

ADRIANA LUCKOW INVITTI

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DO GENE *rpoC* DE  
*Herbaspirillum seropedicae***

Monografia apresentada à disciplina Estágio em Bioquímica – BQ016 como requisito parcial à conclusão do curso de Ciências Biológicas, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientador: Prof. Emanuel Maltempi de Souza

**CURITIBA  
2004**

## AGRADECIMENTOS

Agradeço ao professor Emanuel M. de Souza pela orientação e estímulo para realização deste projeto. Agradeço pelas críticas e sugestões e principalmente pela confiança depositada em mim para a realização deste projeto.

Às professoras Leda S. Chubatsu e Liu Un Rigo pela paciência e compreensão.

À professora Roseli Wassem pelo auxílio e sugestões.

À professora Maria Berenice R. Steffens pelo apoio e incentivo.

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

A todos os professores do Núcleo de Fixação de Nitrogênio.

Aos colegas do laboratório pelo auxílio e momentos de descontração: Lílian, Luiza, Luciano, Gi, Cláudio, Roberta, Renata, Ana C, Marcelo A., Juliana I, Gus, Geraldo, Cynthia, Carol, Michelle D., Andréa, André, Rafa, Rose Adele, Marco, Michelle T., Lys, Helisson, Giovani, Giovana, Humberto, Marcelo M., Daniel, Stefan, Ana Paula, Luis, Thiago e Leo.

A Pati pela ajuda e contribuição.

Aos funcionários e amigos Roseli Prado, D. Jú e Valter, pelo carinho, apoio e atenção.

Em especial, à Fabi, que ajudou muito tanto como co-orientadora quanto como amiga.

À minha família.

Ao Gu pela paciência, auxílio, companheirismo e broncas. Por sempre me incentivar e ficar do meu lado.

## SUMÁRIO

RESUMO .....	v
1 INTRODUÇÃO .....	1
1.1 <i>Herbaspirillum seropedicae</i> .....	1
1.2 RNA polimerase DNA-dependente .....	1
1.2.1 Estrutura .....	2
1.2.2 Funções e Genética.....	4
1.2.3 Transcrição em procariotos .....	7
2 OBJETIVOS .....	13
3 MATERIAIS E MÉTODOS .....	14
3.1 Bactérias e Plasmídios.....	14
3.2 Meios de Cultura e Condições de Cultivo .....	14
3.3 Antibióticos.....	16
3.4 Preparo de Solução de RNase .....	16
3.5 Purificação de DNA.....	17
3.5.1 DNA genômico de <i>H. seropedicae</i> .....	17
3.5.2 DNA plasmidial .....	17
3.6 Construção de Oligonucleotídeos iniciadores.....	18
3.7 Amplificação do gene <i>rpoC</i> .....	22
3.7.1 Gradiente de temperatura de anelamento .....	22
3.7.2 Amplificação do fragmento correspondente ao gene <i>rpoC</i> .....	22
3.8 Eletroforese de DNA em gel de ágar .....	23
3.9 Clivagem do DNA com enzimas de restrição.....	23
3.10 Reação de ligação de DNA.....	23
3.11 Preparo de Células Eletrocompetentes .....	24
3.12 Transformação Bacteriana por Eletroporação .....	24
3.13 Ferramentas de Bioinformática.....	24
4 RESULTADOS E DISCUSSÃO .....	26
4.1 Análise dos operons que possuem os genes que codificam para as subunidades da RNAPolimerase de <i>H. seropedicae</i> .....	26
4.1.1 Operon <i>rplKAJLrpoBC</i> .....	26
4.1.2 Operon <i>rpsMKDrpoArpIQ</i> .....	39
4.2 Comparação da seqüência de aminoácidos da subunidade $\beta'$ de <i>H. seropedicae</i> .....	44

4.3	Identificação dos domínios da subunidade $\beta'$ da RNA polimerase de <i>H. seropedicae</i> .....	48
4.4	Amplificação do gene <i>rpoC</i> .....	49
4.4.1	Determinação da temperatura de anelamento ideal para amplificação do fragmento de interesse.....	49
4.4.2	Amplificação do gene <i>rpoC</i> de <i>H. seropedicae</i> para clonagem ..	50
4.5	Estratégias de clonagem do gene <i>rpoC</i> de <i>H. seropedicae</i> .....	51
4.6	Estratégia de expressão do gene <i>rpoC</i> em <i>H. seropedicae</i> .....	52
5	CONCLUSÕES .....	54
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	55



## RESUMO

A regulação da expressão gênica ocorre principalmente a nível transcricional. O principal alvo deste controle, em bactérias, é a RNA polimerase DNA-dependente em resposta a fatores intrínsecos e extrínsecos. A regulação da transcrição está mais bem caracterizada em *E. coli*. A RNA polimerase de *E. coli* é uma enzima oligomérica que possui duas subunidades  $\alpha$ , uma  $\beta$ , uma  $\beta'$  e uma  $\omega$ . A subunidade  $\beta'$  é responsável pela ligação ao DNA molde e pela interação com os fatores  $\sigma$ . A região C-terminal da subunidade  $\beta'$  não está envolvida com a atividade catalítica ou com a montagem da holoenzima. Poucos estudos sobre a RNA polimerase de outros organismos foram feitos. Neste trabalho, o gene *rpoC* de *Herbaspirillum seropedicae*, que codifica para a subunidade  $\beta'$  da RNA polimerase, foi identificada no Banco de Dados do Programa GENOPAR. Este gene foi amplificado por PCR usando oligonucleotídeos iniciadores complementares a seqüência das regiões que codificam para a extremidade N-terminal e C-terminal da subunidade  $\beta'$  da RNA polimerase de *H. seropedicae*. As condições ótimas de amplificação foram determinadas. O fragmento obtido possui 4242pb e está sendo clonado no vetor pET29a. O clone resultante deverá expressar a subunidade  $\beta'$  ligada a uma cauda de histidina.

## 1 INTRODUÇÃO

### 1.1 *Herbaspirillum seropedicae*

*Herbaspirillum seropedicae* são microrganismos diazotrofos endofíticos, pertencentes à subclasse  $\beta$  das Proteobactérias (Young, 1992), isolados dos órgãos vegetativos de gramíneas como cana de açúcar e arroz (Baldani *et al.*, 1986, 1995). São bactérias Gram-negativas, vibrióides, algumas vezes helicoidais e muito móveis, capazes de fixar nitrogênio em condições de microaerofilia (Baldani *et al.*, 1986).

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

Tendo em vista o potencial biofertilizante de *H. seropedicae* torna-se importante o estudo dos mecanismos celulares destes microrganismos. O estudo da RNA polimerase desse organismo, bem como dos mecanismos de transcrição podem contribuir para esclarecer a regulação da expressão gênica neste organismo, em especial da regulação do metabolismo e mecanismo de invasão e colonização de plantas.

### 1.2 RNA polimerase DNA-dependente

As RNA polimerases DNA-dependentes foram descobertas em 1959 (Von Hippel *et al.*, 1984) e catalisa a síntese de RNA utilizando DNA como

molde, processo denominado transcrição, tanto em procariotos como em eucariotos (Yura e Ishihama, 1979). As semelhanças estruturais existentes entre as RNA polimerases de procariotos com as de organismos superiores indica que a sua função e mecanismo são bastante conservados (Darst, 2001).

### 1.2.1 Estrutura

As RNA polimerases presentes em procariotos podem ser divididas em duas classes: enzimas oligoméricas, que possuem múltiplas subunidades; e enzimas monoméricas, que são induzidas em células bacterianas hospedeiras por alguns tipos de bacteriófagos. As RNA polimerases oligoméricas são constituídas por pelo menos 5 subunidades e podem ser isoladas em duas formas: a holoenzima que possui as subunidades  $\alpha_2\beta\beta'\sigma$ , capaz de transcrever moldes de DNA nativo; e o núcleo da enzima que possui todas as subunidades exceto a subunidade (ou fator)  $\sigma$  que não é capaz de iniciar a síntese de RNA a partir de moldes nativos (Kumar, 1981). A subunidade  $\sigma$ , ou fator sigma, é responsável pelo reconhecimento e ligação a seqüências específicas do DNA, denominadas promotores (Wösten, 1998). As RNA polimerases de alguns organismos possuem ainda a subunidade  $\omega$  ligada às subunidades  $\beta'$ , embora a ausência desta subunidade não afeta a função da RNA polimerase (Kumar, 1981; Ghosh *et al*, 2003). A figura 1 mostra o modelo estrutural da holoenzima de *Thermus aquaticus*.

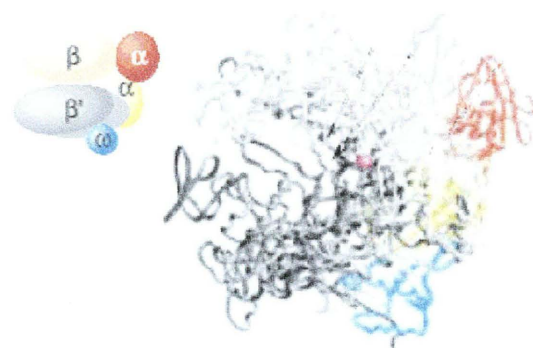


Figura 1. RNA polimerase de *T. aquaticus*. As cores de cada subunidade no esquema à esquerda correspondem às cores do modelo estrutural. Modificado de Cramer, 2002.

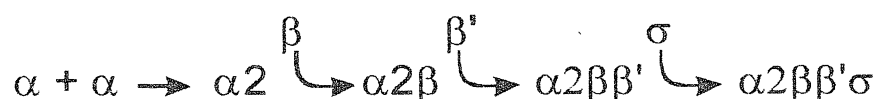
O núcleo da RNA polimerase de *T. aquaticus* possui a forma de uma garra de caranguejo com 150 Å de comprimento, 115 Å altura e 110 Å de

espessura e com um canal central de 27 Å, (Murakami & Darst, 2003). O canal possui um sítio para íons  $Mg^{2+}$  que corresponde ao sítio catalítico da enzima. Uma das partes da garra corresponde a subunidade  $\beta'$  e a outra a subunidade  $\beta$ . As subunidades  $\alpha$  encontram-se fora do canal, no outro lado da garra e mantém as subunidades  $\beta$  e  $\beta'$  unidas (Burukov & Severinov, 2002).

A holoenzima possui dois átomos de  $Zn^{2+}$  ligados por molécula de enzima, esses átomos são quelados por quatro resíduos de cisteína da subunidade  $\beta'$ ; a dissociação das subunidades faz com que esses átomos se desliguem (Kumar, 1981; Von Hippel *et al*, 1984; Burukhov & Nudler, 2003).

A enzima isolada de *E. coli* é a RNA polimerase DNA-dependente melhor estudada e possui uma subunidade  $\beta'$  (155.613 Da), uma  $\beta$  (150.618 Da), uma  $\sigma$  (70.263 Da) e duas  $\alpha$  (30.512 Da) (Mc Clure, 1985). A massa molecular total do núcleo da RNA polimerase de *E. coli* é de aproximadamente 400kDa (Darst, 2001) e as subunidades  $\beta$  e  $\beta'$  contribuem com 60% da massa total da enzima (Burukov & Severinov, 2002). Além das subunidades  $\alpha$ ,  $\beta$  e  $\beta'$  que compõe o núcleo; a RNA polimerase de *E. coli* pode se associar com um dos múltiplos tipos de fator  $\sigma$  existentes (Ishihama, 1990) ou ainda possuir uma pequena subunidade adicional denominada  $\omega$  (10 KDa) (Gosh *et al*, 2003). Os fatores  $\sigma$  de bactérias podem ser divididos em duas famílias: família  $\sigma^{70}$  e família  $\sigma^{54}$ . Proteínas destas duas famílias possuem pouca ou nenhuma identidade de sequência entre si. *E. coli* possui 6 tipos de fatores sigma pertencentes à família  $\sigma^{70}$ :  $\sigma^{70}$ ,  $\sigma^{38}$ ,  $\sigma^{32}$ ,  $\sigma^{28}$ ,  $\sigma^E$  e  $\sigma^{FecI}$  e apenas um fator tipo  $\sigma^{54}$  (Wösten, 1998).

A montagem do núcleo acontece de acordo com a seguinte sequência de reações (Yura & Ishihama, 1979; Kumar, 1981; Ishihama, 1988):



O primeiro complexo  $\alpha_2\beta\beta'$  formado corresponde ao núcleo prematuro (Yura & Ishihama, 1979), que *in vitro* depende da incorporação do fator sigma e incubação em altas temperaturas para que ocorra ativação (Ishihama, 1990). As interações que ocorrem entre as subunidades  $\beta$  e  $\beta'$  do núcleo com o fator  $\sigma$

provocam as mudanças conformacionais necessárias para a formação da holoenzima ativa (Burukhov & Nudler, 2003). A subunidade  $\omega$  está envolvida no processo de rearranjo conformacional do núcleo em holoenzima (Gosh *et al*, 2003).

Deleções na porção C-terminal da subunidade  $\beta$  interferem na montagem do núcleo, indicando que esta região está envolvida na interação entre as subunidades, além disso somente o complexo  $\alpha_2\beta$  é capaz de interagir com  $\beta'$  demonstrando que a interação entre  $\beta$  e as duas subunidades  $\alpha$  é necessária para a ligação de  $\beta'$  (Ishihama, 1990). A porção N-terminal das subunidades  $\alpha$  interage com as outras subunidades (Ishihama, 1990), já a subunidade  $\omega$  interage com  $\beta'$  pela sua região C-terminal (Gosh *et al*, 2003).

### 1.2.2 Funções e Genética

A RNA polimerase é responsável pela síntese de todos os tipos de RNA *in vivo*: mRNA, rRNA e tRNA (Mc Clure, 1985). Dworkin & Losick (2002) demonstraram que além dessas funções a RNA polimerase de *E. coli* possui um papel importante na segregação dos cromossomos bacterianos durante a replicação, principalmente durante a separação inicial da região das origens de replicação.

Cada uma das subunidades da RNA polimerase possui uma função específica e modifica o processo de transcrição. As funções características de cada subunidade e dos fatores associados a RNA polimerase de *E. coli* estão descritos na tabela 1.

**Tabela 1. Subunidades e Fatores Associados à RNA polimerase de *E. coli*.**

Subunidade ou fator associado	Gene	Nº resíduos de aminoácidos	Massa molecular (Da)	Função e características relevantes
$\alpha$	<i>rpoA</i>	329	36.512	Associação das subunidades $\beta$ e $\beta'$ .
$\beta$	<i>rpoB</i>	1342	150.618	Catálise da síntese de RNA (iniciação e alongamento), reconhecimento de terminador, ligação do substrato ribonucleosídeo trifosfato, ligação do RNA produzido, controle da estringência, regulação autógena da síntese de $\beta$ e $\beta'$ , ligação a rifamicina e estreptovaricina.
$\beta'$	<i>rpoC</i>	1407	156.163	Ligação ao DNA molde e a subunidade $\sigma$ . Ligação de poliânions como heparina.
$\omega$	<i>rpoS</i>	91	10.000	Associa-se a subunidade $\beta'$ e atua na renaturação do núcleo da RNA polimerase. Mutantes <i>rpoS</i> apresentam redução modesta na taxa de crescimento.
$\sigma^{70}$	<i>rpoD</i>	613	70.263	Maturação do núcleo e reconhecimento de promotores de expressão regular. Indispensável para a iniciação da transcrição.
$\sigma^{20}$	<i>rpoE</i>	191	20.000	Reconhecimento de promotores de genes ligados ao choque térmico (heat-shock) extremo.
$\sigma^{28}$	<i>rpoF</i>	239	28.000	Reconhecimento de promotores de genes flagelares e de quimiotaxia.

$\sigma^{54}$	<i>rpoN</i>	477	54.000	Reconhecimento de promotores de genes ligados ao metabolismo de nitrogênio.
$\sigma^{32}$	<i>rpoH</i>	284	32.381	Reconhecimento de promotores de genes ligados ao choque térmico (heat-shock).

**Tabela 1** - Referências: Kumar, 1981; Ishihama, 1988; Ishihama, 1990, Wösten, 1998 e Gosh *et al*, 2003

Os genes que codificam para o núcleo da RNA polimerase são *rpoA* ( $\alpha$ ), *rpoB* ( $\beta$ ) e *rpoC* ( $\beta'$ ). O gene *rpoB* foi isolado de mutantes que não possuíam sensibilidade a rifamicinas e estreptovaricinas (Yura e Ishihama, 1979). Os genes *rpoB* e *rpoC* estão localizados em um mesmo operon, enquanto *rpoA* encontra-se em outro operon; esses três genes são co-transcritos com genes que codificam para proteínas ribossomais (Yura e Ishihama, 1979).

O gene *rpoA* faz parte do operon *rpsMrpsKrpsDrpoArplQ* onde *rpsM*, *rpsK*, *rpsD* e *rplQ*, codificam para as proteínas ribossomais S13, S11, S4 e L17, respectivamente. O gene *rpoA* parece estar sob o mesmo controle que os genes que codificam para as proteínas ribossomais adjacentes (Ishihama, 1990). Por outro lado os genes *rpoBC* estão no operon *rplKrplArplJrplLrpoBrpoC*, onde *rplK*, *rplA*, *rplJ* e *rplL* codificam para as proteínas ribossomais L11, L1, L10 e L12, respectivamente (Ishihama, 1990). A expressão de *rplK* e *rplA* é regulada pela concentração de proteína L1 livre na célula enquanto a expressão de *rplJ* e *rplL* é controlada pela quantidade do complexo L10-L7/12 livre na célula (Ralling & Linn, 1984). Existe um promotor bastante forte a montante a *rplJ* e 70% dos transcritos deste promotor param em *rplL* devido a presença de um atenuador entre *rplL* e *rpoB* (Ralling & Linn, 1984).

Embora os genes que codificam para as subunidades  $\alpha$ ,  $\beta$  e  $\beta'$  estejam em operons diferentes sua síntese é estruturalmente coordenada, resultando em concentrações celulares com estequiometria de 2:1:1 (Ishihama, 1990). A transcrição dos genes *rpoBC* sofre uma regulação autógena pela concentração celular de  $\beta$  e  $\beta'$ , complexo  $\alpha_2\beta$  e holoenzima (Yura & Ishihama, 1979; Ishihama, 1990). Existem três formas de controle da expressão dos genes *rpoBC*, a frequência de iniciação nos promotores e de terminação de

transcrição no atenuador, a clivagem do RNAm pela RNaseIII e a frequência de tradução do RNAm (Passador & Linn, 1989). O aumento da concentração celular das subunidades  $\beta$  e  $\beta'$  estimula a transcrição de *rpoA* (Ishihama, 1990). Em *E. coli* o excesso de subunidade  $\beta$  na célula faz com que a síntese de  $\beta'$  seja reduzida em 60% e o excesso de  $\beta'$  provoca uma redução de 40% na síntese de  $\beta$ . Essa regulação ocorre principalmente a nível traducional, sendo a redução da transcrição menos afetada. A superexpressão das subunidades  $\beta$  e  $\beta'$  provocam um atraso no ciclo celular bacteriano, aumentando o tempo necessário para a duplicação. Níveis elevados de  $\beta$  e  $\beta'$  reprimem a tradução do transcrito *rpoBC*, enquanto níveis elevados de holoenzima reprimem a transcrição de *rpoBC* (Dykxhoorn, Pierre & Linn, 1996). A terminação da transcrição no atenuador a montante a *rpoB* não é dependente das proteínas Nus e ocorre pela interação direta de terminador/anti-terminador no RNAm (Steward, Pierre & Linn, 1997). Uma região 29pb a montante do início de tradução de *rpoB* contém um sítio de regulação pós-transcricional, onde as subunidades  $\beta$ ,  $\beta'$  e a holoenzima ligam impedindo a tradução do RNAm (Passador & Linn, 1992).

O gene *rpoD*, que codifica para o fator sigma 70, encontra-se em outra região do cromossomo (Yura & Ishihama, 1979), formando o operon *rpsUdnaGrpoD*; *rpsU* codifica para a proteína ribossomal S21 e *dnaG* codifica para a primase (Ishihama, 1990).

