



**COPPE/UFRJ**

SELEÇÃO, CLONAGEM E EXPRESSÃO DE UMA PROTEÍNA ASSOCIADA À  
VIRULÊNCIA DE *Streptococcus pneumoniae*

Karen Einsfeldt

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Química, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Química.

Orientadores: Tito Lívio Moitinho Alves

Ariane Leites Larentis

Rodrigo Volcan Almeida

Rio de Janeiro

Março de 2010

SELEÇÃO, CLONAGEM E EXPRESSÃO DE UMA PROTEÍNA ASSOCIADA À  
VIRULÊNCIA DE *Streptococcus pneumoniae*

Karen Einsfeldt

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO  
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA  
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE  
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE  
EM CIÊNCIAS EM ENGENHARIA QUÍMICA.

Examinada por:

---

Prof. Tito Lívio Moitinho Alves, D.Sc.

---

Dr<sup>a</sup>. Ariane Leites Larentis, D.Sc.

---

Prof<sup>a</sup>. Leda dos Reis Castilho, Dr.Ing.

---

Prof<sup>a</sup>. Bianca Cruz Neves, Ph.D.

---

Dr. Marco Alberto Medeiros, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2010

Einsfeldt, Karen

Seleção, Clonagem e Expressão de uma Proteína  
Associada à Virulência de *Streptococcus pneumoniae*/  
Karen Einsfeldt. – Rio de Janeiro: UFRJ/COPPE, 2010.

XXIII, 127 p.: il.; 29,7 cm.

Orientadores: Tito Lívio Moitinho Alves

Ariane Leites Larentis

Rodrigo Volcan Almeida

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de  
Engenharia Química, 2010.

Referencias Bibliográficas: p. 118-127.

1. Expressão de proteína recombinante. 2. ClpP 3.  
*Escherichia coli*. 4. *Bacillus subtilis*. 5. Planejamento de  
Experimentos. 6. Estabilidade Plasmidial. I. Alves, Tito  
Lívio Moitinho *et al.* II. Universidade Federal do Rio de  
Janeiro, COPPE, Programa de Engenharia Química. III.  
Titulo.

Aos meus pais

“A mente que se abre a uma nova idéia jamais voltará ao seu tamanho original”

(Albert Einstein)

## AGRADECIMENTOS

Agradeço aos meus orientadores, Tito, Ariane e Rodrigo, pelos ensinamentos, discussões e principalmente pela inspiração e motivação.

Ao Prof. Tito Lívio Moitinho Alves e ao Programa de Engenharia Química/COPPE pela formação acadêmica, estrutura disponibilizada, e suporte para a realização deste trabalho.

A D.Sc. Ariane Leites Larentis pela dedicação, atenção e ensinamentos neste período do mestrado. Ao D.Sc. Marco Alberto Medeiros por ter disponibilizado a estrutura do Laboratório de Tecnologia Recombinante (LATER)/BioManguinhos/Fiocruz e fornecido todo o suporte para a realização de parte deste trabalho. Agradeço também ao pessoal do LATER pelo auxílio neste trabalho, em especial, a Ana Paula, pelos ensinamentos, e a Júlia, pela ajuda em vários experimentos.

Ao Prof. Rodrigo Volcan Almeida pelos ensinamentos, atenção e por ter disponibilizado o laboratório e todo o suporte para a realização deste trabalho. Ao pessoal do LAMMP pelas conversas e pelo auxílio em parte deste trabalho.

Ao D.Sc. Leon Rabinovitch e a M.Sc. Sônia E. Alves da Silva pelo treinamento fornecido em *B. subtilis* e por terem aberto as portas do Laboratório de Fisiologia Bacteriana e Coleção de Culturas do Gênero Bacillus e Gêneros Correlatos.

Ao Laboratório de Produtos Biológicos – INCQS/Fiocruz, por disponibilizar os equipamentos e o espaço para realização das análises de densitometria. Ao Laboratório de AIDS & Imunologia Molecular – IOC/Fiocruz por disponibilizar o uso do seqüenciador e ao Alfredo Jabor pelo auxílio nas análises do seqüenciamento.

Aos grandes amigos que fiz aqui no Rio de Janeiro, em especial ao Aldo, Elis, Felipe, Gisele, Guillermo e Lívia, pelas conversas, risadas, festas, emails engraçados, e pelo apoio nos momentos difíceis no mestrado e no Rio. Ao Guillermo “boludo”, pela preciosa ajuda nos experimentos finais deste trabalho.

A minha família que de alguma forma sempre esteve presente. Aos meus pais por toda a força, dedicação e pelo incentivo para que eu nunca desista dos meus sonhos, por mais difíceis que estes possam parecer. A mudança para o Rio de Janeiro foi difícil não só para mim, mas para eles também, por isso muito obrigada por tudo, sei que todo o esforço será recompensado.

Em especial ao meu namorado, minha família no Rio de Janeiro, que me deu forças para continuar meus estudos, que sempre me encorajou a “querer mais” e a “fazer melhor” e que sempre esteve ao meu lado, suportando meu nervosismo, mau humor e ansiedade principalmente na parte final deste trabalho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

SELEÇÃO, CLONAGEM E EXPRESSÃO DE UMA PROTEÍNA ASSOCIADA À  
VIRULÊNCIA DE *Streptococcus pneumoniae*

Karen Einsfeldt

Março/2010

Orientadores: Tito Lívio Moitinho Alves

Ariane Leites Larentis

Rodrigo Volcan Almeida

Programa: Engenharia Química

Infecções causadas por *S. pneumoniae* são uma das principais causas de morte no mundo e o desenvolvimento de uma vacina composta por proteínas é uma alternativa para resolver este problema. Neste trabalho objetivou-se selecionar, clonar e expressar, utilizando *E. coli* e *B. subtilis* como sistemas de expressão, um fator de virulência de *S. pneumoniae*, sorotipo 14, com alta prevalência e conservação entre os sorotipos, bem como otimizar a expressão da proteína empregando planejamento experimental e estudar a segregação plasmidial do sistema. A proteína selecionada, ClpP, foi clonada em *E. coli* e sua expressão foi otimizada através de planejamento fatorial 2<sup>2</sup>, variando a concentração de indutor e de canamicina e analisando, como respostas, concentração de proteína heteróloga, crescimento celular e fração de células com plasmídeo. Em todos os experimentos, a proteína ClpP foi expressa em concentrações similares e de forma solúvel, com média no ponto central de 240,4 mg/L. Pode-se concluir que é possível diminuir em 10 vezes a concentração de indutor e retirar o antibiótico do sistema, reduzindo custos e mantendo a expressão em níveis similares. O sistema apresentou segregação plasmidial mesmo nas concentrações habituais de antibiótico. Não foram obtidas células recombinantes utilizando *B. subtilis* como hospedeiro, provavelmente devido à instabilidade do vetor, reforçando a ideia que as ferramentas para a utilização deste sistema de expressão precisam ser melhor estudadas e desenvolvidas.



Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SELECTION, CLONING AND EXPRESSION OF A PROTEIN ASSOCIATED  
WITH VIRULENCE OF *Streptococcus pneumoniae*

Karen Einsfeldt

March/2010

Advisors: Tito Lívio Moitinho Alves

Ariane Leites Larentis

Rodrigo Volcan Almeida

Department: Chemical Engineering

Infections caused by *S. pneumoniae* are one of the main causes of death in the world and the development of a vaccine composed of proteins is an alternative to tackle this problem. This work aimed to select, clone and express, using *E. coli* and *B. subtilis* as expression systems, a virulence factor of *S. pneumoniae*, serotype 14, with high prevalence and conservation among the serotypes, as well to optimize the protein expression using experimental design and to study the plasmid segregation of the system. The selected protein, ClpP, was cloned in *E. coli* and its expression was optimized by a 2<sup>2</sup> factorial design, varying the inducer and kanamycin concentrations and evaluating as responses the concentration of heterologous protein, cell growth and fraction of cells with plasmid. In all experiments, the ClpP protein was expressed at similar concentrations and in soluble form, with an average concentration of 240,4mg/L in the central point. It can be concluded that it is possible to decrease 10 times the concentration of inducer and remove the antibiotic from the system, reducing costs and keeping the expression at similar levels. The system showed plasmid segregation even in normal concentrations of antibiotics. When *B. subtilis* was used as host, no recombinant cells were obtained, probably due to vector instability, reinforcing the idea that the tools for using this expression system must be better studied and developed.

## SUMÁRIO

<b>RESUMO</b>	viii
<b>ABSTRACT</b>	ix
<b>LISTA DE FIGURAS</b>	xiii
<b>LISTA DE TABELAS</b>	xx
<b>LISTA DE ABREVIACÕES</b>	xxii
<b>1 INTRODUÇÃO</b>	1
<b>2 REVISÃO DA LITERATURA</b>	5
<b>2.1 Fatores de virulência</b>	5
2.1.1 Cápsula polissacarídica	5
2.1.2 Proteínas	6
2.1.2.1 <i>ClpP</i>	9
2.1.2.2 <i>IgA1 protease</i>	12
2.1.2.3 <i>Hialuronato liase (Hyl)</i>	13
<b>2.2 Sistemas de expressão de proteínas recombinantes</b>	16
2.2.1 <i>Escherichia coli</i> como sistema de expressão de proteínas recombinantes	18
2.2.2 <i>Bacillus subtilis</i> como sistema de expressão de proteínas recombinantes	21
2.2.3 Estabilidade plasmidial	23
<b>2.3 Planejamento de experimentos em sistemas recombinantes</b>	25
<b>3 METODOLOGIA</b>	27
<b>3.1 Seleção da proteína</b>	27
<b>3.2 Clonagem e expressão do gene <i>clpP</i> de <i>S. pneumoniae</i> em <i>Escherichia coli</i></b>	28

3.2.1 Extração de DNA genômico de <i>Streptococcus pneumoniae</i>	28
3.2.2 Amplificação do gene através da técnica de PCR	30
3.2.2.1 Oligonucleotídeos iniciadores	30
3.2.2.2 Reação de PCR	32
3.2.2.3 Purificação do gene <i>clpP</i> amplificado	33
3.2.3 Clonagem	34
3.2.4 Preparação de células eletrocompetentes	37
3.2.5 Transformação de <i>Escherichia coli</i>	38
3.2.6 Seleção de clones e preparação de estoques em glicerol	39
3.2.7 Extração de DNA plasmidial	40
3.2.8 Confirmação dos clones	41
3.2.8.1 Digestão com enzimas de restrição	41
3.2.8.2 Seqüenciamento	42
3.2.9 Preparação de lote de trabalho e viabilidade celular	42
3.2.10 Crescimento celular e curva de massa seca de células	44
3.2.11 Expressão e preparação dos extratos protéicos para análise de solubilidade	44
3.2.12 Otimização da expressão por planejamento de experimentos	46
3.2.13 Avaliação do crescimento celular e quantificação da proteína expressa em SDS-PAGE	48
3.2.14 Análise da segregação plasmidial	49
3.2.15 Validação da condição otimizada no planejamento fatorial	50
3.2.16 Purificação	50
<b>3.3 Clonagem utilizando <i>Bacillus subtilis</i> como sistema de expressão</b>	51
3.3.1 Caracterização da cepa de <i>Bacillus subtilis</i> WB600	51
3.3.2 Desenho do gene <i>clpP</i> sintético e vetores utilizados	52
3.3.3 Preparo de células competentes para transformação por choque térmico	54
3.3.4 Clonagem	55
3.3.5 Transformação por choque térmico	57

<b>4 RESULTADOS E DISCUSSÃO</b>	58
<b>4.1 Seleção da proteína</b>	58
4.1.1 Comparação das proteínas: ClpP, IgA1 protease e Hialuronato liase	66
<b>4.2 Amplificação do gene <i>clpP</i> de <i>S. pneumoniae</i></b>	67
4.2.1 Extração de DNA genômico de <i>S. pneumoniae</i>	67
4.2.2 Amplificação do gene <i>clpP</i> por PCR (Reação em Cadeia da Polimerase)	68
<b>4.3 Clonagem e expressão utilizando <i>Escherichia coli</i> como sistema de expressão</b>	69
4.3.2 Clonagem utilizando o vetor pET28b	69
4.3.3 Seqüenciamento	71
4.3.4 Expressão da proteína recombinante ClpP	72
4.3.5 Otimização da expressão e estudos de segregação plasmidial	74
4.3.5.1 Validação da condição otimizada pelo planejamento de experimentos - Estudo da cinética de produção da proteína e segregação plasmidial	82
4.3.6 Purificação	90
<b>4.4 Clonagem e expressão utilizando <i>Bacillus subtilis</i> como sistema de expressão</b>	93
4.4.1 Caracterização da cepa <i>Bacillus subtilis</i> WB600	93
4.4.2 Clonagem	94
<b>5 CONCLUSÕES</b>	98
<b>APÊNDICE A</b>	101
<b>APÊNDICE B</b>	103
<b>APÊNDICE C</b>	114
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	118

## LISTA DE FIGURAS

- Figura 2.1** Representação do monômero de ClpP de *S. pneumoniae*. Adaptado de GRIBUN *et al.* (2005) 9
- Figura 2.2** (A) Estrutura da proteína ClpP de *S. pneumoniae* depositada no PDB (*Protein Data Bank*) sob o número de identificação 1Y7O (GRIBUN *et al.*, 2005). (B) Representação esquemática da estrutura da molécula tetradecamérica de ClpP de *Helicobacter pilory*, a qual apresenta seqüências e estrutura razoavelmente bem conservadas. Composição da molécula semelhante a de ClpP de *S. pneumoniae* (KIM E KIM, 2008) 10
- Figura 2.3** Desenho esquemático da região próxima ao sítio catalítico da proteína ClpP de seis diferentes microrganismos. Os desenhos em verde e lilás representam duas cepas de *H. pilory*, em marrom *E. coli*, em azul *S. pneumoniae*, em laranja *P.falciparum* e em amarelo *M. tuberculosis* (KIM E KIM, 2008) 10
- Figura 2.4** Representação esquemática de IgA1 protease típica de *Streptococcus*. Adaptado de PAOLIS *et al.* (2007) 13
- Figura 2.5** Degradação do ácido hialurônico e formação da unidade dissacarídica (principal produto da ação da Hyl bacteriana sobre o ácido hialurônico). Adaptado de AKHTAR E BHAKUNI (2004). 14
- Figura 2.6** Estrutura da Hyl de *Streptococcus pneumoniae* complexada com dissacarídeo de ácido hialurônico, depositada no PDB (*Protein Data Bank*) sob o número de identificação 1C82 (PONNURAJ E JEDRZEJAS, 2000) 15
- Figura 3.1** Sítio de clonagem e expressão do vetor pET28b (Novagen). Os círculos vermelhos indicam os sítios das enzimas de restrição, também adicionados aos oligonucleotídeos iniciadores, onde o gene *clpP* amplificado irá se ligar. O círculo verde indica o códon de terminação, ou seja, a sinalização para o

termino da tradução, possibilitando assim a expressão da proteína fusionada a cauda de histidina. 32

**Figura 3.2** Desenho esquemático das etapas realizadas na clonagem (construção do vetor pET28b/*clpP*). 35

**Figura 3.3** Mapa do vetor pET28a, igual ao pET28b (Novagen). O mapa mostra a origem de replicação (*ori*), o repressor *lac* (*lacI*), o gene de resistência à canamicina (*Kan*) e o sítio de múltipla clonagem (seta em preto sólido). 35

**Figura 3.4** Desenho esquemático dos procedimentos realizados para a análise de viabilidade celular. 43

**Figura 3.5** Esquema dos procedimentos realizados nos experimentos de expressão e análise de solubilidade da proteína. 45

**Figura 3.6** Desenho esquemático das etapas realizadas na análise de segregação plasmidial. 49

**Figura 3.7** Fotos das placas de LB ágar, com e sem canamicina, das análises de segregação plasmidial do experimento PC8 do planejamento fatorial. 50

**Figura 3.8** Mapa do plasmídeo pBSK ClpP717. 53

**Figura 3.9** Mapa do plasmídeo pHCMC03 (NGUYEN *et al.*,2005) 54

**Figura 3.10** Desenho esquemático das etapas realizadas para a construção do vetor utilizada na clonagem em *B. subtilis*. 55

**Figura 4.1.** Comparação da seqüência de aminoácidos da proteína ClpP de *S. pneumoniae* CGSP14, depositada no *GenBank* sob o código GI: 182629000, com outras seqüências depositadas no banco de dados de proteínas, através da ferramenta Blastp. A seta indica onde a análise avalia que além das três seqüências

demonstradas existem mais 52 seqüências com 100% de identidade. 59

**Figura 4.2.** Alinhamento múltiplo, realizado com o auxílio do programa CLUSTAL W, de oito seqüências da proteína ClpP, de diferentes cepas de *S. pneumoniae*, depositadas no *GenBank*. O alinhamento foi realizado com as seguintes cepas de *S. pneumoniae*: 70585 (GI: 225858557), TIGR4 (GI: 930696), R6 (GI: 934194), CGSP14 (GI:182629000), Hungary19A-6 (GI:168994986), D39 (GI: 116076792), P1031 (GI: 225856422), G54 (GI:194358110). 60

**Figura 4.3.** Regiões transmembranares presentes na proteína IgA1 protease de *S. pneumoniae* CGSP14, segundo predição do programa CLC Workbench 4. (A) Desenho esquemático da região amino terminal de IgA1 protease, apresentando as regiões transmembranares e o peptídeo sinal. (B) Esquema da disposição da proteína na membrana celular, apresentando as três regiões que atravessam a membrana. 63

**Figura 4.4.** Análise da região amino terminal da proteína IgA1 protease de *S. pneumoniae* CGSP14. Predição do peptídeo sinal através do programa SignalP 3.0. *S score* (linha verde) mostra a probabilidade de um aminoácido estar presente no peptídeo sinal, *C score* (linha vermelha) e *Y score* (linha azul) mostram onde seja o provável começo da proteína. 64

**Figura 4.5.** Esquema da disposição da Hyl na membrana celular, de acordo com o programa CLC Workbench 4. As duas regiões da proteína que atravessam a membrana (transmembranares) são representadas no esquema. 65

**Figura 4.6.** Análise da região amino terminal da proteína Hyl de *S. pneumoniae* CGSP14. Predição do peptídeo sinal através do programa SignalP 3.0. *S score* (linha verde) mostra a probabilidade de um aminoácido estar presente no peptídeo sinal, *C score* (linha vermelha) e *Y score* (linha azul) mostram onde seja o provável começo da proteína. 66

**Figura 4.7** Gel de agarose 0,8% em TAE mostrando o perfil eletroforético da

solução resultante da extração de DNA genômico de *S. pneumoniae*. **1 e 2**, DNA genômico extraído de *S. pneumoniae* (amostras 3 e 4 respectivamente); **3 e 4**, DNA genômico extraído de *S. pneumoniae* diluído 10 vezes (amostras 3 e 4 respectivamente); **5 e 6**, DNA genômico extraído de *S. pneumoniae* diluído 100 vezes (amostras 3 e 4 respectivamente); **M**, padrão de tamanho de fragmento (DNA  $\lambda$  digerido com *Hind* III). A seta indica a banda do marcador com 23130bp. 68

**Figura 4.8** (A) Gel de agarose 0,8% em TAE da amplificação do gene *clpP* utilizando diferentes temperaturas de anelamento no programa do PCR. **M**, padrão de tamanho de fragmento ( $\Phi$ X174 DNA digerido com *Hae* III); Gene *clpP* amplificado por PCR utilizando temperatura de anelamento dos oligonucleotídeos iniciadores igual a 54°C (**1**), 55,1°C (**2**), 56°C (**3**), 55,9°C (**4**); Controle negativo da reação de PCR (**5**); A seta indica o banda do marcador com 603bp, próximo ao tamanho esperado do gene *clpP* amplificado. (B) Gel de agarose 1,5% em TAE do produto de PCR purificado. **M**, padrão de tamanho de fragmento ( $\Phi$ X174 DNA digerido com *Hae* III); **1**, Purificação do gene *clpP* amplificado; A seta indica o banda do marcador com 603bp, próximo ao tamanho esperado do gene *clpP* amplificado. 69

**Figura 4.9** Confirmação dos clones - Digestão dos plasmídeos extraídos das cepas transformadas *E. coli* TOP 10 com o vetor construído. **M**, padrão de tamanho de fragmento 1 kb DNA Ladder (Invitrogen); **1, 3, 5 e 7**, plasmídeos extraídos dos clones C1N, C8N, C2P e C4P, respectivamente; **2, 4, 6 e 8**, plasmídeos digeridos extraídos dos clones C1N, C8N, C2P e C4P, respectivamente; **9 e 11**, *clpP* e pET28b, respectivamente; **10** digestão do gene *clpP* com *Pst*I (controle positivo); **12**, digestão do vetor pET28b com *Eco*RV (controle positivo). As setas indicam as bandas do padrão de tamanho de fragmento mais próximas as bandas esperadas na digestão de clones positivos. 70

**Figura 4.10** Digestão dos plasmídeos extraídos das cepas transformadas *E. coli* BL21 Star (DE3). **M**, padrão de tamanho de fragmento 1 kb DNA Ladder (Invitrogen); **1, 3, 5 e 7**, plasmídeos extraídos dos clones C2P/1, C2P/2, C2P/3, C2P/4, respectivamente; **2, 4, 6 e 8**, plasmídeos digeridos extraídos dos clones



C2P/1, C2P/2, C2P/3, C2P/4, respectivamente; **9** e **11**, *clpP* e pET28b, respectivamente; **10** digestão do gene *clpP* com *PstI* (controle positivo); **12**, digestão do vetor pET28b com *EcoRV* (controle positivo). As setas indicam as bandas do padrão de tamanho de fragmento mais próximas as bandas esperadas na digestão de clones positivos. 71

**Figura 4.11** Seqüenciamento do Plasmídeo pET28b/*clpP*-C2P. A figura mostra a seqüência de nucleotídeos do gene *clpP* (preto), o sítio de restrição da enzima *XhoI* (rosa), a seqüência da cauda de histidina (azul) e o *códon* de terminação (vermelho), presentes no plasmídeo construído e seqüenciado. 71

**Figura 4.12** Gel de SDS-PAGE 12,5% dos extratos protéicos totais e de suas frações, solúvel e insolúvel, obtidos através da expressão da proteína ClpP por *E. coli* BL21 Star (DE3). **M1** e **M2**, padrão de massa molar; **1** e **5**, amostras do cultivo não induzido dos clones C2P/2 e C2P/4, respectivamente; **2** e **6**, extratos protéicos totais dos clones C2P/2 e C2P/4, respectivamente; **3** e **7**, frações insolúveis dos clones C2P/2 e C2P/4, respectivamente; **4** e **8**, frações solúveis dos clones C2P/2 e C2P/4, respectivamente. As amostras aplicadas no gel foram normalizadas por  $Abs_{600nm}$ . 74

**Figura 4.13** Gel de SDS-PAGE 12,5% com o extrato total das proteínas dos experimentos do planejamento fatorial. (A) **M**, padrão de massa molar LMW (Amersham Bioscience); **1**, **3**, **5**, e **7**, extrato total sem indução dos experimentos 1, 2, 5(PC) e 7(PC), respectivamente; **2**, **4**, **6** e **8**, extrato total com 4h de indução dos experimentos 1, 2, 5(PC) e 7(PC), respectivamente. (B) **M**, padrão de massa molar LMW (Amersham bioscience); **1**, **3**, **5**, e **7**, extrato total sem indução dos experimentos 3, 4, 6(PC) e 8(PC), respectivamente; **2**, **4**, **6** e **8** extrato total com 4h de indução dos experimentos 3, 4, 6(PC) e 8(PC), respectivamente. As amostras foram normalizadas pela  $Abs_{600nm}$ . 78

**Figura 4.14** Gráfico de *F* (fração de células com plasmídeo) em função da concentração de IPTG, mostrando o efeito não linear do sistema. 81

**Figura 4.15** Cinética de crescimento celular das réplicas da condição otimizada, experimentos A, B, C, D e E, e do experimento N (controle negativo). A seta mostra o ponto em que foi adicionado o indutor IPTG. 83

**Figura 4.16** Ajuste linear dos pontos durante a fase exponencial de crescimento celular a partir da indução da expressão. Cálculo da velocidade específica de crescimento de *E. coli* durante a expressão da proteína ClpP. 84

**Figura 4.17** Cinética da produção de proteína. Os quadrados representam a concentração média de ClpP ao longo do tempo utilizando a condição otimizada pelo planejamento de experimentos (0,1mM IPTG e 0µg/mL). As barras representam os desvios padrões nos pontos do experimento na condição otimizada. 86

**Figura 4.18** Gráfico da variável  $F$  (fração de células com plasmídeo) em função do tempo. 87

**Figura 4.19** Gráfico do Fator de rendimento de massa de ClpP produzida por massa de célula ( $Y_{P/X}$ ) e da Fração de células com plasmídeo ( $F$ ) em função do tempo. 90

**Figura 4.20** Gel de SDS-PAGE 12,5% da purificação da proteína ClpP. **M**, padrão de massa molar LMW (Amersham Bioscience); **1**, extrato celular total lisado; **2**, fração não adsorvida (amostra sem imidazol); **3**, eluído com a primeira concentração de imidazol (amostra sem imidazol), **4**, eluído com 220 mM de imidazol (amostra sem imidazol); **5**, eluído com 300 mM de imidazol (amostra sem imidazol); **6**, fração não adsorvida (amostra com 1mM de imidazol); **7**, eluído com a primeira concentração de imidazol (amostra com 1 mM de imidazol); **8**, eluído com 220 mM de imidazol (amostra com 1mM de imidazol); **9**, eluído com 300 mM de imidazol (amostra com 1 mM de imidazol). 91

**Figura 4.21** Gel de SDS-PAGE 12,5% do teste de estabilidade da proteína. (A) gel das amostras logo após a purificação; **M**, padrão de massa molar LMW

(Amersham Bioscience); **1** e **2**, amostras de ClpP purificada. (B) gel das amostras armazenadas a -20°C durante 7 meses após a purificação; **M**, padrão de massa molar LMW (Amersham Bioscience); **1**, **2**, e **3**, amostras de ClpP purificadas e armazenadas a -20°C por 7 meses.

92

**Figura 4.22** Alinhamento múltiplo do gene *clpP* sintético utilizado para a clonagem em *B. subtilis*, sem a otimização de códons (submetido) e otimizado. Em verde são mostrados os nucleotídeos que sofreram alterações no gene otimizado. A seqüência do peptídeo sinal está realçada em amarelo. A seqüência do sítio de clivagem está sublinhada.

95

**Figura 4.23** (A) gel de agarose 1% das digestões dos vetores pBSKClpP717 e pHCMC03. **M**, marcador padrão de tamanho de fragmento 1kb DNA ladder; **1**, vetor pBSKclpP717; **2** e **3**, vetor pBSKclpP717 digerido com as enzimas *BamHI* e *XbaI*; **4**, vetor pHCMC03; **5** e **6**, vetor pHCMC03 digerido com as enzimas *BamHI* e *XbaI*. (B) gel de agarose 1% das purificações das digestões. **M**, marcador padrão de tamanho de fragmento 1kb DNA ladder; **1**, vetor pBSKclpP717; **2**, inserto (*clpP*) digerido e purificado; **3**, vetor pHCMC03; **4**, **5** e **6** vetor pHCMC03 linearizado e purificado.

96

## LISTA DE TABELAS

- Tabela 3.1** Oligonucleotídeos iniciadores utilizados na reação de PCR para amplificação do gene *clpP* a partir de DNA de *S. pneumoniae* CGSP14. Os sítios de clivagem das enzimas de restrição são mostrados sublinhados nas seqüências, sítio de *NcoI* na seqüência senso e *XhoI* na anti-senso. 31
- Tabela 3.2** Tabela com as condições dos experimentos do planejamento fatorial 2<sup>2</sup>. Os números entre parênteses são os valores das variáveis independentes normalizados. 47
- Tabela 4.1.** Dados gerados pelo alinhamento múltiplo, realizado no programa CLUSTAL W, de oito seqüências de aminoácidos da proteína IgA1 protease de *S. pneumoniae* depositadas no *GenBank*. 62
- Tabela 4.2.** Dados gerados pelo alinhamento múltiplo, realizado no programa CLUSTAL W, de nove seqüências de aminoácidos da proteína Hyl de *S. pneumoniae* depositadas no *GenBank*. 65
- Tabela 4.3** Mutações detectadas no seqüenciamento do plasmídeo pET28b/*clpP*-C2P quando comparada a seqüência obtida pelo seqüenciamento e a seqüência de *S. pneumoniae* CGSP14 depositada no *GenBank*. 72
- Tabela 4.4** Experimentos do planejamento fatorial e suas variáveis de resposta 75
- Tabela 4.5** Efeitos das variáveis sobre o crescimento celular (mg/mL de massa seca de células) 77
- Tabela 4.6** Efeitos das variáveis sobre concentração de ClpP (mg/L) 79
- Tabela 4.7** Efeitos das variáveis sobre o valor de *F* (fração de células com plasmídeo), desconsiderando o valor de *F* do experimento PC6. 79

<b>Tabela 4.8</b> Efeitos das variáveis sobre o valor de $F$ (fração de células com plasmídeo), considerando o valor de $F$ do experimento PC6.	80
<b>Tabela 4.9</b> Crescimento celular dos experimentos de validação do planejamento fatorial	82
<b>Tabela 4.10</b> Concentração de ClpP recombinante produzida ao longo das 4 horas de expressão nos experimentos da validação.	85
<b>Tabela 4.11</b> Fatores de rendimento ( $Y_{P/X}$ ) ao longo do tempo de indução	86
<b>Tabela 4.12</b> Valores da variável $F$ (fração de células com plasmídeo) ao longo do tempo nas réplicas dos experimentos da validação	88
<b>Tabela 4.13</b> Resultado dos testes bioquímicos realizados com as cepas <i>Bacillus subtilis</i> WB600 e <i>Bacillus subtilis</i> LFB732 (utilizada para comparação).	94

## LISTA DE ABREVIACOES E SMBOLOS

aa	Aminocido
Abs <sub>600nm</sub>	Absorbncia a 600nm
BLAST	<i>Basic Local Alignment Search Tool</i>
DNA	cido desoxirribonuclico
DTT	Ditiotreitol
EDTA	cido etilenodiaminotetractico
<i>F</i>	Frao de clulas com plasmdeo
GI	Nmero de identificao do gene no <i>GenBank</i>
GRAS	<i>generally recognized as safe</i>
HIV	Vrus da Imunodeficincia Humana
HPLC	Cromatografia lquida de alta performance
IPTG	Isopropil- $\beta$ -D-tiogalactosdeo
LB	Luria-Bertani
LMW	<i>Low Molecular Weight</i>
LPS	Lipopolissacardeo
Mops	Morfolinopropanosulfnico
mRNA	cido ribonuclico mensageiro
NCBI	<i>National Center for Biotechnology</i>
<i>p</i>	<i>p</i> -valor
pb	Pares de base
PBS	Tampo fosfato salina
PC	Ponto Central
PCR	Reao em Cadeia da Polimerase
PDB	<i>Protein Data Bank</i>
PCV-7	Vacina conjugada heptavalente
RNA	cido ribonuclico
SDS	Dodecil sulfato de sdio

SDS-PAGE	Gel de Eletroforese Desnaturante de Poliacrilamida
TAE	Tampão Tris-Acetato-EDTA
Tris	Tris(hidroximetil)aminometano
TSB	<i>Tryptic Soy Broth</i>
UFC	Unidade Formadora de Colônia
$Y_{P/X}$	Fator de Rendimento (produto por célula)

## 1 INTRODUÇÃO

Uma das principais causas de mortalidade e morbidade em todo o mundo, principalmente em países em desenvolvimento, são infecções causadas por *Streptococcus pneumoniae* (GARCÍA-SUÁREZ *et al.*, 2006; OMS, 2007; BRAIDO *et al.*, 2008). Este microrganismo é uma bactéria encapsulada, gram-positiva, anaeróbia facultativa, que coloniza o trato respiratório superior de seres humanos saudáveis e é transmitido por aerossol. O exterior da superfície bacteriana é coberto por uma cápsula polissacarídica. Existem mais de 90 diferentes tipos de polissacarídeos capsulares, indentificando pelo menos 90 diferentes sorotipos (BRAIDO *et al.*, 2008; PLETZ *et al.*, 2008).

A virulência de *S. pneumoniae* é largamente dependente da sua cápsula de polissacarídeo, que é muito heterogênea entre os sorotipos existentes e representa um sério obstáculo para a concepção de uma vacina com alta eficiência (BAROCCHI *et al.* 2007). No entanto, inúmeras proteínas da bactéria também têm sido consideradas fatores de virulência, sendo envolvidas na patogenicidade das doenças pneumocócicas (JEDRZEJAS, 2001; GARCÍA-SUÁREZ *et al.*, 2006; BRAIDO *et al.*, 2008).

A espécie *S. pneumoniae* coloniza assintomaticamente aproximadamente 30-70% das crianças saudáveis e 6 % dos adultos saudáveis, aumentando para 18-30% dos adultos quando estes convivem com crianças (BRAIDO *et al.*, 2008). Esta bactéria é a principal causadora de pneumonia bacteriana e é freqüentemente envolvida em doenças como otite, sinusite e meningite, além de quadros clínicos de septicemia e bacteremia (JEDRZEJAS, 2001; BOGAERT, 2004; BRAIDO *et al.*, 2008). Dentre todas as infecções invasivas causadas por este microrganismo, cerca de 90% são representadas por pneumonia e 5% por meningite (BRAIDO *et al.*, 2008).

Nos países desenvolvidos, a incidência das doenças pneumocócicas é maior entre crianças com menos de 2 anos e idosos. Condições associadas à deficiência imune, como infecção pelo vírus do HIV, aumentam significativamente a probabilidade de contrair doenças causadas por *S. pneumoniae*. Ou seja, as taxas mundiais de doenças pneumocócicas são muito altas, porém são ainda mais preocupantes em grupos de risco, como crianças, idosos e imunocomprometidos. Dados da Organização Mundial da



Saúde (OMS) estimam que 1,6 milhões de pessoas morrem todo o ano no mundo devido a doenças pneumocócicas, sendo que destas, em torno de 1 milhão são crianças com menos de 5 anos, e a maioria dos casos ocorre em países em desenvolvimento (OMS, 2007).

Outro fato que preocupa é que nos últimos anos vem aumentando o número de cepas resistentes aos mais diversos tipos de antibióticos, sendo encontradas cepas resistentes até mesmo à Vancomicina, o que torna o tratamento das infecções pneumocócicas um grave problema de saúde pública (NOVAK *et al.*, 1999; JEDRZEJAS, 2001; BRAIDO *et al.*, 2008).

Atualmente existem dois tipos de vacinas contra *S. pneumoniae* disponíveis no mercado mundial: vacinas polissacarídicas e vacinas conjugadas. Ambas utilizam como antígenos polissacarídeos capsulares. A vacina polissacarídica é composta por 23 polissacarídeos capsulares purificados de diferentes sorotipos de *S. pneumoniae* (sorotipos 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F e 33F) (BRAIDO *et al.*, 2008; PLETZ *et al.*, 2008). Esta vacina é eficaz em adultos e crianças maiores de 2 anos, no entanto, tem baixa eficácia em crianças com menos de 2 anos, imunocomprometidos e idosos. Além disso, tem seu período de proteção limitado (BOGAERT *et al.*, 2004; GARCÍA-SUÁREZ *et al.*, 2006; PLETZ *et al.*, 2008). A vacina polissacarídica não é eficiente em menores de 2 anos, devido à ausência de linfócitos B maduros que induzem uma melhor resposta contra carboidratos. Em adultos é preciso fazer a re-vacinação depois de 5-6 anos. A vantagem desta vacina é o alto número de sorotipos abrangidos (PLETZ *et al.*, 2008).

Atualmente é comercializada a vacina conjugada heptavalente (PCV-7), composta por polissacarídeos capsulares de 7 sorotipos (4, 6B, 9V, 14, 18C, 19F, 23F) os quais representam os mais frequentemente envolvidos em infecções invasivas (PLETZ *et al.*, 2008). Os polissacarídios presentes são conjugados ao toxóide diftérico CRM 197 (detoxificado geneticamente) (BRAIDO *et al.*, 2008). Esta vacina tem se mostrado eficaz em crianças, no entanto, ela possui cobertura limitada de sorotipos e custo elevado, o que torna seu uso proibitivo para um programa nacional de vacinação em países em desenvolvimento (BOGAERT *et al.*, 2004; BAROCCHI *et al.*, 2007). A PCV-7 pode alcançar uma cobertura de 65-80% dos sorotipos relacionados as doenças

pneumocócicas invasivas em crianças nos países industrializados ocidentais, nos quais os sorotipos presentes nesta vacina são os mais prevalentes. No entanto, esta cobertura pode chegar a níveis muito mais baixos em alguns países em desenvolvimento (OMS, 2007).

Além da vacina conjugada heptavalente já comercializada, no ano de 2009 estavam em fase final dos testes clínicos as vacinas conjugadas 10-valente e 13-valente (DAGAN E FRASCH, 2009). A vacina 13-valente, desenvolvida pela empresa Wyeth, além dos 7 sorotipos presentes na PCV-7, apresenta também os sorotipos 1, 3, 5, 6A, 7F e 19A. Os polissacarídeos presentes nesta vacina também são conjugados ao toxóide diftérico CRM 197 (detoxificado geneticamente). É estimado que esta vacina aumente a cobertura das doenças pneumocócicas invasivas em 89% na Europa, 92% nos Estados Unidos e Canadá, 86% na Oceania, 87% na África e América Latina e 73% na Ásia (GRIMPREL, 2009). A vacina 10-valente, desenvolvida pela empresa GlaxoSmithKline Biologicals (GSK), apresenta mais 3 sorotipos (1, 5 e 7F) além daqueles presentes na PCV-7. Esta vacina possui 8 polissacarídeos capsulares conjugados com a proteína D de *H. influenzae* (WYSOCKI *et al.*, 2009). Através de acordo de transferência de tecnologia entre o Ministério da Saúde brasileiro e a empresa GSK, a vacina 10-valente fará parte do Calendário Básico de Vacinação do Programa Nacional de Imunização (PNI) a partir de março deste ano (LOBATO, 2010).

Diante dos problemas apresentados, o desenvolvimento de uma vacina anti-pneumocócica eficaz contra uma ampla gama de sorotipos tornou-se uma das prioridades no mundo inteiro (JEDRZEJAS, 2001). Como alternativa para a obtenção de uma vacina que apresente maior cobertura e menor custo, nas últimas décadas, vários grupos vêm investigando diversas proteínas associadas à virulência, como candidatas à formulação de uma vacina protéica (PATON, 1998; JEDRZEJAS, 2001; DI GUILMI E DESSEN, 2002; BOGAERT *et al.*, 2004; GARCIA-SUÁREZ *et al.*, 2006; OGUNNIYI *et al.*, 2007; BAROCCHI *et al.*, 2007). Uma futura vacina protéica provavelmente será uma combinação de proteínas recombinantes que apresentem alta imunogenicidade e sejam protetoras, alta conservação e estejam presentes em todas as cepas de *Streptococcus pneumoniae* (GARCIA-SUÁREZ *et al.*, 2006).

Em face do exposto acima, um dos objetivos deste trabalho foi selecionar, clonar e expressar, utilizando *Escherichia coli* e *Bacillus subtilis* como sistemas de expressão, uma proteína de *Streptococcus pneumoniae*, diretamente associada à sua virulência, com elevado nível de prevalência e conservação entre os vários sorotipos. Além disso, objetivou-se otimizar a expressão da proteína recombinante empregando ferramentas de planejamento experimental, bem como estudar a segregação plasmidial do sistema. Deste modo, buscou-se contribuir para o estudo desta proteína recombinante como possível antígeno para a composição de uma vacina contra *Streptococcus pneumoniae*, além de auxiliar na compreensão do sistema de produção em nível industrial.

## 2 REVISÃO DA LITERATURA

Para subsidiar a discussão sobre a seleção, clonagem e expressão da proteína, bem como sobre a otimização da expressão e os estudos do comportamento do sistema, este capítulo apresentará uma revisão abordando os fatores de virulência de *S. pneumoniae* e os sistemas de expressão utilizados neste trabalho.

### 2.1 Fatores de virulência

Nesta seção serão apresentados alguns dos fatores de virulência de *S. pneumoniae*, em especial as proteínas ClpP protease, IgA1 protease e Hialuronato liase (Hyl). São descritos como fatores de virulência componentes celulares que atuam na invasão, aderência e colonização ou que são capazes de produzir uma resposta inflamatória (SALYERS E WHITT *apud* SILVA, 2005, p.26).

O principal fator de virulência de *S. pneumoniae* é a cápsula polissacarídica, sendo que cepas que não possuem a cápsula têm sua patogenicidade diminuída drasticamente quando comparadas com cepas encapsuladas. Além da cápsula polissacarídica, algumas proteínas têm sido sugeridas como fatores de virulência do microrganismo. Sabe-se ainda que os fatores de virulência que seriam os melhores antígenos para composição de uma vacina são aqueles que, quando retirados de *S. pneumoniae* (cepas mutantes), fazem com que sua virulência seja atenuada (JEDRZEJAS, 2001).

#### 2.1.1 Cápsula polissacarídica

A camada mais externa de *S. pneumoniae* é formada pela cápsula polissacarídica, a qual possui em torno de 200 a 400 nm de espessura (KADIOGLU *et al.*, 2008). A cápsula é importantíssima para a virulência do microrganismo, sendo considerado seu principal fator de virulência (KADIOGLU *et al.*, 2008, GARCIA-SUÁREZ *et al.*, 2006, BOGAERT *et al.*, 2004).

Durante a colonização das vias aéreas superiores, a cápsula evita o aprisionamento no muco nasal, permitindo o acesso do microrganismo às superfícies epiteliais (KADIOGLU *et al.*, 2008). A cápsula polissacarídica é fortemente anti-fagocitária, formando uma espécie de escudo, evitando a fagocitose por macrófagos (KADIOGLU *et al.*, 2008, BOGAERT *et al.*, 2004). Além disso, ela é capaz de proporcionar resistência à autólise espontânea ou induzida por antibióticos, contribuindo assim para a tolerância das cepas aos antibióticos. A capacidade infecciosa de *S. pneumoniae* difere entre os diferentes sorotipos capsulares, os quais possuem diferentes composições de polissacarídeos capsulares (KADIOGLU *et al.*, 2008).

### 2.1.2 Proteínas

Algumas proteínas ou enzimas, que são de alguma forma expostas na superfície do microrganismo, podem contribuir para a patogenicidade e estar envolvidas no processo de infecção da doença (JEDRZEJAS, 2001; BOGAERT *et al.*, 2004). Muitas vezes, estas proteínas estão envolvidas diretamente nas interações com o tecido do hospedeiro ou nos mecanismos para encobrir a superfície bacteriana do sistema de defesa do hospedeiro (JEDRZEJAS, 2001). No caso de *S. pneumoniae*, estima-se que este possua mais de 100 proteínas de superfície, sendo que muitas destas têm algum tipo de função que contribui para a patogenicidade do microrganismo (BAROCCHI *et al.*, 2007).

Nos últimos anos, estudos têm sugerido diversas proteínas associadas à virulência como candidatos a antígenos para formulação de uma vacina protéica (PATON, 1998; JEDRZEJAS, 2001; DI GUILMI E DESSEN, 2002; BOGAERT *et al.*, 2004; GARCIA-SUÁREZ *et al.*, 2006; OGUNNIYI *et al.*, 2007; BAROCCHI *et al.*, 2007), sendo provável que uma futura vacina seja uma combinação de proteínas recombinantes que apresentem alta imunogenicidade e sejam prevalentes em todas as cepas de *S. pneumoniae* (GARCIA-SUÁREZ *et al.*, 2006). Algumas destas proteínas são pneumolisina, neuromidases, autolisinas, proteínas de superfície ligantes à colina (como a PspA), lipoproteínas (como a PsaA), hialuronidases, entre outras.

A pneumolisina, pertencente ao grupo conhecido como citolisinas colesterol-dependentes, é uma enzima citoplasmática que é liberada devido à ação de outra enzima de *S. pneumoniae*, a autolisina. A pneumolisina inibe o movimento ciliar nos brônquios e desorganiza a monocamada epitelial bronquial, dificultando a retirada de muco e facilitando o espalhamento da infecção (JEDRZEJAS, 2001; BOGAERT *et al.*, 2004). Além disso, ela também inibe a função fagocítica sendo, portanto, importante na patogênese da doença (BOGAERT *et al.*, 2004). A pneumolisina é uma proteína altamente conservada e é um antígeno capaz de estimular a proteção imunológica (GARCIA-SUÁREZ *et al.*, 2006).

Outro fator de virulência de *S. pneumoniae* é a proteína Neuromidase, amplamente presente nas cepas do microrganismo, que se apresenta pelo menos em duas formas, NanA e NanB. Estas proteínas mudam o padrão de glicosilação da célula do hospedeiro, deixando as superfícies da célula mais expostas (JEDRZEJAS, 2001), e diminuem a viscosidade das superfícies mucosas, facilitando a adesão (BOGAERT *et al.*, 2004).

Outra proteína estudada como fator de virulência de *S. pneumoniae* é a PspA (*Pneumococcal surface protein A*), pertencente a um grupo de proteínas de superfície ligantes à colina (JEDRZEJAS, 2001; BOGAERT *et al.*, 2004). Ela situa-se ancorada na parede celular de *S. pneumoniae*, sendo que os domínios ligadores de colina são os responsáveis pela ligação da proteína na superfície da bactéria. A PspA é amplamente presente nas cepas de *S. pneumoniae* (JEDRZEJAS, 2001), no entanto, apresenta um elevado polimorfismo devido à seleção imunológica, já que sua localização é de fácil acesso para os anticorpos (GARCIA-SUÁREZ *et al.*, 2006). Esta proteína tem a capacidade de reduzir os níveis de fagocitose do microrganismo (JEDRZEJAS, 2001) e tem sido uma das proteínas mais estudadas como antígeno para composição de uma vacina (BOGAERT *et al.*, 2004), demonstrando bons resultados em modelo animal (JEDRZEJAS, 2001; BOGAERT *et al.*, 2004; GARCIA-SUÁREZ *et al.*, 2006).

