e seco foi ressuspensionado em 150 µL de água ultrapura e foi adicionado 1 µL de solução de RNase a 10 mg/mL. O sistema foi incubado a 37°C por 1 h para a degradação de RNA contaminante. Logo depois, este material foi desproteinizado pela adição de 0,5 volume de solução de fenol-clorofórmio-álcool isoamílico (25:24:1), seguida de agitação vigorosa e centrifugação a 12.800 xg por 5 min. A fase aquosa foi coletada em um tubo novo e foram adicionados 75 µL de solução de acetato de amônio 7,5 mol/L e 3 volumes de etanol a 95% (v/v). A mistura foi homogeneizada, incubada à temperatura ambiente por 20 min e centrifugada a 12.800 xg por 20 min. O sobrenadante foi descartado e o precipitado foi lavado duas vezes com 1 mL de etanol a 70% (v/v), seco em câmara de vácuo e dissolvido em 20 µL de água ultrapura.

2.13.3 Reações de seqüenciamento

O DNA-inserto dos plasmídeos originais e de seus derivados foi seqüenciado pelo método enzimático (SANGER *et al.*, 1977). Para tanto, foram utilizados um “kit” de seqüenciamento que contém terminadores de cadeia marcados com corantes fluorescentes (“ABI Prism™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit”, Perkin-

Elmer, Foster, EUA) e os oligonucleotídeos iniciadores (“primers”) universal e T7 adquiridos de Amersham Pharmacia Biotech Europe GmbH (Uppsala, Suécia) ou sintetizados no aparelho “Gene Assembler Special” (Amersham Pharmacia Biotech Europe GmbH, Uppsala, Suécia), conforme instruções do fabricante.

As reações de seqüenciamento foram realizadas de acordo com as recomendações do fabricante do “kit” de seqüenciamento. Em um sistema de reação de 20 μ L, foram empregados 1-2 μ L de solução de DNA-molde (200-400 ng), 20-30 pmol de oligonucleotídeo iniciador universal ou T7, 8 μ L de “BigDye™ Terminator Ready Reaction Mix” (Perkin-Elmer, Foster, EUA) e água ultrapura em quantidade suficiente para 20 μ L. O DNA foi amplificado em termociclador, sendo submetido aos seguintes ciclos de temperatura: 95°C, 5 min (1 ciclo); 95°C, 30 s; 40°C, 30 s; 60°C, 6 min (30 ciclos). Em seguida, o volume da reação foi transferido para um tubo novo, adicionando-se 10 μ L de solução de acetato de amônio 7,5 mol/L e 90 μ L de etanol a 95% (v/v). A mistura foi agitada, incubada em banho de gelo por 20 min e centrifugada a 12.800 xg por 20 min. O sobrenadante foi descartado e o precipitado foi lavado uma vez com 1 mL de etanol a 70% (v/v), seco em câmara de vácuo e ressuspensionado em 15 μ L de solução TSR (“template suppression reagent”, Perkin-Elmer, Foster, EUA) ou em 5 μ L de solução de formamida deionizada-“blue dextran”/EDTA 5:1 (Perkin-Elmer, Foster, EUA). A separação e a leitura dos produtos das reações de seqüenciamento foram realizadas em Seqüenciador Automático ABI310 (Perkin-Elmer, Foster, EUA) (ressuspensão em solução TSR) ou ABI377 (Perkin-Elmer, Foster, EUA) (ressuspensão em solução de formamida deionizada-“blue dextran”/EDTA).

2.14 EDIÇÃO E ANÁLISE DE SEQÜÊNCIAS

As seqüências obtidas a partir dos plasmídeos originais e de seus derivados foram alinhadas e editadas no programa Auto Assembler 1.4 (Perkin-Elmer, Foster, EUA). Prováveis seqüências codificadoras de proteína (ORFs) foram identificadas pelo método

de preferência de códons do programa Analyseq (STADEN, 1982). Para a utilização deste método, foi empregada uma tabela de preferência de códons obtida a partir das seqüências dos genes *nifA*, *nifB*, *nifH*, *nifD*, *nifK*, *glnA*, *glnB* e *ntrC* de *H. seropedicae* (CHUBATSU, 1998) recuperadas do banco de dados GenBank. As prováveis ORFs foram comparadas com genes descritos para outros organismos através do programa Blast 2.0 (ALTSCHUL *et al.*, 1997).

As seqüências de nucleotídeos das ORFs identificadas foram traduzidas para seqüências de aminoácidos com o auxílio do programa Strider 1.3 (Perkin-Elmer, Foster, EUA). As seqüências de aminoácidos deduzidas foram alinhadas pelo programa Clustal X (THOMPSON *et al.*, 1997) com seqüências de aminoácidos de proteínas similares, coletadas do banco de dados GenBank. Os alinhamentos obtidos foram submetidos ao programa Antheprot 4.9 (DELEAGE *et al.*, 1988) para assinalar os aminoácidos idênticos e similares. As porcentagens de identidade e similaridade entre as proteínas alinhadas foram obtidas pelo programa Multalin (CORPET, 1988). Motivos e domínios característicos das proteínas identificadas foram determinados por análise manual. A massa molecular e o ponto isoelétrico destas proteínas foram estimados pelos programas BioEdit 4.8.8 (HALL, 1999) e “EMBL WWW Gateway to Isoelectric Point Service” do banco de dados EMBL, respectivamente.

Promotores, sítios para ligação de proteínas, sítios de ligação de ribossomo e terminadores foram encontrados pela análise manual da seqüência de nucleotídeos.

3 RESULTADOS E DISCUSSÃO

3.1 OBTENÇÃO DE SUBCLONES PARA SEQUENCIAMENTO DE DNA

Dois fragmentos contíguos *Bam*HI/*Sal*I de 3,5 kb do DNA-inserto do plasmídeo pEMS1 foram subclonados nos vetores pTZ19R e pSport1 (figura 4). O fragmento que contém a região proximal do gene *nifA* foi empregado na construção dos plasmídeos pELV2B e pELV2C, enquanto o fragmento que dista cerca de 2,7 kb do gene *nifA* foi utilizado na construção dos plasmídeos pELV3B e pELV3C.

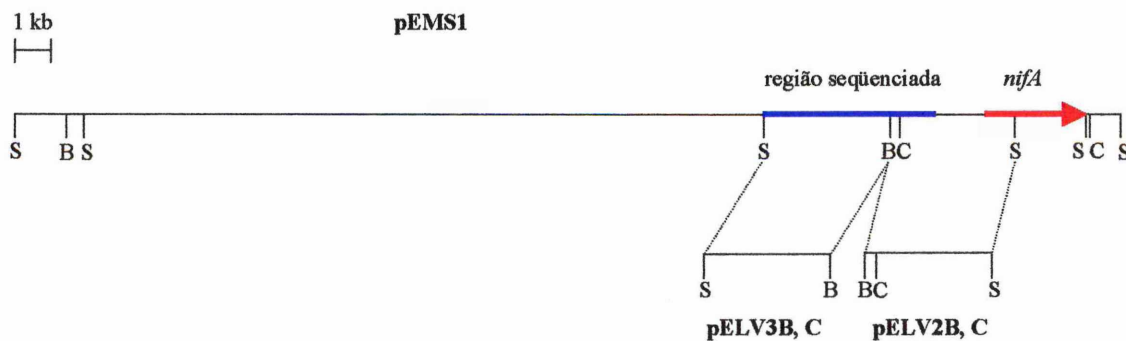


Figura 4. Mapa de restrição do plasmídeo pEMS1 e de plasmídeos derivados. pVK102 (anexo 1) é o vetor de pEMS1. pTZ19R (anexo 2) é o vetor de pELV2B e pELV3B. pSport1 (anexo 3) é o vetor de pELV2C e pELV3C. O gene *nifA* está representado por uma seta vermelha, que indica a direção de transcrição. A região sequenciada a montante do gene *nifA* está representada por um traço azul. São indicados os sítios *Bam*HI (B), *Cla*I (C) e *Sal*I (S).

Dois fragmentos do DNA-inserto do plasmídeo pELV2B foram subclonados no vetor pTZ19R (figura 5). Os fragmentos *Bam*HI/*Pst*I e *Eco*RI/*Eco*RI de 0,9 kb foram utilizados na construção dos plasmídeos pELV21B e pELV22B, respectivamente. O DNA-inserto destes plasmídeos apresenta uma sobreposição correspondente ao segmento *Eco*RI/*Pst*I de 0,2 kb.

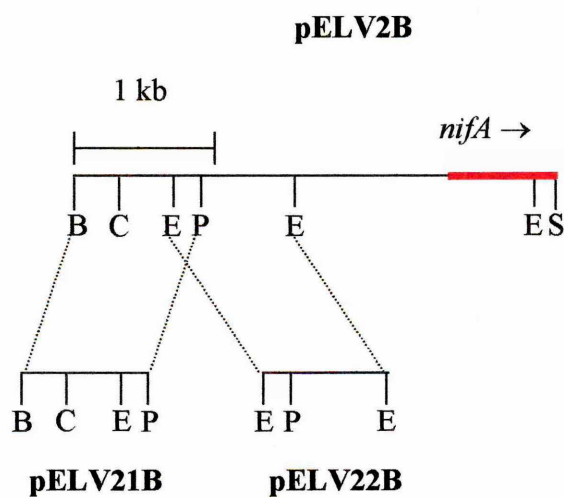


Figura 5. Mapa de restrição do plasmídeo pELV2B e de plasmídeos derivados. pTZ19R é o vetor de pELV2B, pELV21B e pELV22B. A região proximal do gene *nifA* está representada por um traço vermelho. A seta indica a direção de transcrição do gene *nifA*. São indicados os sítios *Bam*HI (B), *Cla*I (C), *Eco*RI (E), *Pst*I (P) e *Sal*I (S).

O DNA-inserto dos plasmídeos pELV2C, pELV3B e pELV3C foi submetido à deleção por exonuclease III para a redução progressiva do seu tamanho. Por meio deste procedimento, foram obtidos 32 subclones. A deleção de pELV2C no sentido *SalI* → *BamHI* originou os plasmídeos pELV2C-10P, 15P, 7R, 6P, 15R, 3P, 14P, 17R, 12R e 11P (figura 6). A deleção de pELV3B no sentido *BamHI* → *SalI* resultou nos plasmídeos pELV3B-22J, 25K, 2L, 7L, 7K, 14K, 6M, 3L, 5M, 30J e 22K (figura 7). Finalmente, a deleção de pELV3C no sentido *SalI* → *BamHI* originou os plasmídeos pELV3C-21V, 25V, 20U, 12V, 22V, 12U, 17U, 1Z, 2Z, 1W e 2W (figura 8). Os subclones produzidos por este procedimento apresentaram entre si uma diferença de tamanho de 300-400 bp em média.

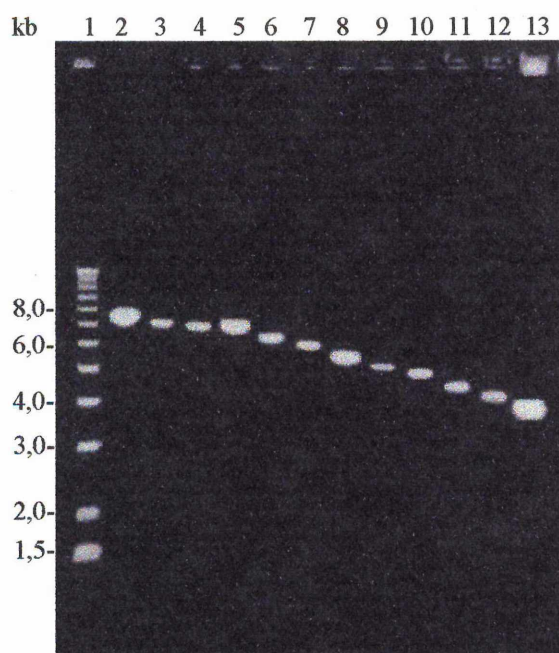


Figura 6. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV2C.

Coluna 1: marcador de tamanho molecular “1 kb DNA ladder”; coluna 2: plasmídeo pELV2C clivado com *Bam*HI; colunas 3 a 12: plasmídeos pELV2C-10P, 15P, 7R, 6P, 15R, 3P, 14P, 17R, 12R e 11P clivados com *Bam*HI; coluna 13: pSport1 clivado com *Bam*HI. A corrida eletroforética foi realizada em gel de agarose a 0,7% (p/v) em tampão TBE 1x concentrado e conduzida a 2 V/cm por 3-4 h. O gel foi corado com solução de brometo de etídeo (0,5µg/mL), irradiado com luz ultravioleta (312 nm) e fotografado (filme ISO/ASA 100).

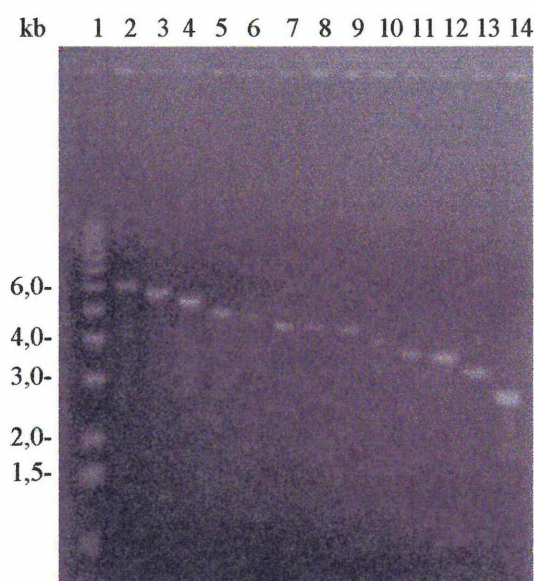


Figura 7. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV3B.

Coluna 1: marcador de tamanho molecular “1 kb DNA ladder”; coluna 2: plasmídeo pELV3B clivado com *Sa*I; colunas 3 a 13: plasmídeos pELV3B-22J, 25K, 2L, 7L, 7K, 14K, 6M, 3L, 5M, 30J e 22K clivados com *Sa*I; coluna 13: pTZ19R clivado com *Sa*I. A corrida eletroforética foi realizada em gel de agarose a 0,7% (p/v) em tampão TBE 1x concentrado e conduzida a 2 V/cm por 3-4 h. O gel foi corado com solução de brometo de etídeo (0,5 µg/mL), irradiado com luz ultravioleta (312 nm) e fotografado (filme ISO/ASA 100).

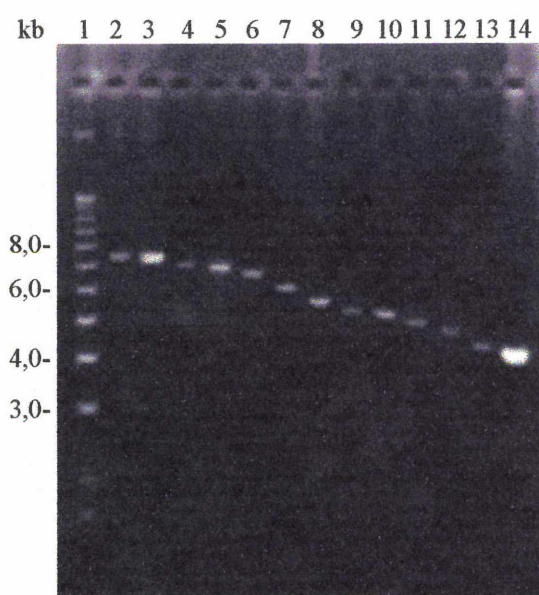


Figura 8. Perfil eletroforético dos plasmídeos derivados do plasmídeo pELV3C.

Coluna 1: marcador de tamanho molecular “1 kb DNA ladder”; coluna 2: plasmídeo pELV3C clivado com *HindIII*; colunas 3 a 13: plasmídeos pELV3C-21V, 25V, 20U, 12V, 22V, 12U, 17U, 1Z, 2Z, 1W e 2W clivados com *HindIII*; coluna 14: pSport1 clivado com *HindIII*. A corrida eletroforética foi realizada em gel de agarose a 0,7% (p/v) em tampão TBE 1x concentrado e conduzida a 2 V/cm por 3-4 h. O gel foi corado com solução de brometo de etídeo (0,5µg/mL), irradiado com luz ultravioleta (312 nm) e fotografado (filme ISO/ASA 100).