### 1.2.3 Transcrição em procariotos

A síntese do RNA é catalisada pelo núcleo da RNA polimerase, mas o reconhecimento de promotores e início da transcrição dependem da presença da subunidade  $\sigma$ . Somente a holoenzima é capaz de provocar as mudanças conformacionais na região do promotor necessárias para separar as duas fitas de DNA e iniciar a transcrição (Ishihama, 1990). O mecanismo básico de catálise da RNA polimerase é a transferência de um ribonucleosídeo monofosfato para a ponta OH 3' de RNA nascente utilizando como substrato ribonucleosídeos trifosfato (Von Hippel, 1984).

O processo de transcrição pode ser dividido em quatro etapas (Kumar, 1981; Von Hippel, 1984): reconhecimento do DNA molde, iniciação,



alongamento e terminação (Figura 2).

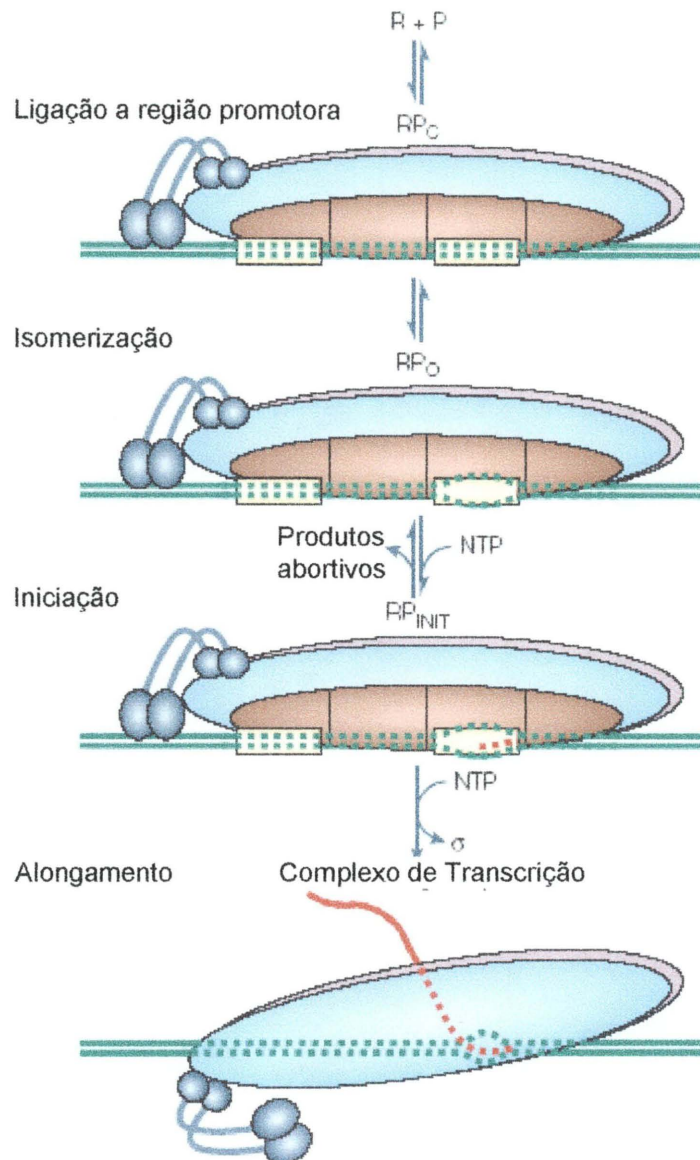


Figura 2. Três primeiras fases da transcrição em procariotos. A RNA polimerase (R) interage com a região promotora no DNA (P) formando o complexo fechado ( $RP_C$ ). As linhas tracejadas verdes indicam as regiões promotoras do DNA onde se liga a holoenzima. A dupla fita de DNA está desenrolada na região de transcrição (Bolha de transcrição) formando o complexo aberto ( $RP_O$ ). O complexo de iniciação é formado ( $RP_{INIT}$ ) e a síntese de RNA (representado por linhas tracejadas vermelhas) é iniciada com formação da ligação fosfodiéster entre o nucleotídeo inicial e um NTP. O alongamento é a última fase demonstrada, onde o tamanho da cadeia de RNA aumenta (demonstrado com uma linha vermelha contínua). Modificado de Browning & Busby, 2004.

O reconhecimento do molde inicia-se com a ligação da holoenzima ao DNA dupla fita na região do promotor, dando origem ao complexo binário fechado. Então as fitas de DNA são separadas para permitir o pareamento de bases da fita molde com os ribonucleotídeos. A bolha de transcrição é formada pela desnaturação da fita de DNA que se inicia na região onde a RNA polimerase está ligada, formando o complexo binário aberto (Von Hippel, 1984 e Lewin, 2000).

A iniciação corresponde a ligação dos primeiros nucleotídeos do RNA. A enzima permanece ligada ao promotor durante a incorporação dos primeiros nucleotídeos. A formação da primeira ligação fosfodiéster entre os dois primeiros ribonucleotídeos origina o complexo ternário (Lewin, 2000). A iniciação é prolongada pela ocorrência de inícios abortivos, onde a enzima produz transcritos curtos que são liberados e então ela inicia novamente a síntese do RNA, desde a primeira base (Hsu *et al*, 2003). Durante o ciclo de iniciação abortiva a fita de DNA molde se acumula no sítio catalítico enquanto o DNA dupla fita da região promotora continua ligado a RNA polimerase (Cheetham, Jeruzalmi & Steitz, 1999). O acúmulo de DNA molde no sítio catalítico da enzima provoca um aumento da bolha de transcrição. Essa distorção nos complexos de transcrição iniciais provoca o retorno da RNA polimerase e a liberação da fita de RNA nascente, caracterizando os inícios abortivos. Entretanto, quando uma quantidade suficiente de DNA molde entra no sítio ativo da enzima, o movimento das fitas pode forçar o re-enrolamento do DNA molde a montante, permitindo que a enzima se desloque para a região a jusante liberando a região promotora e iniciando a transcrição produtiva (Vo *et al*, 2003b).

Existem duas grandes classes de complexos de iniciação, os produtivos, que após algumas iniciações abortivas passam a transcrição completa do DNA molde e os improdutivos, onde a RNA polimerase está presa a região promotora e promove a transcrição de apenas dois ou três nucleotídeos (Vo *et al*, 2003a). Uma terceira classe é constituída pelos chamados complexos de iniciação moribundos, geralmente inativados na forma de complexo binário. Estes complexos podem ser reativados, ou seja, tornarem-se transcricionalmente produtivos, na presença dos fatores GreA e GreB. Entretanto, os complexos de iniciação improdutivos não se tornam produtivos

na presença dos fatores Gre. A proporção de formação de complexos produtivos e improdutivos está relacionada a concentração de NTPs. Se a concentração de NTPs for alta a formação de complexos improdutivos é reduzida (Vo *et al*, 2003a).

Os inícios abortivos são ferramentas essenciais para a iniciação da cadeia de RNA e provavelmente são decorrentes de bloqueios no estágio de iniciação de transcrição (Hsu *et al*, 2003). Um dos domínios do fator sigma, que penetra no sítio catalítico do núcleo, é provavelmente o responsável pelos inícios abortivos. Este domínio choca-se inevitavelmente com a cadeia de RNA nascente e pode causar a expulsão de transcritos curtos do complexo. O ciclo de iniciações abortivas é um preço que as RNA polimerases DNA-dependentes pagam por serem capazes de iniciar uma síntese dependente de molde oligonucleotídeo iniciador (Burukov & Severinov, 2002).

A força de um promotor é determinada pela sua similaridade a seqüência consenso, quanto maior a similaridade mais estável será o complexo aberto formado, entretanto uma grande similaridade confere uma ligação muito forte entre a RNA polimerase e a fita de DNA molde dificultando o movimento da enzima e saída da região promotora (Vo *et al*, 2003 a,b). O tamanho dos transcritos abortivos, a posição em que o transcrito será liberado e a posição em que o fator sigma é liberado são determinados por cada tipo de promotor. Existem promotores que possuem taxas de iniciação produtiva extremamente baixas, realizando apenas uma síntese produtiva para cada 165 inícios de transcrição (Hsu *et al*, 2003).

A fase seguinte corresponde ao alongamento. Durante esta fase o núcleo da RNA polimerase move-se pela fita de DNA e estende a cadeia de RNA nascente. A fita de DNA a jusante é desenrolada enquanto a enzima se move para expor um novo segmento molde. Os nucleotídeos são covalentemente ligados à extremidade 3' do RNA nascente e formam um híbrido RNA-DNA na região desenrolada. A fita de DNA retoma a forma de dupla hélice na região a montante à enzima. O RNA fica livre na forma de simples fita (Lewin, 2000). A estabilização do complexo ternário é dada principalmente pelas interações proteína-DNA e proteína-RNA (Nudler *et al*, 1996). Pareamentos de base incorretos na fita híbrida de DNA-RNA provocam desestabilização do complexo de alongamento e fazem com que a RNA

polimerase retroceda, clive a fita de RNA que está sendo transcrito e recomeça a transcrição (Nudler *et al* 1997).

Existem dois modelos para a estrutura do híbrido DNA-RNA na bolha de transcrição. O modelo clássico propõe um heteroduplex de cerca de 12pb que tem papel fundamental na manutenção da estabilidade da bolha de transcrição. O modelo alternativo propõe que o heteroduplex estenda-se apenas por 2 ou 3pb e sua função principal é formar a ponta 3'-OH. Entretanto, estudos recentes fornecem um modelo onde o híbrido RNA-DNA possui de 8 a 9pb e a sua função é a de manter a fidelidade da sequência transcrita (Nudler *et al* 1997).

A terminação é a última etapa e envolve o reconhecimento do ponto onde deve haver a parada da adição de bases à cadeia de RNA nascente. O fim da transcrição só ocorre se a formação de novas ligações fosfodiéster cessar e o complexo de transcrição for separado (Lewin, 2000). Uma sequência denominada terminadora é necessária para que a bolha de transcrição se desmonte. Esta sequência desestabiliza o híbrido de RNA-DNA permitindo que o DNA retorne a forma de dupla fita e a cadeia de RNA e a RNA polimerase sejam liberadas (Platt & Richardson, 1992).

Existem dois mecanismos de terminação conhecidos em *E. coli*: a liberação do RNA transcrito e da RNA polimerase pela formação de grampos no RNA nascente provocando a interrupção transitória da RNA polimerase, e pela ação de um fator protéico conhecido como Rho. A proteína Rho atua ligando-se ao RNA nascente, em regiões ricas em resíduos de citosina, seguido de deslocamento pelo RNA acoplado à hidrólise de ATP em direção a RNA polimerase. Quando a proteína Rho alcança a bolha de transcrição ocorre o desenrolamento do heteroduplex de RNA-DNA, a liberação da fita de RNA e a dissociação do complexo de transcrição (Richardson, 1990 e Platt & Richardson, 1992).

A RNA polimerase possui uma plasticidade muito grande, entretanto não realiza movimentos mede-palmo durante a transcrição. Nas primeiras etapas da transcrição ocorre uma mudança conformacional do sítio ativo da RNA polimerase para acomodar a cadeia de RNA nascente (8 a 9 nucleotídeos), essa conformação é mantida durante todo o processo de alongação, inclusive quando existe retrocesso. A enzima retorna a conformação inicial quando a

transcrição termina e o complexo de alongamento se desfaz (Nudler *et al* 1997).

A RNA polimerase pode ainda interagir com diversos fatores durante o processo de transcrição. Esta interação pode ocorrer em qualquer uma das fases de transcrição podendo provocar aumento ou diminuição da produção de transcrito. Proteínas acessórias como NusA, NusG, GreA e GreB podem atuar em diferentes momentos no processo de alongamento (Erie, 2002). NusA parece estar primariamente envolvida na terminação Rho-dependente; ela diminui a taxa de alongamento durante a transcrição aumentando as pausas da RNA polimerase e inibe ou estimula a terminação Rho-dependente (Carlomagno & Nappo, 2003). NusG atua como inibidora de pausa e as proteínas GreA e GreB estimulam a clivagem do RNA nascente (Erie, 2002). Cramer (2002) sugere que a RNA polimerase seja capaz de detectar a incorporação de um nucleotídeo incorreto ao RNA nascente e retornar corrigindo o erro quando associada às proteínas acessórias GreA e GreB.

A transcrição pode ser inibida por diversos mecanismos, tais como: repressores, inibição da ligação da RNA polimerase ao promotor, inibição da desnaturação do promotor e inibição do desligamento do promotor. Todos estes mecanismos são bastante eficientes e dependem do tipo de regulação do promotor. Os repressores de transcrição não possuem somente a capacidade de ligar DNA, mas também de modificar a estrutura do DNA e interagir com a RNA polimerase (Rojo, 2001)

## 2 OBJETIVOS

- Identificar o gene *rpoC* de *H. seropedicae* no banco de dados de seqüências;
- Identificar módulos regulatórios do gene *rpoC* de *H. seropedicae*;
- Amplificar o gene *rpoC* (codifica para a subunidade  $\beta'$  da RNA polimerase) de *H. seropedicae* por PCR utilizando oligonucleotídeos específicos.

### 3 MATERIAIS E MÉTODOS

#### 3.1 Bactérias e Plasmídios

A estirpes de bactérias e plasmídios utilizados neste trabalho estão listados na tabela 2.

Tabela 2 – Estirpes e Plasmídios

Estirpe ou Plasmídio	Genótipo ou Características relevantes	Referências
<i>Herbaspirillum seropedicae</i> - SmR1	derivada da estirpe Z78, Sm <sup>R</sup> , Nif <sup>+</sup>	Souza <i>et. al.</i> , 1995
<i>Escherichia coli</i> – TOP10	F <sup>-</sup> <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80/ <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara-leu</i> ) 7697 <i>galU galK</i> <i>rpsL</i> (Str <sup>R</sup> ) <i>endA1 nupG</i>	Invitrogen
pET-29a	vetor de expressão, gera proteína de fusão com segmento His-Tag, promotor T7, Km <sup>R</sup>	Novagen

#### 3.2 Meios de Cultura e Condições de Cultivo

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

A composição do meio de cultura Nfb é a seguinte:

	Concentração (g/L)
MgSO <sub>4</sub> .7H <sub>2</sub> O	2 x 10 <sup>-1</sup>
NaCl	1 x 10 <sup>-1</sup>
CaCl <sub>2</sub>	2 x 10 <sup>-2</sup>
Ácido nitrilo-triacético	5,6 x 10 <sup>-2</sup>
FeSO <sub>4</sub> . 7H <sub>2</sub> O	2 x 10 <sup>-2</sup>
Malato de sódio	5
Biotina	1 x 10 <sup>-4</sup>

Na <sub>2</sub> MoO <sub>4</sub> .2 H <sub>2</sub> O	2 x 10 <sup>-3</sup>
MnSO <sub>4</sub> . H <sub>2</sub> O	2,35 x 10 <sup>-3</sup>
H <sub>3</sub> BO <sub>3</sub>	2,8 x 10 <sup>-3</sup>
CuSO <sub>4</sub> .5 H <sub>2</sub> O	8 x 10 <sup>-5</sup>
ZnSO <sub>4</sub> .7 H <sub>2</sub> O	2,4 x 10 <sup>-4</sup>

O meio Nfb sólido foi obtido com a adição de 15 g/L de ágar a mistura de meio líquido. O meio completo é denominado NfbHPN (Nfb com adição de NH<sub>4</sub>Cl 20 mmol/L e 50 mmol/L de solução de Fosfatos - 159,4 g/L de KH<sub>2</sub>PO<sub>4</sub> e 17,8 g/L de K<sub>2</sub>HPO<sub>4</sub>).

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

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

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

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

A composição do meio de cultura Terrific Broth é:

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



A composição do meio de cultura SOB é:

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

A composição do meio de cultura SOC é:

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

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

### 3.3 Antibióticos

As soluções de antibióticos foram preparadas em água milli-Q estéril e esterilizadas por filtração em filtro Millipore HAWP 0,45 µm e estocadas a -20°C. As concentrações de antibióticos no meio de cultura foram: estreptomicina (Sm) 80 µg/mL e canamicina (Km) 50 µg/mL.

### 3.4 Preparo de Solução de RNase

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

### 3.5 Purificação de DNA

#### 3.5.1 DNA genômico de *H. seropedicae*

A purificação de DNA genômico para utilização nas reações de amplificação de DNA (PCR) foi realizada segundo um protocolo modificado de Souza (1990). Uma cultura de células (1500  $\mu\text{L}$ ) foi centrifugada em microcentrífuga Eppendorf minispin a 13000 rpm por 1 minuto. O sobrenadante foi descartado e o precipitado de células foi ressuspenso com a pipeta em 500  $\mu\text{L}$  de tampão GET (glucose 50 mmol/L, Tris-HCl 25 mmol/L pH 8,0 e EDTA 10 mmol/L pH 8,0). A essa mistura adicionou-se 50  $\mu\text{L}$  de SDS 10% e a mistura foi incubada à 60°C por 20 minutos para que houvesse lise celular. Ao lisado celular foram adicionados 200  $\mu\text{g/mL}$  de Pronase E (solução estoque 20 mg/mL pré-tratada com incubação à 37°C por 1 hora) seguido de incubação a 37°C por 24 horas. A mistura foi tratada com 100  $\mu\text{L}$  de fenol:clorofórmio:álcool isoamílico (25:24:1), homogeneizada por inversão, cerca de 100 vezes, e posteriormente centrifugado a 13000 rpm por 15 minutos. A fase aquosa foi transferida para outro tubo e o DNA precipitado com 2,5 volumes de etanol absoluto e recuperado por centrifugação por 20 minutos a 13000 rpm. O "pellet" de DNA foi lavado duas vezes com etanol 80%, seco à vácuo e dissolvido em 200  $\mu\text{L}$  de água milli-Q autoclavada. A solução de DNA foi conservada em geladeira.

Após a purificação, a concentração e pureza do DNA foram determinadas. O DNA purificado foi diluído (50 vezes) em água milli-Q autoclavada. A relação  $D.O._{260} / D.O._{280}$  indicativa do grau de pureza do DNA, apresentou o valor de 1,93. Valores entre 1,8 e 2,0 indicam DNA com alto grau de pureza. A absorvância a 260 nm permite a determinação da concentração de DNA através da relação: absorvância 1 corresponde a 50  $\mu\text{g/mL}$  de DNA dupla fita.

#### 3.5.2 DNA plasmidial

A purificação de plasmídeos foi realizada pela técnica de lise alcalina conforme descrito em SAMBROOK *et al.*(1989). A extração rápida de plasmídeos para seleção de recombinantes foi realizada em placa pelo método modificado de Monteiro *et al* (2001). Culturas de células (200  $\mu\text{L}$ ) foram

colocadas em placas de fundo em V e centrifugadas a 4000 rpm por 7 minutos. O sobrenadante foi descartado e o “pellet” de células foi resuspenso em 30 µL de solução RL (glicerol 10%, RNase 5 µg/mL e azul de bromofenol 0,05%) com ajuda do agitador vórtex. Adicionou-se então 10 µL de SDS 10% e manteve-se a placa em água fervente por 1 minuto. Foram adicionados a mistura 5 µL de KCl 3 mol/L, para precipitação das proteínas, e a placa foi incubada em gelo por 10 minutos. A placa foi então submetida a uma nova centrifugação (4000rpm/7minutos) e 15 µL da fração sobrenadante foram submetidos a eletroforese em gel de ágar 1%, utilizando-se o plasmídeo pET29a íntegro para comparação.

### 3.6 Construção de Oligonucleotídeos iniciadores

Foram planejados 2 oligonucleotídeos iniciadores a partir da seqüência do gene *rpoC* de *Herbaspirillum seropedicae* depositada no banco de dados do Programa de Sequenciamento Genômico GENOPAR. O oligonucleotídeo *rpoCN* é complementar a seqüência de nucleotídeos que codifica para a região amino terminal da proteína RpoC e possui a inserção do sítio de restrição para a enzima *NdeI* e o oligonucleotídeo *rpoCC* é complementar a seqüência de nucleotídeos que codifica para a região carboxi-terminal da proteína RpoC, possuindo o sítio de restrição para a enzima *HindIII*. Esses iniciadores encontram-se listados abaixo, os sítios de restrição estão destacados em vermelho.

*rpoCN* – 5' CCTGGAGTGACATATGAAAGCA 3'

*rpoCC* – 5' GCGAAAGCTTCGCTTCGCCTTC 3'

A posição dos iniciadores e as alterações de seqüência inseridas estão mostradas na figura 3.