As autolisinas são capazes de degradar peptideoglicanos, ou seja, estas enzimas degradam a parede celular. Um exemplo destas autolisinas é a enzima conhecida como

LytA amidase (*N-acetylmuramoyl-L-alanine amidase*), também pertencente ao grupo de proteínas de superfície ligantes à colina, que está envolvida na patogenicidade do microrganismo e já mostrou bons resultados quando utilizada como antígeno em modelo animal (JEDRZEJAS, 2001). No entanto, o papel exato das autolisinas na patogênese ainda precisa ser melhor estudado assim como as propriedades de proteção desta proteína (BOGAERT *et al.*, 2004).

Outra proteína de superfície também pertencente ao grupo de proteínas de superfície ligantes à colina é a CbpA (*Choline binding protein A*). Ela é exposta na superfície de *S. pneumoniae* e está envolvida na aderência e colonização dos tecidos do hospedeiro (JEDRZEJAS, 2001).

A PsaA (*pneumococcal surface adhesin A*), membro da família das chamadas lipoproteínas ligantes de metal, é outra proteína relacionada à virulência de *S. pneumoniae* e importante candidata a uma vacina (BOGAERT *et al.*, 2004; BAROCCHI *et al.*, 2007). Esta proteína faz parte de um sistema de transporte de manganês e zinco para o citoplasma da bactéria e está ancorada na membrana celular (JEDRZEJAS, 2001).

Outros fatores de virulência são as proteínas pneumocócicas homólogas de superfície celular denominadas Pht (*pneumococcal histidine triad*) (BOGAERT *et al.*, 2004; BAROCCHI *et al.*, 2007), as quais são altamente conservadas, implicando em um potencial de ampla cobertura para uma vacina (BOGAERT *et al.*, 2004).

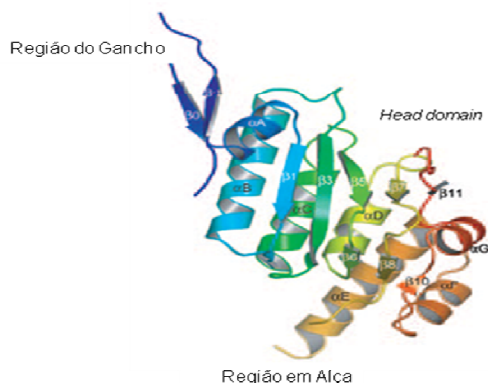
A Hialuronato liase (Hyl), outra proteína investigada como fator de virulência, pertencente ao grupo das hialuronidasas, hidrolisa ácido hialurônico, o qual é um importante componente de tecidos conectivos (JEDRZEJAS, 2001). Recentemente, novas proteínas têm sido descritas, como as proteínas de *pili* que contribuem para a aderência, virulência e induzem resposta inflamatória (BAROCCHI *et al.*, 2006), a proteína de choque térmico ClpP, que modula a expressão dos genes de virulência (KWON *et al.*, 2004), e a proteína Imunoglobulina A1 protease (IgA1 protease), que desempenha um importante papel na aderência e na resistência do patógeno à resposta imune do hospedeiro (ROMANELLO *et al.*, 2006).

### 2.1.2.1 ClpP

A ClpP é uma serino-protease (GRIBUN *et al.*, 2005) pertencente ao grupo conhecido como proteínas de choque térmico (HSPs) (KWON *et al.*, 2004). Esta proteína é altamente conservada tanto em procariotos quanto em eucariotos (CAO *et al.*, 2007), sendo que, entre os vários sorotipos de *S. pneumoniae*, sua seqüência é muito conservada (CAO *et al.*, 2009).

Esta proteína de choque térmico forma um complexo com proteínas ATP-dependentes, como a ClpA e ClpX, sendo que o papel funcional deste complexo é a regulação de proteínas através de proteólise, atuando como uma espécie de controle de qualidade de proteínas (KIM E KIM, 2008). As proteínas ClpA e ClpX são responsáveis por desdobrar e conduzir as proteínas até o sítio ativo da ClpP, onde serão degradadas (GRIBUN *et al.*, 2005; KIM E KIM, 2008). Além disso, a ClpP por si so é capaz de degradar apenas peptídeos de 6 aminoácidos, sendo que grandes polipeptídeos só podem ser degradados quando esta proteína está complexada com ClpX ou ClpA (GRIBUN *et al.*, 2005).

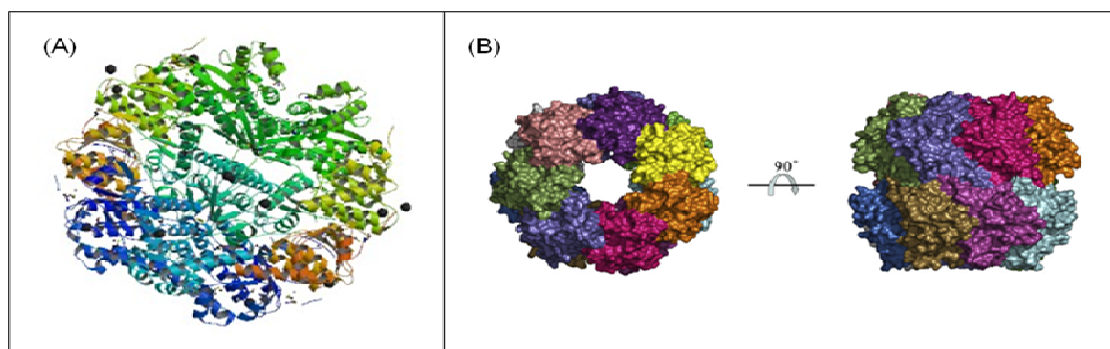
A ClpP é composta por dois anéis heptaméricos, onde cada anel heptamérico é formado por sete unidades principais da ClpP (monômero), sendo cada monômero com aproximadamente 21kDa. O monômero de ClpP, mostrado na figura 2.1, pode ser dividido em duas principais regiões, uma região em alça que é responsável pelas interações entre os anéis e um domínio principal (*head domain*) (GRIBUN *et al.*, 2005).



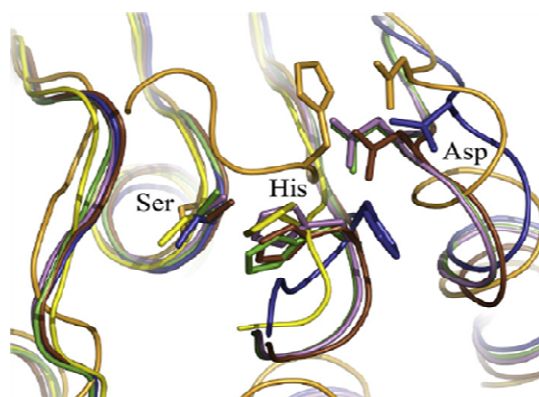
**Figura 2.1** Representação do monômero de ClpP de *S. pneumoniae*. Adaptado de GRIBUN *et al.* (2005)



A região em alça de um anel heptamérico se une à região em alça do outro anel, formando a estrutura tetradecamérica da ClpP, mostrada na figura 2.2. Na região N-terminal de cada monômero, a qual possui uma seqüência altamente conservada nos diversos reinos da vida, a proteína apresenta um gancho, que é essencial para a formação do complexo ClpAP ou ClpXP. Entre as regiões em alça e o domínio principal se localiza o sítio catalítico da ClpP, a tríade catalítica Ser<sup>111</sup>, His<sup>136</sup> e Asp<sup>185</sup>, característica de serino-proteases. A figura 2.3 mostra a região da tríade catalítica de proteínas ClpP de algumas bactérias, inclusive de *S. pneumoniae* (GRIBUN *et al.*, 2005).



**Figura 2.2** (A) Estrutura da proteína ClpP de *S. pneumoniae* depositada no PDB (*Protein Data Bank*) sob o número de identificação 1Y70 (GRIBUN *et al.*, 2005). (B) Representação esquemática da estrutura da molécula tetradecamérica de ClpP de *Helicobacter pilory*, a qual apresenta seqüências e estrutura razoavelmente bem conservadas. Composição da molécula semelhante à de ClpP de *S. pneumoniae* (KIM E KIM, 2008)



**Figura 2.3** Desenho esquemático da região próxima ao sítio catalítico da proteína ClpP de seis diferentes microorganismos. Os desenhos em verde e lilás representam duas cepas de *H. pilory*, em marrom *E. coli*, em azul *S. pneumoniae*, em laranja *P.falciparum* e em amarelo *M. tuberculosis* (KIM E KIM, 2008).

Sabe-se que as proteínas de choque térmico, como a ClpP, protegem a bactéria contra efeitos adversos, aumentando seus níveis de sobrevivência. Quando o microrganismo passa para a corrente sanguínea do hospedeiro, ocorrem mudanças morfológicas, na expressão de genes, e, além disso, ocorre um estresse térmico, pois na corrente sanguínea a temperatura é maior, sendo acionado um aumento na síntese das proteínas de choque térmico. Cepas mutantes na proteína ClpP mostram-se sensíveis a altas temperaturas, corroborando a ideia de que esta proteína protege a bactéria de situações adversas. A ClpP se localiza no citoplasma, no entanto, em virtude do choque térmico, esta é direcionada para a parede celular (KWON *et al.*, 2004). Em seus estudos, CAO *et al.* (2008) confirmaram a localização da ClpP na superfície.

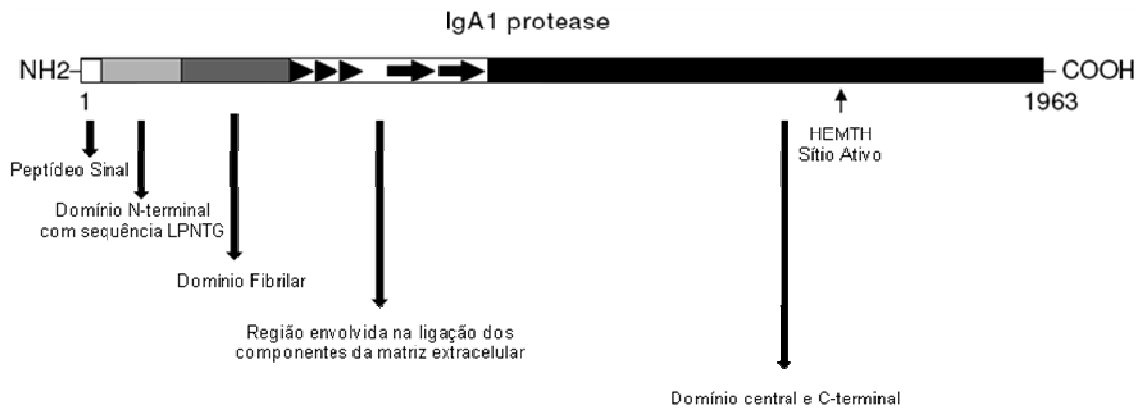
A ClpP desempenha alguns papéis na virulência de *S. pneumoniae*. Um deles é a modulação da expressão de genes de virulência, assim como ocorre com outras proteínas de choque térmico. Sugere-se também que a ClpP seja requerida para a sobrevivência intracelular e aumente a estabilidade de mRNA a altas temperaturas (KWON *et al.*, 2004). Cepas mutantes nesta proteína, quando comparadas a cepas selvagens, mostraram-se menos virulentas, o que reforça a ideia de que a ClpP seria uma excelente candidata a antígeno para a composição de uma vacina (KWON *et al.*, 2004).

Em vários trabalhos, a ClpP protease tem sido clonada e expressa, e a sua imunogenicidade tem sido testada em modelo animal. KWON *et al.* (2004) demonstraram que ratos imunizados com ClpP produzem uma forte e específica resposta de anticorpos ao antígeno. CAO *et al.* (2007) também demonstraram que a ClpP protease, expressa em *E. coli*, conferiu proteção em modelo animal contra infecções pneumocócicas, sendo que, quando misturada a outras duas proteínas, PspA e PspC, produziu níveis ainda maiores de proteção. Além destes estudos, CAO *et al.* (2009) demonstraram a capacidade da proteína ClpP de proteger sozinha contra infecções invasivas de *S. pneumoniae* pertencentes a diversos sorotipos.

### 2.1.2.2 IgA1 protease

A IgA1 protease é uma enzima que cliva ligações peptídicas específicas da imunoglobulina A, subclasse 1, conhecida como IgA1. Esta protease é um importante fator de virulência que está diretamente ligado à colonização e infecção das mucosas (MISTRY E STOCKLEY, 2006), sendo de grande importância para a resistência de *S. pneumoniae* à resposta imunológica do hospedeiro. Além disso, a IgA1 protease é expressa em todas as cepas de *S. pneumoniae* (PAOLIS *et al.*, 2007).

A IgA1 protease de *S. pneumoniae* é uma grande proteína de superfície (ROMANELLO *et al.*, 2006), identificada como uma Zn-metaloprotease (CHIAVOLINI *et al.*, 2003) pois apresenta a seqüência HEMMH a qual é o sítio ativo da enzima. Esta protease possui 1964 aminoácidos, podendo variar de 130 a 200 kDa, mantendo a atividade proteolítica, devido a modificações pós-traducionais na parte N-terminal (ROMANELLO *et al.*, 2006). Esta parte N-terminal da cadeia apresenta três regiões hidrofóbicas. A primeira região é uma seqüência que se apresenta em procariotos como um peptídeo sinal, a segunda junto com a seqüência LPNTGS funciona como ponto de fixação da protease na parede celular (POULSEN *et al.*, 1998; ROMANELLO *et al.*, 2006). A cadeia apresenta ainda um domínio fibrilar de 200 aminoácidos, uma região com segmentos envolvidos na ligação de componentes da matriz extracelular e ainda domínio central e C-terminal, onde se localiza o sítio ativo da enzima (PAOLIS *et al.*, 2007). Uma representação esquemática da IgA1 protease é mostrada na figura 2.4. Sabe-se que a IgA1 protease é capaz de se transportar sozinha para fora da célula pois ela possui em sua cadeia todas as informações necessárias para este transporte (MISTRY E STOCKLEY, 2006).



**Figura 2.4** Representação esquemática de IgA1 protease típica de *Streptococcus*. Adaptado de PAOLIS *et al.* (2007)

O substrato da enzima IgA1 protease, a IgA1 humana, está presente nas mucosas, e, no soro, é o principal isotipo de IgA (PAOLIS *et al.*, 2007). A IgA1 protease é altamente heterogênea entre as cepas de um microrganismo e acredita-se que isto seja um mecanismo de proteção do microrganismo contra as defesas do hospedeiro. Apesar desta alta heterogenicidade, esta proteína é capaz de provocar uma forte resposta imunológica *in vivo* (MISTRY E STOCKLEY, 2006).

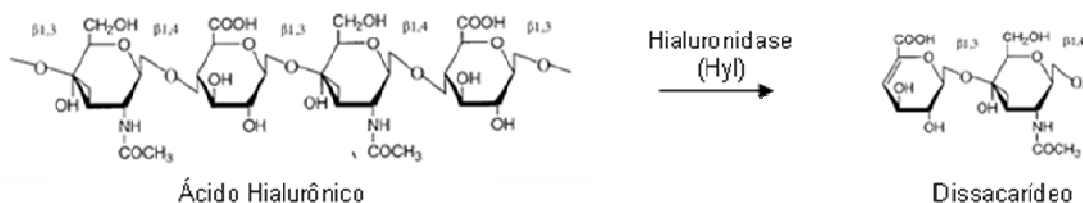
Em seus estudos, ROMANELLO *et al.* (2006) clonaram e expressaram três diferentes fragmentos de IgA1 protease. Por combinação destes fragmentos, obtiveram a proteína praticamente inteira, com exceção dos primeiros aminoácidos, os quais compreendem a região do peptídeo sinal e a parte N-terminal correspondente ao domínio de fixação da proteína. Apenas o fragmento do final da sequência, correspondendo aos aminoácidos 1032 a 1964, não teve atividade enzimática. Além disso, os três fragmentos se mostraram imunogênicos quando submetidos a soros de pacientes infectados com *S. pneumoniae*, indicando que a IgA1 protease pode ser um antígeno na formulação de uma vacina contra *S. pneumoniae*.

### 2.1.2.3 Hialuronato liase (*Hyl*)

A hialuronidase é uma enzima capaz de degradar ácido hialurônico, o qual é um importante componente e o mais abundante glicosaminoglicano da matriz extracelular de tecidos conectivos (AKHTAR E BHAKUNI, 2004). O ácido hialurônico é

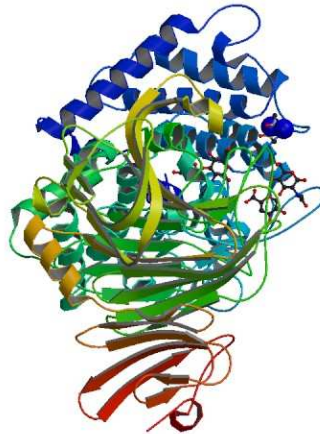
encontrado nos tecidos e fluidos de seres humanos e animais superiores e, além da função estrutural, também está presente em processos como fertilização, desenvolvimento embrionário, diferenciação celular, cicatrização e metástase de células tumorais (JEDRZEJAS, 2001; AKHTAR E BHAKUNI, 2004).

A hialuronidase encontrada em bactérias é a hialuronato liase (Hyl), que catalisa e quebra as ligações glicosídicas do ácido hialurônico pelo processo de  $\beta$ -eliminação, resultando como produto final principalmente dissacarídeos, como mostrado na figura 2.5. Além da Hyl existem pelo menos mais dois tipos de hialuronidasas que degradam o ácido hialurônico por mecanismos diferentes (JEDRZEJAS, 2001; AKHTAR E BHAKUNI, 2004).



**Figura 2.5** Degradação do ácido hialurônico e formação da unidade dissacarídica (principal produto da ação da Hyl bacteriana sobre o ácido hialurônico). Adaptado de AKHTAR E BHAKUNI (2004).

A estrutura da Hyl é dividida em dois domínios aproximadamente do mesmo tamanho, o domínio  $\alpha$ -hélice e o domínio de folhas  $\beta$ -pregueadas (JEDRZEJAS, 2001; AKHTAR E BHAKUNI, 2004). O domínio  $\alpha$ -hélice é responsável pela ligação e degradação do substrato, enquanto o domínio de folhas  $\beta$ -pregueadas é provavelmente responsável pela manutenção da fenda catalítica e por modular o acesso a mesma (JEDRZEJAS, 2001). A estrutura da Hyl pode ser visualizada na figura 2.6.



**Figura 2.6** Estrutura da Hyl de *Streptococcus pneumoniae* complexada com dissacarídeo de ácido hialurônico, depositada no PDB (*Protein Data Bank*) sob o número de identificação 1C82 (PONNURAJ E JEDRZEJAS, 2000).

Hialuronidase é uma das principais proteínas de superfície de *S. pneumoniae* que pode contribuir para sua virulência. Sabe-se que pelo menos uma parte da hialuronidase é liberada nos tecidos do hospedeiro com o objetivo de facilitar a invasão, aumentando a permeabilidade dos tecidos através da degradação de componentes da matriz extracelular, ou seja, esta enzima está diretamente envolvida nos mecanismos de invasão de *S. pneumoniae* (JEDRZEJAS, 2001; AKHTAR E BHAKUNI, 2004). Como a hialuronidase está diretamente envolvida nos mecanismos de invasão do microrganismo, esta representa outra alternativa para composição de uma vacina (JEDRZEJAS, 2001). A hialuronidase já foi clonada e expressa por BERRY *et al.* (1994), os quais obtiveram a proteína expressa com aproximadamente 89kDa, e com atividade enzimática. No entanto, experimentos com cepas mutantes nesta proteína mostraram que estes mutantes não têm sua virulência atenuada, o que indica que a hialuronidase pode ser utilizada em uma vacina contra *S. pneumoniae* apenas quando estiver combinada com outros fatores de virulência (JEDRZEJAS, 2001). Outro ponto que reforça o uso desta enzima na composição de uma vacina é o fato da maioria das cepas de *S. pneumoniae* produzirem a Hialuronato liase. (AKHTAR e BHAKUNI, 2004).

## 2.2 Sistemas de expressão de proteínas recombinantes

Para atender a demanda industrial, proteínas de interesse podem ser produzidas em altos níveis através da utilização da engenharia genética (DEMAIN E VAISHNAV, 2009). A expressão de proteínas recombinantes pode ser realizada em bactérias, em leveduras ou em células de insetos, plantas ou mamíferos. Para a expressão da proteína de forma bem sucedida, a escolha do sistema de expressão é de fundamental importância e deve levar em consideração a estrutura da proteína, sua funcionalidade e complexidade, e também a produtividade desejada (LARENTIS *et al.*, 2006; DEMAIN E VAISHNAV, 2009). Nesta seção serão apresentados alguns aspectos importantes na clonagem e expressão de proteínas recombinantes e, em especial, das duas bactérias utilizadas como sistemas de expressão neste trabalho, *Escherichia coli* e *Bacillus subtilis*.

Para a produção da proteína heteróloga de interesse, o gene da proteína a ser expresso deve primeiramente ser inserido em um vetor de expressão. Os vetores de expressão são capazes de se auto-replicar e regular a expressão dos genes neles codificados. As estruturas que formam um vetor são: promotores, origens de replicação, sítios iniciadores e terminadores tanto da transcrição como da tradução, marcadores seletivos, sítios de múltipla clonagem ou de ligação do gene isolado. Para a clonagem de genes são utilizados três tipos de vetores: plasmídeos, bacteriófagos e cosmídeos (LARENTIS *et al.*, 2006). Como neste trabalho foram utilizados plasmídeos como vetores de expressão, estes serão o foco de uma breve revisão.

Com capacidade de replicação autônoma, plasmídeo é um DNA circular não cromossomal de fita dupla. Eles podem ser naturalmente encontrados em bactérias e, no processo de divisão celular, são replicados e passados para as novas células geradas. Já os promotores presentes nos plasmídeos regulam a expressão do gene. Os plasmídeos são muito utilizados para clonagem de genes com até 5 kb (LARENTIS *et al.*, 2006).

Os plasmídeos podem ser classificados de acordo com seu número de cópias, que é o número de cópias do plasmídeo por célula, sendo que esta classificação pode variar muito na literatura. De acordo com FRIEHS (2004), classificam-se como

plasmídeos com baixo número de cópias aqueles que apresentam 1 a 10 plasmídeos por célula, plasmídeos com médio número de cópias apresentam 10 a 20 plasmídeos por célula e plasmídeos de alta cópia podem chegar a aproximadamente 700 cópias por célula. Já SCHUMANN E FERREIRA (2004) mostram uma classificação um pouco diferente, onde plasmídeos de muito alto número de cópias estão presentes em mais de 100 plasmídeos por célula, plasmídeos de alto número de cópias apresentam 15 a 60 plasmídeos por célula, plasmídeos de médio número de cópias apresentam cerca de 10 plasmídeos por célula e plasmídeos de baixo número de cópias que apresentam de 1 a 2 plasmídeos por célula.

Através de marcadores seletivos inseridos no vetor é realizada a seleção das cepas contendo o vetor de interesse. Entende-se por marcadores seletivos genes que conferem alguma característica específica que diferencia as células transformadas das demais. Normalmente, os plasmídeos apresentam um gene que confere resistência a algum antibiótico, como forma de selecionar as células com o plasmídeo das demais células que não o possuem (pressão seletiva) (LARENTIS *et al.*, 2006).

A produção de proteína recombinante não é um processo natural da célula hospedeira, sendo assim, muitas vezes causa problemas à célula, dificultando a produção da proteína heteróloga. Desta forma, a produção da proteína recombinante envolve vários fatores, como características da proteína a ser produzida, vetores e hospedeiros utilizados, escolha do processo utilizado e purificação da proteína produzida, sendo que todos estes fatores precisam ser profundamente conhecidos. Ainda é necessário considerar o compartimento celular em que a proteína será expressa, assim como as estratégias utilizadas na purificação, já que a proteína pode ser ligada, durante a clonagem, a um polipeptídeo de fusão para facilitar a purificação (LARENTIS *et al.*, 2006). O nível de expressão da proteína pode ser influenciado por alguns fatores como as características estruturais do gene a ser expresso, a estabilidade e eficiência do mRNA, o enovelamento correto e eficiente da proteína, a toxicidade da proteína para a célula hospedeira, a degradação da proteína por proteases e também os códons utilizados na seqüência do gene (SCHUMANN E FERREIRA, 2004).



Para a expressão, a célula é cultivada sob pressão seletiva, na maioria dos casos através da adição de antibiótico no meio de cultivo e, desta forma, apenas as células que possuem o plasmídeo com o gene de resistência ao antibiótico poderão crescer. Um dos problemas do uso de antibiótico é a contaminação da biomassa e do produto, o que não seria aceitável para proteínas utilizadas para fins médicos (BANEYX, 1999; FRIEHS, 2004). Além disso, este método pode se tornar inviável em escala industrial, pois os custos da adição de antibióticos poderiam ser enormes e, para produtos com fins alimentícios e farmacêuticos, os antibióticos devem ser removidos no processo de purificação (FRIEHS, 2004).

### 2.2.1 *Escherichia coli* como sistema de expressão de proteínas recombinantes

A bactéria gram-negativa *E. coli* é um dos mais utilizados sistemas de expressão de proteínas recombinantes, inclusive em escala comercial (BANEYX, 1999; TERPE, 2006), isto porque, além de oferecer várias vantagens, como a capacidade de crescer rapidamente em altas concentrações celulares e em substratos baratos, possui genética bem caracterizada e existem vários vetores e cepas mutantes disponíveis comercialmente (BANEYX, 1999; PETI E PAGE, 2007). Além disso, o genoma de *E. coli* pode ser modificado com facilidade, as células de *E. coli* têm capacidade de acumular mais de 80% da sua massa seca em proteína recombinante e ainda sobrevivem em uma variedade de condições de cultivo. Esta bactéria pode produzir proteínas recombinantes funcionais não-glicosiladas, sendo que, normalmente, estas proteínas, quando possuem fins terapêuticos, são produzidas em *E. coli* ou em leveduras (DEMAIN E VAISHNAV, 2009).

Uma das desvantagens de utilizar *E. coli* como sistema de expressão é que em alguns casos esta cepa apresenta dificuldade de produzir proteínas heterólogas funcionais, biologicamente ativas e solúveis. A produção de proteínas complexas, com pontes de enxofre, múltiplas subunidades ou modificações pós-traducionais é dificultada neste microrganismo pela falta da maquinaria necessária para fazer estes tipos de modificações (BANEYX, 1999). Altas densidades celulares podem se tornar tóxicas devido à formação de acetato, o que pode ser contornado através de controles do

processo, como o controle de oxigênio e alimentação exponencial de glicose (DEMAIN E VAISHNAV, 2009).

A superprodução de proteínas heterólogas no citoplasma muitas vezes ocasiona a perda da conformação e tendência à agregação, formando agregados insolúveis, conhecidos como corpos de inclusão (BANEYX, 1999; SCHUMANN E FERREIRA, 2004; TERPE, 2006). Este fenômeno provavelmente ocorre, pois a capacidade do sistema de chaperonas - as quais mediam o correto enovelamento, a secreção das proteínas e o controle das proteínas com má formação através de ação proteolítica - é extrapolada durante a produção em altas densidades da proteína heteróloga (KURLAND E GALLANT, 1996; MERGULHÃO *et al.*, 2005). A formação de corpos de inclusão é vista por muitos como um problema, no entanto, há também aqueles que afirmam que a formação destes corpos de inclusão pode simplificar a purificação. Entretanto, após a purificação é necessário o re-enovelamento da proteína *in vitro*, o que pode se tornar um problema na produção de grandes quantidades da proteína na sua forma biologicamente ativa (BANEYX, 1999; SWARTZ, 2001). Além disso, estes agregados insolúveis não são compostos apenas por cadeias da proteína recombinante, eles possuem em sua composição diversas impurezas (SCHUMANN E FERREIRA, 2004). Algumas alternativas para reduzir a formação destes corpos de inclusão podem ser adotadas, como reduzir a temperatura de expressão (reduzindo a taxa de expressão), co-expressar as chaperonas, mudar algumas condições de cultivo, como o pH (TERPE, 2006).

Quando a proteína expressa em *E. coli* é destinada ao uso na saúde humana, o acúmulo de lipopolissacarídeos (LPS) por parte deste microrganismo pode se tornar uma dificuldade, porque os LPSs atuam como endotoxinas pirogênicas para humanos e outros mamíferos. Se as proteínas têm finalidade de uso humano, elas devem passar por uma etapa de purificação capaz de reter estas toxinas (TERPE, 2006).

Normalmente as proteínas são produzidas no citoplasma de *E. coli*, no entanto elas podem ser exportadas para o periplasma, para facilitar a purificação, através da fusão com um peptídeo que sinalize a translocação para o periplasma. Todavia, a produção no citoplasma normalmente apresenta rendimentos maiores (DEMAIN E

VAISHNAV, 2009), e o sinal de transporte da proteína para o periplasma nem sempre funciona (LARENTIS *et al.*, 2006).

Para superar as limitações apresentadas, muitas pesquisas têm sido feitas para melhorar o sistema de expressão em *E. coli*. O plasmídeo está fortemente associado à eficiência da produção da proteína heteróloga dependendo do promotor utilizado e sua regulação, das regiões de terminação e iniciação, da eficiência da sequência de ligação ao ribossomo, da origem de replicação e conseqüentemente do número de cópias do plasmídeo, dos marcadores de resistência utilizados, dos códons utilizados, da estabilidade do mRNA, da estabilidade da proteína e da cepa utilizada como hospedeira (MAKRIDES, 1996; SØRENSEN E MORTENSEN, 2005; JANA E DEB, 2005). Promotores com diversas características têm sido desenvolvidos para melhorar a eficiência do sistema utilizando *E. coli*. Um bom promotor deve ser forte, ser fortemente regulado, a indução deve ser simples, de baixo custo e independente dos componentes normalmente utilizados no meio de cultura (TERPE, 2006). Muitas vezes são utilizados promotores derivados do sistema *lac* de regulação da bactéria (BANEYX, 1999), muitos promotores foram construídos empregando o elemento regulador do sistema *lac* (TERPE, 2006). O sistema *lac* é formado por uma região composta por promotor/operador que antecede os genes a serem transcritos. Na ausência de um indutor, o repressor do sistema, codificado pelo gene *lacI*, liga-se ao operador, impedindo a transcrição (SCHUMANN E FERREIRA, 2004). Para alcançar altas concentrações de proteína, vários plasmídeos utilizam o sistema de promotor do bacteriófago T7 RNA polimerase (BANEYX, 1999). A enzima T7 RNA polimerase é capaz de estender cadeias cinco vezes mais rápido que a mesma enzima de *E. coli*. O gene que codifica a T7 RNA polimerase foi inserido no cromossomo da cepa BL21 (DE3) sob controle de um promotor derivado do sistema *lac*, permitindo forte indução da produção desta enzima pela adição de IPTG, mesmo na presença de glicose (TERPE, 2006).

Além de explorar as características genéticas das cepas de *E. coli* e dos plasmídeos utilizados, devem também ser utilizadas outras estratégias para maximizar o rendimento da proteína de interesse. Uma destas estratégias seria aumentar a densidade celular no cultivo, manipulando as variáveis: temperatura, composição do meio de

cultivo, modo de operação do bioreator e aeração (HANNING E MAKRIDES, 1998; BANEYX, 1999; JANA E DEB, 2005; SØRENSEN E MORTENSEN, 2005; PETI E PAGE, 2007).

### 2.2.2 *Bacillus subtilis* como sistema de expressão de proteínas recombinantes

A produção de proteínas heterólogas em outros hospedeiros, que não *E. coli*, tem se tornado cada vez mais freqüente, provavelmente devido ao aumento do conhecimento em genômica, que entre outras vantagens, permite comparar os códons utilizados pelo hospedeiro com os do organismo produtor original da proteína (TERPE, 2006).

O gênero *Bacillus* é o mais utilizado como sistema de expressão bacteriano, depois de *E. coli*, sendo utilizado para a expressão de muitas proteínas de importância farmacêutica. Quando comparadas a *E. coli*, as cepas de *Bacillus* apresentam a vantagem de não possuírem lipopolissacarídeos (LPS) na sua membrana externa. Outro fator importante é o fato de terem naturalmente a capacidade de secretar as proteínas para o meio de cultivo em uma única operação (TERPE, 2006).

O microrganismo gram-positivo *B. subtilis* possui, como uma de suas maiores vantagens, a capacidade de secretar as proteínas diretamente no meio de cultivo (LI *et al.*, 2004; FERREIRA *et al.*, 2005; TERPE, 2006; FU *et al.*, 2007). *B. subtilis* é capaz de crescer rápido em substratos simples e baratos (FERREIRA *et al.*, 2005). O genoma de *B. subtilis* já foi seqüenciado e esta cepa não produz exotoxinas ou endotoxinas prejudiciais. A capacidade de secretar as proteínas para o meio de cultivo e não formar corpos de inclusão facilita a purificação, elimina a necessidade de romper a célula para extrair a proteína, aumentando os rendimentos da produção (DEMAIN E VAISHNAV, 2009). Esta bactéria é de fácil manipulação e reconhecida como GRAS (*generally recognized as safe*) (LI *et al.*, 2004; FERREIRA *et al.*, 2005; FU *et al.*, 2007; DEMAIN E VAISHNAV, 2009). Além disso, como este microrganismo é utilizado industrialmente para a produção de muitas enzimas, a sua tecnologia de cultivo é bem conhecida e várias ferramentas genéticas têm sido desenvolvidas para o uso na produção de proteínas recombinantes (LI *et al.*, 2004). Sabe-se também que o sistema de expressão utilizando *B. subtilis* tem sido empregado para a produção de proteínas

heterólogas com atividades farmacológicas e imunológicas, inclusive para a expressão de antígenos com possível aplicação em vacina humana (FERREIRA *et al.*, 2005).

O uso de *B. subtilis* como sistema de expressão normalmente pode apresentar dois inconvenientes que limitam a sua aplicação: a instabilidade dos plasmídeos, tanto estrutural quanto segregacionista, e o baixo nível de proteína produzida (LI *et al.*, 2004; FERREIRA *et al.*, 2005). Outras limitações apresentadas na expressão em *B. subtilis* são o correto enovelamento da proteína, os elevados níveis de proteases naturalmente secretadas para o meio, o processamento do peptídeo sinal e secreção da proteína. O correto enovelamento pode se tornar um problema na expressão de altos níveis de proteína. *B. subtilis* apresenta chaperonas intracelulares e extracitoplasmáticas. As chaperonas intracelulares estão envolvidos no enovelamento, minimizando a agregação e mantendo a conformação adequada das proteínas para a secreção. Devido à atividade limitada das chaperonas intracelulares, a proteína pode formar agregados insolúveis no citoplasma. A proteína PrsA é uma chaperona extracitoplasmática responsável pelo correto enovelamento da proteína madura em uma conformação ativa e estável. Para aumentar a produção de proteína heteróloga, um nível celular alto destas chaperonas pode ajudar (LI *et al.*, 2004).

Uma das principais desvantagens deste microrganismo são os níveis elevados de proteases também secretadas para o meio, no entanto, este tem sido o foco de muitos desenvolvimentos nos últimos anos (LI *et al.*, 2004; TERPE, 2006). Atualmente já existem cepas deficientes em 6 ou 8 proteases. A cepa WB600 é deficiente em seis das principais proteases extracelulares de *B. subtilis* (protease neutra A, subtilisina, protease extracelular, metaloprotease, bacilopeptidase F e protease neutra B) e apresenta um nível muito baixo de atividade de proteases quando comparada à cepa selvagem (WU *et al.*, 1991). Já a cepa WB800 é deficiente em 8 proteases extracelulares (TERPE, 2006).

O processo de transporte das proteínas através da membrana plasmática ocorre em três estágios: reconhecimento, deslocamento através da membrana e liberação no meio de cultivo. O reconhecimento é feito no citoplasma, através de reconhecimento do peptídeo sinal. A proteína é transportada através da membrana por um canal formado por um transportador específico, a *translocase*. Então o peptídeo sinal é removido pela

peptidase sinal e a proteína madura é liberada do transportador. As peptidases sinal podem se tornar um fator limitante no processo de produção da proteína (LI *et al.*, 2004), assim como a estrutura do peptídeo sinal. Uma das possibilidades para melhorar a quantidade de proteína secretada seria a construção de vetores de expressão nos quais fossem incluídas seqüências sinalizadoras. Sabe-se também que o aparato celular para secreção de proteínas pode ser saturado pela expressão de altos níveis de proteína, diminuindo a produção da proteína heteróloga (FU *et al.*, 2007).

### 2.2.3 Estabilidade plasmidial

Durante a produção de proteínas recombinantes uma das questões fundamentais é a estabilidade plasmidial, principalmente durante o escalonamento do processo ou produção industrial (GUPTA *et al.*, 1995). A estabilidade do plasmídeo é caracterizada em dois tipos: a estrutural e a segregacionista. A instabilidade estrutural é o resultado de alterações estruturais no plasmídeo durante o cultivo, como alterações da seqüência de nucleotídeos de um plasmídeo, inserções, deleções ou rearranjos no DNA plasmidial. Já a segregação plasmidial é a perda do plasmídeo durante a divisão celular, resultando em uma das células filha sem o plasmídeo (GUPTA *et al.*, 1995; FRIEHS, 2004). Na maioria das vezes, a expressão “instabilidade plasmidial” se refere à segregação plasmidial. A replicação correta dos plasmídeos durante a divisão celular é um problema central em sistemas recombinantes. A taxa de crescimento específico de células sem plasmídeo é maior quando comparado à de células com plasmídeo, logo, se houver algumas células sem plasmídeo no início do cultivo, estas células crescerão mais que as células com plasmídeo e se tornarão dominantes no cultivo, levando a culturas mistas. Desta forma, a produtividade da proteína recombinante será drasticamente afetada (FRIEHS, 2004). Segundo FRIEHS (2004), a segregação plasmidial é tão importante que, sem um estudo para determiná-la no processo em questão, nenhuma conclusão real pode ser feita sobre a produtividade da proteína recombinante.

Muitos fatores podem influenciar a estabilidade plasmidial. Alguns destes fatores são a taxa de crescimento celular, o número de cópias do plasmídeo, o tamanho do inserto e do plasmídeo, o nível de expressão da proteína recombinante, a toxicidade da proteína recombinante, a formulação do meio de cultivo, a concentração de oxigênio

dissolvido, o pH, a temperatura e o modo de operação do cultivo (GUPTA *et al.*, 1995; FRIEHS, 2004; LARENTIS *et al.*, 2006; XU *et al.*, 2006). Alguns autores afirmam que a carga metabólica imposta à célula pela replicação do plasmídeo, transcrição do gene e síntese da proteína recombinante é responsável pela segregação plasmidial (CORCHERO E VILLAVERDE, 1998).

A composição do meio de cultivo pode afetar a estabilidade plasmidial através de diferentes rotas metabólicas e sistemas regulatórios, isto porque a composição do meio influencia as atividades metabólicas da célula (LARENTIS *et al.*, 2006). Segundo GUPTA *et al.*, (1995) meios de cultura mais complexos levam a diminuição da estabilidade plasmidial. Sabe-se também que cultivos em meios mínimos apresentam menor segregação plasmidial que aqueles em meios com aminoácidos, além disso, a adição de extrato de levedura diminui a estabilidade plasmidial (FRIEHS, 2004)

Influenciam também na segregação plasmidial o mecanismo de replicação (FRIEHS, 2004; XU *et al.*, 2006) e o tamanho do plasmídeo, sendo que plasmídeos maiores podem apresentar maior instabilidade segregacionista (FRIEHS, 2004). O número de cópias do plasmídeo também tem influência sobre a segregação, visto que a estabilidade segregacionista significa que pelo menos um plasmídeo esteja presente em cada célula depois da divisão celular, então, um alto número de cópias por célula aumentaria as chances de pelo menos um plasmídeo estar presente na célula (FRIEHS, 2004; LARENTIS *et al.*, 2006). No entanto, alguns autores relatam que um número muito alto de cópias do plasmídeo pode aumentar a perda do plasmídeo (BANEYX, 1999).

Condições de cultivo, como a concentração de oxigênio dissolvido e o pH, podem influenciar na segregação. Sabe-se que a diminuição da concentração de oxigênio dissolvido leva a um aumento da instabilidade do plasmídeo (FRIEHS, 2004). A disponibilidade de oxigênio pode afetar a taxa de crescimento, além de apresentar outros efeitos complexos sobre as rotas metabólicas. A replicação e a transcrição de um plasmídeo com múltiplas cópias requer uma grande quantidade de energia, porém, a diminuição da concentração de oxigênio no meio diminui a geração de energia pelo catabolismo celular e pode afetar a replicação e partição do plasmídeo. A temperatura

também afeta a estabilidade plasmidial, já que esta afeta a taxa de crescimento celular, as rotas metabólicas e os sistemas regulatórios (LARENTIS *et al.*, 2006). Além disso, a perda do plasmídeo pode aumentar quando o cultivo de células é feito em altas densidades ou de forma contínua (BANEYX, 1999).

A indução da expressão da proteína sempre diminui a estabilidade plasmidial, provavelmente devido aos efeitos tóxicos da proteína para a célula (FRIEHS, 2004; LARENTIS *et al.*, 2006; XU *et al.*, 2006). Quando o nível de expressão é aumentado, a estabilidade plasmidial tende a diminuir e isto pode ser ocasionado pela repressão da replicação (LARENTIS *et al.*, 2006). Os dados mostrados por XU *et al.* (2006) indicam que a instabilidade plasmidial é muito mais dependente da indução em si do que da presença do antibiótico (pressão seletiva). Além disso, o modo de indução do promotor também tem grande influência sobre a estabilidade. Por exemplo, sabe-se que a indução com IPTG apresenta maior taxa de perda do plasmídeo que aquela feita com lactose (FRIEHS, 2004).

### **2.3 Planejamento de experimentos em sistemas recombinantes**

As variáveis que influenciam na expressão de proteínas recombinantes são numerosas e interagem umas com as outras. A estratégia normalmente aplicada na área para avaliar os fatores que influenciam o processo é variar um fator por vez mantendo os demais constantes (SWALLEY *et al.*, 2006). No entanto, esta estratégia pode não ser eficiente, não permitindo avaliar as interações entre as variáveis do processo e acarretando em um número maior de experimentos. As técnicas de planejamento de experimentos reduzem o número de experimentos envolvidos, diminuem o tempo gasto e o custo final do processo, melhoram a qualidade das informações obtidas, permitem analisar os fatores simultaneamente e suas interações, avaliar os erros experimentais e ainda otimizar mais de uma variável ao mesmo tempo (RODRIGUES E IEMMA, 2005).

O planejamento de experimentos tem sido aplicado muito recentemente na avaliação e otimização da expressão de proteínas recombinantes em bactérias. Esta técnica ainda não é um procedimento usual para avaliação dos fatores envolvidos no



processo de produção de proteína recombinantes, e seu potencial pode ser mais explorado para a avaliação da composição dos meios de cultura (NIKEREL *et al.*, 2005; NIKEREL *et al.*, 2006; REN *et al.*, 2006; ZHANG *et al.*, 2006) e para as variáveis envolvidas na indução, como temperatura, tempo pós-indução, concentração celular, concentração do indutor e cepa empregada na expressão (LEÓN *et al.*, 2004; WANG *et al.*, 2005; CAO *et al.*, 2006; SWALLEY *et al.*, 2006; MALDONADO *et al.*, 2007; ISLAM *et al.*, 2007; LO *et al.*, 2007).

## 3 METODOLOGIA

### 3.1 Seleção da proteína

Neste item, além de descrever os procedimentos para a seleção da proteína, para um melhor entendimento das análises realizadas, também serão descritos alguns termos das ferramentas de bioinformática utilizadas neste trabalho.

Através de revisão da literatura foram pré-selecionadas 3 proteínas: ClpP, IgA1 protease e Hialuronato Liase (Hyl). Para seleção da proteína utilizada neste trabalho, dentre estas três pré-selecionadas, foram feitas análises através de ferramentas de bioinformática e também foram levados em consideração dados obtidos da literatura. A seleção da proteína utilizou os seguintes critérios: conservação e prevalência entre os sorotipos de *S. pneumoniae*, imunogenicidade e presença de peptídeos sinais e regiões transmembranares – o que dificultaria a clonagem e a expressão da proteína de forma solúvel, principalmente em *E. coli*.

Para avaliar a conservação e prevalência das proteínas entre as cepas de *S. pneumoniae*, análises de bioinformática foram utilizadas. Primeiramente, as seqüências de nucleotídeos e aminoácidos das proteínas, pertencentes à cepa *S. pneumoniae* CGSP14, sorotipo 14, foram buscadas do *GenBank*, um banco de dados onde se encontram seqüências públicas disponíveis de nucleotídeos e proteínas, sendo que este pode ser acessado pelo sitio do NCBI (*National Center for Biotechnology*). As seqüências foram então submetidas a um alinhamento com outras disponíveis no banco de dados, através da ferramenta BLAST (*Basic Local Alignment Search Tool*), um algoritmo para fazer alinhamentos (processo de comparar duas ou mais seqüências buscando o máximo de similaridade ou conservação entre elas) entre a seqüência em estudo e outras depositadas no banco de dados, fazendo um alinhamento local, comparando uma base com outra base, ou um aminoácido com outro (ALTSUL *et al.*, 1990). Existem vários tipos de BLAST, comparando seqüências de proteínas (blastp), seqüências de nucleotídeos (blastn), e seqüências de nucleotídeos com proteínas ou vice-versa (blastx e tblastn). Neste estudo foram realizados alinhamentos utilizando blastp com o objetivo de identificar outras seqüências homólogas, disponíveis nos

bancos de dados, das proteínas de *S. pneumoniae* em questão. As seqüências encontradas foram utilizadas para a realização de um múltiplo alinhamento através do programa CLUSTAL W (THOMPSON *et al.*, 1994), capaz de fazer alinhamentos globais de múltiplas seqüências. Desta forma, foi possível identificar o grau de conservação e prevalência das proteínas nas diferentes cepas de *S. pneumoniae*.

Para a obtenção de dados gerais das proteínas, bem como de regiões transmembranares, o programa CLC Workbench 4 foi utilizado. Para avaliar a presença de regiões transmembranares é realizada uma predição de regiões hidrofóbicas na seqüência de aminoácidos da proteína.

