

Universidade Federal de Santa Catarina
Centro Tecnológico
Departamento de Engenharia Química e Engenharia de Alimentos
Programa de Pós-Graduação em Engenharia Química

**Produção de Celulose Bacteriana: Identificação do
Operon *bcs* e Produção de Biofilme Celulósico por
*Chromobacterium violaceum***

DERCE DE OLIVEIRA SOUZA RECOUVREUX

Dissertação apresentada ao Programa de Pós-
Graduação em Engenharia Química da
Universidade Federal de Santa Catarina como
requisito parcial para obtenção do grau de
Mestre em Engenharia Química

Orientador: Prof. Dr. Luismar Marques Porto (EQA/UFSC)
Co-orientadora: Profa. Dra. Regina Vasconcellos Antonio (BQM/UFSC)

Florianópolis-SC

2004

Derce de Oliveira Souza Recouvreux

**Produção de Celulose Bacteriana: Identificação do Operon *bcs* e
Produção de Biofilme Celulósico por *Chromobacterium violaceum***

Esta dissertação foi julgada e aprovada para a obtenção de grau de **Mestre em Engenharia Química no Programa de Pós-Graduação em Engenharia Química** da Universidade Federal de Santa Catarina

Área de Concentração
Engenharia Genômica e Engenharia Biomédica

Prof. Dr. Luismar Marques Porto
Orientador

Profa. Dra. Regina Vasconcellos Antônio
Co-Orientadora

Prof. Dr. Agenor Furigo Junior
Coordenador do Programa de Pós-Graduação em Engenharia Química

Banca examinadora:

Prof. Dr. Luismar Marques Porto

Profa. Dra. Regina Vasconcellos Antônio

Prof. Dr. Agenor Furigo Junior

Prof. Dr. André Oliveira de Souza Lima

Florianópolis - SC, novembro de 2004.

Recouvreux, Derce de Oliveira Souza

Produção de Celulose Bacteriana: Identificação do Operon *bcs* e
Produção de Biofilme Celulósico por *Chromobacterium violaceum*.
124p.

Tese (Mestrado) – Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia Química.

1. Biofilme – 2. Celulose bacteriana – 3. Operon *bcs* – 4.
Chromobacterium violaceum

Este trabalho é parte integrante das pesquisas realizadas pelo Grupo de Engenharia Genômica e foi desenvolvido no Laboratório de Tecnologias Integradas (Intelab) do Departamento de Engenharia Química e Engenharia de Alimentos e no Laboratório de Bioquímica e Biologia Molecular de Microrganismos (LBBMM) do Departamento de Bioquímica, Universidade Federal de Santa Catarina.

Dedico este trabalho aos meus pais, Herculano e Maria,
ao meu marido Sergio, e aos meus filhos Ricardo e
Catherine, por todo amor, compreensão e carinho
dedicados neste momento tão importante de minha vida.

Amo vocês!

AGRADECIMENTOS

Agradeço ao Professor Luismar Marques Porto, pela oportunidade de ingressar numa área tão fascinante, pela orientação, conhecimentos transmitidos, atenção, sugestões e principalmente por sua amizade.

À Professora Regina Vasconcellos Antônio, agradeço pela co-orientação, pela paciência, dedicação, amizade e por todo conhecimento transmitido.

À Ana Paula Miqueleto que mesmo a distância prestou grande ajuda para a realização deste trabalho.

Agradecimento especial aos colegas Ana Kelly, Danielle, Claudimir, Itamar, Artiva e Cauê pela amizade, companheirismo e pela incansável ajuda na realização deste trabalho. A ajuda de vocês foi essencial. Sempre lembrarei de vocês com muito carinho. E também agradeço a Luciani pelo apoio dedicado, a Cristiana, Simão e Diogo pela acolhida no grupo.

A todos os colegas do laboratório Intelab e LBBMM, que de alguma forma contribuíram para que este trabalho fosse realizado.

À Coordenadoria de Pós-Graduação em Engenharia Química, em especial ao Edivilson, por todo seu carinho e atenção.

A Novozymes Latin America Ltda que gentilmente forneceu as enzimas para este trabalho, em especial à Sabrina Ganem por toda atenção. Ao Professor Willibaldo Schmidell Netto, a Professora Valéria Reginatto Spiller e a Professora Edna Regina Amante pela ajuda com as enzimas.

Aos meus pais, meu marido, meus filhos Ricardo e Catherine agradeço por todo amor e carinho com que me apoiaram, desculpem-me pela ausência. A toda a minha família (irmãos e sobrinhos) obrigada pelas palavras de incentivo e principalmente a minha irmã Dilceia e a querida Haíra pelo apoio no inglês.

SUMÁRIO

ÍNDICE DE FIGURAS	x
ÍNDICE DE TABELAS	xiv
LISTA DE ABREVIATURAS	xv
RESUMO	xvi
ABSTRACT	xvii
CAPÍTULO I	1
Introdução, Motivação e Justificativa	1
CAPÍTULO II.....	5
Revisão Bibliográfica	5
2.1 Biofilme Bacteriano	5
2.1.1 Por que os microrganismos formam biofilmes?	6
2.1.2 Ciclo de desenvolvimento de um biofilme bacteriano	7
2.1.3 Exopolissacarídeos.....	8
2.1.4 Monitoramento da densidade populacional (<i>Quorum Sensing</i>)	8
2.2 Celulose	9
2.3 Celulose Bacteriana	13
2.3.1 Aplicações da celulose bacteriana	13
2.3.2 Biossíntese de celulose em bactérias	15
2.3.3 Principais proteínas do complexo celulose sintase em bactérias.....	17
2.3.3.1 Proteína YhjQ	18
2.3.3.2 Proteína BcsA: celulose sintase	18
2.3.3.3 Proteína BcsB: proteína de ligação de c-di-GMP	21
2.3.3.4 Proteína BcsZ: celulase	22
2.3.3.5 Proteína BcsC.....	23
2.3.4 Produção de celulose bacteriana.....	23
2.3.5 Proteínas com motivo GGDEF	24

2.4 Microrganismo Modelo: <i>Chromobacterium violaceum</i>	25
2.5 Elementos da Transcrição em Procariotos	26
2.5.1 Elementos estruturais da seqüência do DNA e RNA.....	26
2.5.2 O Código Genético.....	27
2.5.3 Quadro Aberto de Leitura	28
2.5.4 Características estruturais da transcrição e tradução	29
2.5.5 Tradução do mRNA em proteína	31
2.5.6 Família de proteínas, domínios e motivos	32
2.6 Operon	33
2.6.1 Operon <i>bcs</i>	34
2.7 Bioinformática e Genômica Comparativa	34
2.7.1 Bancos de dados biológicos	35
2.7.2 Genômica comparativa	36
2.7.3 Ferramentas de bioinformática e de genômica comparativa	36
2.7.3.1 Plataformas <i>on-line</i> para pesquisa biológicas.....	37
2.7.3.2 Bancos de dados de seqüência de DNA e estrutura de proteínas	40
2.7.3.3 Ferramentas de genômica comparativa e para análise de seqüências	
.....	44
CAPÍTULO III	48
Materiais e Métodos	48
3.1 Ferramentas de bioinformática utilizadas	48
3.2 Procedimentos Experimentais Computacionais	49
3.2.1 Identificação das enzimas da via de biossíntese de celulose em	
<i>Chromobacterium violaceum</i>	49
3.2.2 Identificação do operon de biossíntese de celulose em <i>Chromobacterium</i>	
<i>violaceum</i>	49
3.2.3 Identificação de proteínas com motivo GGDEF em <i>Chromobacterium</i>	
<i>violaceum</i>	52
3.3 Procedimentos Experimentais Microbiológicos	52
3.3.1 Microrganismo utilizado na etapa laboratorial.....	52
3.3.2 Materiais e equipamentos utilizados nos ensaios	52
3.3.2.1 Meio de cultivo LB (Luria Bertani) para o cultivo de <i>Chromobacterium</i>	
<i>violaceum</i>	53
3.3.2.2 Enzima utilizada para a hidrólise enzimática.....	53
3.3.2.3 Materiais e equipamentos utilizados para a determinação de glicose	
em amostras do cultivo	53
3.3.2.4 Solução tampão utilizada na hidrólise enzimática	53
3.3.3 Cultivo de <i>Chromobacterium violaceum</i> para produção de biofilme	54

3.3.3.1 Pré-Inóculo	54
3.3.3.2 Inóculo e condições de cultivo	54
3.3.4 Coleta das amostras para análise.....	54
3.3.5 Hidrólise enzimática.....	55
3.3.6 Determinação da concentração de glicose.....	56
CAPÍTULO IV	57
Resultados e Discussão	57
4.1 Análise dos Resultados do Procedimento Computacional.....	57
4.1.1 Via de biossíntese de celulose em <i>Chromobacterium violaceum</i>	57
4.1.2 Identificação do operon <i>bcs</i> de <i>Chromobacterium violaceum</i>	59
4.1.3 Características estruturais de cada proteína/enzima codificada pelos genes que compõem o operon <i>bcs</i> de <i>Chromobacterium violaceum</i>	63
4.1.3.1 ORF CV2679.....	63
4.1.3.2 Proteína BcsA: celulose sintase	65
4.1.3.3 Proteína BcsB: proteína de ligação de c-di-GMP.....	69
4.1.3.4 Proteína BcsZ: celulase	72
4.1.3.5 Proteína BcsC.....	74
4.1.4 Predição de região regulatória de transcrição do operon <i>bcs</i> de <i>Chromobacterium violaceum</i>	75
4.1.5 Organização dos genes de biossíntese de celulose em diferentes grupos de bactérias Gram-negativas	77
4.1.6 Proteínas com motivo GGDEF em <i>Chromobacterium violaceum</i>	79
4.2 Produção de Biofilme Celulósico por <i>Chromobacterium violaceum</i>.....	82
4.2.1 Resultados do cultivo de <i>Chromobacterium violaceum</i>	82
4.2.2 Concentração de glicose em amostras de exopolissacarídeo celulósico produzido por <i>Chromobacterium violaceum</i>	84
CAPÍTULO V	87
Conclusões e Sugestões	87
5.1 Conclusões.....	87
5.2 Sugestões para Trabalhos Futuros.....	89
CAPÍTULO VI	90
Referências Bibliográficas.....	90

ÍNDICE DE FIGURAS

- Figura 1 – Etapas do ciclo de vida de um biofilme microbiano: aderência a uma superfície adequada; agregação formando colônias; produção EPS formando um biofilme maduro e desprendimento do biofilme retornando a estilo de vida livre. Durante a fase de vida livre as bactérias movimentam-se livremente em um meio líquido utilizando flagelos. Os pilos são utilizados para aderência a uma superfície e também para movimentarem-se sobre ela associando-se a outras bactérias. O sistema *quorum sensing* (QS) pode induzir a produção de EPS. Adaptado de O'TOOLE *et al.* (2000)..... 7
- Figura 2 – Estrutura da cadeia $\beta(1\rightarrow4)$ glicana (celulose). A unidade de repetição celobiose está indicada entre colchetes. Adaptado de STEPHENS (2004)..... 10
- Figura 3 – Elementos que compõem a microfibrila de celulose da parede celular de planta. Adaptado de HJEMMESIDE (2004)..... 11
- Figura 4 – Estrutura roseta associada à biogênese da microfibrila da celulose vegetal (aumento de 222.000 vezes). No canto superior esquerdo amplificação de uma das rosetas (aumento de 504.545 vezes). Obtida de DELMER (1999). Créditos: de Mark J. Grimson e Candace H. Haigler, Departamento de Ciências Biológicas, Texas Tech University, Lubbock, Texas. 12
- Figura 5 – Modelo para a estrutura roseta mostrando a formação de microfibrila de celulose. BcsA representa a celulose sintase. A subunidade roseta é formada por seis celuloses sintase e a roseta é composta por seis subunidades rosetas dando origem a microfibrila. Modificado de DOBLIN (2002)..... 12
- Figura 6 – (a) Aplicações da celulose bacteriana como pele artificial – BioFill® – Curativo biológico. (b) Tratamento de paciente com úlcera isquêmica e lesão necrótica de calcâneo e área do tendão de Aquiles como um novo substituto da pele. Úlcera fechada após três meses do início do tratamento. Figuras obtidas em BIOFILL (2004)..... 14
- Figura 7 – Via de biossíntese da celulose em *A. xylinum* a partir da glicose como fonte de carbono, conforme proposto por ROSS *et al.* (1991). 16
- Figura 8 – Modelo de regulação e biossíntese de celulose em *Acetobacter xylinum* a partir do substrato glicose (ROSS *et al.*, 1987; 1990; 1991; SKINNER e CANNON, 2000; SHODA & ANO Laboratory, 2004)..... 17

Figura 9 – Modelo da região catalítica da celulose sintase apresentado por Brown <i>et al.</i> (1996), a partir da seqüência de aminoácidos de celulose sintases (AcsAB, BcsA e AcsAII) de três linhagens de <i>A. xylinum</i> , mostrando os domínios A e B, conforme descrito no texto.	20
Figura 10 – Estrutura tridimensional hipotética da proteína BcsA. As hélices transmembrana formam um poro por onde passa a cadeia $\beta(1\rightarrow4)$ glicana que está sendo polimerizada. A figura foi adaptada de DELMER (1999) e RICHMOND (2000).	21
Figura 11 – Modelo da função da celulase na biossíntese de celulose apresentado por MATTHYSSE <i>et al.</i> (1995). O modelo sugere a transferência de oligossacarídeo ligado a lipídeo para o crescimento da cadeia $\beta(1\rightarrow4)$ glicana, após ser clivado pela celulase.....	23
Figura 12 – Reatores para produção de celulose bacteriana. (a) e (b) reatores do tipo RDF (<i>rotating disc fermentor</i>) (KRYSTYNOWICZ <i>et al.</i> , 2002) e em (c) um reator do tipo <i>airlift</i> (SHODA & ANO Laboratory, 2004).	25
Figura 13 – Elementos estruturais da seqüência de DNA e outros elementos envolvidos na transcrição do DNA para RNA e tradução em proteína.....	32
Figura 14 – Genomas bacterianos seqüenciados, depositados no GenBank até 24 de novembro de 2004.....	37
Figura 15 – Etapas para a identificação do operon <i>bcs</i> no genoma de <i>Chromobacterium violaceum</i>	50
Figura 16 -Via de biossíntese de celulose em <i>C. violaceum</i> a partir do substrato glicose.	59
Figura 17 – Esquema geral da estrutura do operon <i>bcs</i> de <i>C. violaceum</i> . As setas indicam os genes ou ORFs que compõem o operon. O nome dentro da seta corresponde ao gene ou ORF. O número acima da seta corresponde ao número de acesso do gene no GenBank do NCBI. Imediatamente abaixo das setas está indicada a massa molecular de cada polipeptídeo, bem como, o número de aminoácidos que o compõem. Os números dentro dos círculos correspondem às distâncias intergênicas. As anotações sob as setas indicam a posição de início e de término de cada gene/ORF e também o códon de início e de parada. As informações apresentadas neste esquema foram obtidas do BRGene e NCBI.....	60
Figura 18 – Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da ORF CV2679 de <i>C. violaceum</i> (NP_902349), <i>S. typhimurium</i> (NP_462521) e <i>E. coli</i> (NP_417991).....	64

- Figura 19 – Esquema gráfico dos domínios da proteína YhjQ de *S. typhimurium* obtido do PFAM. O número de acesso no SwissPfam é YHJQ_SALTY..... 64
- Figura 20 – Modelo da região catalítica da celulose sintase apresentado por BROWN *et al.* (1996), a partir da seqüência de aminoácidos de proteínas BcsA de três linhagens de *A. xylinum*, mostrando os domínios A e B, conforme descrito no texto. Ao modelo foi inserida a proteína BcsA de *C. violaceum*, (que como pode ser observado na Figura 21 apresentada a seguir), está de acordo com o mesmo número de resíduos entre os resíduos catalíticos D e motivo conservado QXXRW..... 65
- Figura 21 – Regiões conservadas obtidas do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsA de *C. violaceum* (NP_902348), *A. xylinum* (A) (AAA21884), *A. xylinum* (B) (CAA38487), *A. tumefaciens* (AAC41435), *R. leguminosarum* (AAD28574), *S. typhimurium* (NP_462520), *E. coli* (NP_417990), *P. putida* (NP_744779), e *P. fluorescens* (AAL71842)..... 67
- Figura 22 – Esquema gráfico dos domínios da proteína BcsA de *C. violaceum* obtido do PFAM. Q7NUL9 é o número de acesso no SwissPfam. 68
- Figura 23 – Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsB de *C. violaceum* (NP_902347), *S. typhimurium* LT2 (NP_462519), *E. coli* (NP_417989), *P. putida* (NP_744780), e *P. fluorescens* (AAL71843). Os retângulos salientam os resíduos de cisteína e aspartato conservados entre as espécies..... 70
- Figura 24 – Resíduos de aminoácidos conservados entre as proteínas BcsB (CelB) de *A. tumefaciens*, *R. leguminosarum* *bv.* *trifolii* e *S. typhimurium* (CAC44016.1) obtido do estudo realizado por RÖMLING (2002), sendo X_n é qualquer resíduo de aminoácido (o índice n representa o número de repetições do resíduo X)..... 71
- Figura 25 – Região N-terminal obtida do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsB de *C. violaceum* (NP_902347), *S. typhimurium* LT2 (NP_462519), *E. coli* (NP_417989), *P. putida* (NP_744780), e *P. fluorescens* (AAL71843)..... 71
- Figura 26 – Esquema gráfico dos domínios da proteína BcsB de *C. violaceum* obtido do PFAM. O número de acesso no SwissPfam é Q7NUM0..... 72
- Figura 27 – Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsZ de *C. violaceum* (NP_902347), *A. tumefaciens* (AAC41433), *R. leguminosarum* (AAD28576), *S. typhimurium* LT2 (NP_462518), *E. coli* (NP_417988), *P. putida* (NP_744781), e *P. fluorescens* (AAL71844)..... 73