Figura 3. Sequência do gene *rpoC* de *H. seropedicae*

```

1      TCGAGCGGCACCCAGCAAGCAGAGGCTTCGGCTCTCCCGCCGATGCGAACAGGGTTAAGCAGGGTGAACCGGCCACAC      80
                                     M K A L L D L F K Q V Q Q N
81      GTAAATTTAGTTTTTTCACGCAACCTGGAGTGATACATGAAAGCACTGCTCGATCTATTCAAGCAAGTCCAGCAAAACG      160
      rpoCN - CCTGGAGTGACATATGAAAGCA→

      E Q F D A I K I G L A S P D K I R S W S F G E V K K E
161     AACAAATTCGACGCGATCAAGATCGGTCTGGCTTCGCCTGACAAAGATCCGCTCGTGGTCCTTCGGCGAAGTCAAGAAGCCG      240
  
```

E T I N Y R T F F K P E R D G L F C A K I F G P I K D Y  
 241 GAAACCATCAACTACCGTACCTTCAAGCCCGAGCGGACGGCCTGTCTGCGCCCAAGATCTTCGGCCCATCAAGGACTA 320

E C L C G K Y K R L K H R G V I C E K C G V E V T L  
 321 TGAGTGCCTGTGCGGCAAATACAAGCGCCTGAAGCACCGTGGCGTGATCTGCGAGAAGTGGCGCGTGGAAAGTGACACTGG 400

A K V R R E R M G H I E L A S E T A H I W F L K S L F  
 401 CCAAGGTGCGTCTGTGAGCGCATGGGCCACATCGAGCTGGCGTCCCCGACTGCCACATCTGGTTCCCTGAAAGTGGCTGCGG 480

S R L G M V L D M T L R D I E R V L Y F E A Y V V T D  
 451 TCCCGTCTGGGCATGGTGTGTGGACATGACCCCTGCGCGACATCGAGCGCGTGTCTGTACTTCGAAGCCTACCTGCTGACCGA 560

P G M T P L K K C Q I M S E D D Y A A K Y E E Y G D  
 561 TCCTGGCATGACCCCGCTGAAGAAGTGCCAGATCATGTCCGAAGACGACTACGCGGCCAAGTACGAAGAGTACGGGCGCG 640

D F I A F M G A E G I R E L L R S I D I D R E A E T L  
 641 ACTTCATCGCCTTCATGGGCGCCGAAGGTATCCGCGAAGTGTGCGCTCGATCGATATCGACCGCGAAGCGGAAACCCCTG 720

R Q E L K D S K S E A K I K K Y A K R L K V L E A F Q  
 721 CGCCAGSAACTGAAGGACTCCAAGTCCGAAGCCCAAGATCAAGAAGTACGCCAAGCGCCTGAAGGTGCTGGAGGCGTTCCA 800

R S G I K P E W M I M E V L F V L P P E L R P L V P  
 801 GCGCTGGGCATCAAGCCCGAGTGGATGATCATGGAAGTGTGCGCGTGTGCGCGGGAAGTGGCTGCGCTGGTGGCCGC 880

L D G G R F A T S D L N D L Y R R V I N R N N R L K R  
 881 TGGATGGCGCCCGCTTCGCGACCTCCGACCTGAACGACCTGTACCGCGCGTGCATCAACCGTAACAACCGTCTGAAGCGC 960

L M E L R A P E I I T R N E K R M L Q E A V D S L L D  
 961 CTGATGGAAGTGGCGCGCCTGAAATCATCACGCGCAACGAAAAGCGCATGTTGCGAGGAAGCGGTGGACTCGCTGCTGGA 1040

N G R R G K A M T G A N K R P L K S L A E M I K G K  
 1041 CAACGGTGTGCGCGCAAGGCCATGACCGCGCCCAACAAGCGTCCGCTGAAGTGTGCTGGCCGAAATGATCAAGGGCAAGG 1120

G G R F R Q N L L G K R V D Y S G R S V I V V G P Q L  
 1121 GCGGTGCTTCCGTCAGAACCTGCTGGGCAAGCGCGTGGACTACTCGGGCCGTTCCGGTCATCGTGGTGGGCCCCGAGCTC 1200

K L H Q C G L P K L M A L E L F K P F I F N K L E L M  
 1201 AAGTGCACCAAGTGGCGCTGCCAAGCTGATGGCCCTGGAAGTGTTCAGCCCTTCATCTTCAACAAGCTGGAAGTATGAT 1280

G L S T T I K A A K K L V E A Q E P V V W D I L E E  
 1281 GGGTCTGTGACCAACCATCAAGCGCGCAAGAAGCTGGTGGAAAGCGCAAGAGCCGCTGGTCTGGGACATCCTGGAAGAAG 1360

V I R E H P V M L N R A P T L H R L G I Q A F E P V L  
 1361 TCATCCCGGAACACCCGCTCATGTGAACCGCGCGCGACCCCTGCACCGCTCTGGGCATCCAGGCGTTCGAGCCGCTCCCTG 1440

I E G K A I Q L H P L V C A A F N A D E D G D Q M A V  
 1441 ATCGAAGGCAAGGCGATCCAGCTGCACCCGCTGGTGTGCGCGCGTTCACCGCGACTTCGACGGGAGCCAGATGGCCGT 1520

H V P L S I E A Q M E A R T L M L A S N N I L F P S  
 1521 GCACGTGCGCTGTGATCGAAGCGCAGATGGAAGCCCGCACCCCTGATGCTGGCCTCCAACAACATCCTGTTCCCGTCCA 1600

N G E P S I V P S Q D I V L G L Y Y A T R E A I N G K  
 1601 ACGGCGAACCSTCGATCGTGCCTGCCAGGATATCGTGGGTCTGTACTATGCCACCCGTGAAGCCATCAACGGCAAG 1680

N E G M M F Q D V S E V I R A Y D N K E V E L A T R I  
 1681 AACGAAGGCATGATGTTCCAGGACGTGTGGGAAGTCATCCGGCCTACGACAACAAGGAAGTCGAGCTGGCCACCCGTAT 1760

T V R I V E H P K D P V T G E F V R T V K R Y E T T  
 1761 CACGGTSCGTATCGTGGAACACCCGAAGGACCCGGTCAACGGCGAATTCGTGGGTACCGTCAAGCGCTACGAGACCACCA 1840

I G R A I L S E I L S K G L P F S V L N R A L K K K E  
 1841 TCGGCCGCGCATCCTCTCCAGATCCTGCCAAGGGCCTGCCGTTCTCGGTGCTGAACCGCGCGTTGAAGAAGAAGGAA 1920

I S K L I N T S F R K C G L R A T V V F A D Q L M Q S  
 1921 ATCTCGAAGTCGATCAACACGTCCTTCCGCAAGTGGCGTCTGCGTGCCACCGTGTGTTGCCGACCAGCTGATGCAGTC 2000

G F R L A T R A G I S I C V D D M L V P P Q K A T L  
 2001 GGGCTTCCGCTGGCGACCCGCGCCGGTATCTCCATCTGCGTGGACGACATGCTTGTGCCGCGCAAAAGGCCACCCTGA 2080

I A A A E S E V K Q I E Q Q Y S S G L V T A G E R Y N  
 2081 TCGCCGCTGCCGAGTCCGAAGTCAAGCAGATCGAACAGCAGTACTCCTCGGGTCTGGTGACCGCAGGCGAGCGCTACAAC 2160

K V V D I W G K A G D E V G K A M M D Q L K V E D V I  
 2161 AAGSTGTTGGACATCTGGGGCAAGGCCGGCGACGAAGTCCGCAAGGCCATGATGGACCAGCTGAAGGTGGAAGACGTCAT 2240

T R E G K K T T Q E S F N A I Y M M A D S G A R G S  
 2241 CACCCGCGAAGGCAAGAAGACCACCCAGGAATCGTTCAACGCGATTTACATGATGGCCGACTCCGGCGCCCGTGGTTCCG 2320

A A Q I R Q L A G M R G L M A K P D G S I I E T P I T  
 2321 CTGCCCAGATCCGCCAGCTGGCCGSTATGCCGCGCCTGATGGCCAAGCCTGACGGCTCCATCATCGAGACSCCCATTACC 2400

A N F R E G L N V L Q Y F I S T H G A R K G L A D T A  
 2401 GCGAACTTCCGCGAAGGCCTGAACGTGTTGCAGTACTTCATCTCGACGCACGGCCCGTAAGGGTCTGGCCGACACCCG 2480

L K T A N S G Y L T R R L V D V T Q D L V V I E D D  
 2481 GCTGAAGACCCCAACTCCGGTTACCTGACCCGTCGTCTGGTGGACGTGACCCAGGATCTGGTCTGATCGAAGACGATT 2560

C G T S N G S A M K A L V E G G E V I E A L R D R I L  
 2561 GCGGCACCTCCAACGGTTCGGCCATGAAGGCGCTGGTCCGAAGGCGGTGAAGTTATCGAAGCCCTGCGCGACCCGTATCCTC 2640

G R V A A T D I V N P E D Q S T L F E A G T L L D E D  
 2641 GGCCTGTGGCGCGACCCSATATCGTCAACCCGGAAGATCAATCCACCCTGTTGGAAGCCGGCACCCCTGCTGSAAGAGA 2720

K V E E I E R M G I D E V K V R T P L T C D T R Y G  
 2721 CAAGTCCGAAGAGATCGAACGCATGGGTATCGACGAAGTGAAGGTCCGTACTCCGCTGACCTGCGACACCCGCTATGGCC 2800

L C A K C Y G R D L G R G L L V N S G E A V G V V A A  
 2801 TGTSCGCCAAGTCTATGGCCGCGACCTGGGCGTGGCTGCTGGTGAACCTCCGGCGAAGCCGTCGGCGTGGTGGCAGCG 2880

Q S I G E P G T Q L T M R T F H I G G A A S R A A V A  
 2881 CAGTCCATCGGTGAACCCGGGTACCCAGCTGACCATGCGTACCTTCCACATCGGTGGTGGCGCATCGCGTGCAGCGGTGGC 2960



S S V E A K S N G T V R F T A T M R Y V T N G K G E  
 2961 CTCTCGGTGGAAGCCAAGTCCAACGGCACCGTGCCTTCAACGGCCACGATGCGTTACGTGACCAATGGCAAGGGCGAGC 3040

Q I V I S R S G E V L I T D D L G R E R E R H K V F Y  
 3041 AGATCGTCATTTCCCGTTCCGGCGAAGTGCTGATCACCGATGACCTGGGCGCTGAGCGGAGCGTCACAAAGTGCCGTAC 3120

G A T L I V K D G L T I K A G T A L A T W D P L T R P  
 3121 GGGCCACCGTATCGTCAAGGATGGCCTGACCATCAAGGCGCGTACCGCCCTGGGACCTGGGATCCGCTGACCCGTCC 3200

I I T E Y S G T V K F E N V E E G S T V A K Q I D E  
 3201 GATCATCACCGAGTACTCCGGTACCGTGAAGTTCGAAAACGTGGAAGAAGGCTCCACCGTGGCCAAGCAGATCGACGAAG 3280

V T G L S T L V V I D A K R R G F S S K V M R P Q V K  
 3281 TCACCGGTCTGTGCAACCGTGGTGGTCAATCGATGCGAAGCGTCTGGTCCCGTCTCCAAAGGTCAATGCGTCCGCGAGGTGAAA 3360

L L N E Q G E E V K I A G T E H S V T I G F Q V G A L  
 3361 CTGCTGAACGAACAAGGCGAAGAAGTGAAGATCGCCGGCACCGAACACTCGGTGACCATCGGCTTCCAGGTCCGGCGCGCT 3440

I T V K D G Q Q V H V G E V L A R I P T E S Q K T R  
 3441 GATCACCGTCAAGGACGGCCAGCAAGTGCATGTGGSTGAAGTGTGGCCCGTATCCCGACCGAATCGCAGAAGACCCGTG 3520

D I T G G L P R V A E L F E A R S P K D A G M L A E V  
 3521 ACATTACCGGTGGTCTGCGCGCGCTGGCCGAGCTGTTCGAAGCCCGTTCGCCCAAGGATGCGCGCATGCTGGCCGAAGTC 3600

T G T V A F G K E T K G K Q R L E I T D M E G T K H E  
 3601 ACTGGTACCGTTGCCCTTCGGCAAGGAACCAAGGGCAAGCAGCGTCTGGAATCACCGACATGGAAGGCACCAAGCACGA 3680

F L I T K D K Q V L V H D G Q V V N K G E M I V D G  
 3681 GTTCTGATCACCAAGGACAAGCAAGTGTCTGGTGCACGACGGCCAGGTGGTGAACAAGGGCGAAATGATCGTGGACGGTC 3760

P A D P Q D I L R L L G I E A L A R Y I V D E V Q D V  
 3761 CGGCTGACCCGACAGGACATCCTGCGCTGTTGGGTATCGAAGCGCTGGCACGTTACATCGTCCGACGAAGTGCAGGACGTG 3840

Y R L Q G V K I N D K H I E V I V R Q M L R R V Q V V  
 3841 TACCGTCTGCAGGGCGTGAAGATCAACGACAAGCACATCGAAGTGTGCTGCGCCAGATGCTGCGCCGCGTGCRAAGTGGT 3920

E P G D T N Y I T G E Q V E R S E L L D E N D R V T  
 3921 GGAGCCGGGGACACCAACTACATCACCGGTGAGCAGGTCGAGCGTTCGAACTGCTGGACGAGAACGACCCGCTGACCG 4000

A E G K I P A T Y E N V L L G I T K A S L S T D S F I  
 4001 CCGAAGGCAAGATCCCGCGACCTACGAAAACGTGCTGCTGGTATTACCAAGGCATCGCTGTGACCCGACTCCTTCATC 4080

S A A S F Q E T T R V L T E A A I M G K K D G L R G L  
 4081 TCGGCGCATCGTTCCAGGAAACCACCGCGTGTGACCGAAGCCGCGATCATGGGCAAGAAGGACGGCCCTGCGCGGCCT 4160

K E N V I V G R L I F A G T G L A F H R A R K E K D  
 4161 GAAGGAPAAAGTCAATCGTCCGCGCCTGATCCCGGCTGGTACTGGCCTGGCCTTCCACCGCGCCCGCAAGGAAAAGGATG 4240

A W E A E E R A A L L E A E R A S R S E F G D E E I S  
 4241 CATGGGAAGCCGAAGAGCGCGCAGCACTGCTCGAAGCCGAACGTGCTTCCCGCTCCGAATTCGGCGACGAGGAAATCTCC 4320

S T E S I G G G E G E A \*

4321 TCGACCGAATCCATCGGTGGCGGGCGAAGGCGAAGCGTAAGCCTCGCATCGCCCGGGTGGTCCGCTTGTATCGGCCAACC 4400

← CTTCCGCTTCGCTTCGAAAGCG - *rpoCC*

4401 ACACGGCAGAAAGCTAACAGCTGCAAGAGAAAGCGGCGTCGGAAGTCATTCCGG 4452

**Legenda:** A primeira linha corresponde a seqüência de aminoácidos codificada pela seqüência de nucleotídeos indicada na segunda linha. Estão indicados os iniciadores *rpoCN* e *rpoCC*. Em azul estão indicadas as bases que foram modificadas para que fossem inseridos os sítios de restrição. O asterisco (\*) representa o códon de parada de tradução.

### 3.7 Amplificação do gene *rpoC*

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

#### 3.7.1 Gradiente de temperatura de anelamento

O sistema de amplificação possuía 63 ng/μL de DNA genômico de *Herbaspirillum seropedicae*, tampão de PCR (Tris-HCl 20 mmol/L pH 8,4 e KCl 50 mmol/L), 1,5 mmol/L de MgCl<sub>2</sub>, 0,2 mmol/L de dNTPs, 0,5 pmol/ μL de cada oligonucleotídeo iniciador e 2,5 U de *Taq* DNA polimerase em um volume final de 10 μL. Doze tubos de reação foram submetidos a um gradiente de temperatura de anelamento de 38°C à 58,7°C. Os parâmetros utilizados para amplificação foram: um ciclo de 95°C por 2 minutos, trinta ciclos de 95°C por 15 segundos, 38°C-58°C por 30 segundos e 72°C por 6 minutos e um ciclo de 72°C por 10 minutos.

#### 3.7.2 Amplificação do fragmento correspondente ao gene *rpoC*

O sistema de amplificação possuía 63 ng/ μL de DNA genômico de *Herbaspirillum seropedicae*, tampão de PCR (Tris-HCl 20 mmol/L pH 8,4 e KCl 50 mmol/L), 1,5 mmol/L de MgCl<sub>2</sub>, 0,2 mmol/L de dNTPs, 0,4 pmol/ μL de cada oligonucleotídeo iniciador, 4 U de *Taq* DNA polimerase e 0,5 U de *Pfx* DNA polimerase em um volume final de 25 μL. A enzima *Pfx* DNA polimerase foi adicionada ao sistema para minimizar a taxa de erros na cópia do DNA molde durante a amplificação (Cheng *et al*, 1994). Os parâmetros utilizados para amplificação foram: um ciclo de 95°C por 2 minutos, trinta e cinco ciclos

de 95°C por 15 segundos, 58,7°C por 30 segundos e 72°C por 6 minutos e um ciclo de 72°C por 10 minutos.

### 3.8 Eletroforese de DNA em gel de ágar

As eletroforeses foram realizadas em gel de ágar conforme descrito por SAMBROOK *et al.* (1989) e a visualização do DNA foi realizada sob luz UV (302 nm) após tratamento com brometo de etídeo (0,5 µg/mL).

### 3.9 Clivagem do DNA com enzimas de restrição

O DNA plasmidial, vetor pET29a, e o produto de PCR foram clivados em tampão M (Tris-HCl 10 mmol/L, MgCl<sub>2</sub> 10 mmol/L, NaCl 50 mmol/L e Ditioneritritol – DTE 1 mmol/L; pH 7,5) com 10 U de cada enzima (*Nde*I e *Hind*III), o sistema de restrição do DNA plasmidial possuía ainda 2,5 µg/mL de RNase. A incubação do sistema de clivagem do vetor foi de 4 horas. O produto de PCR foi tratado com fenol:clorofórmio:álcool isoamílico (25:24:1), precipitado com etanol absoluto e ressuspensão em água milli-Q autoclavada, antes de sofrer clivagem, para que houvesse completa inativação da DNA polimerase. Ao produto de PCR foi adicionado 5 U de cada enzima a cada 12 horas incubado por 36 horas. Esse tempo prolongado de incubação é o recomendado para que as enzimas clivassem completamente o produto de PCR.

### 3.10 Reação de ligação de DNA

Vetor e produto de PCR foram tratados com fenol:clorofórmio:álcool isoamílico (25:24:1), precipitados e ressuspensos em água milli-Q autoclavada, antes de realizar-se o sistema de ligação. O sistema de ligação foi montado com uma proporção molar de vetor:inserto de 1:5. Foram utilizados tampão de T4 DNA ligase (50 mmol/L Tris-HCl pH 7,6; 10 mmol/L MgCl<sub>2</sub>; 1 mmol/L DTT; 0,5 mmol/L ATP e 5% PEG 8000) e 0,5 U de T4 DNA ligase em um sistema de volume final 20 µL. A reação foi incubada a 37°C por 1 hora e precipitada com 2,5 volumes de etanol absoluto, o DNA foi seco à vácuo e ressuspensão em 5 µL de água milli-Q autoclavada.



### 3.11 Preparo de Células Eletrocompetentes

Uma cultura com 40mL de *E. coli* crescida por 3 a 5 horas (DO 0,8) foi centrifugada a 5000 rpm por 5 minutos, em centrifuga gelada (4°C). O sobrenadante foi descartado e o sedimento ressuspense em 40 mL de água milli-Q estéril gelada. Essa suspensão foi centrifugada novamente a 5000 rpm por 5 minutos (4°C), e o sedimento ressuspense, repetindo-se o procedimento anterior. Após descartar o sobrenadante as células foram lavadas em solução de glicerol 15% estéril e gelado. Finalmente, após centrifugação de 5000 rpm por 5 minutos, as células foram ressuspensas em 300 µL de glicerol 15% estéril e gelado, aliquotadas (20 µL) e conservadas à -70°C.