As possíveis seqüências sinalizadoras (peptídeos sinais) presentes na proteína foram verificadas através do programa SignalP 3.0. Este programa permite fazer predições de peptídeos sinais e da localização do sítio de clivagem entre a proteína e o peptídeo sinal, de gram-negativos, gram-positivos e eucariotos, utilizando combinações de redes neurais e de modelos escondidos de Markov (BENDTSEN *et al.*, 2004).

Com o auxílio de dados disponíveis na literatura, as proteínas foram avaliadas quanto à sua imunogenicidade, ou seja, a capacidade de uma substância induzir uma resposta imunológica eficaz.

### **3.2 Clonagem e expressão do gene *clpP* de *S. pneumoniae* em *Escherichia coli***

#### **3.2.1 Extração de DNA genômico de *Streptococcus pneumoniae***

O DNA genômico foi extraído da cepa *Streptococcus pneumoniae* 113/95, sorotipo 14, depositada no Instituto Adolfo Lutz. Sabe-se que um dos sorotipos que mais causa doenças pneumocócicas invasivas no mundo é o sorotipo 14 (DING *et al.*, 2009). A extração foi feita a partir de um cultivo de 16h a 37°C e 100rpm, em meio TSB (*Tryptic Soy Broth*), com concentração de 47,24 g/L (pH 7,6), suplementado com 0,630 g/L de MgSO<sub>4</sub>.7 H<sub>2</sub>O, 0,236 g/L de Hidrocloreto de cisteína e 0,0205 g/L de CaCl<sub>2</sub> . 2H<sub>2</sub>O. Foram pesadas 2,25 g de células úmidas deste cultivo e a elas adicionados 40 mL de TE (Tris 10mM, EDTA 1mM, pH8,0), 20% m/v de sacarose, 200

mg de lisozima, sendo que a lisozima age na quebra dos carboidratos da membrana celular. A amostra foi incubada por 1 hora a 37°C. Depois da incubação a amostra foi dividida em duas alíquotas de 24 mL e a cada alíquota adicionados 0,2 mg/mL de proteinase K e 1,2 mL de SDS 20%. As amostras foram novamente incubadas a 56°C por 1 hora, sendo agitadas delicadamente a cada 15 minutos, evitando o rompimento do DNA total celular. A proteinase K age na desnaturação de proteínas que se ligam ao DNA e tem sua atividade aumentada com a adição de SDS e incubação em temperaturas elevadas (50°C-60°C). Posteriormente foram adicionados 0,2 mg/mL de RNAse em cada alíquota e então estas incubadas a 37°C por 30 minutos. Após estes procedimentos de lise celular, as proteínas se encontravam solúveis na fase aquosa junto com o DNA.

A partir desta etapa foram iniciadas extrações sucessivas com fenol/clorofórmio/álcool isoamílico (25:24:1). Para cada 24 mL da amostra foram adicionados 24 mL de fenol/clorofórmio/álcool isoamílico. A mistura resultante foi homogeneizada delicadamente, evitando o rompimento do DNA, e centrifugada a 4000rpm por 10 minutos em centrífuga Eppendorf 5810R (rotor A-4-62). A fase aquosa superior foi retirada e utilizada para fazer nova extração, repetindo o procedimento com fenol/clorofórmio/álcool isoamílico. Após as duas extrações, clorofórmio foi adicionado à fase aquosa superior na mesma proporção de volume e a mistura centrifugada por 10 minutos a 4000 rpm em centrífuga Eppendorf 5810R (rotor A-4-62). A fase aquosa superior foi retirada e a ela adicionado 0,1 volumes de acetato de sódio 3M (pH5,3) e 0,7 volumes de isopropanol. O material resultante foi incubado por 16 horas a 10°C e posteriormente centrifugado por 20 minutos a 12000 rpm em centrífuga Eppendorf 5810R (rotor A-4-62). O sobrenadante foi retirado cuidadosamente e descartado. O precipitado foi lavado com 1 mL de etanol 70%, incubado a temperatura ambiente por 5 minutos e centrifugado por 5 minutos a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). Após a retirada do sobrenadante, o precipitado contendo DNA foi submetido à secagem em concentrador centrífugo a vácuo (Vacufuge<sup>TM</sup>Eppendorf) a 30°C por 3 minutos. Após este procedimento foram adicionados 50 µL de TE (Tris 10 mM, EDTA 1 mM, pH8,0) à amostra e esta incubada a 37°C por 15 minutos. Como no início do procedimento a amostra foi dividida em duas alíquotas, foram obtidas duas amostras de DNA extraído, denominados amostra 3 e amostra 4. Ao final, as amostras foram estocadas a -20°C.

A análise das amostras de DNA total extraído foi realizada através de eletroforese em gel de agarose 0,8% m/v utilizando solução tampão de corrida TAE (0,04 M de Tris-baseÁcido acético glacial e 0,01 M de EDTA). Foram utilizadas na eletroforese as amostras da extração de DNA total, concentradas e diluídas 10 e 100 vezes. As amostras de DNA extraído e diluído 100 vezes foram quantificadas no equipamento Qubit Fluorometer (Invitrogen). Após a confirmação da extração do DNA em gel de agarose e quantificação, este foi utilizado como molde para o isolamento e amplificação do gene que codifica a proteína ClpP através da técnica de reação em cadeia da polimerase (PCR), a qual será descrita na próxima seção.

### 3.2.2 Amplificação do gene através da técnica de PCR

A amplificação do gene *clpP* foi feita utilizando a técnica de PCR (Reação em Cadeia da Polimerase). A Reação em Cadeia da Polimerase foi desenvolvida na metade da década de 1980 por Kary Mullis. Esta técnica é capaz de produzir cópias de regiões específicas do DNA, ou seja, é capaz de amplificar apenas a região do DNA correspondente ao gene desejado, produzindo muitas cópias (WATSON *et al.* 1992). Esta reação simula *in vitro* a replicação da molécula de DNA que ocorre *in vivo* utilizando a enzima DNA polimerase, oligonucleotídeos iniciadores e temperatura (WATSON *et al.* 1992; LARENTIS *et al.*, 2006). A reação ocorre em vários ciclos, sendo que cada ciclo é composto por basicamente 3 etapas. Na primeira etapa, ocorre a abertura da fita molde de DNA, na segunda etapa ocorre a ligação (anelamento) dos oligonucleotídeos iniciadores, que são complementares a cada uma das regiões 3' do gene a ser copiado, nas fitas simples. Na terceira etapa, ocorre a síntese das novas fitas de DNA pela extensão dos oligonucleotídeos iniciadores sob ação da DNA polimerase (LARENTIS *et al.*, 2006).

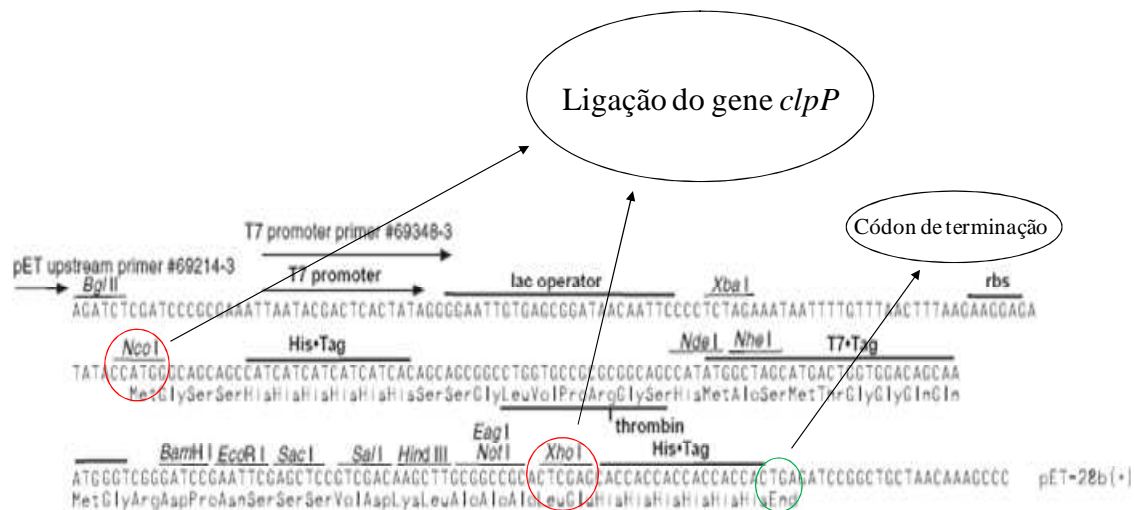
#### 3.2.2.1 Oligonucleotídeos iniciadores

Os oligonucleotídeos iniciadores utilizados neste trabalho para amplificação do gene *clpP* foram desenhados de acordo com o gene de *S. pneumoniae* CGSP14 e podem ser visualizados na tabela 3.1.

**Tabela 3.1** Oligonucleotídeos iniciadores utilizados na reação de PCR para amplificação do gene *clpP* a partir de DNA de *S. pneumoniae* CGSP14. Os sítios de clivagem das enzimas de restrição são mostrados sublinhados nas seqüências, sítio de *NcoI* na seqüência senso e *XhoI* na anti-senso.

Oligonucleotídeo iniciador	Seqüência
pET28bHisClpPf (Senso)	5' - <u>CCCATGGTTCCTGTAGTTATTGAACAAAC</u> - 3'
pET28bHisClpPr (Anti-senso)	5' - <u>CACTCGAGGTTCAATGAATTGTTGGC</u> - 3'

Os oligonucleotídeos iniciadores foram desenhados de forma a parearem com as extremidades 3' do gene de interesse. Nas extremidades 5' dos oligonucleotídeos iniciadores foram colocadas seqüências reconhecidas e clivadas por enzimas de restrição (sítios de restrição) de acordo com a estratégia de clonagem e vetor utilizados. As enzimas de restrição são capazes de reconhecer certas seqüências de nucleotídeos e catalisar a hidrólise das ligações fosfodiéster de bases específicas da seqüência (ZAHA *et al.*, 1996). Foram inseridas as seqüências dos sítios de restrição da enzimas *NcoI* e *XhoI* nos oligonucleotídeos senso e anti-senso, respectivamente. Estas enzimas foram escolhidas, de acordo com o vetor utilizado, neste caso pET28b (Novagen), de forma que, depois da clonagem, o vetor permita a expressão da proteína fusionada com uma cauda de histidina na sua extremidade carbóxi-terminal para facilitar posterior purificação. A figura 3.1 mostra a seqüência do sítio de clonagem e expressão do vetor pET28b, com uma representação da seqüência das enzimas de restrição onde foi inserido o gene *clpP*.



**Figura 3.1** Sítio de clonagem e expressão do vetor pET28b (Novagen). Os círculos vermelhos indicam os sítios das enzimas de restrição, também adicionados aos oligonucleotídeos iniciadores, onde o gene *clpP* amplificado irá se ligar. O círculo verde indica o códon de terminação, ou seja, a sinalização para o termino da tradução, possibilitando assim a expressão da proteína fusionada a cauda de histidina.

### 3.2.2.2 Reação de PCR

A amplificação do gene *clpP* foi realizada utilizando como molde a amostra 4 do DNA extraído e diluído 100 vezes, com concentração de aproximadamente 30 ng/μL. Foram utilizados os oligonucleotídeos iniciadores desenhados para esta reação e mostrados na seção anterior.

A reação foi feita empregando 30 ng do DNA extraído de *S. pneumoniae*, 1U de DNA polimerase de alta fidelidade (Platinum® Taq DNA Polymerase High Fidelity, Invitrogen), 2,5 μL de tampão específico para a enzima DNA polimerase utilizada (10X High Fidelity PCR Buffer, Invitrogen), 1 μL de MgSO<sub>4</sub> 50 mM, 200 μM de cada tipo de dNTP e 5 pmol de cada oligonucleotídeo (senso e anti-senso) específicos para esta amplificação em 25 μL de reação, sendo que foi utilizada como solvente da reação água ultrapura. Para cada reação de PCR, foi realizado um controle negativo, ou seja, uma reação idêntica, exceto pela substituição do volume empregado de DNA pelo mesmo volume de água ultrapura.

Primeiramente foi realizado um PCR em gradiente, no qual foram feitas quatro reações variando em cada uma delas a temperatura de anelamento dos oligonucleotídeos

iniciadores de 54°C até 56°C, com o objetivo de avaliar a temperatura ideal de anelamento. Nesta etapa as condições empregadas na reação foram as seguintes: 94°C por 5 minutos, 30 ciclos compostos por 3 etapas cada ciclo (94°C por 1 minuto; 54°C, 55,1°C, 55,9° e 56°C, uma temperatura para cada uma das 4 reações, por 30 segundos; 68°C por 1 minuto), 68°C por 7 minutos e resfriamento a 4°C.

Após a verificação da temperatura ideal de anelamento, em torno de 55°C, as reações seguintes de PCR feitas neste trabalho foram realizadas com as seguintes condições: 94°C por 5 minutos, 30 ciclos compostos por 3 etapas cada ciclo (abertura da dupla fita de DNA: 94°C por 1 minuto, anelamento dos oligonucleotídeos: 55°C por 30 segundos, extensão dos oligonucleotídeos e síntese de novas fitas de DNA: 68°C por 1 minuto), 68°C por 7 minutos e resfriamento a 4°C.

Tanto a verificação da melhor temperatura de anelamento, quanto a confirmação da amplificação do gene *clpP*, com o tamanho esperado, foram feitas através de eletroforese em gel de agarose 0,8% m/v.

### 3.2.2.3 Purificação do gene *clpP* amplificado

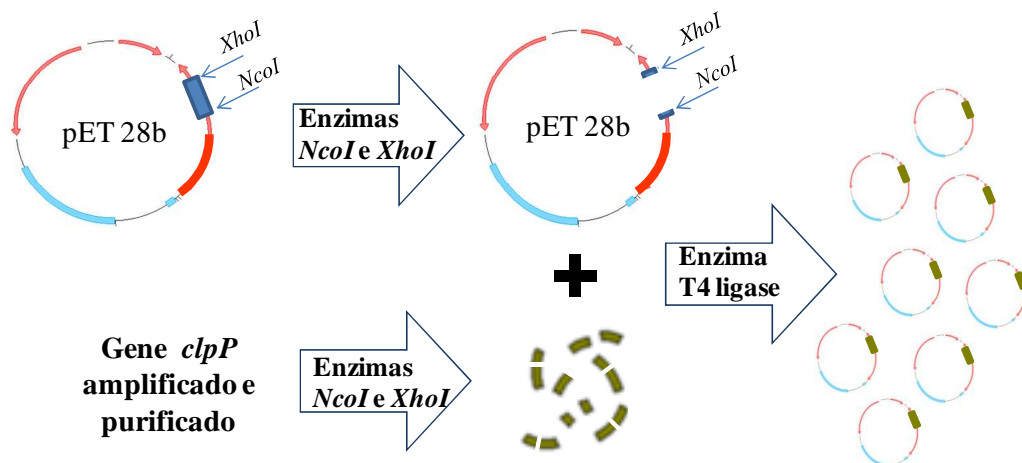
Após a confirmação da amplificação no tamanho esperado do gene *clpP* através de eletroforese em gel de agarose, o produto da reação de PCR foi purificado para a remoção de nucleotídeos, oligonucleotídeos iniciadores e possíveis amplificações não específicas com concentrações muito baixas, as quais não possibilitam a visualização em gel de agarose. A purificação foi feita através do *kit* Wizard® SV Gel and PCR Clean-Up System (Promega). Para tanto, o produto da reação de PCR foi submetido à eletroforese em gel de agarose 0,8%. Utilizando um bisturi, a banda de DNA foi cortada e retirada do restante do gel de agarose. Foram recortados pequenos pedaços da banda de DNA e pesados dentro de tubos de microcentrífuga de 1,5 mL, de forma que cada pedaço não apresentasse massa superior a 350 mg. Para cada 10 mg de gel pesados nos tubos de microcentrífuga, foram adicionados 10 µL de *Membrane Binding Solution* (4,5M de isotiocianato de guanidina, 0,5M de acetato de potássio), e estes agitados vigorosamente e incubados a 55°C até a completa solubilização da agarose. Depois da solubilização a solução resultante foi transferida para uma minicoluna com membrana,



capaz de reter fragmentos de DNA, e incubada à temperatura ambiente por 1 minuto. A minicoluna foi centrifugada a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11) por 1 minuto e a solução percolada foi descartada. A minicoluna foi então lavada com 700 µL de *Wash Solution* (10mM de acetato de potássio, pH 5,0; 80% de etanol; 16,7 µM de EDTA, pH 8,0) e centrifugada novamente a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11) por 1 minuto. Novamente o percolado foi descartado e a coluna lavada com 500 µL de *Wash Solution* (contendo etanol) e centrifugada novamente a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11) por 5 minutos. A solução que percolou foi descartada e a minicoluna vazia foi centrifugada por 1 minuto a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). A minicoluna foi transferida para um tubo de microcentrífuga limpo. Então foram adicionados 30 µL de água ultrapura autoclavada na minicoluna, com o objetivo de eluir os fragmentos de DNA retidos. A coluna foi incubada por 1 minuto à temperatura ambiente e centrifugada por 1 minuto a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). O percolado contendo uma solução aquosa dos fragmentos de DNA, retido no tubo de microcentrífuga, foi estocado a -20°C e analisado em gel de agarose 1,5%.

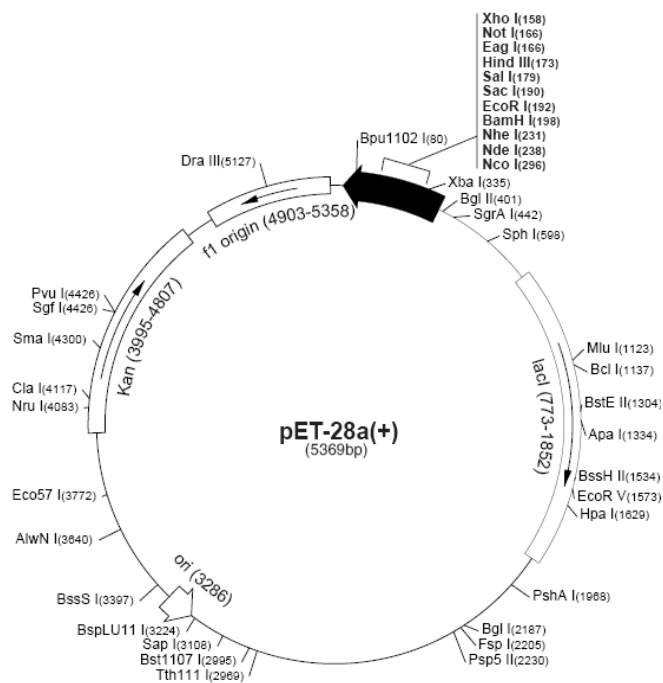
### 3.2.3 Clonagem

A estratégia de clonagem utilizada neste trabalho foi a clonagem clássica, empregando enzimas de restrição para clivar o gene e o vetor, e ligá-los através de DNA ligase. As enzimas DNA ligases são capazes de unir fitas de DNA clivadas por enzimas de restrição, ou seja, catalisam a formação de ligações fosfodiésteres de fragmentos de DNA (ZAHA *et al.*, 1996). Um desenho esquemático das etapas da clonagem é mostrado na figura 3.2.



**Figura 3.2** Desenho esquemático das etapas realizadas na clonagem (construção do vetor pET28b/*clpP*).

O vetor utilizado nesta clonagem, foi o pET28b (Novagen). O vetor pET28b utiliza o promotor do bacteriófago T7 RNA polimerase e marcador de resistência à canamicina. O mapa do vetor pET 28b é mostrado na figura 3.3.



**Figura 3.3** Mapa do vetor pET28a, igual ao pET28b (Novagen). O mapa mostra a origem de replicação (*ori*), o repressor *lac* (*lacI*), o gene de resistência à canamicina (*Kan*) e o sítio de múltipla clonagem (seta em preto sólido).

Após a purificação do gene *clpP* amplificado, foram realizadas as digestões do gene *clpP* e do vetor pET28b com as enzimas de digestão *XhoI* e *NcoI*. Na reação de digestão do gene *clpP*, empregando água ultrapura como solvente, foram utilizados aproximadamente 3 µg do gene, 5U de *XhoI* (Invitrogen), 3 µL de solução tampão *react2* (Invitrogen), em um volume total de reação de 30 µL. Na reação de digestão do vetor pET28b, empregando água ultrapura como solvente, foram utilizados aproximadamente 0,5 µg do gene, 5U de *XhoI* (Invitrogen), 4 µL de solução tampão *react2* (Invitrogen), em um volume total de reação de 40 µL. As reações foram incubadas por 2 horas a 37°C. Após este tempo foram retirados 2 µL de cada reação e adicionados 5U da enzima *NcoI* (Invitrogen) e 2U de enzima *XhoI*. À reação com o gene *clpP* foram ainda adicionados mais 1µL de solução tampão *react2* (Invitrogen) e o volume completado com água ultrapura a 40µL e à reação com o vetor foram ainda adicionados mais 1µL de solução tampão *react2* (Invitrogen) e o volume completado com água ultrapura a 50 µL. As reações foram incubadas por 1 hora e 30 minutos a 37°C. Após este tempo as reações foram incubadas por 20 minutos a 65°C para a inativação das enzimas. Os produtos das digestões foram purificados através do kit Wizard® SV Gel and PCR Clean-Up System (Promega) da mesma forma exemplificada na seção 3.2.2.3, exceto pelo fato de que a solução *Membrane Binding Solution* foi adicionada diretamente a solução com os produtos da digestão e não ao gel de agarose. As reações de digestão purificadas foram confirmadas em eletroforese de gel de agarose 0,8%.

Com o gene *clpP* e o vetor pET28b digeridos, foi realizada a construção do vetor pET28b/*clpP* através de ligação com a enzima T4 DNA ligase (Invitrogen), respeitando na reação de ligação a razão molar inserto(*clpP*)/vetor de 5/1. Foram realizadas duas reações de ligação, uma com o vetor e o gene, e outra (controle negativo) utilizando apenas o vetor. As reações foram compostas por 25 ng do gene *clpP* (exceto no controle negativo), 50 ng do vetor pET28b, 1U de T4 DNA ligase e 4 µL da solução tampão 5X *Ligase Reaction Buffer* (Invitrogen) sendo o volume completado até 20µL com água ultrapura autoclavada. As reações foram incubadas por 1 hora a 25°C. Esta reação foi utilizada para fazer a transformação em célula eletrocompetente *E. coli* TOP10.

### 3.2.4 Preparação de células eletrocompetentes

A cepa *Escherichia coli* TOP 10 (Invitrogen) foi utilizada como hospedeiro do vetor construído nas rotinas de clonagem e a cepa *Escherichia coli* BL21 Star (DE3) (Invitrogen) foi empregada como hospedeiro para o vetor construído nas rotinas de expressão da proteína. A cepa *Escherichia coli* TOP 10 é ideal para a propagação do plasmídeo, permitindo a replicação estável de plasmídeos de alto número de cópias e possui genótipo: F- *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\phi$ 80*lacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*araleu*) 7697 *galU galK rpsL* (StrR) *endA1 nupG*. Já a cepa *Escherichia coli* BL21 Star (DE3) possui genótipo: F- *ompT hsdS<sub>B</sub>* (*r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>*) *gal dcm rne131* (DE3). Esta cepa possui o gene que codifica a T7RNA polimerase sob controle do promotor *lacUV5*, o qual pode ser induzido por IPTG. Além disso, esta cepa é deficiente nas proteases *lon* e *OmpT*, reduzindo a degradação das proteínas heterólogas expressas, e possui uma mutação no gene *rne*, que codifica para a RNase E, aumentando a estabilidade do mRNA.

A transformação dos hospedeiros com o plasmídeo construído pET28b/*clpP* foi feita por eletroporação das células. Para isso as células passaram por um processo com lavagens sucessivas para a retirada do sal presente no meio de cultivo. Para obter as células eletrocompetentes, uma colônia isolada foi inoculada em meio LB (Luria-Bertani), 5g/L de extrato de levedura (BD), 10g/L de triptona (BD), 10g/L de NaCl (Merck), pH 7, e então, incubada a 37°C e 200 rpm por 16 horas. Depois deste período, 2,5 mL deste pré-inóculo foram inoculados em 250 mL de LB e incubado a 37°C e 200 rpm até a biomassa atingir a fase exponencial de crescimento ( $Abs_{600nm}$  aproximadamente 0,7). O cultivo foi então colocado no gelo por 15 minutos e, logo após, dividido em duas alíquotas, as quais foram centrifugadas por 20 minutos a 4000rpm e 4°C em centrífuga Eppendorf 5810R (rotor A-4-62). O sobrenadante foi descartado e as células de cada alíquota ressuspensas com o mesmo volume de água ultrapura gelada. As células lavadas foram divididas em alíquotas de 40 mL e foram novamente centrifugadas nas mesmas condições anteriores. O sobrenadante foi descartado e as células ressuspensas delicadamente em 20mL de água ultrapura gelada. As células foram centrifugadas (nas mesmas condições) e o sobrenadante descartado. Nesta etapa as células foram ressuspensas com 1 mL de glicerol 10% gelado e cada

duas alíquotas foram reunidas em uma, para aumentar o volume das amostras. As três alíquotas resultantes foram centrifugadas por 15 minutos a 4000rpm e 4°C em centrífuga Eppendorf 5810R (rotor A-4-62) e resuspensas com um total de 1mL glicerol 10% gelado (dividido entre as 3 alíquotas). As células foram aliquotadas em amostras de 100 µL e congeladas rapidamente em gelo seco com álcool etílico (-45°C). As alíquotas com as células eletrocompetentes foram estocadas a -70°C.

### 3.2.5 Transformação de *Escherichia coli*

As células de *E. coli* eletrocompetentes preparadas foram utilizadas como hospedeiros para o plasmídeo construído pET28b/*clpP*. A inserção deste plasmídeo nas células foi feita por eletroporação. O método da eletroporação consiste na inserção do plasmídeo na célula mediante um choque elétrico, o qual produz rompimentos transientes na membrana celular.

Para a eletroporação foram misturados delicadamente, para evitar a formação de bolhas, 10 µL da reação de ligação (seção 3.2.3) e 100 µL de *E. coli* TOP 10 eletrocompetente. A mistura foi transferida para uma cubeta de 0,2cm de espessura entre os eletrodos, e então submetida a uma descarga elétrica por cerca de 5ms, com voltagem de 2,5 kV, capacitância de 25 µF e resistência de 200 Ω em um eletroporador Gene Pulser® II (Bio-Rad). Após o choque elétrico, rapidamente foram adicionados 1 mL de meio SOC estéril e a mistura foi transferida para um recipiente de 15 mL estéril e incubada a 37°C, sob agitação de 200 rpm, por 1 hora. Depois deste período 200 µL da mistura foram plaqueados em placa de LB Agar contendo 50 µg/mL de canamicina. As placas foram incubadas em estufa a 37°C por 16 horas. A placa contendo antibiótico permite fazer a seleção das colônias que foram transformadas com o plasmídeo contendo o gene de resistência a canamicina. No entanto, algumas células podem ser transformadas apenas com o vetor pET28b, sem o gene *clpP*. Por isso, se faz necessário o uso de outras técnicas, além da pressão seletiva na placa, para confirmar as cepas transformadas com a construção pET28b/*clpP* que deram origem as colônias na placa. Além disso, todo o procedimento de transformação também foi efetuado com a reação de ligação apenas com o vetor pET28b (controle negativo mostrado na seção 3.2.3).

Após a confirmação do clone positivo, transformado com a construção pET28b/*clpP* de maneira correta, este plasmídeo é extraído e transformado, pelo mesmo procedimento apresentado acima, na cepa *E. coli* BL21 Star (DE3) utilizada para expressão da proteína ClpP.

### 3.2.6 Seleção de clones e preparação de estoques em glicerol

A seleção dos clones que foram transformados com o plasmídeo foi feita através do plaqueamento das células em meio LB ágar com pressão seletiva (uso de antibiótico, canamicina). Apenas as células com o plasmídeo, o qual possui gene de resistência à canamicina, foram capazes de crescer em meio com este antibiótico. No entanto, como mencionado na seção anterior, algumas colônias transformadas podem ter recebido o plasmídeo com algum tipo de não conformidade, como por exemplo, com o gene de interesse ligado de forma errada, alguma base incorporada de maneira errônea no PCR ou até mesmo sem o gene. Para fazer a seleção, algumas colônias presentes na placa foram selecionadas, transferidas para outra placa com LB ágar e 50 µg/mL de canamicina, e ainda inoculadas em meio LB líquido com 50 µg/mL de canamicina e incubadas a 37°C e 200 rpm por 16 horas. A partir destes cultivos foram feitos os estoques de glicerol (chamados de banco mãe) e as extrações plasmidiais para a confirmação dos clones (padrão de digestão e seqüenciamento).

O estoque de glicerol foi realizado utilizando 300µL do cultivo, citado anteriormente, de cada clone e a este volume foram adicionados 300µL de glicerol 50% estéril. Os estoques foram armazenados a -70°C.

Os mesmos procedimentos (estoque de glicerol e extração plasmidial) foram repetidos para a cepa *E. coli* BL21 Star (DE3) após a sua transformação com a construção pET28b/*clpP*-C2P.

### 3.2.7 Extração de DNA plasmidial

A extração de DNA plasmidial foi utilizada em algumas partes deste trabalho, como para a digestão dos plasmídeos e seqüenciamento dos mesmos para confirmação dos clones e para verificação da correta seqüência de gene *clpP* no vetor pET28b/*clpP*.

Para a extração de DNA plasmidial, foi utilizado o *kit* Wizard® Plus SV Minipreps DNA Purification System (Promega). Este sistema é baseado na separação do DNA plasmidial através de uma membrana de sílica presente dentro de uma coluna. As extrações foram feitas a partir de 10 mL de um cultivo de 16 horas das células contendo o DNA plasmidial. O cultivo foi centrifugado por 5 minutos a 4000 rpm em centrífuga Eppendorf 5810R (rotor A-4-62) e o sobrenadante descartado. As células foram resuspensas com 250µL da solução *Cell Resuspension* (50 mM de Tris-HCl (pH 7.5), 10 mM de EDTA, 100 µg/ml RNase A) e, logo após, adicionados 250 µL da solução *Cell Lysis* (0,2 M de NaOH, 1% SDS). A mistura foi invertida quatro vezes, cuidadosamente, e então adicionados 10 µL da solução de Protease alcalina. Novamente a mistura foi invertida quatro vezes, cuidadosamente, e incubada por 5 minutos a temperatura ambiente. Após estes 5 minutos, foram adicionados 350 µL da solução *Neutralization* (4,09 M de hidrocloreto de guanidina, 0,759 M de acetato de potássio, 2,12 M de ácido acético glacial), a mistura foi invertida quatro vezes e centrifugada a 14000 rpm por 10 minutos em centrífuga Eppendorf 5417R (rotor F45-30-11). Depois da centrifugação o sobrenadante clarificado foi transferido para a coluna com a membrana e esta centrifugada por 1 minuto a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). O percolado foi descartado e a coluna lavada com 750 µL de *Wash solution* (60% de etanol, 60 mM de acetato de potássio, 8,3 mM de Tris-HCl, 0,04 mM de EDTA) e centrifugada por 1 minuto a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). Novamente o percolado foi descartado, a coluna foi lavada desta vez com 250 µL de *Wash solution* e centrifugada a 14000 rpm por 2 minutos em centrífuga Eppendorf 5417R (rotor F45-30-11). A coluna foi transferida para um tubo de microcentrífuga de 1,5 mL e a ela foram adicionados 50 µL de ultrapura autoclavada, para eluir o DNA plasmidial, e centrifugada por 1 minuto a 14000 rpm em centrífuga Eppendorf 5417R (rotor F45-30-11). A solução percolada, contendo DNA plasmidial,

foi estocada a  $-20^{\circ}\text{C}$ . O DNA plasmidial foi visualizado em eletroforese em gel de agarose.

### 3.2.8 Confirmação dos clones

A confirmação dos clones foi feita primeiramente através do padrão de digestão dos plasmídeos extraídos, com enzimas de restrição específicas, e depois através de seqüenciamento do plasmídeo.

#### 3.2.8.1 Digestão com enzimas de restrição

Para a confirmação dos clones, os plasmídeos extraídos foram submetidos à digestão. A digestão foi realizada para confirmar a presença do gene no plasmídeo e para confirmar a ligação do inserto de maneira correta. Para isto foram selecionadas enzimas de restrição de maneira que uma fosse capaz de digerir o gene e outra de digerir o vetor, formando dois fragmentos facilmente identificáveis em eletroforese de gel de agarose. Para a digestão foram selecionadas as enzimas *PstI* e *EcoRV*, capazes de digerir o gene *clpP* e o vetor pET28b, respectivamente. Desta forma, para a confirmação dos clones, ao digerir os plasmídeos extraídos com as duas enzimas são esperadas duas bandas de 1697pb e 4134pb.

As reações de digestão foram realizadas utilizando 8 $\mu\text{L}$  da extração plasmidial (cerca de 200 – 250 ng de DNA), 2U de *EcoRV* (Invitrogen), 2U de *PstI* (Invitrogen), 1  $\mu\text{L}$  do Tampão *React2* (Invitrogen), e o volume completado com água ultrapura até 10  $\mu\text{L}$ . Também foram realizadas digestão apenas do vetor pET28b com *EcoRV* e apenas do gene *clpP* com *PstI*, como controle positivo para as enzimas. A reação apenas com o gene foi realizada utilizando aproximadamente 300 ng de DNA, 2U de *PstI*, 1  $\mu\text{L}$  do Tampão *React2* e o volume completado com água ultrapura até 10  $\mu\text{L}$ . Para a reação com o vetor pET28b foram utilizados aproximadamente 300 ng de DNA, 2U de *EcoRV*, 2  $\mu\text{L}$  do Tampão *React2* e o volume completado com água ultrapura até 20  $\mu\text{L}$ . Todas as reações de digestão foram incubadas a  $37^{\circ}\text{C}$  por 16 horas.



### 3.2.8.2 Seqüenciamento

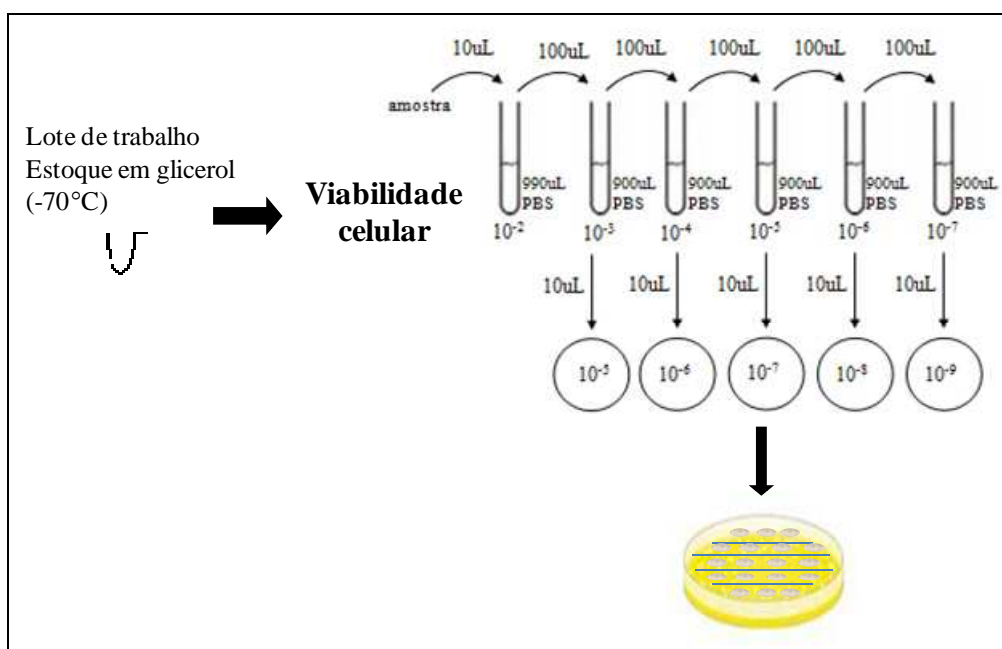
O plasmídeo pET28b/*clpP* extraído e confirmado por digestão, foi seqüenciado para verificação do correto posicionamento do gene no plasmídeo e para confirmação da correta incorporação de bases no gene durante o PCR. O seqüenciamento foi realizado com terminadores fluorescentes (Big Dye, Applied Biosystems) utilizando o seqüenciador capilar ABI PRISM® 3100 (Applied Biosystems) do Laboratório de AIDS & Imunologia Molecular – IOC/Fiocruz. Para a reação de seqüenciamento foram feitas seis reações, quatro contendo os oligonucleotídeos iniciadores capazes de amplificar a partir de uma região do plasmídeo (T7 senso e anti-senso), duas com cada, uma com o oligonucleotídeo pET28bHisClpPf (Senso) e uma com o oligonucleotídeo pET28bHisClpPr (anti-senso). Além destas seis reações, foi realizada uma reação, como controle positivo, utilizando o oligonucleotídeo iniciador MI3 e o vetor pGEM como DNA a ser seqüenciado. As reações, utilizando água ultrapura como solvente, foram compostas por: 150 ng de DNA, 2 µL de solução tampão MgCl<sub>2</sub> (400 mM Tris pH 9,0, 10 mM MgCl<sub>2</sub>), 2 µL de Big Dye (Applied Biosystems), 3,2 pmol do oligonucleotídeo iniciador e o volume completado até 10 µL. O programa utilizado no termociclador para a extensão das fitas foi de 1 minuto a 95°C para a desnaturação das fitas e depois 30 ciclos composto pelas seguintes etapas: 96°C por 10 segundos, 50°C por 5 segundos e 60°C por 4 minutos. As reações foram armazenadas a 4°C, protegidas da luz, até o momento da precipitação. O DNA foi precipitado com isopropanol e seco em concentrador centrífugo a vácuo (Vacufuge™Eppendorf). As amostras foram então submetidas ao seqüenciador capilar ABI PRISM® 3100 (Applied Biosystems). A seqüência foi analisada e comparada com as seqüências depositadas no *GenBank* utilizando programa CLUSTAL W.

### 3.2.9 Preparação de lote de trabalho e viabilidade celular

Após as confirmações do clone positivo, o plasmídeo pET28b/*clpP*-C2P extraído deste clone foi transformado em *E. coli* BL21 Star (DE3). Os clones desta cepa foram confirmados e a partir deles foi preparado um lote de trabalho, estocado em glicerol. A partir deste lote que foram realizadas todas as expressões da proteína ClpP deste trabalho.

O lote de trabalho foi preparado a partir do estoque de glicerol (banco mãe) dos quatro clones selecionados e confirmados de *E. coli* BL21 Star (DE3)/pET28b/*clpP*. As cepas foram crescidas em 10 mL de meio LB (5 g/L de extrato de levedura (BD); 10 g/L de triptona (BD); 10 g/L de NaCl (Merck), pH 7), enriquecido com 1% de glicose (Merck), 0,4% de glicerol e 50 µg/mL de canamicina (Sigma), a 37°C e agitação de 200 rpm, até atingirem  $Abs_{600nm}$  de aproximadamente 1,0. Foram preparadas várias alíquotas com 100 µL do cultivo e 100 µL de glicerol 50% estéril. As alíquotas com (lote de trabalho) foram estocadas a -70°C.

As expressões foram sempre realizadas a partir deste banco de trabalho, sendo que as alíquotas sempre que utilizadas eram avaliadas quanto à viabilidade celular, apresentando em torno de  $10^8$  UFC/mL. A viabilidade celular do lote de trabalho da cepa recombinante *E. coli* BL21 Star (DE3)/pET28b/*clpP* foi avaliada a partir da contagem de UFC (Unidades Formadoras de Colônia). Foram feitas diluições seriadas em PBS pH 7,4, e estas diluições foram plaqueadas em placas de Petri contendo LB Ágar e canamicina 50 µg/mL. Na figura 3.4 é mostrado um desenho esquemático das etapas realizadas na análise de viabilidade celular.



**Figura 3.4** Desenho esquemático dos procedimentos realizados para a análise de viabilidade celular.

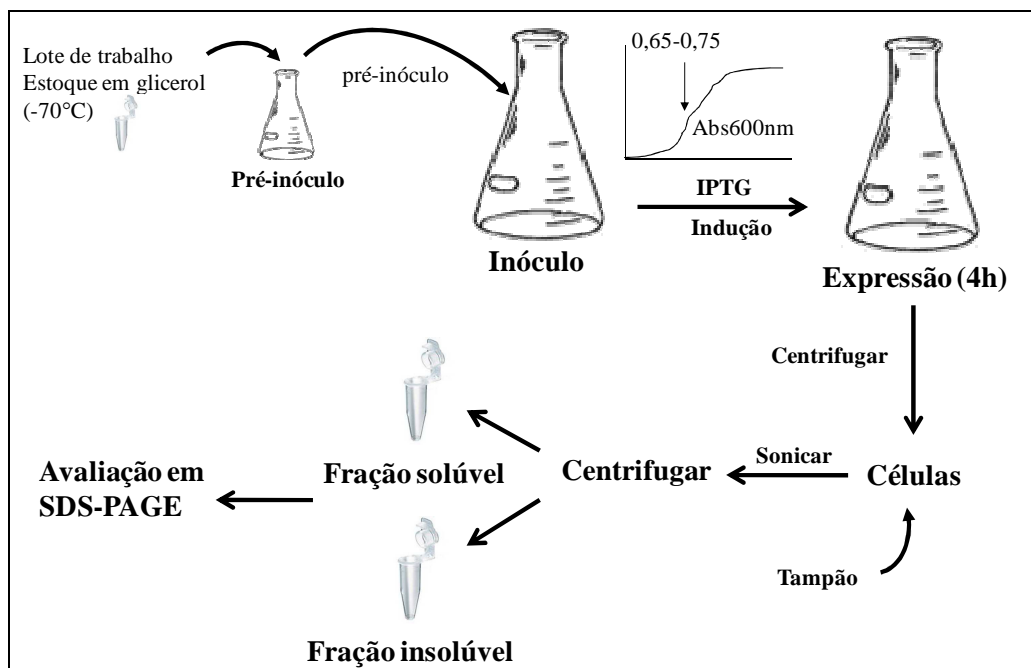
### 3.2.10 Crescimento celular e curva de massa seca de células

A avaliação do crescimento celular foi realizada através de medidas de  $Abs_{600nm}$ . No entanto, todas estas leituras foram convertidas para massa seca de células. Para tanto foi realizada uma curva de absorvância em função da concentração de células em massa seca para realizar as conversões.

Para fazer a curva foi utilizado um precipitado celular proveniente de 200 mL do cultivo da cepa *E. coli* BL21 Star (DE3)/pET28b/*clpP* com 4 horas de expressão. O precipitado foi lavado exaustivamente com água destilada gelada até a obtenção de um sobrenadante límpido. Com este precipitado lavado foi preparada um suspensão em balão volumétrico com concentração final de 21,29 mg de massa úmida de células/mL. Foram transferidos 5 mL desta suspensão para uma placa previamente pesada, e este conjunto levado a estufa a 60°C até obtenção de massa constante. Com esta massa foi possível obter a concentração de massa seca de células/mL da suspensão preparada. A partir da suspensão de 21,29 mg de massa úmida de células/mL foram preparadas outras suspensões diluídas em água destilada 100, 70, 50, 30, 25 e 20 vezes. Estas suspensões foram submetidas à leitura de  $Abs_{600nm}$ . Com isso foi traçado um gráfico da  $Abs_{600nm}$  em função da concentração de células em massa seca e assim obtido através de regressão linear a equação que correlaciona  $Abs_{600nm}$  com concentração de massa seca de células.

### 3.2.11 Expressão e preparação dos extratos protéicos para análise de solubilidade

Primeiramente para realizar a expressão da proteína foram utilizados os lotes de trabalho dos quatro clones de *E. coli* BL21 Star (DE3)/pET28b/*clpP*. Um esquema geral dos procedimentos realizados para fazer a expressão e a análise de solubilidade da proteína heteróloga é mostrado na figura 3.5.



**Figura 3.5** Esquema dos procedimentos realizados nos experimentos de expressão e análise de solubilidade da proteína.

Para fazer o pré-inóculo, 10  $\mu\text{L}$  do lote de trabalho da bactéria recombinante *E. coli* BL21 Star (DE3)/pET28b/clpP foram inoculados em 10 mL de meio LB (5 g/L de extrato de levedura (BD); 10 g/L de triptona (BD); 5 g/L de NaCl (Merck), pH 7) enriquecido com 1% de glicose (Merck), 0,4% de glicerol e 50  $\mu\text{g/mL}$  de canamicina (Sigma). O pré-inóculo foi incubado por 16 horas a 37°C e 200 rpm, em frascos agitados de 50 mL. Após as 16 horas, o inóculo foi preparado, em frascos de 500 mL, com 2 mL do pré-inóculo em 100 mL do meio LB enriquecido com 1% de glicose (Merck), 0,4% de glicerol e 50  $\mu\text{g/mL}$  de canamicina (Sigma). O cultivo foi incubado a 37°C e 200 rpm até atingir a fase exponencial de crescimento ( $\text{Abs}_{600\text{nm}}$  entre 0,65 e 0,75). Neste ponto, a expressão foi induzida com 1 mM de IPTG (isopropil- $\beta$ -D-tiogalactopiranosídeo) por 4 horas. Amostras de 1 mL do cultivo foram retiradas, antes e ao final das 4 horas de expressão, e centrifugadas, para avaliar a expressão da proteína ClpP e também sua solubilidade.

Para avaliar a expressão da proteína foram utilizados os precipitados celulares obtidos das amostras de 1 mL retiradas dos cultivos antes (não induzido) e após as 4 horas de expressão. Aos precipitados foi adicionado tampão Tris-HCl 60 mM pH6,8, 10% de glicerol, 5%  $\beta$ -mercaptoetanol, 2% de SDS e 0,5% de azul de bromofenol,

seguinto a razão de 25  $\mu\text{L}$  de tampão para cada 0,1 de  $\text{Abs}_{600\text{nm}}$ . As amostras foram incubadas em banho de água fervente por 5 minutos e aplicadas em Gel de Eletroforese Desnaturante de Poliacrilamida com SDS 12,5% (SDS-PAGE). Foram realizadas corridas a 120 V por 90 minutos e então os géis foram corados com Coomassie Blue R-250.