Figura 28 – Esquema gráfico dos domínios da proteína BcsZ de <i>C. violaceum</i> obtido do PFAM. O número de acesso no SwissPfam é Q7NUM1.	74
Figura 29 – Esquema gráfico dos domínios da proteína BcsC de <i>C. violaceum</i> obtido do PFAM. O número de acesso no SwissPfam é Q7NUM2.	74
Figura 30 – Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsC de <i>C. violaceum</i> (NP_902347), <i>A. xylinum</i> (A) (AAA21886), <i>A. xylinum</i> (B) (CAA38489), <i>S. typhimurium</i> (NP_462517), <i>E. coli</i> (YP_026226), <i>P. putida</i> (NP_744782), e <i>P. fluorescens</i> (AAL71845), que abrangem o domínio BCSC_C.....	76
Figura 31 – Possíveis regiões regulatórias de transcrição do operon <i>bcs</i> de <i>C. violaceum</i> . Em destaque as prováveis Box -35 e -10 e a região SD.	77
Figura 32 – Organização dos genes de biossíntese de celulose em diferentes grupos de bactérias Gram-Negativas.....	78
Figura 33 – Proteínas com motivo GGDEF em <i>C. violaceum</i> . Arquitetura de domínios obtidos do PFAM.	81
Figura 34 – Culturas estática e agitada realizadas com <i>C. violaceum</i> em meio LB. Coleta com 6 dias de cultivo e precipitação do meio líquido com etanol. ..	83
Figura 35 – Determinação da concentração de glicose em amostras do produto da hidrólise enzimática, utilizando o método enzimático colorimétrico do kit de glicose enzimática. a1, e1 e p1 são amostras tratadas com celulase, enquanto a2, e2, p2 são as mesmas amostras submetidas às condições de hidrólise na ausência de celulase.....	84
Figura 36 – Concentração de glicose em amostras de exopolissacarídeos celulósico produzido por <i>C. violaceum</i> , em cultura estática e cultura agitada.	85

ÍNDICE DE TABELAS

Tabela 1 – O Código Genético.....	28
Tabela 2 – Ferramentas de bioinformática importantes e úteis para esta abordagem.	38
Tabela 3 – Ferramentas de bioinformática, genômica comparativa e para análise de seqüências utilizadas neste trabalho.....	48
Tabela 4 – Amostras isoladas de biofilme.	55
Tabela 5 – Proteínas/enzimas presentes no genoma de <i>C. violaceum</i> envolvidas na via de biossíntese de celulose.	58
Tabela 6 – Bactérias selecionadas para estudo comparativo. Número de acesso no GenBank de cada proteína/enzima do operon <i>bcs</i> das bactérias selecionadas e o nome do gene/ORF destas proteínas/enzimas.....	62
Tabela 7 – Proteínas com domínio GGDEF em <i>C. violaceum</i>	80
Tabela 8 – Concentração de glicose em amostras de exopolissacarídeos celulósico produzido por <i>C. violaceum</i> , em cultura estática e cultura agitada.	85

LISTA DE ABREVIATURAS

acil HSL	= acil homoserina lactona
ATP	= adenosina trifosfato
ADP	= adenosina difosfato
<i>bcs</i>	= <i>b</i> acterial <i>c</i> ellulose <i>s</i> ynthesis
c-di-GMP	= diguanilmonofosfato cíclico
CMC	= carboximetil celulose
COG	= <i>cluster of orthologous groups of proteins</i> ou Conjunto do grupo de proteínas ortólogas
D	= aspartato ou ácido aspártico
EC	= <i>enzyme classification</i> ou classificação de enzimas
EPS	= exopolissacarídeos
GTP	= guanosina trifosfato
G6P	= glicose-6-fosfato
G1P	= glicose-1-fosfato
LB	= Luria Bertani
ORF	= <i>open reading frame</i> ou quadro aberto de leitura
OHHL	= N-3-(oxohexanoil) homoserina lactona
PPi	= pirofosfato
TPR	= <i>tetratric peptide repeat</i>
UTP	= uridina trifosfato
UDP	= uridina difosfato
UDP-glicose	= uridina difosfato glicose

RESUMO

RECOUVREUX, Derce de Oliveira Souza. **Produção de Celulose Bacteriana: Identificação do Operon *bcs* e Produção de Biofilme Celulósico por *Chromobacterium violaceum***. 2004. 124p. Dissertação (Mestrado em Engenharia Química) - Programa de Pós-Graduação em Engenharia Química, UFSC, Florianópolis, SC.

Exopolissacarídeos (EPS) são os principais constituintes de biofilme bacteriano. Esses polímeros, componentes da matriz extracelular bacteriana, têm despertado grande interesse em aplicações industriais e na área médica. Entretanto, têm sido associado com mecanismos de combate bacteriano. A celulose tem sido identificada como um dos EPS da matriz extracelular produzido por várias espécies de bactérias durante a formação de biofilmes. A bactéria Gram-negativa *Chromobacterium violaceum*, linhagem ATCC 12472, teve seu genoma recentemente seqüenciado pelo *Brazilian National Genome Project Consortium*. A análise do genoma da *C. violaceum* mostrou a presença de genes relacionados à biossíntese de celulose, até então desconhecidos. Neste trabalho, analisou-se a estrutura e organização dos genes diretamente envolvidos na biossíntese de celulose, e buscou-se evidências experimentais da presença de celulose na matriz extracelular do biofilme formado por esta bactéria. A metodologia empregada envolveu o uso de ferramentas de bioinformática que exploram informações do seu genoma, como estrutura dos genes, elementos de regulação, domínios e motivos conservados. Ensaio laboratoriais foram utilizados para se obter evidências experimentais das informações genômicas. Estas estratégias permitiram determinar a existência de um operon que compreende cinco genes estruturais codificando as proteínas e enzimas do complexo celulose sintase, como também propor uma via metabólica de biossíntese de celulose a partir de glicose. Identificaram-se ainda 43 ORFs (*open reading frames*) com domínio GGDEF normalmente associado à atividade de síntese de diguanilmonofosfato cíclico (c-di-GMP) a partir de duas moléculas de guanosina trifosfato (GTP). A molécula c-di-GMP funciona como um componente regulatório de diversas atividades celulares, entre elas a ativação da síntese de celulose. Experimentalmente foi analisado o produto do biofilme formado durante o cultivo de *C. violaceum* em meio Luria Bertani (LB), culturas estática e agitada. Os resultados obtidos sugerem a presença de celulose como componente do biofilme.

Palavras Chave: Biofilme, celulose bacteriana, operon *bcs*, *Chromobacterium violaceum*.

ABSTRACT

RECOUVREUX, Derce de Oliveira Souza. **Produção de Celulose Bacteriana: Identificação do Operon *bcs* e Produção de Biofilme Celulósico por *Chromobacterium violaceum*.** 2004. 124p. Dissertação (Mestrado em Engenharia Química) - Programa de Pós-Graduação em Engenharia Química, UFSC, Florianópolis, SC.

Exopolysaccharides (EPS) are the mainly constituent of bacterial biofilm. These polymers components of the bacterial extracellular matrix have raised much interest for potential industrial and medical applications. However they have been associated with immunological mechanisms that fight bacteria infections. Cellulose has been identified as one of EPS in the extracellular matrix produced by several species of bacteria during the formation of biofilms. The Gram-negative bacterium *Chromobacterium violaceum*, strain ATCC 12472, had its genome recently sequenced by the Brazilian National Genome Project Consortium. The analysis of the genome of *C. violaceum* revealed the presence of genes involved in cellulose biosynthesis. In this work, was analyzed the structure and organization of those genes directly involved in cellulose biosynthesis, and search for experimental evidence of the presence of cellulose in the extracellular matrix of biofilms produced by this bacterium. The methodology employed involved use of bioinformatics tools that explore genome information, such as gene structure, regulatory, domain and conserved sequences. Laboratory assays were used, in order to obtain experimental evidence of the genomic information. These strategies allowed to determine the existence of an operon that comprises five structural genes coding for enzymes of the cellulose synthase complex as well as to propose a metabolic pathway for the biosynthesis of cellulose from glucose. It was identified 43 ORFs (*open reading frames*) with a GGDEF domain, normally associated the activity for the synthesis of cyclic diguanylate (c-di-GMP) from two molecules of guanosine triphosphate (GTP). The c-di-GMP molecule functions as a regulatory component of diverse cellular activities, among them the activation of cellulose synthesis. The biofilm product formed during the culture of *C. violaceum* in Luria Bertani (LB) medium was experimentally analyzed, in both static and agitated cultures. The obtained results strongly suggested the presence of cellulose as a component of the biofilm.

Keywords: Biofilm, bacterial cellulose, operon *bcs*, *Chromobacterium violaceum*.

CAPÍTULO I

Introdução, Motivação e Justificativa

Na maioria dos ambientes naturais, uma estrutura conhecida como biofilme está associada ao estilo de vida microbiano prevalente (WATNICK e KOLTER, 2000). O biofilme bacteriano consiste de colônias de bactérias inseridas em uma matriz polimérica extracelular por elas mesmas produzida (COSTERTON *et al.*, 1999). Essa estrutura constitui uma forma de proteção que permite ao microrganismo sobreviver em ambientes adversos. Dos componentes da matriz extracelular os exopolissacarídeos (EPS) têm despertado grande interesse em aplicações industriais e na área médica (XAVIER *et al.*, 2003), como também têm sido considerados como um problema, interferindo nos mecanismos imunológicos contra infecção bacteriana (COSTERTON *et al.*, 1999).

Embora os EPS sejam os componentes da matriz extracelular mais estudados, existe uma grande variedade de EPS que ainda permanece não caracterizada (SUNTHERLAND, 2001). A celulose tem sido identificada como um EPS da matriz extracelular, que é produzido por várias espécies de bactérias durante a formação de biofilmes (SPIERS *et al.*, 2003; ZOGAJ *et al.*, 2001).

A celulose é o polímero orgânico mais abundante encontrado na Natureza. É de grande importância na economia global, sendo o principal constituinte do algodão (acima de 94%) e o principal componente estrutural da madeira (acima de 50%). Juntos esses produtos somam as maiores fontes de celulose para diversas aplicações como: papel, tecidos, materiais de construção entre outros (BROWN, 2004).

Já a algum tem-se demonstrado que vários organismos como plantas, algas, fungos, alguns animais e algumas bactérias, tais como as α -proteobactérias *Acetobacter xylinum*, *Agrobacterium tumefaciens*, *Rhizobium spp* e a bactéria Gram-positiva *Sarcina ventriculli* são capazes de produzir celulose (ROSS *et al.*, 1991). Recentemente observou-se a produção de celulose em γ -proteobactérias como *Salmonella typhimurium*, *Escherichia coli*, *Pseudomonas fluorescens* SBW25, entre outras (SPIERS *et al.*, 2003; ZOGAJ *et al.*, 2001). A fibra de celulose bacteriana possui alta pureza química, cristalinidade, alta força de tensão, elasticidade, durabilidade, elevada capacidade de retenção e absorção de água, biodegradabilidade. Além de ser não tóxica e não alergênica (SAXENA *et al.*, 1994; JONAS e FARAH, 1998; VANDAMME *et al.*, 1998). Devido as suas características peculiares, a celulose bacteriana, tem sido utilizada em inúmeras aplicações industriais bem como na área médica.

Com a crescente explosão populacional, e conseqüentemente, com a grande demanda na industrialização de celulose, gera-se um desequilíbrio ecológico. A maioria dos produtos derivados de celulose é obtido a partir de fonte vegetal (principalmente do eucalipto e do pinheiro). O processo de purificação da celulose vegetal (liberação da lignina e hemicelulose), requer produtos químicos altamente poluidores. A celulose bacteriana é quimicamente pura (livre de lignina e hemicelulose) (BROWN, 2004). Estudos sobre a produção de celulose bacteriana contribuem para a identificação de outras fontes fornecedoras de celulose.

A via de biossíntese da celulose envolve algumas enzimas individuais e um complexo de proteínas, onde a celulose sintase apresenta-se como a principal enzima desse complexo. Em diversas bactérias observou-se que os genes que codificam as proteínas desse complexo estão organizados em um operon. O processo da biossíntese da celulose inclui a síntese de uridina difosfato glicose (UDP-glicose), a qual é substrato da celulose sintase, seguida

de sua polimerização com formação de cadeia $\beta(1\rightarrow4)$ glicana e a liberação de uridina difosfato (UDP). As associações de cadeias nascentes irão formar as fibras de celulose (WONG *et al.*, 1990; BROWN *et al.*, 1996; SAXENA *et al.*, 1990; 1991; 1994; 1995;).

A α -proteobactéria Gram-negativa *A. xylinum* é o produtor de celulose mais eficiente conhecido. Por essa razão essa bactéria tem sido utilizada como organismo modelo para a elucidação de características básicas da biossíntese de celulose, incluindo-se estudos de biossíntese de celulose em plantas (ROSS *et al.*, 1991).

A bactéria *Chromobacterium violaceum*, linhagem ATCC 12472, teve seu genoma recentemente seqüenciado pelo BRGene (*Brazilian National Genome Project Consortium*) (BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM, 2003), e os dados do seqüenciamento estão disponíveis em servidores Web (www.brgene.lncc.br; www.ncbi.nlm.nih.gov: número de acesso NC_005085). Esta bactéria tem sido caracterizada, há algum tempo, como capaz de produzir EPS (MARTIN e RICHARDS, 1963; CORPE, 1964). Estudos realizados por MARTINELLI *et al.* (2002) apresentaram evidências da formação de biofilme nesta bactéria ser regulada por um processo de comunicação intercelular bacteriano, conhecido como *quorum sensing*. A análise do genoma da *C. violaceum* mostrou a presença de três genes relacionados à biossíntese de celulose, CV2675, CV2677 e CV2678 (BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM, 2003), até então desconhecidos.

Este trabalho teve como objetivo estudar elementos genômicos da biossíntese de celulose por *C. violaceum*, bem como obter evidências experimentais da produção de biofilme celulósico por esta bactéria.

Objetivos específicos:

- A partir de informações da literatura sobre a formação de biofilme celulósico em outros microrganismos buscar informações no genoma de *C. violaceum* que evidenciem a formação do mesmo
- Identificar genes e elementos conservados nas seqüências gênicas (domínios e motivos) envolvidos na biossíntese de celulose por *C. violaceum*, utilizando ferramentas de genômica comparativa
- Identificar características gênicas, tais como: elementos regulatórios de transcrição, ordem dos genes no genoma, fita codificadora e distâncias intergênicas de genes envolvidos na biossíntese de celulose por *C. violaceum*
- Elucidar a via metabólica envolvida na biossíntese de celulose por *C. violaceum*
- Realizar cultivos de *C. violaceum*, a fim de obter evidências experimentais de produção de exopolissacarídeo celulósico como um componente do biofilme formado por esta bactéria.

CAPÍTULO II

Revisão Bibliográfica

2.1 Biofilme Bacteriano

Por muito tempo foi considerado que as bactérias viviam de forma isolada. No entanto, nas últimas décadas observou-se que ocorre não apenas associação, como também a comunicação entre as bactérias. Esse tipo de comportamento comunitário permite a formação de uma estrutura multicelular complexa denominada biofilme (FUQUA *et al.*, 1996). O estilo de vida microbiano na maioria dos ambientes naturais está associado a formação de biofilme (WATNICK e KOLTER, 2000). O estudo das interações entre bactérias pode ter um impacto significativo no desenvolvimento de produtos de interesse biotecnológico, como por exemplo, na indústria de alimentos (produção de fermentados e de iogurtes), na prevenção de doenças, em problemas de corrosão na indústria petroleira e naval, em tratamento de água e efluentes e na agricultura (SUNTHERLAND, 2001; COSTERTON *et al.*, 1999; O'TOOLE *et al.*, 2000; XAVIER *et al.*, 2003).

O biofilme pode ser constituído por apenas uma espécie de bactéria ou por várias espécies. A predominância na maioria dos ambientes é biofilme de múltiplas espécies, como por exemplo, as placas bacterianas encontradas em nossos dentes. Os biofilmes constituídos de uma única espécie existem em uma variedade de infecções e implantes médicos (O'TOOLE *et al.*, 2000). Esta observação tem requerido atenção especial por parte de vários pesquisadores da área médica, uma vez que a formação de biofilme confere resistência a agentes antimicrobianos, implicando em infecções de difícil tratamento. Como consequência o tratamento de tais infecções requer altas dosagens de

antibióticos e em determinados casos estas se tornam crônicas (COSTERTON *et al.*, 1999).