### 3.12 Transformação Bacteriana por Eletroporação

Alíquotas de células eletrocompetentes de *E. coli* foram descongeladas em banho de gelo e a elas adicionados de 0,2-1 µL de solução de DNA ou reação de ligação. Essa mistura foi homogeneizada, transferida para câmaras de eletroporação e submetida a um pulso elétrico de 4 kV, 200 Ω, 330 µF. A seguir, as células foram incubadas com 1 mL de meio SOC, sem antibióticos, a 37°C - 130 rpm por 40 minutos para permitir sua recuperação. Alíquotas de 50-250 µL de células foram plaqueadas em meio LA com antibióticos adequados à seleção das células recombinantes e as placas incubadas de 16-24 horas à 37°C.

### 3.13 Ferramentas de BioInformática

O gene *rpoC* foi localizado da sequência genômica de *H. seropedicae* no Banco de Dados do programa Genopar (<http://www.genopar.org>); utilizando o programa GAAT (Genomic Assembly and Analysis Tool) (Favetti *et al*, 2003).

Os alinhamentos das seqüências de aminoácidos das proteínas foram realizados com a utilização do programa ClustalW (Higgins, *et al* 1994). Este programa realiza alinhamentos entre seqüências de proteínas, identificando similaridades.

Os operons *rpsMKDrpoArpIQ* e *rpIKAJrpoBC* foram analisados utilizando o programa FramePlot 2.3.2 (Ishikawa & Hotta, 1999), que localiza as

prováveis regiões codificadoras de proteína de acordo com o conteúdo GC da região. As seqüências selecionadas no gráfico gerado pelo FramePlot foram traduzidas e comparadas com o banco de dados de seqüências de proteínas GenBank utilizando o aplicativo BlastP (<http://www.ncbi.nlm.nih.gov/blast>). As mudanças de fase e erros de sequenciamento foram corrigidos pela análise dos eletroferogramas visualizados no programa Consed (Gordon, Abajian & Green, 1998).

Os domínios estruturais da subunidade  $\beta'$  da RNA polimerase de *H. seropedicae* foram identificados pelo programa SMART (Simple Modular Architecture Research Tool) (Schultz *et al*, 1998 e Letunic *et al*, 2002) e Pfam (Protein families database) (<http://www.sanger.ac.uk>). O ponto isoelétrico e a massa molecular teórica da proteína nativa e com cauda de resíduos de histidina foram calculados utilizando-se o aplicativo Compute pI/MW Tool (<http://us.expasy.org>) (Bjellqvist *et al*, 1993; Bjellqvist *et al*, 1994 e Wilkins *et al*, 1998).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise dos operons que possuem os genes que codificam para as subunidades da RNAPolimerase de *H. seropedicae*

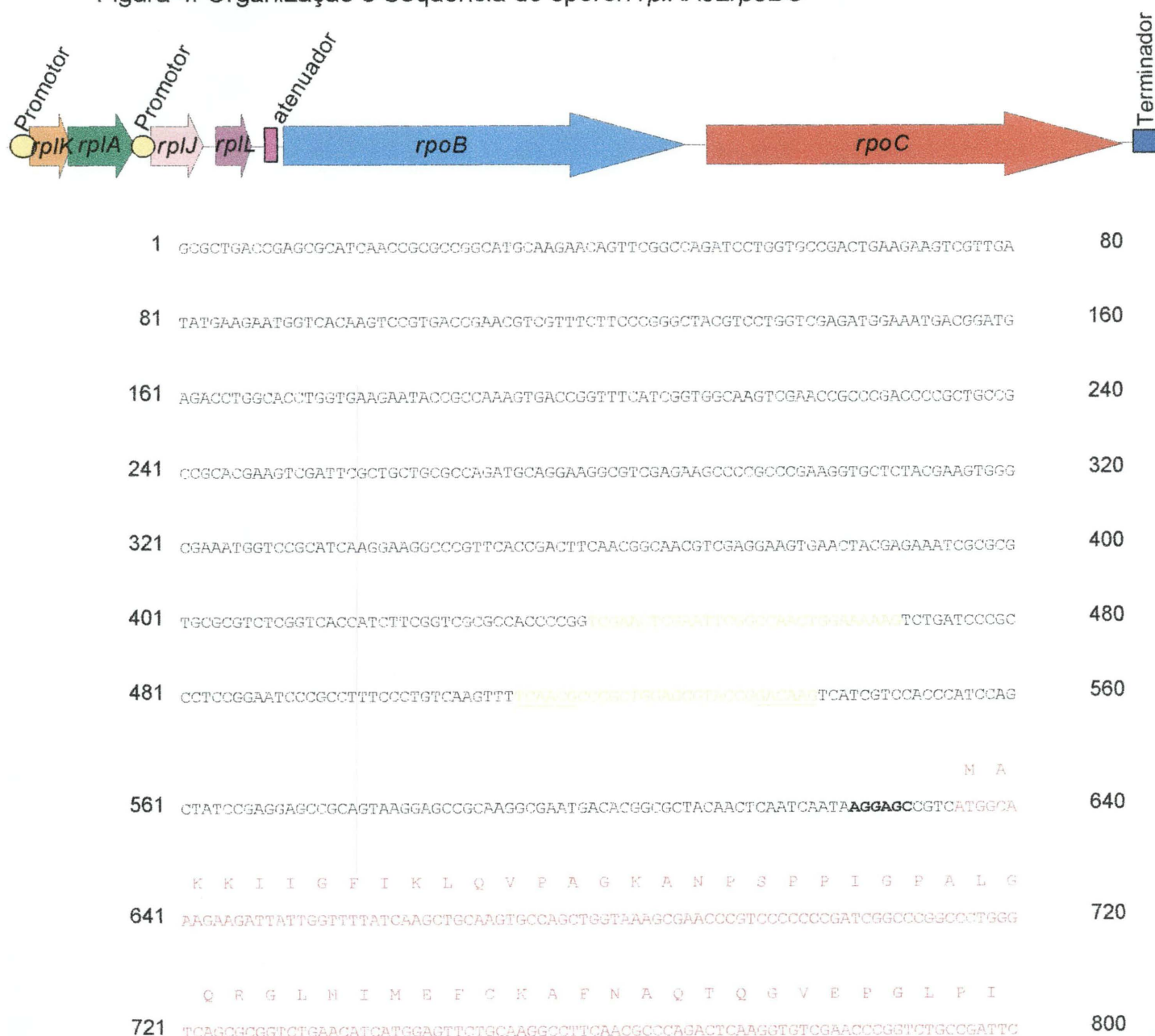
#### 4.1.1 Operon *rplKAJLrpoBC*

A seqüência de nucleotídeos da posição 7500 até 19900 do contig 194 (montagem de 18/08/2003) foi obtida no banco de dados do programa Genopar e analisada pelo programa FramePlot 2.3.2. A análise demonstrou que a organização do operon *rplKAJLrpoBC* em *H. seropedicae* também é idêntica a organização encontrada em *E. coli* (Figura 4).

Três prováveis promotores tipo sigma 70 foram identificados, dois deles na região a montante ao início de tradução do primeiro gene do operon (TCGAACTCGAATTCGGCCAACTGGAAAAAG e TCAACGCCCGCTGGAGCGTACCGGACAAG) e um na região intergênica a montante ao início de tradução do gene *rplJ* (TGCCCTATGCTGCAGTGA ACTCCTAAACG). O primeiro gene do operon é o gene *rplK* (429pb) que codifica para a proteína ribossomal L11 (142aa), 5 pares de base a montante ao códon de início de tradução encontra-se um provável sítio de ligação de ribossoma, AGGAG. O gene seguinte é denominado *rplA* (693pb) que codifica para a proteína ribossomal L1(230aa) e possui um provável sítio de ligação de ribossoma, GGAGG, 7pb a montante ao códon de início de tradução. Não existe região intergênica entre esses genes, o códon de parada coincide com o códon de início de tradução do próximo gene. O terceiro gene que compõe este operon é o gene *rplJ* (534pb) que codifica para a proteína ribossomal L10 (177aa) e possui um provável sítio de ligação de ribossoma, AGGCAAG, 6pb a montante ao códon de início de tradução. O próximo gene deste operon é o gene *rplL* (369pb) que codifica para a proteína ribossomal L12 (122aa) e também possui um provável sítio de ligação de ribossoma, GGAG, 6pb a montante ao códon de início de tradução. O penúltimo gene deste operon é o gene *rpoB* (4101pb), que codifica para a subunidade  $\beta$  da RNA polimerase (1366aa) e possui um sítio de ligação de ribossoma, GGAG, 5pb a montante ao códon de início de tradução. O último gene deste operon é o gene *rpoC* (4242pb), que

codifica para a subunidade  $\beta'$  da RNA polimerase (1413aa) e possui um sítio de ligação de ribossoma, GGAG, 6pb a montante ao códon de início de tradução. Prováveis regiões formadoras de grampos, que podem atuar como terminadores, foram encontradas após o códon de final de tradução do último gene do operon (AGCGCAGCGCAGGCATTGCCAACGAAGA, TTCATGGCGGCCTGCGTATGCA e TGGCGCTGCAGCAGTGGCTATGCGT). Também foi identificado um provável grampo na região intergênica entre *rplL* e *rpoB*, que pode funcionar como um atenuador de transcrição (TTGAGTGCCGATGGCCCGGACT).

Figura 4. Organização e Seqüência do operon *rplKAJLrpoBC*



E V V I T A F A D K S F T F V M K T P E A T I L I K K  
 801 CGGTGGTGGATCACCAGCCTTGGGGGACAAAGTCTTCCACTTCCGTGATGAAGACCCCGCGGGCGACCATCCTGATCAAGAAE 880  
  
 A A G I Q K G S S K P H T D K V G K I T R K Q A E E I  
 881 GCTGCTGGTATCCAGAAGGGTTCCCTCCAAGCGGCATACCGACAAGGTGGGCAAGATCACCCGCAAGCAAGCGGAAGAGAT 960  
  
 A T Q K K P D L T A A D L E A A V R T I A G S A R S  
 961 CGCAACCCAGAAGAAGCCGGACCTGACCGCTGCTGATCTGGAAAGCTGCCGTGCGTACCATCGCTGGTTCCGACCGTTCCGA 1040  
  
 M G I T V E G L \* M A K V S K R V F A M K E K V D R T  
 1041 TGGGCATCACGGTGGAGGCTCTGTAAATGGCTAAGGTATCCAAAGCGGTCAAGGCAATGAAGGAAAAGSTCGATCGTACCA 1120  
  
 K A Y P F D N A V S L I K E F A T A K E N E S I D V S  
 1121 AGGCCTATCCGTTCCACAATGCCCTTCCCTGATCAAGGAATTCGCAACCCGCCAAGTTCACAAGAATCGATCGACGPTCC 1200  
  
 V Q L S V D A K K S D Q V V R G S V V L P A G T G K T  
 1201 GTCCAGCTGGGGCTGGACGGGAAGAAGTCGGACCAAGTGGTCCGTGGTTCCGGTGGTGCCTGCCGTACCGGCAAGAC 1280  
  
 V R V A V F A S G D K A S A A K A A G A D V V G M E  
 1281 CGTTGGCGTGGCCCGTGTTCGCTCCGGGACAAAGGCTGAAGCTGCCAAGGCAGCCCGGCGTCTGACGTGGTCCGGCATGGAAG 1360  
  
 D L A E R V K A G D M P F D I V I A S P D T M R I V G  
 1361 ATCTGGCAGAGCCCGTGAAGGCCGGCCGACATGCCCTTCGACATCGTTCATCGCTTCCGCGACACCATCGSTATCGTCCGT 1440  
  
 T L S Q I L S F R G L M P N P K V G T V T P D V A T A  
 1441 ACCCTGGGTGAGATCCTGGGCCCCCGTGGCTCATGCTAAACCCGAAAGCTTGGCACCGTCAACCCGGACGTTGGCTACTGC 1520  
  
 V K N A K A G Q V Q Y R T D K S G I I H A T I G R K  
 1521 CGTCAAGAAACGCCAAGGCTGGTCAAGTGCARTACCGTACCGACAAGTCCGGIATCATCCAGGCCACCATCGGGCCGCAAGT 1600  
  
 S F S D A D L K A N L L A L I D A I N K A K P A S S K  
 1601 CGTTCTCCGACGCCGATCTGAAGGCCAACCTGCTGGCCCTGATCGATGCCCTGAACCAAGGCCAAGCCGGCCAGGCAAG 1680  
  
 G V Y L R K I A L S S T N G A G V R V D Q A S L A A \*  
 1681 GGTGTGTACCTGCCAAGATCGCTCTGTCTCCACCATGGGGCGCTGGCGTTCGCGTGGACCAAGCTTCCCTGGCCGCTA 1760  
  
 A G C A A G C A A T A C A G T C C C C G T T C C C T G A G G A A G C G G G C A T C T C T T G G G C T G T C G C A G T G C A T T G C A A T C A G C G A T G G  
 1761 1840  
  
 A T T G C G G C A G G C C T T C A A A G A C C G T T G G G C G G A A G C C G G C A G C A A C A A G C C G G T C G A A G T T A A A A A C G C T G T A C G C A A G T  
 1841 1920

1921 GCGGTGGACCCCAACGCAGATGGTGAACCCGAGCAAGTTTGGTAGTCTCAGGCAGCATATTCAGTTT**TGGCTATGATCTA** 2000

M A C V V A E Y Y

2001 **GTGAACTCCTAAAC**TCGGACGCCGTGTTGGAACCCGGTGCA**AGGCAAG**CAGGCAATGGCCCTGTGTAGTGGCCGAGTACTA 2080

F W R L I L S L N L N D K K A V V A E V S A K V A S

2081 TTTTGGAGGTTGATCTTGAGTCTCAATCTGAATGACAAGAAGGCCCGTCCGCGCCGAAGTTTCGGCAAAGGTTGCATCTG 2160

A Q T I V V A E Y R G I Q V G S L T Q L R A Q A R A Q

2161 CGCAAACGATCGTCCGTGGCCGAATACCGTGGCATCCAGGTTGGTTCCCTTGACGCAACTGCGCGCACAAAGCGCGTGGCCAA 2240

G V Y L R V L K N T L A R R A I E G T K F A D L A P Q

2241 GGCSTGTACCTGCGCGTGTGAAAAACACCGTGGCTCGTCCGCGCATCGAAGGCACGAAGTTTGCCGACCTCGCCCGCA 2320

L T G P L I Y S I S E D A V A A A K V V S D F A K T

2321 ACTGACCGGTCCGCTGATCTACTCGATCTCGGAAGATGCCGTGGCCGCTGGCCAAAGTCGTGTCCGACTTTGCCAAGACCA 2400

N D K L V V K A G N Y E G K Q L D K A A V A S L A N I

2401 ACSACAAGCTGGTTGTGAAGGCAGGTAACACTCGAAGGCCAAGCAGCTGACAAAGGCTGCTGTGGCTTCGCTGGCAAACATC 2480

P S R E V L L A Q V L G M M Q A P V S G F A R A L A A

2481 CCGTCTCGGSAAGTTCTGCTGGCCAGGTTCTGGGCATGATGCAGGCTCCGGTGTCCGGCTTTCCGCGGCTCTGGCTGC 2560

L A A K K E A E A A \*

2561 CCTGGCAGCCAAAGAAGGAAGCCGAAGCCGCATGAGGCAGGGCCGCTGGAAGCGGCCTCGCTCTGCAGTTTCGACGCTTTC 2640

M A I S K E D I L E

2641 GCGTCACTACCAATCTGATGTTATTACCGAAATAAATTTAGGAGTTTCAAATGGCAATTAGCAAAGAAGACATCCTGGAA 2720

A V G A L S V M E L N D L V K A F E E K F G V S A A A

2721 GCGGTTGGCGCGCTGTCGGTGATGGAACCTGAATGACCTGGTCAAGGCATTCGAAGAGAAGTTTGGTGTTCGGCTGCCGC 2800

M A A P A A G G A A G G G A A A A E E Q T E F T V V

2801 GATGGCCGCTCCGCGCCCGGTTGGCGCTGCTGGTGGTGGTGGTGGCGCTGCTGAAGAACAGACCGAATTCACCGTTGTGC 2880

L A E V G A N K V G V I K A V R E I T G L G L K E A R

2881 TGGCTGAAGTCCGGCCCAACCAAGGTCGGGCTCATCAAGGCCSTCCGTGAAATCACCGCCCTGGGTCTGAAGGAAGCAAGG 2960



I W L T V H R S R S R K A S P R L T L K L P R R S W K  
 2961 ATCTGGTTGACGGTGCACCCGAAGCCGGTCAAGGAAGGCATCGCCAAGGCTGACGCTGAAGCTGCCAAGAAGAAGCTGGAA 3040  
 K L A P R P N \*  
 3041 GAAGCTGBCGCCAAGGCCGAAGTGAAGTAATTCTCTTGCCCTACCAAGGCGAGTCATTACTCCGGAGTCAAAGTGCAGAAT 3120  
 3121 CCCTGAAAAAGGGATTCCGGCTTTGGCTCCTTTGTGCTTTCTTATTTTGTCTGGTGACGTTGACGCGCCGCGCAGTATTCGTT 3200  
 3201 GTCGGGAATATTAAGCGCGATCTCAACAAAAACTAGGGTAGACATTTGAAGATTTGAGTGCCGATGGCCCGGACTGCTT 3280  
 M H Y S  
 3281 TGCCGGCAAGTCCGCGCGTCTGATCGAATCTTGAATCTCTATCCTTCCCTGTCACCTCACGGAGTGTCCATGCACTACTCA 3360  
 F T E K K R I R K S F A K R A N V H N V P F L L A T Q  
 3361 TTTACTGAGAAGAAGCGCATTCCGCAAATCCTTCGGGAAACGCGCAAACGTCACACACGTTCCGTTCCCTGCTCGCGACTCA 3440  
 I E S Y Q N F L Q P E R A A S E R K N E G L Q S A F  
 3441 GATCGAGTCGTATCAGAATTTCTTCAACCCGAACGTGCAGCATCGAGCGCAAGAACGAAGGCCTGCAATCGGCTTTCA 3520  
 T S I F P I V S H N G F A R L E F L S Y V L G A P P F  
 3521 CCTCGATTTCCCATCGTGTGCGCACATGGTTTTGCGCGCCTCGAATTCCTGTCGTATGTGCTGGGCGCTCCGCGCTT 3600  
 D V K E C Q Q R G L T F A S P L R A K V R L V I L D K  
 3601 GACGTCAAAGAGTGCCAAACAGTGGCCTGACCTTCGCGTCCCGCTGCGCGCCAAAGGTGCGCCTGGTGATCCTGGACAA 3680  
 E S P T K P V V K E M K E Q E V Y M G E L P L M T T  
 3681 GGAATCGCCACCAAGCCGGTCTGTCAGGAATGAAGGAACAGGAAGTCTACATGGGCGAACTGCCCTGATGACCACCA 3760  
 T G S F V I N G T E R V I V S Q L H R S P G V F F E H  
 3761 CCGGTTCTTCGTATCAACGGCACCGAGCGCGTGTGCTCTCCAGCTGCACCGCTCCCGGGCGTGTCTTTCGAACAC 3840  
 D R G K T H S S G K L L F S A R I I P Y R G S W L D F  
 3841 GACCGGGCAAGACGATTCGTCCGCAAGCTGCTGTTCTCGGCCCGCATCATTCCTTACCGTGGCTCCPBGCTGGACTT 3920  
 E F D P K D I L F E R V D R R R K M P V T I L L K A  
 3921 CGAATTCGATCCGAAGGACATCCTGTTCTTCGCGTCCGACCGCCCGCAAGATGCCGGTGCAGGATCCTGCTCAAGGCCA 4000  
 I G M T S E Q I L A N F F V F D N F N L H A E G A D M  
 4001 TTGGCATGACCTCCGAGCAGATCCTGGCGAACTTCTTCTGCTCTCGACAACTTCAACCTGCACGCGGAAGGCGGGACATG 4080

E F V S E R L R G E V A R F D I T D K S G K V M V A K  
 4081 GAGTTCGTCTCCGAGCGCCTGCGCGGTGAAGTGGCACGTTTCGACATCACCCACAAGTCGGGCAAGGTGATGGTCCGCAA 4160  
  
 D K R I N A K H V R D I E A A G I K H I S V P E D Y  
 4161 GGACAAGCGCATCAACGCTAAGCACGTCGCTGACATCGAAGCCGCGGCATCAAGCACATCTCCSTGCTGAGGACTACC 4240  
  