Para a análise de solubilidade da proteína expressa foram utilizadas as amostras de 1mL retiradas ao final das 4 horas de expressão dos cultivos realizados a partir dos clones *E. coli* BL21 Star (DE3)/pET28b/*clpP*-C2P/2 e *E. coli* BL21 Star (DE3)/pET28b/*clpP* -C2P/4. As amostras foram resuspendidas em tampão de lise (Tris 20 mM; EDTA 1 mM, pH8,0), seguindo a razão 25  $\mu\text{L}$  de tampão para cada 0,1 de  $\text{Abs}_{600\text{nm}}$  (normalizando pela  $\text{Abs}_{600\text{nm}}$ ), obtendo os extratos totais das proteínas. Os extratos totais foram submetidos a ultra-som por cinco ciclos de 10 segundos com amplitude de 30% em sonicador *Sonics&Material Inc*. Através de centrifugação foram separadas as frações solúveis e insolúveis das proteínas totais dos cultivos. A estas frações foi adicionado tampão Tris-HCl 60 mM pH6,8, 10% de glicerol, 5%  $\beta$ -mercaptoetanol, 2% de SDS e 0,5% de azul de bromofenol. As amostras foram incubadas em banho de água fervente por 5 minutos e então aplicadas em gel de poliacrilamida 12,5%. Foi realizada corrida a 120 V por 90 minutos, e após a corrida, o gel foi corado com Coomassie Blue R-250.

### 3.2.12 Otimização da expressão por planejamento de experimentos

Com o objetivo de otimizar alguns parâmetros importantes na expressão da proteína ClpP foi realizado um planejamento fatorial a dois níveis com duas variáveis (2<sup>2</sup>). Neste trabalho foram avaliadas as influências de duas variáveis do processo, concentração de canamicina e concentração de IPTG, sobre o crescimento celular, sobre a concentração de proteína e sobre a estabilidade plasmidial. Antibióticos, como a canamicina, são largamente utilizados em escala laboratorial para fazer a pressão seletiva no meio, evitando a segregação do plasmídeo, já que a maioria dos plasmídeos utilizados possui um gene marcador de resistência a algum antibiótico. A segregação plasmidial pode ter reflexos muito importantes na produtividade da proteína recombinante principalmente em escala industrial. No entanto, em escala industrial o

uso destes antibióticos se torna inviável por vários motivos. Os antibióticos possuem um custo elevado e, além disso, contaminam o produto e precisam ser completamente removidos nos processos de purificação dos produtos com fins farmacêuticos ou alimentícios (FRIEHS, 2004). Por estes motivos o estudo da variável canamicina é tão importante neste processo, no entanto, a variação do antibiótico no cultivo pode ter reflexos sérios na instabilidade plasmidial. O IPTG também é uma variável muito importante no processo, principalmente em larga escala, já que também possui custo elevado e a sua toxicidade o torna contra indicado para proteínas de uso humano (HANNING E MAKRIDES, 1998).

A matriz do planejamento fatorial utilizado neste estudo é mostrada na tabela 3.2. Foram realizados 8 experimentos, sendo que destes experimentos 4 foram réplicas no ponto central. Os 8 experimentos do planejamento foram realizados com o clone *E. coli* BL21 Star (DE3)/pET28b/*clpP*-C2P/4. Para fazer o pré-inóculo, 10µL do lote de trabalho da bactéria recombinante *E. coli* BL21 Star (DE3)/pET28b/*clpP*-C2P/4 foram inoculados em 10 mL de meio LB (5 g/L de extrato de levedura (BD); 10 g/L de triptona (BD); 5 g/L de NaCl (Merck), pH 7) enriquecido com 1% de glicose (Merck), 0,4% de glicerol e 50 µg/mL de canamicina (Sigma), e incubados por 16 horas a 37°C e 200 rpm. A cada pré-inóculo realizado, a viabilidade celular do lote de trabalho utilizado foi avaliada conforme descrito na seção 3.2.9, sendo todas avaliadas em torno de 10<sup>8</sup> UFC/mL.

**Tabela 3.2** Tabela com as condições dos experimentos do planejamento fatorial 2<sup>2</sup>. Os números entre parênteses são os valores das variáveis independentes normalizados.

Experimento	IPTG (mM)	Canamicina (µg/mL)
1	0,1 (-1)	0 (-1)
2	0,1 (-1)	50(1)
3	1 (1)	0 (-1)
4	1 (1)	50(1)
PC5	0,55 (0)	25 (0)
PC6	0,55 (0)	25 (0)
PC7	0,55 (0)	25 (0)
PC8	0,55 (0)	25 (0)

O inóculo foi preparado em frascos de 500 mL, com 2 mL do pré-inóculo saturado em 100 mL do meio LB com a mesma composição e concentrações do pré-inóculo, exceto a concentração do antibiótico canamicina que variou de acordo com os experimentos do planejamento fatorial (tabela 3.2). O cultivo foi incubado a 37°C e 200 rpm. A indução da expressão da proteína recombinante com IPTG (isopropil- $\beta$ -D-tiogalactopiranosídeo) foi feita por 4 horas a partir da fase exponencial de crescimento (com  $Abs_{600nm}$  entre 0,65 e 0,75), variando a concentração de IPTG (Promega) de acordo com os experimentos do planejamento fatorial (tabela 3.2).

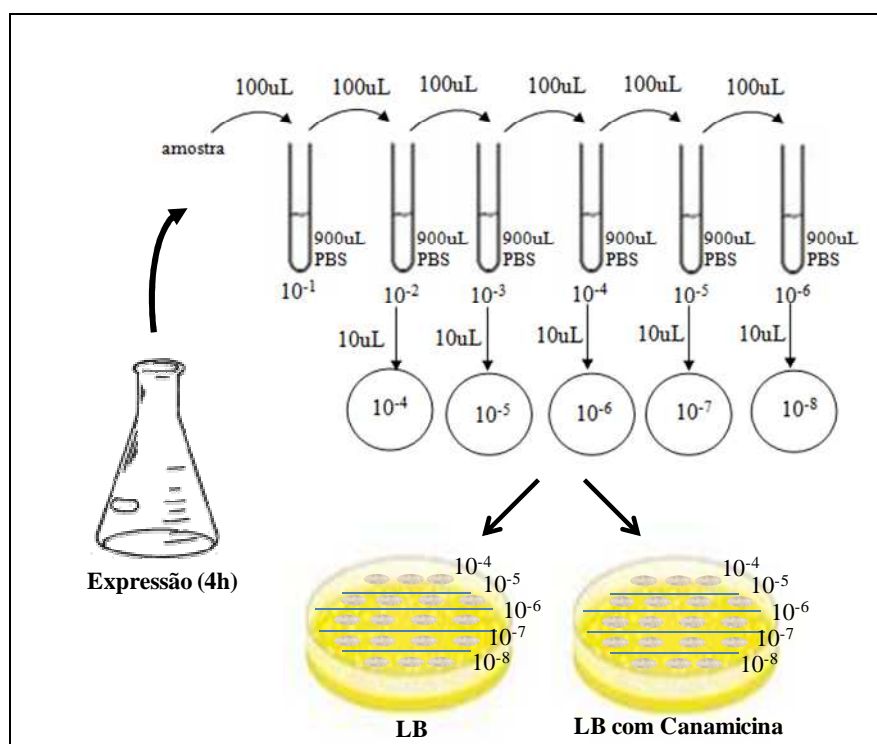
As análises estatísticas foram feitas através do programa STATISTICA 6.1, empregando as variáveis normalizadas. O efeito de cada um dos fatores foi estimado, assim como o erro padrão, e avaliado pelo teste *t*, considerando estatisticamente significativos aqueles com  $p < 0,1$  (RODRIGUES E IEMMA, 2005).

### 3.2.13 Avaliação do crescimento celular e quantificação da proteína expressa em SDS-PAGE

Após 4 horas de indução, o crescimento foi medido por  $Abs_{600nm}$  ( $Abs_{final}$ ). Amostras de células antes da indução e ao final do cultivo de cada condição de expressão do planejamento experimental foram resuspensas em tampão de amostra, obtendo os extratos totais das proteínas, seguindo a razão 25 $\mu$ L de tampão para cada 0,1 de  $Abs_{600nm}$ , normalizando as amostras pela  $Abs_{600nm}$ . Estas amostras foram aplicadas em gel de poliacrilamida 12,5%, corado com Coomassie Blue R-250. Neste mesmo gel também foi aplicado um marcador de massa molar LMW (Amersham Bioscience), com bandas de 97kDa, 66kDa, 45kDa, 30kDa, 20,1kDa e 14,4kDa, e concentrações conhecidas de proteína em cada banda, para comparação com as bandas correspondentes a proteína ClpP. A quantidade de proteína expressa em cada condição foi analisada por densitometria em equipamento Bio-Rad (GS-800 Calibrated Densitometer) e programa QuantityOne 4.4.1. Através do marcador de massa molar foi feita uma curva padrão, com a qual foi calculada a concentração de proteína expressa.

### 3.2.14 Análise da segregação plasmidial

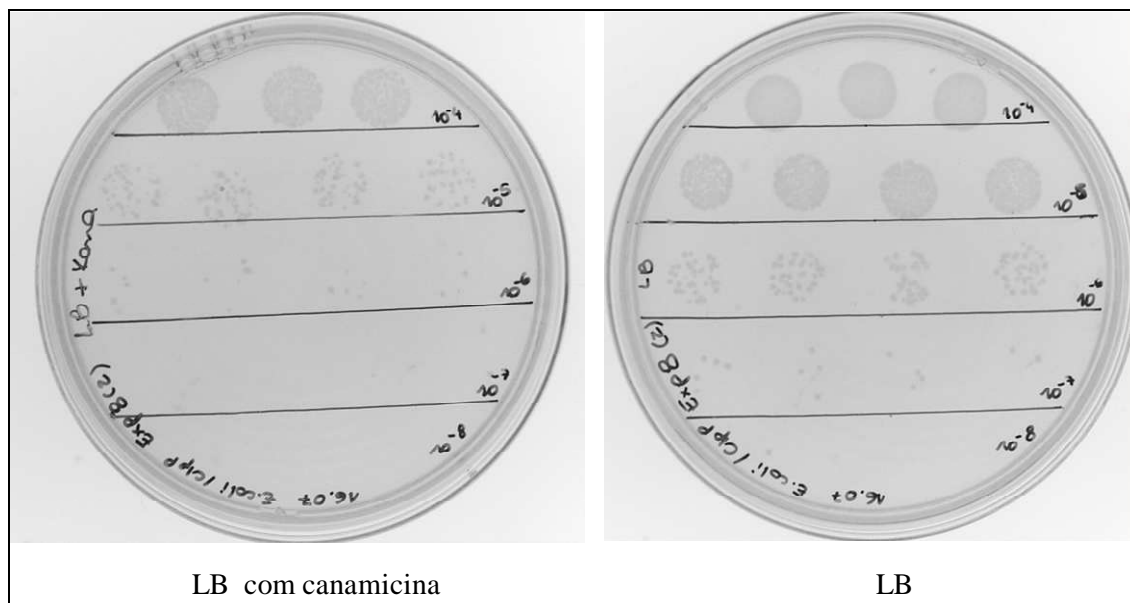
A segregação plasmidial foi avaliada através da contagem de UFC (unidades formadoras de colônia) na presença e ausência de canamicina (BAHERI *et al.*, 2001; TOMAZETTO *et al.*, 2007). As análises foram realizadas através de 100  $\mu\text{L}$  do cultivo, os quais passaram por diluições seriadas em PBS estéril até a diluição  $10^{-6}$ . Foram colocados 10  $\mu\text{L}$  de cada diluição, em no mínimo 3 réplicas, em placas de LB Ágar, com canamicina 50  $\mu\text{g}/\text{mL}$  e também sem o antibiótico. A estabilidade plasmidial foi mensurada através de  $F$  (fração de células com plasmídeo), para tanto foi realizada a razão entre a contagem de colônias da placa com antibiótico (apenas células com plasmídeo) e a contagem de colônias da placa sem antibiótico (total de células presentes). A figura 3.6 mostra um desenho esquemático das etapas realizadas na análise da segregação plasmidial.



**Figura 3.6** Desenho esquemático das etapas realizadas na análise de segregação plasmidial.

As diluições foram feitas de modo que na placa sempre estivessem presentes uma diluição sem células e uma diluição com crescimento conflente. Na figura 3.7, para exemplificar, é mostrada uma foto das placas realizadas no experimento PC8 do planejamento fatorial realizado neste trabalho.





**Figura 3.7** Fotos das placas de LB ágar, com e sem canamicina, das análises de segregação plasmidial do experimento PC8 do planejamento fatorial.

### 3.2.15 Validação da condição otimizada no planejamento fatorial

Após as análises do planejamento fatorial, foram realizadas réplicas da condição de cultivo otimizada para validar os resultados obtidos no planejamento. Após a indução foram retiradas amostras a cada hora para avaliar a cinética de produção da proteína ClpP, o crescimento celular e a segregação plasmidial. Estas análises foram realizadas conforme descrito anteriormente.

### 3.2.16 Purificação

A purificação da proteína ClpP clonada de forma a ser expressa fusionada a uma cauda de histidina foi realizada por cromatografia de afinidade através de coluna empacotada com resina contendo íons níquel imobilizados (HisTrap™ HP, GE Healthcare) de 1 mL, em HPLC ÄKTA Purifier 10 system (Amersham Biosciences).

Primeiramente, a purificação foi realizada com amostras preparadas de duas maneiras distintas, a partir de um precipitado celular proveniente de uma expressão de 4 horas a 37°C e 200 rpm. Em uma das amostras foi utilizada a solução tampão Tris 20 mM, EDTA 1 mM, imidazol 20 mM, pH8,0, tanto para fazer a lise celular quanto

para equilibrar a coluna utilizada na purificação. Na outra amostra foi utilizada a solução tampão Tris 20 mM, EDTA 1 mM, imidazol 5 mM, pH8,0 também para fazer a lise celular e para equilibrar a coluna utilizada na purificação.

As amostras foram aplicadas na coluna (cerca de 12 mg de proteína total) e o material que atravessou a coluna, não adsorvido, foi recolhido para posterior análise como sendo a fração não adsorvida. Após, a coluna foi eluída com solução tampão Tris 20 mM, EDTA 1 mM, pH8,0, fazendo um gradiente de 0 a 1 M de imidazol, para a desorção das proteínas. Amostras foram sendo recolhidas do material eluído da coluna em diferentes concentrações de imidazol para análises posteriores.

Depois destas duas amostras, ainda foram realizadas mais duas análises com outras duas amostras. No entanto, desta vez foi utilizada solução tampão Tris 20 mM, EDTA 1 mM, DTT (Ditiotreitol) 1 mM, pH8,0 para equilibrar a coluna e para fazer a lise celular de uma das amostras e solução tampão Tris 20 mM, EDTA 1 mM, DTT 1 mM, 1 mM imidazol, pH8,0, para equilibrar a coluna e fazer a lise celular da outra amostra. A purificação destas duas amostras foi realizada em coluna de níquel da mesma forma como nas duas primeiras purificações.

As amostras eluídas nas quatro purificações foram analisadas através de SDS-PAGE 12,5%.

### **3.3 Clonagem utilizando *Bacillus subtilis* como sistema de expressão**

#### **3.3.1 Caracterização da cepa de *Bacillus subtilis* WB600**

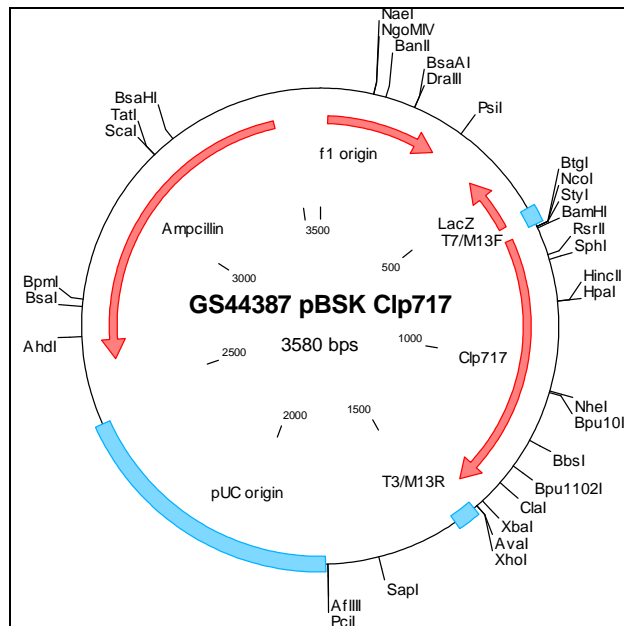
A cepa de *Bacillus subtilis* WB600 (WU *et al.*, 1991) empregada neste trabalho foi gentilmente cedida pelo professor Dr. Sui-Lam Wong da Universidade de Calgary – Canadá. Esta cepa é deficiente de seis das principais proteases extracelulares presentes em *Bacillus subtilis*. Os esporos da cepa foram abertos no Laboratório de Fisiologia Bacteriana (LFB) e Coleção de Culturas do Gênero *Bacillus* e Gêneros Correlatos (CCGB) do Instituto Oswaldo Cruz – FIOCRUZ. Estes esporos foram inoculados em meio LB com 5 µg/mL de eritromicina e lincomicina e incubados a 37°C por

aproximadamente 16 horas. Através desta cultura foram feitos novos estoques de esporos em papel filtro, estoques em LB ágar (com 5 µg/mL de eritromicina e lincomicina) inclinado, e isolamento de colônias em placas também de LB ágar (com 5 µg/mL de eritromicina e lincomicina). Em parceria com o LFB e CCGB foram realizados ensaios de caracterização desta cepa de gênero e espécie.

A morfologia das colônias foi caracterizada através de sua visualização nas placas e a morfologia celular foi caracterizada por coloração de Gram e também por visualização em microscópio de lâminas a fresco. Foram ainda realizados ensaios bioquímicos, como redução de nitrato, Voges-Proskauer (VP), utilização de citrato, arabinose, xilose e glicose, crescimento em diferentes concentrações de NaCl, degradação da caseína, do amido e da gelatina, β-hemólise em placas de ágar sangue, entre outros. Foram também realizados testes bioquímicos através do sistema de identificação Api 50 CHB (Biomerieux).

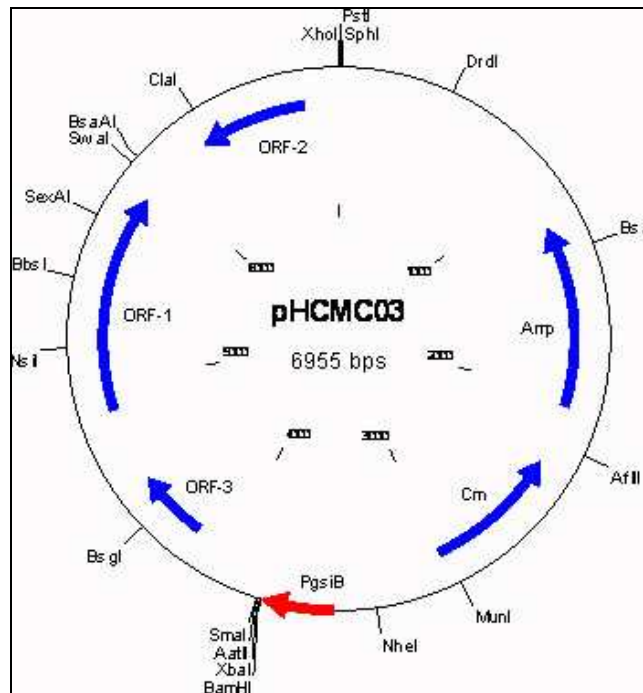
### 3.3.2 Desenho do gene *clpP* sintético e vetores utilizados

Para fazer a expressão da proteína em *B. subtilis* um gene sintetizado quimicamente foi utilizado. O gene foi desenhado de acordo com a seqüência da proteína ClpP obtida no seqüenciamento da construção pET28b/*clpP*-C2P. A esta seqüência ainda foi adicionado a seqüência do peptídeo sinal *Epr* (BROCKMEIER *et al.*, 2006) e um sítio de clivagem entre o peptídeo sinal e a seqüência da proteína. Os códons utilizados no gene sintético foram otimizados para os códons mais freqüentes utilizados por *B. subtilis*. O gene sintético, flanqueado pelas enzimas de restrição *Xba*I e *Bam*HI, foi inserido no vetor pBluescript II SK para a manutenção do gene, o qual passou a se chamar pBSKClpP717. O mapa do vetor pBSKClpP717 é mostrado na figura 3.8, e sua seqüência juntamente com a seqüência do gene sintético se encontram no apêndice A.



**Figura 3.8** Mapa do plasmídeo pBSK ClpP717.

Para as rotinas de clonagem em *E. coli* JM109 e posterior expressão em *B. subtilis*, o vetor utilizado foi o pHCMC03 (NGUYEN *et al.*,2005), gentilmente cedido pela pesquisadora Rita C. Ferreira do Departamento de Microbiologia da Universidade de São Paulo - Brasil. Este vetor possui o gene marcador de resistência à ampicilina e é capaz de se replicar tanto em *E. coli* quanto em *B. subtilis*. O pHCMC03 utiliza o promotor *PgsiB* que pode ser induzido por estresse térmico, ácido ou por etanol. A figura 3.9 mostra o mapa do vetor pHCMC03.



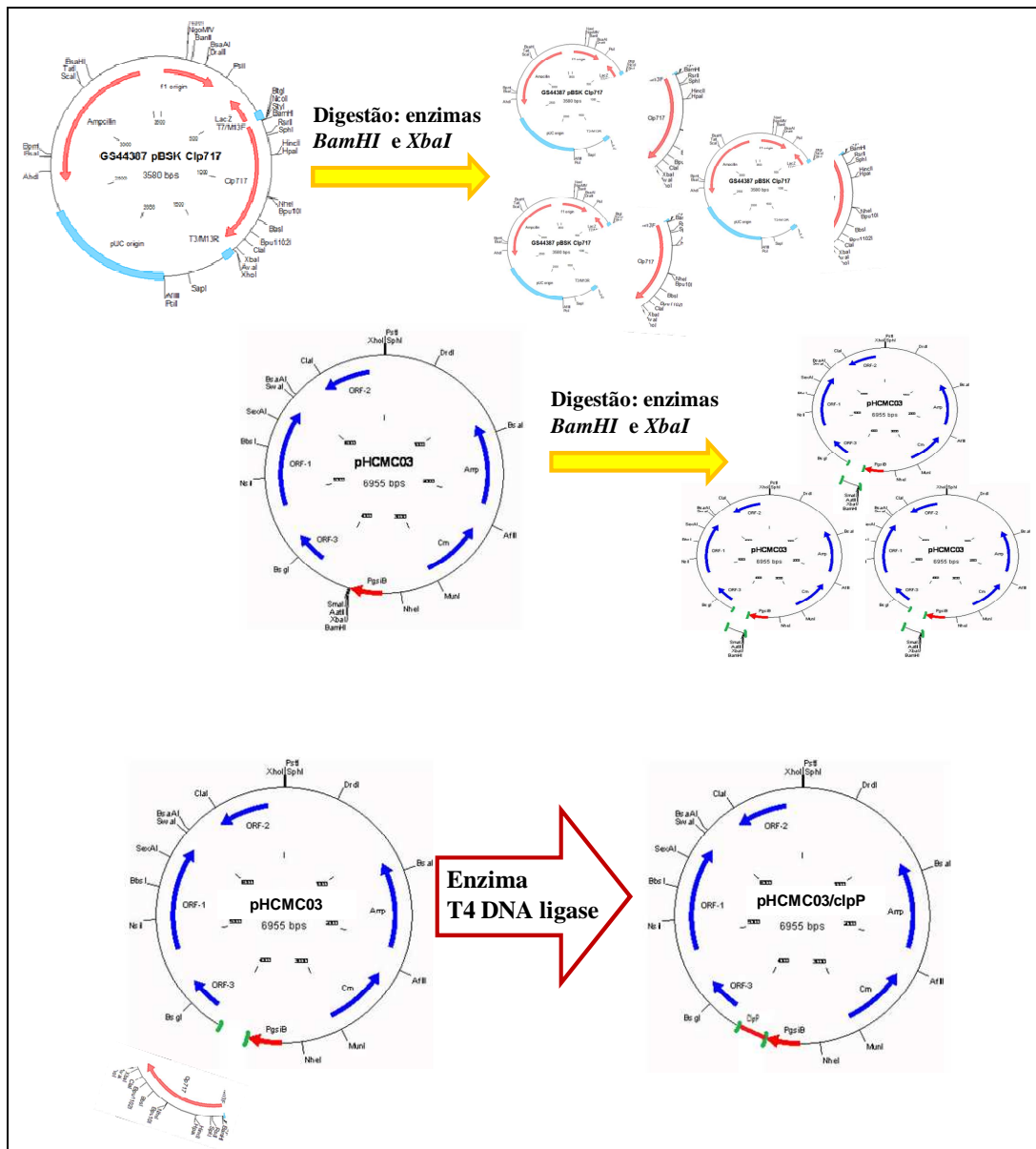
**Figura 3.9** Mapa do plasmídeo pHCMC03 (NGUYEN *et al.*,2005)

### 3.3.3 Preparo de células competentes para transformação por choque térmico

A cepa *E. coli* JM109 (Promega) utilizada para as rotinas de clonagem nesta parte do trabalho possui o seguinte genótipo: *endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17* ( $r_k^-$ ,  $m_k^+$ ), *relA1*, *supE44*,  $\Delta(lac-proAB)$ , [*F' traD36*, *proAB*, *laqI<sup>9</sup>Z* $\Delta$ M15]. Esta cepa, *E. coli* JM109, foi inoculada em LB e incubada por 16 horas à 37°C e 200 rpm. Após as 16 horas, 0,5 mL desta cultura foram inoculados em 50 mL de meio LB com 20 mM de MgSO<sub>4</sub> e incubados à 37°C e 200 rpm até atingir a fase exponencial de crescimento ( $Abs_{600nm}$  aproximadamente 0,6). A cultura foi centrifugada a 4000 rpm e 4°C por 5 minutos. As células foram resuspensas em 20 mL de uma solução estéril gelada de 30 mM de acetato de potássio, 100 mM de RbCl, 10 mM de CaCl<sub>2</sub>, 50 mM de MnCl<sub>2</sub> e 15% de glicerol, com pH 5,8. As células resuspensas foram mantidas em gelo por 5 minutos e então centrifugadas por 5 minutos à 5000 rpm e 4°C. Nesta etapa, as células foram resuspensas em solução estéril gelada contendo 10 mM de Mops (Morfolinopropanosulfônico), 10 mM de RbCl, 75 mM de CaCl<sub>2</sub>, e 15% de glicerol, com pH 6,5. As células resuspensas foram incubadas por 45 minutos em gelo. Após este período foram feitas alíquotas de 100  $\mu$ L e congeladas rapidamente em gelo seco. As alíquotas foram estocadas a -70°C.

### 3.3.4 Clonagem

Nesta parte do trabalho a clonagem foi feita utilizando o gene sintético *clpP* e o vetor pHCMC03 (NGUYEN *et al.*, 2005). Um esquema geral das etapas realizadas na construção do vetor para esta clonagem é mostrado na figura 3.10.



**Figura 3.10** Desenho esquemático das etapas realizadas para a construção do vetor utilizado na clonagem em *B. subtilis*.

Para fazer a construção do vetor pHCMC03/*clpP* foram realizadas reações de digestão com o vetor pHCMC03, para a sua linearização, e com o vetor pBSKClpP717,

para a liberação do inserto *clpP*. Na reação de digestão do vetor pBSKClpP717 empregando água ultrapura como solvente, foram utilizados aproximadamente 2 µg do vetor, 10U de *BamHI* (Fermentas), 4 µL de solução tampão *Tango* - 33mM de Tris-acetato (pH 7.9 a 37°C), 10 mM de acetato de magnésio, 66 mM acetato de potássio e 0.1 mg/mL de BSA - (Fermentas), em um volume total de reação de 20 µL. Na reação de digestão do vetor pHCMC03, empregando água ultrapura como solvente, foram utilizados aproximadamente 2 µg do gene, 10U de *BamHI* (Fermentas), 4 µL de solução tampão *Tango* (Fermentas), em um volume total de reação de 20 µL. As reações foram incubadas por 8 horas a 37°C. Após este tempo a enzima foi inativada sendo incubada à 80°C por 20 minutos. Foram então adicionados 10U da enzima *XbaI* (Fermentas) em cada reação de digestão. As digestões foram então incubadas a 37°C por 16 horas. As digestões foram visualizadas através de eletroforese em gel de agarose 1%. As digestões foram purificadas através do kit Wizard® SV Gel and PCR Clean-Up System (Promega) da mesma forma exemplificada na seção 3.2.2.3, exceto pelo fato de que a solução *Membrane Binding Solution* foi adicionada diretamente à solução com os produtos da digestão e não ao gel de agarose. As reações de digestão purificadas foram confirmadas em eletroforese de gel de agarose 1%.

Com o gene *clpP* digerido (liberado do vetor pBSKClpP717) e o vetor pHCMC03 também digerido, foi realizada a construção do vetor pHCMC03/*clpP* através de ligação com a enzima T4 DNA ligase (Fermentas), respeitando na reação de ligação a razão molar inserto(*clpP*)/vetor de 5/1. Foram realizados dois tipos de reações de ligação, uma com o vetor e o gene, e outra (controle negativo) utilizando apenas o vetor. As reações foram compostas por 42 ng do gene *clpP* (exceto no controle negativo), 81 ng do vetor pHCMC03, 5U de T4 DNA ligase e 2 µL da solução tampão 10X *T4 DNA Ligase Buffer* (Fermentas) sendo o volume completado até 20 µL com água ultrapura autoclavada. As reações de ligação foram incubadas a 22°C por diferentes tempos, 1 hora, 2 horas e 16 horas. Após a incubação a enzima foi inativada por incubação a 65°C por 10 minutos. Estas reações foram utilizadas para fazer a transformação em célula competente *E. coli* JM109.

### 3.3.5 Transformação por choque térmico

As células competentes de *E. coli* JM109 foram utilizadas como células hospedeiras para a construção pHCMC03/clpP. Para a transformação, 10 µL da reação de ligação foram adicionados a 100 µL das células *E. coli* JM109 competentes. A mistura foi incubada em gelo por 1 hora e após este tempo, foi realizado um choque térmico a 37°C por 45 segundos. Depois do choque térmico, a mistura foi incubada novamente em gelo por 2 minutos e logo após, foram adicionados 400 µL de meio LB contendo 0,4% de glicose e incubados a 37°C sob agitação vigorosa por 30 minutos. As células foram centrifugadas e ressuspensas em 50µL de LB, sendo então plaqueadas em LB ágar com 100 µg/mL de ampicilina.



## 4 RESULTADOS E DISCUSSÃO

### 4.1 Seleção da proteína

A seleção das proteínas foi realizada utilizando ferramentas de bioinformática e dados disponíveis na literatura. Foram pré-selecionadas e analisadas três proteínas, ClpP, IgA1 protease e Hialuronato Liase (Hyl).

Através da utilização do programa Blastp (ALTSHUL *et al.*, 1990), a sequência de aminoácidos da proteína ClpP de *S. pneumoniae* CGSP14, sorotipo 14, depositada no *GenBank* sob o código GI: 182629000, foi alinhada com outras contidas nos bancos de dados, sendo o resultado deste alinhamento mostrado na figura 4.1. Neste alinhamento foram encontradas 55 seqüências com 100% de identidade com a seqüência da cepa CGSP14, o que indica que a ClpP é altamente conservada entre as diversas cepas de *S. pneumoniae*. Destas, foram escolhidas oito seqüências de aminoácidos da proteína ClpP e realizado o alinhamento, através do programa CLUSTAL W (Figura 4.2). As seqüências escolhidas são pertencentes às seguintes cepas: *S. pneumoniae* P1031 (GI: 225856422), *S. pneumoniae* 70585 (GI: 225858557), *S. pneumoniae* D39 (GI: 116076792), *S. pneumoniae* Hungary19A-6 (GI:168994986), *S. pneumoniae* CGSP14 (GI:182629000), *S. pneumoniae* G54 (GI:194358110), *S. pneumoniae* R6 (GI: 934194) *S. pneumoniae* TIGR4 (GI: 930696).

```

> ref|NP\_345245.1| G ATP-dependent Clp protease proteolytic subunit [Streptococcus pneumoniae TIGR4]
ref|NP\_358250.1| G ATP-dependent Clp protease proteolytic subunit [Streptococcus pneumoniae R6]
ref|ZP\_01408244.1| G hypothetical protein SpneT_02001290 [Streptococcus pneumoniae TIGR4]
▶52 more sequence titles
Length=196

GENE ID: 930696 clpP | ATP-dependent Clp protease proteolytic subunit [Streptococcus pneumoniae TIGR4] (10 or fewer PubMed links)

Score = 404 bits (1038), Expect = 3e-111, Method: Compositional matrix adjust.
Identities = 196/196 (100%), Positives = 196/196 (100%), Gaps = 0/196 (0%)

Query 1 MIPVVEQTSRGERSDIYSRLLKDRIIMLTGPFVEDNMANSVIAQLFLDAQDSTKDIYL 60
Sbjct 1 MIPVVEQTSRGERSDIYSRLLKDRIIMLTGPFVEDNMANSVIAQLFLDAQDSTKDIYL 60

Query 61 YVNTPGGSVSAGLAIVDTMNFIAKADVQTIIVMGMAASMGTVIASSGAKGKRFMLPNAEYMI 120
Sbjct 61 YVNTPGGSVSAGLAIVDTMNFIAKADVQTIIVMGMAASMGTVIASSGAKGKRFMLPNAEYMI 120

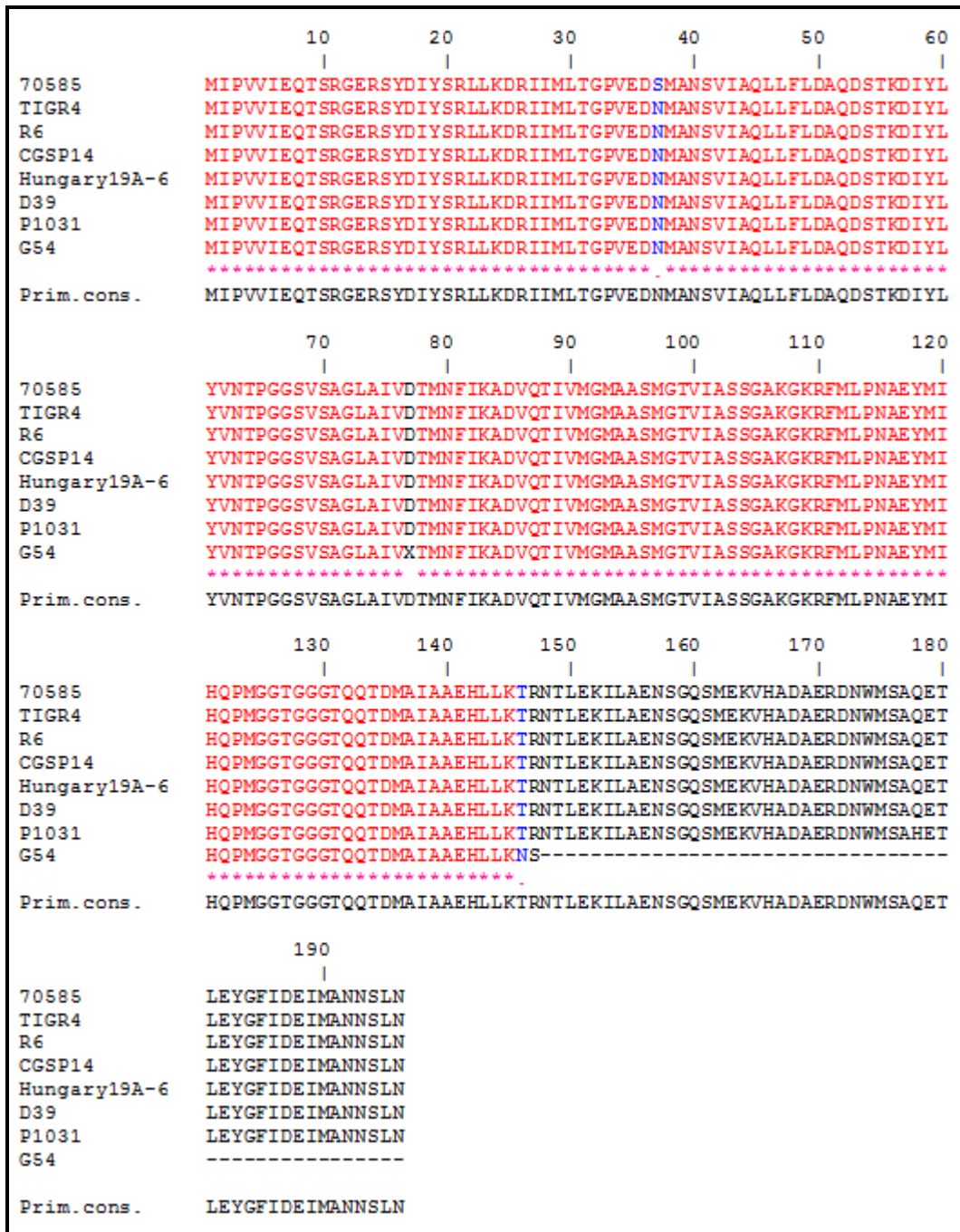
Query 121 HQPMGGTGGGTQQTDMAIAAEHLLKTRNTLEKILAENSGQSMEKVHADAERDNWMSAQET 180
Sbjct 121 HQPMGGTGGGTQQTDMAIAAEHLLKTRNTLEKILAENSGQSMEKVHADAERDNWMSAQET 180

Query 181 LEYGFIDEIMANNSLN 196
Sbjct 181 LEYGFIDEIMANNSLN 196

```

**Figura 4.1.** Comparação da sequência de aminoácidos da proteína ClpP de *S. pneumoniae* CGSP14, depositada no *GenBank* sob o código GI: 182629000, com outras sequências depositadas no banco de dados de proteínas, através da ferramenta Blastp. A seta indica onde a análise avalia que além das três sequências demonstradas existem mais 52 sequências com 100% de identidade.

O alinhamento múltiplo feito pelo programa CLUSTAL W apresentou tamanho de 196 aminoácidos, com 72,96% de resíduos idênticos (mostrados em vermelho na figura 4.2), 1,02% de resíduos com fraca similaridade (mostrados em azul na figura 4.2) e 26,02% de resíduos diferentes (mostrados em preto na figura 4.2). O alto percentual de resíduos diferentes ocorreu pois a sequência pertencente à cepa *S. pneumoniae* G54 possui apenas 147 aminoácidos, enquanto as demais possuem 196 aminoácidos. No entanto, se retirarmos apenas a sequência pertencente à cepa *S. pneumoniae* G54 (GI:194358110), o alinhamento das outras sete sequências passa a apresentar 98,98% de resíduos idênticos, 0,52% de resíduos com alta similaridade e 0,52% de resíduos com baixa similaridade.



**Figura 4.2.** Alinhamento múltiplo, realizado com o auxílio do programa CLUSTAL W, de oito seqüências da proteína ClpP, de diferentes cepas de *S. pneumoniae*, depositadas no GenBank. O alinhamento foi realizado com as seguintes cepas de *S. pneumoniae*: 70585 (GI: 225858557), TIGR4 (GI: 930696), R6 (GI: 934194), CGSP14 (GI:182629000), Hungary19A-6 (GI:168994986), D39 (GI: 116076792), P1031 (GI: 225856422), G54 (GI:194358110).

Os dados obtidos no alinhamento múltiplo feito com o programa CLUSTAL W, juntamente com os dados do alinhamento feito pelo programa Blastp (55 seqüências com 100% de identidade com a seqüência da cepa CGSP14), indicam que a proteína

ClpP possui elevada identidade entre as cepas de *S. pneumoniae*, sendo altamente conservada. Por conta de sua elevada identidade, as próximas análises continuaram a ser realizadas utilizando a seqüência da proteína ClpP de *S. pneumoniae* CGSP14 depositada no GenBank sob o código GI: 182629000.

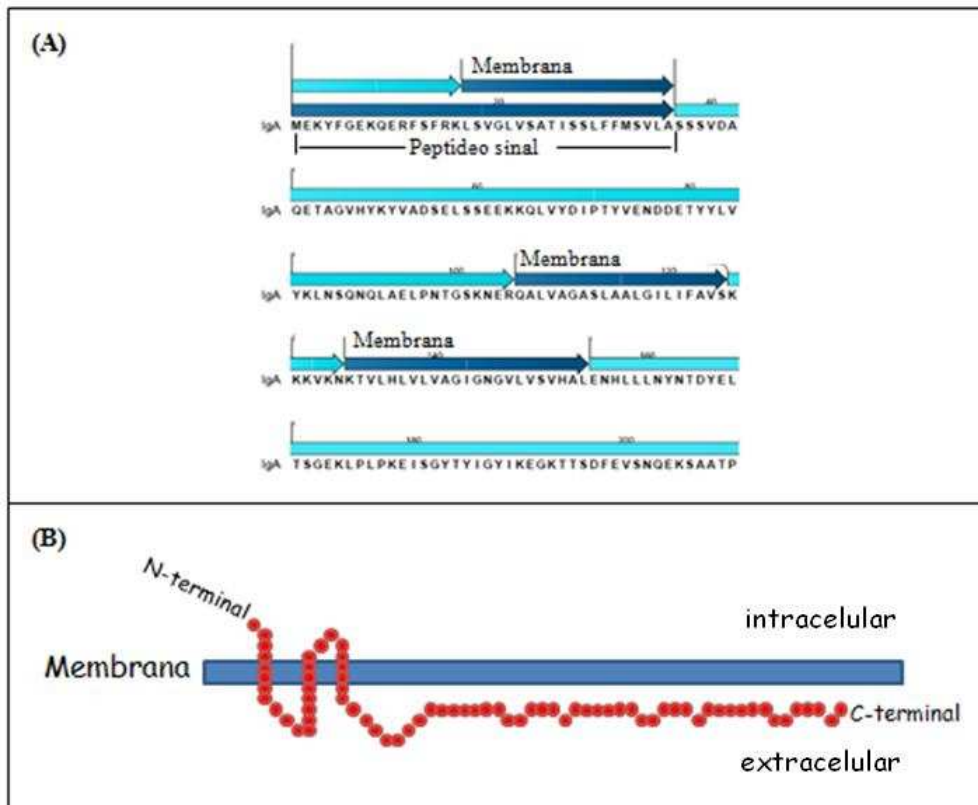
A proteína ClpP, com 196 aminoácidos, de acordo com o programa CLC Workbench 4, possui massa molar estimada em 21,5kDa, o que está de acordo com dados da literatura (KWON *et al.*, 2004), e ponto isoelétrico em pH 5,11. Além disso, a proteína não possui regiões transmembranares (hidrofóbicas). Para a análise de possíveis seqüências sinalizadoras presentes na proteína, a seqüência de aminoácidos foi submetida ao programa SignalP 3.0. Através dos resultados obtidos pode-se inferir que esta proteína não possui peptídeo sinal.

Outra proteína analisada foi a proteína IgA1 protease. A seqüência de aminoácidos da proteína IgA1 protease de *S. pneumoniae* CGSP14, depositada no *GenBank* sob o código GI:182629447, foi alinhada com outras seqüência disponíveis no banco de dados, utilizando o programa Blastp, com o objetivo de identificar outras seqüências de IgA1 protease. A partir do resultado, foram selecionadas oito seqüências de IgA1 protease de *S. pneumoniae*. Com estas seqüências, um alinhamento múltiplo foi realizado no programa CLUSTAL W. O alinhamento múltiplo mostrou identidade de 51,08% entre as seqüências, confirmando que a proteína é altamente heterogênea entre os sorotipos (MISTRY E STOCKLEY, 2006). Os demais dados deste alinhamento são mostrados na tabela 4.1, e o alinhamento completo é mostrado no apêndice B. Assim como realizado com a proteína ClpP, as demais análises foram realizadas com a seqüência da cepa *S. pneumoniae* CGSP14.

**Tabela 4.1.** Dados gerados pelo alinhamento múltiplo, realizado no programa CLUSTAL W, de oito seqüências de aminoácidos da proteína IgA1 protease de *S. pneumoniae* depositadas no *GenBank*.

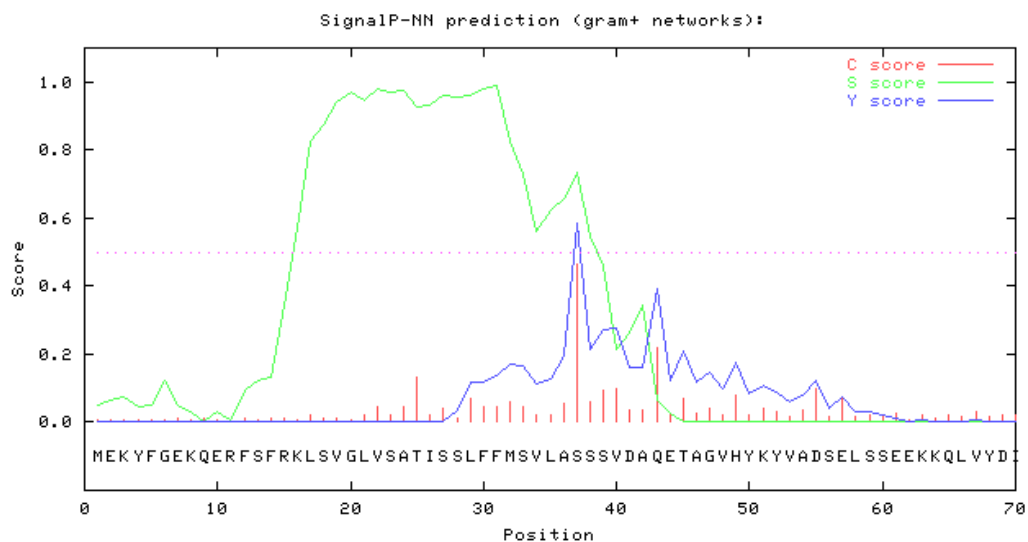
Dados do Alinhamento Múltiplo	
Tamanho do Alinhamento:	2081 aa
Identidade:	1063 aa , 51,08%
Alta similaridade:	293 aa, 14,08%
Baixa similaridade:	153 aa, 7,35%
Diferente:	572 aa, 27,49 %
Seqüências utilizadas:	<i>S. pneumoniae</i> SP195, GI:168488981 ( 1999 aa) <i>S. pneumoniae</i> CDC1087-00, GI:168486448 (1928 aa) <i>S. pneumoniae</i> G54, GI:194398492 ( 1945 aa) <i>S. pneumoniae</i> CGSP14, GI:182629447 ( 1908 aa) <i>S. pneumoniae</i> R6, GI:15903086 ( 1963 aa) <i>S. pneumoniae</i> 70585, GI:225858942 ( 1942 aa) <i>S. pneumoniae</i> TIGR4, GI:15901019 ( 2004 aa) <i>S. pneumoniae</i> Hungary19A-6, GI:169833458 ( 1892 aa)

A proteína IgA1 protease, com 1908 aminoácidos, foi também analisada no programa CLC Workbench 4, através do qual estimou-se o tamanho da proteína em 214,2kDa, valor um pouco acima dos dados encontrados na literatura, os quais indicavam uma proteína variando de 130 a 200 kDa (ROMANELLO *et al.*, 2006). Esta diferença provavelmente ocorre devido à elevada heterogeneidade das seqüências de aminoácidos das diferentes cepas. O ponto isoelétrico da proteína foi estimado em pH 5,61. Utilizando o mesmo programa, foram preditas três regiões transmembranares (hidrofóbicas) na proteína, as quais são mostradas na figura 4.3.