A formação de biofilme tem sido investigada a nível molecular em algumas espécies bacterianas (GENEVAUX *et al.*, 1996). Em estudos realizados por O'TOOLE e KOLTER (1998) foram analisadas culturas com organismos mutantes deficientes na produção de biofilme. Esses estudos demonstraram que a interação inicial da bactéria com uma superfície é mediada por pilos e flagelos. Em contato com a superfície as bactérias utilizam pilos para locomoverem-se até encontrarem outras bactérias, formarem colônias e iniciarem a produção de EPS. Em estudos realizados por COSTERTON *et al.* (1999) verificaram ainda que durante a fase em que vivem associadas em um biofilme as bactérias expressam genes que não são observados quando estas vivem de forma isolada.

2.1.1 Por que os microrganismos formam biofilmes?

O'TOOLE *et al.* (2000) sugerem que as bactérias iniciam a formação de biofilme em resposta a sinais ambientais, tais como disponibilidade e tipo de nutrientes, temperatura, pH e oxigênio. Os organismos que formam biofilmes em meio rico em nutrientes, quando em situações de escassez se desprendem do biofilme e retornam ao modo de vida isolado, provavelmente em busca de novas fontes de nutrientes. Essas condições variam amplamente entre os organismos. As bactérias associadas em um biofilme possuem vantagens nutricionais, proteção e facilidade de comunicação (WATNICK e KOLTER, 2000).

2.1.2 Ciclo de desenvolvimento de um biofilme bacteriano

O desenvolvimento de biofilme bacteriano envolve etapas com diferenciações estruturais, morfológicas e de expressões gênicas. Um modelo genérico de formação de biofilme é ilustrado na Figura 1, baseado no modelo proposto por O'TOOLE *et al.* (2000). As bactérias movimentam-se livremente em um meio líquido, utilizando-se de flagelos. Quando encontram uma superfície adequada podem a ela se aderir. Em contato com a superfície as bactérias movimentam-se através de pilos, associam-se a outras, e multiplicam-se formando colônias. Pela densidade populacional das colônias podem induzir por *quorum sensing* (descrito na sessão 2.1.4) a produção de EPS, formando um biofilme maduro. Posteriormente, as bactérias podem desprenderem-se do biofilme, retornando ao estilo de vida isolado completando, assim, o ciclo de desenvolvimento do biofilme.

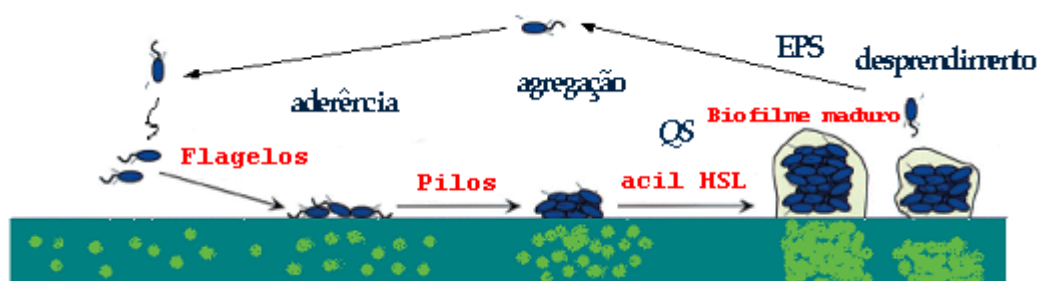


Figura 1 - Etapas do ciclo de vida de um biofilme microbiano: aderência a uma superfície adequada; agregação formando colônias; produção EPS formando um biofilme maduro e desprendimento do biofilme retornando a estilo de vida livre. Durante a fase de vida livre as bactérias movimentam-se livremente em um meio líquido utilizando flagelos. Os pilos são utilizados para aderência a uma superfície e também para movimentarem-se sobre ela associando-se a outras bactérias. O sistema *quorum sensing* (QS) pode induzir a produção de EPS. Adaptado de O'TOOLE *et al.* (2000).

2.1.3 Exopolissacarídeos

Os EPS secretados por bactérias, possuem um papel fundamental na estrutura do biofilme bacteriano e têm demonstrado relação com a virulência de algumas bactérias (COSTERTON *et al.*, 1999). Embora os EPS (alginato, goma xantana, goma curdlana, gelana, succinoglicana, entre outros) sejam os componentes da matriz extracelular mais estudados, existem uma grande variedade de EPS que permanecem não caracterizados (SUNTHERLAND, 2001). A celulose tem sido identificada como um dos EPS da matriz extracelular que é produzido por várias espécies de bactérias durante a formação de biofilmes (SPIERS *et al.*, 2003; ZOGAJ *et al.*, 2001).

2.1.4 Monitoramento da densidade populacional (*Quorum Sensing*)

Muitas espécies de bactérias utilizam sinais químicos para se comunicarem e coordenarem suas atividades. As bactérias Gram-negativas utilizam-se de um sistema de comunicação regulador denominado *quorum sensing*. Esta regulação acontece por meio da produção de uma molécula sinal: acil homoserina lactona (acil HSL), a qual confere a bactéria a habilidade de monitorar sua densidade populacional e controlar a expressão de diversas funções fisiológicas.

O sistema *quorum sensing* foi primeiramente observado em *Vibrio fischeri* envolvendo reguladores de transcrição. (FUQUA *et al.*, 1994; 1996). A *Vibrio fischeri* é uma bactéria marinha que vive em simbiose com certos tipos de peixes e lulas. Quando a bactéria atinge alta densidade populacional produz luminescência, como resultado de um processo de autoindução, mediado pela molécula N-3-(oxohexanoil) homoserina lactona (OHHL) (EBERHARD *et al.*, 1981).

A luminescência em *Vibrio fischeri* é regulada pelos produtos dos genes *luxR* e *luxI*. O gene *luxI* codifica a proteína LuxI (proteína homoserina lactona sintetase) que sintetiza a OHHL. Quando a densidade populacional é elevada ocorre maior produção de OHHL e, conseqüentemente, há o estímulo da transcrição do gene da proteína LuxI, num processo chamado de autoindução. A molécula OHHL se liga à proteína LuxR. A ligação da OHHL à proteína LuxR, provoca a mudança conformacional dessa proteína, permitindo que esta ligue-se à região do DNA chamada de lux box (FUQUA *et al.*, 1994; 1996), estimulando a expressão do operon *luxICDABEG*. A expressão do operon *luxICDABEG* leva a síntese de compostos luminescentes.

Dentre os fenômenos regulados pelo sistema *quorum sensing* incluem a produção de EPS por bactérias, tais como: *Ralstonia solanacearum*, *Pantoea stewartii* subsp. *stewartii*, *Pseudomonas aeruginosa*, *Sinorhizobium meliloti* (PUSKAS *et al.*, 1997; VON BODMAN *et al.*, 1998; GLESSNER *et al.*, 1999; PARSEK e GREENBERG, 2000; MARKETON *et al.*, 2003).

2.2 Celulose

As diferenças entre os polissacarídeos estruturais como a celulose e os armazenadores de energia como o amido e o glicogênio, são devidas ao tipo de ligações glicosídicas que ocorrem entre as moléculas de glicose que os formam. No glicogênio e amido (amilose e amilopectina) as ligações glicosídicas são do tipo $\alpha(1\rightarrow4)$ e $\alpha(1\rightarrow6)$ ou apenas $\alpha(1\rightarrow4)$, como no caso da amilose (VILLELA *et al.*, 1978).

Já a celulose é homopolissacarídeo neutro, formado por cadeias lineares não ramificadas de monossacarídeos de β -D-glicose, unidos por ligação $\beta(1\rightarrow4)$, e a unidade estrutural de repetição é a celobiose, formada pela união de

duas moléculas de glicose (BOBBIO e BOBBIO, 2001). Cada resíduo de glicose está rotacionado aproximadamente 180° em relação ao resíduo vizinho. A cadeia glicana pode ter de 2000 a mais de 25000 resíduos de glicoses (RAVEN *et al.*, 2001; BROWN, 1999). A Figura 2, mostra a estrutura da cadeia $\beta(1\rightarrow4)$ glicana denominada de celulose. A hidrólise total da celulose produz moléculas de glicose, enquanto que a hidrólise parcial produz moléculas de celobiose (RAVEN *et al.*, 2001; BROWN *et al.*, 1996).

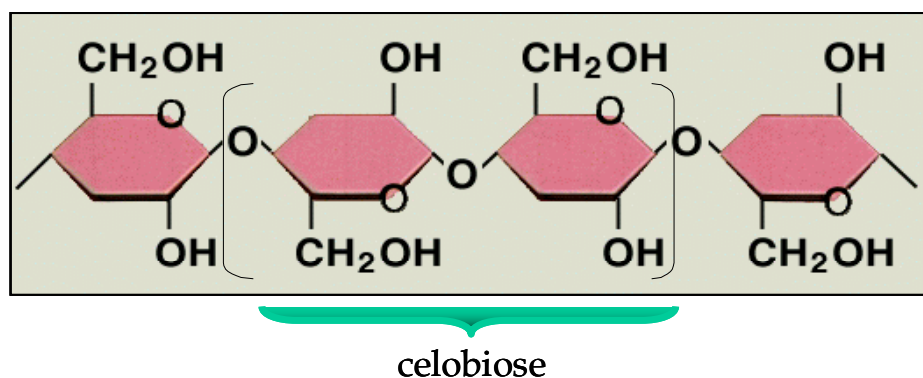


Figura 2 - Estrutura da cadeia $\beta(1\rightarrow4)$ glicana (celulose). A unidade de repetição celobiose está indicada entre colchetes. Adaptado de STEPHENS (2004).

Na Natureza, a celulose não existe como uma única cadeia. As moléculas longas e rígidas de celulose combinam-se para formar as microfibrilas, cada uma consistindo de várias cadeias de celulose. As cadeias de celulose que compõem a microfibrila estão orientadas paralelamente, formando ligações de hidrogênio intramolecular (RAVEN *et al.*, 2001). A Figura 3, ilustra os elementos que compõem a microfibrila de celulose da parede celular de plantas.

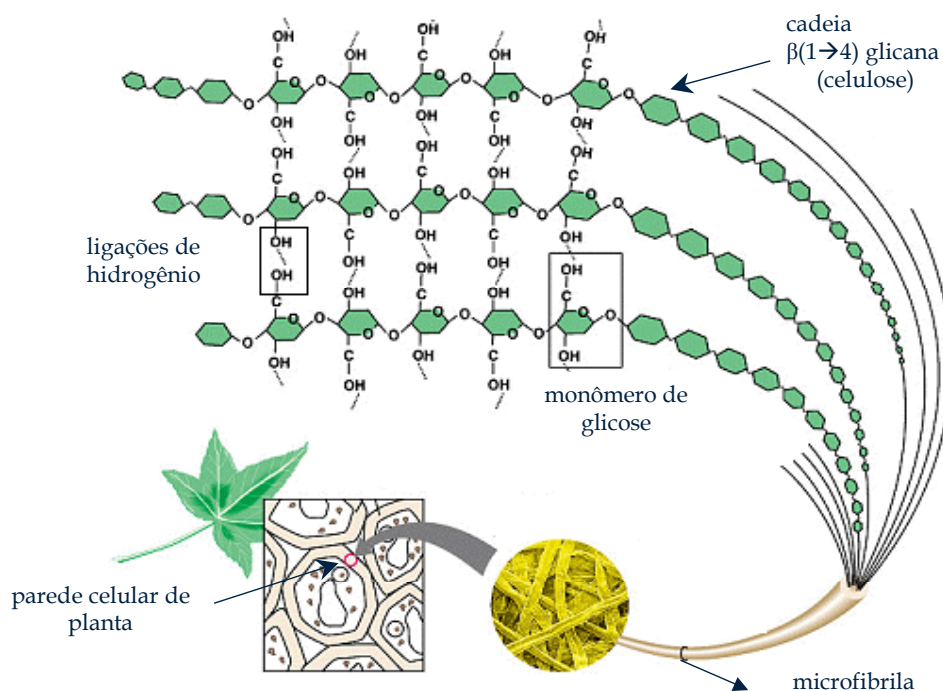


Figura 3 - Elementos que compõem a microfibrila de celulose da parede celular de planta. Adaptado de HJEMMESIDE (2004).

A microfibrila pode ter aproximadamente 36 cadeias de celulose, sendo que a celulose produzida por algas pode formar microfibrilas com mais de 1200 cadeias $\beta(1\rightarrow4)$ glicana. A estrutura da celulose tem sido identificada através da microscopia eletrônica, em membranas celulares de vários organismos. Complexos arranjados linearmente em uma única fileira ou múltiplas fileiras são observados em bactérias enquanto em plantas, é observada uma estrutura sextavada denominada roseta, conforme pode ser visualizado na Figura 4 (DELMER, 1999).

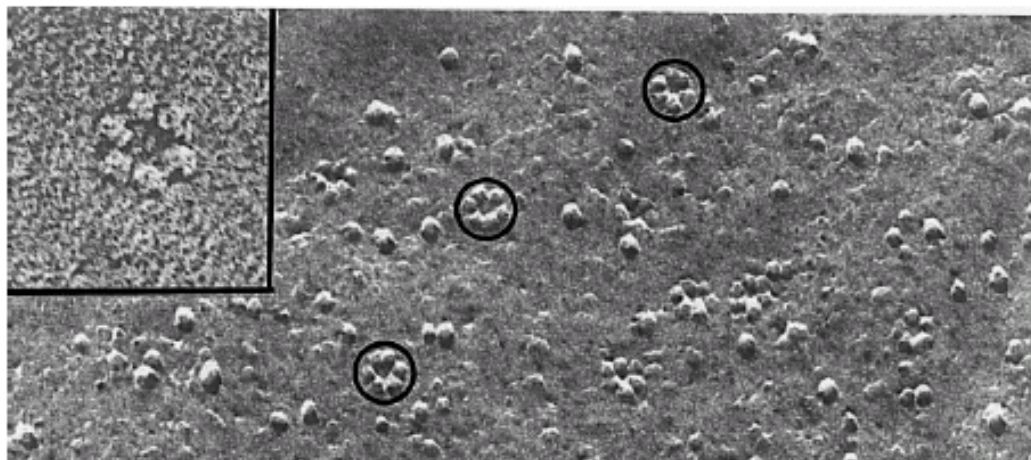


Figura 4 – Estrutura roseta associada à biogênese da microfibrila da celulose vegetal (aumento de 222.000 vezes). No canto superior esquerdo amplificação de uma das rosetas (aumento de 504.545 vezes). Obtida de DELMER (1999). Créditos: de Mark J. Grimson e Candace H. Haigler, Departamento de Ciências Biológicas, Texas Tech University, Lubbock, Texas.

Um modelo para a estrutura roseta apresentado por DOBLIN (2002) é mostrado na Figura 5. O modelo propõe uma estrutura formada por seis subunidades catalíticas de celulose sintase (BcsA) interagindo para formar uma subunidade roseta. E seis subunidades rosetas, sintetizando seis cadeias $\beta(1\rightarrow4)$ glicana interagem para formar a microfibrila.

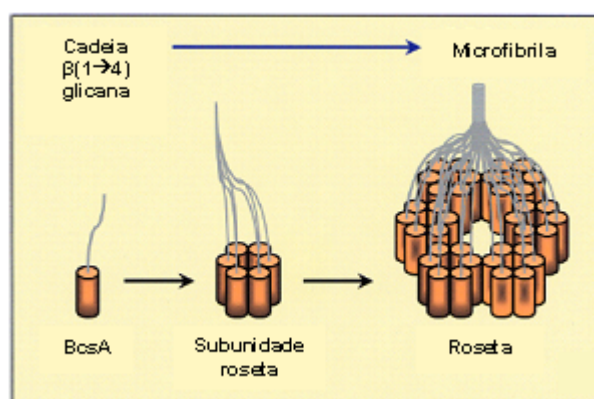


Figura 5 – Modelo para a estrutura roseta mostrando a formação de microfibrila de celulose. BcsA representa a celulose sintase. A subunidade roseta é formada por seis celulosos sintase e a roseta é composta por seis subunidades rosetas dando origem a microfibrila. Modificado de DOBLIN (2002).

2.3 Celulose Bacteriana

Embora a produção de celulose sintetizada por bactéria já venha sendo estudada desde a década de 30, avanços genéticos moleculares só aconteceram a partir de 1958, quando GLASER (1958) identificou o envolvimento da UDP-glicose envolvida na biossíntese de celulose. Estudos intensivos da produção de celulose bacteriana, usando a bactéria *A. xylinum* como modelo, iniciaram a partir do descoberta por ROSS *et al.* (1987) do regulador da biossíntese de celulose, diguanilmonofosfato cíclico (c-di-GMP). Em 1989, LIN e colaboradores identificaram três polipeptídeos com massa molecular de 83, 93 e 97 kDa, em um sistema de produção de celulose *in vitro* por *A. xylinum*. Estes autores sugeriram o possível envolvimento destes polipeptídeos na polimerização da glicose em celulose. Em 1990, estes resultados foram confirmados por LIN *et al.*, (1990), através da identificação do polipeptídeo de 83 kDa, como sendo a subunidade catalítica da celulose sintase. WONG *et al.* (1990) em estudos paralelos fizeram a mesma observação. O isolamento e seqüenciamento do gene da celulose sintase foi realizado por SAXENA *et al.* (1990). Em 1995, SAXENA e colaboradores publicaram o primeiro trabalho propondo uma arquitetura multidomínio para as β -glicosiltransferases, classe de enzimas à qual pertence a celulose sintase.