3.2 SEQUENCIAMENTO E ANÁLISE ESTRUTURAL DA REGIÃO A MONTANTE DO GENE *nifA*

A partir dos plasmídeos pELV2B, pELV21B, pELV22B, pELV3B e de 15 subclones produzidos por deleção, foram obtidas seqüências com uma média de 500 bp. Os plasmídeos pELV2B, pELV3B-22J, 2L, 7K, 3L e 30J foram seqüenciados utilizando o oligonucleotídeo iniciador universal (“primer” universal). Os plasmídeos pELV21B, pELV22B, pELV2C-15P, 6P, 15R e 17R, pELV3C-12U, 17U, 1Z, 2Z, 1W e 2W foram seqüenciados usando o oligonucleotídeo iniciador T7 (“primer” T7). Finalmente, o plasmídeo pELV3B foi seqüenciado utilizando ambos os oligonucleotídeos iniciadores. Cada seqüenciamento foi repetido 3-5 vezes. O alinhamento das seqüências parciais permitiu a determinação de parte da seqüência do DNA-inserto dos plasmídeos pELV2B e pELV2C e a seqüência completa do DNA-inserto dos plasmídeos pELV3B e pELV3C, totalizando 4777 bp (figura 9).

A região seqüenciada está localizada aproximadamente 1,3 kb a montante do gene *nifA* (figura 4). A seqüência de uma das extremidades do DNA-inserto dos plasmídeos pELV2B e pELV2C já havia sido determinada por SOUZA *et al.* (1991). Esta seqüência inclui a região proximal do gene *nifA* e parte da região intergênica a montante deste gene, totalizando cerca de 1,2 kb. Neste trabalho, foi determinada a seqüência da outra extremidade do DNA-inserto dos plasmídeos pELV2B e pELV2C, correspondente a aproximadamente 1,2 kb. A seqüência da região central do DNA-inserto destes plasmídeos, correspondente a cerca de 1,1 kb, não foi determinada. Os subclones produzidos por deleção pELV2C-3P e 14P, os quais poderiam fornecer seqüências desta região, falharam nos experimentos de seqüenciamento. Além destes, outros subclones obtidos por deleção (dados não mostrados), de tamanho molecular semelhante, também falharam nestes experimentos.

orf4 →

TTGATCTGGCCGCACGCCAGGATTACGGATTGATGGCCACACCCTCCGCC 50
 GATCTTGAAGCAGGTGCTCACGCGGGTGGCTGCACAGCGTGCGCAGCGCGG 100
 TCACTTCCCCTTTGCCATCGCCCGCTTTTTCATTGCGAACCCGAGGCGC 150
 ACAAGGCGCTTGCCCATGGTCTGGCAGGTGTGCGTGCGTATCCGCGCATG 200
 ATGCAAGTGCAATTCGCGCCCGGTGGCACCCCGCCCATGTTCCAGGCCGA 250
 TGCCTCCGATCAAGTCATCCTGGCCAACGCCATCATCGGCGACCCGGCGC 300
 AATGCGTGGCACGCATCGGCCAGCTGCAAGAAAACTGGGGCCGCATCGC 350
 CTGCTGCTCAAGCCTGCCAGCCACGATCCCCAGCAAGCGCGCGCAGCCCT 400

modE →

GAGCTTGCTGGCGGGCGCGCTGGGACTGCAGCCATCGGCAAGGACATGAG 450
 TACCTTGCTGCCACCGATAACACCTTGTGAGCAGCGAACTGAACTGG 500
 TGCATCGCCTGGACCAGCGCTTCTTTGCCCTGCTCGACGCCATCGCGCAA 550
 ACCGGCTCGATCAATCGCGCGGCCAGTACGGCGGGCTACAGCTACAAGGG 600
 CGCATGGATGCTGCTGGAGTCAGCCGGCAACCTGGTCAATGGCGCATTGA 650
 TTGAAACGGTCACCGGCGGCAAAGGCGGCGGCGGCACCCGGCTCACGCCC 700
 GCCGCGGTGGAAGTCTGCGCCGTGTGGCGCGAGCTGCAACGGCGCAATCT 750
 CGAATTCCTGCATCGGCAGGAGACATGGCTTAATCAACTGCCAGCCTTGG 800
 CTGGCCTGCTCAGGAGAATGTCAATGAAAACAGCGCACGTAATCAATTC 850
 GCCGGGGTCATCAGTGCATCGATACCGGCCCTGTCACTACGCAAGTCAC 900
 CGTCACCATCGCCGGCGCGCAAGAGATCGTGGCCACCATGACTACGACCG 950
 CCGCCAATCGGCTCAAGCTCAGGATCGGCAGCAATGCGATTGCCCTGATC 1000
 AAGTCCTCGGCGGTGGTGTGCTGGTCACCGATTTTGCCGGTTTTTCCCTCTC 1050
 GGCACGCAATCAGTTTGAAGGCACGGTCTCGCGCGTAGAGCGTGGCGCGG 1100
 TATCGTCGCTTGTGGTGTGACCTTGCCGCGCGCCTGTATGACCGCC 1150
 AGCCTGACCAATGACGCCATCGATGCGCTGTGCTGGCGGTGCGGCAGAC 1200
 CGCTACCGCCGTCTTCAAGGCGTATGCAAGTGGTGGCCGTGCAGCAGG 1250
 ATTGATTACCTACTGCGGTATCCTTCATTCTGGGTGCTGACACCAGTG 1300
 TGTGAGGGTCAAGTGCCGGATCCCGCACATAAACCGACCGTATCTTCTGC 1350
 GTCGCCCCCGTCCGCAGAACTGGCCCCCTCAATGATAGGCCAGTTCATTTT 1400
 TTCCAGGCGAGCGATCACCTTCTCCAACGGCACGGACGCCATGAAGCACA 1450
 GGTCCAGCGCGCCCGGGCACAGGTAGATGCGCTTTCGGCTCGAATTCCTT 1500
 GCCGCGTACATGCAGGTTGATTTTCTGGTTACCGAAACGAAAGGCGATAC 1550
 GCCCTTGCCCGAAGGTCTCCAGCGTCATGCCAGCACTTCCACATAGAAG 1600
 TGGGTGCTGGCGTCGGGGTTGATCGAAGTCAGTACGAGATGGTCAAGATG 1650
 GTCAATCAGGCTCATGTGAGGTCCGTGGAGTGTTCAGCCTCAAGCGGATA 1700
 GAGCAGATGAGTGACCAATCCCTGCGGGGTCTTTCACGCTTCAATCCGCT 1750

GATATAGCCGCAGCAGATAAAAACAGGAACGATCGGGATTAATTAAAAAG 1800
 CTGCAAAACCTTTTAACTCATATTAAGTGGCACCGACCGATGCCGAGCGA 1850
 TGCGAACCTGATATTAAGTATACAGCTGATGCAATGATGGGAATATCCC 1900
 ATCCGGAAAAACAGGAATAAGCGCCGATCTGGCCTGCAGATCGGAACGG 1950
 GAAGAAACAAGAATGACCTCTCTCCCTTCCCGTTCTTTGTTTCAGCTTAT 2000
 TTCTGCATATTTCTGCTTGTTTCTCTTATTGCTGTTCCCTGATCCTGGCTG 2050
 ATCAGGTATTTCTTTTTCAGGCGTTTCGCTGACTTCGCCATTCTTGTCCTT 2100
 GAGCCACAATGGAATACGGCCACGACCGCTCCAGGTCAGGCCATCCTCAT 2150
 TCTGATAAAGCGCGGGCAGTGTGGTTTTCCCGGCACTTGCCTTGCCCGCC 2200
 TTGCCGATTTTCGATTTGAAACCAATTCTTTCAGCGTGATGTTGTGAAG 2250
 CGCGATCAAGGCCTTGATCTGCTCGATGATGGGCTCTCTTTCCTTGGCCC 2300
 GTGCTTCTTCAATTTTCTGCGCCAATTCTTCTTGCTGCGCCAACAGTTTT 2350
 GCCAATTCGCTATTTGCCATAGATCCTCCGGGGGGAAAAAATTCAGGTGG 2400
 GACCGACAATAAGCCAGCTGGCAAACCTAGGGAAGGTGCGAACAAGTCCCT 2450

orf5 →

GATATGAGAT**CATG**TTTGTCTGAGCCATAGAACAGGGTTCATCATG 2500
 AGTCATCAACTTACCTTCGCCGACAGTGAATTCAGCAGTAAGCGCCGTCA 2550
 GACCAGAAAAGAGATTTTCTTGTCCCGCATGGAGCAGATTCTGCCATGGC 2600
 AAAACATGGTGGAAAGTCATCGAGCCGTTTTACCCCAAGGCTGGTAATGGC 2650
 CGGCGACCTTATCCGCTGGAAACCATGCTACGCATTCACTGCATGCAGCA 2700
 TTGGTACAACCTGAGCGATGGCGCGATGGAAGATGCTCTGTACGAAATCG 2750
 CCTCCATGCGTCTGTTTGCCCGGTTATCCCTGGATAGCGCCTTGCCGGAC 2800
 CGCACCACCATCATGAATTTCCGCCACCTGCTGGAGCAGCATCAACTGGC 2850
 CCGCCAATTGTTCAAGACCATCAATCGCTGGCTGGCCGAAGCAGGCGTCA 2900
 TGATGACTCAAGGCACCTTGGTCGATGCCACCATCATTGAGGCACCCAGC 2950
 TCGACCAAGAACAAGAGCAGCAACGCGATCCGGAGATGCATCAGACCAA 3000
 GAAAGGCAATCAGTGGCACTTTGGCATGAAGGCCACATTGGTGTGATG 3050
 CCAAGAGTGGCCTGACCCACAGCCTGGTCACCACCGCGGCCAACGAGCAT 3100
 GACCTCAATCAGCTGGGTAATCTGCTGCATGGAGAGGAGCAATTTGTCTC 3150
 AGCCGATGCCGGCTACCAAGGGGCGCCACAGCGCGAGGAGCTGGCCGAGG 3200
 TGGATGTGGACTGGCTGATCGCCGAGCGCCCCGGCAAGGTAAGAACCTTG 3250
 AAACAGCATCCACGCAAGAACAACGCGCCATCAACATCGAATACATGAA 3300
 AGCCAGCATCCGGGCCAGGGTGGAGCACCCATTTTCGCATCATCAAGCGAC 3350
 AGTTCGGCTTCGTGAAAGCCAGATACAAGGGGTGCTGAAAAACGATAAC 3400
 CAACTGGCGATGTTATTACGCTGGCCAACCTGTTTCGGGCGGACCAAAT 3450
 GATACGTCAGTGGGAGAGATCTCACTAAAAACTGGGGATAACGCCTTAAA 3500
 TGGCGAAGAAACGGTCTAAATAGGCTGATTCAAGGCATTTACGGGAGAAA 3550

```

AAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA 3600
TTTCCCCGCTTATTCGCACCTTCCCTAGAGCAACTATTTTTTCAATTTAT 3650
TCTTTCAGGATGATAAAAAGACTTTTATCTGGCGCTGCTGCTGAGCTGAT 3700
GTAAATTGCGGATAAGGCAATCCATTGAGCGTTCTTATCTTCATTCTGCC 3750
TGCGCAGAAAATAACAAAAGCAAATAAAAGCAAATAAAAAACAAGAAGAAA 3800
GAATTCAAAAAATCAGAAAGAAAGTCGGTATTGGCGCCCTTCCAAAAAG 3850
TTCCCCTTAACGGCTCACCATAACAGTTCGCCCCAGCGGGGTGCGGAAGA 3900
CCCCGCGGAAGCTGCCTGCCGAAACCATTACCATCCCACGAAACACCCA 3950
CAGGTGCCGGCAGCAAGGCTGCGCCGGCTAGTGGGTACCGTCAGCAAGA 4000
CCTTGCCCTTGCTGGTCTGGCCATCGGTCCAGCTGGTCTGGGTGGCCGCG 4050
GGCGGGTGAAGCTGGCCATGGCCTGCACCAGCTTGTCACGTCGCTGTC 4100
GTAGAGCGTCTTGCCATCGGCAGTCTTGATGCGCTCGACGCGATTATCCG 4150
CGCCGGCATAACCAATCCTTGACGGTGACCTGGTCAGTGCTTCCAGCACG 4200
TCGATCTGCAAGTCCTTGCCGACCTGACTGAACCACAGGTTGGTTTGCCT 4250
GGCGTCGCTGAACCTTGATCGTATCAATATTGCCCCGCCGTGGCATCCTGGT 4300
CCACCACCAGATCGCGACCATCGCCCTTACCGAAGACGTAGGTGTCATTG 4350
CCCGCTTGCCCGGTCAGCGTATCGTTACCCTTGGCGCCGGTGAGGAGGTT 4400
GTTATCCTGATTGCCACCAGGATGTTGTCGAGTTCGTTGCCGGTGGCGC 4450
TGCTGTCGCGATGGAGCAGCTTGGCCGGGTCCACGCTGTAGATGAGGGCT 4500
GATTCGTGCGCGTCCATGTTGCGCAAGTCCCCTACTGACAGGAAGCGGTT 4550
CATTTGCGTGTAACGTCTTGGTCCCGAGTCGGTGATATAGAAACCCAGCA 4600
CCTCACCCGTCTTGGCCGAGCAAGCCACGCCGCTCACCGCCACCGCATGG 4650
AATCCCACTCGCTTCCCCGCAAGGTAGCCACTCATGTCCCGGAACACCGC 4700
GGCGTCCGCGCACAAGGTAACACATGGCCGCCCTTGAGCAACTTGACCA 4750

← orf6

GCCGTGCCTCGTCGTATCTGAGTCGAC 4777

```

Figura 9. Sequência de nucleotídeos da região sequenciada a montante do gene *nifA* de *Herbaspirillum seropedicae*. São mostradas as sequências completas dos genes *modE* e *orf5* e as sequências parciais dos genes *orf4* e *orf6*. Um possível sítio de ligação de ribossomo está sublinhado com linha pontilhada. Os códons de início de tradução estão em negrito. Os códons de término de tradução estão sublinhados com linha contínua. As setas indicam a direção de transcrição.

A sequência de nucleotídeos da região sequenciada foi analisada pelo método de preferência de códons do programa Analyseq (STADEN, 1982) para a identificação de prováveis regiões codificadoras de proteína (ORFs). Por meio deste método, os códons encontrados nas três fases de leitura da região sequenciada, em ambas as direções, foram comparados com uma tabela de preferência de códons de *H. seropedicae* (CHUBATSU, 1998). Esta comparação foi representada por gráficos contendo três seções, referentes às fases de leitura 1, 2 e 3 (figuras 10 e 11). Nestas seções, as prováveis ORFs correspondem a regiões do gráfico situadas acima da linha pontilhada, a qual representa a probabilidade de 50% de ocorrência de uma ORF na sequência, conforme a preferência de códons de *H. seropedicae*.

A análise da região sequenciada pelo método de preferência de códons permitiu a identificação de quatro prováveis ORFs, designadas *orf4* (figura 10, fase de leitura 3), *modE* (figura 10, fase de leitura 2), *orf5* (figura 10, fase de leitura 2) e *orf6* (figura 11, fase de leitura 3). ORFs denominadas *orf1*, *orf2* e *orf3* foram previamente descritas para *H. seropedicae* por KLASSEN (1998) e KLASSEN *et al.* (1999). Estes resultados foram confirmados através da comparação das sequências de nucleotídeos correspondentes a cada ORF com as sequências depositadas no banco de dados GenBank pelo programa Blast 2.0 (ALTSCHUL *et al.*, 1997), para verificação da sua identidade e similaridade. Além disso, estas sequências de nucleotídeos foram traduzidas para sequências de aminoácidos pelo programa Strider 1.3 (Perkin-Elmer, Foster, EUA) para confirmar a localização de códons de início e término de tradução.

Os genes *modE* e *orf5* foram completamente sequenciados e os genes *orf4* e *orf6* foram sequenciados parcialmente. Entre os genes *modE* e *orf5*, foi encontrada uma extensa região intergênica composta por 1205 bp (figuras 9 e 10). A organização destes genes na região sequenciada está representada na figura 12.