**Figura 4.3.** Regiões transmembranares presentes na proteína IgA1 protease de *S. pneumoniae* CGSP14, segundo predição do programa CLC Workbench 4. (A) Desenho esquemático da região amino terminal de IgA1 protease, apresentando as regiões transmembranares e o peptídeo sinal. (B) Esquema da disposição da proteína na membrana celular, apresentando as três regiões que atravessam a membrana.

Para a verificação de possíveis seqüências sinalizadoras presentes na proteína, a região amino terminal da proteína IgA1 protease foi analisada através do programa SignalP 3.0. A figura 4.4 mostra a análise feita pelo programa SignalP 3.0. Conforme indicado nesta figura, a linha verde mostra a probabilidade de um aminoácido fazer parte da seqüência sinalizadora, enquanto as linhas vermelha e azul indicam o provável começo da proteína. Através desta análise, pode-se inferir que os 36 primeiros aminoácidos da proteína têm elevada probabilidade de fazer parte do peptídeo sinal.



**Figura 4.4.** Análise da região amino terminal da proteína IgA1 protease de *S. pneumoniae* CGSP14. Predição do peptídeo sinal através do programa SignalP 3.0. *S score* (linha verde) mostra a probabilidade de um aminoácido estar presente no peptídeo sinal, *C score* (linha vermelha) e *Y score* (linha azul) mostram onde seja o provável começo da proteína.

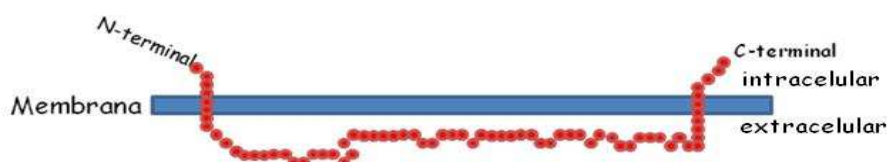
Continuando a análise das proteínas, a sequência de aminoácidos da proteína Hialuronato liase (Hyl) de *S. pneumoniae* CGSP14, depositada no *GenBank* sob o código GI: 182683292, também foi alinhada com outras sequências disponíveis no banco de dados, utilizando o programa Blastp, com o objetivo de identificar sequências da proteína de outras cepas do microrganismo. A partir deste alinhamento, foram selecionadas nove sequências da proteína de diferentes cepas de *S. pneumoniae*. Utilizando estas sequências foi realizado um alinhamento múltiplo no programa CLUSTAL W, com o objetivo de verificar os níveis de identidade entre diversas cepas de *S. pneumoniae*. Os dados do alinhamento são mostrados na tabela 4.2 e o alinhamento completo é mostrado no apêndice B. Os dados do alinhamento, de 1088 aminoácidos, mostram uma identidade de 95,95% entre as sequências, o que sugere que a Hyl é conservada entre diversos sorotipos de *S. pneumoniae*.

Como a proteína Hyl apresenta elevada identidade entre as diferentes cepas de *S. pneumoniae*, e assim como ocorreu na análise das outras duas proteínas, as próximas análises foram realizadas utilizando a sequência da proteína Hyl de *S. pneumoniae* CGSP14 depositada no *GenBank* sob o código GI:182683292.

**Tabela 4.2.** Dados gerados pelo alinhamento múltiplo, realizado no programa CLUSTAL W, de nove seqüências de aminoácidos da proteína Hyl de *S. pneumoniae* depositadas no *GenBank*.

Dados do Múltiplo Alinhamento	
Tamanho do Alinhamento:	1078 aa
Identidade:	1030 aa , 95.55%
Alta similaridade:	11aa, 1,02%
Baixa similaridade:	12 aa, 1,11%
Diferente:	25 aa, 2,32 %
Seqüências utilizadas:	<i>S. pneumoniae</i> D39, GI:116076553 (1067 aa) <i>S. pneumoniae</i> R6, GI:15457838 (1078 aa) <i>S. pneumoniae</i> Taiwan19F-14, GI: 225860354 (1067 aa) <i>S. pneumoniae</i> Hungary19A-6, GI:169833327 (1067 aa) <i>S. pneumoniae</i> CGSP14, GI:182683292 (1078 aa) <i>S. pneumoniae</i> G54, GI:194396877 (1067 aa) <i>S. pneumoniae</i> P1031, GI: 225856058 (1067 aa) <i>S. pneumoniae</i> CDC1087-00, GI: 168486599 ( 1066 aa) <i>S. pneumoniae</i> TIGR4, GI: 15900247 ( 1066 aa)

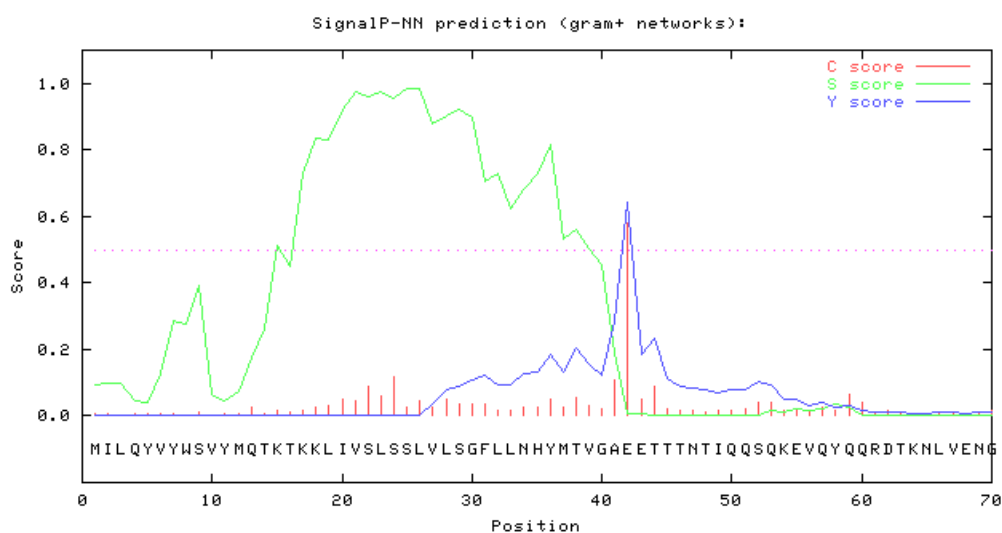
Da mesma forma que no caso das outras proteínas, o tamanho da Hyl foi estimado em 122,1kDa e o ponto isoelétrico em pH 6,07. Foram preditas duas regiões transmembranares (hidrofóbicas) na proteína, uma na região amino terminal e outra na região carbóxi terminal. Um esquema da disposição da proteína na membrana é mostrado na figura 4.5.



**Figura 4.5.** Esquema da disposição da Hyl na membrana celular, de acordo com o programa CLC Workbench 4. As duas regiões da proteína que atravessam a membrana (transmembranares) são representadas no esquema.



Para a verificação da presença de peptídeos sinais na proteína, a região amino terminal da Hyl foi analisada através do programa SignalP 3.0. A figura 4.6 mostra a análise feita pelo programa SignalP 3.0, conforme indicado nesta figura, a linha verde mostra a probabilidade de um aminoácido fazer parte da seqüência sinalizadora, enquanto as linhas vermelha e azul indicam o provável começo da proteína. Através desta análise, pôde-se verificar que os 41 primeiros aminoácidos da proteína têm elevada probabilidade de fazer parte de um peptídeo sinal.



**Figura 4.6.** Análise da região amino terminal da proteína Hyl de *S. pneumoniae* CGSP14. Predição do peptídeo sinal através do programa SignalP 3.0. S score (linha verde) mostra a probabilidade de um aminoácido estar presente no peptídeo sinal, C score (linha vermelha) e Y score (linha azul) mostram onde seja o provável começo da proteína.

#### 4.1.1 Comparação das proteínas: ClpP, IgA1 protease e Hialuronato liase

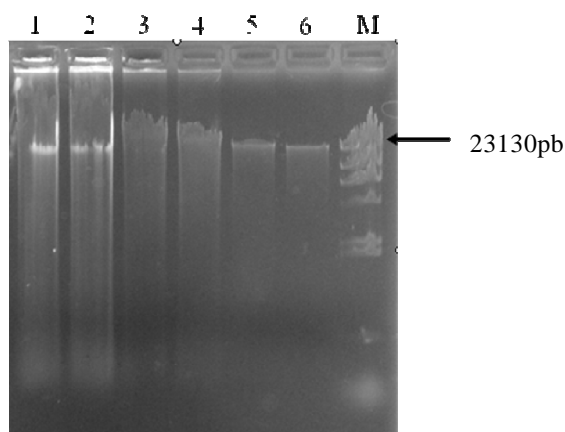
As análises apresentadas acima mostram que a proteína ClpP não possui regiões transmembranares, nem seqüências sinalizadoras, e pelos alinhamentos foi possível perceber que, assim como indicado na literatura (CAO *et al.*, 2009), a ClpP é altamente conservada entre os vários sorotipos de *S. pneumoniae*. Além disso, dados da literatura mostram que esta proteína possui alta imunogenicidade (KWON *et al.*, 2004; CAO *et al.*, 2007; CAO *et al.*, 2009). Já a proteína IgA1 protease possui peptídeo sinal e regiões transmembranares, o que dificultaria a clonagem e expressão da proteína de forma solúvel. Esta proteína se mostra pouco conservada entre os sorotipos, corroborando os

dados que afirmam sua heterogeneidade (MISTRY E STOCKLEY, 2006). No entanto, a IgA1 protease tem mostrado ser imunogênica (ROMANELLO *et al.*, 2006). As análises feitas com a proteína Hyl, mostraram que ela também possui peptídeo sinal e regiões transmembranares, no entanto apresenta alta conservação entre os sorotipos de *S. pneumoniae*. Apesar disso, estudos mostraram que cepas mutantes nesta proteína não tiveram a virulência atenuada, o que indica que ela só pode ser usada em uma vacina quando combinada com outros fatores de virulência (JEDRZEJAS, 2001). Através do exposto acima, a proteína ClpP foi selecionada para a realização deste trabalho devido a sua alta conservação, imunogenicidade e à ausência de peptídeo sinal e regiões transmembranares, visando o alto nível de expressão em sistemas bacterianos.

## **4.2 Amplificação do gene *clpP* de *S. pneumoniae***

### 4.2.1 Extração de DNA genômico de *S. pneumoniae*

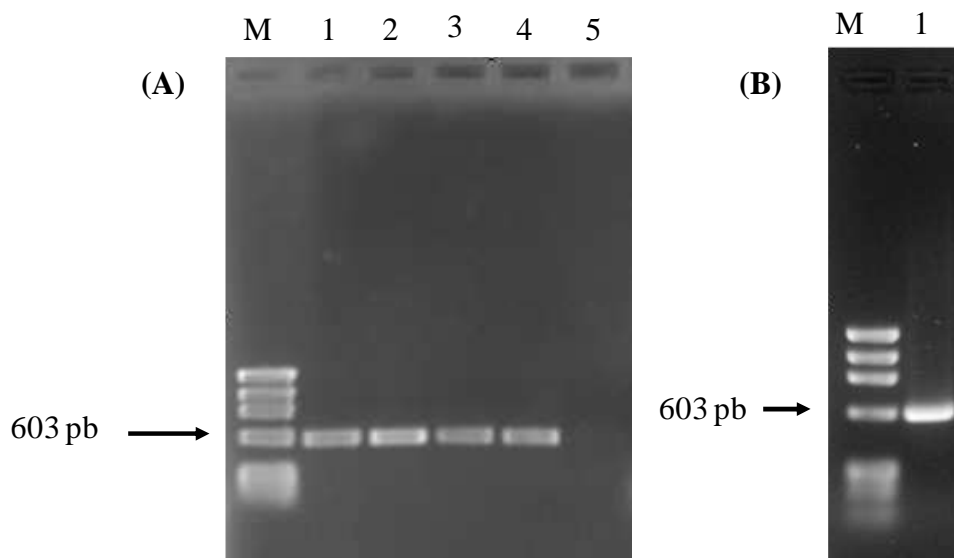
A extração de DNA genômico foi feita a partir do cultivo da cepa *Streptococcus pneumoniae* 113/95, sorotipo 14, depositada no Instituto Adolfo Lutz. O DNA extraído por fenol/clorofórmio foi analisado em gel de agarose 0,8% (Figura 4.7). O gel de agarose 0,8% mostra que o DNA foi extraído por fenol/clorofórmio de maneira eficiente e em alta concentração. Devido à concentração elevada de DNA extraído, este foi utilizado diluído 100 vezes para o isolamento e amplificação do gene *clpP* através de PCR (reação em cadeia da polimerase), isto porque altas concentrações de DNA podem inibir a reação de PCR (LARENTIS *et al.*, 2006). Através desta diluição foi possível a obtenção de aproximadamente 59 ng/μL de DNA (amostra 3) e 30 ng/μL de DNA (amostra 4).



**Figura 4.7** Gel de agarose 0,8% em TAE mostrando o perfil eletroforético da solução resultante da extração de DNA genômico de *S. pneumoniae*. **1 e 2**, DNA genômico extraído de *S. pneumoniae* (amostras 3 e 4 respectivamente); **3 e 4**, DNA genômico extraído de *S. pneumoniae* diluído 10 vezes (amostras 3 e 4 respectivamente); **5 e 6**, DNA genômico extraído de *S. pneumoniae* diluído 100 vezes (amostras 3 e 4 respectivamente); **M**, padrão de tamanho de fragmento (DNA  $\lambda$  digerido com *Hind* III). A seta indica a banda do marcador com 23130bp.

#### 4.2.2 Amplificação do gene *clpP* por PCR (Reação em Cadeia da Polimerase)

Através da técnica de PCR, utilizando os oligonucleotídeos desenhados especificamente para esta amplificação, foi possível amplificar o gene *clpP* a partir de DNA genômico extraído de *S. pneumoniae*. A reação de PCR amplificou fragmentos de 599 pb (Figura 4.8A), o que era esperado de acordo com as seqüências do gene depositadas no *GenBank*, utilizadas nas análises de bioinformática e para o desenho dos oligonucleotídeos iniciadores. A reação de PCR gradiente foi analisada através de eletroforese em gel de agarose 0,8%, mostrado na figura 4.8A. É possível perceber que em todas as temperaturas de anelamento utilizadas, o gene foi amplificado de forma específica e com bom rendimento. Como todas as temperaturas de anelamento apresentaram rendimento similar, a temperatura de anelamento de 55°C foi escolhida para as próximas reações de PCR realizadas neste trabalho. A temperatura de 55°C foi escolhida por ser a intermediária entre as testadas visando à maior eficiência da reação, visto que temperaturas baixas diminuem a especificidade e temperaturas altas podem diminuir a eficiência do anelamento. Após a definição das condições empregadas na reação de PCR e a confirmação da amplificação do gene *clpP*, o produto de PCR foi purificado. O gene purificado foi visualizado em gel de agarose (Figura 4.8B), também com tamanho esperado de 599 pb.



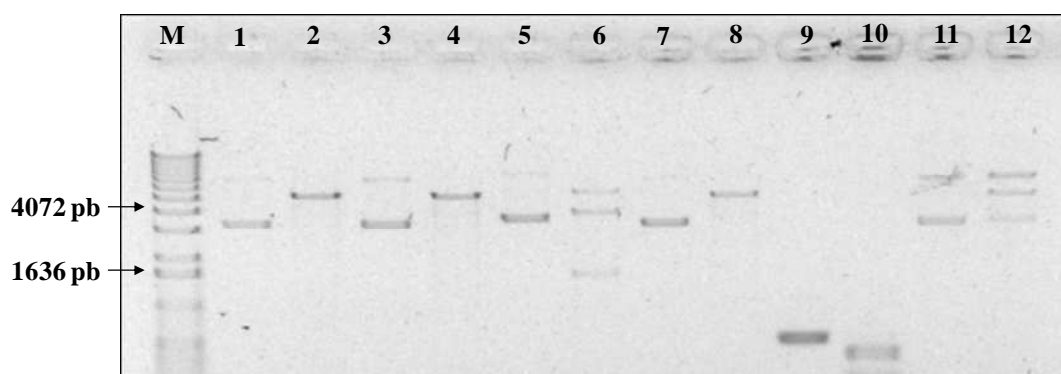
**Figura 4.8** (A) Gel de agarose 0,8% em TAE da amplificação do gene *clpP* utilizando diferentes temperaturas de anelamento no programa do PCR. **M**, padrão de tamanho de fragmento ( $\Phi$ X174 DNA digerido com *Hae* III); Gene *clpP* amplificado por PCR utilizando temperatura de anelamento dos oligonucleotídeos iniciadores igual a 54°C (1), 55,1°C (2), 56°C (3), 55,9°C (4); Controle negativo da reação de PCR (5); A seta indica o banda do marcador com 603bp, próximo ao tamanho esperado do gene *clpP* amplificado. (B) Gel de agarose 1,5% em TAE do produto de PCR purificado. **M**, padrão de tamanho de fragmento ( $\Phi$ X174 DNA digerido com *Hae* III); **1**, Purificação do gene *clpP* amplificado; A seta indica o banda do marcador com 603bp, próximo ao tamanho esperado do gene *clpP* amplificado.

### 4.3 Clonagem e expressão utilizando *Escherichia coli* como sistema de expressão

#### 4.3.2 Clonagem utilizando o vetor pET28b

Através do gene amplificado e purificado foram realizadas as reações de digestão e ligação ao vetor pET28b. A estratégia utilizada para amplificação do gene e ligação ao vetor possibilita a expressão da proteína com cauda de histidina na posição C-terminal da proteína, já que o códon de terminação da tradução, desta forma, ficou localizado após a sequência que codifica a cauda de histidina. A célula eletrocompetente *E. coli* TOP10 foi transformada com o vetor construído, denominado pET28b/*clpP*, por eletroporação. Após a transformação, as células foram plaqueadas em meio LB com antibiótico. Foram selecionadas quatro colônias desta placa, como sendo possíveis cepas recombinantes com a construção pET28b/*clpP* correta. Estes clones tiveram seus plasmídeos extraídos e submetidos à digestão com enzimas de restrição específicas para

sua confirmação. As reações de digestão foram analisadas em gel de agarose (Figura 4.9).

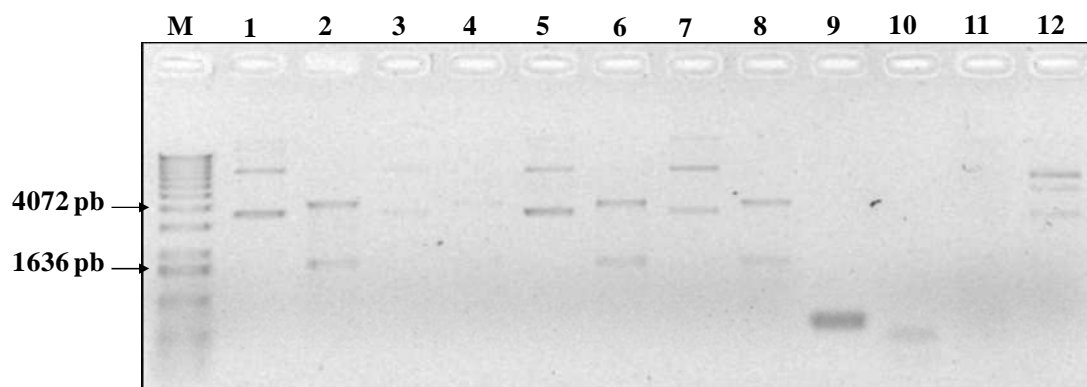


**Figura 4.9** Confirmação dos clones - Digestão dos plasmídeos extraídos das cepas transformadas *E. coli* TOP 10 com o vetor construído. **M**, padrão de tamanho de fragmento 1 kb DNA Ladder (Invitrogen); **1, 3, 5 e 7**, plasmídeos extraídos dos clones C1N, C8N, C2P e C4P, respectivamente; **2, 4, 6 e 8**, plasmídeos digeridos extraídos dos clones C1N, C8N, C2P e C4P, respectivamente; **9 e 11**, *clpP* e pET28b, respectivamente; **10** digestão do gene *clpP* com *PstI* (controle positivo); **12**, digestão do vetor pET28b com *EcoRV* (controle positivo). As setas indicam as bandas do padrão de tamanho de fragmento mais próximas as bandas esperadas na digestão de clones positivos.

Através da análise da figura 4.9 é possível visualizar que apenas o plasmídeo extraído do clone *E. coli* TOP10/pET28b/*clpP*-C2P apresentou os padrões de bandas esperados, 1697pb e 4134pb, ao ser digerido pelas enzimas *EcoRV* no plasmídeo e *PstI* no gene. Os plasmídeos das demais cepas recombinantes apresentaram apenas uma banda característica do vetor pET28b linearizado pela enzima *EcoRV*, indicando que estas cepas possuíam apenas o plasmídeo pET28b, sem o gene *clpP*. A partir deste resultado pode-se inferir que apenas o clone C2P possui o plasmídeo com o inserto *clpP*, sendo possível confirmar esta cepa recombinante como positiva. O plasmídeo extraído deste clone foi seqüenciado e também transformado na cepa de expressão *E. coli* BL21 Star (DE3), previamente submetida ao protocolo de obtenção de células competentes.

As colônias, geradas pela cepa *E. coli* BL21 Star (DE3) transformada com a construção pET28b/*clpP*-C2P, passaram pela mesma confirmação que as cepas de clonagem. Os plasmídeos de quatro colônias foram extraídos e confirmados por digestão com enzimas de restrição. Todas as quatro colônias se mostraram positivas por

apresentarem o padrão de bandas da digestão esperado. Este resultado pode ser visualizado na figura 4.10.



**Figura 4.10** Digestão dos plasmídeos extraídos das cepas transformadas *E. coli* BL21 Star (DE3). **M**, padrão de tamanho de fragmento 1 kb DNA Ladder (Invitrogen); **1, 3, 5 e 7**, plasmídeos extraídos dos clones C2P/1, C2P/2, C2P/3, C2P/4, respectivamente; **2, 4, 6 e 8**, plasmídeos digeridos extraídos dos clones C2P/1, C2P/2, C2P/3, C2P/4, respectivamente; **9 e 11**, *clpP* e pET28b, respectivamente; **10** digestão do gene *clpP* com *Pst*I (controle positivo); **12**, digestão do vetor pET28b com *EcoRV* (controle positivo). As setas indicam as bandas do padrão de tamanho de fragmento mais próximas as bandas esperadas na digestão de clones positivos.

#### 4.3.3 Sequenciamento

A construção pET28b/*clpP*-C2P foi sequenciada a fim de confirmar a seqüência correta do gene a ser expresso. O resultado do sequenciamento (seqüência consenso) é mostrado na Figura 4.11.

```

ATGGTTCCTGTAGTTAT TGAACAAACAAGCCGTGGAGAACGTTCTTACGATATTTACTCACGTCTTCTCAAAGACCGCATC
ATTATGCTGCACAGGTCGGTT GAAGACAATA TGCTAATCTGTTATTGCCCAATT GCTTTTCTTGGATGCCCAAGATAGT
ACGAAAGATATT TACCTTTATGTCAA TACACCAGGTGGTTCGTTTT CAGCTGGTTTGGCAA TCGTAGATACCATGAACTTT
ATCAAGGCAGATGTCCAAACCATTGT TATGGGAATGGCTGCATCTATGGGGACTGT CATCGCATCAAGTGGAGCAAAGGC
AAACGTTTCATGCTTCCAAATGCTGAATACATGATT CACCAACCAATGGGCGGTACAGGTGGTGGTACCCAACAACTGAT
ATGGCTATCGCTGCAGAACACTTGCT CAAAACCTGTAATACCTTGGAAAAAATCTTGGCTGAAAATTCAGGTGAGTCAATG
GAAAAAGTCCATGCAGATGCAGAACGTGATAACTGGATGAGCGCCAGGAAACACT TGAATATGGCTTTATTGATGAAATT
ATGGCCAACAATTCATTGAACCTCGAGCACCACCACCACCACCCTGA
  
```

**Figura 4.11** Sequenciamento do Plasmídeo pET28b/*clpP*-C2P. A figura mostra a seqüência de nucleotídeos do gene *clpP* (preto), o sítio de restrição da enzima *Xho*I (rosa), a seqüência da cauda de histidina (azul) e o *códon* de terminação (vermelho), presentes no plasmídeo construído e sequenciado.

Para a análise do seqüenciamento, a seqüência consenso obtida foi alinhada com outras seqüências do gene *clpP*, disponíveis no *GenBank*, através do programa CLUSTAL W. Através deste alinhamento foram encontrados 5 nucleotídeos diferentes entre o consenso e a seqüência de *S. pneumoniae* CGSP14 depositada no *GenBank*. A troca do nucleotídeo na posição 4 da seqüência, adenina por guanina, foi feita propositalmente quando os oligonucleotídeos foram desenhados para conseguir inserir a seqüência da enzima *NcoI* e manter a temperatura de fusão dos oligonucleotídeos baixa. Esta mudança resultou na troca do segundo aminoácido da cadeia, isoleucina por valina. As outras quatro mudanças de nucleotídeos são mostradas na tabela 4.3. As mudanças são todas silenciosas, ou seja, não acarretam mudanças de aminoácidos. Além disso, alguns destes nucleotídeos se mostram diferentes na cepa CGSP14, mas idênticos em outras utilizadas no alinhamento. Sendo assim, é possível que as mudanças ocorridas não tenham sido por nucleotídeos erroneamente incorporados pela ação da DNA polimerase durante a PCR, mas, que estes nucleotídeos já estavam presentes no DNA utilizado como molde na reação de amplificação do gene, já que este DNA se refere ao da cepa depositada no Instituto Adolfo Lutz e cuja seqüência não está depositada no *GenBank*.

**Tabela 4.3** Mutações detectadas no seqüenciamento do plasmídeo pET28b/*clpP*-C2P quando comparada a seqüência obtida pelo seqüenciamento e a seqüência de *S. pneumoniae* CGSP14 depositada no *GenBank*.

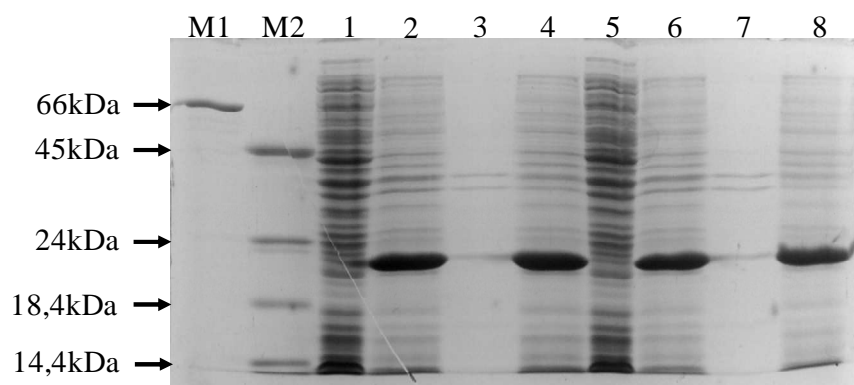
Posição	Nucleotídeo	Aminoácido	Mutação
Nucleotídeo 136	C → T	Leucina	Silenciosa
Nucleotídeo 165	A → G	Treonina	Silenciosa
Nucleotídeo 294	A → G	Glicina	Silenciosa
Nucleotídeo 297	A → T	Treonina	Silenciosa

#### 4.3.4 Expressão da proteína recombinante ClpP

Com o objetivo de verificar a expressão nos quatro clones selecionados e confirmados da cepa *E. coli* BL21 Star (DE3), estes foram cultivados a 37°C sob agitação de 200 rpm, em meio LB com 50 µg/mL de canamicina até a fase exponencial

de crescimento ( $Abs_{600nm}$  aproximadamente 0,7). Nesta fase, a expressão da proteína foi induzida, pela adição de 1 mM de IPTG (isopropil- $\beta$ -D-tiogalactosídeo), por 4 horas. As análises da expressão da proteína, feitas através de SDS-PAGE, mostraram que os quatro clones expressam a proteína ClpP em concentrações semelhantes e com o tamanho esperado (cerca de 21,5 kDa). Com isso foram utilizados dois destes clones para fazer as análises de expressão e solubilidade da proteína recombinante. Os clones C2P/2 e C2P/4 foram cultivados e expressos nas mesmas condições anteriores. Ao final da expressão foram retiradas amostras, com as quais os extratos protéicos foram preparados, sendo separadas as frações solúvel e insolúvel das proteínas totais. Estas amostras também foram analisadas em SDS-PAGE, mostrado na figura 4.12. Através desta análise é possível perceber que os dois clones expressam a proteína, visualizada no gel entre as bandas de 18,4 kDa e 24 kDa do marcador de massa molar. Além disso, esta mesma banda não é visualizada antes da indução da expressão (amostra não induzida), não havendo “vazamentos” na expressão, já que é utilizado o sistema do promotor do bacteriófago T7 RNA polimerase, que é fortemente regulado e permite a expressão da proteína recombinante apenas quando adicionado o indutor. A proteína é expressa em níveis semelhantes pelos dois clones, em altas concentrações e de forma solúvel, não ocorrendo a formação de corpos de inclusão. De acordo com BANEYX (1999), um dos problemas da superprodução de proteínas heterólogas dentro do citoplasma desta bactéria é a formação de proteínas insolúveis em forma de agregados (corpos de inclusão) ocasionados pela perda da conformação da proteína, contudo isto não foi verificado no presente trabalho.





**Figura 4.12** Gel de SDS-PAGE 12,5% dos extratos protéicos totais e de suas frações, solúvel e insolúvel, obtidos através da expressão da proteína ClpP por *E. coli* BL21 Star (DE3). **M1** e **M2**, padrão de massa molar; **1** e **5**, amostras do cultivo não induzido dos clones C2P/2 e C2P/4, respectivamente; **2** e **6**, extratos protéicos totais dos clones C2P/2 e C2P/4, respectivamente; **3** e **7**, frações insolúveis dos clones C2P/2 e C2P/4, respectivamente; **4** e **8**, frações solúveis dos clones C2P/2 e C2P/4, respectivamente. As amostras aplicadas no gel foram normalizadas por  $Abs_{600nm}$ .

Após estas análises, como os dois clones apresentados no gel acima mostram os mesmos níveis de expressão, foram feitos estoques em glicerol para um banco de trabalho com a cepa *E. coli* BL21 Star (DE3)/pET28b/clpP-C2P/4. O banco de trabalho, utilizado para fazer todas as expressões a partir deste ponto, foi feito a partir de um cultivo em fase de crescimento exponencial, com  $Abs_{600nm}$  aproximadamente 1,0.

#### 4.3.5 Otimização da expressão e estudos de segregação plasmidial

Para analisar a influência de algumas variáveis importantes na expressão da proteína ClpP, foi utilizada a ferramenta de planejamento de experimentos, através da realização de um planejamento fatorial a dois níveis com duas variáveis ( $2^2$ ) e quatro réplicas no ponto central. As duas variáveis independentes utilizadas foram: a concentração de indutor do sistema, IPTG, e a concentração do antibiótico canamicina, o qual é responsável pela pressão seletiva do sistema. Através do planejamento foi avaliada a influência destas variáveis independentes sobre o crescimento celular, sobre a produção de proteína e sobre a segregação plasmidial. Os experimentos foram realizados como descrito na seção 3.2.10. A expressão da proteína foi realizada por 4 horas e ao final deste tempo foram retiradas amostras para leituras de  $Abs_{600nm}$ , as quais foram convertidas para massa seca de células conforme descrito na seção 3.2.10, com o

objetivo de avaliar o crescimento celular. Foram também retiradas amostras para a quantificação da concentração de proteína expressa (por densitometria) e para análises de segregação plasmidial (através de diluições sucessivas e crescimento em placas com e sem antibiótico). As condições empregadas em cada um dos experimentos do planejamento fatorial  $2^2$  são mostradas na tabela 4.4, juntamente com as respostas das variáveis dependentes analisadas.

**Tabela 4.4** Experimentos do planejamento fatorial e suas variáveis de resposta

<b>Experimento</b>	<b>IPTG (mM)</b>	<b>Canamicina (<math>\mu\text{g/mL}</math>)</b>	<b>Massa seca de células (mg/mL)</b>	<b>ClpP (mg /L)</b>	<b><i>F</i> (fração de células com plasmídeo)</b>
<b>1</b>	0,1	0	0,91	247,3	0,44
<b>2</b>	0,1	50	0,87	251,1	0,55
<b>3</b>	1	0	0,71	271,6	0,05
<b>4</b>	1	50	0,76	225,9	0,09
<b>PC5</b>	0,55	25	0,74	266,3	0,07
<b>PC6</b>	0,55	25	0,72	244,2	0,23*
<b>PC7</b>	0,55	25	0,77	229,7	0,06
<b>PC8</b>	0,55	25	0,76	221,2	0,11

\* possível outlier

Através dos dados mostrados na tabela 4.4, é possível perceber que a concentração de proteína se mantém em níveis semelhantes em todas as condições empregadas no planejamento. Já os experimentos com menos IPTG apresentaram crescimento celular acima dos demais, independente da concentração de canamicina. Também apresentaram maior estabilidade plasmidial, já que os valores de *F* (fração de células com plasmídeo) dos experimentos 1 e 2 estão acima dos outros experimentos. É possível verificar uma tendência dos valores de *F*, já que no menor valor de IPTG a fração de células com plasmídeo se apresenta entre 44% e 55%, enquanto nos outros experimentos, induzidos com concentrações maiores de IPTG o valor de *F* está em uma faixa de 5% a 11%, desconsiderando o experimento PC6, ou seja, experimentos que utilizam concentrações menores de IPTG apresentam menor segregação plasmidial. Este

efeito do IPTG sobre a segregação é relatado na literatura, pois são encontrados alguns trabalhos que discutem a influência negativa de indutores utilizados em promotores baseados no operon *lac*, como o IPTG, sobre a estabilidade plasmidial, reduzindo o valor de  $F$  (FRIEHS, 2004).

O valor de  $F$  do experimento PC6 pode tratar-se de um *outlier*, pois ele apresenta um valor diferente da tendência apresentada pelos demais valores de  $F$  das réplicas no ponto central. Um *outlier* é definido como um ponto experimental que parece não se adequar a uma distribuição particular de probabilidades definida pela grande maioria dos demais pontos experimentais (SCHWAAB E PINTO, 2007). O cálculo para detecção de um *outlier* pode ser realizado através do cálculo do intervalo:  $\mu \pm 2 \cdot \sigma$ , onde  $\mu$  é a média no ponto central e  $\sigma$  é o desvio padrão. Através deste cálculo, utilizando os valores de  $F$  dos três pontos centrais (PC) na faixa entre 0,06 e 0,11 para o cálculo da média e desvio padrão, obtém-se um intervalo de dados entre 0,02 e 0,14. Sendo assim, o valor de  $F$  do experimento PC6 está fora do intervalo calculado e pode ser considerado um *outlier*. Desta forma, valor de  $F$  do experimento PC6 pode ser desconsiderado nas análises estatísticas. No entanto, a análise dos efeitos será realizada tanto utilizando quanto não utilizando o valor de  $F$  do experimento PC6, a fim de comparar os resultados obtidos.

Através dos experimentos com 0,55 mM de IPTG e 25  $\mu\text{g/mL}$  de canamicina (réplicas no ponto central) obteve-se média de 240,4 mg/L de ClpP com desvio padrão de 19,7 e desvio padrão relativo de 8%. Obteve-se também média de 0,75mg/mL de massa seca de células com desvio padrão de 0,02 e desvio padrão relativo de 2,5%. A variável  $F$  apresentou média de 0,08 com desvio padrão de 0,03 e desvio padrão relativo de 36%.

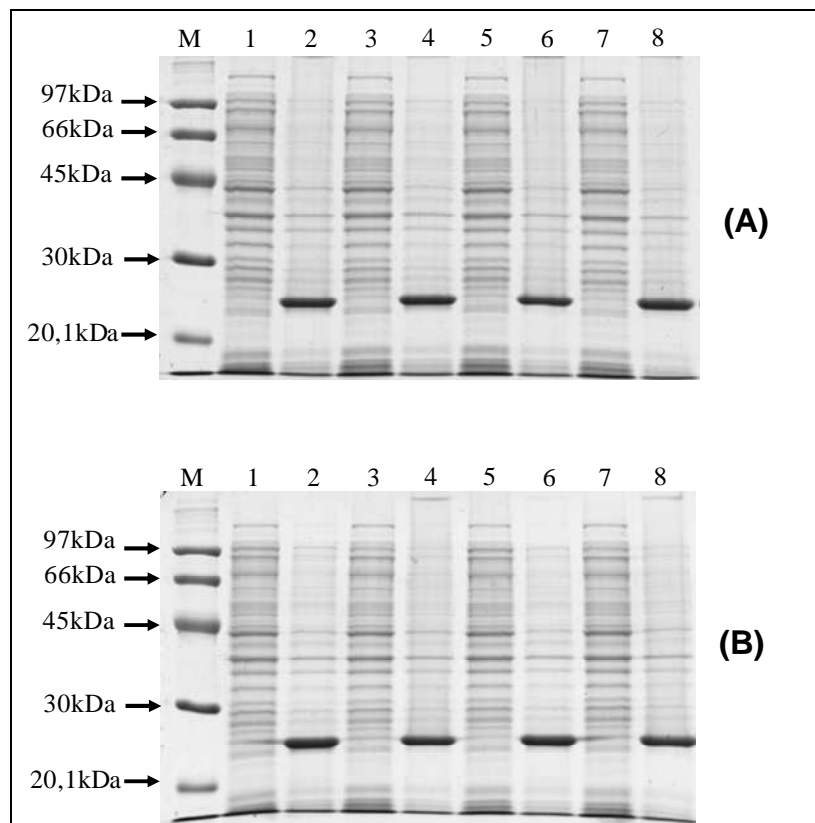
Para analisar estatisticamente as respostas obtidas no planejamento, o programa STATISTICA 6.1 foi utilizado considerando significativos os efeitos com  $p < 0,1$ . Primeiramente foram avaliados os efeitos das variáveis IPTG e canamicina sobre o crescimento celular (Tabela 4.5). Analisando a tabela 4.5 observa-se que o  $p$ -valor para a variável IPTG é menor que 0,1, ou seja, a variável IPTG influencia significativamente o crescimento celular, de forma negativa. Logo pode-se inferir, com 90% de confiança,

que a concentração de IPTG influencia significativamente de forma negativa o crescimento celular. Alguns autores já relataram o efeito tóxico do IPTG (ANDERSSON *et al.*, 1996) sendo que a indução com altas concentrações de IPTG pode diminuir drasticamente o crescimento celular (LÉON *et al.*, 2004). Já o fator canamicina apresentou *p*-valor maior que 0,1, logo a concentração deste antibiótico não influencia significativamente o crescimento celular.

**Tabela 4.5** Efeitos das variáveis sobre o crescimento celular (mg/mL de massa seca de células)

	Efeitos	Erro Padrão	t(5)	<i>p</i> -valor
Média	0,779	0,016	47,31	<0,0001
IPTG	-0,152	0,046	-3,27	0,022
Canamicina	0,008	0,046	0,185	0,86

O valor da concentração de proteína de cada experimento foi obtido através de densitometria. Os géis de SDS-PAGE dos extratos protéicos totais dos experimentos do planejamento, utilizados para a análise de densitometria, são mostrados na figura 4.13. Observando os géis é possível verificar, mais uma vez, que a proteína ClpP foi expressa em níveis similares em todos os experimentos. Além disso, a proteína é expressa em todos os experimentos em altas concentrações e de forma solúvel a 37°C e 200rpm, o que é vantajoso, já que *E. coli* normalmente expressa proteínas insolúveis nestas condições, como já discutido na seção 4.3.4. A concentração de ClpP obtida nestes experimentos está dentro da faixa de concentrações obtidas por outros trabalhos da literatura, onde são realizados experimentos de otimização de expressão de outras proteínas em *E. coli*, utilizando frascos agitados e cultivos em batelada (CHOI *et al.*, 2006; MALDONADO *et al.*, 2007; CHUAN *et al.*, 2008; HAN *et al.*, 2008). Dados da literatura mostram que concentrações mais elevadas de proteínas produzidas em *E. coli*, em termos de g/L, são alcançadas quando a expressão da proteína heteróloga é realizada e otimizada em bioreatores, onde as condições de cultivo podem ser controladas (CHOI *et al.*, 2006).



**Figura 4.13** Gel de SDS-PAGE 12,5% com o extrato total das proteínas dos experimentos do planejamento fatorial. (A) **M**, padrão de massa molar LMW (Amersham Bioscience); **1, 3, 5, e 7**, extrato total sem indução dos experimentos 1, 2, 5(PC) e 7(PC), respectivamente; **2, 4, 6 e 8**, extrato total com 4h de indução dos experimentos 1, 2, 5(PC) e 7(PC), respectivamente. (B) **M**, padrão de massa molar LMW (Amersham bioscience); **1, 3, 5, e 7**, extrato total sem indução dos experimentos 3, 4, 6(PC) e 8(PC), respectivamente; **2, 4, 6 e 8** extrato total com 4h de indução dos experimentos 3, 4, 6(PC) e 8(PC), respectivamente. As amostras foram normalizadas pela  $Abs_{600nm}$ .

Através das análises estatísticas foram calculados os efeitos dos fatores sobre a concentração de proteína. Os resultados, visualizados na tabela 4.6, indicam que tanto a concentração de IPTG, quanto a concentração de canamicina, nas faixas testadas, não influenciam significativamente na expressão da proteína, já que o *p*-valor em ambos os casos é maior que 0,1. Por meio destas análises é possível indicar que, nas condições estudadas, pode-se diminuir a concentração do indutor em até 10 vezes (vantajoso também pelo efeito negativo sobre o crescimento celular), e até mesmo retirar a canamicina do sistema sem prejudicar significativamente a concentração de proteína, desta forma reduzindo custos e mantendo a expressão em níveis similares.

**Tabela 4.6** Efeitos das variáveis sobre concentração de ClpP (mg/L)

	Efeitos	Erro Padrão	t(5)	<i>p</i> -valor
Média	244,68	6,95	35,21	<0,0001
IPTG	-0,486	19,65	-0,024	0,981
Canamicina	-20,98	19,65	-1,067	0,334

Da mesma forma que para o crescimento celular e para a concentração de proteína, foram avaliados os efeitos das variáveis independentes IPTG e canamicina sobre a estabilidade plasmidial através da análise da variável dependente *F* (fração de células com plasmídeo). Para tanto, foram considerados significativos os efeitos com  $p < 0,1$ , ou seja, as análises foram feitas com grau de confiança de 90%, assim como nos demais casos analisados anteriormente. Na tabela 4.7 são mostrados os efeitos das variáveis sobre a estabilidade plasmidial, considerando o valor de *F* do experimento PC6 um *outlier*, ou seja, desconsiderando este valor nas análises estatísticas. Através destas análises é possível concluir, com 90% de confiança, que a concentração de IPTG, na faixa testada, possui efeito negativo sobre a estabilidade plasmidial, e que a concentração de canamicina no sistema não apresenta efeito significativo sobre *F*, dentro das faixas estudadas. Isso significa que quanto mais próxima de 0,1 mM for a concentração de IPTG utilizada, menor será a segregação plasmidial (perda do plasmídeo), em qualquer concentração de canamicina dentro da faixa estudada.

**Tabela 4.7** Efeitos das variáveis sobre o valor de *F* (fração de células com plasmídeo), desconsiderando o valor de *F* do experimento PC6.

	Efeitos	Erro Padrão	t(4)	<i>p</i> -valor
Média	0,194	0,05	3,825	0,019
IPTG	-0,419	0,134	-3,11	0,036
Canamicina	0,074	0,134	0,557	0,607

Os efeitos das variáveis independentes, IPTG e canamicina, sobre a estabilidade plasmidial também foram avaliados utilizando o valor de *F* do experimento PC6 nas análises estatísticas. Os efeitos das variáveis sobre o valor de *F* utilizando os valores do experimento PC6 são mostrados na tabela 4.8. Assim como na análise anterior, é possível concluir, através da análise da tabela 4.8, que a concentração de IPTG possui

efeito negativo sobre a estabilidade plasmidial, e que a concentração de canamicina no sistema não apresenta efeito significativo sobre  $F$ , mesmo utilizando o valor de  $F$  do experimento PC6. Ou seja, considerando ou desconsiderando o valor de  $F$  do experimento PC6 nas análises estatísticas, o resultado da avaliação dos efeitos das variáveis sobre a estabilidade plasmidial é o mesmo: o IPTG possui efeito negativo sobre a estabilidade plasmidial e a canamicina não apresenta efeito significativo sobre esta variável.