2.3.1 Aplicações da celulose bacteriana

A aplicabilidade da celulose bacteriana é tão extensa que, no período de 1990 a 1996, mais de 50 patentes foram registradas (JONAS e FARAH, 1998), sendo a mesma apontada como a próxima *commodity* bioquímica a ser comercializada. Suas características peculiares abrangem um amplo campo de aplicações: na indústria têxtil; indústria alimentícia como espessante; em empresas de aparelhos eletroacústicos utilizados como diafragma em fones de ouvido (Sony Corporation); na medicina como substituto temporário da pele

humana. Como substituto temporário da pele a celulose bacteriana tem sido comercializada com a marca BioFill® (Ver Figura 6), produto patenteado que promete ser revolucionário em processos de recuperação de lesões, queimaduras e ferimentos de pele (JONAS e FARAH, 1998). O uso da película de celulose tem sido investigado recentemente em outros procedimentos da área médica, tais como: na cardiologia como revestimento de *stent*, uma pequena malha metálica introduzida na artéria durante a angioplastia para impedir que a artéria volte a se fechar; na neurologia como substituição da dura-máter em casos de traumas e tumor; na odontologia para recuperação de tecido periodontal; e até mesmo para o desenvolvimento de materiais para a fabricação de placas blindadas e compostos utilizados em coletes à prova de balas (ERENO, 2004).

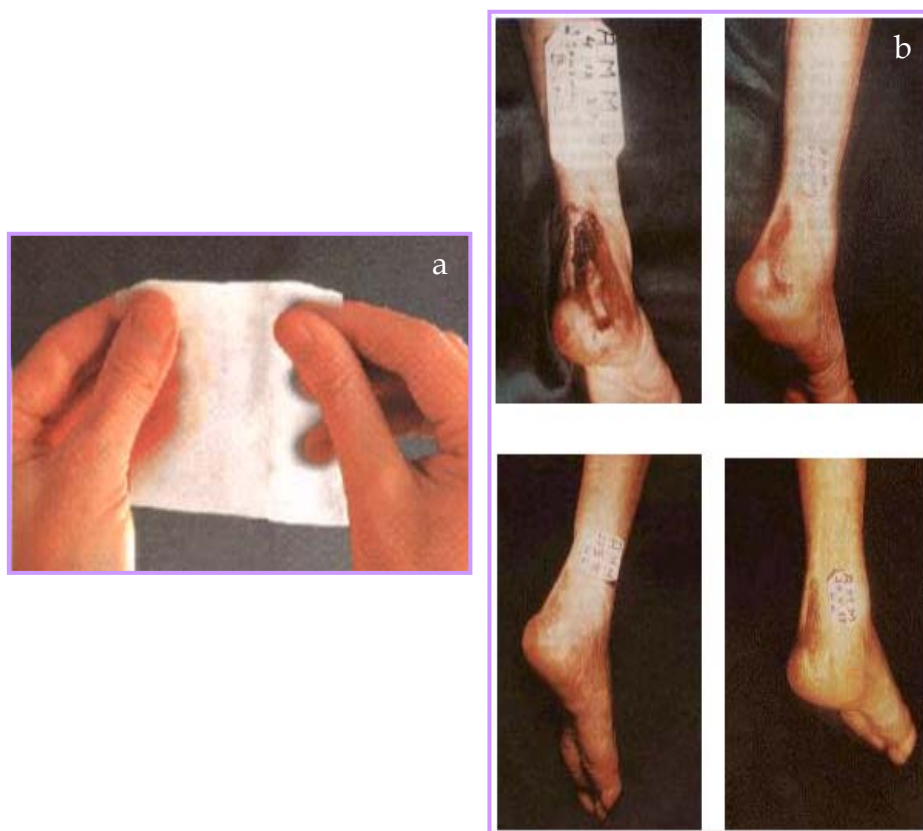


Figura 6 - (a) Aplicações da celulose bacteriana como pele artificial - BioFill® - Curativo biológico. (b) Tratamento de paciente com úlcera isquêmica e lesão necrótica de calcâneo e área do tendão de Aquiles como um novo substituto da pele. Úlcera fechada após três meses do início do tratamento. Figuras obtidas em BIOFILL (2004).

2.3.2 Biossíntese de celulose em bactérias

A biossíntese de celulose em bactérias é um processo que ocorre através de várias etapas, envolvendo enzimas individuais e um operon denominado *bcs* (*bacterial cellulose synthesis*). Este operon codifica proteínas e enzimas que compõem uma associação para a polimerização da glicose e a formação de celulose, sendo conservado entre várias espécies de bactérias Gram-negativas.

A via biossintética da celulose é mais compreendida na bactéria *A. xylinum*. A via de conversão do substrato glicose ao produto celulose envolve várias reações, onde a glicose é primeiramente fosforilada em glicose-6-fosfato (G6P) pela enzima glicocinase com a participação de uma molécula de adenosina trifosfato (ATP) e liberado uma molécula de adenosina difosfato (ADP); numa segunda reação a G6P é convertida a glicose-1-fosfato (G1P) pela enzima fosfoglicomutase; G1P, através de uma reação de isomerilação catalisada pela enzima UDP-glicose-pirofosforilase é convertida em uridina difosfato glicose (UDP-glicose), utilizando uma molécula de uridina trifosfato (UTP) e liberando uma molécula de pirofosfato (PPi). A molécula UDP-glicose é, finalmente, a precursora da biossíntese da celulose. A polimerização de glicose em cadeias $\beta(1\rightarrow4)$ glicana é realizada pela celulose sintase. Associação de várias cadeias nascentes formam as fibras de celulose (ROSS *et al.*, 1990, BROWN *et al.* 1996). A Figura 7 apresenta a via de biossíntese de celulose em *A. xylinum* a partir da glicose como fonte de carbono, conforme proposto por ROSS *et al.* (1991).

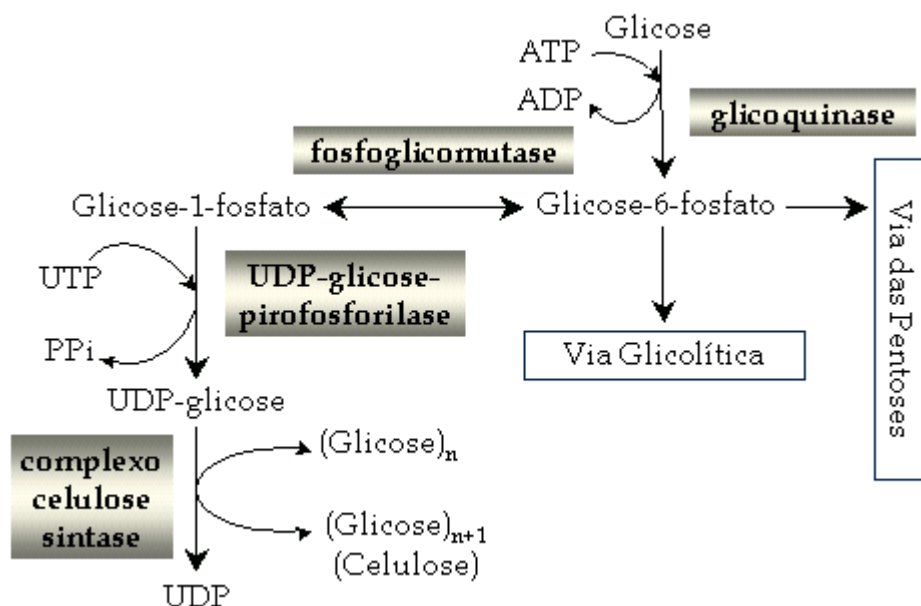


Figura 7 - Via de biossíntese da celulose em *A. xylinum* a partir da glicose como fonte de carbono, conforme proposto por ROSS *et al.* (1991).

A enzima celulose sintase em *A. xylinum* é ativada alostericamente pela molécula c-di-GMP, conforme pode ser observado no modelo de regulação e biossíntese de celulose bacteriana apresentado na Figura 8. O ativador c-di-GMP é sintetizado pela enzima diguanilato ciclase a partir de duas moléculas de guanosina trifosfato (GTP). A concentração de c-di-GMP intracelular é controlada pela ação oposta da enzima fosfodiesterase que degrada a molécula ativadora c-di-GMP. Os genes que codificam as enzimas que sintetizam e degradam c-di-GMP em *A. xylinum* estão organizados em um operon contendo três diguanilato ciclase e três fosfodiesterase (CHANG *et al.*, 2001). Tanto a diguanilato ciclase como a fosfodiesterase possuem em sua estrutura primária o domínio GGDEF. Tem sido demonstrado que proteínas que possuem o domínio GGDEF possuem atividade diguanilato ciclase (AUSMEES *et al.*, 2001; PEI e GRISHIN, 2001; GALPERIN *et al.*, 2001; 2004; PAUL *et al.*, 2004).

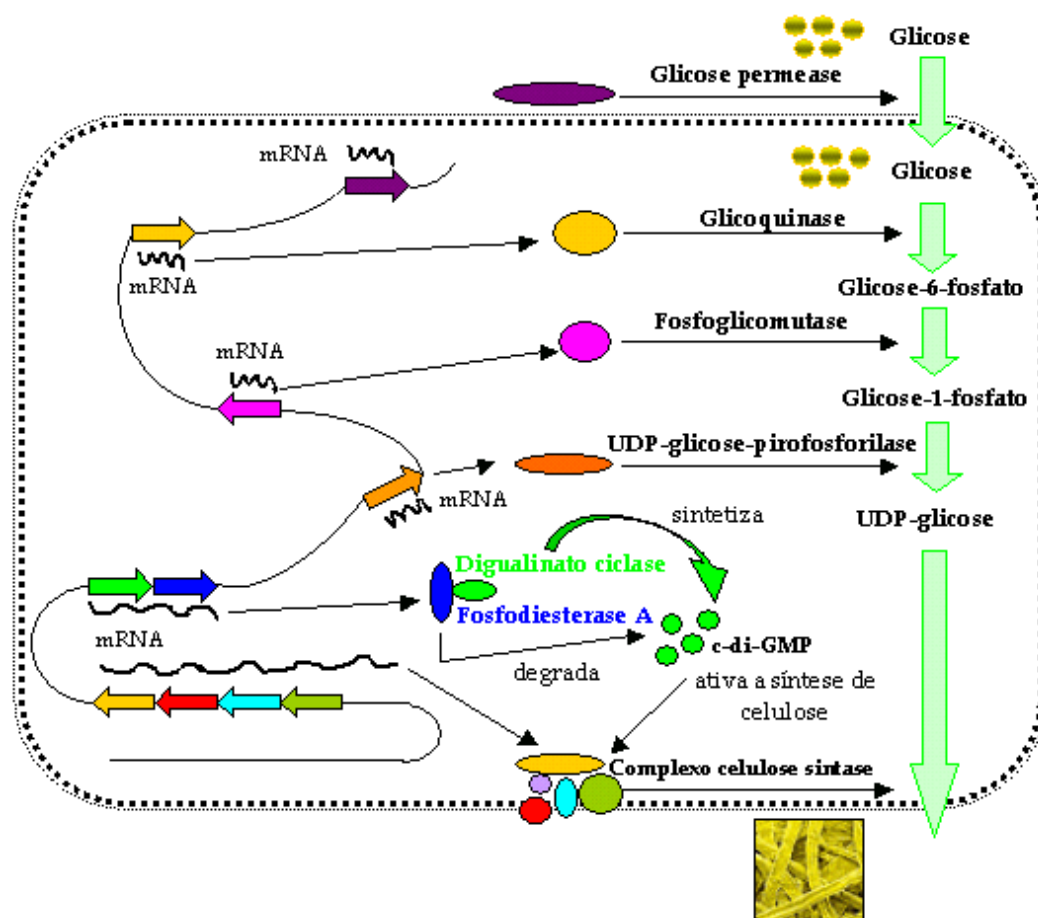


Figura 8 - Modelo de regulação e biossíntese de celulose em *Acetobacter xylinum* a partir do substrato glicose (ROSS *et al.*, 1987; 1990; 1991; SKINNER e CANNON, 2000; SHODA & ANO Laboratory, 2004).

2.3.3 Principais proteínas do complexo celulose sintase em bactérias

A seguir são apresentadas as características estruturais das proteínas e enzimas que compõem o complexo responsável pela polimerização da cadeia $\beta(1\rightarrow4)$ glicana. Estas proteínas e enzimas são codificadas pelos genes que compõem o operon *bcs* em diversas espécies bacterianas filogeneticamente relacionadas. Elas são as proteínas designadas como: YhjQ, BcsA, BcsB, BcsZ, BcsC.

2.3.3.1 Proteína YhjQ

O gene que codifica a proteína YhjQ em *S. typhimurium* LT2 e *E. coli* K12 está localizado na região a montante do operon *bcs*. Esta proteína tem sido descrita como uma ATPase envolvida na divisão do cromossomo. Ensaio realizados por KIM *et al.* (2002) de inativação desse gene nessas bactérias, mostraram que causa retardo na divisão celular e produz células filamentosas de vários tamanhos. Esses autores sugerem que essa proteína pode afetar a duplicação e morfologia celular.

2.3.3.2 Proteína BcsA: celulose sintase

A proteína BcsA, celulose sintase, é a subunidade catalítica, pertencendo à família 2 das β -glicosiltransferases processivas. Glicosiltransferases são enzimas que catalisam a transferência de resíduos de açúcar para moléculasceptoras específicas, formando ligações glicosídicas. As glicosiltransferases podem agir como enzimas processivas ou não-processivas. As glicosiltransferases processivas que freqüentemente referem-se às enzimas que sintetizam polissacarídeos, catalisam a adição de múltiplos resíduos de açúcar na extremidade crescente de uma cadeia linear de polissacarídeos. As glicosiltransferases não-processivas, por outro lado, catalisam a adição de um único resíduo de açúcar em posições específicas de uma molécula acceptora (HENRISSAT e DAVIES, 2000; SAXENA *et al.*, 2001).

SAXENA *et al.* (1995) propuseram um modelo da arquitetura multidomínio de β -glicosiltransferase, sugerindo dois tipos de domínios, A e B. No domínio do tipo A estão inseridos dois resíduos catalíticos de aspartato (D,D), e no domínio do tipo B encontra-se um resíduo catalítico de aspartato (D) e o motivo conservado glutamina, dois resíduos de aminoácidos quaisquer,

arginina e triptofano (QXXRW). Nas β -glicosiltransferases processivas encontram-se ambos os domínios (A e B), enquanto nas não-processivas encontram-se somente o domínio do tipo B. A troca dos resíduos de aspartato (D,D) por outros quaisquer aminoácidos implica na perda da atividade da celulose sintase em *A. xylinum*, o que sugere que esses resíduos são essenciais para a atividade catalítica da enzima. Já o motivo QXXRW está relacionado com o comprimento da cadeia polimérica. O motivo QRXRW sintetiza cadeias longas de polissacarídeos e está presente na celulose e na quitina sintase, enquanto que o motivo QQXRW sintetiza oligossacarídeos de cadeia curta. A análise da seqüência de aminoácidos das β -glicosiltransferases realizada por SAXENA *et al.* (1995), mostra que a distância entre o resíduo D e o motivo QXXRW do domínio B é praticamente constante, variando de 33 a 38 resíduos de aminoácidos; a distância entre os resíduos DD do domínio do tipo A é de 46 resíduos de aminoácidos para a celulose sintase e de 120 resíduos de aminoácidos para a quitina sintase; já a distância entre o resíduo D do domínio do tipo A e o resíduo D do domínio do tipo B varia de 92 a 138 resíduos de aminoácidos. Baseando-se nesta análise, os proponentes do modelo sugeriram um motivo comum entre as β -glicosiltransferases, como sendo: D,D,D35Q(R,Q)XRW. Este motivo pode ser utilizado para prever uma β -glicosiltransferase.

BROWN *et al.* (1996) realizaram alinhamento entre seqüências de aminoácidos de celulose sintases de diferentes linhagens de *A. xylinum* e observaram que o número de resíduos de aminoácidos entre os três resíduos de ácido aspártico (D,D,D) e o motivo conservado QRXRW nessas enzimas são de 46, 96 e 35 resíduos, respectivamente, conforme pode ser visto na Figura 9.

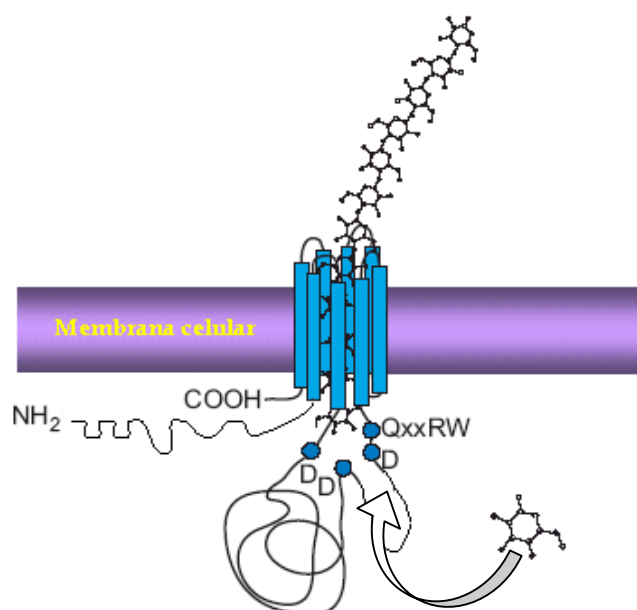


Figura 10 - Estrutura tridimensional hipotética da proteína BcsA. As hélices transmembrana formam um poro por onde passa a cadeia $\beta(1\rightarrow4)$ glicana que está sendo polimerizada. A figura foi adaptada de DELMER (1999) e RICHMOND (2000).