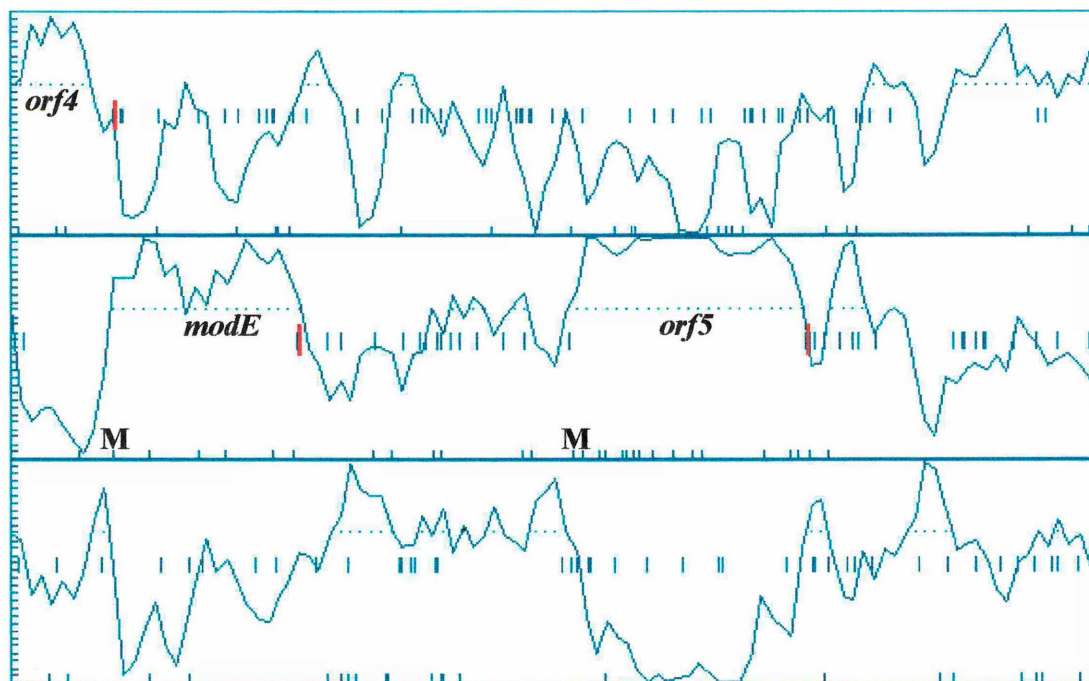


Figura 10. Identificação das prováveis regiões codificadoras de proteína dos genes *orf4*, *modE* e *orf5* de *Herbaspirillum seropedicae* através do método de preferência de códons do programa Analyseq (STADEN, 1982). A sequência de nucleotídeos obtida foi analisada utilizando-se uma tabela de preferência de códons de *H. seropedicae* (CHUBATSU, 1998). Cada seção do gráfico representa uma fase de leitura (fases de leitura 1, 2 e 3, de baixo para cima). O primeiro aminoácido (metionina) está representado pela letra M. Os códons de término de tradução estão representados por um traço vermelho no centro da seção.

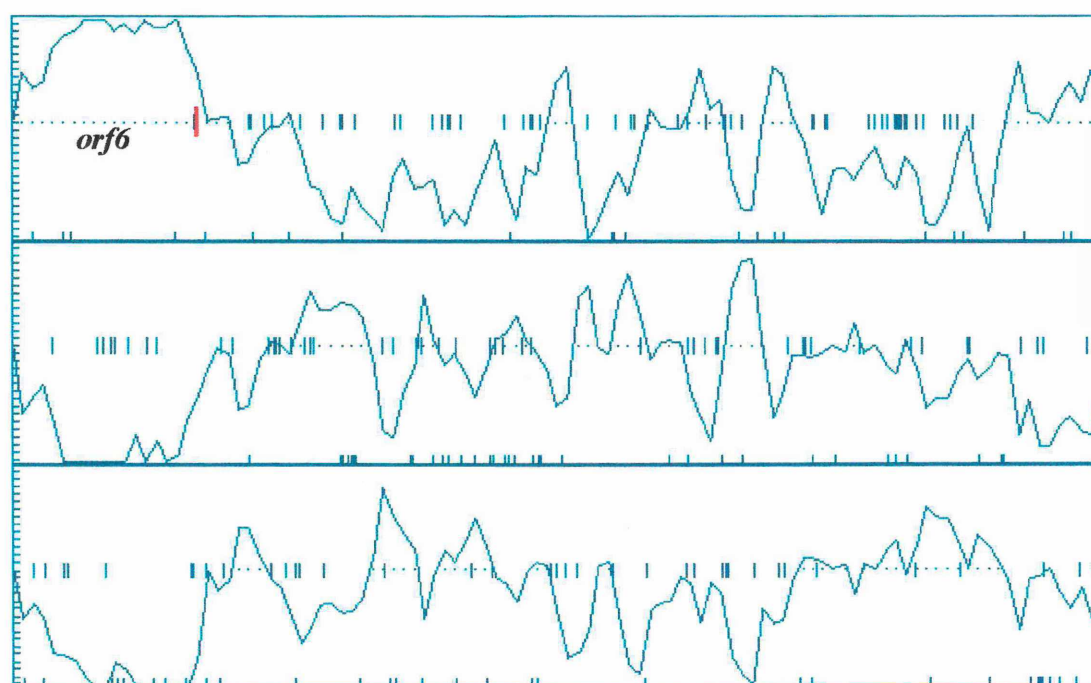


Figura 11. Identificação da provável região codificadora de proteína do gene *orf6* de *Herbaspirillum seropedicae* através do método de preferência de códons do programa Analyseq (STADEN, 1982). A sequência de nucleotídeos obtida foi analisada utilizando-se uma tabela de preferência de códons de *H. seropedicae* (CHUBATSU, 1998). Cada seção do gráfico representa uma fase de leitura (fases de leitura 1, 2 e 3, de baixo para cima). O códon de término de tradução está representado por um traço vermelho no centro da seção.

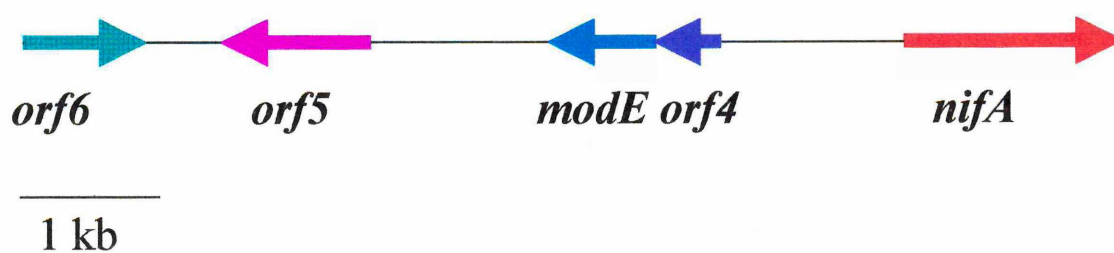


Figura 12. Representação das ORFs identificadas em *Herbaspirillum seropedicae* a montante do gene *nifA*. A região seqüenciada corresponde à região entre os genes *orf6* e *orf4*. As setas indicam a direção de transcrição.

3.2.1 *orf4*

A região distal da *orf4* (nucleotídeos 1-449, figura 9) foi identificada a aproximadamente 1,3 kb do gene *nifA*, apresentando direção de transcrição oposta a deste gene. O códon de término de tradução (TGA) da *orf4* está sobreposto ao códon de início de tradução (ATG) da ORF seguinte, o gene *modE*. A jusante do códon de término de tradução, não foi encontrado terminador típico (figura 13). Estes dados sugerem que os genes *orf4* e *modE* fazem parte de um operon.

```

GAT CTG GCC GCA CGC CAG GAT TAC GGA TTG ATG GCC ACA CCC TCC GCC GAT CTT GAG CAG
D   L   A   A   R   Q   D   Y   G   L   M   A   T   P   S   A   D   L   E   Q
GTG CTC ACG CGG GTG GCT GCA CAG CGT GCG CAG CGC GGT CAC TTC CCG TTT GCC ATC GCC
V   L   T   R   V   A   A   Q   R   A   Q   R   G   H   F   P   F   A   I   A
CGC TTT TTC CAT TGC GAA CCC GAG GCG CAC AAG GCG CTT GCC CAT GGT CTG GCA GGT GTG
R   F   F   H   C   E   P   E   A   H   K   A   L   A   H   G   L   A   G   V
CGT GCG TAT CCG CGC ATG ATG CAA GTG CAA TTC GCG CCC GGT GGC ACC CCG CCC ATG TTC
R   A   Y   P   R   M   M   Q   V   Q   F   A   P   G   G   T   P   P   M   F
CAG GCC GAT GCC TCC GAT CAA GTC ATC CTG GCC AAC GCC ATC ATC GGC GAC CCG GCG CAA
Q   A   D   A   S   D   Q   V   I   L   A   N   A   I   I   G   D   P   A   Q
TGC GTG GCA CGC ATC GGC CAG CTG CAA GAA AAA CTG GGG CCG CAT CGC CTG CTG CTC AAG
C   V   A   R   I   G   Q   L   Q   E   K   L   G   P   H   R   L   L   L   K
CCT GCC AGC CAC GAT CCC CAG CAA GCG CGC GCA GCC CTG AGC TTG CTG GCG GGC GCG CTG
P   A   S   H   D   P   Q   Q   A   R   A   A   L   S   L   L   A   G   A   L
GGA CTG CAG CCA TCG GCA AGG ACA TGA GTA CCT TGC CTG CCA CCG ATA CCA CCT TGT TGA
G   L   Q   P   S   A   R   T   *
GCA GCG AAC TGA AAC TGG TGC ATC GCC TGG ACC AGC GCT TCT TTG CCC TGC TCG ACG CCA
TCG CGC AAA CCG GCT CGA TCA ATC GCG CGG CCA GTA CGG CGG GCT ACA GCT ACA AGG GCG
CAT GGA TGC TGC TGG AGT CAG CCG GCA ACC TGG TCA ATG GCG CAT TGA TTG AAA CGG TCA

```

Figura 13. Seqüência de nucleotídeos da região distal da *orf4* e seqüência de aminoácidos deduzida da região C-terminal da proteína Orf4 de *Herbaspirillum seropedicae*. O códon de término de tradução está marcado com um asterisco.

A sequência de aminoácidos deduzida da região distal da *orf4* apresentou baixa similaridade com proteínas identificadas em espécies filogeneticamente distantes, cujas funções não são relacionadas. A partir destes dados, não foi possível deduzir a potencial função da *orf4* em *H. seropedicae*. No entanto, esta ORF pode estar envolvida com o transporte de molibdênio, uma vez que é contígua ao gene *modE* e provavelmente é expressa conjuntamente com este gene, relacionado com este processo. O seqüenciamento da região proximal desta ORF e a análise fisiológica do seu produto poderão auxiliar na determinação da sua função.

3.2.2 Gene *modE*

O gene *modE* (nucleotídeos 446-1255, figura 9) foi localizado a cerca de 1,8 kb do gene *nifA* e é transcrito na direção oposta. A montante do códon de início de tradução (ATG), foi encontrado um possível sítio de ligação de ribossomo (AGGA) e a jusante do códon de término de tradução (TGA), foi identificado um provável terminador intrínseco (figura 14).

A sequência de aminoácidos deduzida do gene *modE* apresentou similaridade com proteínas ligantes de molibdênio, principalmente com o repressor de transcrição ModE de *Azotobacter vinelandii* (37% de identidade e 53% de similaridade) (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995), de *E. coli* (33% de identidade e 57% de similaridade) (GRUNDEN *et al.*, 1996) e de *Haemophilus influenzae* (26% de identidade e 63% de similaridade) (FLEISCHMANN *et al.*, 1995) e com as proteínas ligantes de molibdopterinas MopB (33% de identidade e 59% de similaridade) e MopA (31% de identidade e 63% de similaridade) de *Rhodobacter capsulatus* (WANG *et al.*, 1993).

```

TGA TCT GGC CGC ACG CCA GGA TTA CGG ATT GAT GGC CAC ACC CTC CGC CGA TCT TGA GCA
GGT GCT CAC GCG GGT GGC TGC ACA GCG TGC GCA GCG CGG TCA CTT CCC GTT TGC CAT CGC
CCG CTT TTT CCA TTG CGA ACC CGA GGC GCA CAA GGC GCT TGC CCA TGG TCT GGC AGG TGT
GCG TGC GTA TCC GCG CAT GAT GCA AGT GCA ATT CGC GCC CGG TGG CAC CCC GCC CAT GTT
CCA GGC CGA TGC CTC CGA TCA AGT CAT CCT GGC CAA CGC CAT CAT CGG CGA CCC GGC GCA
ATG CGT GGC ACG CAT CGG CCA GCT GCA AGA AAA ACT GGG GCC GCA TCG CCT GCT GCT CAA
GCC TGC CAG CCA CGA TCC CCA GCA AGC GCG CGC AGC CCT GAG CTT GCT GGC GGC CGC GCT
GGG ACT GCA GCC ATC GGC AAG GAC ATG AGT ACC TTG CCT GCC ACC GAT ACC ACC TTG TTG
                                M   S   T   L   P   A   T   D   T   T   L   L
AGC AGC GAA CTG AAA CTG GTG CAT CGC CTG GAC CAG CGC TTC TTT GCC CTG CTC GAC GCC
S   S   E   L   K   L   V   H   R   L   D   Q   R   F   F   A   L   L   D   A
ATC GCG CAA ACC GGC TCG ATC AAT CGC GCG GCC AGT ACG GCG GGC TAC AGC TAC AAG GGC
I   A   Q   T   G   S   I   N   R   A   A   S   T   A   G   Y   S   Y   K   G
GCA TGG ATG CTG CTG GAG TCA GCC GGC AAC CTG GTC AAT GGC GCA TTG ATT GAA ACG GTC
A   W   M   L   L   E   S   A   G   N   L   V   N   G   A   L   I   E   T   V
ACC GGC GGC AAA GGC GGC GGC GGC ACC CGG CTC ACG CCC GCC GCG GTG GAA CTG CTG GCC
T   G   G   K   G   G   G   G   T   R   L   T   P   A   A   V   E   L   L   A
GTG TGG CGC GAG CTG CAA CGG CGC AAT CTC GAA TTC CTG CAT CGG CAG GAG ACA TGG CTT
V   W   R   E   L   Q   R   R   N   L   E   F   L   H   R   Q   E   T   W   L
AAT CAA CTG CCA GCC TTG GCT GGC CTG CTC AGG AGA ATG TCA ATG AAA ACC AGC GCA CGT
N   Q   L   P   A   L   A   G   L   L   R   R   M   S   M   K   T   S   A   R
AAT CAA TTC GCC GGC GTC ATC AGT GCG ATC GAT ACC GGC CCT GTC ACT ACG CAA GTC ACC
N   Q   F   A   G   V   I   S   A   I   D   T   G   P   V   T   T   Q   V   T
GTC ACC ATC GCC GGC GCG CAA GAG ATC GTG GCC ACC ATG ACT ACG ACC GCC CAA ATC CGG
V   T   I   A   G   A   Q   E   I   V   A   T   M   T   T   T   A   A   N   R
CTC AAG CTC AGG ATC GGC AGC AAT GCG ATT GCC CTG ATC AAG TCC TCG GCG GTG GTG CTG
L   K   L   R   I   G   S   N   A   I   A   L   I   K   S   S   A   V   V   L
GTC ACC GAT TTT GCC GGT TTT TCC CTC TCG GCA CGC AAT CAG TTT GAA GGC ACG GTC TCG
V   T   D   F   A   G   F   S   L   S   A   R   N   Q   F   E   G   T   V   S
CGC GTA GAG CGT GGC GCG GTA TCG TCG CTT GTG GTG TTG ACC TTG CCT GGC GGC GCC TGT
R   V   E   R   G   A   V   S   S   L   V   V   L   T   L   P   G   G   A   C
ATG ACC GCC AGC CTG ACC AAT GAC GCC ATC GAT GCG CTG TCG CTG GCG GTC GGG CAG ACC
M   T   A   S   L   T   N   D   A   I   D   A   L   S   L   A   V   G   Q   T
GCT ACC GCC GTC TTC AAG GCG TAT GCA GTG ATG GTG GCC GTG CAG CAG GAT TGA TTA CCT
A   T   A   V   F   K   A   Y   A   V   M   V   A   V   Q   Q   D   *
ACT GCG GTC ATC CTT CAT TCT GGG TTG CTG ACA CCA GTG TGT CAG GGT CAA GTG CCG GAT
CCC GCA CAT AAA CCG ACC GTA TCT TCT GCG TCG CCC CCG TCC GCA GAA CTG GCC CCT CAA
TGA TAG GCC AGT TCA TTT CTT CCA GGC GAG CGA TCA CCT TCT CCA ACG GCA CGG ACG CCA
TGA AGC ACA GGT CCA GCG CGC CCG GGC ACA GGT AGA TGC GCT TTC GGC TCG AAT TCC TTG
CCG CGT ACA TGC AGG TTG ATT TTC TGG TTA CCG AAA CGA AAG GCG ATA CGC CCT TGC CCG
AAG GTC TCC AGC GTC ATG CCC AGC ACT TCC ACA TAG AAG TGG GTC GTG GCG TCG GGG TTG
ATC GAA GTC AGT ACG AGA TGG TCA AGA TGG TCA ATC AGG CTC ATG TCA GGT CCG TGG AGT
GTT CAG CCT CAA GCG GAT AGA GCA GAT GAG TGA CCA ATC CCT GCG GGG TCT TTC ACG CTT
CAA TCC GCT GAT ATA GCC GCA GCA GAT AAA AAC AGG AAC GAT CGG GAT TAA TTA AAA AGC

```

Figura 14. Sequência de nucleotídeos do gene *modE* e sequência de aminoácidos deduzida da proteína ModE de *Herbaspirillum seropedicae*. Um possível sítio de ligação de ribossomo está destacado em azul. O códon de término de tradução está marcado com um asterisco. Um provável terminador intrínseco está destacado em verde.