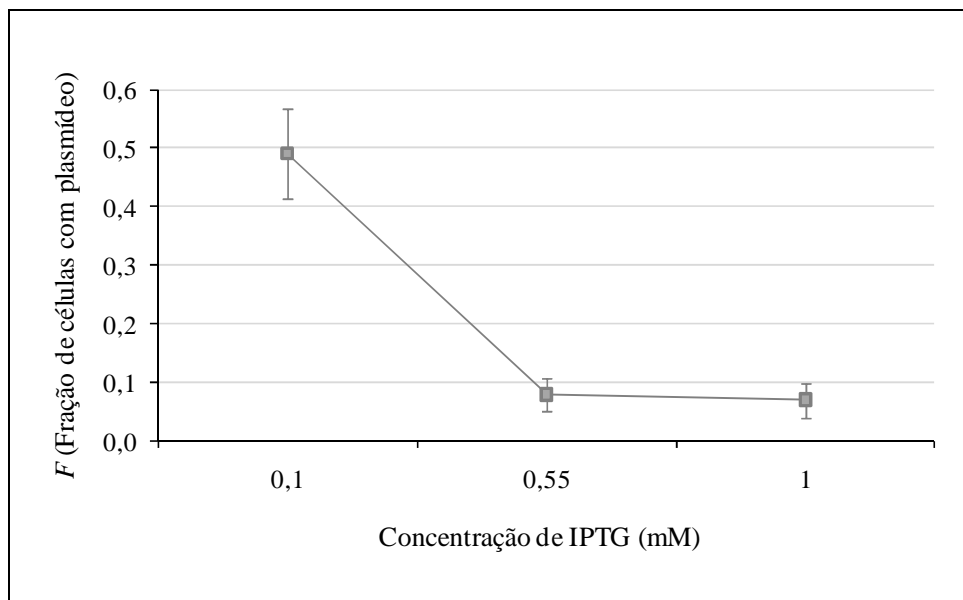
**Tabela 4.8** Efeitos das variáveis sobre o valor de  $F$  (fração de células com plasmídeo), considerando o valor de  $F$  do experimento PC6.

	Efeitos	Erro Padrão	t(5)	$p$ -valor
Média	0,199	0,004	4,65	0,006
IPTG	-0,419	0,121	-3,44	0,018
Canamicina	0,074	0,121	0,616	0,564

O efeito negativo da indução com IPTG sobre a segregação é mencionado em alguns trabalhos na literatura (FRIEHS, 2004). Em seus estudos, MARÍ *et al.* (1999) demonstram que na presença de IPTG ocorre uma diminuição na estabilidade plasmidial quando comparada a culturas sem o indutor, na presença ou ausência de antibiótico. Portanto, este trabalho indica que a estabilidade plasmidial é dependente da variável IPTG e independente do marcador de seleção (antibiótico) (MARÍ *et al.*, 1999). Além disso, os dados apresentados por XU *et al.* (2006) indicam que a segregação plasmidial é muito mais dependente da indução do que da presença ou ausência de antibiótico. Isto porque, em seus estudos XU *et al.* (2006) mostram que a estabilidade plasmidial em culturas não induzidas apresenta valores em torno de 95% com antibiótico e 90% sem antibiótico; já em culturas induzidas, a estabilidade diminui drasticamente, em torno de 15% com antibiótico e 10% sem antibiótico.

Através dos valores de  $F$  (fração de células com plasmídeo) mostrados na tabela 4.4, pode-se perceber que  $F$  não apresenta um comportamento linear, o que foi confirmado pelo baixo valor do coeficiente de ajuste ao modelo ( $R^2$ ), igual a 0,72. Como para realizar as análises de um plano fatorial  $2^2$ , como o empregado, só é possível

avaliar os coeficientes de regressão linear de cada variável, o valor baixo de  $R^2$  indica que o modelo linear não se ajusta bem aos dados. O comportamento não linear do sistema, também pode ser visualizado pelo gráfico apresentado na figura 4.14. Assim, de acordo com as faixas estudadas, para obter menores níveis de segregação plasmidial, deve-se utilizar o indutor IPTG em concentração de 0,1mM. Estes dados ainda sugerem a possibilidade de existir um ponto ótimo, menor que 0,1mM de IPTG, onde a segregação seja mínima, e a expressão da proteína continue em altos níveis. Para confirmar esta hipótese e avaliar a concentração mínima de IPTG, novos estudos precisam ser realizados, ajustando esta concentração em relação ao número de células, uma vez que o indutor liga-se ao promotor T7.



**Figura 4.14** Gráfico de  $F$  (fração de células com plasmídeo) em função da concentração de IPTG, mostrando o efeito não linear do sistema.

Através das análises do planejamento fatorial foi possível inferir a condição otimizada dentro dos limites estudados. A condição otimizada com base no planejamento foi a utilizada no experimento 1 (0,1mM de IPTG e 0 $\mu$ g/mL de canamicina). Utilizando esta condição é possível reduzir 10 vezes a concentração de indutor e retirar a canamicina do sistema, apresentando elevada concentração de proteína e crescimento celular, e ainda mantendo a estabilidade plasmidial em níveis que não prejudiquem a produção da proteína recombinante nas 4 horas de expressão.



*4.3.5.1 Validação da condição otimizada pelo planejamento de experimentos - Estudo da cinética de produção da proteína e segregação plasmidial.*

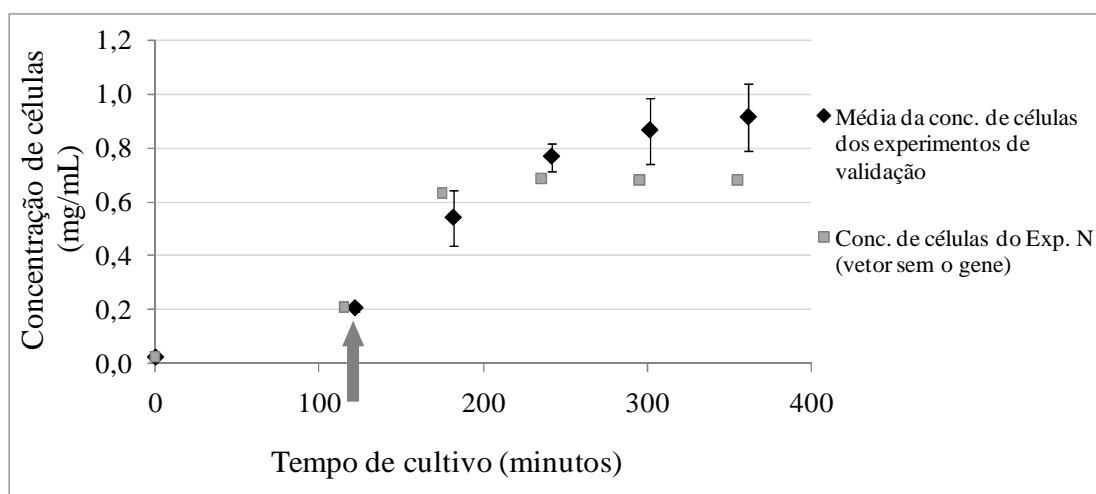
Para validar a condição otimizada pelo planejamento foram realizadas 5 réplicas de cultivos nesta condição. Foram realizados 5 experimentos (induzidos com 0,1 mM IPTG e 0 µg/mL canamicina), denominados A, B, C, D e E. Além destes experimentos também foi realizado um cultivo nas mesmas condições, no entanto, utilizando a cepa transformada apenas com o vetor pET28b (sem o gene), como controle negativo, denominado experimento N. Os cultivos foram realizados até atingirem fase exponencial de crescimento ( $Abs_{600nm}$  aproximadamente 0,7) e então induzidos com 0,1 mM de IPTG. A partir deste ponto foram retiradas amostras de hora em hora para as análises, até o término das 4 horas de indução. Os experimentos A, B, C, D, E e N (sem o gene) foram analisados quanto ao crescimento celular (massa seca de células - mg/mL) e quanto à produção de proteína recombinante ao longo do tempo. Para a análise de segregação plasmidial foram utilizados apenas os experimentos B, D, E e N (sem o gene), isto porque esta é uma análise mais demorada.

Os resultados dos crescimentos celulares das réplicas são mostrados na tabela 4.9. O crescimento celular médio (expresso em massa seca de células – mg/mL) obtido nestas réplicas foi 0,91mg de células/mL ao final das 4 horas de indução, com desvio padrão relativo de 13%, o que está de acordo com o valor obtido no experimento 1 do planejamento fatorial. O crescimento celular também foi acompanhado durante todo o tempo de cultivo e o gráfico da cinética do crescimento celular é mostrado na figura 4.15.

**Tabela 4.9** Crescimento celular dos experimentos de validação do planejamento fatorial

Tempo de cultivo (minutos)	Tempo de indução (horas)	Exp. A Massa seca de células (mg/mL)	Exp. B Massa seca de células (mg/mL)	Exp. C Massa seca de células (mg/mL)	Exp. D Massa seca de células (mg/mL)	Exp. E Massa seca de células (mg/mL)
0	-	0,020	0,027	0,019	0,023	0,025
122	0	0,211	0,219	0,191	0,204	0,202
182	1	0,603	0,681	0,531	0,434	0,451
242	2	0,753	0,841	0,772	0,697	0,769
302	3	0,758	1,065	0,791	0,825	0,880
362	4	0,794	1,121	0,927	0,858	0,863

A figura 4.15 mostra o gráfico de concentração de células (em massa seca de células - mg/mL) em função do tempo de cultivo dos experimentos do ponto otimizado e do crescimento utilizando a cepa *E. coli* BL21 Star (DE3) apenas com o vetor pET28b sem o gene (experimento N – controle negativo). Pode-se perceber, que após 2 horas de indução (242 minutos de cultivo), as células começam a chegar à fase estacionária de crescimento. Alguns autores afirmam que quando são utilizados sistemas com promotores fortes, como é o caso do promotor T7, ao induzir o sistema, a taxa de crescimento diminui devido à sobrecarga do metabolismo da célula hospedeira (TOKSOY *et al.*, 2002). Já o experimento N apresentou um crescimento um pouco menor que os demais. Este crescimento menor pode representar os mesmos números que os demais devido aos erros presentes no experimento, já que o experimento N foi realizado apenas uma vez e não foi possível mensurar a magnitude do erro associado a ele. No entanto, levando em consideração o erro de 13% presente nos últimos pontos das réplicas e supondo um erro de 13% também no experimento N, poder-se-ia dizer que estatisticamente as duas curvas de crescimento são similares.



**Figura 4.15** Cinética de crescimento celular das réplicas da condição otimizada, experimentos A, B, C, D e E, e do experimento N (controle negativo). A seta mostra o ponto em que foi adicionado o indutor IPTG.

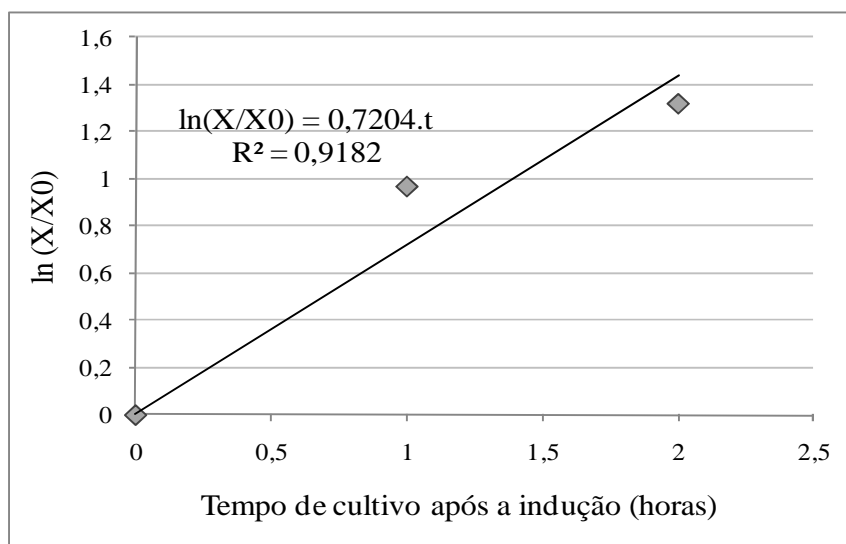
A velocidade específica de crescimento ( $\mu$ ) nestes experimentos foi calculada a partir da equação de conservação de biomassa ( $X$ ) para processos em batelada (1).

$$\frac{dX}{dt} = X \cdot \mu \quad (1)$$

Integrando a equação (1) de  $X_0$  a  $X$  e de  $0$  a  $t$ , sendo  $X_0$  a massa seca de células no tempo de indução da expressão ( $t=0$ ), e logo após, aplicando logaritmo natural nos dois lados da igualdade da equação resultante, obteve-se a equação (2).

$$\ln(X / X_0) = \mu \cdot t \quad (2)$$

Utilizando a equação (2) e realizando ajuste linear dos pontos durante a fase exponencial de crescimento a partir da indução da expressão (Figura 4.16), obteve-se a velocidade específica de crescimento.



**Figura 4.16** Ajuste linear dos pontos durante a fase exponencial de crescimento celular a partir da indução da expressão. Cálculo da velocidade específica de crescimento de *E. coli* durante a expressão da proteína ClpP.

A velocidade específica de crescimento obtida neste trabalho foi de  $0,72 \text{ h}^{-1}$  e o tempo de geração de  $0,96$  horas. ALMEIDA (2001) obteve em seu trabalho velocidade específica de crescimento de *E. coli*, durante a expressão da proteína heteróloga, de aproximadamente  $0,5 \text{ h}^{-1}$ . Estes valores são inferiores às taxas de crescimento descritas para cepas selvagens de *E. coli*. Como discutido na literatura (TOKSOY *et al.*, 2002), a

diminuição da taxa de crescimento está relacionada à sobrecarga metabólica imposta à célula hospedeira pela expressão da proteína heteróloga.

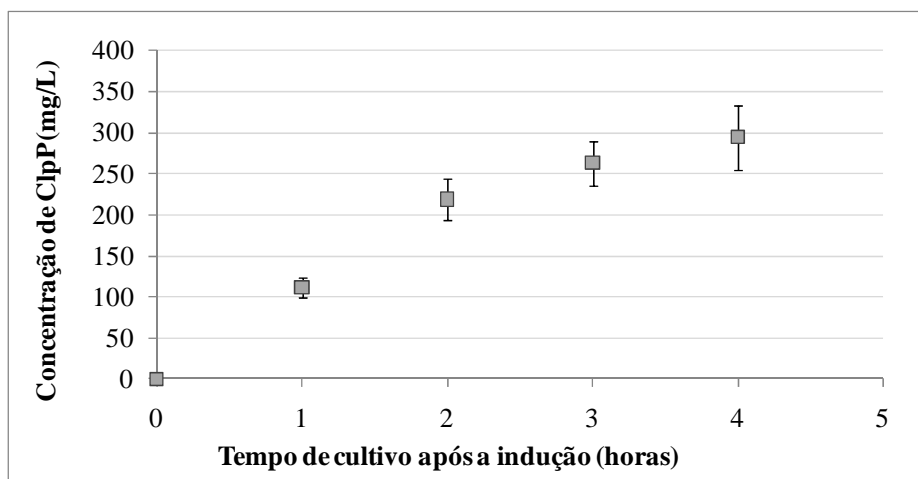
Também foram analisadas as quantidades da proteína ClpP produzida ao longo do cultivo, nas cinco réplicas do ponto otimizado pelo planejamento. As concentrações de proteína de cada uma das réplicas são mostradas na tabela 4.10.

**Tabela 4.10** Concentração de ClpP recombinante produzida ao longo das 4 horas de expressão nos experimentos da validação.

Tempo de indução (horas)	Exp. A Conc. ClpP (mg/L)	Exp. B Conc. ClpP (mg/L)	Exp. C Conc. ClpP (mg/L)	Exp. D Conc. ClpP (mg/L)	Exp. E Conc. ClpP (mg/L)
0*	0	0	0	0	0
1	105,4	134,3	102,7	107,7	107,4
2	190,3	256,6	201,2	225,2	221,6
3	235,4	273,6	238,3	266,8	300,9
4	238,3	306,6	321,1	270,6	335,7

\* Instante em que ocorre a adição do indutor

A média de proteína expressa nos experimentos, de aproximadamente 294mg/L, apresentou concentração um pouco mais elevada que a obtida no experimento 1 do planejamento fatorial. No entanto, levando em consideração os erros associados às medidas de densitometria que variaram entre 10 e 13 % nestes experimentos e o erro estimado em 8% do experimento 1 do planejamento, pode-se afirmar que os valores obtidos na validação são similares aos do planejamento. Na figura 4.17 é mostrado o gráfico da média da concentração da proteína nos experimentos da validação em função do tempo de expressão.



**Figura 4.17** Cinética da produção de proteína. Os quadrados representam a concentração média de ClpP ao longo do tempo utilizando a condição otimizada pelo planejamento de experimentos (0,1mM IPTG e 0 $\mu$ g/mL). As barras representam os desvios padrões nos pontos do experimento na condição otimizada.

Pode-se perceber que após a segunda hora de indução a taxa de produção de proteína começa a diminuir e se aproxima de uma fase estacionária na quarta hora de indução. Pode-se concluir que tempos de expressão mais longos não seriam vantajosos já que a concentração de proteína se manteria constante, reduzindo a produtividade do processo.

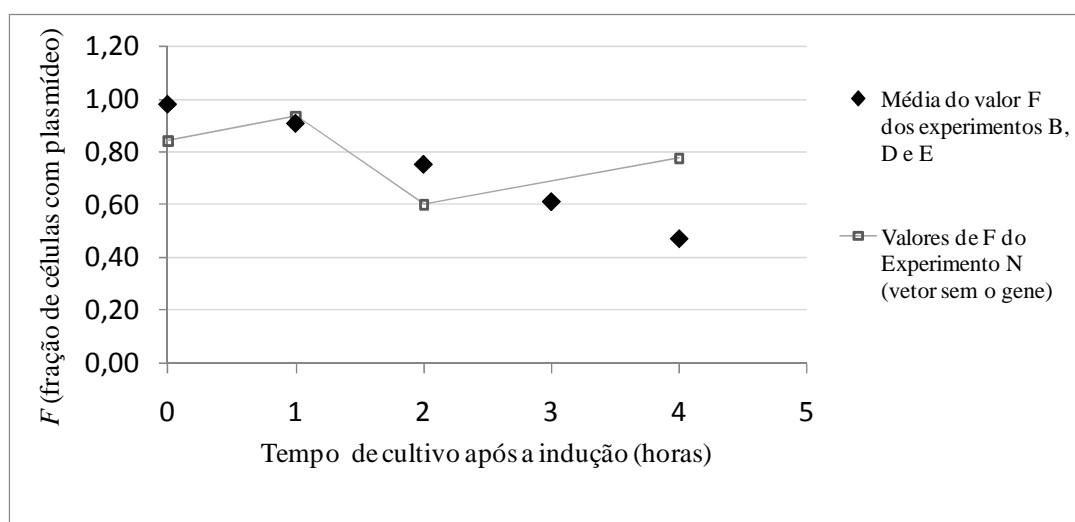
Através da razão entre a concentração de proteína e a massa seca de células, pode-se calcular o fator de rendimento  $Y_{P/X}$  (produção de produto por célula) ao longo do tempo de indução. Estes dados são mostrados na tabela 4.11. É possível perceber que, após a segunda hora de indução, o rendimento não aumenta mais na mesma proporção, indicando mais uma vez que tempos mais longos de expressão não seriam tão vantajosos.

**Tabela 4.11** Fatores de rendimento ( $Y_{P/X}$ ) ao longo do tempo de indução

Tempo de indução (horas)	Média da Concentração de ClpP (mg/L)	Média da Massa seca de células (mg/mL)	( $Y_{P/X}$ ) Fator de rendimento (mg ClpP/mg célula)
0*	0	0,21	-
1	111,5	0,54	0,21
2	219,0	0,77	0,29
3	263,0	0,86	0,30
4	294,5	0,91	0,32

\* Instante em que ocorre a adição do indutor

Outra análise realizada foi o estudo de segregação plasmidial dos cultivos ao longo do tempo a partir da indução. A figura 4.18 mostra o gráfico da variável  $F$  (fração de células com plasmídeo) em função do tempo de cultivo. Em 4 horas de expressão a segregação média chega a níveis de 50 %, o que está de acordo com a segregação obtida no experimento 1 do planejamento fatorial. No gráfico também é mostrada a segregação ocorrida no experimento N, utilizando a cepa transformada apenas com o vetor. Este experimento foi realizado com o objetivo de verificar se o gene poderia interferir na segregação. Os valores de  $F$  obtidos no experimento N foram um pouco superiores aos valores obtidos nas réplicas da validação, no entanto, devido aos erros em alguns pontos, é difícil afirmar que o gene pode influenciar na segregação, já que o experimento foi realizado sem réplicas. Para poder avaliar a influência do gene na segregação, novas análises precisam ser realizadas, no mínimo em triplicata, para conseguir mensurar a magnitude dos erros presentes.



**Figura 4.18** Gráfico da variável  $F$  (fração de células com plasmídeo) em função do tempo.

Os valores de estabilidade plasmidial obtidos em todos os experimentos, juntamente com suas médias e desvios padrão, podem ser visualizados na tabela 4.12.

**Tabela 4.12** Valores da variável  $F$  (fração de células com plasmídeo) ao longo do tempo nas réplicas dos experimentos da validação

Tempo (h)	$F$ (Exp. B)	$F$ (Exp. D)	$F$ (Exp. E)	Média	Desvio Padrão
0*	1,24	0,88	0,82	0,98	0,23
1	1,10	0,79	0,84	0,91	0,17
2	0,40	1,20	0,66	0,75	0,41
3	0,31	0,63	0,89	0,61	0,29
4	0,50	0,47	0,44	0,47	0,03

\* Instante em que ocorre a adição do indutor

Os valores de  $F$  obtidos são similares aos obtidos por SUNITHA *et al.* (2000) em seu trabalho expressando uma fitase e também utilizando um vetor da série pET. Neste trabalho os autores relatam que depois da indução os níveis de estabilidade caíram para 10% em 4 horas. Os autores ainda relatam que, antes da indução, o plasmídeo foi estável por mais de 96h; no entanto, após a indução, o plasmídeo começou a apresentar segregação. TOMAZETTO *et al.* (2007) também mostraram que o vetor pET101 apresenta maior estabilidade em cultivos não induzidos. A maior instabilidade depois da indução pode ser atribuída ao fato de que ao induzir a expressão da proteína recombinante o metabolismo celular fica sobrecarregado, aumentando os níveis de segregação. Neste mesmo trabalho (TOMAZETTO *et al.* 2007), foi demonstrado que induzindo um sistema com lactose, a estabilidade plasmidial também é baixa, chegando a cerca de 30% quando o pH não é controlado e cerca de 60% mantendo o pH em 7,0, ao final de 4 horas. Estes números também indicam que o pH pode estar influenciando na baixa estabilidade obtida neste trabalho, já que o pH dos cultivos não foi mantido constante. Nos experimentos de validação da condição otimizada pelo planejamento fatorial, os cultivos apresentaram pH inicial igual a 7,0 e, após as 4 horas de expressão, apresentaram pH em torno de 5,1. Como já discutido anteriormente, alguns autores ainda apontam o IPTG como um fator que aumenta a segregação plasmidial (FRIEHS, 2004). No entanto, outros fatores podem ainda ser associados à baixa estabilidade plasmidial encontrada nestes cultivos. Ainda segundo FRIEHS (2004), a diminuição do oxigênio dissolvido pode diminuir a estabilidade plasmidial. Como os experimentos foram conduzidos em frascos agitados e neles não se tem o controle do oxigênio dissolvido no meio, esta pode ser uma das causas para os

altos níveis de segregação encontrados ao longo do cultivo. Para controlar os níveis de aeração, pH, assim como monitorar outras variáveis de processo, bioreatores devem ser empregados.

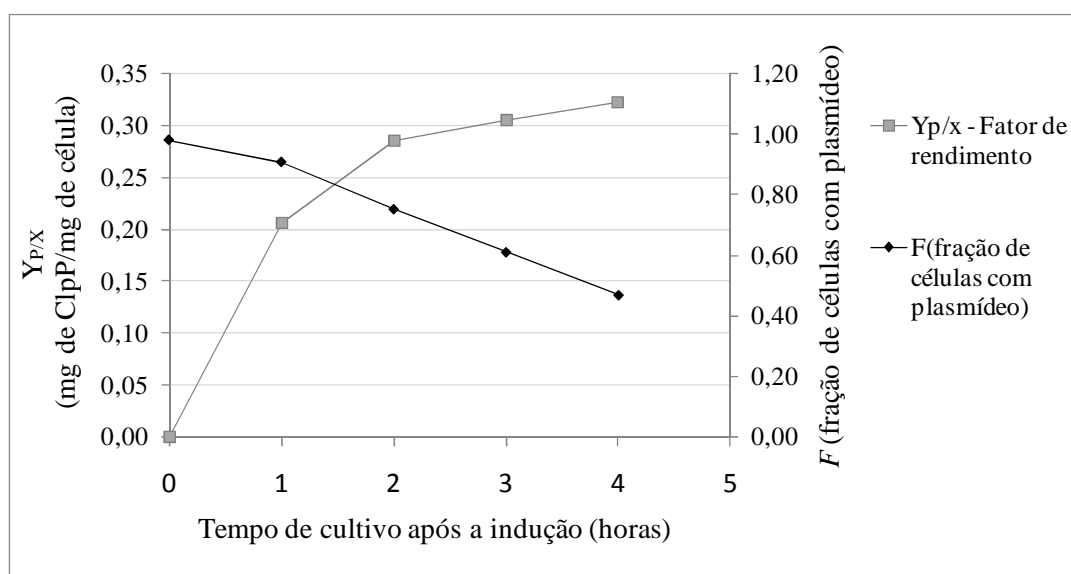
Outros trabalhos ainda mostram que a composição da construção do vetor e os promotores utilizados podem influenciar muito a segregação. TOKSOY *et al.* (2002) obtiveram níveis de estabilidade plasmidial de 80, 72 e 0,1% ao final de 4 horas de indução, variando as construções (plasmídeo/gene) e plasmídeos utilizados, sendo que o plasmídeo, derivado do vetor pET28a sob controle do promotor T7, foi o que apresentou maior estabilidade.

Além dos fatores apresentados, muitos outros são envolvidos nos níveis de estabilidade plasmidial. Segundo GUPTA *et al.* (1995), meios de cultura mais complexos levam à diminuição da estabilidade plasmidial. Outros fatores que podem afetar a estabilidade são: taxa de crescimento, número de cópias do plasmídeo, tamanho do inserto e nível de expressão da proteína recombinante (GUPTA *et al.*, 1995). O vetor utilizado neste trabalho, pET28b, segundo SILVA (2005) apresenta de 15 a 20 cópias por célula, sendo um plasmídeo de médio a alto número de cópias (FRIEHS, 2004; SCHUMANN E FERREIRA, 2004), o que seria positivo para manter pelo menos 1 plasmídeo em cada célula após a divisão celular (FRIEHS, 2004).

Na figura 4.19 é mostrado o gráfico do fator de rendimento ( $Y_{P/X}$ ) e da fração de células com plasmídeo ( $F$ ) em função do tempo. Como esperado, à medida que a segregação aumenta, a taxa de formação de produto por massa de célula diminui e o fator de rendimento ( $Y_{P/X}$ ) se aproxima de um estado estacionário. O fator de rendimento, mesmo a uma taxa mais lenta, ainda mostra um aumento nas últimas 2 horas de expressão. Isto pode ocorrer devido a um aumento da produção da proteína pelas células que ainda possuem plasmídeo. Provavelmente, o fator de rendimento começaria a diminuir nos próximos pontos, depois das 4 horas de indução. Em seus estudos, expressando fitase em *E. coli*, SUNITHA *et al.* (2000) mostraram que nas duas primeiras horas da indução a produção de fitase aumenta de 0 até 800 U/L enquanto a estabilidade plasmidial diminui para 60% neste período, ou seja, mesmo com segregação de 40% a produção de fitase continua aumentando. Em duas horas e meia de



indução, a produção de fitase aumenta ainda mais, para 1000 U/L, enquanto a segregação aumenta para 80%, e só após este ponto é que a atividade de fitase começa a diminuir. Como mostrado na figura 4.19, em quatro horas de indução, a segregação ainda estava em torno de 55%, e o fator de rendimento ainda estava aumentando. No entanto, assim como mostrado por SUNITHA *et al.* (2000), quando a segregação atingisse níveis ainda maiores, o fator de rendimento começaria a diminuir, o que provavelmente seria visualizado nas próximas horas de indução, se o cultivo tivesse sido mantido por mais de 4 horas de indução.



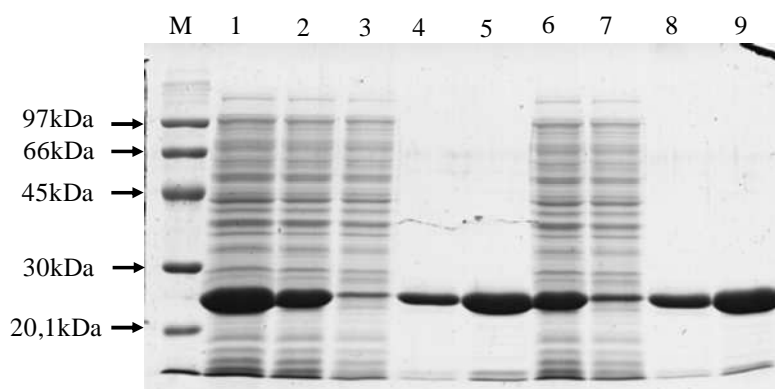
**Figura 4.19** Gráfico do Fator de rendimento de massa de ClpP produzida por massa de célula ( $Y_{p/x}$ ) e da Fração de células com plasmídeo ( $F$ ) em função do tempo.

#### 4.3.6 Purificação

Ensaio preliminares foram realizados para purificação da proteína expressa fusionada à cauda de histidina, utilizando cromatografia de afinidade, empregando coluna empacotada com resina contendo íons de níquel imobilizados. A proteína, quando fusionada a uma cauda de histidina, tem sua purificação facilitada, pois interage de forma pseudo-específica com ligantes metálicos, como o níquel (LARENTIS *et al.*, 2006). Alguns trabalhos já expressaram a proteína ClpP fusionada à cauda de histidina e esta proteína fusionada foi testada contra células rompidas de *S. pneumoniae* em *Western blot*, mostrando especificidade para os antígenos protéicos (KWON *et al.*,

2004; CAO *et al.*, 2007). Estes dados indicam que a princípio a cauda de histidina não influencia na estrutura da proteína, mantendo o reconhecimento da mesma pelos antígenos protéicos.

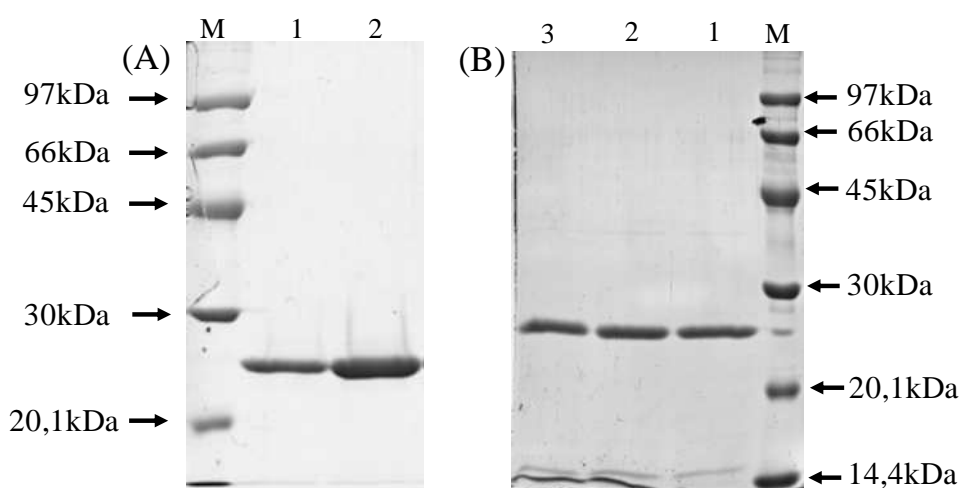
A solução tampão utilizada na purificação, tanto para o rompimento celular quanto para o equilíbrio e eluição da coluna, foi a solução 20 mM Tris/1 mM EDTA. Primeiramente, foram utilizadas para a purificação soluções tampão com 20 mM e 5 mM de imidazol, para equilibrar a coluna e fazer o rompimento celular das amostras a serem purificadas. No entanto, ao passar as amostras pela coluna, uma grande quantidade da proteína de interesse não foi adsorvida na coluna (fração não adsorvida). Isto pode ter ocorrido devido à proteína estar enovelada de tal forma que a cauda de histidina não esteja exposta na superfície, impedindo a sua ligação com o níquel. Por isso, foram feitos novos testes, utilizando uma amostra lisada com solução tampão sem imidazol, mas com 1 mM de DTT, e outra lisada com solução tampão com 1 mM de imidazol e 1 mM de DTT. O imidazol compete com as proteínas pela ligação com o metal (neste caso o níquel), logo, o próprio imidazol pode ter prejudicado a ligação da ClpP à coluna. A coluna com as proteínas adsorvidas foi eluída com solução tampão em gradiente de 0-500 mM de imidazol. As análises desta purificação foram feitas em SDS-PAGE (Figura 4.20).



**Figura 4.20** Gel de SDS-PAGE 12,5% da purificação da proteína ClpP. **M**, padrão de massa molar LMW (Amersham Bioscience); **1**, extrato celular total lisado; **2**, fração não adsorvida (amostra sem imidazol); **3**; eluído com a primeira concentração de imidazol (amostra sem imidazol), **4**, eluído com 220 mM de imidazol (amostra sem imidazol); **5**, eluído com 300 mM de imidazol (amostra sem imidazol); **6**, fração não adsorvida (amostra com 1mM de imidazol); **7**, eluído com a primeira concentração de imidazol (amostra com 1 mM de imidazol); **8**, eluído com 220 mM de imidazol (amostra com 1mM de imidazol); **9**, eluído com 300 mM de imidazol (amostra com 1 mM de imidazol).

A proteína ClpP foi eluída com cerca de 220 a 300mM de imidazol. Como pode ser visualizado na figura 4.20, foram obtidas frações da proteína ClpP com mais de 98% de pureza, já que praticamente não existem bandas de outras proteínas, segundo as análises de densitometria do gel. Futuros estudos podem ser realizados para a determinação do rendimento da purificação e para melhorar a fração purificada.

A proteína purificada foi submetida à diálise contra tampão PBS para retirada de imidazol, sendo verificada precipitação da ClpP no processo de diálise. A proteína purificada se mostrou estável por aproximadamente 7 meses, armazenada a -20°C. A figura 4.21 apresenta o gel de poliacrilamida a 12,5% com a amostra da proteína purificada e desta mesma amostra armazenada por sete meses a -20°C. Nesta figura é possível visualizar a banda íntegra da proteína ClpP.



**Figura 4.21** Gel de SDS-PAGE 12,5% do teste de estabilidade da proteína. (A) gel das amostras logo após a purificação; **M**, padrão de massa molar LMW (Amersham Bioscience); **1** e **2**, amostras de ClpP purificada. (B) gel das amostras armazenadas a -20°C durante 7 meses após a purificação; **M**, padrão de massa molar LMW (Amersham Bioscience); **1**, **2**, e **3**, amostras de ClpP purificadas e armazenadas a -20°C por 7 meses.

## 4.4 Clonagem e expressão utilizando *Bacillus subtilis* como sistema de expressão

### 4.4.1 Caracterização da cepa *Bacillus subtilis* WB600

A cepa de *Bacillus subtilis* WB600 (WU *et al.*, 1991) empregada neste trabalho, deficiente de 6 proteases extracelulares, foi caracterizada, quanto ao gênero e à espécie, em parceria com o Laboratório de Fisiologia Bacteriana (LFB) e Coleção de Culturas do Gênero *Bacillus* e Gêneros Correlatos (CCGB) do Instituto Oswaldo Cruz – FIOCRUZ.

As colônias da cepa *Bacillus subtilis* WB600 em meio LB ágar apresentaram-se irregulares e opacas, o que está de acordo com a morfologia das colônias de *B. subtilis* (SNEATH e BERGEY, 1986). As células se mostraram gram-positivas e nas lâminas a fresco também apresentaram a morfologia celular esperada para *B. subtilis*. Foram também realizados testes bioquímicos com a cepa *Bacillus subtilis* WB600 e com a cepa *Bacillus subtilis* LFB732 (utilizada para comparação). Os resultados dos testes bioquímicos, exceto os realizados através do sistema de identificação Api 50 CHB (Biomerieux), são mostrados na tabela 4.13. As diferenças nos resultados de 3 testes bioquímicos da cepa WB 600, quando comparados aos resultados da cepa LFB732, provavelmente se devem à ausência na cepa WB600 de seis das principais proteases extracelulares de *B. subtilis*. No entanto, os testes bioquímicos, de acordo com SNEATH e BERGEY (1986), e também o teste feito com o sistema de identificação Api 50 CHB, caracterizaram que a cepa WB600 é de *Bacillus subtilis*.

**Tabela 4.13** Resultado dos testes bioquímicos realizados com as cepas *Bacillus subtilis* WB600 e *Bacillus subtilis* LFB732 (utilizada para comparação).

Testes Bioquímicos	<i>B. subtilis</i> WB600	<i>B. subtilis</i> LFB732
Amido	+	+
Caseína	-	+
Tirosina	-	-
Gelatina	-	+
Lectinase	-	-
β-Hemólise em ágar	-	-
Uréia	-	-
Citrato	-	+
VP	+	+
Glicose	+	+
Arabinose	-	-
Xylose	-	-
Nitrato	+	+
NaCl 5%	+	+
NaCl 7%	+	+
NaCl 10%	-	-

#### 4.4.2 Clonagem

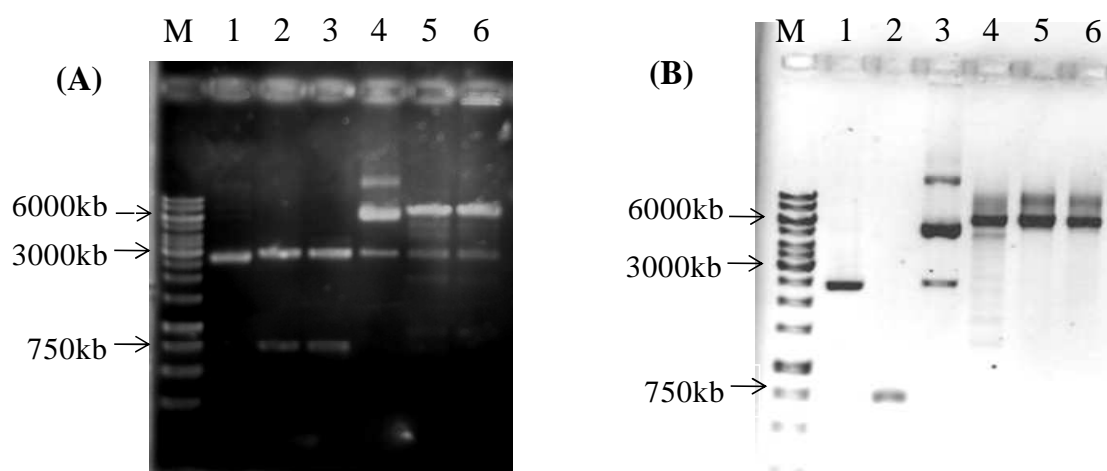
A fim de obter um sistema que fosse capaz de expressar a proteína ClpP extracelularmente, *Bacillus subtilis* foi escolhido como hospedeiro para clonar e expressar a proteína. Este microrganismo, que também tem sido utilizado como sistema de expressão de proteínas de uso farmacológico e imunológico, é atrativo por algumas razões, como secretar as proteínas diretamente no meio de cultura, ser considerado GRAS (*generally regarded as safe*), produzir esporos (facilitando a preservação das cepas) e ter habilidade de crescer e expressar as proteínas em meios baratos (FERREIRA *et al.*, 2005). Para tanto o gene *clpP* foi sintetizado quimicamente fusionado a um peptídeo sinal, para a expressão extracelular da proteína. O gene sintético utilizou como base a mesma seqüência obtida no seqüenciamento do vetor pET28b/*clpP*-C2P, clone positivo de *E. coli* obtido neste trabalho, incluindo a cauda de histidina. Ao gene sintético ainda foram adicionados uma seqüência sinalizadora e um sítio de clivagem entre esta seqüência e o gene. Além disso, o gene sintético utilizado ainda teve seus *códons* otimizados. A otimização foi realizada levando em

consideração: a frequência de códons utilizados em *B. subtilis* (alterando os códons com frequência menor que 15%) e a formação de fortes estruturas secundárias, alterando os códons para evitar este problema. Na figura 4.22 é mostrado um alinhamento múltiplo do gene *clpP* sintético antes e depois da otimização de códons. As frequências em que os códons ocorrem em *B. subtilis*, tanto para o gene sem otimização quanto para o gene otimizado, são mostradas no apêndice C. O gene sintético com códons otimizados foi inserido no vetor pBSK com marcador de resistência à ampicilina.



**Figura 4.22** Alinhamento múltiplo do gene *clpP* sintético utilizado para a clonagem em *B. subtilis*, sem a otimização de códons (submetido) e otimizado. Em verde são mostrados os nucleotídeos que sofreram alterações no gene otimizado. A seqüência do peptídeo sinal está realçada em amarelo. A seqüência do sítio de clivagem está sublinhada.

Para a construção do vetor foram realizadas reações de digestão com o vetor pBSKclpP717, para a liberação do inserto, e com o vetor pHCMC03, para a linearização do vetor. As reações de digestão foram visualizadas em gel de agarose (Figura 4.23). Foram obtidas bandas nos tamanhos esperados tanto do inserto, 717pb, quanto do vetor pHCMC03 linearizado. As bandas correspondentes ao inserto (gene *clpP*) e ao vetor linearizado foram purificadas e novamente visualizadas em gel de agarose (Figura 4.23). Foram obtidos fragmentos nos tamanhos esperados.



**Figura 4.23** (A) gel de agarose 1% das digestões dos vetores pBSKclpP717 e pHCMC03. M, marcador padrão de tamanho de fragmento 1kb DNA ladder; 1, vetor pBSKclpP717; 2 e 3, vetor pBSKclpP717 digerido com as enzimas *Bam*HI e *Xba*I; 4, vetor pHCMC03; 5 e 6, vetor pHCMC03 digerido com as enzimas *Bam*HI e *Xba*I. (B) gel de agarose 1% das purificações das digestões. M, marcador padrão de tamanho de fragmento 1kb DNA ladder; 1, vetor pBSKclpP717; 2, inserto (*clpP*) digerido e purificado; 3, vetor pHCMC03; 4, 5 e 6 vetor pHCMC03 linearizado e purificado.

A reação de ligação foi realizada, utilizando o gene e o vetor digeridos e purificados, através da enzima T4 DNA ligase. O vetor construído nesta ligação foi transformado por choque térmico em *E. coli* JM109 competente. Não foram obtidas colônias transformantes, nem mesmo no controle negativo (reação de ligação apenas com o vetor). As reações de ligação foram realizadas em algumas condições diferentes e mesmo assim não se obteve êxito nas transformações. Algumas possibilidades são levantadas, como o fato de o vetor utilizado não ser um vetor comercial e poder não ter todos os mecanismos necessários para se manter linearizado. Além disso, após a

digestão, o vetor pode ter formado alças e mantido a conformação enovelada, dificultando a ligação com o inserto. A instabilidade estrutural dos plasmídeos tem sido citada em trabalhos como o motivo pela falta de sucesso na utilização de *B. subtilis* como sistema de expressão (LI *et al.*, 2004; NGUYEN *et al.*, 2005).



## 5 CONCLUSÕES E PERSPECTIVAS

O gene *clpP* foi clonado em *E. coli*, utilizando o vetor pET28b (Novagen), sendo que a proteína foi expressa de forma eficiente, em altos níveis e de forma solúvel. Além de *E. coli*, objetivou-se o emprego de *B. subtilis* como sistema de expressão, utilizando para esta clonagem o vetor pHCMC03 (NGUYEN *et al.*, 2005), no entanto, não foram obtidas células recombinantes para este sistema de expressão. Uma das possibilidades levantadas para a falta de êxito nesta clonagem, é o fato de o vetor utilizado (pHCMC03) não ser comercial e poder não ter todos os mecanismos necessários para se manter linearizado ou ter sofrido algum tipo de mutação ou deleção em sua seqüência, causada por uma instabilidade estrutural. Como mencionado no capítulo de discussão, a instabilidade estrutural dos plasmídeos tem sido um dos obstáculos na utilização de *B. subtilis* como sistema de expressão (LI *et al.*, 2004; NGUYEN *et al.*, 2005).

Quando se trata de *E. coli* como sistema de expressão se pode dizer que existem várias ferramentas genéticas, bem desenvolvidas, sendo comercializadas. São encontrados vetores e cepas com as mais diversas especificações, para serem utilizados de acordo com a proteína a ser produzida e o processo utilizado. Mesmo este sistema ainda apresentando algumas limitações, como a produção da proteína heteróloga muitas vezes em forma de agregados insolúveis, ainda é o mais atraente do ponto de vista da eficiência e facilidade da obtenção da proteína recombinante, e também dos conhecimentos adquiridos e técnicas disponíveis para o uso deste sistema. Já a utilização de *B. subtilis* como sistema de expressão pode ser dificultada, pois apesar de ser citado na literatura como um sistema atrativo por secretar as proteínas diretamente no meio de cultura, ser considerado GRAS (*generally regarded as safe*), produzir esporos facilitando a preservação das cepas e ter habilidade de crescer e expressar as proteínas em meios baratos (FERREIRA *et al.*, 2005), este sistema ainda apresenta algumas falhas e precisa ser melhor estudado e desenvolvido. Apesar de já existirem cepas para serem utilizadas na expressão de proteínas recombinantes, estas ainda precisam ser aprimoradas e estarem disponíveis comercialmente, além disso, os vetores precisam ser desenvolvidos para serem mais estáveis e de fácil utilização.