 L L G R V L A K N I V D P E T G E V V A S A N D E L T  
 4241 TGCTGGGCCGCGTGTGGCCAAGAACATCGTCGATCCGGAACCGGTGAAGTCGTGGCCAGCGCCAACGACGAGCTGACC 4320  
  
 E E L L A K L R E A N I S A I Q T L Y T N D L D Q G G  
 4321 GAAGAACTGCTGGCCAAGCTGCGCGAAGCCAAACATCTCGGGCATCCAGACCCGTGACACCAACGACCTGGACCAGGGCGG 4400  
  
 Y I S Q T L R T D D T A D Q T A A R V A I Y R M M R  
 4401 CTACATCTCGCAGACCCTGCGCACCGACGACACCCGCGACCAGACCCTGCGCAGCGTGGCGATCTATCCCATGATGCGTC 4480  
  
 P G E P P T E E S V E A L F N G L F Y S E D R Y D L S  
 4481 CTGGCGAACACCACCGAAGAGTCCGTGGAAGCCCTGTTCAACCGCCTGTTCTACAGCGAAGACCGCTACGACCTGTGCG 4560  
  
 A V G R M K F N R R V G R D E L T G T M T L S N D D I  
 4561 GCCGTGGGCCCGCATGAAGTCAACCGTCCGCTCGGTGCGGATGAGCTGACCCGGCACCATGACCCCTGTCCARTGACGACAT 4640  
  
 L A V I K I L V E L R N G R G E V D D I D H L G N R  
 4641 CCTGGCCGTCATCAAGATCCTGGTCAACTGCGTAACGGCCCGGGCGAAGTGGATGACATCGACCACCTGGGTAAACCGTC 4720  
  
 R V R C V G E L A E N Q F R A G L V R V E R A V K E R  
 4721 GCCTGCGTTGCTGGGCAACTGGCCGAGAACCACTCCCTGCAGGTCTCGTGGCGCTGGAGCGCGCTGTCAAGGAACCG 4800  
  
 L G Q A E A D N L M P H D L I N S K P I S A A I R E E  
 4801 CTGGGCCAAGCCGAGGCGGACAACCTGATGCGGCACGACCTGATCAACTCCAGCCGATCTCGGCCGCGCATCCGTGAGTT 4880  
  
 F G S S Q L S Q F M D Q T N F L S E I T H K R R V S  
 4881 CTTCCGTTCTGTCGAGTTGTCCGAGTTCATGGACCAGACCAACCCGCTGTCCGGAATCAGGCACAAGCGCCGCGTTTCCG 4960  
  
 A L G P G G L T R E R A G F E V R D V H P T H Y G R V  
 4961 CTCTGGGCCCTGGCGGTCTGACCCCGGAACCGCCCGGCTTCGAAGTCCGCGACGTGCACCCGAOCCACTACGGCCGCGTC 5040  
  
 C P I E T P E G P N I G L I N S L A L Y A R L N E Y G  
 5041 TGCCCGATCGAAACGCTGAAGGCCCGAACRTCGGCCTGATCAACTCGCTGGCGCTGTAGGCCCGCTGAACGAATAACGG 5120



F L E T P Y R K V E G S K V T N Q I D Y L S A I E E  
 5121 TTTCCTGGAAACCCCGTACCGCAAGGTCGAGGGCAGCAAGGTCACCAACCAGATCGACTACCTGTGGCCATCGAAGAAG 5200

G R Y V I A Q A N A T I D G E G K L S D E L V S A R E  
 5201 GTGGTTACGTGATCGCCAGGCGAAGCGACCATCGACGGCGAAGGCAAGCTGTCCGACGAACTGGTGTGGGCCCGTGAA 5280

A G E T I L V S P E R V Q Y M D V A P G Q V V S V A A  
 5281 GCCGGCGAAACCATCCTGCTCTCGCCGGAAGCGGTGCACTACATGGACGTGGCCCCGCGAGGTGGTGTGGGTGGCGGC 5360

S L I P F F L E H D D A N R A L M G A N M Q R Q A V F  
 5361 CTCGCTGATTCGGTTCCCTGGAGCACGATGACCGGAACCGTGCCTGATGGGCGCCACATGCAGCGCCAGGCGGTGCCCT 5440

C L R P E K P L V G T G I E R T V A V D S G T T V Q A  
 5441 GCCTGCGTCCGAAAAGCGCTGGTGGCCACCGGCATCGAGCGTACCGTGGCAGTGCCTCGGTACCACCGTGCACAGCC 5520

L R G G V V D Y V D A G R V V I R V N D D E A Q A G E  
 5521 CTGCGTGGCGCGCTGGTGGACTACGTGATGCAGGCGGTGTGGTGTATCCGGGTCAACGACGACGAAGCGCAAGCTGGCGA 5600

V G V D I Y N L I K Y T R S N Q N T N I N Q R P I V  
 5601 AGTGGGTGGGACATCTACAACCTGATCAAGTACACCCGTCCCAACCAGAACACCAACATCAACCAGCGCCCGATCGTCA 5680

K I G D R V A K H D V I A D G A S T D M G E L A L G Q  
 5681 AGATCGGTGACCGCGTGGCCAAAGCATGACGTGATCGCCGACGGCGCTCGAACCGACATGGGCGAGCTGGCGCTGGGCCAG 5760

N M L V A F M F W N G Y N F E D S I L I S E K V V A D  
 5761 AACATGCTGGTGGCGTTTCATGCCCTGGAAACGGCTACAACCTCGAAGACTCGATCCTGATCTCGAAAAGGTCGTGGCTGA 5840

D R Y T S I H I E E L S V V A R D T K L G A E E I T  
 5841 TGACCGCTACACCTCGATCCACATCGAGGAACTGTGGTGGTGGCGCGTGACACCAAGCTGGGTGCAGAAGAAATCACCC 5920

R D I S N L A E N Q L A R L D E S G I V Y I G A E V E  
 5921 GCGACATCTCCAACCTGGCCGAGAACCAACTGGCCCGTCTGGATGAGTCCGGTATCGTCTACATCGGCGCGGAAGTCSAA 6000

A G D T L V G K V T P K G E T Q L T P E E K L L R A I  
 6001 GCTGGCGACACCCCTGGTGGTAAGGTACCCCCAAGGGCGAGACCCAGCTGACTCCGGAAGAGAAGCTGCTGCGCGCGAT 6080

F G E K A S D V K D T S L R V P S G M V G T V I D V  
 6081 CTTGGCGAGAAAGGCTCCGACGTGAAGGACACCTGGCTGGCGCTGCCCTCTGGCATGGTGGCCACCGTCATCGACGTGC 6160

	Q V F T R E G I Q R D K R A Q Q I I D D E L K R Y R L	
6161	AGGTGTTTCACCCGCGAAGGCATCCAGCGCGACAAGCGCGCCAGCAGATCATCGACGATGAACCTCAAGCGCTATCGCCTG	6240
	D L N D Q L R I V E G D A F Q R L E R L L I G K V A N	
6241	GACCTGAACGACCAGCTGCGTATCGTGGAAAGCGGATGCCTTCCAGCGTCTGGAGCGTCTCTTGATCGCCAAAGGTCGCCAA	6320
	G G P K K L V K G T K L D K A Y L D D L D K F H W F	
6321	CGGCGGCCCAAGAAGCTGGTCAAGCGCACCAAGCTGGACAAGGCCTACCTGGACGATCTGGACAAGTTCCTACTGGETTCG	6400
	D I R P A D D D I A N A L E A I K E S I A E K R H Q F	
6401	ACATCCGCCCGGCCGACGACGACATCGCCACCGCACTGGAAGCCATCAAGGAATCCATCGCCGAGAAGCGTCAACCAGTTC	6480
	D L A F E E K R K K L T Q G D E L P P G V Q K M V K V	
6481	GACCTGGCCTTCGAAGAGAAGCGCAAGAAGCTGACCCAGGGCGATGAACTGCCGCGGGCGTGCAGAAGATGGTCAAGGT	6560
	Y L A V K R R L Q P G D K M A G R H G N K G V V S R	
6561	CTACCTGGCTGTGAAGCGTCCCTGCAGCCTGGCGACAAGATGGCCGGTCTGCACGGTAACAAGGGTGTGGTTTCGCGCA	6640
	I L P I E D M F H M A D G T P A D I V L N P L G V P S	
6641	TCCTGCCGATCGAAGACATGCCCGACATGGCCGACGGTACCCCGCCGACATCGTGTGAACCCGCTGGGCGTGCCTCG	6720
	R M N V G Q V L E V H L G W A A K G L G L R L G E M V	
6721	CGTATGAACGTTGGCCAGGTGCTGGAAGTTCACCTGGGCTGGGCCGCCAAGGGTCTGGGCTGCGTCTGGGCGAGATGGT	6800
	Q A Q A K V A E L R K F L T T I Y N E S G K Q E E L	
6801	GCAGGCGCAAGCCAAGGTTCGCGAAGTTCACCTGGGCTGGGCCGCCAAGGGTCTGGGCTGCGTCTGGGCGAGATGGT	6880
	D S F T D E E I I E L V E N L K K G V P F A T P V F D	
6881	ATTGTTTCACCGATGAAGAAATCATCGAGCTGGTGGAAAACCTGAAGAAGGGCGTGCCTTCGCGACCCCGGTGTTTCGAC	6960
	G A H E E E T R R M L D L A Y P D P I A K Q L G M T P	
6961	GGCGCGCATGAAGAAGAAACCCGCGCATGCTCGATCTGGCCTATCCGGACCCGATCGCCAAGCAGCTGGGCATGACCCC	7040
	S K N Q V T M Y D G R T G E A F E R T V T V G Y M H	
7041	GTCCAAGAACCAGGTCACGATGTATGACGGCCGTACCSTGAAAGCCTTCGAGCGCACCGTGACCGTCCGGCTACATGCACT	7120
	Y L K L H H L V D D K M H A R S T G P Y S L V T Q Q P	
7121	ACCTGAAGCTGCACCACCTGGTCCGACGACAAGATGCACGCCCGTTCACCCGGCCCGTACTCGCTGGTGACCCAGCAGCCG	7200

L G G K A Q F G G Q R F G E M E V W A L E A Y G A S Y  
 7201 CTGGGCGGCAAGGCGCAGTTCCGGCGGCCAGCGTTTCGGCGAAATGGAAAGTGTGGGCGCTGGAAGCGTACGGCGCGTCCCTA 7280  
  
 V L Q E M L T V K S D D V N G R T K V Y E N L V K G  
 7281 CGTCCTGCAGGAAATGCTGACGGTCAAGTCCSACGACGTBAACCGCCGTACCAAGGTCTACGAGAACCTGGTCAAGGGCG 7360  
  
 D H V I D A G M P E S F N V L V K E I R S L G I D I D  
 7361 ACCACGTGATCGATGCCGGSTATGCCSGAATCCTTCAACCGTGTGGTCAAGGAAATCCSCTCGCTGGGTATCGACATCGAC 7440  
  
 L E R E \*  
 7441 CTGGAACCGGAATGATGTAAAGGGGTGCGCGGCGCAGGCTGCGCACCCCGCGGTTTTTCCGGCGAACCAAGAATTC AAG 7520  
  
 AATTCAGTATTTCAGTTTTGACGCATCGCTCCACGGGCGATGTGAACGAATCAAGAGATTCAAGAGATTCAAGCGAAGTAC 7600  
  
 GGCAGTAAGCTGGGGAGGTCTGTCCGGTATCGAGCGGCACCAAGCAGAGGCTTCGGCTCTCCCGCGATGCGA 7680  
  
  
 M K A L  
 7681 ACAGGGTTAAGCAGGSETGAACCGGCCACAGTAAATTTAGTTTTTTCACGCCAACCTGGAGTGATACATGAAAGCACTGC 7760  
  
 L D L F K Q V Q Q N E Q F D A I K I G L A S F D K I R  
 7761 TCGATCTATTCAAGCAAGTCCAGCAAAACGAACAATTCGACGCGATCAAGATCGGTCTGGCTTCGCCCTGACAAGATCCGC 7840  
  
 S W S F G E V K K P E T I N Y R T F K P E R D G L F C  
 7841 TCGTGGTCTTCGGCGAAGTCAAGAAGCCGGAACCATCAACTACCGTACCTTCAAGCCCGAGCGCGACGGCTGTTCG 7920  
  
 A K I F G P I K D Y E C L C G K Y K R L K H R G V I  
 7921 CGCCAAGATCTTCGGCCCCATCAAGSACTATGAGTGCCTGTGCGGCAAATACAAGCGCCTGAAACACCGTGGCGTGAICT 8000  
  
 C E K C G V E V T L A K V R R E R M G H I E L A S P T  
 8001 GCGAGAAGTGGGCGTGGAAAGTGACACTGGCCAAAGGTGCGTCTGTGAGCGCATGGGCCACATCGAGCTGGCGTCCCCGACT 8080  
  
 A H I W F L K S L P S R L G M V L D M T L R D I E R V  
 8081 GCCCACATCTGGTTCCTGAAGTCCGCTGCCGTCCCGTCTGGGCATGGTGTGTSACATGACCCTGCGCGACATCGAGCGCGT 8160  
  
 L Y F E A Y V V T D P G M T P L K K C Q I M S E D D  
 8161 GCTGTACTPCGAAGCCTACGTGCTGACCGATCCTGGCATGACCCCGCTGAAGAAGTCCAGATCATGTCCGAAGACGACT 8240  
  
 Y A A K Y E E Y G D D F I A F M G A E G I R E L L R S  
 8241 ACGCCGCCAAGTACGAAGAGTACGGCGACGACTTCATCGCCTTCATGGGCGCGAAGGTATCCGCGAAGTGTGCGCTCG 8320

I D I D R E A E T L R Q E L N D S K E E A K I K K Y A  
 8321 ATCGATATCGACCCGGAAGCGGAAACCCTGGGCGAGGAACTGAAGGACTCCAAGTCCGAAGCCAGATCAAGAAGTACGC 8400

K R L K V L E A E Q R S G I K P E W M I M E V L P V  
 8401 CAAGCGCTGAAGGTGCTGGAGGGTTCCAGCGCTCGGGCATCAAGCCGAGTGGATGATCATGGAAGTGGCTGCCCGTGC 8480

L P P E L R P L V P L D G G R F A T S D L N D L Y R R  
 8481 TGCCCGCGAACTGGCTCCGCTGGTCCCGCTGGATGGCGCCCGCTTCGCGACCTCCGACCTGAACGACCTGTACCCGCGC 8560

V I N R N N R L K R L M E L R A P E I I T R N E K R M  
 8561 GTCATCAAGCGTAACAACCGTCTGAAAGCGCTGATGGAAGTGGCGCGCCGAAATCATCACGCGCAACGAAAAGCGCAT 8640

L Q E A V D S L L D N G R R G K A M T G A N K R P L  
 8641 GTTGCAGGAAGCGGTGGACTCGCTGCTGGACAACGGTCTGCGCGCAAGGCCATGACCGCGCCCAACAAGCGTCCGCTGA 8720

K S L A E N I K G K G G R F R Q N L L G K R V D Y S G  
 8721 AGTCGCTGGCCGAAATGATCAAGGGCAAGGGCGGTCTGTTCCCTCAGAACCTGCTGGGCAAGCGCGTGGACTACTCGGGC 8800

R S V I V V G P Q L K L H Q C G L P K L M A L E L F K  
 8801 CGTTCGCTCATCGTGGTGGGCCCGCAGCTCAAGCTGCACCAGTGGCGCCTGCCCAAGCTGATGCCCCTGGAAGTGTTCGA 8880

P F I F N K L E L H G L S T T I K A A K K L V E A Q  
 8881 GCCCTTCATCTCAACAAGGTGGAAGTGGTCTGTGACACCACATCAAGGCCGCCAAGAAGCTGGTGGAAAGCGCAAG 8960

E P V V W D I L E E V I R E H P V M L N R A P T L H R  
 8961 AGCCGGTGGTCTGGGACATCCTGGAAGAAGTCAACCGCAACACCCGGTCAATGCTGAACCGCGCGCGACCCCTGCACCGT 9040

L G I Q A F E P V L I E G K A I Q L H P L V C A A F N  
 9041 CTGGGCATCCAGGCGTTCCAGCCGGTCCIGATCGAAGGCAAGGCGATCCAGCTGCACCCGCTGGTGTGCGCCGCGSTTCAA 9120

A D E D G D Q M A V H V P L S I E A Q M E A R T L M  
 9121 CGCCGACTTCGACGGCGCACAGATGGCCGTGCACGTGCCGCTGTGATCGAAGCGCAGATGGAAGCCCGCACCCCTGATGC 9200

L A S N N I L F P S N G E P S I V P S Q D I V L G L Y  
 9201 TGGCCTCCAACAACATCCTGTTCCCGTCCAACGGCGAACCCTGATCGTGCCTGCCAGGATATCGTCTGGGTCTGTAC 9280

Y A T R E A I N G K N E G M M F Q D V S E V I R A Y D  
 9281 TATGCCACCCGTGAAGCCATCAACGGCAAGAAGCAAGGCATGATGTTCCAGGACGTGTGGGAAGTCAACCGCGCTACGA 9360

N K E V E L A T R I T V R I V E H P K D F V T G E F  
 9361 CAACAAGGAAGTCSAGCTGCCACCCGTATCAGGGTGGCGTATCGTGGAAACACCCGAAGGACCCGGTCAACGGCGAATTCCG 9440  
 V R T V K R Y E T T I G R A I L S E I L P K G L P F S  
 9441 TGGGTACCGTCAAGCGCTACGAGACACCATCGGCCCGGCCATCCTCTCCGAGATCCTGCCCAAGGGCCTGCCGTTCTCCG 9520  
 V L N R A L K K K E I S K L I N T S F R K C G L R A T  
 9521 GTGCTGAACCGCGCGTTGAAGAAGAAGGAAATCTCGAAGCTGATCAACACGTCCTTCCGCAAGTGGCGTCTGCCGTGCCAC 9600  
 V V F A D Q L M Q S G F R L A T R A G I S I C V D D  
 9601 CGTCTGTTTGGCCAGCCAGCTGATGCAGTCCGGCTTCGGCCTGGCGACCCCGCCGGTATCTCCATCTGCCGTGGAGGACA 9680  
 M L V P P Q K A T L I A A A E S E V K Q I E Q Q Y S S  
 9681 TGCTTGTGCCCGCCAAAAGGCCACCCCTGATCGCCGCTGCCGAGTCCGAASTCAAGCAGATCGAACAGCAGTACTCCTCCG 9760  
 G L V T A G E R Y N K V V D I W G K A G D E V G K A M  
 9761 GGTCTGGTGAACCGCAGGCGAGCGCTACAACAAGGTCTGGGACATCTGGGGCAAGGCCCGGCGACGAAGTCCGCCAAGGCCAT 9840  
 M D Q L K V E D V I T R E G K K T T Q E S F N A I Y  
 9841 GATGGACCAGCTGAAGGTGGAAGACGTCAACCCCGGAAGGCAAGAAGACCCACCCAGGAATCGTTCAACCGCATTTACA 9920  
 M M A D S G A R G S A A Q I R Q L A G M R G L M A K P  
 9921 TGATGGCCGACTCCGGCGCCCGTGGTTCCGGTGGCCAGATCCGCCAGCTGGCCGGTATGCCGCGCCTGATGGCCCAAGCCT 10000  
 D G S I I E T F I T A N F R E G L N V L Q Y F I S T H  
 10001 GACCGCTCCATCATCGAGACGCCCATTAACCGCAACTTCCCGGAAGGCTGAAACGTTGTCAGTACTTCATCTCGACCGCA 10080  
 G A R K G L A D T A L K T A N S G Y L T R R L V D V  
 10081 CCGCGCCCGTAAGGGTCTGGCCGACACCCCGCTGAAGACCGCCAACTCCGGTTAAGTACCCGCTCGTCTGGTGGAGCTGA 10160  
 T Q D L V V I E D D C G T S N G S A M K A L V E G G E  
 10161 CCCAGGATCTGGTCTGATCGAAGACGATTGCGGCACCTCCAAAGGTTCCGGCCATGAAGGCGCTGGTGAAGGCGGTGAA 10240  
 V I E A L R D R I L G R V A A T D I V N P E D Q S T L  
 10241 GTTATCGAAGCCCTGCGCGACCGTATCCTCGGCCGTGGCGCGACCGATATCCTCAACCCGGAAGATCAATCCACCCCT 10320  
 F E A G T L L D E D K V E E I E R M G I D E V K V R  
 10321 GTTCGAAGCCGGCACCCCTGCTGGACGAAGACAAAGTTCGAAGAGATCGAACGCATGGGTATCGACGAAGTGAAGGTCCGTA 10400