Os genes *mod* estão envolvidos com o transporte de molibdênio sob a forma de íons molibdato (LUQUE *et al.*, 1993; WANG *et al.*, 1993; MOUNCEY *et al.*, 1995; MAUPIN-FURLOW *et al.*, 1995). Em *A. vinelandii* (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995), *E. coli* (MAUPIN-FURLOW *et al.*, 1995) e *R. capsulatus* (WANG *et al.*, 1993), os genes *modABC* codificam para proteínas que integram um sistema de transporte de íons molibdato de alta afinidade, as quais fazem parte da família de transportadores ABC. Nestes organismos, ModA é uma proteína periplasmática ligante de molibdato, ModB é uma proteína transmembrana formadora de canal e ModC é uma proteína ligante de ATP.

Em *A. vinelandii* (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995) e *E. coli* (GRUNDEN *et al.*, 1996), o gene *modE* codifica para um repressor de transcrição dos genes *modABC*. Em *E. coli*, a proteína ModE se liga ao íon molibdato quando este se encontra em altas concentrações, formando o complexo ModE-molibdato. Este complexo forma um dímero e se liga a um operador localizado a montante do gene *modA*, impedindo a transcrição dos genes *modABC* (McNICHOLAS *et al.*, 1998; GRUNDEN *et al.*, 1999).

Os genes *mopA* e *mopB* de *R. capsulatus* (WANG *et al.*, 1993) e *mopI*, *mopII* e *mopIII* de *Clostridium pasteurianum* (HILTON *et al.*, 1987) são similares ao gene *modE* de *A. vinelandii* (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995) e de *E. coli* (GRUNDEN *et al.*, 1996) e aparentemente desempenham funções semelhantes.

Em *H. influenzae* (FLEISCHMANN *et al.*, 1995), foram identificados genes similares aos genes *modABC* e *modE* descritos para as espécies citadas.

Os genes *modG* de *A. vinelandii* (MOUNCEY *et al.*, 1995), *modD* (MAUPIN-FURLOW *et al.*, 1995) e *modF* (GRUNDEN *et al.*, 1996) de *E. coli* e *modD* de *R. capsulatus* (WANG *et al.*, 1993) codificam para proteínas cuja função ainda não foi determinada.

Os genes *mod* foram descritos inicialmente para *R. capsulatus* (WANG *et al.*, 1993), *A. vinelandii* (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995), *H. influenzae*

(FLEISCHMANN *et al.*, 1995) e *E. coli* (MAUPIN-FURLOW *et al.*, 1995; GRUNDEN *et al.*, 1996). Nestes organismos, os genes *mod* estão organizados em operons contíguos e orientados em direções opostas, ocupando um único *locus* no genoma (GRUNDEN *et al.*, 1996) (figura 15). Recentemente, os genes *modABC* foram identificados em *H. seropedicae* a jusante do operon *nifHDKENXorf1orf2*, compondo o operon *nifQmodABC* (KLASSEN, 2000). Neste trabalho, o gene *modE* de *H. seropedicae* foi identificado a montante do gene *nifA* e a jusante da *orf4*. Os genes *orf4* e *modE* são transcritos na mesma direção e o códon de término de tradução da *orf4* é sobreposto ao códon de início de tradução do gene *modE*. Como não foi encontrado promotor típico a montante da sequência estrutural do gene *modE*, é possível que os genes *orf4* e *modE* façam parte de um operon. Assim sendo, os genes *mod* estariam organizados em *H. seropedicae* de uma forma diferente da descrita para outras espécies, ocupando, pelo menos, dois *loci* distintos no genoma (figura 15). Estas diferenças estruturais provavelmente refletem diferenças filogenéticas entre estes organismos.

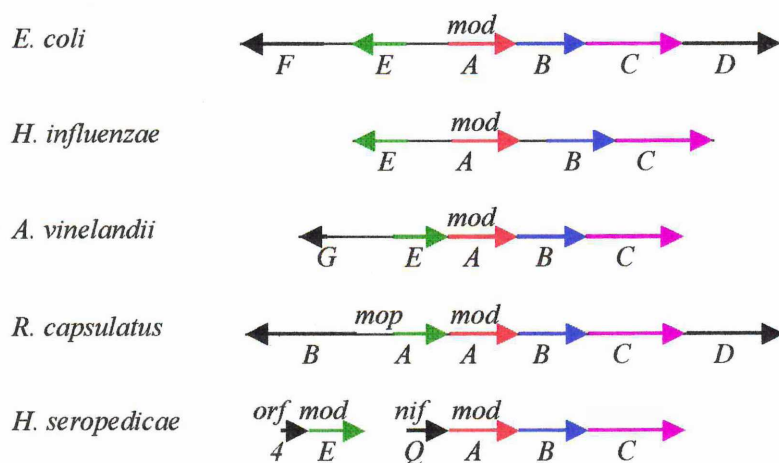


Figura 15. Comparação da organização estrutural dos genes *mod* e/ou *mop* em *E. coli*, *H. influenzae*, *A. vinelandii*, *R. capsulatus*, e *H. seropedicae*. As setas indicam a direção de transcrição destes genes. Genes que apresentam funções similares estão marcados com a mesma cor. Genes que possuem funções distintas ou indeterminadas estão representados em preto. Adaptado de GRUNDEN *et al.* (1996) e KLASSEN (2000).

A proteína ModE de *H. seropedicae* é formada por 269 resíduos de aminoácidos e apresenta massa molecular de cerca de 28,3 kDa, calculada pelo programa BioEdit 4.8.8 (HALL, 1999). O seu ponto isoelétrico foi estimado em aproximadamente pH 9,8 pelo programa “EMBL WWW Gateway to Isoelectric Point Service” do banco de dados EMBL. Este valor indica que a proteína ModE deve apresentar caráter básico e, deste modo, afinidade por DNA.

A proteína ModE de *H. seropedicae* foi comparada com as proteínas ModE de *A. vinelandii* (LUQUE *et al.*, 1993; MOUNCEY *et al.*, 1995), de *H. influenzae* (FLEISCHMANN *et al.*, 1995) e de *E. coli* (GRUNDEN *et al.*, 1996) e com as proteínas MopA e MopB de *R. capsulatus* (WANG *et al.*, 1993) (figura 16). Através desta comparação, foi possível identificar motivos e domínios característicos na proteína ModE de *H. seropedicae*. Na região N-terminal, foi encontrado um provável motivo hélice-volta-hélice, que pode estar envolvido com a ligação desta proteína ao DNA (McNICHOLAS *et al.*, 1998; GRUNDEN *et al.*, 1999). Na região C-terminal, foram identificados dois motivos conservados (SARNQ), os quais podem estar relacionados com a dimerização da proteína ModE (McNICHOLAS *et al.*, 1998). Estes motivos foram localizados nos domínios MopE1 e MopE2 (McNICHOLAS *et al.*, 1998), cuja sequência de aminoácidos é semelhante a das proteínas MopI, MopII e MopIII de *C. pasteurianum* (HILTON *et al.*, 1987). Nos domínios MopE1 e MopE2, pode ocorrer a ligação de íons molibdato (McNICHOLAS *et al.*, 1998).

A semelhança estrutural entre a proteína ModE de *H. seropedicae* e a de outros organismos sugere que esta proteína pode regular a expressão dos genes *modABC* em resposta à concentração de íons molibdato. Experimentos de mutagênese sítio-dirigida e análise fisiológica do gene *modE* poderão esclarecer o seu papel na regulação da expressão dos genes *modABC* nesta espécie.

		10	20	30	40	50			
1 Rhoca_MopB	MAATKQGG	GDDGRCAR	GVVLERTGA	RMGAERVALLA	AI	GRTGSISAAAR	49		
2 Rhoca_MopA	-----	MNEQPLIA	AALSIO	RAGAPRV	GGDRIRLLE	AIARHGTTIAGAA	42		
3 Azovi_Mode	-----	MTATRFLA	MSIDTDV	GTALSD	TRIRLLE	AIERE	GSINRAAK	42	
4 Herse_Mode	-----	MSTLPATD	TTLLSSE	LKLVHRLD	QORFFALLD	AI	AO	TSINRAAS	44
5 Ecoli_Mode	-----	MQAEI	LLTIK	LQOKL	FADPRRIS	LLKHIALS	GSISQ	GAK	39
6 Haein_Mode	-----	MKNTEI	LLTIK	LQOAL	FIDPKRV	LLKEIQ	QCGSIN	QAAK	40
7 Consenso	-----		LLSL L	A R D	RIRLL	AIAR	GSIN	AAK	
		60	70	80	90	100			
1 Rhoca_MopB	EVGLSYKA	AWDGVQAM	NNLLAA	PVVTAAP	GGKAGGG	AVLTPAGE	KLIAAY	99	
2 Rhoca_MopA	EVGLSYKT	AWDAVGT	LNNLFEQ	PLVEAAP	GGRTGGN	ARVTEAG	QALIAGF	92	
3 Azovi_Mode	VVPLSYKA	AWDAIDT	MNNLAPE	PLVVRVAG	GRQGGGT	QLTDYGR	RIVAMY	92	
4 Herse_Mode	TAGYSYK	GAWMLLES	AGNLVNG	ALIVT	TGGKGGG	TRLTPAA	VELLAW	94	
5 Ecoli_Mode	DAGISYKS	AWDAINEM	NQLSEH	ILVERAT	GGKGGG	AVLTRY	QORLIQLY	89	
6 Haein_Mode	NAKVSYS	AWDHL	EAMNKIS	PRPLERNT	GGKNGG	TALT	TYAERLLQLY	90	
7 Consenso	VGLSYK	AWDA	MNNL	PLVERAT	GGK	GGGA	LT AG	RLIA Y	
		110	120	130	140	150			
1 Rhoca_MopB	GAIEAGVAK	LLSSFEK	-----	SINLDPAE	-VLRGLSLRT	SARNAWACKV	142		
2 Rhoca_MopA	GLLEGALT	KALGVLE	GGV----	SAPEKALN	-TLWSLT	MRTSNRNT	TLRCTV	137	
3 Azovi_Mode	RALEIEYQ	SALDRLS	ERINEVT	GGDIQAF	QRLMHS	MSMKT	SARNOFAGIV	142	
4 Herse_Mode	RELQRRNL	EFLHRQ	ETWLN----	QLPALAGL	LRRMSMKT	SARNOFAGVI	139		
5 Ecoli_Mode	DLLAQIQK	KAFDVL	SDDD----	ALPLNSLL	AAISRF	SLO	SARNQWFGTI	135	
6 Haein_Mode	DLLERTQE	HAFHILQ	DE----	SVPLDSIL	TATARF	SLO	SARNQFFGRV	135	
7 Consenso	LLE	KAL	LE	----	S PL AL	L R SL	SARNOFAG V		
		160	170	180	190	200			
1 Rhoca_MopB	WSVAADDV	AAQVRMR	LGEQD	-LTAVIT	TARSA	AEMLAPG	SEVLALVKSN	191	
2 Rhoca_MopA	TRVTLGAV	NAEVELA	LT	DGHS	-LTAVIT	TERSATEM	GLAPGVEVFALIKAS	186	
3 Azovi_Mode	TGLRVGGV	DYEVRI	RDAENE	-IAAVIT	TKASA	ENLELA	IGKEVFALVKSS	191	
4 Herse_Mode	SAIDTGPV	TTQVTV	TTAGAQE	-IVATMT	TTAANRL	KLRIGS	NAIALIKSS	188	
5 Ecoli_Mode	TARDHDDV	QQHV	DLADG	KTRIKVAIT	AQSGARL	GLDEG	KEVLILLKAP	185	
6 Haein_Mode	AQQRIDS	RCVVDV	NVQGLPT	PIQVSIT	TKSSARL	KLITE	KEVMLMFKAP	185	
7 Consenso	T	GDV	V V L G	-L AVIT	SAARL	LA	GKEV AL KSS		
		210	220	230	240	250			
1 Rhoca_MopB	FVLLAGAGV	PERLSVR	NVRGRVIER	IDAPLSSE	VTDLGG	GKTI	TATIT	241	
2 Rhoca_MopA	FVMLAAGG	DPGRIS	ACNRLTG	IVAARTD	GPVNTEI	ILD	GNCKSITAVIT	236	
3 Azovi_Mode	SVMLTTEP	-SLKLT	ARNQLWGE	VIDIHE	GPVNNEV	TIALPS	GRSVTCVVT	240	
4 Herse_Mode	AVVLVTDF	AGFSL	SARNQFEGT	VS	SRVERG	AVSSLV	VTLP	GGACMTASLT	238
5 Ecoli_Mode	WVGITQDE	-AVAQ	NADNQLPGI	ISHIERG	AEQCEV	LMALPD	GQTL	CATVP	234
6 Haein_Mode	WVKISEQP	-LANQ	---PNQFP	---VNIKS	LN	EEAILQ	FAESNIE	FCATV	228
7 Consenso	V L	-	LSARNQL	G V I	GPV	EV L LP G	TATVT		

		260	270	280	
1	Rhoca_MopB	RDSAEMLDLHPGVET	TALIKSSHVILALP	--	270
2	Rhoca_MopA	HTSADALGLAPGP	PATALFKASHVILAMP	--	265
3	Azovi_Mode	ADSCALGLAPGVA	CAFFKSSSVILAVYG		270
4	Herse_Mode	NDAIDALSLAVGQT	TATAVFKAYAVMVAVQ	QD	269
5	Ecoli_Mode	--VNEATSLQQGQ	NVTAYFNADSVIIATLC		262
6	Haein_Mode	---HQPNQWQIEQQ	VWIHIDQEQIILATLG	--	255
7	Consenso	DS	AL LAPGV ATA FKAS VILA		

Figura 16. Alinhamento da sequência de aminoácidos das proteínas ModE de *H. seropedicae* (Herse_ModE), de *A. vinelandii* (Azovi_ModE), de *H. influenzae* (Haein_ModE) e de *E. coli* (Ecoli_ModE) e das proteínas MopB (Rhoca_MopB) e MopA (Rhoca_MopA) de *R. capsulatus*. O alinhamento foi feito utilizando o programa Antheptrot 4.9 (DELÉAGE *et al.*, 1988). Em vermelho, são mostrados os resíduos de aminoácidos com 100% de identidade; em azul, acima de 75% de similaridade; em verde, acima de 50%; em preto, abaixo de 50%. Motivos conservados das proteínas ModE, MopA e MopB (McNICHOLAS *et al.*, 1998) estão delimitados por retângulos. Um provável motivo hélice-volta-hélice (McNICHOLAS *et al.*, 1998; GRUNDEN *et al.*, 1999) está sublinhado com linha contínua. Os domínios MopE1 e MopE2 (McNICHOLAS *et al.*, 1998) estão sublinhados com linha pontilhada e tracejada, respectivamente.

3.2.3 *orf5*

A *orf5* (nucleotídeos 2462-3478, figura 9) foi identificada a aproximadamente 3,7 kb do gene *nifA* e é transcrita na direção oposta. A montante do códon de início de tradução (ATG), foi encontrado um possível sítio de ligação de ribossomo (GAGA). Além disso, a montante e a jusante da sequência estrutural, foram localizadas repetições invertidas (figura 17).