2.3.3.3 Proteína BcsB: proteína de ligação de c-di-GMP

A proteína BcsB está firmemente associada à subunidade catalítica celulose sintase, proteína BcsA. Em *A. xylinum*, linhagem P21877, as proteínas BcsA e BcsB estão codificadas num mesmo gene que codifica a proteína AcsAB.

A proteína BcsB é considerada a subunidade de regulação da celulose sintase por ligar-se ao ativador alostérico da síntese de celulose, c-di-GMP, o qual foi identificado primeiramente em *A. xylinum*. Nessa bactéria o papel do c-di-GMP livre na célula é ativar alostericamente a proteína BcsA - celulose sintase. Acredita-se que a maioria do c-di-GMP celular está reversivelmente ligado à proteína BcsB. Como a proteína BcsB está associada à proteína BcsA formando um complexo multienzimático - inseridos na membrana celular - é possível que esta proximidade espacial seja necessária para liberar c-di-GMP diretamente da proteína BcsB para a proteína BcsA (RÖMLING, 2002). É

importante salientar que o nível celular de c-di-GMP depende do equilíbrio entre a atividade da enzima diguanilato ciclase, que sintetiza c-di-GMP a partir de duas moléculas de GTP, e a ação oposta da enzima fosfodiesterase, que degrada c-di-GMP (CHANG *et al.*, 2001).

2.3.3.4 Proteína BcsZ: celulase

A proteína BcsZ pertence à família 8 das glicosil hidrolases e possui atividade endoglicanase, mais comumente conhecida como celulase. As celulases são enzimas que catalisam a hidrólise de ligações internas $\beta(1\rightarrow4)$ -D-glicosídicas da celulose. Podem hidrolisar também ligações $\beta(1\rightarrow4)$ em D-glicanas que contenham ligações $\beta(1\rightarrow3)$. Seus substratos naturais são a celulose e a xiloglicana (possui em sua cadeia principal ligações $\beta(1\rightarrow4)$ D-glicosídicas). Além disso apresenta especificidade variável sobre carboximetil celulose (CMC) (derivada da celulose com estrutura $\beta(1\rightarrow4)$ D-glicopiranoose), Avicel (celulose cristalina com ligações $\beta(1\rightarrow4)$ linear), β -glicana (polissacarídeo linear não ramificado com ligações $\beta(1\rightarrow3)$ e $\beta(1\rightarrow4)$ D-glicopiranoose) e xilana (polissacarídeos mais comum entre as hemiceluloses com ligações $\beta(1\rightarrow4)$ D-xilopiranosil) (WULFF, 2002; HENRISSAT *et al.*, 1989).

A função da celulase no operon *bcs* ainda não está completamente esclarecida. Estudo realizado por PENG *et al.* (2002) observaram em plantas a associação da atividade da celulase com a biossíntese da celulose uma vez que um sitosterol- β -glicosídeo serviria como iniciador para a polimerização da cadeia $\beta(1\rightarrow4)$ glicana, sendo que uma celulase poderia participar na clivagem do sitosterol- β -glicosídeo possibilitando o alongamento do polímero.

No entanto, procariotos não possuem esteróis e, portanto, podem ou não usar iniciador para síntese de celulose, ou usar iniciadores com outra função orgânica, diferente do sitosterol (DOBLIN, 2002). Em *A. tumefaciens*, conforme

mostrado na Figura 11, uma celulase está envolvida na clivagem de um oligossacarídeo ligado a um lipídeo que será utilizado como iniciador na biossíntese de celulose (MATTHYSSE *et al.*, 1995).

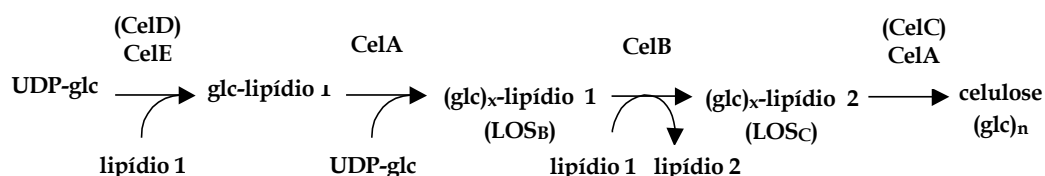


Figura 11 - Modelo da função da celulase na biossíntese de celulose apresentado por MATTHYSSE *et al.* (1995). O modelo sugere a transferência de oligossacarídeo ligado a lipídeo para o crescimento da cadeia $\beta(1\rightarrow4)$ glicana, após ser clivado pela celulase.

2.3.3.5 Proteína BcsC

A proteína BcsC está envolvida na biossíntese de celulose em várias espécies bacterianas e sua função ainda permanece desconhecida. As proteínas BcsC possuem pelo menos um TPR (*tetratricopeptide repeat*). O motivo TPR é um motivo estrutural, identificado por uma seqüência de 34 aminoácidos presentes em uma variedade de proteínas. Podem ocorrer de 3 a 16 motivos na proteína, podendo estar dispersos ao longo da seqüência. Acredita-se que formam a estrutura envolvida na interação proteína-proteína, e freqüentemente atua para manter unido um complexo protéico (DAS *et al.*, 1998).

2.3.4 Produção de celulose bacteriana

Como salientado anteriormente, a razão pelo qual os microrganismos produzem celulose ainda mantém-se desconhecida, porém existem algumas hipóteses. Acredita-se que bactérias aeróbicas produzam este biofilme para manter-se na superfície líquida. Outra suposição é a de que as bactérias produzam o biofilme como uma estratégia de sobrevivência, para proteger-se

dos raios ultravioletas, de outros organismos e de metais-pesado. Além disso, no biofilme os nutrientes podem ser supridos facilmente por difusão.

As células produtoras de celulose estão envolvidas pela matriz polimérica extracelular que suporta a população bacteriana na superfície líquida. A matriz do polímero participa na adesão das células a uma superfície sólida ou líquida. Por causa de suas propriedades absorventes, facilita a acessibilidade aos nutrientes, uma vez que sua concentração na rede polimérica é maior do que no meio líquido circundante (ROSS *et al.*, 1991).

A produção de celulose por *A. xylinum* tem sido realizada principalmente em culturas estáticas. No entanto, para se obter aumento na produtividade, estudos têm sido desenvolvidos envolvendo culturas agitadas, que podem ser realizadas em um reator do tipo RDF (*rotating disc fermentor*), ou em um reator de fluxo contínuo "*airlift*", como os apresentados na Figura 12 (CHENG *et al.*, 2002; SKINNER e CANNON, 2000).

2.3.5 Proteínas com motivo GGDEF

O domínio conservado denominado GGDEF existe em várias proteínas de procariotos, freqüentemente envolvidas com componentes regulatórios sensoriais (AUSMEES *et al.*, 2001). O domínio GGDEF foi identificado primeiramente na bactéria *Caulobacter crescentus* em uma proteína de regulação de resposta, PleD, que controla a diferenciação na transição celular, causando perda de mobilidade e fixação em uma superfície formando uma espécie de talo (HECHT e NEWTON, 1995). Alguns indícios sobre o domínio GGDEF surgiram de estudo da regulação da biossíntese da celulose em *A. xylinum*. Observou-se que proteínas possuindo o domínio GGDEF apresentam atividade diguanilato ciclase (AUSMEES *et al.*, 2001). Os sistemas que são mediados por estes reguladores de resposta incluem os responsáveis pela interação das bactérias

com o ambiente (fímbria, proteínas extracelulares, virulência) e a formação de biofilme (GALPERIN *et al.*, 2001).

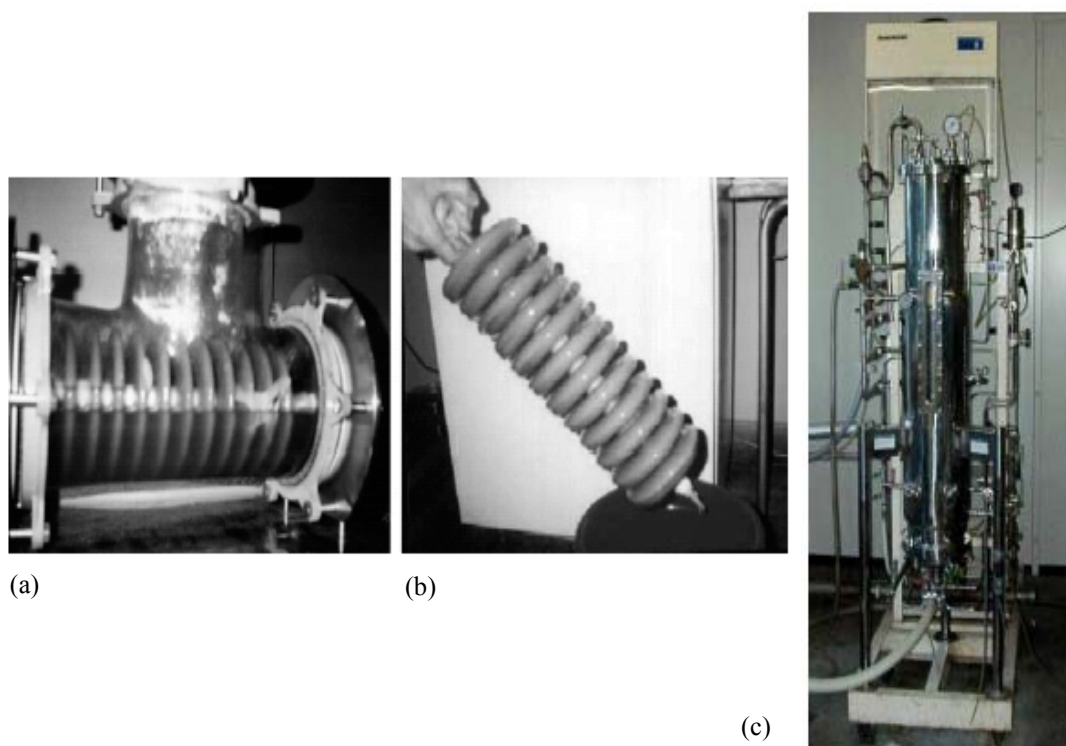


Figura 12 - Reatores para produção de celulose bacteriana. (a) e (b) reatores do tipo RDF (*rotating disc fermentor*) (KRYSTYNOWICZ *et al.*, 2002) e em (c) um reator do tipo *airlift* (SHODA & ANO Laboratory, 2004).

2.4 Microrganismo Modelo: *Chromobacterium violaceum*

A bactéria *Chromobacterium violaceum*, microrganismo modelo utilizado neste estudo, é um organismo particularmente interessante, uma vez que é β -proteobactéria, Gram-negativa que habita solo e água de uma variedade de regiões tropicais e subtropicais e é encontrada em abundância nas águas do Rio Negro, região da Amazônia Brasileira (DURÁN e MENCK, 2001). A análise de seu genoma deve fornecer informações importantes sobre adaptação fisiológica de organismos de vida livre, porém, patógeno oportunista.

A bactéria *C. violaceum* possui a característica marcante de produzir um pigmento violeta denominado violaceína que tem demonstrado o seu potencial como atividade antibiótica, antitumoral, antiparasitária e antiviral (ANTÔNIO, 1994; DE SOUZA *et al.*, 2002; DURÁN *et al.*, 1994; LEON *et al.*, 2001; DURÁN e MENCK, 2001; ANDRIGHETTI-FROHNER *et al.*, 2003). Além disso, *C. violaceum* possui a capacidade de biossíntese do homopolímero de 3-hidroxicinovelato (polihidroxicinovelato ou PHV), bem como outros tipos de polihidroxicinovelatos (PHA's) (STEINBÜCHEL *et al.*, 1993; PIEMOLINI *et al.*, 2003).

A formação de biofilme por esta bactéria tem sido pouco explorada. Na década de 60 nos EUA, alguns estudos relataram a formação de EPS por uma linhagem de *C. violaceum* obtida do solo (MARTIN e RICHARDS, 1963; CORPE, 1964). Recentemente, tem sido reportada a formação de biofilme pela linhagem ATCC 31532 envolvendo o sistema *quorum sensing* de comunicação intercelular (MARTINELLI *et al.*, 2002; MARTINELLI *et al.*, 2004).

O seqüenciamento do genoma da *C. violaceum* (linhagem ATCC 12472) realizado por um projeto brasileiro apresentou as ORFs CV2675, CV2677 e CV2678 envolvidas na biossíntese de celulose, (BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM, 2003).

2.5 Elementos da Transcrição em Procariotos

2.5.1 Elementos estruturais da seqüência do DNA e RNA

O DNA (ácido desoxirribonucléico) é uma molécula polimérica não ramificada, constituída de uma dupla hélice de longas cadeias de nucleotídeos unidos por ligações fosfodiéster. As cadeias de nucleotídeos são antiparalelas, interconectadas por ligações de hidrogênio que se formam entre bases

complementares, púricas e pirimidínicas. Quatro tipos de nucleotídeos se apresentam ao longo da cadeia, de acordo com a base nitrogenada que possuem: purinas (adenina (A), guanina (G)) e pirimidinas (timina (T), citosina (C)) (CHAMPE e HARVEY, 2002).

O DNA contém o mapa da vida de todo organismo. Toda a informação genética está contida na seqüência de ácidos nucléicos do DNA; porém, é a partir de uma cópia do DNA em um ácido ribonucléico (RNA) que estas informações serão expressas. O RNA difere do DNA em alguns aspectos: são consideravelmente menores, contém uma pentose (ribose) e uracil (U) ao invés de timina (T). Três tipos de RNA participam da síntese de proteínas: RNA de transferência (tRNA), RNA ribossômico (rRNA), e RNA mensageiro (mRNA). O processo de expressão da informação genética envolve a transcrição de certas regiões do DNA (genes) em seqüências de aminoácidos (cadeias polipeptídicas) formando as proteínas. O fluxo da informação do DNA ao RNA e à proteína estabelece o chamado “dogma central da biologia molecular” (LEHNINGER *et al.*, 2000; LEWIN 1995).

2.5.2 O Código Genético

O código genético contém a correspondência entre uma seqüência de bases nitrogenadas e um determinado aminoácido, convertendo a informação hereditária dos genes em proteínas. As quatro bases nitrogenadas (adenina (A), guanina (G), citosina (C) e uracil (U)) combinam-se de forma diferente em uma seqüência de três bases nitrogenadas adjacentes no DNA ou mRNA correspondendo a um aminoácido específico na proteína. A esta seqüência de bases nitrogenadas dá-se o nome de códon. Dessa forma existem sessenta e quatro combinações diferentes de bases nitrogenadas, embora existam apenas vinte tipos de aminoácidos para serem representados (ver Tabela 1) (BONATO, 2004).

2.5.3 Quadro Aberto de Leitura

Um quadro aberto de leitura (ORF) consiste em uma série de códons adjacentes que especifica uma seqüência de aminoácido com possibilidade de ser um gene codificando para uma proteína. Uma ORF começa com um códon de início - geralmente (mas nem sempre) ATG - e finaliza com um códon de término: TAA, TAG ou TGA (BROWN, 2002; KANEHISA, 2000).

1ª posição	2ª posição								3ª posição
	U	C	A	G					
U	UUU	Phe(F) Fenilalanina (aromático)	UCU	Ser(S) Serina	UAU	Tyr(Y) Tirosina (aromático)	UGU	Cys(C) Cisteína	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu(L) Leucina	UCA		UAA	TERM	UGA	TERM	A
	UUG		UCG		UAG		UGG	Trp(W) Tryptofano (aromático)	G
C	CUU	Leu(L) Leucina	CCU	Pro(P) Prolina	CAU	His(H) Histidina	CGU	Arg(R) Arginina	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln(Q) Glutamina	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile(I) Isoleucina	ACU	Thr(T) Treonina	AAU	Asn(N) Asparagina	AGU	Ser(S) Serina	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys(K) Lisina	AGA	Arg(R) Arginina	A
	AUG	Met(M) Metionina	ACG		AAG		AGG		G
G	GUU	Val(V) Valina	GCU	Ala(A) Alanina	GAU	Asp(D) Aspartato	GGU	Gly(G) Glicina	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu(E) Glutamato	GGA		A
	GUG		GCG		GAG		GGG		G

Verde: aminoácidos não polares (hidrofóbicos); **vermelho:** aminoácidos polares não carregados; **azul:** aminoácidos polares com carga positiva (básicos); **amarelo:** aminoácidos polares com carga negativa (ácidos); **verde claro,** aminoácido não polar: Metionina (codificado pelo códon de início da tradução); **rosa:** códons de término da tradução. Modificado de BONATO (2004).