T P L T C D T R Y G L C A K C Y G R D L G R G L L V N  
 10401 CTCCGCTGACCTGCGACACCCGCTATGGCCCTGTGCGCCAAGTGCTATGGCCGCGACCTGGGCGCTGGCCCTGCTGGTGAAC 10480

S G E A V G V V A A Q S I G E P G T Q L T M R T F H I  
 10481 TCCGCGGAAGCCCTCGGCGTGGTGGGAGCGCAGTCCATCGGTGAACCGGGTACCCAGCTGACCATGCGTACCTTCCACAT 10560

G G A A S R A A V A S S V E A K S N G T V R F T A T  
 10561 CGGTGGTGCGGCATCGCGTGCAGCGTGGCCCTCCTCGGTGGAAAGCCAAAGTCCAACGGCACCGTGCCTTCACCGCCACGA 10640

M R Y V T N G K G E Q I V I S R S G E V L I T D D L G  
 10641 TCGCTTACCTGACCAATGGCAAGGGCGAGCAGATCGTCAATTCGCCGTTCCGGCGAAGTGCTGATCACCGATGACCTGGGC 10720

R E R E R H K V P Y G A T L I V K D G L T I K A G T A  
 10721 CGTGAGCGGAGCGTCACAAAGTGCCTACGGCGCCACCCTGATCGTCAAGGATGGCCCTGACCATCAAGGCGGGTACCGC 10800

L A T W D P L T R P I I T E Y S G T V K F E N V E E  
 10801 CCTGGCGACCTGGGATCCGCTGACCCGTCGGATCATCACCGAGTACTCCGGTACCCTGAAGTTCGAAAACGTGGAAGAAG 10880

G S T V A K Q I D E V T G L S T L V V I D A K R R G P  
 10881 GCTCCACCGTGGCCAAAGCAGATCGACGAAGTCAACCGGCTCTGTGACCCCTGGTGGTTCATCGATCGAAGCGTCTGGTCCG 10960

S S K V M R P Q V K L L N E Q G E E V K I A G T E H S  
 10961 TCCTCCAAGGTCATGCGTCCGCGAGGTGAAACTGCTGAACGAACAAGGCGAAGAAGTGAAGATCGCCGGCACCGAACACTC 11040

V T I G F Q V G A L I T V K D G Q Q V H V G E V L A  
 11041 GGTGACCATCGGCTTCCAGGTCGGCGCGCTGATCACCGTCAAGGACGGCCAGCAAGTGCATGTGGGTGAAGTCTGGCCC 11120

R I P T E S Q K T R D I T G G L P R V A E L F E A R S  
 11121 GTATCCCGACCGAATCGCAGAGAACCCGTCACATTACCGGTGGTCTGCGCGCGTGGCCGAGCTGTTCAAGSCCCGTTG 11200

P K D A G M L A E V T G T V A F G K E T K G K Q R L E  
 11201 CCCAAGGATGCGGCGATGCTGGCCGAAGTCACTGGTACCCTTGCCTTCGGCAAGGAAACCAAGGGCAAGCAGCGTCTGGA 11280

I T D M E G T K H E F L I T K D K Q V L V H D G Q V  
 11281 AATCACCGACATGGAAGGCACCAAGCACGAGTTCCTGATCACCAAGGACAAGCAAGTGTCTGGTGCACCACGGCCAGGTGG 11360

V N K G E M I V D G P A D P Q D I L R L L G I E A L A  
 11361 TGAACAAGGGCGAATGATGCTGGACGGTCCGGCTGACCCGCGAGGACATCCTGCGCCTGTTGGGTATCGAAGCGCTGGCA 11440

11441	R Y I V D E V Q D V Y R L Q G V K I N D K H I E V I V CGTTACATCGTTCGACGAAGTGCAGGACGTGTACCGTCTGCAGGGCGTGAAGATCAACGACAGCACATCGAAGTGATCGT	11520
11521	R Q M L R R V Q V V E P G D T N Y I T G E Q V E R S GCGCCAGATGCTGCGCCCGTGCRAAGTGGTGGAGCCGGGCGACACCACTACATCACCGGTGAGCAGTTCGAGCGTCCG	11600
11601	E L L D E N D R V T A E G K I P A T Y E N V L L G I T AACTGCTGGACGAGAACGACCCCGTSAACGCGGAAGGCAAGATCCCGGCGACCTACGAAAACGTGCTGCTGGGTATTACC	11680
11681	K A S L S T D S F I S A A S F Q E T T R V L T E A A I AAGGCATCGCTGTGACCGACTCCTTCATCTCGGCCGATCGTTCCAGGAAAACCCCGCGTGTGACCGAAGCCCGCGAT	11760
11761	M G K K D G L R G L K E N V I V G R L I P A G T G L CATGGGCAAGAAGGACCGCCTGCGCGCCTGAAGGAAAACGTCAATCGTCCGCCCGCTGATCCCGGCTGCTACTGGCCTGG	11840
11841	A F H R A R K E K D A W E A E E R A A L L E A E R A S CCTTCCACCCGCGCCGCAAGGAAAAGGATGCATGGGAAGCCGAAGAGCGCGCAGCACTGCTCGAAGCCGAACGTGCCCTCC	11920
11921	R S E F G D E E I S S T E S I G G G E G E A * CGCTCCGAATTCGGCGACGAGGAAATCTCCTCGACCGAATCCATCGGTGGCGGCGAAGGCGAAGCGTAAGCCTCGCATCG	12000
12001	CCCGGGTGGTCCGCTTGTATCGGCCACCCGACACGGCAGAAAGCTAACAGCTGCAAGAGAAGCGGGCTCGGAAGTCATTCC	12080
12081	GGCGCCGTTTTTTTTTCSTCCCGCAAGTCCCTTGTCTTGTCTGGCGTAGGGGGAGCGCAGCGCAGGCATTGCCAACGAAG	12160
12161	AGCGAATCTCCTATCATCTTCAAATAATAACGATCTCATCTGTGCGCGCATGGGTGAGTGAAAGACAGCATTCATGGC	12240
12241	GGCCTGCGTATGCGAGTCTTTCTTTTCCAAGAGGTAGCCAGCAGTCCGCTAGTAACCGATCCAACCAAGGGGTGGTTACA	12320
12321	GTGTCTCATGCAAATCGAGGTGTGCGGCAGCCGTTCTCCTGGCGCTGCAGCAGTGGCTATGCGTCCGCGCCTATGCGCA	12400
12401	AGCGGATCAGCCGCAATCCGTTCAATCCAGTTCATCAGATCCTTCCGTCGTGCTGGCGCAGGCCAGCGGCAAGAAGCTGC	12480

**Legenda:** *rplK* (429pb) laranja; *rplA* (693pb) verde; *rplJ* (534pb) rosa; *rplL* (396pb) roxo; *rpoB* (4101) azul e *rpoC* (4242) vermelho. Os prováveis sítios de ligação de ribossoma, seqüências de Shine-Dalagarno, estão em negrito. As prováveis regiões promotoras estão em amarelo e sublinhado, o atenuador está em rosa escuro e os terminadores em azul escuro. Asteriscos (\*) indicam o códon de parada de tradução.

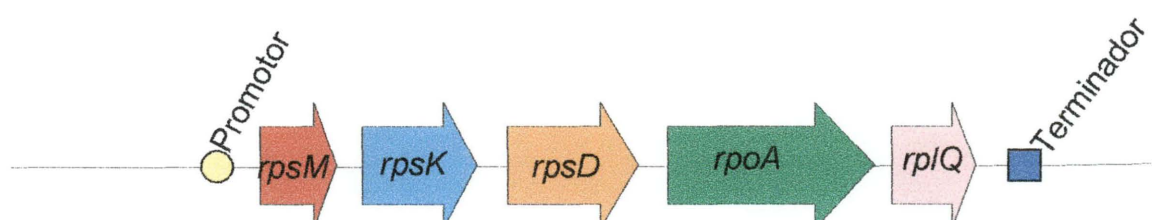


#### 4.1.2 Operon *rpsMKDrpoArpIQ*

Parte da seqüência do contig 282 (posição 5670 até 9200, montagem de 18/08/2003) obtida no banco de dados do programa Genopar foi analisada pelo programa FramePlot 2.3.2. Os resultados revelaram a organização do operon operon *rpsMKDrpoArpIQ* em *H. seropedicae* que é idêntica a organização encontrada em *E. coli* (figura 5).

Dois prováveis promotores do tipo sigma 70 (GTGTTCCGAACCAAGTAACAGTGAAACATT e GTCAAGCGGATCTGCCGC AACTGCAAAAT) foram identificados na região a montante ao início de tradução do primeiro gene do operon. O primeiro gene do operon é o gene *rpsM* (363pb) que codifica para a proteína ribossomal S13 (120aa), 7 pares de base a montante ao códon de início de tradução encontra-se um provável sítio de ligação de ribossoma, AGGA. O gene seguinte é denominado *rpsK* (542pb) que codifica para a proteína ribossomal S11 (184aa) e possui um provável sítio de ligação de ribossoma, AGGA, 7pb a montante ao códon de início de tradução. O terceiro gene que compõe este operon é o gene *rpsD* (621pb) que codifica para a proteína ribossomal S4 (120aa) e possui um provável sítio de ligação de ribossoma, AGGA, 6pb a montante ao códon de início de tradução, que neste caso é o códon alternativo GTG. O penúltimo gene deste operon é o gene *rpoA* (978pb) que codifica para a subunidade  $\alpha$  da RNA polimerase (325aa) e também possui um provável sítio de ligação de ribossoma, AGGA, 6pb a montante ao códon de início de tradução. O último gene deste operon é o gene *rplQ* (396pb), que codifica para a proteína ribossomal L17 (131aa) e possui um sítio de ligação de ribossoma 7pb a montante ao códon de início de tradução. Foi identificado um provável terminador independente da proteína Rho (AAAAAGCCAGGCCTTGCCTGGCTTTTT) 24pb após o códon de final de tradução do último gene do operon.

Figura 5. Organização e seqüência do operon *rpsMKDrpoArpIK*



1 CGTTATTCAGATGCAGGGCGAGATCTTGAAAAATCTGCCAACCGGACATTCAGGGTTAAGTTGGAAAACGGTCACGTTG 80  
 81 TACTCGGCCATATTTCCGGCAAGATGCGCATGAATTACATCCGATCTTGCCCGCGGATAAGGTAACGGTGGAACTGACG 160  
 161 CCTTACGATTGAGCCGGGCAAGTATC **CTTTGAGGATCTCTTAAKYSZAAAT** GAAAGAAAAGGGGCCACAATGAA 240  
 241 AGTGCTCGCATCA **CTTTGAGGATCTCTTAAKYSZAAAT** CATCAAGCGCAATGGCATCGTTTCGGGTGATCTGCACCG 320  
 M A R I A G V N  
 321 AACCCCGCCATAAGCAGCGCCAGGGTTAACTTAACGTTATTGATCG**AGGA**AATAACGARTGGCACGTATTGCAGGGGTCAA 400  
 I F N H Q H T V I G L T A I Y G I G R P R A Q D I C  
 401 CATCCCAACCATCAGCACACCGTGTGATCGGCCCTGACCGCCATCTATGGTATTGGCCGTCCGCGGCACAAAGACATTTGCG 480  
 A A T G V P T N K K V K D L D D N E L E K L R D E I A  
 481 CTGCTACCGCGGTTCCGACCAACAAGAAGGTCAAGGACCTGGACGATAACGAGCTGGAAAAGCTGGGTGACGAAATCGCC 560  
 K F V V E G D L R R E L S M N I K R L M D L G C Y R G  
 561 AAGTTCGTGCGTGAAGGCGATCTGCGCCGTGAGTTGTCCATGAACATCAAGCGTTTGATGGAACTGGGTGCTACCGGG 640  
 L R H R R G L P V R G Q R T R T N A R T R K G P R K  
 641 TCTGCGTCACCGTCTGTTGCTGCGGGTCCGTGGTCAACGTACCCGCAACCAACGCCCGTACTCGCAAGGGTCCGCGCAAG 720  
 A A Q S L K K \* M A K S P  
 721 CCGCTCAATCGCTGAAGAAATAATCCCAGCAGGCGCAAGCACTGGTCCGATTCA**AGGA**AGAAATTATGGCAAAGTCCCC 800  
 N N A A A S R V R K K V K K N V A E G I A H V H A S F  
 801 AACAAACGACGAGCATCGCGTGTTCGTAAGAAAGTTAAGAAGARCGTTGCTGAAGGCATCGCGCACGTTCAACGCTTCGTT 880  
 N N T I I T I T D R Q G N A L S W A T S G G A G F K  
 881 CAACAACACCATCATCAGGATCACCGACCGTCAAGGCAACGGTCTCTGTCGTGGGCGACCTCGGGTGGTGTGCTGGCTCAAG 960  
 G S R K S T P F A A Q V A A E S A G K V A I E C G I K  
 961 GTTCGCGCAAGTCCACTCCSTTCGCGAGCCAGGTGCGCCGCTGAAAGCGCCGCGCAAGSTTGCCATCGAATGCGGCATCAAG 1040  
 N L E V R I K G P G F G R E S A V R A L N N L G I K I  
 1041 AACCTGGAAGTGCATCAAGGGCCCCGGCCCGGGCCGTGAATCCGCGAGTGGCGGCTCTGAAACAACCTGGGTATCARGAT 1120

T Q I Q D V T P V P H N G C P S A E A P S H L S C S  
 1121 CACCCAGATCCAGGACGTGACCCCGGTCCCGCACAACGGCTGCCCGTCCGCCGAAGCGCCGTCCGCATCTAAGTTGTAGTA 1200  
 I I L R F R L F V R Q K R R Q I T R Q S P W R V L F L  
 1201 TAATCTTGGCCTTTCCGCTTTCGTTCAGGCAAAAAGCGCAGGCAAATCACCCCECAATCACDCTTGGCGGGTTTTGTCTTA 1280  
 R P L S G R M Q V R L A S G N M G R H \* V A  
 1281 AGACCACTGTCTGGCCGCATGCAGGTCAGACTAGCGTCCGGCAACATGGGACGTCATTGATAAAGGAATACCGTGGCAC 1360  
 R Y I G P K A K L S R R E G T D L F L K S A R R S L D  
 1361 GTTATATCGGACCGAAAGCAAACCTCTCCCGCGCGAAGSCACCGACCTGTTCCCTGAAGAGCGCACGCCGCTCGCTGGAT 1440  
 S K C K L D S K P G Q H G R T S G A R T S D Y G N Q L  
 1441 TCCAAGTGCAAGCTGGATTCCAAGCCGGGTGAGCACGGCCGCACCTCGGGCGCCCGCACCTCCGACTACGGCAACCASCT 1520  
 R E K Q K V K R M Y G V L E R Q F R R Y F A E A D R  
 1521 GCGCGAAAAGCAGAAGGTCAAGCGCATGTACGGCGTCCCTCGAGCGCCAGTTCCGCGCGCTACTTCCCTGAAGCCGACCGCC 1600  
 R K G N T G E T L L K L L E S R L D N V V Y R M G F G  
 1601 GCAAGGGCAACACCGCGAACCCTGCTGAAGCTGCTGGAATCGCGCTGGACAACTGCTCTACCGCATGGGCTTCCGCG 1680  
 S T R A E A R Q L V S H K A L T V N G Q V V N I A S Y  
 1681 TCGACCCGCGTGAAGCGCGTCAAGCTGGTGTCCACAAAGGCCCTGACCGTGAACGGTCAAGTCCGTGAACATCGCTTCCTA 1760  
 S V K A G D V I A V R E K A K K Q V R I A E A L S L  
 1761 CTCGGTCAAGGCAGGCGACGTCATCGCCGTGCGTGAAAAGGCCAAGAGCAAGTGCCTATCGCGAAGCCCTGTCCGCTGG 1840  
 A E S N G F P Q W V S V D S K K L E G T F K S S P D R  
 1841 CCGAGTCGAACCGTTTTCCCGCAGTGGGTTTTCCGTGGATTGGAAGAAGCTGGAAGGCACCTTCAAGTCCCTCCCGGACCGT 1920  
 S E I A A D V N E S L I V E L Y S R \*  
 1921 AGCGAAATTTGCTGCCGACGTCACGAATCGCTGATCGTCAACTGTACTCGCGTTAATTCGTAGTACCGCCGTACCCCTG 2000  
 2001 TGCCCCCTCATGGTGGGCATTTTTCCGAACTGCCATCAGCCTTATCGGTGTAACGAGCCGAGGGTATTGAAAAGGACAA 2080  
 M Q N S L L K P R I I E V E T L A P G H A K V V M E  
 2081 TTCATGCAAAACAGTTTTGTTGAAGCCCGCCTCATCGAAGTGGAAACCCTGGCCCCCGTCAACGCAAAGGTCGTGATGGA 2160  
 P F E R G Y G H T L G N A L R R V L L S T M S G Y A  
 2161 GCGGTTGAGCGTGGTTATGGTCAACCCCTGGGCAACGCGCTGCGTCCGCTGCTGCTGTCGACCATGAGCGGTTACGCTC 2240

P T E V T I A G V V H E Y S S L D G V Q E D V V D I L  
 2241 CCACCGAAGTGAACATCGCTGGTGTGGTGCACGAACTACGTCCTGGACGGCGTGCAGGAAGACGTCGTGGACATCGCTG 2320  
  
 L N L K G V V F K L H N R D E V T L T L K K E G E G A  
 2321 CTGAACCTCAAGGGCGTGGTCTTCAAGCTGCACAACCGTGACGAAAGTCAACCTGACCCCTGAAGAAGGAAGGCGAAGGGCGC 2400  
  
 V L A S D I D L P H D V E L I N P D H V I A N L T G  
 2401 CGTCTGGGCTTCCGATATCGATCTGCCGCAAGAGCTCGAACTGATCAACCCGGATCACGTCATCGCAACCTGACCCGGTG 2480  
  
 G G K L D M Q I K V E K G R G Y V E G N V R R L S E D  
 2481 GCGGCAAGCTGGACATGCAGATCAAGGTGCAAAAAGGGCCGTGGCTATGTGCCCGGTAACGTGCCTCGTCTGTCCGAAGAC 2560  
  
 T N K T I G R I I L D A S F S P V R R V S Y A V E S A  
 2561 ACCAACAAGACCATCGGCCGCATCATTCTGGATGCCTCGTCTCGCCGGTGCCTCGCSTGTCCCTACGGGTCGAATCCGC 2640  
  
 R V E Q R T D L D K L V M N I E T N G V I T F E E A  
 2641 GCGTGTGGAACAGCGTACCGACCTGGACAAGCTGCTCATGAACATCGAGACCAACGGCGTGCATCACCCCGGAAGAAGOGA 2720  
  
 I R Q S A R V L V D Q L N V F A A L E G T E A T A E A  
 2721 TTCGCCAGTCCGGCTCGTGTGCTGGTTCGATCAGCTCAACGTGTTGCTGCCCTGGAAGGCCACCGAAGCSACTGCAGAAGCA 2800  
  
 F S R A P Q V D P I L L R E V D D L E L T V R S A N C  
 2801 CCGTCCGCGCGCACCCGAGGTGATCCTATCCTGCTGCGTCCGCTCGACGACCTGGAGTTGACCGTCCGTTCCGGCAAACCTG 2880  
  
 L K A E N I Y Y I G D L I Q R S E N E L L K T P N L  
 2881 CCTGAAGGCCGAGAACATCTACTACATCGGCGACCTGATCCAGCGCAGCGAAAACGAAGTCTGAAGACCCCGAACCTGG 2960  
  
 G R K S L N E I K E V L A S R G L T L G M K L E N W P  
 2961 GTCGCAAGTCCGCTGAACGAAATCAAGGAAGTCTGGCTTCCCCTGGCCTGACCCCTGGGCATGAAGCTGGAAAACCTGGCCG 3040  
  
 P A G L E K \*  
 3041 CCGGCCGCTCTGGAAAAGTAATTCACGTCACCTGGCGGGTGCAGCGCAAGCCGCGCCGCGCAGGCTGGCATGAGGCAAG 3120  
  
 3121 ACCTGAAGTATCCGTCGTACCGTAGCAAAGAAGCTCTTATCCGTCATTTACCGTCCGCGGTCCGGACAGCCTGGCTGGCC 3200  
  
 M R H R H G L R K L H  
 3201 TGCCGATAGAAGAAGTGGATCAAAAAGTGAACCTGAAAGGAATACCATGCGTCAACCGTCAUGGTCTCCGCRAACTGAA 3280



	R T S S H R L A M L R N M T V S L L R H E A I K T T	
3281	TGGTACTTCGTOCCACCGTCTGGCCATGCTGCGCAACATGACCGTGTCCCTGCTGCGTCATGAAGCCATCAAGACCACCC	3360
	L P K A K E L R R V V E F I L T L G K T D T L A N K R	
3361	TGCCGAAGGCCAAGGAACTECGTCGCGTCTGTCGAGCCCATCCTGACCCCTGGGCAAGACCGACACCCCTGGCCCAACGAGCGC	3440
	L A E N R L R D R E I V G K L F S E L G P R Y A A R N	
3441	CTGGCCCTTCAACCGCCTGCGCGACCGGGAAATCGTCGGCAAGCTGTCTCTCGAACTGGGCGCGCGCTACGCTGCCCGTAA	3520
	G G Y L R I L K M G F R Q G D H A P H A F V E L V D	
3521	CGGCGGCTACCTGGCGCATCCTGAAGATGGGCTTCGGTCAAGGGCGACAACGCACCGATGGCCCTTCGTGAGCTGGTGGATC	3600
	R F E V T D A A D A P T A E *	
3601	GTCCGGAAAGTCACCGACGCTGCTGACGCCCCGACCGCTGAGTAACTCTCAOGGACGTGCGCATCACCAAGAAAAAGCCAGGC	3680
3681	<b>CTTGCCTGGCTTTTT</b> TATTTGGCGCGGCAAGTTTTCGCTGATGTCTCTGAGCTTGTATGGCGACGCGCGAGCTCAGGATGTG	3760
3761	GTTGAATACGG	3772

**Legenda:** *rpsM* (363pb) vermelho; *rpsK* (542pb) azul claro; *rpsD* (621pb) laranja; *rpoA* (978pb) verde e o *rpIQ* (396pb) rosa. O provável promotor está destacado em amarelo e sublinhado. Os prováveis sítios de ligação de ribossoma, seqüências de Shine-Dalagarno, estão em negrito. Um provável terminador não dependente de *Rho* está destacado em azul escuro. Acima da seqüência de nucleotídeos encontra-se o produto de tradução, onde os asteriscos representam o códon de término de tradução.