A sequência de aminoácidos deduzida da *orf5* apresentou similaridade com proteínas codificadas por sequências de inserção (IS), incluindo as transposases Trs5 de *E. coli* (100% de identidade) (ENGLER e BREE, 1981; SCHONER e KAHN, 1981) e Tra6 de *Neisseria meningitidis* (20% de identidade e 50% de similaridade) (KNIGHT *et al.*, 1992).

```

TTG AGC CAC AAT GGA ATA CGG CCA CGA CCG CTC CAG GTC AGG CCA TCC TCA TTC TGA TAA
AGC GCG GGC AGT GTG GTT TTC CCG GCA CTT GCC TTG CCC GCC TTG CCG GAT TTC GAT TTG
AAA CCC AAT TCC TTG AGC GTG ATG TTG TGA AGC GCG ATC AAG GCC TTG ATC TGC TCG ATG
ATG GGC TCT CTT TCC TTG GCC CGT GCT TCT TCA ATT TTC TGC GCC AAT TCT TCT TGC TGC
GCC AAC AGT TTT GCC AAT TCG CTA TTT GCC ATA GAT CCT CCG GGG GGA AAA AAT TCA GGT
GGG ACC GAC AAT AAG CCA GCT GGC AAA CTA GGG AAG GTG CGA ACA AGT CCC TGA TAT GAG
ATC ATG TTT GTC ATC TGG AGC CAT AGA ACA GGG TTC ATC ATG AGT CAT CAA CTT ACC TTC
M F V I W S H R T G F I M S H Q L T F
GCC GAC AGT GAA TTC AGC AGT AAG CGC CGT CAG ACC AGA AAA GAG ATT TTC TTG TCC CGC
A D S E F S S K R R Q T R K E I F L S R
ATG GAG CAG ATT CTG CCA TGG CAA AAC ATG GTG GAA GTC ATC GAG CCG TTT TAC CCC AAG
M E Q I L P W Q N M V E V I E P F Y P K
GCT GGT AAT GGC CGG CGA CCT TAT CCG CTG GAA ACC ATG CTA CGC ATT CAC TGC ATG CAG
A G N G R R P Y P L E T M L R I H C M Q
CAT TGG TAC AAC CTG AGC GAT GGC GCG ATG GAA GAT GCT CTG TAC GAA ATC GCC TCC ATG
H W Y N L S D G A M E D A L Y E I A S M
CGT CTG TTT GCC CGG TTA TCC CTG GAT AGC GCC TTG CCG GAC CGC ACC ACC ATC ATG AAT
R L F A R L S L D S A L P D R T T I M N
TTC CGC CAC CTG CTG GAG CAG CAT CAA CTG GCC CGC CAA TTG TTC AAG ACC ATC AAT CGC
F R H L L E Q H Q L A R Q L F K T I N R
TGG CTG GCC GAA GCA GGC GTC ATG ATG ACT CAA GGC ACC TTG GTC GAT GCC ACC ATC ATT
W L A E A G V M M T Q G T L V D A T I I
GAG GCA CCC AGC TCG ACC AAG AAC AAA GAG CAG CAA CGC GAT CCG GAG ATG CAT CAG ACC
E A P S S T K N K E Q Q R D P E M H Q T
AAG AAA GGC AAT CAG TGG CAC TTT GGC ATG AAG GCC CAC ATT GGT GTC GAT GCC AAG AGT
K K G N Q W H F G M K A H I G V D A K S
GGC CTG ACC CAC AGC CTG GTC ACC ACC GCG GCC AAC GAG CAT GAC CTC AAT CAG CTG GGT
G L T H S L V T T A A N E H D L N Q L G
AAT CTG CTG CAT GGA GAG GAG CAA TTT GTC TCA GCC GAT GCC GGC TAC CAA GGC GCG CCA
N L L H G E E Q F V S A D A G Y Q G A P
CAG CGC GAG GAG CTG GCC GAG GTG GAT GTG GAC TGG CTG ATC GCC GAG CGC CCC GGC AAG
Q R E E L A E V D V D W L I A E R P G K
GTA AGA ACC TTG AAA CAG CAT CCA CGC AAG AAC AAA ACG GCC ATC AAC ATC GAA TAC ATG
V R T L K Q H P R K N K T A I N I E Y M
AAA GCC AGC ATC CGG GCC AGG GTG GAG CAC CCA TTT CGC ATC ATC AAG CGA CAG TTC GGC
K A S I R A R V E H P F R I I K R Q F G
TTC GTG AAA GCC AGA TAC AAG GGG TTG CTG AAA AAC GAT AAC CAA CTG GCG ATG TTA TTC
F V K A R Y K G L L K N D N Q L A M L F
ACG CTG GCC AAC CTG TTT CGG GCG GAC CAA ATG ATA CGT CAG TGG GAG AGA TCT CAC TAA
T L A N L F R A D Q M I R Q W E R S H *
AAA CTG GGG ATA ACG CCT TAA ATG GCG GAG AAA CGG TCT AAA TAG GCT GAT TCA AGG CAT
TTA CGG GAG AAA AAA TCG GCT CAA ACA TGA AGA AAT GAA ATG ACT GAG TCA GCC GAG AAG
AAT TTC CCC GCT TAT TCG CAC CTT CCC TAG AGC AAC TAT TTT TTC AAT TTA TTC TTT CAG
GAT GAT AAA AAG ACT TTT ATC TGG CGC TGC TGC TGA GCT GAT GTA AAT TGC GGA TAA GGC
AAT CCA TTG AGC GTT CTT ATC TTC ATT CTG CCT GCG CAG AAA ATA ACA AAA GCA AAT AAA
AGC AAA TAA AAA CAA GAA GAA AGA ATT CCA AAA AAT CAG AAA GAA AGT CGG TAT TGG CGC
CCT TCC AAA AAG TTC CCC TTA ACG GCT CAC CAT ACC AGT TCG CCC CAG CGG GGT CGC GAA

```

Figura 17. Sequência de nucleotídeos da *orf5* e sequência de aminoácidos deduzida da proteína Orf5 identificada em *Herbaspirillum seropedicae*. Um possível sítio de ligação de ribossomo está marcado em azul. As repetições invertidas terminais estão destacadas em rosa. Os sítios alvo (sequências duplicadas) estão marcados em verde. O códon de término de tradução está indicado com um asterisco.

As seqüências de inserção são seqüências de DNA de até 2,5 kb capazes de se inserir em sítios específicos de moléculas hospedeiras e modificar a sua posição ao longo das mesmas (ENGLER e BREE, 1981; SCHONER e KAHN, 1981; MAHILLON e CHANDLER, 1998). São geralmente formadas por seqüências relacionadas apenas à sua mobilidade. As regiões terminais apresentam repetições invertidas de 10-40 bp e a região central codifica para uma transposase (MAHILLON e CHANDLER, 1998). As repetições invertidas estão envolvidas com a inserção das IS em sítios alvo das moléculas hospedeiras. Estes sítios são seqüências de 2-14 bp que geralmente sofrem duplicação durante o processo de inserção (ENGLER e BREE, 1981; SCHONER e KAHN, 1981; MAHILLON e CHANDLER, 1998). Já as transposases ou recombinases são enzimas que reconhecem e clivam as repetições invertidas, possibilitando a movimentação das IS ao longo das moléculas hospedeiras (MAHILLON e CHANDLER, 1998). As IS podem mediar inversões, deleções ou fusões de fragmentos de DNA e podem alterar a expressão de genes contíguos pela introdução de seqüências de início ou término de transcrição (ENGLER e BREE, 1981). As IS também podem agir na transferência de genes envolvidos com vias catabólicas alternativas e determinantes de virulência (MAHILLON e CHANDLER, 1998).

A proteína Orf5 identificada em *H. seropedicae* é constituída por 338 resíduos de aminoácidos e apresenta massa molecular de cerca de 39,3 kDa, calculada pelo programa BioEdit 4.8.8 (HALL, 1999). O seu ponto isoelétrico foi estimado em aproximadamente pH 9,9 através do programa “EMBL WWW Gateway to Isoelectric Point Service” do banco de dados EMBL. Este valor indica que a proteína Orf5 deve apresentar caráter básico e, portanto, afinidade por DNA, substrato das transposases.

A seqüência de nucleotídeos da *orf5* foi comparada com a do gene *trs5*, contido na IS5 de *E. coli* (ENGLER e BREE, 1981; SCHONER e KAHN, 1981) (figura 18). Esta comparação mostrou que a seqüência de nucleotídeos da *orf5* é idêntica à do gene *trs5*. Além disso, a análise da seqüência de nucleotídeos a montante e a jusante da seqüência estrutural da *orf5* revelou que as repetições invertidas encontradas são iguais às

identificadas na IS5 de *E. coli* e que estas seqüências estão localizadas entre seqüências duplicadas (CTAG), sítios alvo desta IS (figura 17) (ENGLER e BREE, 1981; SCHONER e KAHN, 1981). Assim sendo, conclui-se que esta seqüência encontrada em *H. seropedicae* corresponde à IS5 de *E. coli*.

```

1 Herse_orf5 ATGTTTGTCTCATCTGGAGCCATAGAACAGGGTTCATCATGAGTCATCAACTTACCTTCGCC 60
  |||
2 Ecoli_trs5 ATGTTTGTCTCATCTGGAGCCATAGAACAGGGTTCATCATGAGTCATCAACTTACCTTCGCC 60

1 Herse_orf5 GACAGTGAATTCAGCAGTAAGCGCCGTCAGACCAGAAAAGAGATTTTCTTGTCCTCGCATG 120
  |||
2 Ecoli_trs5 GACAGTGAATTCAGCAGTAAGCGCCGTCAGACCAGAAAAGAGATTTTCTTGTCCTCGCATG 120

1 Herse_orf5 GAGCAGATTCTGCCATGGCAAAACATGGTGGAAAGTCATCGAGCCGTTTTACCCCAAGGCT 180
  |||
2 Ecoli_trs5 GAGCAGATTCTGCCATGGCAAAACATGGTGGAAAGTCATCGAGCCGTTTTACCCCAAGGCT 180

1 Herse_orf5 GGTAATGGCCGGCGACCTTATCCGCTGGAAACCATGCTACGCATTCACTGCATGCAGCAT 240
  |||
2 Ecoli_trs5 GGTAATGGCCGGCGACCTTATCCGCTGGAAACCATGCTACGCATTCACTGCATGCAGCAT 240

1 Herse_orf5 TGGTACAACCTGAGCGATGGCGCGATGGAAGATGCTCTGTACGAAATCGCCTCCATGCGT 300
  |||
2 Ecoli_trs5 TGGTACAACCTGAGCGATGGCGCGATGGAAGATGCTCTGTACGAAATCGCCTCCATGCGT 300

1 Herse_orf5 CTGTTTGCCCGGTTATCCCTGGATAGCGCCTTGCCGGACCGCACCACCATCATGAATTTC 360
  |||
2 Ecoli_trs5 CTGTTTGCCCGGTTATCCCTGGATAGCGCCTTGCCGGACCGCACCACCATCATGAATTTC 360

1 Herse_orf5 CGCCACCTGCTGGAGCAGCATCAACTGGCCCGCCAATTGTTCAAGACCATCAATCGCTGG 420
  |||
2 Ecoli_trs5 CGCCACCTGCTGGAGCAGCATCAACTGGCCCGCCAATTGTTCAAGACCATCAATCGCTGG 420

1 Herse_orf5 CTGGCCGAAGCAGGCGTCATGATGACTCAAGGCACCTTGGTCGATGCCACCATCATTGAG 480
  |||
2 Ecoli_trs5 CTGGCCGAAGCAGGCGTCATGATGACTCAAGGCACCTTGGTCGATGCCACCATCATTGAG 480

1 Herse_orf5 GCACCCAGCTCGACCAAGAACAAGAGCAGCAACGCGATCCGGAGATGCATCAGACCAAG 540
  |||
2 Ecoli_trs5 GCACCCAGCTCGACCAAGAACAAGAGCAGCAACGCGATCCGGAGATGCATCAGACCAAG 540

1 Herse_orf5 AAAGGCAATCAGTGGCACTTTGGCATGAAGGCCACATTGGTGTGCGATGCCAAGAGTGGC 600
  |||
2 Ecoli_trs5 AAAGGCAATCAGTGGCACTTTGGCATGAAGGCCACATTGGTGTGCGATGCCAAGAGTGGC 600

1 Herse_orf5 CTGACCCACAGCCTGGTCACCACCGCGCCAACGAGCATGACCTCAATCAGCTGGGTAAT 660
  |||
2 Ecoli_trs5 CTGACCCACAGCCTGGTCACCACCGCGCCAACGAGCATGACCTCAATCAGCTGGGTAAT 660

```

```

1 Herse_orf5 CTGCTGCATGGAGAGGAGCAATTTGTCTCAGCCGATGCCGGCTACCAAGGGGCGCCACAG 720
  |||||
2 Ecoli_trs5 CTGCTGCATGGAGAGGAGCAATTTGTCTCAGCCGATGCCGGCTACCAAGGGGCGCCACAG 720

1 Herse_orf5 CGCGAGGAGCTGGCCGAGGTGGATGTGGACTGGCTGATCGCCGAGCGCCCCGGCAAGGTA 780
  |||||
2 Ecoli_trs5 CGCGAGGAGCTGGCCGAGGTGGATGTGGACTGGCTGATCGCCGAGCGCCCCGGCAAGGTA 780

1 Herse_orf5 AGAACCTTGAAACAGCATCCACGCAAGAACAAAACGGCCATCAACATCGAATACATGAAA 840
  |||||
2 Ecoli_trs5 AGAACCTTGAAACAGCATCCACGCAAGAACAAAACGGCCATCAACATCGAATACATGAAA 840

1 Herse_orf5 GCCAGCATCCGGGCCAGGGTGGAGCACCCATTTGCGCATCATCAAGCGACAGTTCGGCTTC 900
  |||||
2 Ecoli_trs5 GCCAGCATCCGGGCCAGGGTGGAGCACCCATTTGCGCATCATCAAGCGACAGTTCGGCTTC 900

1 Herse_orf5 GTGAAAGCCAGATACAAGGGGTTGCTGAAAAACGATAACCAACTGGCGATGTTATTACAG 960
  |||||
2 Ecoli_trs5 GTGAAAGCCAGATACAAGGGGTTGCTGAAAAACGATAACCAACTGGCGATGTTATTACAG 960

1 Herse_orf5 CTGGCCAACCTGTTTCGGGCGGACCAAATGATACGTCAGTGGGAGAGATCTCACTAA 1017
  |||||
2 Ecoli_trs5 CTGGCCAACCTGTTTCGGGCGGACCAAATGATACGTCAGTGGGAGAGATCTCACTAA 1017

```

Figura 18. Alinhamento da sequência de nucleotídeos da *orf5* identificada em *H. seropedicae* (*Herse_orf5*) e do gene *trs5* de *E. coli* (*Ecoli_trs5*). O alinhamento foi feito utilizando o programa Blast 2.0 (ALTSCHUL *et al.*, 1997). A sequência de nucleotídeos destes genes apresentou 100% de identidade.

As seqüências de nucleotídeos a montante e a jusante da *orf5* foram analisadas. Quando a seqüência correspondente à IS5 foi extraída e as seqüências contíguas foram unidas, a seqüência resultante não apresentou similaridade significativa com nenhuma seqüência depositada no banco de dados GenBank, o que sugere que a IS5 foi inserida em uma região intergênica.

A IS5 de *E. coli* pode ter sido adquirida por transferência horizontal em condições naturais (MAHILLON e CHANDLER, 1998; DAVISON, 1999) ou durante o processo de construção da biblioteca genômica de *H. seropedicae* Z78, a qual o plasmídeo pEMS1 pertence (SOUZA *et al.*, 1990). A transferência em condições naturais poderia ocorrer no solo ou no interior de uma planta hospedeira. Como *H. seropedicae* é um endófito obrigatório (BALDANI *et al.*, 1997), a sua permanência no solo deve ser efêmera, dificultando o processo de transferência neste ambiente. No entanto, a IS5 de *E. coli* poderia ser adquirida de uma outra bactéria endofítica portadora desta seqüência, durante sua coexistência com *H. seropedicae* nos tecidos de uma mesma planta hospedeira. Já a transferência em condições de laboratório poderia ocorrer durante a permanência do plasmídeo pEMS1 na célula hospedeira, *E. coli* HB101. Um processo semelhante foi descrito anteriormente por SCHONER e KAHN (1981). A transferência da IS5 do genoma de *E. coli* HB101 para um fragmento de DNA de *Haemophilus haemolyticus* clonado no vetor pBR322 ocorreu durante a sua permanência nesta célula hospedeira. Experimentos de hibridização do fragmento correspondente à *orf5* com o DNA cromossômico de *H. seropedicae* Z78 poderão informar sobre a ocorrência desta seqüência no seu genoma.

3.2.4 *orf6*

A região distal da *orf6* (nucleotídeos 3978-4777, figura 9) foi encontrada a cerca de 5,3 kb do gene *nifA* e está orientada na mesma direção de transcrição deste gene. A jusante do códon de término de tradução (TAG) não foi identificado terminador típico (figura 19).