2.5.4 Características estruturais da transcrição e tradução

A transcrição dos genes em procariotos acontece de forma diferenciada dos organismos eucariotos. Em bactérias, só uma RNA polimerase sintetiza todos os RNAs, exceto os iniciadores de RNA. A RNA polimerase é formada por duas subunidades alfa, uma beta e uma beta' e uma subunidade sigma, também conhecida como fator sigma. As subunidades alfa e beta formam a enzima central. A polimerase completa inclui o fator sigma e esse conjunto é denominado de holoenzima. O fator sigma está presente no momento da interação da RNA polimerase com o DNA, e assim que inicia a transcrição ele é liberado, permanecendo somente a enzima central para a realização do alongamento da molécula (CHAMPE e HARVEY, 2002; LEHNINGER *et al.*, 2000; LEWIN, 1995).

O fator de transcrição sigma reconhece o promotor, em bactérias. Diferentes fatores sigmas reconhecem diferentes promotores. Em *Escherichia coli*, por exemplo, pelo menos seis diferentes fatores sigmas são sintetizados. O mais importante em *E. coli* é o sigma 70 que está envolvido no metabolismo geral da célula. Outras bactérias possuem fator sigma 70 homólogo ao de *E. coli*. Uma região promotora denominada box -10, também conhecida como Pribnow Box, e outra de box -35 são regiões de ligação com o DNA pelo fator sigma 70 de *E. coli*. Observa-se que estas regiões também estão presentes em outras bactérias. Apesar destas seqüências não serem perfeitamente idênticas entre os diferentes organismos, alguns nucleotídios são encontrados com maior freqüência do que outros, numa determinada posição, constituindo o que se denomina seqüência consenso. Dois consensos, com 6 nucleotídios cada, foram determinados a partir desse tipo de análise. A seqüência consenso -10 é TATAAT. A seqüência consenso -35 é TTGACA. Os consensos estão separados

entre si por cerca de 16-18 nucleotídios em 90% dos casos analisados (CHAMPE e HARVEY, 2002; LEHNINGER *et al.*, 2000; LEWIN, 1995).

Em especial, na *E. coli*, a transcrição envolve as etapas de iniciação, alongamento e término. Na etapa de iniciação, a RNA polimerase se liga reversivelmente ao DNA na região promotora, boxes -10 e -35. Esta ligação envolve contatos sobre uma face do DNA entre as posições -55 e -5. Através de experiências denominadas *footprint*, esses resultados podem ser demonstrados. A RNA polimerase, após se ligar à região promotora, altera a sua conformação estrutural, e passa a cobrir regiões entre -55 e +20. A base nitrogenada em que geralmente ocorre o início da transcrição é a adenina (A) e a menos comum é a citosina (C), e localiza-se aproximadamente de 5 a 9 bases à jusante do box -10. Assim que são transcritos 7 a 12 nucleotídeos, o fator sigma é liberado e a transcrição continua num processo contínuo de alongamento até o término (LEWIN, 1995).

Em bactérias a transcrição e a tradução ocorrem simultaneamente. Tão logo o RNA esteja sendo sintetizado pela RNA polimerase, ribossomos já estão se ligando para iniciarem a tradução. O mesmo transcrito poderá ser retranscrito várias vezes durante o ciclo celular. Uma outra RNA polimerase pode reiniciar a transcrição a partir do promotor, assim que esse transcrito for liberado do DNA (LEWIN, 1995).

O processo de alongamento da cadeia continua até que um sinal de terminação seja atingido, ou pode ocorrer espontaneamente. O sinal de terminação pode ser constituído por uma estrutura secundária em forma de “grampo de cabelo ou haste-alça”, em inglês denominado *hairpin*, seguida por uma seqüência de aproximadamente 8 resíduos U, que ocorre após o códon de parada. A estrutura *hairpin* é formada por uma região rica em G e C, aproximadamente 7 a 20 bases, que exibe simetria dupla pela presença de um

palíndromo. A formação do grampo leva a RNA polimerase a diminuir a sua velocidade ou até a interromper a síntese. A interação U-A é fraca e, portanto, a seqüência de Us permite a RNA polimerase se dissociar do DNA molde. Esse tipo de terminação é denominado terminação rô-independente (CHAMPE e HARVEY, 2002; LEHNINGER *et al.*, 2000; LEWIN, 1995).

Outro tipo de terminação é chamado de rô-dependente, que requer a participação de uma proteína especial chamada fator rô. Em terminadores rô-dependentes também é encontrada uma região de simetria dupla, rica em C e fraca em G e, portanto apresenta baixa complementaridade entre as bases, formando um grampo "fraco" no RNA. Além disto, não existe uma seqüência rica em Us após o grampo. Neste caso, torna-se necessária à participação da proteína auxiliar (fator rô) para a liberação da RNA polimerase (CHAMPE e HARVEY, 2002; LEHNINGER *et al.*, 2000; LEWIN, 1995).

2.5.5 Tradução do mRNA em proteína

Assim como acontece na transcrição, a tradução também envolve três etapas: iniciação, alongamento e terminação. A iniciação inclui a reunião dos componentes envolvidos na tradução: as duas subunidades ribossômicas, o mRNA a ser traduzido, e o tRNA com o aminoácido específico para o primeiro códon. Envolve também GTP, que fornece energia para o processo, e fatores de iniciação, facilitadores do processo. O ribossomo se liga à seqüência de ligação do ribossomo RBS (*Ribosome Binding Site*), também conhecida como Shine-Dalgarno (SD), localizada de 6 a 10 bases à montante do códon de início. Em seguida, um tRNA iniciador especial reconhece o códon de início, e o alongamento da cadeia polipeptídica promove a adição de aminoácidos na extremidade carboxila da cadeia. Ao formar a ligação peptídica, o ribossomo avança três nucleotídeos em direção ao terminal 3' do mRNA. A terminação da tradução ocorre quando o códon de parada é reconhecido (CHAMPE e

HARVEY, 2002). A Figura 13 ilustra a estrutura da seqüência de DNA, o processo de transcrição de um gene, e a tradução em proteína.

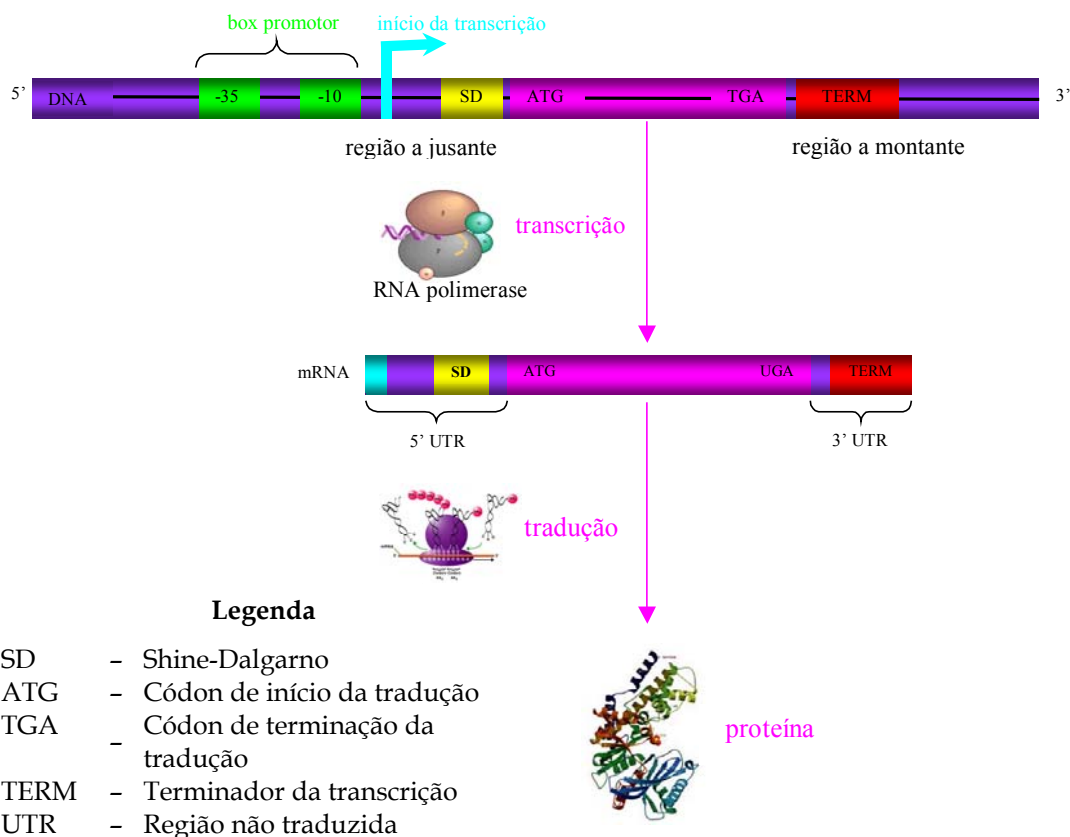


Figura 13 - Elementos estruturais da seqüência de DNA e outros elementos envolvidos na transcrição do DNA para RNA e tradução em proteína.

2.5.6 Família de proteínas, domínios e motivos

Um grupo de proteínas com similaridade na seqüência de aminoácidos freqüentemente possuem estrutura tridimensional similar, compartilham funções biológicas comum, e relacionam-se filogeneticamente. Este grupo de proteínas tem sido denominado de família de proteínas (KANEHISA, 2000).

Domínio é uma região que constitui uma unidade estrutural distinta em uma seqüência de aminoácidos. Uma grande seqüência de aminoácidos pode ter vários domínios. O centro de um domínio é formado por combinações de elementos estruturais denominados motivos. Cada domínio tem características

de uma proteína globular pequena e compacta que é estruturalmente independente dos outros domínios da cadeia polipeptídica (KANEHISA, 2000).

2.6 Operon

Genomas bacterianos estão organizados em unidades de expressão envolvidas por sítios onde a transcrição do DNA em RNA inicia e termina. Uma unidade de expressão, também chamada de unidade de transcrição, pode conter mais de um gene com um único promotor realizando a regulação de sua transcrição, e dando origem a um transcrito único contendo todos os genes a serem traduzidos. Neste caso trata-se de um operon. Um operon é um agrupamento de genes adjacentes sob controle do mesmo promotor sendo, portanto, uma unidade genética de expressão coordenada. É uma forma de otimizar o processo, colocando próximos os genes que executam funções relacionadas (BONATO, 2004).

Quando um operon é induzido, todos os genes que dele fazem parte são transcritos numa molécula única de mRNA. Tal molécula é denominada RNA policistrônico. A tradução ocorre de forma independente. O ribossomo se liga ao RBS e cada gene é traduzido. A tradução de cada gene, também pode ocorrer várias vezes (LEWIN, 1995).

A estrutura de um operon é uma importante característica organizacional de genomas bacterianos. Muitos conjuntos de genes ocorrem na mesma ordem em múltiplos genomas. Estes grupos de genes conservados representam candidatos a operons. As características de um grupo de genes candidatos a um operon incluem: compartilhamento de determinados elementos regulatórios; box -35 e -10 localizados na região a montante do operon; genes arranjados em seqüência na mesma fita; genes separados por distâncias curtas, em geral no máximo 150 bases; genes conservados em dois ou

mais genomas filogeneticamente relacionados e suas funções estão geralmente relacionadas (CHEN *et al.*, 2004).

Vários métodos computacionais têm sido desenvolvidos para predição de operons em genomas bacterianos baseados nas características citadas acima (OVERBEEK *et al.*, 1999; SALGADO *et al.*, 2000; ERMOLAEVA *et al.*, 2001; ZHENG *et al.*, 2002; MORENO-HAGELSIEB e COLLADO-VIDES, 2002a; 2002b; CHEN *et al.*, 2004).

2.6.1 Operon *bcs*

A etapa de polimerização da glicose para a síntese de moléculas de celulose envolve enzimas que são codificadas em um operon conservado entre várias espécies de bactérias Gram-negativas. O mecanismo molecular da biossíntese de celulose e a regulação desse processo foram primeiramente elucidados na α -proteobactéria Gram-negativa *A. xylinum*. Posteriormente, o mecanismo também foi desvendado nas bactérias *A. tumefaciens*, *R. leguminosarum* e, recentemente, nas γ -proteobactérias *S. typhimurium* e *E. coli* (RÖMLING, 2002). As características estruturais das proteínas e enzimas, codificadas pelos genes que compõem o operon *bcs* foram descritas na sessão 2.3.3.

2.7 Bioinformática e Genômica Comparativa

A bioinformática é uma nova ciência que envolve a união de diversas linhas de conhecimento: engenharia de softwares, matemática, estatística, ciência da computação e biologia molecular. Os especialistas em bioinformática devem ter o conhecimento necessário para saber quais são os problemas biológicos reais e quais as opções viáveis de desenvolvimento e abordagem computacional para o problema em questão (PROSDOCIMI *et al.*, 2003), e neste

sentido podem ajudar a tornar mais eficiente o planejamento experimental para pesquisas a serem comprovadas em laboratório.

A Internet é uma fonte de informação muito útil para as pesquisas biológicas. As informações mais abrangentes são os bancos de dados biológicos e as ferramentas de busca, disponíveis publicamente. Os repositórios de dados para novos tipos de dados biológicos, disponíveis na Internet, vêm se multiplicando rapidamente.

2.7.1 Bancos de dados biológicos

Os bancos de dados biológicos proporcionam à comunidade científica uma forma de tornar os dados produzidos em todo o mundo acessíveis, de forma mais fácil e rápida. Esses bancos de dados podem ser classificados em primários e secundários. Os bancos de dados primários são os que apresentam resultados de dados experimentais que são publicados com alguma interpretação, não havendo uma análise cuidadosa desses dados com relação aos outros publicados anteriormente. Como exemplo de bancos primários pode-se citar GenBank (Banco de dados do NCBI) e o EMBL (*Nucleotide Sequence Database*). Já nos bancos de dados secundários há uma compilação e interpretação dos dados de entrada de forma que podem ser obtidos dados mais representativos. Esses são os bancos de dados curados, como o SWISS-PROT (*Protein knowledgebase*) e o TrEMBL (*Translations of EMBL*). Outra classificação que também pode ser atribuída aos bancos de dados biológicos é quanto a função e estrutura das proteínas. Um exemplo de bancos de dados funcional é o KEGG (*Kyoto Encyclopedia of Genes and Genomes*) que disponibiliza acesso a mapas metabólicos de organismos cujos genomas tenham sido completamente ou parcialmente seqüenciados. Já os bancos de dados estruturais possuem dados relativos à estrutura de proteínas (PROSDOCIMI *et al.* 2003).

2.7.2 Genômica comparativa

O crescente número de seqüências completas de genomas bacterianos disponíveis publicamente (196 disponíveis em 24 de novembro de 2004 no *site* http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html; observe a Figura 14), representa um aumento na compreensão biológica desses microrganismos, uma vez que amplia as informações sobre metabolismo, organização celular e evolução para análise comparativa. A comparação entre genomas aumenta a compreensão de cada genoma em particular. Até poucos anos atrás não era possível tal análise, e conclusões errôneas aconteciam por falta de dados para comparações. A análise comparativa de genomas nos oferece informações sobre famílias de proteínas conhecidas, perfis filogenéticos, funções, organizações em operons ou grupo de genes, enfim, as informações comparativas podem até ajudar a predizer funções de ORFs desconhecidas no organismo que estamos analisando, comparando com outros organismos onde o gene já possui função esclarecida. A genômica comparativa utiliza genomas-modelo para a identificação de genes e vias relevantes para a fisiologia e patologia. Os bancos de dados de metabolismo e de classificação filogenética representam uma grande ajuda nesta tarefa (KOONIN e GALPERIN, 1997; GALPERIN e KOONIN, 2000, KANEHISA, 2000).

2.7.3 Ferramentas de bioinformática e de genômica comparativa

Aqui são apresentadas brevemente as plataformas *on-line* para pesquisas biológicas, os bancos de dados biológico e as ferramentas de genômica comparativa e para análise de seqüências úteis e importantes para esta abordagem, e que estão mostrados na Tabela 2. Em seguida é apresentada uma breve descrição dos itens desta tabela.

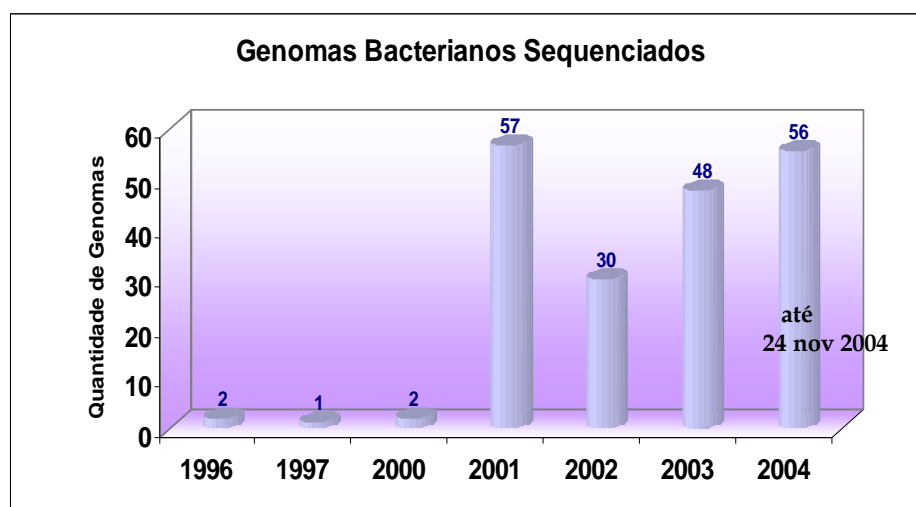


Figura 14 - Genomas bacterianos sequenciados, depositados no GenBank até 24 de novembro de 2004.