Os genes que codificam para as subunidades do núcleo da RNA polimerase estão organizados em dois operons em *E. coli*; o operon *rpsMKDrpoArpIQ* e o operon *rpIKAJLrpoBC* (Yura & Ishihama, 1979). Squires *et al* (1981) identificaram a região terminadora deste operon e um atenuador de transcrição foi identificado entre os genes *rpIL* e *rpoB* (Morgan, Kellet & Hayward). Dois promotores tipo sigma 70 bastante fortes foram identificados neste operon, uma a montante a *rpIK* e outro a montante a *rpIJ* (Ralling & Linn, 1984). A organização destes operons encontrada em *H. seropedicae* é muito semelhante àquela descrita para *E. coli*. Essa semelhança indica que provavelmente a regulação transcricional e pós-transcricional dos genes que codificam para as subunidades do núcleo da RNA polimerase de *H. seropedicae* sejam parecidas com a descrita para *E. coli*.



Hs LPSRLGMVLD MTLRDIERVL YFEAYVVDTP GMTPLKKQOI MSEDYAAKY EYGG-DDFIA  
 Ec LPSRIGLLLD MFLRDIERVL YFESYVVTEG GMTMLERQOI LTEEQYLDAL EEEG-DEPDA  
 Rs LPSRLGMVLD MTLRDIERVL YFEAFVVVEP GMTPLKKSQI MSEDYLAKE DEYGEGETVA  
 Bj LPSRIGLLLD MTLKDLERIL YFEYVVLEP GLTALKDRQL LSEDEYLKAO DEYQDSFTA  
 Consenso \*\*\*\*:\*:\*:\* \*.\*:\*:\*:\* \*\*\* :\*\* : \*\* \* : \* : :\*:\*\* :\*:\* ..\* \*

Hs FMGAEGIREL LRSIDIDREA ETLRQELKDS KSEAKIKKYA KRLKVLEAFQ RSGIKPEWMI  
 Ec KMGAEAIQAL LKSMDLQEQ EQLREELNET NSETKRRKLT KRKLEAFV QSGNKPEWMI  
 Rs FMGAEGIREL LRGLDIEKQI ETIRAEQAT GSEAKIKKFA KRLKVLEAFQ RSGIKPDWMI  
 Bj MIGAEGIREL LKGMDLKLE ASLRVEMQET DSDIKHKKLA KRLKIVEAFR HSGNKPEWMI  
 Consenso :\*\*\*.\*: \* \*.:\*:\*: :\* \*.: : \* : \* \* : \*\*:\*:\*:\* :\*\* \*\*:\*

Hs MEVLPVLPPE LRPLVPLDGG RFATSDLNLD YRRVINRNR LKRLMELRAP EIITRNEKRM  
 Ec LTVLPVLPPE LRPLVPLDGG RFATSDLNLD YRRVINRNR LKRLDLAAP DIIVRNEKRM  
 Rs LEVLPVLPPE LRPLVPLDGG RFATSDLNLD YRRVINRNR LKRLLEKAP EIIVRNEKRM  
 Bj MTVVPVLPPE LRPLVPLDGG RFATSDLNLD YRRVINRNR LKRLMELRAP DIIITRNEKRM  
 Consenso : \*:\*:\*:\*: \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*:\* \*\* :\*\* \*\*\*\*\*

Hs LQEAVDSLDD NGRRGKAMTG ANKRPLKSLA EMIKGKGRF RQNLGKRVY YSGRSVIVVG  
 Ec LQEAVDALLD NGRRGRAITG SNKRPLKSLA DMIKGKGRF RQNLGKRVY YSGRSVITVG  
 Rs LQEAVDSLDD NGRRGKAMTG ANKRPLKSLA EMIKGKGRF RQNLGKRVY YSGRSVIVVG  
 Bj LQEAVDALED NGRRGKAMTG ANKRPLKSLA DMLKGKGRF RQNLGKRVY YSGRSVIVVG  
 Consenso \*\*\*\*\*:\* \* \*\*\*\*\*:.\* :\*\*\*\*\* :\*:\* \*\* \*\*\*\*\* \*\*\*\*\*.\*

Hs PQLKLHQCGL PKMALELEK PFIENKLELM GLSTTIKAAK KIVEAQEPV WDILEEVIRE  
 Ec PTLRLHQCGL PKMALELEK PFIYKLELR GLATTIKAAK KMVERENAVV WDILDEVIRE  
 Rs PTLKLHQCGL PKMALELEK PFIHKLETM GIATTIKAAK KEVESQTPVW WDILEEVIRE  
 Bj PELRLHQCGL PKMALELEK PFIYSRLDAK GLSTTVKQAK KIVEKERPEV WDILDEVIRE  
 Consenso \* \*:\*\*\*\*\*\* \*\* \*\*\*\*\* \*\*\*: :\* :\*:\*:\* \*\* \* \*\* : . \* \*\*\*\*\*

Hs HFVMLNRAPT LHRIGIQAFE PVLIEGKAIQ LHPLVCAAFN ADFDGDQMAV HVPLSIEAQM  
 Ec HFVLLNRAPT LHRIGIQAFE PVLIEGKAIQ LHPLVCAAYN ADFDGDQMAV HVPLTLEAQL  
 Rs HFVMLNRAPT LHRIGIQAFE PVLIEGKAIQ LHPLVCAAFN ADFDGDQMAV HVPLSIEAQM  
 Bj HFVLLNRAPT LHRIGIQAFE PVLIEGKAIQ LHPLVCAAFN ADFDGDQMAV HVPLSIEAQL  
 Consenso \*\*\*:\*\*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*:\* \*\*\*\*\* \*\*\*\*\*:.\*:

Hs EARTLMLASN NILFPSPGEP SIVPSQDIVL GLYYATREAI NGKMEGMMFQ DVSEVIRAYD  
 Ec EARALMMSTN NILSPANGEP IIVPSQDVVL GLYYMTRDCV NAKGEGMVLV GPKEAERLYR  
 Rs EARTLMLASN NVLFPANGEP SIVPSQDVVL GLYYTTRDKI NGRGEGMTER DISEVIRAYE  
 Bj EARVLMSTN NILHEANGQP IIVPSQDIVL GLYYVSINKE GLPGEKIFG DMAELEHALH  
 Consenso \*\*\*.\*:\*:\* \*.\* \*:\*:\* \*\*\*\*\*:\* \*\*\*\*\* : . \*\* : . \* :

Hs NKEVELATRI TVRIVEH--- -PKDPVTGEF VRTVKRYETT IGRAILSEIL SKG--LPEFSV  
 Ec SGLASLHARV KVKITEV--- --EKDANGEL XAKTSLXDTT VGRAILMIV PKG--LEPTI  
 Rs NKEVELASKV NVRITEYDLV NPEADGDARF APKITLQATT VGRAILSEIL PKG--LPEFSV  
 Bj AKVIHLHTKI KYRWQGM--- ----DETQKV S--TRWIETT AGRVMLGNLL FKNPRISYEI  
 Consenso \* : : . \* ... \*\* \*\*:\* : : . \* : :

Hs LNRALKKKEI SKLINTSERK CGLRATVVEA DQLMQSGFRL ATRAGISICV DDMLVPPQKA  
 Ec VNQALGKKAI SKMLNTCYRI LGLKPTVIFA DQIMYTGFAV AARSGASVGI DDMVIPEKIH  
 Rs LNKPLKKEI SRLINTAFK CGLRETVIFA DKLLQSGFRL ATRAGISIAI DDMLVPPAKE





```

Hs      TRDITGGLPR VAELEEARSE KDAAGMLAEVT GTVAFGKETK GKQRLEITDM EG--TKHEFL
Ec      TKDITGGLPR VADLEEARRE KEPAILAEIS GIVSFGKETK GKRRLVITEV DGS-DFYEEM
Rs      TRDITGGLPR VAELEEARSE KDAAVLAEVT GTTSFGKDTK GKQRLVITDL DG--NAHEFL
Bj      TRDITGGLPR VAELEEARRE KDAAILAEIA GTIRFGRDYK NKRRISIEEM DMTDEPREYL
Consenso *:***** **:***** * *...:~*~: * **~: * .*~: * : : * :

Hs      ITKDKQVLVH DGQVVKNGEM IVDGPADEPD ILRLLGIEAL ARYIVDEVQD VYRLQGVKIN
Ec      IPKWRQLNVF EGERVERGDV ISDGPEAPHD ILRLRGVHAV TRYIVNEVQD VYRLQGVKIN
Rs      IAKEKQVLVH DGQVVKNGEM IVEGPADEPD ILRLKGV EEL ATYIVDEVQD VYRLQGVKIN
Bj      IPKKGKHIHLQ DGDVVEKGDV IVEGNPAPHD ILAVKGIEEL AAYLVNEIQE VYRLQGVLIN
Consenso *.~ :~ : :~: ~:~:~. *~*~ *~*~ ** :~*~ :~ :~ :~:~:~:~ ***** **

Hs      DKHIEVIVRQ MLRRVQVVEP GDTNYITGEQ VERSELLEN DRVTAEGKIP ATYENVLLGI
Ec      DKHIEVIVRQ MLRKATIVNA GSSDFLEGEQ VEYSRVKIAN RELEANGKVG ATYSRDLLEGI
Rs      DKHIEVIVRQ MLRRVQIVDV GDTKFIPEGEQ VERSELLEN DKVIAEGKRP ATYENLLLEGI
Bj      DKHIEVIVRQ MLQKVEVTDQ GDTDMISGEQ VDKIEFDALN EKAKEEGKPI ATGTPVLLGI
Consenso *****~ **~:~ :~:~ *~:~ :~*~ *~ :~ .. * . :~*~ ** *****

Hs      TKASLSTDSF ISAASFQETT RVLTEAAIMG KKDGLRGLKE NVIVGRLIPA GTGLAFHRAR
Ec      TKASLATESF ISAASFQETT RVLTEAAVAG KRDELRLGLKE NVIVGRLIPA GTGYAYHQDR
Rs      TKASLSTDSF ISAASFQETT RVLTEAAIMG KVODLRGLKE NVIVGRLIPA GTGLAYHRAR
Bj      TKASLQTRSF ISAASFQETT RVLTEAAVNG KVDPLEGLKE NVIVGRLIPA GTGASHAKIR
Consenso ***** * ** :*****~ *****: * * * *.***** *****~ *** : : *

Hs      KEKDAWEAEE RAALLEAERA SRSEFGDEEI SSTESIGGGE GEA
Ec      MRRRAAGEAP AAFQVTAEDA SAS----LAE LLNAGLGGSD NE-
Rs      KAKEAADRDR AAAIAEBEAA S-----I FETPAVQQ-E GDA
Bj      EVAMKRDKLI LDEREKQAAV VSP----APE AELPALPPAE ---
Consenso . . . . . : : :

```

**Legenda:** Aminoácidos idênticos em todas as cinco seqüências estão indicados por asteriscos (\*), substituições conservativas estão indicadas por dois pontos (:), e substituições semi-conservativas são indicadas por um ponto (.). As falhas estão indicadas por (-).

A RNA polimerase de *E. coli* foi purificada e manteve-se ativa após a purificação utilizando como estratégia a inserção de 6 resíduos de histidina na região C-terminal da subunidade  $\beta'$  (Kashlev *et al.*, 1993). A análise da seqüência de aminoácidos da subunidade  $\beta'$  de *E. coli* e *H. seropedicae* demonstrou uma alta identidade entre essas proteínas (64,8%), indicando que a inserção dos 6 resíduos de histidina na região C-terminal provavelmente não irá provocar alterações na montagem da RNA polimerase e em sua atividade.

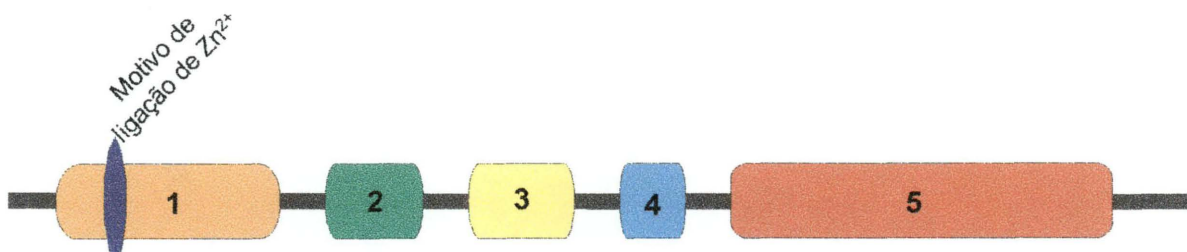
Essa estratégia deverá permitir a recuperação de proteínas que interagem com a subunidade  $\beta'$ , tais como as subunidades  $\beta$ ,  $\alpha$  e  $\omega$  da RNA polimerase, fatores  $\sigma$  e outros fatores de transcrição. Essa metodologia poderá

ser aplicada para identificar interação proteína-proteína em *H. seropedicae*, substituindo a utilização de plasmídios.

#### 4.3 Identificação dos domínios da subunidade $\beta'$ da RNA polimerase de *H. seropedicae*

Para confirmar que o gene *rpoC* de *H. seropedicae* identificado neste trabalho codifica para a subunidade  $\beta'$  da RNA polimerase uma análise de domínios estruturais foi realizada *in silico*. A combinação dos resultados obtidos pelo programa SMART e pelo programa Pfam permitiram identificar os 5 domínios característicos da subunidade  $\beta'$  da RNA polimerase (Severinov *et al*, 1996) e o motivo de coordenação de íons  $Zn^{2+}$  (figura 7). A organização observada é idêntica àquela descrita para *T. aquaticus* (código de entrada SwissPfam Q8RQE8).

Figura 7. Domínios estruturais da subunidade  $\beta'$  da RNA polimerase de *H. seropedicae*.



**Legenda:** Domínio 1, representa uma estrutura em braçadeira (laranjado); Domínio 2, parte do sítio catalítico (verde); Domínio 3, forma um poro (amarelo); Domínio 4, forma um funil (azul) e Domínio 5, parte da fenda onde se encaixa a dupla fita de DNA (vermelho). O motivo para coordenação de íons zinco encontra-se no domínio 1 (roxo).

O domínio 1, resíduo de aminoácido 14 ao 342, apresenta uma estrutura em forma de braçadeira. Esse domínio é uma região móvel da proteína envolvido no posicionamento do DNA, na manutenção da estabilidade da bolha de transcrição e no posicionamento da cadeia de RNA nascente. Este domínio contém ainda o motivo de coordenação de íons  $Zn^{2+}$ , que possui a seqüência consenso C-X(1-6)-H-X-C-X3-C(H/C)-X(3-4)-(H/C)-X(1-10)-C.

O domínio 2, resíduo de aminoácido 344 ao 486, corresponde ao domínio que contém o sítio ativo, apresenta o motivo invariável NADFDGD, que é o sítio de coordenação do íon  $Mg^{2+}$ . O domínio 3 (resíduo 489 a 645) é denominado domínio poro. É por este poro que os NTPs entram no sítio catalítico da enzima e é por onde a extremidade 3' do RNA nascente é colocada para fora durante o retorno da RNA polimerase na transcrição.

O domínio 4 (resíduo 670 a 765), forma uma estrutura em funil onde encontram-se sítios de ligação para fatores de alongamento de transcrição. Finalmente, o domínio 5 (resíduo 767 a 1323) forma uma fenda descontínua que faz parte do canal central do oligômero onde se liga a fita de DNA molde.

Estes resultados mostraram a extremidade C-terminal da subunidade  $\beta'$  da RNA polimerase de *H. seropedicae* que provavelmente não contém um domínio importante para a manutenção de sua estrutura e/ou função. Isto reforça a premissa deste projeto de que a inserção de 6 resíduos de histidina nesta região não irá afetar a montagem do oligômero ou a atividade da RNA polimerase de *H. seropedicae*.

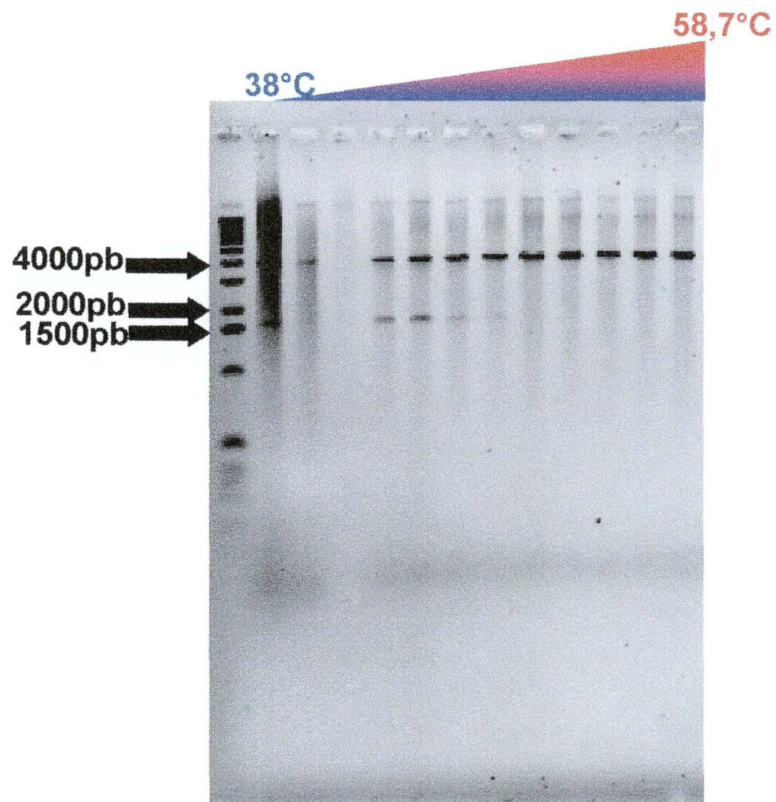
#### 4.4 Amplificação do gene *rpoC*

##### 4.4.1 Determinação da temperatura de anelamento ideal para amplificação do fragmento de interesse.

Para determinar a temperatura ativa de anelamento dos iniciadores ao DNA molde um experimento de PCR foi realizado. Neste ensaio 12 tubos idênticos foram preparados, de acordo com o descrito em Materiais e Métodos, e carregados em termociclador Eppendorf Motorcycle. Este equipamento permite uma variação de temperatura entre as 12 colunas do bloco de tubos. A menor temperatura de anelamento utilizada foi de 38°C e a maior temperatura de anelamento foi de 58,7°C. A temperatura de fusão calculada ( $T_m$ ) para os iniciadores foi de 54°C para o iniciador *rpoCN* e 50°C para o iniciador *rpoCC*. Dois produtos de amplificação, 1800pb e 4200pb, foram obtidos nas temperaturas entre 39,3°C e 51,5°C. A partir de 54°C apenas o produto de 4200pb foi obtido (figura 8)



Figura 8. Eletroforese em gel de agar 1% dos produtos de Amplificação dos iniciadores *rpoCN* e *rpoCC*.