A sequência de aminoácidos deduzida da região distal da *orf6* mostrou similaridade com a região C-terminal de diversas proteínas da família de citotoxinas RTX. Foi verificada maior similaridade com a região C-terminal das hemolisinas-adenilato ciclasas CyaA de *Bordetella pertussis* (30% de identidade e 55% de similaridade) (GLASER *et al.*, 1988) e de *Bordetella bronchiseptica* (28% de identidade e 59% de similaridade) (BETSOU *et al.*, 1995) e com a região C-terminal das proteínas reguladas por ferro FprC (24% de identidade e 58% de similaridade) (THOMPSON *et al.*, 1993b) e FprA (23% de identidade e 59% de similaridade) (THOMPSON *et al.*, 1993a) de *Neisseria meningitidis*.

```

CGA CTC AGA TAC GAC GAG GCA CGG CTG GTC AAG TTG CTC AAG GGC GGC CAT GTG GTT ACC
R   L   R   Y   D   E   A   R   L   V   K   L   L   K   G   G   H   V   V   T
TTG TGC GCG GAC GCC GCG GTG TTC CGG GAC ATG AGT GGC TAC CTT GCG GGG AAG CGA GTG
L   C   A   D   A   A   V   F   R   D   M   S   G   Y   L   A   G   K   R   V
GGA TTC CAT GCG GTG GCG GTG AGC GGC GTG GCT TGC TCG GCC AAG ACG GGT GAG GTG CTG
G   F   H   A   V   A   V   S   G   V   A   C   S   A   K   T   G   E   V   L
GGT TTC TAT ATC ACC GAC TCG GGA CCA GGA CGT TAC AGC GAA ATG AAC CGC TTC CTG TCA
G   F   Y   I   T   D   S   G   P   G   R   Y   S   E   M   N   R   F   L   S
GTA GGG GAC TTG CGC AAC ATG GAC GCG CAC GAA TCA GCC CTC ATC TAC AGC GTG GAC CCG
V   G   D   L   R   N   M   D   A   H   E   S   A   L   I   Y   S   V   D   P
GCC AAG CTG CTC CAT CGC GAC AGC AGC GCC ACC GGC AAC GAA CTC GAC AAC ATC CTG GTG
A   K   L   L   H   R   D   S   S   A   T   G   N   E   L   D   N   I   L   V
GGC AAT CAG GAT AAC AAC CTC CTC ACC GGC AAC GGT AAC GAT ACG CTG ACC GGG CAA
G   N   Q   D   N   N   L   L   T   G   A   K   G   N   D   T   L   T   G   Q
GCG GGC AAT GAC ACC TAC GTC TTC GGT AAG GGC GAT GGT CGC GAT CTG GTG GTG GAC CAG
A   G   N   D   T   Y   V   F   G   K   G   D   G   R   D   L   V   V   D   Q
GAT GCC ACG GCG GGC AAT ATT GAT ACG ATC AAG TTC AGC GAC GCC AGG CAA ACC AAC CTG
D   A   T   A   G   N   I   D   T   I   K   F   S   D   A   R   Q   T   N   L
TGG TTC AGT CAG GTC GGC AAG GAC TTG CAG ATC GAC GTG CTG GGA AGC ACT GAC CAG GTC
W   F   S   Q   V   G   K   D   L   Q   I   D   V   L   G   S   T   D   Q   V
ACC GTC AAG GAT TGG TAT GCC GGC GCG GAT AAT CGC GTC GAG CGC ATC AAG ACT GCC GAT
T   V   K   D   W   Y   A   G   A   D   N   R   V   E   R   I   K   T   A   D
GGC AAG ACG CTC TAC GAC AGC GAC GTG GAC AAG CTG GTG CAG GCC ATG GCC AGC TTC GAC
G   K   T   L   Y   D   S   D   V   D   K   L   V   Q   A   M   A   S   F   D
CCG CCC GCG GCC ACC CAG ACC AGC TGG ACC GAT GGC CAG ACC AGC AAG GGC AAG GTC TTG
P   P   A   A   T   Q   T   S   W   T   D   G   Q   T   S   K   G   K   V   L
CTG ACG GTG ACC CAC TAG CCG GCG CAG CCT TGC TGC CGG CAC CTG TGG GTG TTT CGT GGG
L   T   V   T   H   *
ATG GTG AAT GGT TTC GGC AGG CAG CTT CCG CGG GGT CTT CGC GAC CCC GCT GGG GCG AAC
TGG TAT GGT GAG CCG TTA AGG GGA ACT TTT TGG AAG GGC GCC AAT ACC GAC TTT CTT TCT
GAT TTT TTG GAA TTC TTT CTT CTT GTT TTT ATT TGC TTT TAT TTG CTT TTG TTA TTT TCT
GCG CAG GCA GAA TGA AGA TAA GAA CGC TCA ATG GAT TGC CTT ATC CGC AAT TTA CAT CAG
CTC AGC AGC AGC GCC AGA TAA AAG TCT TTT TAT CAT CCT GAA AGA ATA AAT TGA AAA AAT
AGT TGC TCT AGG GAA GGT GCG AAT AAG CGG GGA AAT TCT TCT CGG CTG ACT CAG TCA TTT
CAT TTC TTC ATG TTT GAG CCG ATT TTT TCT CCC GTA AAT GCC TTG AAT CAG CCT ATT TAG

```

Figura 19. Sequência de nucleotídeos da região distal da *orf6* e sequência de aminoácidos deduzida da região C-terminal da proteína Orf6 de *Herbaspirillum seropedicae*. O códon de término de tradução está marcado com um asterisco.

As citotoxinas RTX são proteínas produzidas por bactérias Gram-negativas que causam patologias em animais. Estas proteínas apresentam atividade de citolisinas ou metaloproteases e são secretadas para o meio extracelular ou permanecem aderidas à superfície celular (THOMPSON *et al.*, 1993a; LOBO e WELCH, 1994). Apresentam em comum um motivo formado por nove resíduos de aminoácidos (LXGGXG^N/_DDX, onde X pode ser qualquer resíduo) repetido em sequência na região C-terminal. Algumas citotoxinas RTX atuam como determinantes de virulência em algumas espécies, mas o seu exato papel no processo de patogênese ainda permanece desconhecido (THOMPSON *et al.*, 1993a; THOMPSON *et al.*, 1993b; ROWE e WELCH, 1994).

B. pertussis (GLASER *et al.*, 1988), *B. bronchiseptica* (BETSOU *et al.*, 1995) e *N. meningitidis* (THOMPSON *et al.*, 1993a) são os agentes etiológicos da tosse comprida humana, da rinite suína e da meningite humana, respectivamente. As proteínas CyaA de *B. pertussis* e de *B. bronchiseptica* aparentemente exercem funções fundamentais no processo de patogênese, sendo consideradas determinantes de virulência destas espécies (BETSOU *et al.*, 1995). A região N-terminal destas proteínas apresenta atividade de adenilato ciclase, enquanto a região C-terminal apresenta atividade hemolítica (GLASER *et al.*, 1988; BETSOU *et al.*, 1995). O exato mecanismo de ação destas proteínas ainda não foi determinado, mas parece que a região C-terminal seria capaz de interagir com a membrana plasmática de células eucariontes, permitindo a translocação da região N-terminal para o citosol destas células (BETSOU *et al.*, 1995).

As funções das proteínas FprA e FprC de *N. meningitidis* não são conhecidas. Contudo, estas proteínas são produzidas em condições limitantes de ferro (THOMPSON *et al.*, 1993a; THOMPSON *et al.*, 1993b). Assim sendo, estas proteínas podem ser produzidas nestas condições para exercer atividade hemolítica, garantindo o fornecimento de ferro para este organismo durante o processo de patogênese (ROWE e WELCH, 1994).

A região C-terminal da proteína Orf6 de *H. seropedicae* foi comparada com a região C-terminal das proteínas CyaA de *B. pertussis* (GLASER *et al.*, 1988) e de *B. bronchiseptica* (BETSOU *et al.*, 1995), FprC (THOMPSON *et al.*, 1993b) e FprA

(THOMPSON *et al.*, 1993a) de *N. meningitidis* (figura 20). Por meio desta comparação, foi possível identificar cinco motivos característicos das citotoxinas RTX, repetidos em sequência na região C-terminal da proteína Orf6 de *H. seropedicae* (figura 20).

	10	20	30	40	50	
1 Borpe_CyaA	-LLGGD G DDQ L S-----	G D A GRDR L YGEAGDDWFFQ D AAN A GN--	LLDG	1500		
2 Borbr_CyaA	-LLGGD G DDQ L S-----	G D A GRDR L YGEAGDDWFFQ D AAN A GN--	LLDG	1499		
3 Neime_FprC	--FTREG N H L I K A K D G S G Q V T V Q S Y F Q N D G S G A Y R I D E I H F D N G K V L D V		1612			
4 Neime_FprA	L T FTREG N H L I K A K D D S G Q V T V Q S Y F Q N D G S G A Y R I D E I H F D N G K V L D V		898			
5 Herse_Orf6	-----R L R-----	Y D E A R L V K L L K G G H V V T L C A D A A V F R D--	MSGY	34		
6 Consenso	- G L ----- G A L G	D A A F N -- L D				

	60	70	80	90	100	
1 Borpe_CyaA	G D G R D T V D F S G P G R G- L D A G A K G V F L S --	L G K G F A S L M D E P E T S N V L R N I	1541			
2 Borbr_CyaA	G D G N D T V D F S G P G R G- L D A G A K G V F L S --	L G K G F A S L M D E P E T T N N V L R H I	1540			
3 Neime_FprC	A T V K K L V Q Q S T D G S D R L Y A Y Q S G N T L N G G L G D D Y L Y G A D G D D L L N G D A G N		1656			
4 Neime_FprA	A T V K E L V Q Q S T D G S D R L Y A Y Q S G S T L N G G L G D D Y L Y G A D G N D L L N G D A G N		942			
5 Herse_Orf6	L A G K R V G F H A V A V S G V A C S A K T G E V L G F Y I T D S G P G R Y S E M N R F L S V G D L		84			
6 Consenso	G K V S G S G L A G L L G D	D E N				

	110	120	130	140	150	
1 Borpe_CyaA	E N A V G S A R D V L I G D A G A N V L N G I A G N D V L S G G A G D D V L L G D E G S D L L S G		1585			
2 Borbr_CyaA	E N A V G S V R D V L I G D A G A N V L N G I A G N D V L S A A P A D D V L L G D E G S D L L S G		1584			
3 Neime_FprC	D S I Y S G N G N D T L N G G E G N D A L Y G Y N G N D V L N G G E G N D H L N G E D G N D T L I G		1700			
4 Neime_FprA	D S I Y S G N G N D T L D G G E G N D A L Y G Y N G N D A L N G G E G N D H L N G E D G N D T L I G		986			
5 Herse_Orf6	R N M D A H E S A L I Y S-----	V D P A K L L H R D S S A T G N E L D N I L V G N Q D N N L L T G	130			
6 Consenso	N D L G G D L G L G N D V L G G E G D D L G G N D L L G					

	160	170	180	190	200	
1 Borpe_CyaA	D A G N D D L F G G Q G D D T Y L F G V G Y G H D T I Y ---	E S G G G H D T I R I N - A G A D Q L	1626			
2 Borbr_CyaA	D A G N D D L F G G Q G D D T Y L F G A G Y G H D T I Y ---	E S G G G H D T I R I N - A G A D Q L	1625			
3 Neime_FprC	G A G N D Y L E G G S G S D T Y V F G K G F G Q D T V Y N Y -	H V D K N S D T M H F K G F K A A D V	1743			
4 Neime_FprA	G A G N D Y L E G G S G S D T Y V F G E G F G Q D T V Y N Y -	H V D K N S D T M H F K G F K A A D V	1029			
5 Herse_Orf6	A K G N D T L T G Q A G N D T Y V F G K G D G R D L V V D Q A T A G N I D T I K F S D A R Q T N L		180			
6 Consenso	A G N D L G G G D T Y V F G G G D T V Y -	G N D T I F A A L				

	210	220	230	240	250	
1 Borpe_CyaA	W F A R Q G N D L E I R I L G T D A L T V H D W Y R D A D H R V E I I H A A N- Q A V D Q A G I E		1669			
2 Borbr_CyaA	W F A R Q G N D L E I R I L G T D A L T V H D W Y R D A D H R V E A I H A A N- Q A I D P A G I E		1668			
3 Neime_FprC	H F I R S G S D L V L S A S - E Q D N V R I S G F F Y G E N H R V D T F V F D D - A A I S N P D F A		1785			
4 Neime_FprA	H F I R S G S D L V L S A S - E Q D N V R I S G F F Y G E N H R V D T F V F D D - A A I S N P D F A		1071			
5 Herse_Orf6	W F S Q V G K D L Q I D V L G S T D Q V T V K D W Y A G A D N R V E R I K T A D G K T L Y D S D V D		230			
6 Consenso	W F R G D L I L G D V T V D W Y G A D H R V E I A D - A I D					

	260	270	280	290	
1 Borpe_CyaA	K-----LVEAMAQYPD	FGAAAAAPPAA	RVPD	TLMQSLAVNWR	1706
2 Borbr_CyaA	K-----LVEAMAQYPD	FGAAAAAPPAA	RVPD	TLMQSLAVNWR	1705
3 Neime_FprC	KYINAGNNLVQSM	SVFGSNTAAT	TGGNV	DANTQSVQQPLLVT	PSA 1829
4 Neime_FprA	KYINAGNNLVQSM	SVFGSNTAAT	TGGNV	DANIQSVQQPLLVT	PSA 1115
5 Herse_Orf6	K-----LVQAMAS	FDPPAATQ	TSWT	DGQTSKGKVL	LTVTTH-- 265
6 Consenso	K-----LVQAMA	F P AA	DA	LLVT	

Figura 20. Alinhamento da sequência de aminoácidos da região C-terminal da proteína Orf6 de *H.seropedicae* (Herse_Orf6) e das proteínas CyaA de *B. pertussis* (Borpe_CyaA), CyaA de *B. bronchiseptica* (Borbr_CyaA) e FrpC (Neime_FprC) e FprA (Neime_FprA) de *N. meningitidis*. O alinhamento foi feito utilizando o programa Antheptrot 4.9 (DELÉAGE *et al.*, 1988). Em vermelho, são mostrados os resíduos de aminoácidos com 100% de identidade; em azul, acima de 75% de similaridade; em verde, acima de 50%; em preto, abaixo de 50%. Motivos característicos das citotoxinas RTX (LXGGXG^N/_DDX, onde X pode ser qualquer resíduo) (THOMPSON *et al.*, 1993a; THOMPSON *et al.*, 1993b) estão sublinhados com linha contínua.

A semelhança estrutural da região C-terminal da proteína Orf6 de *H. seropedicae* com a região C-terminal de citotoxinas RTX sugere que esta proteína faz parte desta família de citotoxinas. No entanto, dados adicionais são necessários para determinar a função desta proteína.

A princípio, *H. seropedicae* não deveria apresentar genes envolvidos com patogênese em animais no seu genoma, uma vez que é um endófito obrigatório (BALDANI *et al.*, 1997). Porém, o gênero *Herbaspirillum* é filogeneticamente relacionado ao gênero *Pseudomonas*, o qual inclui espécies consideradas patógenos oportunistas em animais (SALYERS e WHITT, 1994). A proximidade filogenética destes gêneros pode ser exemplificada por *Herbaspirillum rubrisubalbicans*, um fitopatógeno de algumas espécies de cana-de-açúcar, anteriormente classificado como *Pseudomonas rubrisubalbicans* (PIMENTEL *et al.*, 1991). Os patógenos oportunistas do gênero *Pseudomonas* secretam proteínas que agem como exotoxinas, caracterizadas como determinantes de virulência (SALYERS e WHITT, 1994). Deste modo, pode-se sugerir que a *orf6* foi herdada de um ancestral comum aos gêneros *Herbaspirillum* e *Pseudomonas*. Assim sendo, este gene pode ter perdido a sua função primitiva no decorrer do processo de especiação de *H. seropedicae* e sua adaptação ao hábito endofítico. Alternativamente, a *orf6* pode ter sido adquirida de alguma espécie filogeneticamente distante por transferência horizontal em condições naturais. Neste caso, o processo de transferência poderia ocorrer no solo ou no interior de uma planta hospedeira (item 3.2.3). O seqüenciamento da região proximal da *orf6* poderá contribuir para o esclarecimento da estrutura do seu produto e sua provável função.