2.7.3.1 Plataformas *on-line* para pesquisa biológicas

NCBI (*National Center for Biotechnology Information*)

O NCBI é uma organização do Estados Unidos que cria bases de dados públicas, conduz pesquisas na área de biologia computacional, desenvolve ferramentas para analisar dados de genomas, e divulga informação biomédica para promover uma melhor compreensão dos processos moleculares que afetam a saúde humana e as doenças (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, 2004).

EMBL (*The European Molecular Biology Laboratory*)

O EMBL existe desde 1974 e é mantido por dezessete países incluindo quase toda a Europa ocidental e Israel. O EMBL é uma das grandes instituições de pesquisa do mundo e foi fundado com o objetivo de conduzir pesquisa básica em biologia molecular, fornecer serviços essenciais aos cientistas,

fornecer treinamento de alto nível a seus funcionários, estudantes, e visitantes, e desenvolver nova instrumentação para a pesquisa biológica (THE EUROPEAN MOLECULAR BIOLOGY LABORATORY, 2004).

Tabela 2 - Ferramentas de bioinformática importantes e úteis para esta abordagem.

Plataformas <i>on-line</i> para Pesquisa Biológicas	Endereço Web (URL)
NCBI	http://www.ncbi.nlm.nih.gov/
EMBL	http://www.embl.org/
EBI	http://www.ebi.ac.uk/Information/
INTERPRO	http://www.ebi.ac.uk/interpro
BRGene	http://www.brgene.lncc.br/
ExpASy	http://us.expasy.org/
Bancos de Dados de Seqüências de DNA e Estruturas de Proteínas	Endereço Web (URL)
GenBank	http://www.ncbi.nlm.nih.gov
EMBL-EBI	http://www.ebi.ac.uk/embl/
DDBJ	http://www.ddbj.nig.ac.jp
PDB	http://www.rcsb.org/pdb/
UNIPROT	http://www.expasy.uniprot.org/
SWISS-PROT	http://us.expasy.org/sprot/
TrEMBL	http://us.expasy.org/sprot/
PIR	http://pir.georgetown.edu/
PRODOM	http://prodes.toulouse.inra.fr/prodom/current/html/home.php
PFAM	http://www.sanger.ac.uk/Software/Pfam/
SMART	http://smart.embl-heidelberg.de/
Ferramentas de Genômica Comparativa e para Análise de Seqüências	Endereço Web (URL)
COG	http://www.ncbi.nlm.nih.gov/COG
KEGG	http://www.genome.ad.jp/kegg/
MBGD	http://mbgd.genome.ad.jp/
BLAST	http://www.ncbi.nlm.nih.gov/blast/
CLUSTAL W	http://www.ebi.ac.uk/clustalw/
ARTEMIS	http://www.sanger.ac.uk/Software/Artemis
BIOEDIT	http://www.mbio.ncsu.edu/BioEdit/bioedit.html

EBI (*European Bioinformatics Institute*)

O EBI, é uma organização acadêmica que faz parte do EMBL. O EBI é um centro para a pesquisa e serviços de bioinformática. O instituto controla bancos de dados biológicos que incluem ácidos nucleicos, seqüências de proteínas e estruturas macromoleculares. A missão do EBI é assegurar que as informações da pesquisa em biologia molecular e de genomas sejam de domínio público e que sejam acessíveis livremente a toda comunidade científica a fim de promover o progresso científico (LOPEZ *et al.*, 2003).

INTERPRO (*Integrated resource of Protein Families, Domains and Sites*)

Os bancos de dados secundários de proteínas como PROSITE, PRINTS, SMART, PFAM e PRODOM são recursos vitais para identificar relacionamentos distantes em novas seqüências, e possibilitar predizer a função e a estrutura de proteínas. Infelizmente, estes bancos de dados não compartilham os mesmos formatos e nomenclatura, e cada banco de dados tem seus pontos fortes e fracos. Na tentativa de se obter uma visão integrada dos bancos de dados o EBI, o SIB, a Universidade de Manchester, o Instituto Sanger, o GENE-IT, o CNRS/INRA, a LION Bioscience AG e a Universidade de Bergen unificaram o PROSITE, PRINTS, PRODOM e o PFAM no INTERPRO. Em seguida, o SMART e TIGRFAMs juntaram-se ao projeto, e também o PIR SuperFamily e a SUPERFAMILY (MULDER *et al.*, 2002).

ExPASy (*Expert Protein Analysis System*)

O ExPASy é o servidor proteômico do SIB (*Swiss Institute of Bioinformatics*) dedicado à análise de seqüência e estrutura de proteínas, integrando vários bancos de dados, entre eles SWISS-PROT e TrEMBL (GASTEIGER *et al.*, 2003).

BRGene (*Brazilian Genome Virtual Institute of Genomic Research*)

O BRGene nasceu em 2000, a partir da iniciativa do Ministério da Ciência e Tecnologia (MCT) para mapear o genoma completo de uma bactéria. A bactéria selecionada foi a *Chromobacterium violaceum*, uma bactéria que produz metabólitos secundários com potencial para tratamento de algumas doenças e outros de interesse biotecnológico. Nesse projeto nacional estão envolvidos pesquisadores de diversas regiões do Brasil, cerca de 160 cientistas vinculados a 25 laboratórios, que se uniram, somando aptidões para o sequenciamento genômico e análise das seqüências genéticas. Informações do genoma da *C. violaceum* estão disponibilizadas pelo BRGene (BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM, 2004).

2.7.3.2 Bancos de dados de seqüência de DNA e estrutura de proteínas**GenBank (Banco de dados do NCBI)**

O banco de dados GenBank possui seqüências de DNA, publicamente disponíveis, de diferentes organismos, disponibilizados através da submissão direta de dados da seqüência dos laboratórios individuais e dos grandes projetos de sequenciamento. A troca de dados com *The European Molecular Biology Laboratory* (EMBL) e o *DNA Data Bank of Japan* (DDBJ) ajudaram a assegurar uma cobertura mundial detalhada de todas essas seqüências. Os dados do GenBank são acessíveis através do sistema de recuperação integrado Entrez do NCBI, que integra dados das principais bases de dados de seqüência de DNA e de proteína; junto com a informação da taxonomia do genoma, da estrutura da proteína, e as literaturas biomédicas do PubMed. O GenBank e os seus recursos são de acesso gratuito através da página principal do NCBI (BENSON *et al.*, 2000).

EMBL-EBI (*Nucleotide Sequence Database*)

O banco de dados de seqüências de nucleotídeos do EMBL é mantido pelo EBI com colaboração internacional do DDBJ e do GenBank do NCBI. Os dados são trocados diariamente entre os bancos de dados colaboradores. Os principais contribuintes do EMBL são autores e grupos individuais de projetos de pesquisa genômica. Os formatos de dados do EMBL são compatíveis com o de outros bancos (GenBank, DDBJ) e muitas ferramentas de análise também são comuns. Os alinhamentos múltiplos, as pesquisas por palavras-chave, identificação de padrões específicos (domínios e motivos) são algumas das ferramentas disponíveis para a análise de seqüências. O SWISS-PROT, na Suíça, um banco de dados de proteínas, também é integrado ao EMBL, disponibilizando domínios, descrições de funções, e estruturas de proteínas (STOESSER *et al.*, 2001; REDASCHI *et al.*, 2001).

DDBJ (*DNA Data Bank of Japan*)

O banco de dados DDBJ coleta o máximo possível dos dados de pesquisas genômicas, principalmente de pesquisadores e grupos de pesquisa japoneses. Juntos, o GenBank, o EMBL e o DDBJ formam o chamado *International Nucleotide Sequence Database (INSDB)*, mantendo um importante acordo de cooperação e intercâmbio de dados (TATENO *et al.*, 2002).

PDB (*Protein Data Bank*)

O PDB é um banco de dados com informações de estrutura tridimensional de macromoléculas biológicas. O PDB é o maior repositório do mundo para o processamento e a distribuição de dados da estrutura tridimensional de grandes moléculas de proteínas e ácidos nucléicos. Esse banco de dados contém estruturas tridimensionais de macromoléculas

biológicas que foram determinadas experimentalmente. Os dados são facilmente submetidos ao PDB via web, e também os resultados de pesquisa realizadas neste banco de dados são facilmente acessíveis (BERMAN *et al.*, 2000).

UNIPROT (*Universal Protein Resource*)

O UNIPROT é um repositório central de seqüência e função de proteínas, criado pela junção das informações contidas no SWISS-PROT, no TrEMBL, e no PIR. (APWEILER *et al.*, 2004).

SWISS-PROT (*Protein knowledgebase*)

O SWISS-PROT é um banco de dados curado de seqüência de proteína. Nele são apresentadas a descrição da função de uma proteína, a estrutura dos domínios, locais de modificações pós-tradução, e outras anotações. Este banco de dados concentra seus esforços em manter um nível mínimo de redundância e um alto nível de integração com outros bancos de dados (BOECKMANN *et al.*, 2003).

TrEMBL (*Translations of EMBL*)

O TrEMBL é um suplemento de anotação computacional do SWISS-PROT. Nele estão contidas todas as traduções das seqüências de nucleotídeos do EMBL que ainda não estão integradas no SWISS-PROT (BOECKMANN *et al.*, 2003).

PIR (*Protein Information Resource*)

O PIR é um banco de dados de proteínas anotadas. Recentemente, o PIR, o SWISS-PROT, e o TrEMBL se uniram e formaram o banco de dados UNIPROT (WU *et al.*, 2003).

PRODOM (*Protein Domain*)

O PRODOM é banco de dados de domínios de proteínas que consiste de uma compilação automática de domínios homólogos gerados automaticamente a partir do banco de dados SWISS-PROT e TrEMBL. O banco de dados PRODOM é parte do projeto INTERPRO (SERVANT *et al.*, 2002).

PFAM (*Protein families database*)

O banco de dados PFAM é um conjunto de alinhamentos múltiplos de seqüências, abrangendo domínios comuns de proteínas, e de famílias de proteínas. Para cada família de proteínas no PFAM pode-se: visualizar os alinhamentos múltiplos; visualizar a arquitetura dos domínios das proteínas; examinar a distribuição da espécie; fazer *links* com outros bancos de dados e visualizar estruturas de proteína conhecidas.

O PFAM é um banco de dados composto de duas partes. A primeira parte, denominada PFAM-A é a parte curada, contendo 7459 famílias de proteínas (em agosto de 2004), 74% do total de proteínas do PFAM; o PFAM-B, com 24% das proteínas do PFAM, contém pequenas famílias do banco de dados PRODOM que não estão no PFAM-A. Embora as famílias do PFAM-B sejam de menor qualidade, elas podem ser úteis quando não são encontradas no PfaM-A (BATEMAN *et al.*, 2004).

SMART (*Simple Modular Architecture Research Tool*)

No SMART é possível a identificação e anotação de domínios e a análise da arquitetura de domínios. Cada domínio encontrado no banco de dados de proteína não redundante (nr), assim como os parâmetros de busca e as informações taxonômicas são armazenados em um banco de dados relacional. As interfaces com o usuário permitem buscas por proteínas que contêm domínios específicos (SCHULTZ *et al.*, 2000).

2.7.3.3 Ferramentas de genômica comparativa e para análise de seqüências

A seguir são apresentadas algumas ferramentas de genômica comparativa importantes para esta abordagem.

COG (*Clusters of Orthologous Groups of Proteins*)

O banco de dados COG do NCBI é uma tentativa de classificação filogenética das proteínas codificadas em 21 genomas completos de bactérias, archaea e eucariotos. As famílias de proteínas conservadas no COG contêm grupos de proteínas ortólogas. Duas proteínas são ortólogas se os genes que as codificam evoluíram de um ancestral comum. Tipicamente genes ortólogos codificam proteínas que têm a mesma função. O mesmo não acontece com genes parálogos que se originaram por duplicação e às vezes desenvolveram uma nova função. O banco de dados COG compreende 2091 COGs que incluem 56-83% dos produtos dos genes dos genomas completos de bactérias e archaeas, e aproximadamente 35% do genoma da *Saccharomyces cerevisiae* (TATUSOV *et al.*, 2000).

KEGG (*Kyoto Encyclopedia of Genes and Genomes*)

A KEGG é uma base de conhecimento para a análise sistemática de funções dos genes, ligando a informação genômica à informação funcional. É uma ferramenta centrada na comparação de genomas quanto à sua capacidade para codificar diferentes vias metabólicas. Pode-se obter mapas metabólicos e identificar quais enzimas estão presumidamente codificadas por um determinado genoma. Permite também identificar *clusters* de genes conservados entre espécies. As atualizações do KEGG são diárias (KANEHISA e GOTO, 2000).

MBGD (*Microbial Genome Database*)

O banco de dados MBGD é um sistema para a análise comparativa entre genomas microbianos que tenham sido completamente seqüenciados. A função principal do MBGD é criar uma tabela usando o COG, comparando todos contra todos, relacionando as similaridades entre genes em múltiplos genomas. O MBGD é, assim, uma ferramenta para identificar genes e proteínas homólogas em vários genomas (UCHIYAMA, 2003).

BLAST (*Basic Local Alignment Search Tool*)

O BLAST é uma ferramenta de comparação e alinhamento de seqüências depositadas em banco de dados e utilizada para identificação de similaridades. Os resultados matemáticos permitem uma análise do desempenho deste método, bem como, o significado estatístico dos alinhamentos que o programa gera. O algoritmo básico é simples e robusto; pode ser executado de maneira fácil, e é aplicado em uma variedade de contextos, incluindo a pesquisas de seqüências de DNA e da proteína, motivos, identificação de genes, e na análise

de regiões de múltiplas similaridades em seqüências de DNA. De acordo com o tipo de seqüência de entrada (nucleotídeo ou aminoácido) e com o tipo de resultado esperado, existe um programa BLAST específico (ALTSCHUL *et al.*, 1990): BLASTn: compara seqüências de nucleotídeos; BLASTp: compara seqüências de aminoácidos de proteínas e identifica também possíveis domínios conservados nas proteínas analisadas; BLASTx: compara seqüências de nucleotídeos com a base de dados de proteínas; tBLASTn: traduz uma seqüência de aminoácidos para nucleotídeo e compara com o banco de dados de genes; tBLASTx: traduz uma seqüência de nucleotídeo para aminoácidos e compara com o banco de proteínas; RPS-BLAST: busca por alinhamentos com domínios conservados conhecidos; BLAST 2 Sequences: realiza um alinhamento comparativo entre duas seqüências.

CLUSTAL W (*Multiple Sequence Alignment Program*)

CLUSTAL W é a versão Web de uma das ferramentas gratuitamente disponíveis mais utilizada para se fazer alinhamento múltiplo de seqüências de nucleotídeos e de aminoácidos. Fornece ao usuário uma variedade de opções de parâmetros e diferentes apresentações dos resultados. Os alinhamentos podem ser visualizados e alterados a partir de interface gráfica. Além disso, permite criar também árvores filogenéticas (THOMPSON *et al.*, 1994).

ARTEMIS (*A DNA Sequence Viewer and Annotation Tool*)

Artemis é uma ferramenta de visualização e anotação de seqüência de DNA, que permite visualizar os resultados das análises no contexto da seqüência e de sua tradução dos seis quadros (ORFs). O Artemis é especialmente útil para analisar genomas de bactérias, de archaeas e de alguns eucariotos. O Artemis é executado em Java, e funciona em toda a plataforma

apropriada. As seqüências e as anotações podem ser lidas e escritas em formato EMBL, GenBank e em GFF (*Graphic File Format*) (RUTHERFORD *et al.*, 2000).

BIOEDIT (*Sequence Alignment Editor for Windows*)

Bioedit é uma ferramenta para editar e alinhar seqüências de DNA e de proteínas, gerenciar submissão de seqüências a outros aplicativos, e visualizar os resultados. O Bioedit é uma ferramenta muito confortável para a visualização, com uma variedade de funções de análise das seqüências de nucleotídeos e aminoácidos (TIPPMANN, 2004).

CAPÍTULO III

Materiais e Métodos

Neste capítulo estão descritos os procedimentos para a realização da etapa da investigação computacional do genoma de *Chromobacterium violaceum*, buscando por informações genômicas envolvidas na formação de biofilme celulósico. Além disso, serão descritos os procedimentos laboratoriais da etapa de cultivo desta bactéria, a fim obter evidências da presença de celulose como um elemento do biofilme produzido.

3.1 Ferramentas de bioinformática utilizadas

As plataformas *on-line* para pesquisas biológicas, os bancos de dados biológicos e as ferramentas de genômica comparativa e para análise de seqüências utilizadas neste trabalho estão listadas na Tabela 3.

Tabela 3 - Ferramentas de bioinformática, genômica comparativa e para análise de seqüências utilizadas neste trabalho.