**Legenda:** Linha 1, marcador de massa molecular 1Kb ladder; linhas 2 à 13 produtos de amplificação com gradiente de temperatura de anelamento; o volume todo de reação foi aplicado.

#### 4.4.2 Amplificação do gene *rpoC* de *H. seropedicae* para clonagem

Para amplificar o gene *rpoC* de *H. seropedicae* foi utilizada a temperatura de anelamento de 58,7°C. Uma única banda de interesse, 4200pb, foi obtida (Figura 9). Nesta reação foram utilizadas duas DNA polimerase, *Taq* DNA polimerase e *Pfx* DNA polimerase na proporção de 10:1 (Cheng *et al*, 1994).

Figura 9. Eletroforese em gel de agarose 1% do produto de amplificação com os iniciadores *rpoCN* e *rpoCC*, temperatura de anelamento de 58,7°C, *Pfx* DNA polimerase e *Taq* DNA polimerase.



**Legenda:** Linha 1, marcador de massa molecular 1Kb ladder; linha 2 produtos de amplificação com os iniciadores *rpoCN* e *rpoCC* e temperatura de anelamento de 58,7°C; foram aplicados 2 $\mu$ L da mistura de reação.

A *Pfx* é uma DNA polimerase extraída de *Pirococcus* sp. e possui atividade corretora de erros (exonuclease 3'-5') portanto, a taxa de erro de incorporação de base no DNA produzido é menor ( $10^{-6}$ ) do que a da *Taq* DNA polimerase, não possui atividade corretora. Uma baixa taxa de erro é essencial para produtos de amplificação longos como o gene *rpoC* de *H. seropedicae*. Além disso, a *Pfx* remove o nucleotídeo A adicionado independente de molde pela *Taq* DNA polimerase. O produto de amplificação obtido com a mistura das duas DNA polimerases possui pontas cegas.

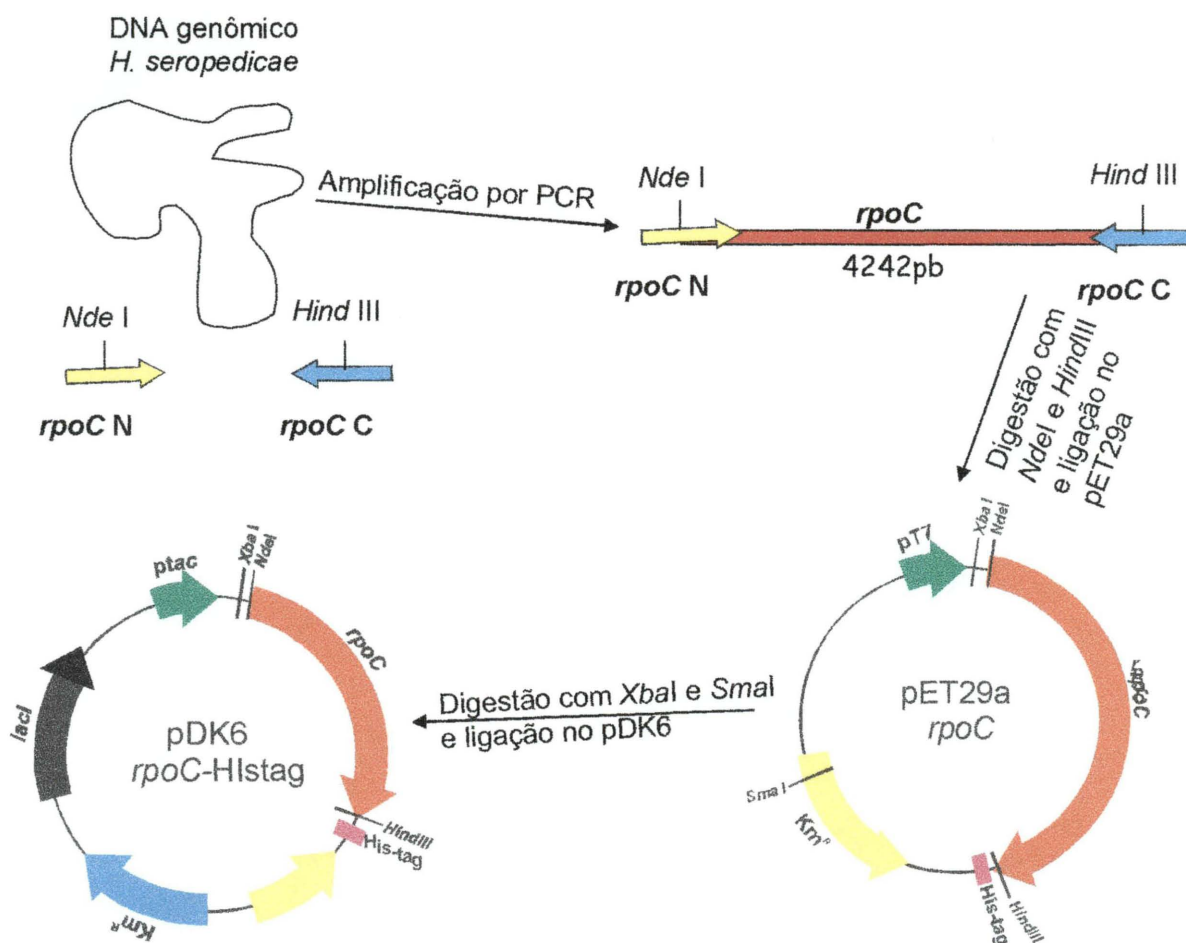
O fragmento obtido, 4,2Kb, foi digerido com as enzimas *NdeI* e *HindIII*, cujos sítios foram introduzidos nos oligonucleotídeos iniciadores e será clonado no vetor pET29a.

#### 4.5 Estratégias de clonagem do gene *rpoC* de *H. seropedicae*.

O produto de amplificação (4200pb) obtido com os iniciadores *rpoCN* e *rpoCC* foi digerido com as enzimas *NdeI* e *HindIII* para clonagem no vetor

pET29a, digerido com as mesmas enzimas. O fragmento *Xba*I/*Sma*I do plasmídio recombinante obtido (gene *rpoC* com fusão de uma cauda que codifica para 6 resíduos de histidina na região C-terminal) deverá ser subclonado no vetor pDK6 digerido com a enzima *Hind*III, tratado com T4 DNA polimerase e digerido com a enzima *Xba*I. O vetor pDK6 foi escolhido por possuir o promotor *tac*, que é funcional em *H. seropedicae*. O esquema de clonagem está mostrado na figura 10.

Figura 10. Esquema de clonagem do gene *rpoC* de *H. seropedicae*



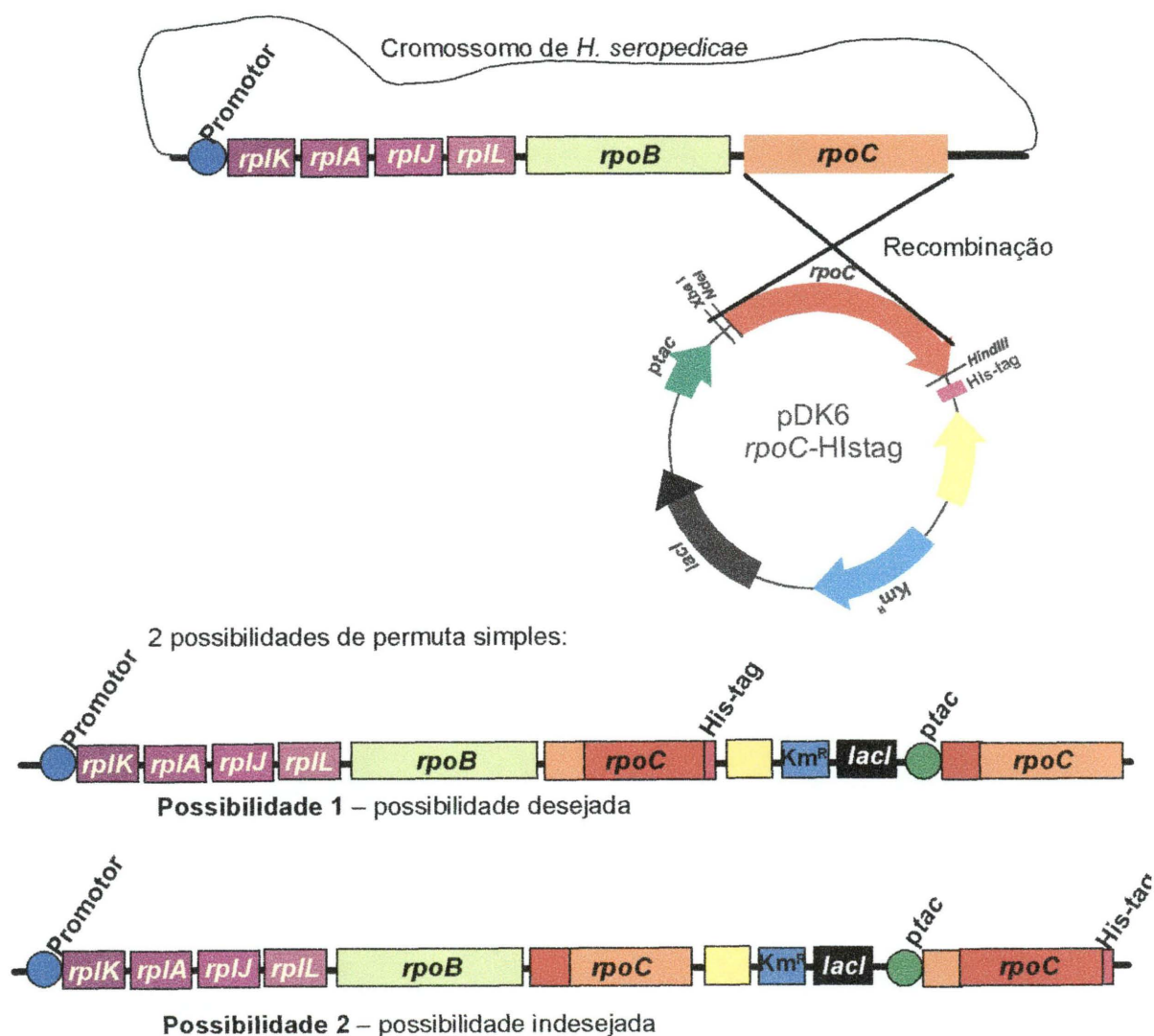
#### 4.6 Estratégia de expressão do gene *rpoC* em *H. seropedicae*.

O plasmídeo contendo o gene *rpoC*-His-tag sob a influência do promotor *tac* (pDK6 – *rpoC*-His-tag) deverá ser transformado em *H. seropedicae* afim de se obter recombinação homóloga do tipo simples permuta (“crossing-over” simples). Os recombinantes deverão ser selecionados pela resistência ao



antibiótico canamicina e analisados por hibridização. O mutante desejado será aquele que possuir o gene *rpoC* fundido com cauda de 6 resíduos de histidina expresso a partir do promotor do operon *rplKAJLrpoBC* nativo de *H. seropedicae*. A superexpressão da subunidade  $\beta'$  da RNA polimerase não é interessante, pois poderia inibir a expressão das outras subunidades, portanto, produziria uma estirpe inviável. O esquema de transformação e recombinação está indicado na figura 11.

Figura 11. Esquema de recombinação do gene *rpoC* de *H. seropedicae*.



## 5 CONCLUSÕES

Os genes que codificam para as subunidades do núcleo da RNA polimerase de *H. seropedicae* estão organizados em dois operons, o operon *rpsMKDrpoArpIQ* e o operon *rplKAJLrpoBC*. Foram identificadas as prováveis regiões regulatórias desses operons; promotores, atenuadores e terminadores de transcrição.

As análises da seqüência de aminoácidos da subunidade  $\beta'$  da RNA polimerase revelaram uma grande similaridade das seqüências de proteínas homólogas de outros organismos, indicando que essa proteína é bastante conservada.

A análise de domínios permitiu a identificação de todos os domínios característicos da subunidade  $\beta'$  da RNA polimerase, reforçando que o gene *rpoC* identificado neste trabalho codifica para esta proteína de *H. seropedicae*.

Um fragmento de PCR de 4200pb, correspondente ao gene *rpoC* de *H. seropedicae* foi obtido e está sendo clonado no vetor pET29a. As condições ótimas para amplificação deste fragmento foram determinadas. O gene *rpoC* de *H. seropedicae* foi amplificado com uma mistura de duas DNA polimerase, uma com atividade corretora de erros (*Pfx* DNA polimerase) e outra sem essa atividade (*Taq* DNA polimerase) para minimizar a taxa de erros e permitir a produção de pontas cegas no fragmento amplificado.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

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

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

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

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

BORUKHOV, S. & NUDLER, E. RNA polymerase holoenzyme: structure, function and biological implications. *Curr. Op. Microbiol.* v.6, p.93-100, 2003.

BORUKHOV, S.; SEVERINOV, K. Role of the RNA Polymerase sigma subunit in transcription initiation. *Res. Microbiol.* v.153, p.557-562, 2002.

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

CARLOMAGNO, M.S. & NAPPO, A. *NusA* modulates intergenic termination by different pathways. *Gene*, v.308, p.115-128, 2003.

CHEETHAM, G.M.T.; JERUZALMI, D.; STEITZ, T.A.; Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature*, v.399, p.80-83, 1999.

CHENG, S.; FOCKLER, C.; BARNES, W. e HIGUCHI, R. Effective amplification of long targets from cloned inserts and human genomic DNA. **Proc. Natl. Acad. Sci.** v.91, p.5695-5699, 1994.

CRAMER, P. Multisubunit RNA polymerase. **Curr. Op. Struc. Biol.** v.12, p.89-97, 2002.

DARST, S. A. Bacterial RNA polymerase. **Curr. Op. Struc. Biol.** v.11, p.155-162, 2001.

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

DWORKIN, J. & LOSICK, R. Does RNA polymerase help drive chromosome segregation in bacteria? **Proc. Nat. Acad. Sci. USA** v. 99, p.14089-94, 2002.

DYKXHOORN, D.M.; PIERRE, R.S.; LINN, T. Synthesis of the  $\beta$  and  $\beta'$  subunits of *Escherichia coli* RNA polymerase is autogenously regulated in vivo by both transcriptional and translational mechanisms. **Mol. Microbiol.** v.19, p.483-493, 1996.

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

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

FUKUDA, R.; TAKETO, M.; ISHIHAMA, A. Autogenous regulation of RNA polymerase  $\beta$  subunit synthesis in vitro. **J. Biol. Chem.** v.253, p.4501-4504, 1978.

GHOSH, P.; RAMAKRISHNAN, C. & CHATTERJI, D. Inter-subunit recognition and manifestation of segmental mobility in *Escherichia coli* RNA polymerase: a case study with  $\omega$ - $\beta'$  interaction. **Biophys. Chem.** v.102, p. 223-237, 2003.

GORDON, D.; ABAJIAN, C. & GREEN, P. Consed: A graphical tool for sequence finishing. **Gen. Res.** v.8, p.195-202, 1998.

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

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

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

ISHIHAMA, A. Promoter selectivity of prokaryotic RNA polymerases. **Trends Genet.** v.4, p.282-86, 1988.

ISHIKAWA, J. & HOTTA, K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. **FEMS Microbiol. Lett.**, v.174(2), p.251-253, 1999.

KASHLEV, M.; MARTIN, E.; POLYAKOV, A.; SEVERINOV, K.; NIKIFOROV, V. e GOLDFARB, A. Histidine-tagged RNA polymerase: dissection of the transcription cycle using immobilized enzyme. **Gene**, v.130, p.9-14, 1993.

KLASSEN, G.; PEDROSA, F.O.; SOUZA, E.M.; FUNAYAMA, S.; RIGO, L.U. Effect of nitrogen compounds on nitrogenase activity in *Herbaspirillum seropedicae*. **Can. J. Microbiol.** v.43, p.887-891, 1997.

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

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

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

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

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

MONTEIRO, R.A.; PEDROSA, F. O.; STEFFENS, M. B. R.; CHUBATSU, L. S. Isolation of Recombinant Plasmids for Rapid Analysis Using a Sodium Dodecyl Sulfate/Potassium Chloride Precipitation. **Anal. Biochem.** v.292, p.169-170, 2001.

MORGAN, B.A.; KELLETT, E.; HAYWARD, R.S. The wild-type nucleotide sequence of the *rpoBC*-attenuator region of *Escherichia coli* DNA, and the implications for the nature of the *rif<sup>d</sup>18* mutation. **Nuc. Ac. Res.** v.12, p.5465-5470, 1984.

MURUKAMI, K.S. & DARST, S.A. Bacterial RNA polymerases: the whole story **Curr. Op. Struc. Biol.** v.13, p.31-39, 2003.

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

NUDLER, E.; MUSTAEV, A.; LUKHTANOV, E.; GOLDFARB, A. The RNA-DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell*, v.89, p.33-41, 1997.

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

PASSADOR, L.; LINN, T. An Internal Region of *rpoB* is Required for Autogenous Translation Regulation of the  $\beta$  Subunit of *Escherichia coli* RNA Polymerase. *J. Bacteriol.* v.174, p.7174-7179, 1992.

PASSADOR, L.; LINN, T. Autogenous Regulation of the RNA Polymerase  $\beta$  Subunit of *Escherichia coli* Occurs at the Translational Level In Vivo. *J. Bacteriol.* v.171, p.6234-6242, 1989.

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

RALLING, G.; LINN, T. Relative Activities of the Transcriptional Regulatory Sites in the *rplKAJLrpoBC* Gene Cluster of *Escherichia coli*. *J. Bacteriol.* v.158, p.279-285, 1984.

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

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



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

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

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

SEVERINOV, K.; MUSTAEV, A.; KUKARIN, A.; MUZZIN, O.; BASS, I.; DARST, S.A.; GOLDFARB, A. Structural Modules of the Large Subunits of RNA Polymerase. **J. Biol. Chem.** v.271, p.27969-27974, 1996.

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

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

SQUIRES, C.; KRAINER, A.; BARRY, G.; SHEN, W.F.; SQUIRES, C.L. Nucleotide sequence at the end of the gene for the RNA polymerase  $\beta'$  subunit (*rpoC*). **Nuc. Acid Res.**, v.9, p.6827-6840, 1981.

STEWART, K.L.; PIERRE, R.S.; LINN, T. Transcription-frequency-dependent modulation of an attenuator in a ribosomal protein-RNA polymerase operon requires an upstream site. **Microbiology**, v.143, p.3501-3511, 1997.

VO, N.V.; HSU, L.; KANE, C.M.; CHAMBERLIN, M.J. In Vitro Studies of Transcripts Initiation by *Escherichia coli* RNA Polymerase. 2. Formation and Characterization of Two Distinct Classes of Initial Transcribing Complexes. **Biochem.**, v.42, p.3787-3797, 2003a.

VO, N.V.; HSU, L.; KANE, C.M.; CHAMBERLIN, M.J. In Vitro Studies of Transcripts Initiation by *Escherichia coli* RNA Polymerase. 3. Influences of Individual DNA Elements within the Promoter Recognition Region on Abortive Initiation and Promoter Escape. **Biochem.**, v.42, p.3798-3811, 2003b.

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

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

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

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

YURA, T. & ISHIHAMA, A. Genetics of Bacterial RNA polymerases. **Ann. Rev. Genet.** v.13, p.59-97, 1979.