4 CONCLUSÕES

1. A região sequenciada a montante do gene *nifA* de *Herbaspirillum seropedicae* apresenta quatro ORFs, denominadas *orf4*, *modE*, *orf5* e *orf6*.
2. A região distal da *orf4* foi localizada a aproximadamente 1,3 kb do gene *nifA* e não apresenta similaridade significativa com nenhum gene identificado em outras bactérias.
3. O gene *modE* foi encontrado a cerca de 1,8 kb do gene *nifA* e provavelmente está envolvido com a regulação da expressão dos genes *modABC* de acordo com a disponibilidade de íons molibdato.
4. A proteína ModE de *H. seropedicae* apresenta um possível motivo hélice-volta-hélice de ligação ao DNA e dois motivos característicos das proteínas ModE (SARNQ), localizados nos prováveis domínios MopE1 e MopE2, envolvidos com a ligação de íons molibdato.
5. Os genes *mod* estão organizados em *H. seropedicae* de forma diferente da descrita para outras bactérias, ocupando, pelo menos, dois *loci* distintos no genoma.
6. A *orf5* foi identificada a aproximadamente 3,7 kb do gene *nifA* e corresponde ao gene *trs5* de *E. coli*, carregado pela sequência de inserção IS5 desta espécie. A IS5 de *E. coli* se inseriu em uma região intergênica do DNA de *H. seropedicae*. Esta sequência de inserção foi provavelmente transferida do cromossomo de *E. coli* HB101 para o plasmídeo pEMS1 durante a sua permanência nesta célula hospedeira.

7. A *orf6* foi localizada a cerca de 5,3 kb do gene *nifA* e provavelmente codifica para uma proteína da família de citotoxinas RTX, as quais estão relacionadas com processos de patogênese em animais. A *orf6* pode ter sido adquirida de um ancestral comum aos gêneros *Herbaspirillum* e *Pseudomonas* ou de uma bactéria filogeneticamente distante por transferência horizontal.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S.F. MADDEN, T.L.; SCHAFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.**, Oxford, v. 25, p. 3389-3402, 1997.
- ANTEZANA, M.A.; KREITMAN, M. The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. **J. Mol. Evol.**, New York, v. 49, p. 36-43, 1999.
- BALDANI, J.I. BALDANI, V.L.D.; SELDIN, L.; DÖBEREINER, 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, J.I. CARUSO, L.; BALDANI, V.L.D.; GOI, S.R.; DÖBEREINER, J. Recent advances in BNF with non-legume plants. **Soil Biol. Biochem.**, Exeter, v. 29, n. 5/6, p. 911-922, 1997.
- BENELLI, E.M.; SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U.; PEDROSA, F.O. Evidence for two possible *glnB*-type genes in *Herbaspirillum seropedicae*. **J. Bacteriol.**, Washington, v. 179, n. 14, p. 4623-4626, 1997.
- BETSOU, F.; SISMEIRO, O.; DANCHIN, A.; GUIZO, N. Cloning and sequence of the *Bordetella bronchiseptica* adenylate cyclase-hemolysin-encoding gene: comparison with the *Bordetella pertussis* gene. **Gene**, Amsterdam, v. 162, p. 165-166, 1995.

- BULMER, M. Coevolution of codon usage and transfer RNA abundance. **Nature**, London, v. 325, p. 728-730, 1987.
- BOYER, H.W.; ROULLAND-DUSSOIX, D. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. **J. Mol. Biol.**, London, v. 41, p. 459-472, 1969.
- BUSBY, S.; EBRIGHT, R.H. Promoter structure, promoter recognition, and transcription activation in prokaryotes. **Cell**, Cambridge, v. 79, p. 743-746, 1994.
- CHUBATSU, L.S. Tabela de preferência de códons de genes de *Herbaspirillum seropedicae*. Curitiba, 10 jun. 1998. Comunicação pessoal.
- CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Res.**, Oxford, v. 16, p. 10881-10890, 1988.
- DAVISON, J. Genetic exchange between bacteria in the environment. **Plasmid**, Orlando, v.42, p. 73-91, 1999.
- DELÉAGE, G.; CLERC, F.F.; ROUX, B.; GAUTHERON, D.C. Antheprot : a package for protein sequence analysis using a microcomputer. **Cabios**, Oxford, v. 4, p. 351-356, 1988.
- ENGLER, J.A.; BREE, M.P. The nucleotide sequence and protein-coding capability of the transposable element IS5. **Gene**, Amsterdam, v. 14, p. 155-163, 1981.

- FASSLER, J.S.; GUSSIN, G.N. Promoters and basal transcription machinery in eubacteria and eukaryotes: concepts, definitions, and analogies. **Methods Enzymol.**, San Diego, v. 273, p. 3-29, 1996.
- FLEISCHMANN, R.D.; ADAMS, M.D.; WHITE, O.; CLAYTON, R.A.; KIRKNESS, E.F.; KERLAVAGE, A.R.; BULT, C.J.; TOMB, J.-F.; DOUGHERTY, B.A.; MERRICK, J.M.; McKENNEY, K.; SUTTON, G.; FRITZHUGH, W.; FIELDS, C.; GOCAYNE, J.D.; SCOTT, J.; SHIRLEY, R.; LIU, L.-I.; GLODEK, A.; KELLEY, J.M.; WEIDMAN, J.F.; PHILLIPS, C.A.; SPRIGGS, T.; HEDBLUM, E.; COTTON, M.D.; UTTERBACK, T.R.; HANNA, M.C.; NGUYEN, D.T.; SAUDEK, D.M.; BRANDON, R.C.; FINE, L.D.; FRITCHMAN, J.L.; FUHRMANN, J.L.; GEOGHAGEN, N.S.M.; GNEHM, C.L.; McDONALD, L.A.; SMALL, K.V.; FRASER, C.M.; SMITH, H.O.; VENTER, J.C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science**, Washington, v. 269, p. 496-512, 1995.
- GALVÃO, C.W.; PEDROSA, F.O.; SOUZA, E.M.; CHUBATSU, L.S.; RIGO, L.U.; STEFFENS, M.B.R. Evidence for a *recX*-like gene in the nitrogen fixing bacteria *Herbaspirillum seropedicae*. In: XXVII^a REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR (1998: Caxambu). **Resumos...** Caxambu: 1998. p. 46.
- GLASER, P.; LADANT, D.; SEZER, O.; PICHOT, F.; ULLMANN, A.; DANCHIN, A. The calmodulin-sensitive adenylate cyclase of *Bordetella pertussis*: cloning and expression in *Escherichia coli*. **Mol. Microbiol.**, Oxford, v. 2, n. 1, p. 19-30, 1988.
- GRALLA, J.D. Activation and repression of *E. coli* promoters. **Curr. Opin. Genet. Dev.**, London, v. 6, p. 526-530, 1996.

- GRANT, S.G.N.; JESSEE, J.; BLOOM, F.R.; HANAHAN, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-resctriction mutants. **Proc. Natl. Acad. Sci.**, Washington, v. 87, p. 4645-4649, 1990.
- GRUNDEN, A.M.; RAY, R.M.; ROSENTEL, J.K.; HEALY, F.G.; SHANMUGAM, K.T. Repression of the *Escherichia coli modABCD* (molybdate transport) operon by ModE. **J. Bacteriol.**, Washington, v. 178, n. 3, p. 735-744, 1996.
- GRUNDEN, A.M.; SELF, W.T.; VILLAIN, M.; BLALOCK, J.E.; SHANMUGAM, K.T. An analysis of the binding repressor protein ModE to *modABCD* (molybdate transport) operator/promoter DNA of *Escherichia coli*. **J. Biol. Chem.**, Baltimore, v. 274, n. 34, p. 24308-24315, 1999.
- HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symp. Ser.**, Oxford, v. 41, p. 95-98, 1999.
- HELMANN, J.D.; CHAMBERLIN, M.J. DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controled by na alternative sigma factor. **Proc. Natl. Acad. Sci.**, Washington, v. 84, n. 18, p. 6422-6424, 1987.
- HENIKOFF, S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. **Gene**, Amsterdam, v. 28, p. 351-359, 1984.
- HENKIN, T.M. Control of transcription termination in prokaryotes. **Ann. Rev. Genet.**, Palo Alto, v. 30, p. 35-57, 1996.

- HERTZ, G.Z.; STORMO, G.D. *Escherichia coli* promoter sequences: analysis and prediction. **Methods Enzimol.**, San Diego, v. 273, p. 30-42, 1996.
- HILTON, S.M.; SLAUGHTER, C.; EISNER, W.; FISHER, T. The molybdenum-pterin binding protein is encoded by a multigene family in *Clostridium pasteurianum*. **Gene**, Amsterdam, v. 54, p. 211-219, 1987.
- HIPPEL, P.H. An integrated model of transcription complex in elongation, termination, and editing. **Science**, Washington, v. 281, p. 660-665, 1998.
- HOOVER, T.R.; SANTERO, E.; PORTER, S.; KUSTU, S. The integration host factor stimulates interaction of RNA polymerase with NifA, the transcriptional activator for nitrogen fixation operons. **Cell**, Cambridge, v. 63, p. 11-22, 1990.
- JAMES, E.K. OLIVARES, F.L.; BALDANI, J.I.; DÖBEREINER, J. *Herbaspirillum*, an endophytic diazotroph colonizing vascular tissue in leaves of *Sorgum bicolor* L. Moench. **J. Exp. Bot.**, Oxford, v. 48, n. 308, p. 785-797, 1997.
- KLASSEN, G. Seqüenciamento de genes a jusante dos genes *nifHDK* de *Herbaspirillum seropedicae*. Curitiba, 25 set. 1998. Comunicação pessoal.
- KLASSEN, G. Análise genética e funcional dos genes *nifENXorf1orf2*, *nifQmodABCfixXC* de *Herbaspirillum seropedicae*. Curitiba, 2000. Tese (Doutorado em Ciências - Bioquímica) – Setor de Ciências Biológicas, Universidade Federal do Paraná.
- KLASSEN, G.; PEDROSA, F.O.; SOUZA, E.M.; YATES, M.G.; RIGO, L.U. Sequencing and functional analysis of the *nifENXorf1orf2* gene cluster of

- Herbaspirillum seropedicae*. **FEMS Microbiol. Lett.**, Amsterdam, v. 118, p. 165-170, 1999.
- KNAUF, V.C.; NESTER, E.W. Wide host range cloning vectors: a cosmid clone bank of an *Agrobacterium* Ti plasmid. **Plasmid**, Orlando, v. 8, p. 45-54, 1982.
- KNIGHT, A.I.; NI, H.; CARTWRIGHT, K.A.; McFADDEN, J.J. Identification and characterization of a novel insertion sequence, IS1106, downstream of the *porA* gene in B15 *Neisseria meningitidis*. **Mol. Microbiol.**, Oxford, v. 6, n. 11, p. 1565-1573, 1992.
- KUSTU, S.; NORTH, A.K.; WEISS, D.S. Prokaryotic transcriptional enhancers and enhancer-binding proteins. **Trends Biochem. Sci.**, Cambridge, v. 16, p. 397-402, 1991.
- LEWIN, B. **Genes**. 6. ed. Oxford: Oxford University Press, 1997.
- LOBO, A.L.; WELCH, R.A. Identification and assay of RTX family of cytolisins. **Methods Enzymol.**, San Diego, v. 235, p. 667-678, 1994.
- LORENZO, V.; PÉREZ-MARTÍN, J. Regulatory noise in prokaryotic promoters: how bacteria learn to respond to novel environmental signals. **Mol. Microbiol.**, Oxford, v. 19, n. 6, p. 1177-1184, 1996.
- LUQUE, F.; MITCHENALL, L.A.; CHAPMAN, M.; CHRISTINE, R.; PAU, R.N. Characterization of genes involved in molybdenum transport in *Azotobacter vinelandii*. **Mol. Microbiol.**, Oxford, v. 7, n. 3, p. 447-459, 1993.

- MACHADO, I.M.P.; YATES, M.G.; MACHADO, H.B.; SOUZA, E.M.; PEDROSA, F.O. Cloning and sequencing of the nitrogenase structural genes *nifHDK* of *Herbaspirillum seropedicae*. **Braz. J. Med. Biol. Res.**, Ribeirão Preto, v. 29, n. 12, p. 1599-1602, 1996.
- MAHILLON, J.; CHANDLER, M. Insertion sequences. **Microbiol. Mol. Biol. Rev.**, Washington, v. 62, n. 3, p. 725-774, 1998.
- MAUPIN-FURLOW, J.A.; ROSENTEL J.K.; LEE, J.H.; DEPPENMEIER, U.; GUNSALUS, R.P.; SHANMUGAM, K.T. Genetic analysis of the *modABCD* (molybdate transport) operon of *Escherichia coli*. **J. Bacteriol.**, Washington, v. 177, n. 17, p. 4851-4856, 1995.
- McCLURE, W.R. Mechanism and control of transcription initiation in prokaryotes. **Ann. Rev. Biochem.**, Palo Alto, v. 54, p. 171-204, 1985.
- McNICHOLAS, P.M.; MAZZOTTA, M.M.; RECH, S.A.; GUNSALUS, R.P. Functional dissection of the molybdate-responsive transcription regulator, ModE, from *Escherichia coli*. **J. Bacteriol.**, Washington, v. 180, n. 17, p. 4638-4643, 1998.
- MEAD, D.A.; SZCZESNA-SKORUPA, E.; KEMPER, B. Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. **Protein Eng.**, Oxon, v.1, p. 67-74, 1986.
- MERRICK, M.J. ; EDWARDS, R.A. Nitrogen control in bacteria. **Microbiol. Rev.**, Washington, v. 59, n. 4, p. 604-622, 1995.

- MILLER, J.H. **A Short Course in Bacterial Genetics – A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria**. New York: Cold Spring Harbor Laboratory Press, 1992.
- MOUNCEY, N.J.; MITCHENALL, L.A.; PAU, R.N. Mutational analysis of genes of the *mod* locus involved in molybdenum transport, homeostasis, and processing in *Azotobacter vinelandii*. **J. Bacteriol.**, Washington, v. 177, n. 18, p. 5294-5302, 1995.
- NOINDORF, L.; SOUZA, E.M.; STEFFENS, M.B.R.; RIGO, L.U.; PEDROSA, F.O.; CHUBATSU, L.S. Evidence for an ammonium transporter gene (*amtB*) in the nitrogen-fixing bacterium *Herbaspirillum seropedicae*. In: XXVIII^a REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR (1999: Caxambu). **Resumos...** Caxambu, 1999. n. F-79.
- PARKER, J. Errors and alternatives in reading the universal genetic code. **Microbiol. Rev.**, Washington, v. 53, n. 3, p. 273-298, 1989.
- PEREIRA, J.A.R. CAVALCANTE, V.A.; BALDANI, J.I.; DÖBEREINER, J. Field inoculation of sorghum and rice with *Azospirillum* spp. and *Herbaspirillum seropedicae*. In: SKINNER, F.A. *et al.* **Nitrogen fixation with non-legumes**. Dordrecht: Kluwer Academic Publishers, 1989. p. 219-224.
- PEREIRA, J.A.R.; BALDANI, J.I. Selection of *Azospirillum* spp. and *Herbaspirillum seropedicae* strains to inoculate rice and maize plants. In: INTERNATIONAL SYMPOSIUM ON SUSTAINABLE AGRICULTURE FOR TROPICS – THE ROLE OF BIOLOGICAL NITROGEN FIXATION. (1995: Angra dos Reis) **Programme and Abstracts**. Angra dos Reis, 1995, p. 220-221.

- PERSUHN, D.C.; SOUZA, E.M.; STEFFENS, M.B.R.; PEDROSA, F.O.; YATES, M.G.; RIGO, L.U. The transcriptional activator NtrC controls the expression and activity of glutamine synthase in *Herbaspirillum seropedicae*. **FEMS Microbiol. Lett.**, Amsterdam, no prelo.
- PIMENTEL, J.P. OLIVARES, F.; PITARD, R.M.; URQUIAGA, S.; AKIBA, F.; DÖBEREINER, J. Dinitrogen fixation and infection of grass leaves by *Pseudomonas rubrisulalbicans* and *Herbaspirillum seropedicae*. **Plant and soil**, Dordrecht, v. 137, p. 61-65, 1991.
- PLATT, T. Transcription termination and the regulation of gene expression. **Ann. Rev. Biochem.**, Palo Alto, v. 55, p. 339-372, 1986.
- REGO, F.G.M. Sequenciamento do gene *nifB* de *Herbaspirillum seropedicae* e caracterização da sua região promotora. Curitiba, 1997. Dissertação (Mestrado em Ciências – Bioquímica) – Setor de Ciências Biológicas, Universidade Federal do Paraná.
- RIGO, L.U.; STEFFENS, M.B.R.; SOUZA, E.M.; MACHADO, H.B.; FUNAYAMA, S.; PEDROSA, F.O. Mutagenesis and nucleotide sequence of the glutamine synthetase structural gene of *Herbaspirillum seropedicae*. In: XXV^a REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR (1996: Caxambu). Resumos... Caxambu, 1996. p. 49.
- REITZER, L.J.; MAGAZANIK, B. Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. **Proc. Natl. Acad. Sci.**, Washington, v. 82, p. 1979-1983, 1985.