Através de planejamento de experimentos, a expressão da proteína ClpP em *E. coli* foi otimizada, variando a concentração de indutor (IPTG) e canamicina. Nestes experimentos foram estudados o crescimento celular, a produção de proteína e a estabilidade plasmidial. A proteína foi expressa, de forma solúvel, em níveis similares em todos os experimentos do planejamento fatorial, sendo que a média de proteína expressa no ponto central foi de  $240,4 \pm 19,7$  mg/L de ClpP com desvio padrão relativo de 8%. As concentrações de IPTG e canamicina não apresentaram efeito significativo sobre a concentração de proteína expressa. Apesar da concentração de proteína variar muito dependendo da proteína que está sendo expressa, a concentração de proteína expressa neste trabalho está dentro da faixa de concentração obtida por outros trabalhos que otimizaram a expressão de outras proteínas em cultivos batelada em frascos agitados (CHOI *et al.*, 2006; MALDONADO *et al.*, 2007; CHUAN *et al.*, 2008; HAN *et al.*, 2008). Para aumentar os níveis de expressão em termos de g/L o cultivo precisa ser feito em bioreatores, pois as variáveis do processo precisam ser bem controladas e o regime de operação do cultivo precisa ser estudado, o que seria alvo para novos estudos. Também no planejamento de experimentos, nos experimentos do ponto central, obteve-se média de  $0,75 \pm 0,02$  mg/mL de massa seca de células com desvio padrão relativo de 2,5%, já a fração de células com plasmídeo (*F*) apresentou média de  $0,08 \pm 0,03$  com desvio padrão relativo de 36% ao final das 4 horas de expressão. O IPTG mostrou efeito negativo sobre o crescimento celular e sobre a estabilidade plasmidial, enquanto a canamicina não apresentou efeito significativo sobre nenhuma destas duas variáveis. O ponto otimizado do sistema foi obtido nas condições de 0,1mM de IPTG e 0µg/mL de canamicina, apresentando concentração da proteína ClpP de 247,3 mg/L, crescimento celular de 0,91 mg/mL de massa seca de células e fração de células com plasmídeo (*F*) de 0,44. Com isso foi possível concluir, dentro das condições e faixas estudadas, que é possível diminuir em 10 vezes a concentração de indutor e retirar o antibiótico do sistema, mantendo a expressão da proteína em níveis similares e reduzindo custos de processo.

Durante os experimentos do planejamento fatorial, a estabilidade plasmidial foi estudada. A estabilidade plasmidial é um dos pontos mais importantes do processo recombinante, já que o plasmídeo é o ponto chave para a produção da proteína heteróloga. Os estudos encontrados na literatura nesta área de estabilidade plasmidial,

muitas vezes são contraditórios, possivelmente devido aos altos erros experimentais associados às técnicas experimentais empregadas. Além disso, a dificuldade de encontrar artigos sobre o assunto, principalmente atuais, e que expliquem o problema da segregação com maior detalhamento e profundidade, é grande, indicando a necessidade de maiores estudos nesta área. Através do estudo de segregação plasmidial realizado neste trabalho, foi possível inferir que ocorre perda do plasmídeo mesmo nas concentrações padrão empregadas de antibiótico. A perda de plasmídeo é reduzida em concentrações mais baixas de IPTG, sendo que no ponto otimizado pelo planejamento de experimento, quando os estudos foram realizados ao longo do tempo, a estabilidade plasmidial atinge média de 47% em 4 horas de indução. Novos estudos precisam ser realizados para a determinação da concentração ótima de IPTG, a concentração mínima possível, onde se consiga o máximo de estabilidade plasmidial e de proteína expressa. Outra possibilidade é o estudo da estabilidade plasmidial utilizando temperaturas mais baixas para a expressão da proteína. Além disso, para melhorar os níveis de estabilidade plasmidial e possivelmente aumentar a concentração de proteína produzida, todas as variáveis descritas como as que influenciam na perda do plasmídeo, precisam ser estudadas ao mesmo tempo. Para tanto ferramentas de planejamento experimental, como os planejamentos fracionados e os delineamentos de Plackett e Burman, podem ser fundamentais permitindo avaliar todas as variáveis ao mesmo tempo, selecionando as realmente significativas (RODRIGUES E IEMMA, 2005). Outra ferramenta que se torna importante para a maximização da produção de proteína e estabilidade plasmidial, assim como para determinar as condições ideais do processo em bioreator, é a modelagem matemática do processo e simulação.

A proteína expressa, fusionada a uma cauda de histidina, foi purificada por cromatografia de afinidade de forma eficiente, utilizando resina com íons de níquel imobilizados. A princípio a cauda de histidina não modifica a conformação da proteína, conforme dados obtidos na literatura (KWON *et al.*, 2004; CAO *et al.*, 2007). A proteína foi purificada com mais de 98% da proteína na fração purificada em gradiente de concentração de imidazol. Novos estudos poderiam ser realizados empregando eluição com pulsos de imidazol e também para determinar o rendimento da purificação.

# APÊNDICE A

## Apêndice A.1 – Sequência do plasmídeo pBSKClpP717

LOCUS GS44387 pBSK Clp717 3580 bp DNA CIRCULAR SYN 31-MAY-2009  
 DEFINITION Ligation of Clp717 into pGS1/SmaI  
 ACCESSION GS44387 pBSK Clp717  
 KEYWORDS .  
 SOURCE Unknown.  
 ORGANISM Unknown  
 Unclassified.  
 REFERENCE 1 (bases 1 to 3580)  
 AUTHORS Self  
 JOURNAL Unpublished.  
 COMMENT SECID/File created by SciEd Central, Scientific & Educational Software  
 FEATURES Location/Qualifiers  
 CDS 21..327  
 /gene="fl origin"  
 misc\_feature 600..643  
 /gene="T7/M13F"  
 CDS complement (460..615)  
 /gene="LacZ"  
 CDS 653..1369  
 /gene="Clp717"  
 misc\_feature complement (1391..1447)  
 /gene="T3/M13R"  
 /product="T3 primer"  
 misc\_feature 1777..2444  
 /gene="pUC origin"  
 CDS complement (2595..3452)  
 /gene="Ampicillin"  
 BASE COUNT 899 a 872 c 882 g 927 t  
 ORIGIN  
 1 CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA  
 61 CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTCTTCCCT TCCTTTCCTG  
 121 CCACGTTCGC CGGCTTTCCT CGTCAAGCTC TAAATCGGGG GCTCCCTTTA GGGTTCGGAT  
 181 TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG  
 241 GGCCATCGCC CTGATAGACG GTTTTTTCGCC CTTTGACGTT GGAGTCCACG TTTCTTTAATA  
 301 GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACCCAT CTCGGTCTAT TCTTTTGATT  
 361 TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAA TGAGCTGATT TAACAAAAAT  
 421 TTAACGCGAA TTTTAACAAA ATATTAACGC TTACAATTG CCATTCGCCA TTCAGGCTGC  
 481 GCAACTGTTG GGAAGGGCGA TCGGTGCGGG CCTCTTCGCT ATTACGCCAG CTGGCGAAAG  
 541 GGGGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG GTTTTCCAG TCACGACGTT  
 601 GTAAAACGAC GGCCAGTGAA TTGTAATACG ACTCACTATA GGGCGACCCC ATGGATCCAT  
 661 GAAGAATATG TCTTGCAAAC TGGTCGTTTC TGTACACTG TTCTTCAGCT TTCTGACTAT  
 721 CGGACCCTTA GCGCATGCAG CTGAGTTTAT GGTGCCTGTC GTCATTGAAC AAACATCTCG  
 781 TCGTGAACGC TCATACGATA TTTATTCTCG TCTGTTGAAA GATCGGATTA TTATGTTAAC  
 841 GGGGCCGGTG GAAGATAACA TGGCCAACAG CGTTATTGCG CAGTTGCTGT TCCTGGACGC  
 901 GCAAGATTCT ACGAAAGATA TCTATTTATA TGTGAACACC CCAGGCGGCT CAGTTAGCGC  
 961 CGGGTTGGCT ATCGTGGACA CCATGAATTT TATCAAGGCC GACGTACAAA CGATTGTAAT  
 1021 GGGCATGGCT GCCAGCATGG GAACAGTAAT TGCTAGCTCA GGAGCGAAAG GCAAAAGATT  
 1081 TATGCTTCCCT AACGCGGAAT ACATGATCCA TCAGCCGATG GGCGGCACGG GGGGAGGAAC  
 1141 ACAGCAAACA GACATGGCCA TTGCGGCGGA GCATCTTCTG AAGACAAGAA ATACTTTAGA  
 1201 AAAGATTCTG GCAGAGAACT CAGGTCAAAG CATGGAGAAA GTTCATGCAG ATGCTGAGCG  
 1261 TGATAATTGG ATGTCAGCAC AGGAGACACT TGAATATGGC TTTATCGATG AAATCATGGC  
 1321 AAATAATTCA TTGAATTTGG AACATCATCA TCATCATCAT TAATCTAGAA TCGGGGATAT  
 1381 CCTCGAGGTT CCCTTTAGTG AGGGTTAATT GCGAGCTTGG CGTAATCATG GTCATAGCTG  
 1441 TTTCTGTGTG GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA  
 1501 AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGGCTCA  
 1561 CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC  
 1621 GCGGGGAGAG GCGGTTTGCG TATGCGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG  
 1681 CCGTCGGTGC TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA  
 1741 TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC  
 1801 AGGAACCGTA AAAAGGCCGC GTTGTGGCG TTTTCCATA GGCTCCGCC CCCTGACGAG  
 1861 CATCAGAAA ATCGACGCTC AAGTCAGAGG TGCGGAAACC CGACAGGACT ATAAAGATAC  
 1921 CAGGCGTTTC CCCCTGGAAG CTCCTCTGTG CGCTCTCTG TTCCGACCT GCCGCTTACC  
 1981 GGATACCTGT CCGCTTTCT CCGCTTCTG AGCGTGGCGC TTTCTCATG CTCACCGTGT  
 2041 AGGTATCTCA GTTCGGTGTG GGTGCTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC  
 2101 GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCGGTAAGA  
 2161 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA

2221 GCGCGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGAACAGTA  
2281 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA  
2341 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTGCAAGCA GCAGATTACG  
2401 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG  
2461 TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC  
2521 TAGATCCTTT TAAATTAATA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT  
2581 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT  
2641 CGTTCATCCA TAGTTGCCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA  
2701 CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCAC GCTCACCAGC TCCAGATTTA  
2761 TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCTTGC AACTTTATCC  
2821 GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT  
2881 AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT  
2941 ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG  
3001 TGCAAAAAAG CGGTTAGCTC CTTCCGGTCTT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA  
3061 GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TFACTGTCAT GCCATCCGTA  
3121 AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG  
3181 CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT  
3241 TTA AAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG  
3301 CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT  
3361 ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA  
3421 ATAAGGGCGA CACGAAAATG TTGAATACTC ATACTCTTCC TTTTTCATA TTATTGAAGC  
3481 ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA  
3541 CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC

## APÊNDICE B

### Apêndice B.1 – Alinhamento Múltiplo de IgA1 protease

	10	20	30	40	50	60
SP195x2	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
CDC1087_00	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
G54xxx4	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
CGSP14	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
R6xxxx0	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
70585x1	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
TIGR4x7	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
Hungary19A_6	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
Prim.cons.	MEKYFGEKQERFSFRKLSVGLV	SATIS	SSLFFMSV	LASSSSVDAQ	ETAGVHYKYV	ADSELSS
*****						
	70	80	90	100	110	120
SP195x2	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
CDC1087_00	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNEM	QALVAGASLA	TLGIL
G54xxx4	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
CGSP14	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
R6xxxx0	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
70585x1	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNEM	QALVAGASLA	ALGIL
TIGR4x7	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	AMGIL
Hungary19A_6	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
Prim.cons.	EEKKQLVYDIPTYVENDDETY	YLVYKLN	SQNQLAELP	NTGSKNER	QALVAGASLA	ALGIL
*****						
	130	140	150	160	170	180
SP195x2	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
CDC1087_00	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
G54xxx4	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
CGSP14	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
R6xxxx0	IFAVSKKKVKNKTVLHLV	LVLVAGM	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
70585x1	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYK	LTSGEKLPLPKDI
TIGR4x7	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
Hungary19A_6	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
Prim.cons.	IFAVSKKKVKNKTVLHLV	LVLVAGI	GNGVLVS	VHALENHLL	LNNTDYEL	TSGEKLPLPKEI
*****						
	190	200	210	220	230	240
SP195x2	SGYTYIGYIKEGKTTSE	EVSNQKSSV	ATPTKQ	QKVDY	NVTPNFV	DHPSTVQAIQE
CDC1087_00	SGYTYIGYIKEENITSE	EVSNQKSS	IVTPTK	QKVDY	SVTPNFV	DHPSTVQAIQE
G54xxx4	SGYTYIGYIKEGKTTSD	FEVSNQEK	SAATPT	KQKVDY	NVTPNFV	DHPSTVQAIQE
CGSP14	SGYTYIGYIKEGKTTSD	FEVSNQEK	SAATPT	KQKVDY	NVTPNFV	DHPSTVQAIQE
R6xxxx0	SGYTYIGYIKEGKTTSD	FEVSNQEK	SAATPT	KQKVDY	NVTPNFV	DHPSTVQAIQE
70585x1	SGYTYIGYIKEENITSE	EVSNQKSSV	ATPTK	QKVDY	SVTPNFV	DHPSTVQAIQE
TIGR4x7	SGYTYIGYIKEGKTTSE	EVSNQKSSV	ATPTK	QKVDY	NVTPNFV	DHPSTVQAIQE
Hungary19A_6	SGYTYIGYIKEGKTTSD	FEVSNQEK	SAATPT	KQKVDY	NVTPNFV	DHPSTVQAIQE
Prim.cons.	SGYTYIGYIKEGKTTSS	2EVSNQ22	SAATPTK	QKVDY	NVTPNFV	DHPSTVQAIQE
*****						
	250	260	270	280	290	300
SP195x2	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
CDC1087_00	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
G54xxx4	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
CGSP14	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
R6xxxx0	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
70585x1	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
TIGR4x7	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
Hungary19A_6	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
Prim.cons.	SSTKPTEVQVVEKPFST	ELINPRKEEK	QSSDSQE	QLAEHKNLE	TKKEEKI	SPKEKTGVNT
*****						

	310	320	330	340	350	360																																																						
SP195x2	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
CDC1087_00	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	S	E	G	I	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
G54xxx4	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
CGSP14	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	T	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
R6xxx0	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
70585x1	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
TIGR4x7	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	M	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
Hungary19A_6	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	V	E	G	I	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V
Prim.cons.	L	N	P	Q	D	E	V	L	S	G	Q	L	N	K	P	E	L	L	Y	R	E	E	T	I	E	T	K	I	D	F	Q	E	E	I	Q	E	N	P	D	L	A	E	G	T	V	R	V	K	Q	E	G	K	L	G	K	K	V	E	I	V

	370	380	390	400	410	420																																																						
SP195x2	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	S	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
CDC1087_00	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	I	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	T	I
G54xxx4	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	I	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
CGSP14	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	S	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
R6xxx0	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	S	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
70585x1	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	I	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
TIGR4x7	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	S	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
Hungary19A_6	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	I	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I
Prim.cons.	R	I	F	S	V	N	K	E	E	V	S	R	E	I	V	S	T	S	T	T	A	P	2	P	R	I	V	E	K	G	T	K	K	T	Q	V	I	K	E	Q	P	E	T	G	V	E	H	K	D	V	Q	S	G	A	I	V	E	P	A	I

	430	440	450	460	470	480																																																				
SP195x2	Q	P	E	L	P	E	A	V	S	D	K	G	E	P	A	V	Q	P	E	L	P	E	A	V	T	D	K	G	E	T	E	V	Q	P	E	S	P	D	T	V	V	S	D	K	G	E	P	E	Q	V	A	P	L	P	E	Y	T	R
CDC1087_00	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	P	K	A	V	T	D	K	G	-----																													
G54xxx4	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	P	E	A	V	T	D	K	G	-----																													
CGSP14	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	S	E	A	V	T	D	K	G	-----																													
R6xxx0	Q	P	E	L	P	E	A	V	S	D	K	G	E	P	E	V	Q	P	T	L	P	E	A	V	T	D	K	G	E	-----																												
70585x1	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	P	E	A	V	T	D	K	G	E	P	A	I	Q	P	E	L	P	E	A	V	T	D	K	G	-----													
TIGR4x7	Q	P	E	L	P	E	A	V	S	D	K	G	E	P	E	V	Q	P	T	L	P	E	A	V	T	D	K	G	E	-----																												
Hungary19A_6	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	S	K	A	V	I	T	D	K	G	E	-----																											
Prim.cons.	Q	P	E	L	P	E	A	V	S	D	K	G	V	P	E	V	Q	P	A	L	P	E	A	V	T	D	K	G	E	2	2	2	Q	P	E	2	2	2	V	V	2	D	K	G	E	P	E	Q	V	A	P	L	P	E	Y	T	R	

	490	500	510	520	530	540																																																	
SP195x2	P	Q	A	G	A	V	V	E	P	A	I	Q	P	E	L	P	E	A	V	T	D	K	G	E	P	-----																													
CDC1087_00	-----																																																						
G54xxx4	-----																																																						
CGSP14	-----																																																						
R6xxx0	-----																																																						
70585x1	-----																																																						
TIGR4x7	-----																																																						
Hungary19A_6	-----																																																						
Prim.cons.	P	Q	A	G	A	V	V	E	P	A	V	Q	P	E	L	P	E	A	V	T	D	K	G	E	P	E	Q	V	A	V	Q	P	E	L	P	E	A	V	T	D	K	G	E	2	E	V	Q	P	2	2	V	V	S	D	K

	550	560	570	580	590	600																																																					
SP195x2	G	E	P	E	Q	V	A	P	L	P	E	Y	T	R	P	Q	A	G	A	V	V	E	P	A	I	Q	P	E	L	P	E	A	V	T	D	K	G	-----																					
CDC1087_00	-	E	P	E	Q	V	A	P	L	P	E	Y	T	R	P	Q	A	G	A	V	V	E	P	A	I	Q	P	E	L	P	E	A	V	T	D	K	G	-----																					
G54xxx4	G	E	P	E	Q	V	A	P	L	P	E	Y	T	R	P	Q	A	G	A	V	V	E	P	A	I	Q	P	E	L	P	E	A	V	T	D	K	G	-----																					
CGSP14	-----																																																										
R6xxx0	-	E	T	P	V	E	K	T	K	E	Q	G	P	E	K	T	E	E	V	P	V	K	P	T	E	T	P	V	N	P	N	E	G	T	T	E	G	T	S	I	Q	E	A	E	N	P	V	Q	P	A	E	E	S	T	T	N	S	E	K
70585x1	-	E	T	P	V	E	K	T	K	E	Q	G	P	E	K	T	E	E	V	P	V	K	P	T	E	T	P	V	N	P	N	E	G	T	T	E	G	T	S	I	Q	E	A	E	N	P	V	Q	P	A	E	E	S	T	T	N	S	E	K
TIGR4x7	-	E	T	P	V	E	K	T	K	E	Q	G	P	E	K	T	E	E	V	P	V	K	P	T	E	T	P	V	N	P	N	E	G	T	T	E	G	T	S	I	Q	E	A	E	N	P	V	Q	P	A	E	E	S	T	T	N	S	E	K
Hungary19A_6	-	E	P	K	Q	V	A	P	L	P	E	Y	T	G	P	Q	A	S	A	I	V	E	P	E	Q	V	A	P	L	P	E	Y	T	G	V	Q	A	G	-----																				
Prim.cons.	G	E	P	2	Q	V	A	P	L	P	E	Y	T	2	P	Q	A	2	A	2	V	E	P	A	2	Q	P	2	L	P	E	A	V	2	T	D	K	G	T	S	I	Q	E	A	E	P	A	V	Q	P	2	L	P	E	A	V	2	D	K

```

          610         620         630         640         650         660
SP195x2    GEPAVQPELPEAVVTDKGETEVQPESPDTVVS---DKGEPKQVAPLPEYTGPQAS-----
CDC1087_00 GE-----TEVQPESPDTVVS---DKGEPKQVAPLPEYTGPQAS-----
G54xxx4    GE-----TEVQPESPDTVVS---DKGEPKQVAPLPEYTGPQAS-----
CGSP14     GE-----TEVQPESPDTVVS---DKGEPKQVAPLPEYTGPQAS-----
R6xxxx0    VS-----PDTSSENTGEVSSNPSDSTTSVGESNKPEHNDSKNENSEKT
70585x1    VS-----PDTSSENTGEVSSNPSDSTTSVGESNQPEKNRTATK-----
TIGR4x7    VS-----PDTSSKNTGEVSSNPSDSTTSVGESNKPEHNDSKNENSEKT
Hungary19A_6 GV-----PEVQPELPEAVVS---DKGVPEVQPELPEAVVSDKG-----
          .:..: . * * * . . . * *
Prim.cons.  GEPAVQPELPEAVVTDKGE2EVQPESPDTVVSNPSDKGEPKQVAPLPEYTGPQASNSEKT

          670         680         690         700         710         720
SP195x2    -----AIVEPEQ-VAPLPEYTGVQAGSIVEPEKVEAPKEYTGKIEQPSA
CDC1087_00 -----AIVEPEQ-VAPLPEYTGVQAGSIVEPEKVEAPKEYTGKIEQPSA
G54xxx4    -----AIVEPEQ-VAPLPEYTGVQAGSIVEPEKVEAPKEYTGKIEQPSA
CGSP14     -----AIVEPEQ-VAPLPEYTGVQAGSIVEPEKVEAPKEYTGKIEQPSA
R6xxxx0    VEEVPVPNPNEGTVEGTSNQTEKPVQPAEETQTNSG--KIANENTGEVSNKPSSDSKPPVE
70585x1    -----P-----ENSGNTTSEN--GTPEPSNGNS-----TEDVSTESTNSN-----
TIGR4x7    VEEVPVPNPNEGTVEGTSNQTEKPVQPAEETQTNSG--KIANENTGEVSNKPSSDSKPPVE
Hungary19A_6 -----EPEQ-VAPLPEYKGNI-----
          .*: *
Prim.cons.  VEEVPVPNPNEGTVEGTAIVEPEQPVAPLPEYTG2QAGSIVEPEKVEAPKEYTGKIEQPSA

          730         740         750         760         770         780
SP195x2    EDTKPENEASSTNGESER-----P
CDC1087_00 EDTKPENEASSTNGESER-----P
G54xxx4    EDTKPENEASSTNGESER-----P
CGSP14     EDTKPENEASSTNGESER-----P
R6xxxx0    ESNQPEKNGTATKPENSG-----N
70585x1    -SNGNE-----IKQEN-----E
TIGR4x7    ESNQPEKNGTATKPENSGNTTSENGQTEPEPSNGNSTEDVSTESNTSNSNGNEEIKQENE
Hungary19A_6 -----EQ-----VKPETP-----
          *: *
Prim.cons.  EDTKPENEASST2GESERNTTSENGQTEPEPSNGNSTEDVSTESNTSNSNGNEEIKQENP

          790         800         810         820         830         840
SP195x2    KDKIKEEKQVDKKLELRNVSNVELYTVENNKYRHITAVDGALDSSLKYFMKVKSENFKDI
CDC1087_00 KDKIKEEKQVDKKLELRNVSNVELYTVENNKYRHITAVDGALDSSLKYFMKVKSENFKDI
G54xxx4    KDKIKEEKQVDKKLELRNVSNVELYTVENNKYRHITAVDGALDSSLKYFMKVKSENFKDI
CGSP14     KDKIKEEKQVDKKLELRNVSNVELYTVENNKYRHITAVDGALDSSLKYFMKVKSENFKDI
R6xxxx0    TTSENGQTEPEKKLELRNVSDIELYSQTNGTYRQHVSLDGIPENTDTYFVKVKSSAFKDV
70585x1    LDPDKMVEEPEKKLELRNVSDIELYSQTNGTYRQHVSLDGIPENTDTYFVKVKSSAFKDV
TIGR4x7    LDPDKMVEEPEKTLELRNVSDLELYSLSNGTYRQHISLEQVPSNPNSYFVKVKSSSFKDV
Hungary19A_6 -VEKLQEEPEKTIELRSVSDVELYSYEANQSKQHTILDKVPNKSAYFIKVKSAFKDV
          .:..:***.***:***: . : : : . **:* **
Prim.cons.  KDKIKEEK222KKLELRNVS2VELY2VENNKYR22TADGA2D2SLKYFMKVKSENFKD2

          850         860         870         880         890         900
SP195x2    MLPVTKIESTTKNNKEVYKIVAHAENLIQHENNVISNDYTYLPKTQQSETGVYTSFKNL
CDC1087_00 MLPVTKIESTTKNNKEVYKIVAHAENLIQHENNVISNDYTYLPKTQQSETGVYTSFKNL
G54xxx4    MLPVTKIESTTKNNKEVYKIVAHAENLIQHENNVISNDYTYLPKTQQSETGVYTSFKNL
CGSP14     MLPVTKIESTTKNNKEVYKIVAHAENLIQHENNVISNDYTYLPKTQQSETGVYTSFKNL
R6xxxx0    YIPVASITEKRNGQSVYKITAKAEKLQELNKYVDNFTFYLDKKAEENTNFTSFSNL
70585x1    YIPVASITEKRNGQSVYKITAKAEKLQELNKYVDNFTFYLDKKAEENTNFTSFSNL
TIGR4x7    YLPVASISERKNDKILYKITAKVEKLQEIESRYKDNFTFYLAKGTEETNFTSFSNL
Hungary19A_6 LLPVSNIEDTTKDGKVVYKITARLEHLKQSVDNQYQDNFTYLPKVVQNDGPIYTSFKKL
          **:.* . : : : ***.***:***: . : : : .**:* **
Prim.cons.  MLPVTKIESTTKNNKEVYKI2AHAENLIQHENNV2S22TYLPKTQQSETGVYTSFKNL

```



	910	920	930	940	950	960
SP195x2	VDAMNSDPNGT	FHLGATMDARE	VELPDDQESY	VKNFEYFKL	LIGENNGKY	YAIYNLKKPLF
CDC1087_00	VDAMNSDPNGT	FHLGATMDARE	VELPDDQESY	VKNFEYFKL	LIGENNGKY	YAIYNLKKPLF
G54xxx4	VDAMNSDPNGT	FHLGATMDARE	VELPDDQESY	VKNFEYFKL	LIGENNGKY	YAIYNLKKPLF
CGSP14	VDAMNSDPNGT	FHLGATMDARE	VELPDDQESY	VKNFEYFKL	LIGENNGKY	YAIYNLKKPLF
R6xxxx0	VKAINQNPSGT	YHLAASLNANE	VELGPDERSY	IKDTFTGR	LIGEKDGKN	YAIYNLKKPLF
70585x1	VKAINQNPSGT	YHLAASLNANE	VELGPDERSY	IKDTFTGR	LIGEKDGKN	YAIYNLKKPLF
TIGR4x7	VKAINQNPSGT	YHLAASLNANE	VELGPDERSY	IKDTFTGR	LIGEKDGKN	YAIYNLKKPLF
Hungary19A_6	VTDMNSRPGD	GTILGATMNA	REFELGENQ	ESYVTKNF	TGTLIGSHQ	GKHYAIYNLKKPLF
Prim.cons.	VDAMNSDPNGT	FHLGATM2ARE	VEL2DDQESY	VKNFE2GKL	LIGENNGKY	YAIYNLKKPLF

	970	980	990	1000	1010	1020
SP195x2	KTLNLTATIQL	NLSIKEANVSS	KEDAAATISK	EAKYNTLID	NVHSDGIIA	GERGIGGLVSKVD
CDC1087_00	KTLNLTATIQL	NLSIKEANVSS	KEDAAATISK	EAKYNTLID	NVHSDGIIA	GERGIGGLVSKVD
G54xxx4	KTLNLTATIQL	NLSIKEANVSS	KEDAAATISK	EAKYNTLID	NVHSDGIIA	GERGIGGLVSKVD
CGSP14	KTLNLTATIQL	NLSIKEANVSS	KEDAAATISK	EAKYNTLID	NVHSDGIIA	GERGIGGLVSKVD
R6xxxx0	ENLSGATVEK	LKSLKNVAIS	GKNDIGSLAN	EATNGTKIK	QVHVDGVL	LAGERGVGGLLAKAD
70585x1	ENLSGATVEK	LKSLKNVAIS	GKNDIGSLAN	EATNGTKIK	QVHVDGVL	LAGERGVGGLLAKAD
TIGR4x7	ENLSGATVEK	LKSLKNVAIS	GKDDIGSLAN	EATNQNTKIK	QVHVDGVL	LAGERGIIGGLLAKAE
Hungary19A_6	NTLNNAIQDL	TLKDVNISG	KNHVASVAME	ATNSSTLDN	VHANGIIA	GELIGGLVQKVD
Prim.cons.	KTLNLTATIQL	NLS2KE2N2S	2KEDAA2I2	2KEAK2NTL	2IDNVHSDGII	2AGERGIGGLVSKVD

	1030	1040	1050	1060	1070	1080
SP195x2	NSRISNSSFT	GRIITNTYD	TTVGYEIGGL	LVGKLSGSL	ASIEKSIAS	IDIASNAKSGDQIVG
CDC1087_00	NSRISNSSFT	GRIITNTYD	TTVGYEIGGL	LVGKLSGSL	ASIEKSIAS	IDIASNAKSGDQIVG
G54xxx4	NSRISNSSFT	GRIITNTYD	TTVGYEIGGL	LVGKLSGSL	ASIEKSIAS	IDIASNAKSGDQIVG
CGSP14	NSRISNSSFT	GRIITNTYD	TTAGYEIGGL	LVGKLSGSL	ASIEKSIAS	IDIASNAKSGDQIVG
R6xxxx0	QSSIAESSFK	GRIIVNTY	ETTDAYNIGGL	LVGHLTGKN	ASIAKSKAT	VTISSNTNRSDQTVG
70585x1	QSNIAESSFK	GRIIVNTY	ETTDAYNIGGL	LVGHLTGKN	ASIAKSKAT	VTISSNTNRSDQTVG
TIGR4x7	QSSITESSFK	GRIINTYET	TAAAYNIGGM	VGHLTGDK	ALLTKSKAT	VAISSNTNTSDQTVG
Hungary19A_6	NSTVRNSSFT	GRIITNTF	VTTSQYAI	IGGLVGLTG	ANARVAHS	ISTVVISSNGEHSNQTTIG
Prim.cons.	NSRISNSSFT	GRIITNTYD	TTVGYEIGGL	LVGKL2GSL	ASIEKSIAS	22DI2SNAKS2DQ2VGV

	1090	1100	1110	1120	1130	1140
SP195x2	GIAGVVEKS	ATIKYSYVEG	NVNVRFHGK	VGGVAGNL	WDRDSQDV	SKSGKLSYVLSDVNV
CDC1087_00	GIAGVVEKS	ATIKYSYVEG	NVNVRFHGK	VGGVAGNL	WDRDSQDV	SKSGKLSYVLSDVNV
G54xxx4	GIAGVVEKS	ATIKYSYVEG	NVNVRFHGK	VGGVAGNL	WDRDSQDV	SKSGKLSYVLSDVNV
CGSP14	GIAGVVEKS	ATIKYSYVEG	NVNVRFHGK	VGGVAGNL	WDRDSQDV	SKSGKLSYVLSDVNV
R6xxxx0	GLAGLVDDA	HIQNSYAEG	DINNKHFGK	VAGVAGYL	WDRDTS	GEEKHAGELTNVLSDVNV
70585x1	GLAGLVDDA	HIQNSYAEG	DINNKHFGK	VAGVAGYL	WDRDTS	GEEKHAGELTNVLSDVNV
TIGR4x7	GLAGLVDRD	AQIQNSYAEG	DINNKHFG	RVAGVAGL	WDRDTS	GDVVRHAGSLTNVLSDVNV
Hungary19A_6	GIVGIVENNA	TVENSYAEGE	INNAKPFAR	VGGVAGSL	WTQEGGTD	NNTGRLTNVLSDMNI
Prim.cons.	GIAGVVEKS	ATIKYSY2EG	N2NNV2HFG	KVGGVAGNL	WDRDS2DV	SKSGKL22VLSDVNV

	1150	1160	1170	1180	1190	1200
SP195x2	TNGNAISGD	NFDNMKED	HAYSSKGNK	VNVVQVDD	DELVTKDS	SDVQRGTVLDADKVKKEKA
CDC1087_00	TNGNAISGD	NFDNMKED	HAYSSKGNK	VNVVQVDD	DELVTKDS	SDVQRGTVLDADKVKKEKA
G54xxx4	TNGNAISGD	NFDNMKED	HAYSSKGNK	VNVVQVDD	DELVTKDS	SDVQRGTVLDADKVKKEKA
CGSP14	TNGNAISGD	NFDNMKED	HAYSSKGNK	VNVVQVDD	DELVTKDS	SDVQRGTVLDADKVKKEKA
R6xxxx0	TNGNAITGY	HHTGMKVANT	FSSKANRV	FNVVLEK	DEVVSKES	SFEERTMLDASQIVSKKA
70585x1	TNGNAITGY	HHTGMKVANT	FSSKANRV	FNVVLEK	DEVVSKES	SFEERTMLDASQIVSKKA
TIGR4x7	TNGNAITGY	HYNEMKVKD	TFSSKANRV	VNVVTLVK	DEVVSKES	SFEERTMLDASQIASKKA
Hungary19A_6	TNGNAISGY	HFNKIKET	NVSNKENK	VNVVQAVG	DEVMSVD	STVERGTEIEKTSVQAKKD
Prim.cons.	TNGNAISGY	2F2GMKED2	TYSSK2NK	VNVVQVDD	EVV2KDS	SDV2RGTVLDADKVKKEKA

	1210	1220	1230	1240	1250	1260
SP195x2	ELVSKHSTKVE	DFDFTSRYTTIYNEVTGYQKSREQVYKNI	EKLLPFYNRE	TIVKYGNLVD		
CDC1087_00	ELVSKHSTKVE	DFDFTSRYTTIYNEVTGYQKSREQVYKNI	EKLLPFYNRE	TIVKYGNLVD		
G54xxx4	ELVSKHSTKVE	DFDFTSRYTTIYNEVTGYQKSREQVYKNI	EKLLPFYNRE	TIVKYGNLVD		
CGSP14	ELVSKHSTKVE	DFDFTSRYNTNYNEVTGYQKSREQVYKNI	EKLLPFYNRE	TIVKYGNLVE		
R6xxx0	EINPLTLPTVE	PLSTSGKKDSDFSKIAHYQANRALVYKNI	EKLLPFYNKSTIVKYGNLVK			
70585x1	EINPLTLPTVE	PLSTSGKKDSDFSKIAHYQANRALVYKNI	EKLLPFYNKSTIVKYGNLVK			
TIGR4x7	EINPLTLPTVE	PLSTSGKKDSDFSKVAYYQAKRNLT	YKNI	EKLLPFYNKATIVKYGNLVN		
Hungary19A_6	EITKKHTITVE	DFTNSVVKQDVDTYKAAHYQSTREVA	YKNI	EKLLPFYNRDTIVKYGNQLS		
Prim.cons.	E2VSKHST2VEDDFD2S2YD	TDYDYN2V2GYQ2SREQVYKNI	EKLLPFYNRE	TIVKYGNLVD		

	1270	1280	1290	1300	1310	1320
SP195x2	ESSELF	TKELLSVVPMKNN	EVITDIN	NKNQEINK	LLLHFE	GNKSQV
CDC1087_00	ESSELF	TKELLSVVPMKNN	EVITDIN	NKNQEINK	LLLHFE	GNKSQV
G54xxx4	ESSELF	TKELLSVVPMKNN	EVITDIN	NKNQEINK	LLLHFE	GNKSQV
CGSP14	DNSDL	FTKLLSVVPMKNN	EVITDIN	NKNQEINK	LLLHFE	GNKSRV
R6xxx0	ENSLLY	QKELLSAVMMKDDQ	VITDIV	SNKQTANK	LLLHYNDH	SSSEKFDLKYQ
70585x1	ENSLLY	QKELLSAVMMKDDQ	VITDIV	SNKQTANK	LLLHYNDH	SSSEKFDLKYQ
TIGR4x7	ENSLLY	QKELLSAVMMKDDQ	VITDIV	SNKQTANK	LLLHYKDDL	SEKLDLKYQ
Hungary19A_6	TDHNL	FKKELLSAVMMKND	EVSDIT	SNKNSIN	HLLHLYKDG	TSEKVDLT
Prim.cons.	ENS2L	FTKELLSVPMKNN	EVITDIN	2NKNQEINK	LLLH2E2NK	SE2L22AYK

	1330	1340	1350	1360	1370	1380
SP195x2	YSIEYQ	GLLYTPNTLLHDYSN	IVDNVLTDL	NSVQYDS	NAIKKIL	DISDKV
CDC1087_00	YSIEYQ	GLLYTPNTLLHDYSN	IVDNVLTDL	NSVQYDS	NAIKKIL	DISDKV
G54xxx4	YSIEYQ	GLLYTPNTLLHDYSN	IVDNVLTDL	NSVQYDS	NAIKKIL	DISDKV
CGSP14	YDIANT	KLMYTPNTMLHDYNN	IVKTI	LNDL	KSVQYSS	ADVRKVL
R6xxx0	YNLGN	TGLLYTPNQFLYDRDS	IVKE	VLPEL	QKLDYQSD	AI
70585x1	YNLGN	TGLLYTPNQFLYDRDS	IVKE	VLPEL	QKLDYQSD	AI
TIGR4x7	YSLGN	TGLLYTPNQFLYDQTS	IIKQ	VL	PDQKVDYH	SEAIRK
Hungary19A_6	YNIEN	TGLVYTPNAFLKNYTP	IIDRV	KNDL	TAVPYNP	TALKN
Prim.cons.	YSIENT	GLLYTPNTFLHDYSN	IV2NVL	2DL2SV	QYDSNAI	2K2L2IS

	1390	1400	1410	1420	1430	1440
SP195x2	FVKTK	ANIKDTLSKLLSADAA	IAENSNS	IIDNYVIQ	KIKQ	NKEALLGLTY
CDC1087_00	FVKTK	ANIKDTLSKLLSADAA	IAENSNS	IIDNYVIQ	KIKQ	NKEALLGLTY
G54xxx4	FVKTK	ANIKDTLSKLLSADAA	IAENSNS	IIDNYVIQ	KIKQ	NKEALLGLTY
CGSP14	FEKTK	ANIEDLSKLLTADAA	IVENN	NKVIDN	YVIEK	IKNKEALLGLTY
R6xxx0	FSKTK	QNLGDSLKLLSADAGLA	-SDNSV	TRGYLVDK	IKNKEALLGLTY	LERWYNFNY
70585x1	FSKTK	QNLGDSLKLLSADAGLA	-SDNSV	TRGYLVDK	IKNKEALLGLTY	LERWYNFNY
TIGR4x7	FAKTK	QLEDSLKLLSADAGLA	-SANP	VT	EGYLV	DKIKR
Hungary19A_6	FTKTK	EHLTETLKKLLSADAAVS	-GNNDI	IIDNYIV	DKIKR	NKEALLGLTY
Prim.cons.	FVKTKAN	2KD2L2KLLSADAAIAENSNS	2IDNYV	2DKIK2N	KNEALLGLTY	LERWYNFKY

	1450	1460	1470	1480	1490	1500
SP195x2	GETKA	KDLVMYHLDFFGKSNSS	ALDNVI	QLGKSGFN	NLLAKNNVI	TYNVLL
CDC1087_00	GETKA	KDLVMYHLDFFGKSNSS	ALDNVI	QLGKSGFN	NLLAKNNVI	TYNVLL
G54xxx4	GETKA	KDLVMYHLDFFGKSNSS	ALDNVI	QLGKSGFN	NLLAKNNVI	TYNVLL
CGSP14	GETNA	KDLIMYHLDFFGKSNSS	ALDNV	IELGKSGFN	NLLAKNNVI	TYNVLL
R6xxx0	GQVNV	KDLVMYHPDFFGKGN	TSPLD	T	IELGKSGFN	NLLAKNNVD
70585x1	GQVNV	KDLVMYHPDFFGKGN	TSPLD	T	IELGKSGFN	NLLAKNNVD
TIGR4x7	GQVNV	KDLVLYHLDFFGKGN	ASPLD	T	IELGKSGFN	NLLAKNNVD
Hungary19A_6	DKASAK	DLMLFHMDFFGKGN	TSPLDAI	IELGKSGY	NNLLAKNNV	TYNALL
Prim.cons.	GETNAK	DLVMYHLDFFGK2NSS	2LDNV	IELGKSGFN	NLLAKNNVI	TYNVLL

	1510	1520	1530	1540	1550	1560
SP195x2	FKALEGYRKVFLP	TTSNNEWFKAQ	TKAYIVEEKSSI	PEVSKKQSE	QGTKYSIGIYD	RLTS
CDC1087_00	FKALEGYRKVFLP	TTSNNEWFKAQ	TKAYIVEEKSSI	PEVSKKQSE	QGTKYSIGIYD	RLTS
G54xxx4	FKALEGYRKVFLP	TTSNNEWFKAQ	TKAYIVEEKSSI	PEVSKKQSE	QGTKYSIGIYD	RLTS
CGSP14	FKALEGYRKVFLP	TISNNEWFKQ	TKAYIVEEKSTI	EGREKQKGE	GTKYSIGVYDR	LTN
R6xxxx0	FSTLEHYRKVFLP	NTSNNDWFKSE	TKAYIVEEKSTI	EVKTKQGLA	GTKYSIGVYDR	ITS
70585x1	FSTLEHYRKVFLP	NTSNNDWFKSE	TKAYIVEEKSTI	EVKTKQGLA	GTKYSIGVYDR	ITS
TIGR4x7	FSTLEHYRKVFLP	NTSNNDWFKSE	TKAYIVEEKSTI	EVKTKQGLA	GTKYSIGVYDR	ITS
Hungary19A_6	FSALEGYRKA	FAPHQTNNE	WFKSQTKAYIVEEKS	NIEVKTQGLV	GTKYSIGVYDR	ITS
Prim.cons.	F2ALEGYRKVFLP	TTSNNEWFKSQ	TKAYIVEEKSTI	EVKTKQGL2	GTKYSIGVYDR	2TS

	1570	1580	1590	1600	1610	1620
SP195x2	PSWKYQSMVLP	LLTLPEEKT	VFMIANISTI	GFGAYDRYRSK	VHPKGDNLN	KFVEDNVREA
CDC1087_00	PSWKYQSMVLP	LLTLPEEKT	VFMIANISTI	GFGAYDRYRSK	VHPKGDNLN	KFVEDNVREA
G54xxx4	PSWKYQSMVLP	LLTLPEEKT	VFMIANISTI	GFGAYDRYRSK	VHPKGDNLN	KFVEDNVREA
CGSP14	PSWKYQSMVLP	LLTLPEEKT	VFMIANISTI	GFGAYDRYRS	EYPKGKLNK	FVEDNKEA
R6xxxx0	ATWKYRNMVLP	LLTLPE-RS	VFVISTMS	SLGFGAYDRY	RSSDHKAGK	ALNDFVEENARET
70585x1	ATWKYRNMVLP	LLTLPE-RS	VFVISTMS	SLGFGAYDRY	RSSDHKAGK	ALNDFVEENARET
TIGR4x7	ATWKYRNMVLP	LLTLPE-RS	VFVISTMS	SLGFGAYDRY	RSSDHKAGK	ALNDFVEENARET
Hungary19A_6	ATWKYRNMVLP	LLTLPE-RS	VFVISTL	SSLGFGAYDRY	RNSEHKAGK	DLNDFVEENARET
Prim.cons.	22WKY22MVL	PLLTLPEEK2	VF2I22IS	22GFGAYDRY	RSS2H22GK	2LN2FVE2NARE2

	1630	1640	1650	1660	1670	1680
SP195x2	AKRFRDHYDYWYK	ILEPENREKLYR	SLLVYDAFKFGR	DNTE	DKVITYQAD	FETDHPAIKYF
CDC1087_00	AKRFRDHYDYWYK	ILELENREKLYR	SLLVYDAFKFGR	DNTE	DKVITYQAD	FETDHPAIKYF
G54xxx4	AKRFRDHYDYWYK	ILEPENREKLYR	SLLVYDAFKFGR	DNTE	DKVITYQAD	FETDHPAIKYF
CGSP14	AKRFRDHYDYWYK	ILDNDNKEKLYR	SILVYDAFKFG	TDKDKDVTH	QATFETDHPA	IKYF
R6xxxx0	AKRQRDHYDYWYR	ILDEQSREKLYR	TILLYDAYKFG	DDTTS	GKATAEAK	FDS
70585x1	AKRQRDHYDYWYR	ILDEQSREKLYR	TILLYDAYKFG	DDTTS	GKATAEAK	FDS
TIGR4x7	AKRQRDHYDYWYR	ILDDNAREKLYR	NILLYDAYKFG	DDNTV	GKATEVAD	FDPNPAMQHF
Hungary19A_6	AKHQRDHYDYWYR	ILDDNAREKLYR	NILLYDAYKFG	DDNTV	GKATEVAD	FDPNPAMQHF
Prim.cons.	AKR2RDHYDYWY2	IILD3ENREKLYR	SIL2YDA2KF	GDDNTE2K2	TYQADF2TD	2PA2KYF

	1690	1700	1710	1720	1730	1740
SP195x2	FGPA	GNNVVHNGHGAYAT	GDAFY	YMA	YRMLDKDGA	VTYTHEMTHNSDREIYLG
CDC1087_00	FGPA	GNNVVHNGHGAYAT	GDAFY	YMA	YRMLDKDGA	VTYTHEMTHNSDREIYLG
G54xxx4	FGPA	GNNVVHNGHGAYAT	GDAFY	YMA	YRMLDKDGA	VTYTHEMTHNSDREIYLG
CGSP14	FGPA	GNNVVHNGHGAYAT	GDAFY	YMA	YRMLDKDGA	VTYTHEMTHNSDREIYLG
R6xxxx0	FGPV	GKVVHNQHGAYAT	GDV	YYMS	YRMLDKDGA	I
70585x1	FGPV	GKVVHNQHGAYAT	GDV	YYMS	YRMLDKDGA	I
TIGR4x7	FGPV	GKVVHNQHGAYAT	GDV	YYMG	YRMLDKDGA	I
Hungary19A_6	FGPV	GKVVHNQHGAYAT	GDV	YYMG	YRMLDKDGA	I
Prim.cons.	FGP2GN2V	VHN2HGAYAT	GDA2Y	YMA	YRMLDKDGA2	TYTHEMTH2SD22IYLG