- RICHARDSON, J.P. Rho-dependent transcription termination. **Biochim. Biophys. Acta**, Amsterdam, v. 1048, p. 127-138, 1990.
- ROWE, G.E.; WELCH, R.A. Assays of hemolytic toxins. **Methods Enzymol.**, San Diego, v. 235, p. 657-667, 1994.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular Cloning – A Laboratory Manual**. 2. ed. New York: Cold Spring Harbor Laboratory Press, 1989.
- SAIER, M.H. Differential codon usage: a safeguard against inappropriate expression of specialized genes? **FEBS Lett.**, Amsterdam, v. 362, p. 1-4, 1995.
- SALYERS, A.A.; WHITT, D.D. **Bacterial pathogenesis: a molecular approach**. Washington: American Society for Microbiology, 1994.
- SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci.**, Washington, v. 74, n. 12, p. 5463-5467, 1977.
- SCHONER, B.; KAHN, M. The nucleotide sequence of IS5 from *Escherichia coli*. **Gene**, Amsterdam, v. 14, p. 165-174, 1981.
- SHINGLER, V. Signal sensing by σ^{54} -dependent regulators: derepression as a control mechanism. **Mol. Microbiol.**, Oxford, v. 19, n. 3, p. 409-416, 1996.
- SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U.; PEDROSA, F.O. Cloning and characterization of the *nifA* gene from *Herbaspirillum seropedicae* strain Z78. **Can. J. Microbiol.**, Ottawa, v. 37, n. 6, p. 425-429, 1990.

- SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U.; YATES, M.G.; PEDROSA, F.O. Sequence and structural organization of a *nifA*-like gene and part of a *nifB*-like gene of *Herbaspirillum seropedicae* strain Z78. **J. Gen. Microbiol.**, Reading, v. 137, p. 1511-1522, 1991.
- STADEN, R. An interactive graphic program for comparing and aligning nucleic acid and amino acid sequences. **Nucleic Acids Res.**, Oxford, v. 10, p. 2951-2961, 1982.
- STEFFENS, M.B.R.; RIGO, L.U.; FUNAYAMA, S.; SOUZA, E.M.; MACHADO, H.B.; PEDROSA, F.O. Cloning of a *recA*-like gene from the diazotroph *Herbaspirillum seropedicae* strain Z78. **Can. J. Microbiol.**, Ottawa, v. 39, p. 1096-1102, 1993.
- STEFFENS, M.B.R.; PEDROSA, F.O.; SOUZA, E.M.; MACHADO, H.B.; PERSUHN, D.C.; RIGO, L.U. Sequence analysis of the regulatory genes *ntrB* and *ntrC* of *Herbaspirillum seropedicae*. In: XXVI^a REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR (1997: Caxambu). **Resumos...** Caxambu, 1997. p. 50.
- TAYLOR, W.E.; STRAUS, D.B.; GROSSMAN, A.D.; BURTON, Z.F.; GROSS, C.A. BURGESS, R.R. Transcription from a heat-inducible promoter causes heat shock regulation of the sigma subunit of *E. coli* RNA polymerase. **Cell**, Cambridge, v. 38, p. 371-381, 1984
- TEIXEIRA, K.R.S. Isolamento e caracterização do operon *glnAntrBC* de *Herbaspirillum seropedicae* Z78. Curitiba, 1991. Dissertação (Mestrado em Ciências – Bioquímica) – Setor de Ciências Biológicas, Universidade Federal do Paraná.

- THOMPSON, J.D. GIBSON, T.J.; PLEWNIK, F.; JEANMOUGIN, F. HIGGINS, D.G.
The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res.**, Oxford, v. 25, n. 24, p. 4876-4882, 1997.
- THOMPSON, S.A.; WANG, L.L.; WEST, A.; SPARLING, P.F. *Neisseria meningitidis* produces iron-regulated proteins related to the RTX family of exoproteins. **J. Bacteriol.**, Washington, v.9, n. 1, p. 85-96, 1993a.
- THOMPSON, S.A.; WANG, L.L.; SPARLING, P.F. Cloning and nucleotide sequence of *fprC*, a second gene from *Neisseria meningitidis* encoding a protein similar to RTX cytotoxins. **Mol. Microbiol.**, Oxford, v.9, n. 1, p. 85-96, 1993b.
- TWERDOCHLIB, A. L. Sequenciamento da região distal do gene *ntrC* de *Herbaspirillum seropedicae*. Curitiba, 15 jul. 1999. Comunicação pessoal.
- WANG, G.; ANGERMÜLLER, S.; KLIPP, W. Characterization of *Rhodobacter capsulatus* genes encoding a molybdenum transport system and putative molybdenum-pterin-binding protein. **J. Bacteriol.**, Washington, v. 175, n. 10, p. 3031-3042, 1993.
- WASSEM, R.; SOUZA, E.M.; YATES, M.G.; PEDROSA, F.O.; BUCK, M. Two roles for integration host factor at an enhancer-dependent *nifA* promoter. **Mol. Microbiol.**, Oxford, v. 35, n. 4, p. 756-764, 2000.
- WEBER, O.B.; BALDANI, V.L.D.; TEIXEIRA, K.R.S.; KIRCHHOF, G.; BALDANI, J.I.; DÖBEREINER, J. Isolation and characterization of diazotrophic bacteria from

banana and pineapple plants. **Plant and Soil**, Dordrecht, v. 210, n. 1, p. 103-113, 1999.

WILSON, K.S.; HIPPEL, P.H. Transcription termination at intrinsic terminators: the role of the RNA hairpin. **Proc. Natl. Acad. Sci.**, Washington, v. 92, p. 8793-8797, 1995.

YOUNG, J.P.W. Phylogenetic classification of nitrogen-fixing organisms. In: STANLEY, G.; BURRIS, R.M.; EVANS, H.S. **Biological nitrogen fixation**. New York: Chapman & Hall, 1992. p. 43-86.

6 ANEXOS

ANEXO 1. Mapa de restrição do vetor pVK100.....	83
ANEXO 2. Mapa de restrição do vetor pTZ19R.....	84
ANEXO 3. Mapa de restrição do vetor pSport1.....	85

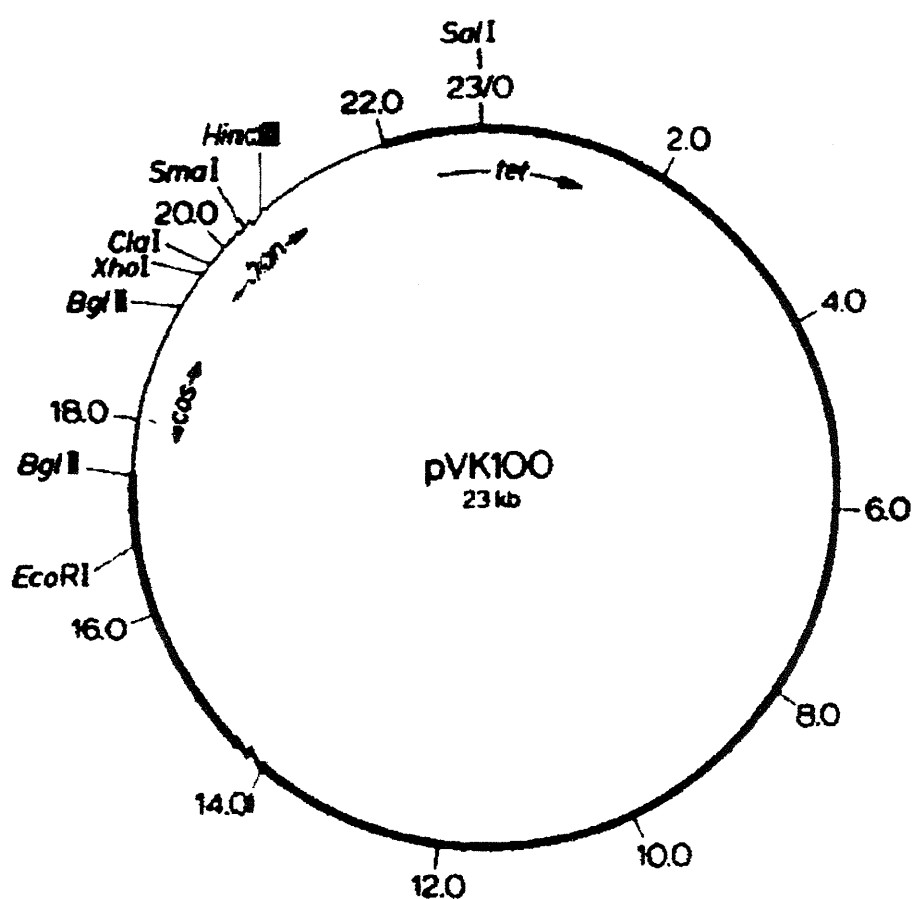


Figura 21. Mapa de restrição do vetor pVK100 (KNAUF e NESTER, 1982). As setas indicam a direção de transcrição dos genes de resistência à canamicina (*kan*) e à tetraciclina (*tet*). São indicados os sítios *Bgl*II, *Cla*I, *Hind*III, *Sal*I, *Sma*I e *Xho*I. O sítio *Eco*RI é ausente no vetor pVK102.

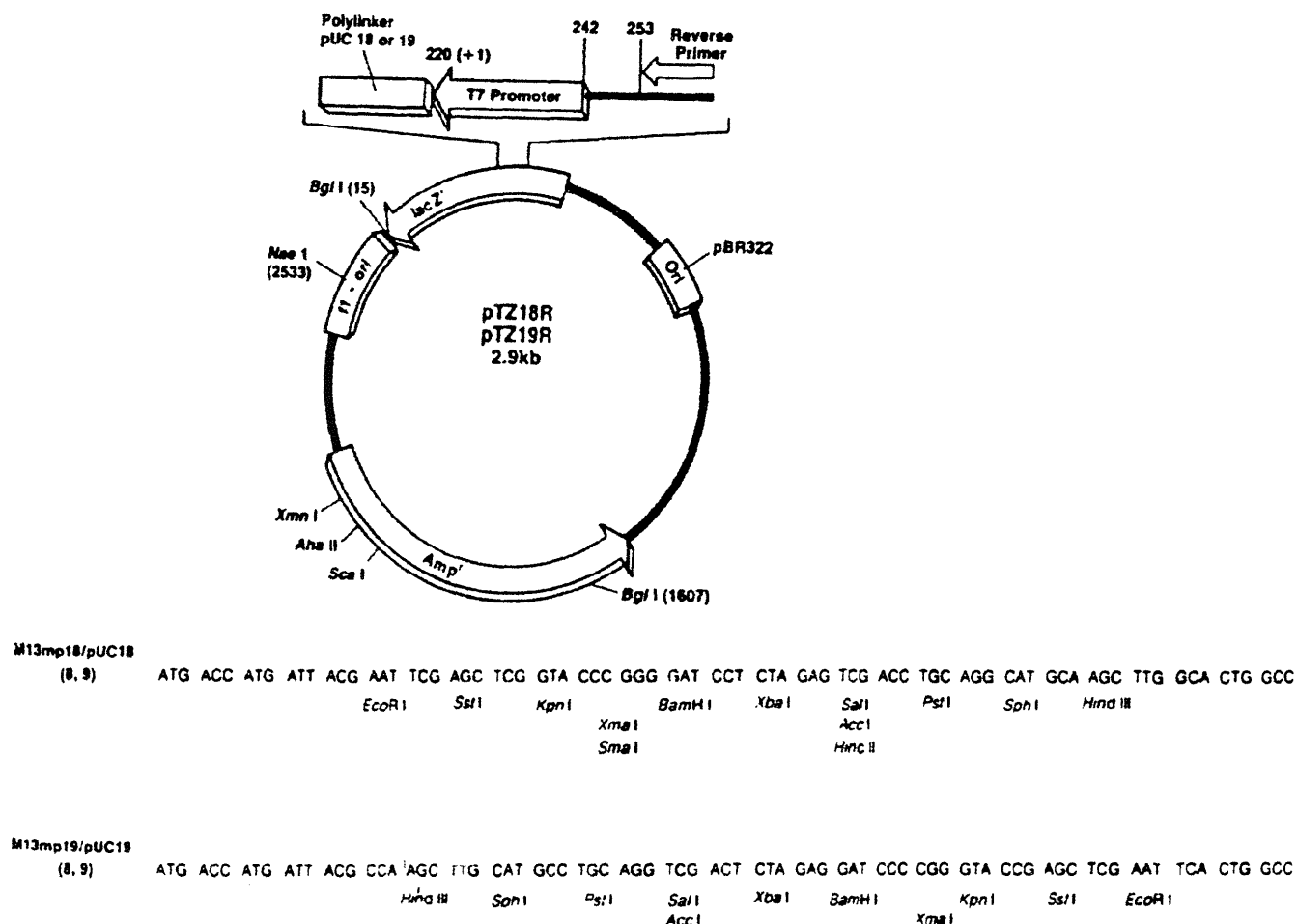


Figura 22. Mapa de restrição do vetor pTZ19R (MEAD *et al.*, 1986). As setas indicam a direção de transcrição dos genes *lacZ* e de resistência à ampicilina (*amp^r*). São indicados os sítios *AccI*, *AhaI*, *BamHI*, *BglI*, *EcoRI*, *HincII*, *HindIII*, *KpnI*, *NaeI*, *PstI*, *SalI*, *ScaI*, *SmaI*, *SphI*, *SstI*, *XbaI*, *XmaI* e *XmnI*. O vetor pTZ19R apresenta sítio de policlonagem idêntico ao do vetor pUC19, indicado na figura.

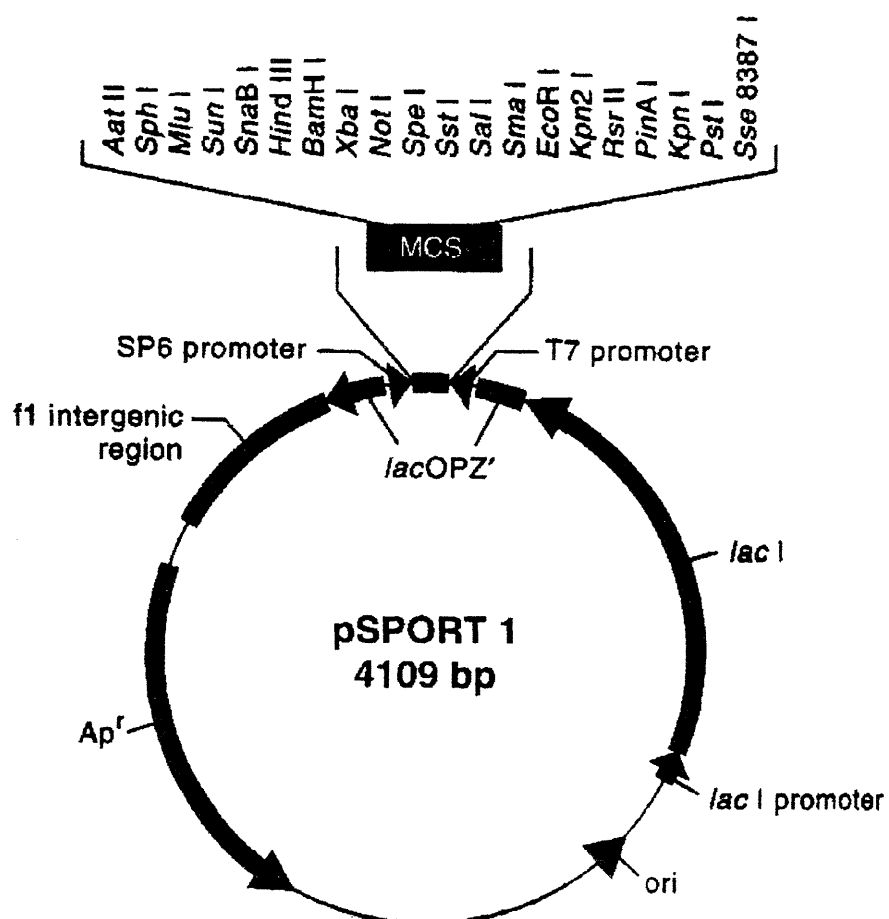


Figura 22. Mapa de restrição do vetor pSport1 (Life Technologies, Gaithersburg, EUA). As setas indicam a direção de transcrição dos genes *lacI*, *lacOPZ'* e de resistência à ampicilina (*ap^r*). São indicados os sítios *AatI*, *BamHI*, *EcoRI*, *HindIII*, *KpnI*, *Kpn2I*, *MluI*, *NotI*, *PinAI*, *PstI*, *RsrI*, *SalI*, *SmaI*, *SnaBI*, *SpeI*, *SphI*, *Sse8387I*, *SstI*, *SunI* e *XbaI*.