	1750	1760	1770	1780	1790	1800
SP195x2	LGPEFYAKGLLQAPD	HPYDPTITINS	VLKYD--DSEN	STR	LQVADPTQRFNSA	EADLHNYM
CDC1087_00	LGPEFYAKGLLQAPD	HPYDPTITINS	VLKYD--DSEN	STR	LQVADPTQRFNSA	EADLHNYM
G54xxx4	LGPEFYAKGLLQAPD	HPYDPTITINS	VLKYD--DSEN	STR	LQVADPTQRFNSA	EADLHNYM
CGSP14	LGPEFYAKGLLQAPD	HPYDPTITINS	VLKYE--DSEN	STR	LQVADPTQRFNSA	EADLHNYM
R6xxxx0	LGPEF	FAKGLLQAPDQ	PSDATITINS	ILKHSKSDS	TEGSRLQVLD	PTERFQNAADLQNYV
70585x1	LGPEF	FAKGLLQAPDQ	PRDATITINS	ILKHSKSDS	TEGSRLQVLD	PTERFQNAADLQNYV
TIGR4x7	LGPEF	FAKGLLQAPDHP	DDATITINS	ILKHSKSDS	TESRRLQVLD	PTTRFNNADDLKQYV
Hungary19A_6	LGPEF	FAKGLLQAPDHP	DDATITINS	ILKHSKSDS	TESRRLQVLD	PTTRFNNADDLKQYV
Prim.cons.	LGPEF2AKGLLQAPDHPYD2	TITINS2LK2SKSDS22	STR	LQV2DPTQRFN2A	EADLHNY2	

	1810	1820	1830	1840	1850	1860
SP195x2	HNMFDLIY	TLEI	ELEGRAVA	-KLDYNAKNDL	LRKIE	NKYKQDP-DGNSVYATNVVRLTMD
CDC1087_00	HNMFDLIY	TLEI	ELEGRAVA	-KLDYNAKNDL	LRKIE	NKYKQDP-DGNSVYATNVVRLTMD
G54xxx4	HNMFDLIY	TLEI	ELEGRAVA	-KLDYNAKNDL	LRKIE	NKYKQDP-DGNSVYATNVVRLTMD
CGSP14	HNMFDVIY	MLEY	LEGKAVA	-NLETNQ	KYEL	LRKIE
R6xxxx0	HNMFDLIY	MMEY	LEGQ	SIVNKL	SVYQ	KMAAL
70585x1	HNMFDLIY	MMEY	LEGQ	SIVNKL	SVYQ	KMAAL
TIGR4x7	HNMFDV	VYMLEY	LEGN	SIL-KLD	TNQ	KQQL
Hungary19A_6	HNMFDV	VYMLEY	LEGN	SIL-KLD	TNQ	KQQL
Prim.cons.	HNMFDLIY	MLEY	LEGR	22ANKLD2	NQKNDL	LRKIE

	1870	1880	1890	1900	1910	1920
SP195x2	EVNKLNSF	DSLIEN	DIITSR	GYKQ	DEYKR	NGYTTID
CDC1087_00	EVNKLNSF	DSLIEN	DIITSR	GYKQ	DEYKR	NGYTTID
G54xxx4	EVNKLNSF	DSLIEN	DIITSR	GYKQ	DEYKR	NGYTTID
CGSP14	EVNKLNSF	DSLIEN	DIITSR	GYKQ	DEYKR	NGYTTID
R6xxxx0	EARNLNSF	ESLIDH	NILSAREY	QSGDY	ERN	NGYTTIK
70585x1	EARNLNSF	ESLIDH	NILSAREY	QSGDY	ERN	NGYTTIK
TIGR4x7	EVERLRSF	NLDIDN	NILSSREY	ASGKY	ERN	NGYTTIK
Hungary19A_6	EVERLRSF	NLDIDN	NILSSREY	ASGKY	ERN	NGYTTIK
Prim.cons.	EVNKLNSF	DSLI	2N2I	22SR	2YK2	2EY2

	1930	1940	1950	1960	1970	1980
SP195x2	AFELLA	AAKGYK	EGMVPY	ISNQY	EKDA	KAAGSK
CDC1087_00	AFELLA	AAKGYK	EGMVPY	ISNQY	EKDA	KAAGSK
G54xxx4	AFELLA	AAKGYK	EGMVPY	ISNQY	EKDA	KAAGSK
CGSP14	AFELLA	AAKGYK	EGMVPY	ISNQY	EKDA	KAAGSK
R6xxxx0	AYELLA	AAKGF	KDGMVPY	ISNQY	EEDAK	QQGQT
70585x1	AYELLA	AAKGF	KDGMVPY	ISNQY	EEDAK	QQGQT
TIGR4x7	AYELLA	AAKGF	KDGMVPY	ISNQY	EEDAK	QQGQT
Hungary19A_6	AYELLA	AAKGF	KDGMVPY	ISNQY	EEDAK	QQGQT
Prim.cons.	A2ELLA	AAK2	K2GM	VVPY	ISNQY	EKDA

	1990	2000	2010	2020	2030	2040
SP195x2	VEFKKAMY	NERIAK	FNNLIS	VSVFNN	PNVSV	WGRTNR
CDC1087_00	VEFKKAMY	NERIAK	FNNLIS	VSVFNN	PNVSV	WGRTNR
G54xxx4	VEFKKAMY	NERIAK	FNNLIS	VSVFNN	PNVSV	WGRTNR
CGSP14	TQFKK	DYKERE	KQF	SKLNR	VNF	INPN
R6xxxx0	AEFKT	AMYQ	ERVD	QF	GNL	KQV
70585x1	AEFKT	AMYQ	ERVD	QF	GNL	KQV
TIGR4x7	SEFKK	AMYQ	ERVD	QF	GNL	KQV
Hungary19A_6	SDFKK	AMYK	ERQDK	F	NKLNK	I
Prim.cons.	VEFKKAMY	2ER	ID2	FNNL	2SV	SN2

	2050	2060	2070	2080
SP195	-FRAQLY	PD	TNSR	V
CDC1087_00	-FRAQLY	PD	TNSR	V
G54	-FRAQLY	PD	TNSR	V
CGSP14	-DVAKFY	P	ETNSR	V
R6	PRWSNY	P	EID	S
70585	TRWSNY	P	EID	S
TIGR4	YHWN	N	P	E
Hungary19A_6	PHWY	N	P	E
Prim.cons.	PFWANY	P	2T	2SR

## Apêndice B.2 – Alinhamento Múltiplo de Hialuronato liase

	10	20	30	40	50	60
D39	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
R6	MILQYVYWSV	YMQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
Taiwan19F_14	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
Hungary19A_6	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
CGSP14	MILQYVYWSV	YMQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
G54	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
P1031	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
CDC1087_00	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
TIGR4	-----	MQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
Prim.cons.	MILQYVYWSV	YMQTKTKKLI	VSLSSLVLS	SGFLLNH	YMTVGA	EETTTNTI
*****;*****						
	70	80	90	100	110	120
D39	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
R6	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
Taiwan19F_14	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
Hungary19A_6	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
CGSP14	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
G54	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
P1031	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
CDC1087_00	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
TIGR4	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
Prim.cons.	RDTK	NLVENG	DFGQ	TEDGSS	PWTG	SKAQGWS
*****						
	130	140	150	160	170	180
D39	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
R6	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
Taiwan19F_14	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
Hungary19A_6	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
CGSP14	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
G54	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
P1031	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
CDC1087_00	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
TIGR4	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
Prim.cons.	EKLRA	AVHRM	VP	IEAK	KKYK	LRFK
****.*:****.* *****;*:*****.*****						
	190	200	210	220	230	240
D39	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
R6	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
Taiwan19F_14	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
Hungary19A_6	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
CGSP14	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
G54	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
P1031	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
CDC1087_00	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
TIGR4	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
Prim.cons.	QTIE	ADYS	PTLD	VVDK	IKLE	LFYET
***** * *:*****						
	250	260	270	280	290	300
D39	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
R6	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
Taiwan19F_14	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
Hungary19A_6	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
CGSP14	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
G54	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
P1031	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
CDC1087_00	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
TIGR4	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
Prim.cons.	IGKK	HVFL	ADYTY	KVEN	PDV	ASV
***** *****						

	310	320	330	340	350	360
D39	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	IYLWEKFSN		
R6	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	IYLWEKFSN		
Taiwan19F_14	DAYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	IYLWEKFSN		
Hungary19A_6	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		
CGSP14	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		
G54	DAYTARLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		
P1031	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		
CDC1087_00	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQAAR	TYLWEKFSN		
TIGR4	DAYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		
Prim.cons.	DTYTDRLDDWNGI	IAGNQYYDSKNEQMAKLNQ	ELEGKVADSLSSISSQADR	TYLWEKFSN		

	370	380	390	400	410	420
D39	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
R6	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
Taiwan19F_14	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
Hungary19A_6	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
CGSP14	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
G54	YKMSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWML	HKHVYNSEKSI
P1031	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
CDC1087_00	YKTSANLTV	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWML	HKHVYNSEKSI
TIGR4	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI
Prim.cons.	YKTSANLTA	TYRKLEEMAKQVTN	PSSRYQDET	VVRTV	RDSMEWMH	HKHVYNSEKSI

	430	440	450	460	470	480
D39	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
R6	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
Taiwan19F_14	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
Hungary19A_6	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
CGSP14	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
G54	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
P1031	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
CDC1087_00	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
TIGR4	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	
Prim.cons.	WDYEIGTPRAINNTLSLMKEYFS	SDEEIKKYTDVIEKFVPD	PEHFRKTTDNP	PFKALG	GNLV	

	490	500	510	520	530	540
D39	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
R6	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
Taiwan19F_14	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
Hungary19A_6	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
CGSP14	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
G54	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
P1031	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
CDC1087_00	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
TIGR4	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	
Prim.cons.	DMGRVKVIAGLLRKDDQEIS	STIRSEIQVFKLVDQ	GEGFYQDGSY	IDHTNVAYT	GAYGNV	

	550	560	570	580	590	600
D39	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
R6	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
Taiwan19F_14	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
Hungary19A_6	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
CGSP14	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
G54	LIDGLS	QLLPVIQKTKSP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
P1031	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
CDC1087_00	LIDGLS	QLLPVIQKTKSP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
TIGR4	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	
Prim.cons.	LIDGLS	QLLPVIQKTKNP	IDKDKMQTMYHWIDK	SFAPLLVNGEL	MDSRGRSISRANSEG	



	610	620	630	640	650	660																																																			
D39	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
R6	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
Taiwan19F_14	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
Hungary19A_6	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
CGSP14	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
G54	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
P1031	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
CDC1087_00	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
TIGR4	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
Prim.cons.	HVA	AEV	LRG	IHR	IADM	SEGETKQRLQSLVKTIVQSDSYDVFKNLKYKDISLMQSLLS																																																			
	670	680	690	700	710	720																																																			
D39	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
R6	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
Taiwan19F_14	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
Hungary19A_6	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
CGSP14	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
G54	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
P1031	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
CDC1087_00	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
TIGR4	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
Prim.cons.	DAG	VAS	VPR	TSY	LSAF	NKMDKTAMYN	AEKGF	GFGL	SLSF	SSR	TLN	YEH	MN	KEN	KRG	WY	TSD																																								
	730	740	750	760	770	780																																																			
D39	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
R6	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
Taiwan19F_14	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
Hungary19A_6	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
CGSP14	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
G54	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANV																																
P1031	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
CDC1087_00	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
TIGR4	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
Prim.cons.	GM	FYL	YNG	DLS	SH	YSD	GY	WPT	VN	PY	KMP	GGT	TET	DA	KRA	DS	TG	KV	LP	SA	FV	GT	SK	LD	DANA																																
	790	800	810	820	830	840																																																			
D39	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
R6	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
Taiwan19F_14	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
Hungary19A_6	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
CGSP14	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
G54	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
P1031	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
CDC1087_00	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
TIGR4	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
Prim.cons.	TAT	MD	F	TN	W	N	Q	T	L	A	H	K	S	W	F	M	L	K	D	K	I	A	F	L	G	S	N	I	Q	N	T	S	T	D	T	A	A	T	I	D	Q	R	K	L	E	S	S	N	P	Y	K	V	Y	V			
	850	860	870	880	890	900																																																			
D39	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
R6	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
Taiwan19F_14	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
Hungary19A_6	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
CGSP14	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
G54	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
P1031	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
CDC1087_00	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
TIGR4	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S
Prim.cons.	ND	KE	AS	L	T	E	Q	E	K	D	Y	P	E	T	Q	S	V	F	L	E	S	S	D	S	K	K	N	I	G	Y	F	F	F	K	K	S	S	I	S	M	S	K	A	L	Q	K	G	A	W	K	D	I	N	E	G	Q	S

	910	920	930	940	950	960
D39	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
R6	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
Taiwan19F_14	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
Hungary19A_6	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
CGSP14	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
G54	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
P1031	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
CDC1087_00	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
TIGR4	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					
	*****	*****	*****	*****	*****	*****
Prim.cons.	DKEVENEFLTISQAHKQNGDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAK					

	970	980	990	1000	1010	1020
D39	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
R6	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
Taiwan19F_14	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
Hungary19A_6	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
CGSP14	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
G54	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
P1031	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
CDC1087_00	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
TIGR4	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					
	*****	*****	*****	*****	*****	*****
Prim.cons.	QGVWGIWKYDSDVSTISNQFQVLKRGVYTIIRKEGDEYKIAAYNPETQESAPDQEVFKKLE					

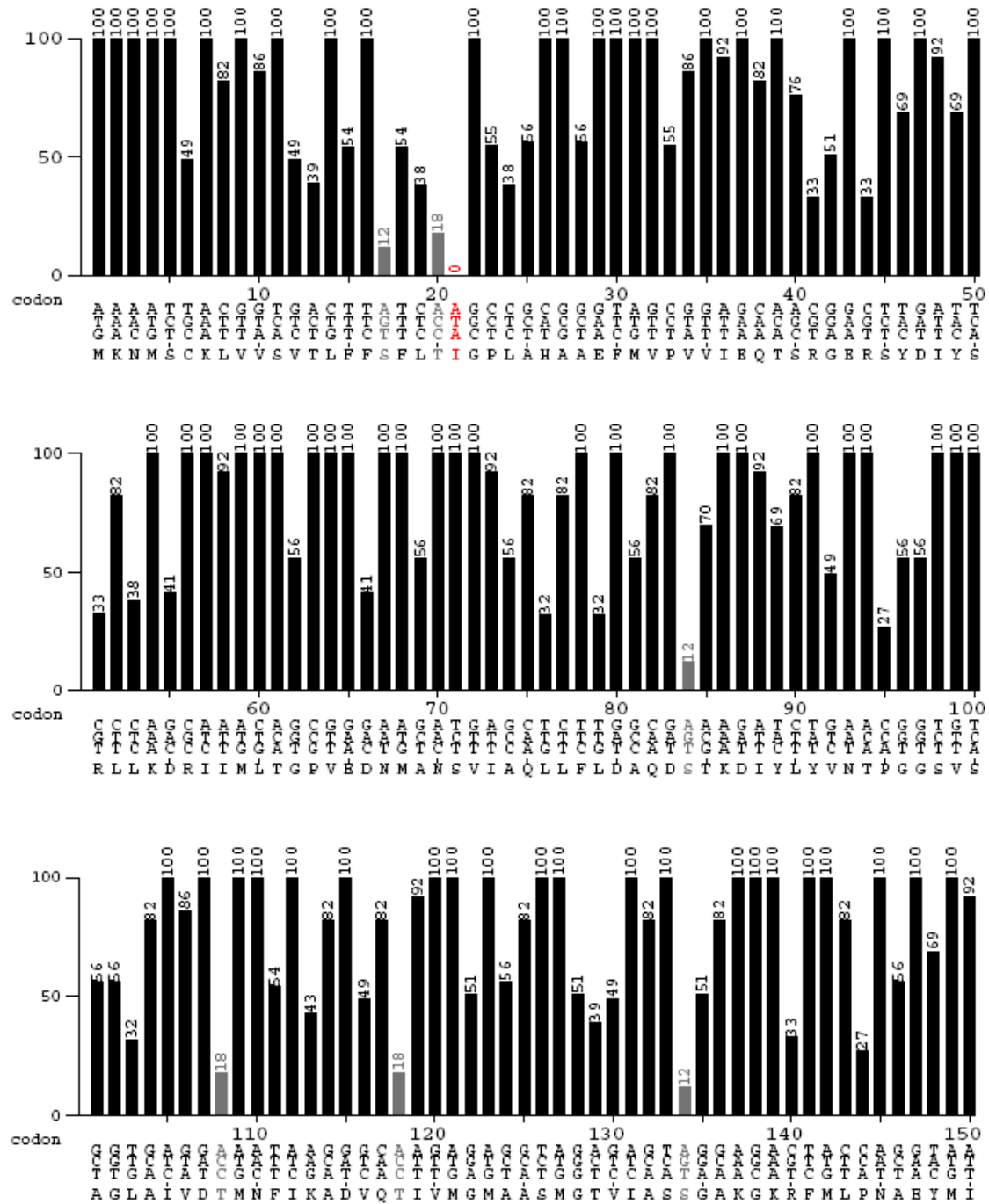
	1030	1040	1050	1060	1070
D39	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
R6	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
Taiwan19F_14	QATQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
Hungary19A_6	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
CGSP14	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
G54	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
P1031	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
CDC1087_00	QAAQLQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
TIGR4	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				
	**:*	*****	*****	*****	*****
Prim.cons.	QAAQPQVQNSKEKEKSEEEKNHSDQKNLPQTGEGQSILASLGFLLLGAFYLFRRGKNN				

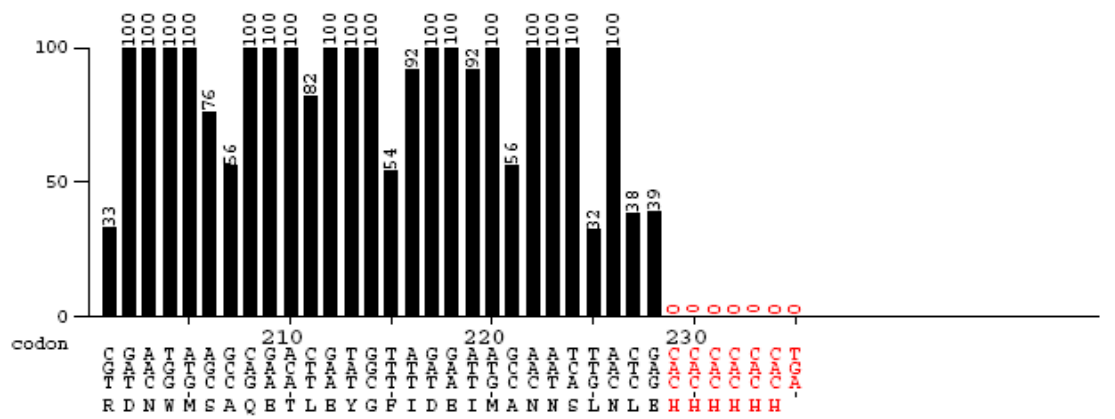
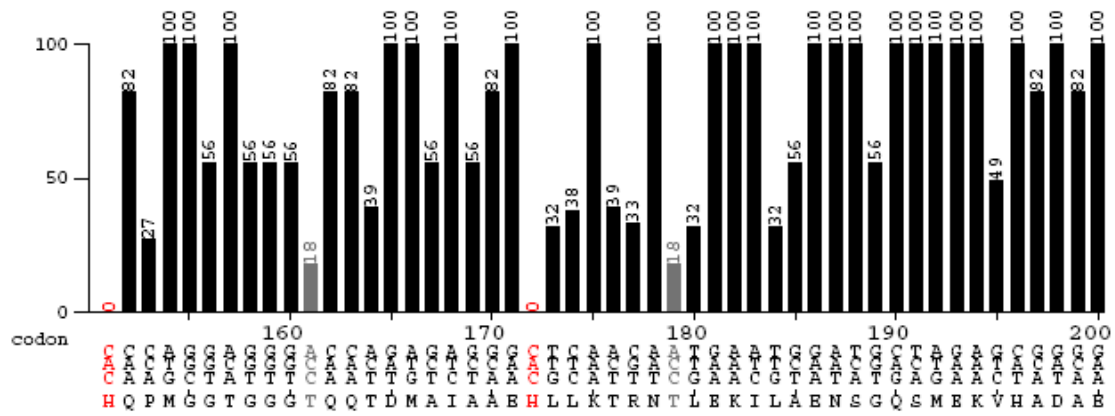


## APÊNDICE C

### Apêndice C.1 – Frequência de códons em *B. subtilis* para a sequência do gene *clpP* antes da otimização

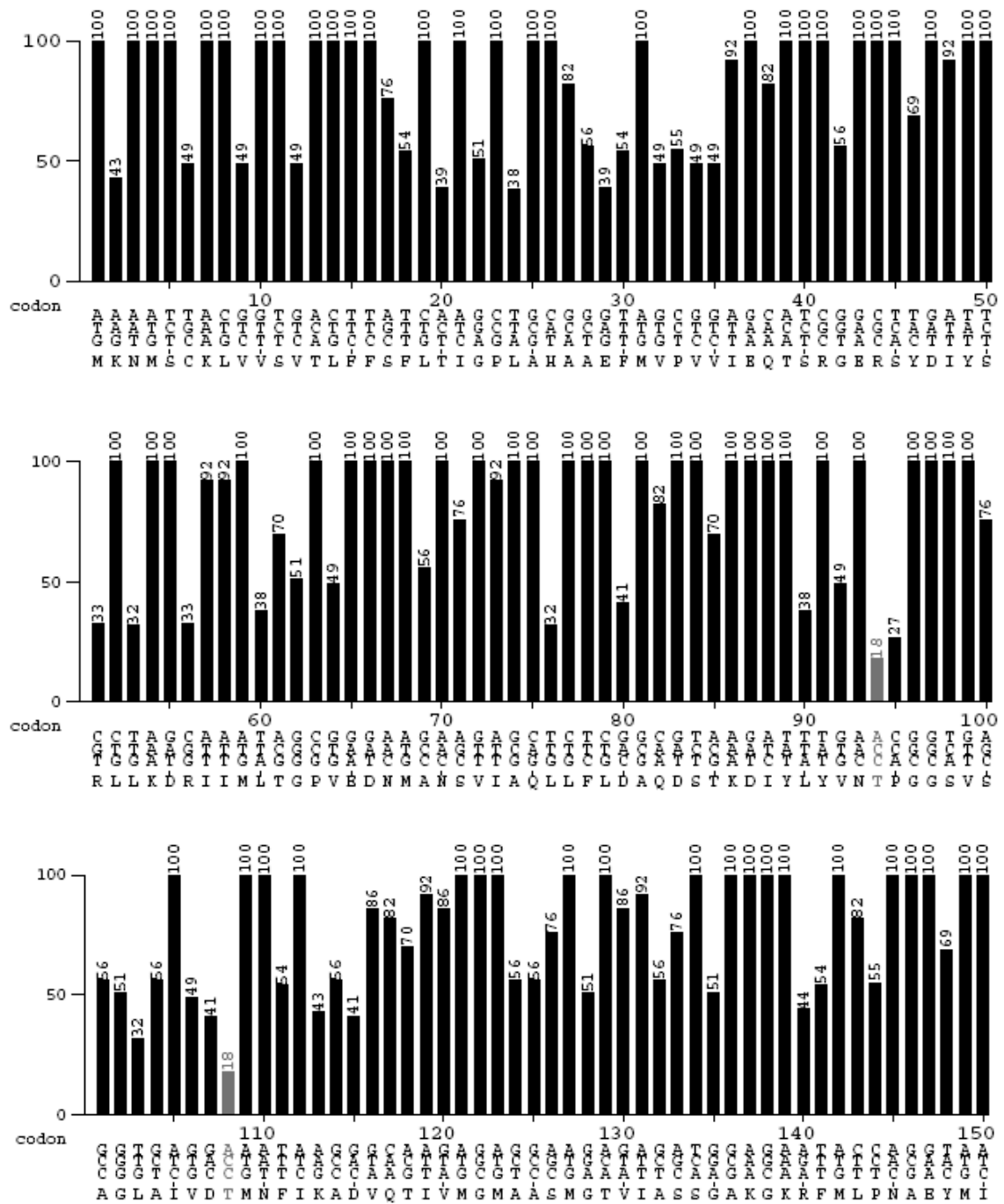
Análise de frequência de códons realizada pelo programa Graphical Codon Usage Analyser 2.0, disponível no sítio [www.gcu.de](http://www.gcu.de).

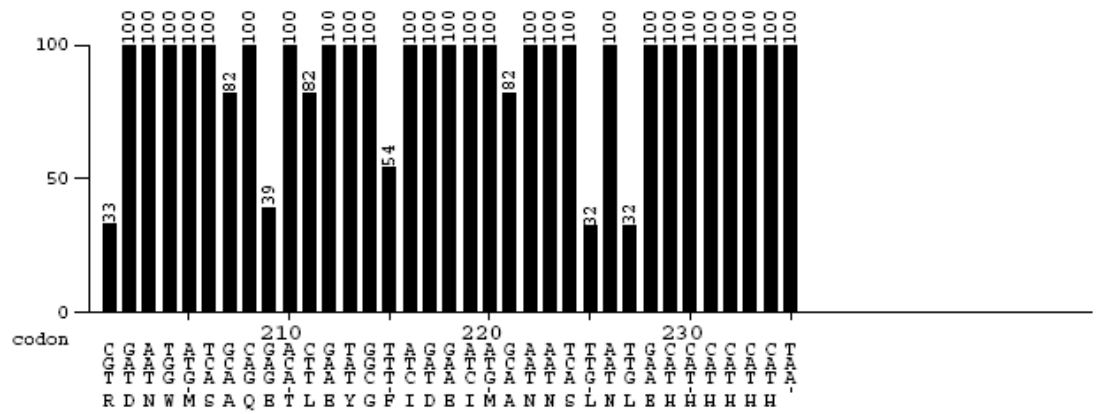
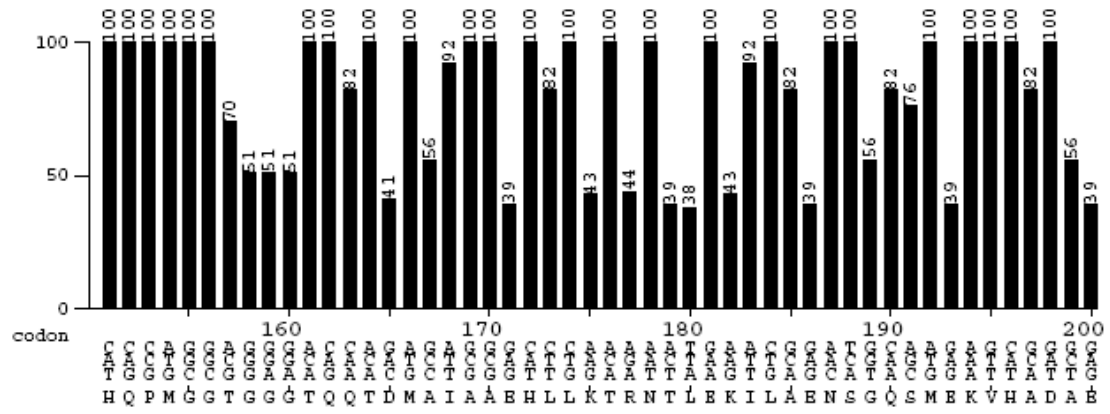




**Apêndice C.2 – Frequência de códons em *B. subtilis* para a seqüência do gene *clpP* otimizado**

Análise de frequência de códons realizada pelo programa Graphical Codon Usage Analyser 2.0, disponível no sítio [www.gcu.a.de](http://www.gcu.a.de).





## REFERÊNCIAS BIBLIOGRÁFICAS

AKHTAR, M.S., BHAKUNI, V., 2004, “*Streptococcus pneumoniae* hyaluronate lyase: An overview”, **Current Science**, v.86 (Jan), n.2, p.285-295.

ALMEIDA, R.V., 2001, **Clonagem e Expressão do Gene de uma Lipase de *Pyrococcus furiosus* em *Escherichia coli***. Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

ALTSCHUL, S. F., GISH, W., MYERS, E. W. *et al.*, 1990, "Basic Local Alignment Search Tool", **Journal of Molecular Biology**, v. 215, n.3, pp. 403-410.

ANDERSSON, L., YANG, S., NEUBAUER, P., 1996, “Impact of plasmid presence and induction on cellular responses in fed batch cultures of *Escherichia coli*”, **Journal of Biotechnology**, v. 46, p.255-263.

BAHERI, H.R., HILL, G.A., ROESLER W.J., 2001, “Modelling plasmid instability in batch and continuous fermentors”, **Biochemical Engineering Journal**, v.8, p.45-50.

BANEYX, F., 1999, "Recombinant protein expression in *Escherichia coli*", **Current Opinion in Biotechnology**, v. 10, p. 411-421.

BAROCCHI, M.A., CENSINI, S., RAPPUOLI, R., 2007, “Vaccines in the era of genomics: The pneumococcal challenge”, **Vaccine**, v.25, p.2963–2973.

BENDTSEN, J.D., NIELSEN, H., VON HEIJNE, G. *et al.*, 2004, “Improved Prediction of Signal Peptides: SignalP 3.0”, **Journal of Molecular Biology**, v. 340, pp. 783-795.

BERRY, A.M., LOCK, R.A., THOMAS, S.M. *et al.*, 1994, “Cloning and Nucleotide Sequence of the *Streptococcus pneumoniae* Hyaluronidase Gene and Purification of the Enzyme from Recombinant *Escherichia coli*”, **Infection and Immunity**, v. 62 (Mar), n.3, p.1101-1108.

BOGAERT, D., HERMANS, P.W.M., ADRIAN, P.V. *et al.*, 2004, “Pneumococcal vaccines: an update on current strategies”, **Vaccine**, v.22, p.2209–2220.

BRAIDO, F., BELLOTTI, M., MARIA, A.D. *et al.*, 2008, “The role of Pneumococcal vaccine”, **Pulmonary Pharmacology and Therapeutics**, v.21, p.608– 615.

CAO, J., LI D., GONG, Y. *et al.*, 2009, “Caseinolytic protease: a protein vaccine which could elicit serotype-independent protection against invasive pneumococcal infection”, **Clinical and Experimental Immunology**, v.156, p.52-60.

CAO, J., CHEN T., LI D. *et al.*, 2008, “Mucosal immunization with purified ClpP could elicit protective efficacy against pneumococcal pneumonia and sepsis in mice”, **Microbes and Infection**, v.10, p.1536-1542.

CAO, J., CHEN, D., XU, W. *et al.*, 2007, “Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP”, **Vaccine**, v.25, p. 4996–5005.

CAO, Y., XIA, Q., FANG, B. *et al.*, 2006, “Optimization of expression of *dhaT* gene encoding 1,3-propanediol oxidoreductase from *Klebsiella pneumonia* in *Escherichia coli* using the methods of uniform design and regression analysis”, **Journal of Chemical Technology and Biotechnology**, v.81, p.109-112.

CHIAVOLINI, D., MEMMI, G., MAGGI, T. *et al.*, 2003, “The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice”, **BMC Microbiology**, v.3(Jul), n.1, p.1-9.

CHUAN, Y. P., LUA, L.H.L., MIDDELBERG A.P.J., 2008, “High-level expression of soluble viral structural protein in *Escherichia coli*”, **Journal of Biotechnology**, v.134, p.64–71.

CHOI, J.H., KEUM, K.C., LEE, S.Y., 2006, "Production of recombinant proteins by high cell density culture of *Escherichia coli*", **Chemical Engineering Science**, v.61, p. 876 – 885.

CORCHERO, J.L., VILLAVERDE A., 1998, "Plasmid Maintenance in *Escherichia coli* Recombinant Cultures is Dramatically, Steadily, and Specifically Influenced by Features of the Encoded Proteins", **Biotechnology and Bioengineering**, v.58, n.6, p.625-632.

DAGAN, R., FRASCH, C., 2009, "Introduction", **The Pediatric Infectious Disease Journal**, v.28 (Apr.), n.4, p.S63-S65.

DEMAIN, A.L., VAISHNAV, P., 2009, "Production of recombinant proteins by microbes and higher organisms", **Biotechnology Advances**, v.27, p.297–306.

DI GUILMI, A.M., DESSEN, A., 2002, "New approaches towards the identification of antibiotic and vaccine targets in *Streptococcus pneumoniae*", **EMBO reports**, v.3, p.728-734.

DING F., TANG P., HSU M., *et al.*, 2009, "Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14", **BMC Genomics**, v.10, n.158, p.1-13.

FERREIRA, L.C.S., FERREIRA, R. C.C., SCHUMANN, W., 2005, "*Bacillus subtilis* as a tool for vaccine development: from antigen factories to delivery vectors", **Anais da Academia Brasileira de Ciências**, v.77, n.1, p.113-124.

FRIEHS, K., 2004, "Plasmid Copy Number and Plasmid Stability" **Advances in Biochemical Engineering Biotechnology**, v.86, p.47–82.

FU, L.L., XU, Z.R., LI, W.F. *et al.*, 2007, "Protein secretion pathways in *Bacillus subtilis*: Implication for optimization of heterologous protein secretion", **Biotechnology Advances**, v.25, p.1-12.

GARCÍA-SUÁREZ, M.M., VÁZQUEZ, F., MÉNDEZ, F. J., 2006, “*Streptococcus pneumoniae* virulence factors and their clinical impact: an update”, **Enfermedades Infecciosas y Microbiología Clínica**, p.512-517, 2006.

GRIBUN, A., KIMBER, M. S., CHING, R. *et al.*, 2005, “The ClpP Double Ring Tetradecameric Protease Exhibits Plastic Ring-Ring Interactions, and the N Termini of Its Subunits Form Flexible Loops That Are Essential for ClpXP and ClpAP Complex Formation”, **The Journal of Biological Chemistry**, v.280, n.16 (abril), p.16185-16196.

GRIMPREL, E., 2009, “Segurança e imunogenicidade de uma vacina pneumocócica conjugada 13-valente administrada com a vacinação pediátrica de rotina em bebês saudáveis na França”. In: **Foco em Vacina Pneumocócica Conjugada: atual e futura**, p.8-10.

GUPTA, R., SHARMA, P., VYAS, V.V., 1995, “Effect of growth environment on the stability of a recombinant shuttle plasmid, pCPPS-31, in *Escherichia coli*”, **Journal of Biotechnology**, v.41, p.29-37.

HAN G.H., SHIN H., KIM, S.W., 2008, “Optimization of bio-indigo production by recombinant *E. coli* harboring *fmo* gene”, **Enzyme and Microbial Technology**, v.42, p.617–623.

HANNING, G., MAKRIDES, S. C., 1998, “Strategies for optimizing heterologous protein expression in *Escherichia coli*”, **Trends in Biotechnology**, v.16, p.54-60.

ISLAM, R.S., TISI, D., LEVY, M.S. *et al.*, 2007, “Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design”, **Biotechnology Progress**, v. 23, p.785-793.

JANA, S., DEB, J. K., 2005, “Strategies for efficient production of heterologous proteins in *Escherichia coli*”, **Applied Microbiology and Biotechnology**, v.67, p. 289-298.



JEDRZEJAS, M.J., 2001, “Pneumococcal Virulence Factors: Structure and Function”, **Microbiology and Molecular Biology Reviews**, v.65 (Jun), n.2, p. 187–207.

KIM, D. Y., KIM, K. K., 2008, “The Structural Basis for the Activation and Peptide Recognition of Bacterial ClpP”, **Journal of Molecular Biology**, V.379, p.760–771.

KADIOGLU, A., WEISER, J.N., PATON, J.C. *et al.*, 2008, “The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease”, **Nature Reviews Microbiology**, v.6 (abril), p.288-301.

KWON, H.Y., OGUNNIYI, A. D., CHOI, M. *et al.*, 2004, “The ClpP Protease of *Streptococcus pneumoniae* Modulates Virulence Gene Expression and Protects against Fatal Pneumococcal Challenge”, **Infection and Immunity**, v.72, n.10 (Out.), p.5646-5653.

LARENTIS, A. L., ALMEIDA, R. V., MARTINS, O. B., ALVES, T. L. M., 2006, **Engenharia de Bioprocessos Recombinantes**. Escola Piloto Virtual do Programa de Engenharia Química/COPPE.

LEÓN, A.D., ISLÃS, H.J., CUEVAS, M.G. *et al.*, 2004, “Analysis of the expression of the *Trichoderma harzianum ech42* gene in two isogenic clones of *Escherichia coli* by surface response methodology”, **Process Biochemistry**, v.39, p.2173-2178.

LI, W., ZHOU, X., LU, P., 2004, “Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*”, **Research in Microbiology**, v.155, p.605–610.

LO, P.K., HASSAN, O., AHMAD, A. *et al.*, 2007, “Excretory over-expression of *Bacillus* sp. G1 cyclodextrin glucanotransferase (CGTase) in *Escherichia coli*: Optimization of the cultivation conditions by response surface methodology”, **Enzyme and Microbial Technology**, v. 40, p.1256-1263.

- LOBATO, F., 2010, “Bio-Manguinhos assina cooperação para introduzir a vacina pneumocócica no calendário nacional e desenvolver imunizantes contra dengue, febre amarela e malária” **BioNotícias - Publicação Bimestral de Bio-Manguinhos/Fiocruz**, n.42 (Jan).
- MAKRIDES, S.C., 1996, “Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*”, **Microbiological Reviews**, v.60 (Sept), n.3, p.512-538.
- MALDONADO, L.M.T.P., HERNÁNDEZ, V.E.B., RIVERO, E.M. *et al.*, 2007, “Optimization of culture conditions for a synthetic gene expression in *Escherichia coli* using response surface methodology: the case of human interferon beta”, **Biomolecular Engineering**, v.24, p.217-222.
- MARÍ, Y.M., ESPINOSA, A.E.S., UBIETA, R. *et al.*, 1999, “Effect of the Selection Marker on the Viability and Plasmid Stability of Two Human Proteins with Neurotrophic Action Expressed in *Escherichia coli*”, **Biochemical and Biophysical Research Communications**, v.258, p. 29–31.
- MERGULHÃO, F.J.M., SUMMERS, D.K., MONTEIRO, G.A., 2005, “Recombinant protein secretion in *Escherichia coli*”, **Biotechnology Advances**, v.23, p.177–202.
- MISTRY, D., STOCKLEY, R. A., 2006, “IgA1 protease”, **The International Journal Of Biochemistry & Cell Biology**, v.38, p. 1248-1248.
- NGUYEN, H. D., NGUYEN, Q. A., FERREIRA, R. C. *et al.*, 2005, “Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability”, **Plasmid**, v.54, p.241–248.
- NIKAREL, I.E., TOKSOY, E., KIRDAR, B. *et al.*, 2005, “Optimizing medium composition for TaqI endonuclease production by recombinant *Escherichia coli* cells using response surface methodology”, **Process Biochemistry**, v.40, p1633-1639.

- NIKAREL, I.E., ÖNER, E., KIRDAR, B. et al., 2006, “Optimization of medium composition for biomass production of recombinant *Escherichia coli* cells using response surface methodology”, **Biochemical Engineering Journal**, v.32, p.1-6.
- NOVAK R, HENRIQUES B, CHARPENTIER E. et al., 1999, “Emergence of vancomycin tolerance in *Streptococcus pneumonia*”, **Nature**, v.399, p.590–593.
- OGUNNIYI, A.D., GRABOWICS, M., BRILES, D.E. et al., 2007, “Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*”, **Infection and Immunity**, v.75, p.350-357.
- OMS (Organização Mundial da Saúde), 2007, “Pneumococcal conjugate vaccine for childhood immunization – WHO position paper”, **Weekly epidemiological record**, v.82, n.12, p.93-104. Disponível em: <<http://www.who.int/wer/2007/wer8212.pdf>>, Acesso em: 18/12/2009.
- PAOLIS, F., BEGHETTO, E., SPADONI, A. et al., 2007, “Identification of a human immunodominant B-cell epitope within the immunoglobulin A1 protease of *Streptococcus pneumonia*”, **BMC Microbiology**, 7:113, p.
- PATON, J.C., 1998, “Novel Pneumococcal surface proteins: role in virulence and vaccine potencial”, **Trends in Microbiology**, v.6, p.85-87.
- PETI, W., PAGE, R., 2007, “Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost”, **Protein Expression and Purification**, v.51, p.1–10.
- PLETZ, M.W., MAUS, U., KRUG, N. et al., 2008, “Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species”, **International Journal of Antimicrobial Agents**, v.32, n.3, p.199-206.

PONNURAJ, K., JEDRZEJAS, M.J., 2000, “Mechanism of hyaluronan binding and degradation: structure of *Streptococcus pneumoniae* hyaluronate lyase in complex with hyaluronic acid disaccharide at 1.7 Å resolution”, **Journal of Molecular Biology**, v.299 (Jun), n.4, p.885-895.

POULSEN, K., REINHOLDT J., JESPERGAARD C. *et al.*, 1998, “A Comprehensive Genetic Study of Streptococcal immunoglobulin A1 Proteases: Evidence for Recombination within and between Species”, **Infection and Immunity**, v.66 (Jan), n.1, p. 181-190.

REN, X., YU, D., HAN, S. *et al.*, 2006, “Optimization of recombinant hyperthermophilic esterase production from agricultural waste using response surface methodology”, **Bioresource Technology**, v. 97, p.2345-2349.

RODRIGUES E IEMMA, 2005, **Planejamento de Experimentos e Otimização de Processos – Uma estratégia sequencial de planejamentos**, 1ed., São Paulo, Brasil, Editora Casa do Pão.

ROMANELLO, V., MARCACCI, M., DAL MOLIN, F. *et al.*, 2006, “Cloning, expression, purification, and characterization of *Streptococcus pneumoniae* IgA1 protease”, **Protein Expression and Purification**, v.45, p.142–149.

SCHUMANN, W., FERREIRA, L. C. S., 2004, “Production of recombinant proteins in *Escherichia coli*”, **Genetics and Molecular Biology**, v.27, p.442-453.

SCHWAAB, M., PINTO, J.C., 2007, **Análise da Dados Experimentais, I: fundamentos de estatística e estimação de parâmetros**. 1ed. Rio de Janeiro, Brasil, Editora e-papers.

SILVA, M., 2005, **Clonagem, Expressão e Purificação das Proteínas de Superfície, PsaA e fragmentos de PspA de *Streptococcus pneumoniae***. Tese de D.Sc., IPT/USP, São Paulo, SP, Brasil.

SNEATH, P. H.A., BERGEY, D. H., 1986, “Gram-positive *Bacteria* other than *Actinomycetes*”. In: Holt J.G. (eds), **Bergey’s Manual of Systematic Bacteriology**, 1 ed., v.2, Baltimore, USA, Williams & Wilkins.

SØRENSEN, H.P., MORTENSEN, K.K., 2005, “Advanced genetic strategies for recombinant protein expression in *Escherichia coli*”, **Journal of Biotechnology**, v.115, p.113–128.

SUNITHA, K., KIM, Y., LEE, J. *et al.*, 2000, “Statistical optimization of seed and induction conditions to enhance phytase production by recombinant *Escherichia coli*”, **Biochemical Engineering Journal**, v.5, p.51-56

SWALLEY, S.E., FULGHUM J.R., CHAMBERS, S.P., 2006, “Screening factors effecting a response in soluble protein expression: Formalized approach using design of experiments”, **Analytical Biochemistry**, v.351, p.122–127.

SWARTZ, J.R., 2001, “Advances in *Escherichia coli* production of therapeutic proteins”, **Current Opinion in Biotechnology**, v.12, p.195–201.

TERPE, K., 2006, “Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems”, **Applied Microbiology and Biotechnology**, v.72, p.211-222.

THOMPSON, J. D., HIGGINS, D. G., GIBSON, T. J., 1994, “Clustal W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weigthing, Position-Specific Gap Penalties and Weigth Matrix Choice”, **Nucleic Acids Research**, v. 22, p.4673-4680.

TOKSOY, E., ÖNSAN, Z.Í., KIRDAR, B., 2002, “High-level production of TaqI restriction endonuclease by three different expression systems in *Escherichia coli* cells using the T7 phage promoter”, **Applied Microbiology and Biotechnology**, v.59, p.239-245.

TOMAZETTO, G., MULINARI, F., STANISÇUASKI, F. *et al.*, 2007, “Expression kinetics and plasmid stability of recombinant *E. coli* encoding urease-derived peptide with bioinsecticide activity”, **Enzyme and Microbial Technology**, v.41, p.821-827.

XU, J., LI, W., WU, J. *et al.*, 2006, “Stability of plasmid and expression of a recombinant gonadotropin-releasing hormone (GnRH) vaccine in *Escherichia coli*”, **Applied Microbiology and Biotechnology**, v.73, p.780–788.

WANG, Y.H., JING, C.F., YANG, B. *et al.*, 2005, “Production of a new sea anemone neurotoxin by recombinant *Escherichia coli*: Optimization of culture conditions using response surface methodology”, **Process Biochemistry**, v.40, p.2721–2728.

WATSON, J. D., GILMAN, M., WITKOWSKI, J. *et al.*, 1992, **Recombinant DNA**. 2 ed. New York, Scientific American Books.

WU, X., LEE, W., TRAN, L. *et al.*, 1991, “ Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases”, **Journal of Bacteriology**, v. 173 (Ago), n. 16, p.4952-4958.

WYSOCKI, J., TEJEDOR, J.C., GRUNERT, D. *et al.*, 2009, “Immunogenicity of the 10-Valent Pneumococcal Non-Typeable *Haemophilus influenza* Protein D Conjugate Vaccine (PHiD-CV) When Coadministered With Different Neisseria meningitidis Serogroup C Conjugate Vaccines”, **The Pediatric Infectious Disease Journal**, v.28 (Apr.), n.4, p.S77-S88.

ZAHA, A., FERREIRA, H. B., PASSAGLIA, L. M. P. *et al.*, 1996, “Técnicas de DNA Recombinante”, **Biologia Molecular Básica**, ed., Capítulo 15, Porto Alegre, Mercado Aberto.

ZHANG, X., LI, Y., ZHUGE, B. *et al.*, 2006, “Optimization of 1,3-propanediol production by novel recombinant *Escherichia coli* using response surface methodology”, **Journal of Chemical Technology e Biotechnology**, v.81, p.1075-1078.