Ferramenta	Endereço Web (URL)	Descrição
NCBI	www.ncbi.nlm.nih.gov/	Organização do USA voltada a pesquisas na área de biologia computacional
BRGene	www.brgene.lncc.br/	Informações do seqüenciamento do genoma de <i>C. violaceum</i>
GenBank	www.ncbi.nlm.nih.gov	Banco de dados de seqüência do NCBI
PFAM	www.sanger.ac.uk/Software/Pfam/	Banco de dados de domínios comuns de proteínas, e de famílias de proteínas
KEGG	www.genome.ad.jp/kegg/	Vias metabólicas
BLAST	www.ncbi.nlm.nih.gov/blast/	Comparação e alinhamento de seqüências
Clustal W	www.ebi.ac.uk/clustalw/	Alinhamento múltiplo de seqüências
ARTEMIS	www.sanger.ac.uk/Software/Artemis	Visualização e anotação de seqüência
BIOEDIT	www.mbio.ncsu.edu/BioEdit/bioedit.html	Edição e alinhamento de seqüências, submissão de seqüências a outros aplicativos, e visualização dos resultados

3.2 Procedimentos Experimentais Computacionais

3.2.1 Identificação das enzimas da via de biossíntese de celulose em *Chromobacterium violaceum*

Para a identificação no genoma de *C. violaceum* dos genes/ORFs que codificam as enzimas envolvidas nas reações da via metabólica de biossíntese de celulose utilizou-se o banco de dados de vias metabólicas KEGG. Para tanto, utilizou-se como modelo as enzimas pertencentes a via de biossíntese de celulose por *A. xylinum*, proposta por ROSS *et al.* (1991),

3.2.2 Identificação do operon de biossíntese de celulose em *Chromobacterium violaceum*

Sabe-se que um dos principais elementos da matriz extracelular de biofilmes bacterianos é o exopolissacarídeos (EPS). Da literatura, verificou-se que celulose é um EPS presente na matriz extracelular de biofilme bacteriano de várias espécies, e que os genes da associação celulose sintase, estão organizados em um operon *bcs*, conservado entre várias espécies bacterianas. O BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM (2003) relatam que no genoma de *C. violaceum* foram identificados três ORFs relacionadas com a biossíntese de celulose (CV2675, CV2677 e CV2678). A partir daí, esta pesquisa buscou elucidar as características genômicas envolvidas na biossíntese de celulose por *C. violaceum*.

A análise dos genes de interesse foi realizada utilizando a plataforma de bioinformática do BRGene, e do NCBI.

A metodologia utilizada para identificação de um provável operon *bcs* em *C. violaceum* foi conduzida em cinco etapas, as quais estão apresentadas no fluxograma da Figura 15.

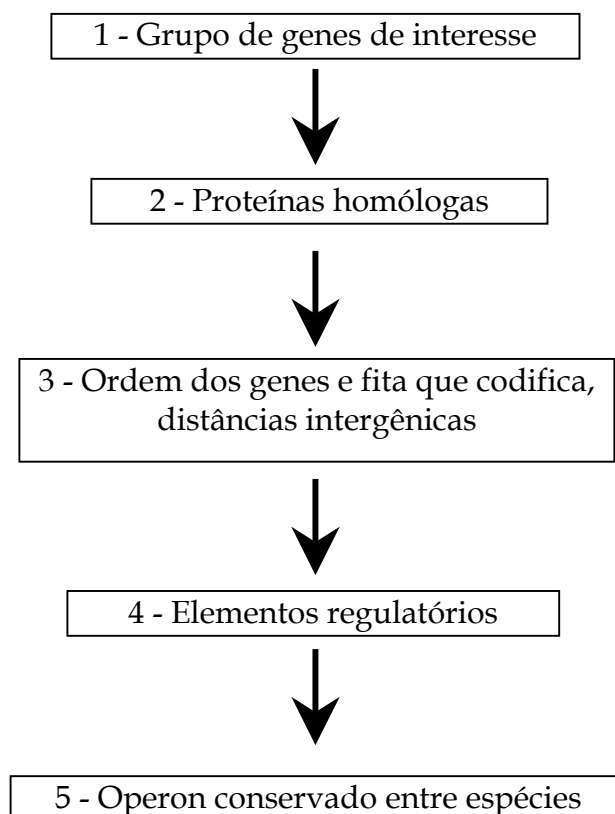


Figura 15 - Etapas para a identificação do operon *bcs* no genoma de *Chromobacterium violaceum*.

Na primeira etapa, foram explorados dados da anotação do genoma de *C. violaceum* (BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM, 2003) que estão disponibilizados publicamente no BRGene, buscando por genes envolvidos na biossíntese de celulose. Foram identificadas as seguintes ORFs: CV2675, CV2676, CV2677, CV2678 e CV2679, que correspondem respectivamente aos genes *bcsC* (NP_902345), *bcsZ* (NP_902346), *bcsB* (NP_902347), *bcsA* (NP_902348) e a proteína hipotética conservada (NP_902349) com probabilidade de compor o operon *bcs*; (entre parênteses, o número de acesso da proteína no GenBank do NCBI).

Em uma segunda etapa, buscando-se por proteínas homólogas, as seqüências de aminoácidos das proteínas, produtos dos genes/ORFs de *C. violaceum* mencionados acima, foram submetidas ao programa BLASTP (ALTSCHUL *et al.*, 1990), a partir do NCBI. Nosso interesse era encontrar proteínas homólogas de bactérias cujo processo de produção de celulose já tinha sido caracterizado experimentalmente. Todas as seqüências de aminoácidos das proteínas das bactérias selecionadas foram obtidas do banco de dados GenBank do NCBI e foram submetidas ao alinhamento múltiplo de seqüências, utilizando o programa Clustal W (THOMPSON *et al.*, 1994), através da interface de bioinformática Bioedit (*Biological sequence alignment editor*) (TIPPMANN, 2004). Do resultado do alinhamento Clustal W buscou-se identificar as identidades e motivos conservados entre os microrganismos selecionados. A interface Bioedit foi utilizada também para gerar o padrão de consenso entre as seqüências, editar e imprimir os alinhamentos. As arquiteturas dos domínios das proteínas foram obtidas do banco de dados de família de proteínas (PFAM).

Na terceira etapa foram realizadas as identificações da ordem dos genes de interesse no genoma de *C. violaceum*, bem como as distâncias intergênicas e outros possíveis genes envolvidos no processo. Nesta etapa também utilizamos o software Artemis V5 (RUTHERFORD *et al.*, 2000), uma ferramenta de anotação e visualização de DNA.

A quarta etapa consistiu em identificar possíveis elementos regulatórios de transcrição do operon. Vários softwares fazem a predição de elementos regulatórios de transcrição em genomas de procariotos. Neste estudo utilizou-se a predição das possíveis regiões de regulação disponibilizada no BRGene pela anotação automática. O software Artemis V5 (RUTHERFORD *et al.*, 2000) foi utilizado para se obter a região a montante do operon. Estas informações

foram utilizadas para elaborar um esquema gráfico do operon, localizando na região a montante os possíveis sítios de regulação da transcrição.

Na quinta etapa foi realizada uma análise dos operons *bcs* das bactérias envolvidas neste estudo, e construído um esquema gráfico alinhando os genes dos diversos operons, visando identificar as similaridades entre eles. Este esquema permitiu uma visualização da organização dos genes de biossíntese de celulose nos diferentes grupos de bactérias.

3.2.3 Identificação de proteínas com motivo GGDEF em *Chromobacterium violaceum*

As enzimas diguanilato ciclase e fosfodiesterase que regulam o nível da molécula c-di-GMP celular possuem o motivo GGDEF. A molécula c-di-GMP é um ativador alostérico da síntese de celulose por ligar-se a proteína BcsB. A fim de se identificar proteínas com o domínio GGDEF em *C. violaceum* foi realizada uma pesquisa no banco de dados PFAM.

3.3 Procedimentos Experimentais Microbiológicos

3.3.1 Microrganismo utilizado na etapa laboratorial

O microrganismo utilizado neste ensaio foi a bactéria *Chromobacterium violaceum*, linhagem ATCC 12472, a mesma linhagem cujo genoma foi seqüenciado.

3.3.2 Materiais e equipamentos utilizados nos ensaios

Para a produção de biofilme celulósico os seguintes materiais foram utilizados:

3.3.2.1 Meio de cultivo LB (Luria Bertani) para o cultivo de *Chromobacterium violaceum*

- ❖ Triptona 10 g
- ❖ Cloreto de sódio 5 g
- ❖ Extrato de levedura 5 g
- ❖ H₂O destilada 1000 mL
- ❖ pH 7

3.3.2.2 Enzima utilizada para a hidrólise enzimática

- ❖ Celluclast 1,5 L (celulase), gentilmente doada pela Novozymes Latin America Ltda.

3.3.2.3 Materiais e equipamentos utilizados para a determinação de glicose em amostras do cultivo

- ❖ Kit de glicose enzimática da Empresa Gold Analisa Diagnóstica
 - Padrão - contém glicose 100 mg/dL
 - Reagente de cor - contém tampão pH 7,5 de fosfato 70 mmol/L, fenol 5,0 mmol/L, glicoseoxidase > 10000 U/L, peroxidase > 1000 U/L e 4-aminoantipirina 0,4 mmol/L
- ❖ Espectrofotômetro - Spectrophotometer 8500 II - Ruilian Science & Technology co., Ltda.

3.3.2.4 Solução tampão utilizada na hidrólise enzimática

- ❖ Citrato de sódio 50 mM (pH 4,5).

3.3.3 Cultivo de *Chromobacterium violaceum* para produção de biofilme

3.3.3.1 Pré-Inóculo

O pré-inóculo foi preparado em um Erlenmeyer de 125 mL, contendo 25 mL de meio de cultura LB a partir de uma colônia *C. violaceum* retirada de cultura em placa de Petri. O frasco foi mantido em agitador orbital (150 rpm) a 32 °C, durante 24 h.

3.3.3.2 Inóculo e condições de cultivo

A cultura empregada na produção de biofilme foi realizada em Erlenmeyer de 250 mL, contendo 50 mL de meio LB, onde foram inoculados 100 µL do pré-inóculo. Objetivando-se obter as melhores condições de cultivo foram realizados dois experimentos: um com Erlenmeyer em condições estáticas em estufa a 32 °C, e outro em agitador orbital (150 rpm) a 32 °C. Cada experimento foi realizado em triplicata.

3.3.4 Coleta das amostras para análise

Após 6 dias de cultivo as amostras foram coletadas para análise. No experimento realizado em condições estáticas, a película formada na superfície líquida foi recolhida, com auxílio de um bastão de vidro, lavada com água destilada e o material insolúvel submetido à secagem em estufa a 50 °C.

A suspensão restante do meio de cultivo (sem a película) foi centrifugada (7000 rpm, 15 min), sendo os sedimentos descartados e ao sobrenadante foi adicionado três vezes o volume de álcool etílico a 92,8%. A precipitação assim obtida foi separada por centrifugação (7000 rpm, 15 min) e submetido à

secagem a 50 °C. Na cultura realizada sob agitação não houve a formação de película. O mesmo procedimento realizado para a porção líquida do meio de cultura estático foi adotado.

3.3.5 Hidrólise enzimática

A fim de certificar que o precipitado recuperado era composto de celulose este foi submetido à ação de celulases e posteriormente a concentração de glicose foi determinada.

A cada 10 mg das amostras isoladas de biofilme (p1, e1 e a1), ver Tabela 4, foram adicionados 40 µL de celulase (Celluclast 1.5 L - Novozymes) em 1 mL de tampão citrato de sódio (50 mM, pH 4,5). As amostras foram incubadas em banho-maria a 50 °C por 1 hora. Paralelamente, 10 mg de cada amostra (p2, e2 e a2), ver Tabela 4, foram incubadas em 1 mL de tampão citrato de sódio (50 mM, pH 4,5), para posterior determinação de glicose contaminante. As amostras foram então centrifugadas (5000 rpm, 5 min). A concentração de glicose liberada por hidrólise enzimática foi determinada no sobrenadante.

Tabela 4 - Amostras isoladas de biofilme.

Amostras	Descrição da amostra
p1	película
e1	precipitado da cultura estática
a1	precipitado da cultura agitada
p2	película
e2	precipitado da cultura estática
a2	precipitado da cultura agitada

3.3.6 Determinação da concentração de glicose

Para a determinação de glicose foi utilizado o kit de determinação de glicose oxidase (Gold Analisa Diagnóstica). Foram coletados 100 μL do sobrenadante de cada tubo de ensaio contendo o produto da hidrólise enzimática, e colocados em tubos de ensaio identificados com p1, p2, e1, e2, a1 e a2. Em cada tubo foram adicionados 900 μL de reagente de cor. Em um tubo identificado com pr foram adicionados 10 μL do padrão, 900 μL de reagente de cor e 90 μL de água. Os tubos foram homogeneizados, e mantidos em temperatura ambiente por 10 minutos. As leituras da absorbância de cada amostra foram realizadas em espectrofotômetro a 500 nm, zerando o aparelho com o branco contendo 1000 μL reagente de cor.

Para o cálculo da concentração de glicose foi utilizada a seguinte expressão:

$$\text{Concentração de glicose da amostra} = \text{absorbância da amostra} \times \text{fator de calibração}$$

onde:

$$\text{Fator de calibração} = \text{concentração do padrão} \times \text{absorbância da solução padrão}.$$

CAPÍTULO IV

Resultados e Discussão

4.1 Análise dos Resultados do Procedimento Computacional

4.1.1 Via de biossíntese de celulose em *Chromobacterium violaceum*

Dos resultados da análise realizada da via de biossíntese de celulose em *A. xylinum*, proposta por ROSS *et al.* (1991), e das informações obtidas no banco de dados de vias metabólicas KEGG, identificou-se no genoma de *C. violaceum* os genes/ORFs que codificam as enzimas envolvidas nas reações da via de biossíntese de celulose em *C. violaceum*, os quais estão apresentados da Tabela 5. Com isto foi possível elaborar esquematicamente uma via de biossíntese de celulose em *C. violaceum* a partir de glicose, conforme apresentado na Figura 16. Em seguida é apresentada uma descrição detalhada das reações desta via.

Na via de biossíntese de celulose em *A. xylinum*, primeiramente a glicose é convertida em G6P, reação catalisada pela enzima glicoquinase com a participação de uma molécula de ATP e liberando um ADP (glicose + ATP \Rightarrow G6P + ADP). Em *C. violaceum*, esta enzima é codificada pelo gene *glk* (CV0147). Na reação seguinte a G6P é convertida em G1P, reação catalisada pela enzima fosfoglicomutase (G6P \Leftrightarrow G1P). No genoma de *C. violaceum* esta enzima é codificada pelo gene *algC* (CV2172). Na seqüência, a reação catalisada pela enzima UDP-glicose-pirofosforilase converte G1P mais uma molécula de uridina trifosfato (UTP) em uridina difosfato glicose (UDP-glicose), liberando uma molécula de PPI. Em *C. violaceum*, esta enzima é codificada pelo gene *galU* (CV3901). A molécula UDP-glicose é precursora da biossíntese da celulose. A polimerização de glicose em cadeia $\beta(1\rightarrow4)$ glicana (celulose) é realizada pelo

complexo celulose sintase usando UDP-glicose como substrato. Em *C. violaceum* as proteínas que compõem a associação celulose sintase, organizam-se em um operon, conforme mostrado neste trabalho, denominado operon *bcs*, formado pelas proteínas e enzimas ORF CV2679, BcsA, BcsB, BcsZ e BcsC. A ORF CV2679 anteriormente anotada como proteína hipotética conservada, sua função está sendo atribuída por genômica comparativa.

Tabela 5 – Proteínas/enzimas presentes no genoma de *C. violaceum* envolvidas na via de biossíntese de celulose.

Proteína/enzima	EC/COG	Identificador da ORF	Gene	Função
Glicoquinase	EC 2.7.1.2 COG0837	CV0147	<i>glk</i>	Catalisa a fosforilação, glicose em G6P
Fosfoglicomutase	EC 5.4.2.2 COG1109	CV2172	<i>algC</i>	Catalisa a isomerilação, G6P em G1P
UDP-glicose-pirofosforilase	EC 2.7.7.9 COG1210	CV3901	<i>galU</i>	Sintetiza UDP-glicose a partir de G1P + UTP, liberando um PPI
Complexo celulose sintase (Associação de proteínas e enzimas)	Não possui EC COG1192	CV2679	novo gene	Provável ATPase envolvida na divisão celular e partição do cromossomo
	EC 2.4.1.12 COG1215	CV2678	<i>bcsA</i>	Subunidade catalítica da celulose sintase; polimeriza glicose em cadeia $\beta(1\rightarrow4)$ glicana
	Não possui EC nem COG	CV2677	<i>bcsB</i>	Subunidade de regulação da celulose sintase, liga-se ao ativador da celulose sintase, c-di-GMP
	EC 3.2.1.4 COG3405	CV2676	<i>bcsZ</i>	Celulase; catalisa a hidrólise de ligações internas $\beta(1\rightarrow4)$ glicosídicas da celulose
	EC 2.4.1.12 COG0457	CV2675	<i>bcsC</i>	Envolvida na biossíntese de celulose em várias espécies bacterianas; sua função ainda permanece desconhecida

EC = Classificação de Enzimas; COG = Conjunto do grupo de proteínas ortólogas

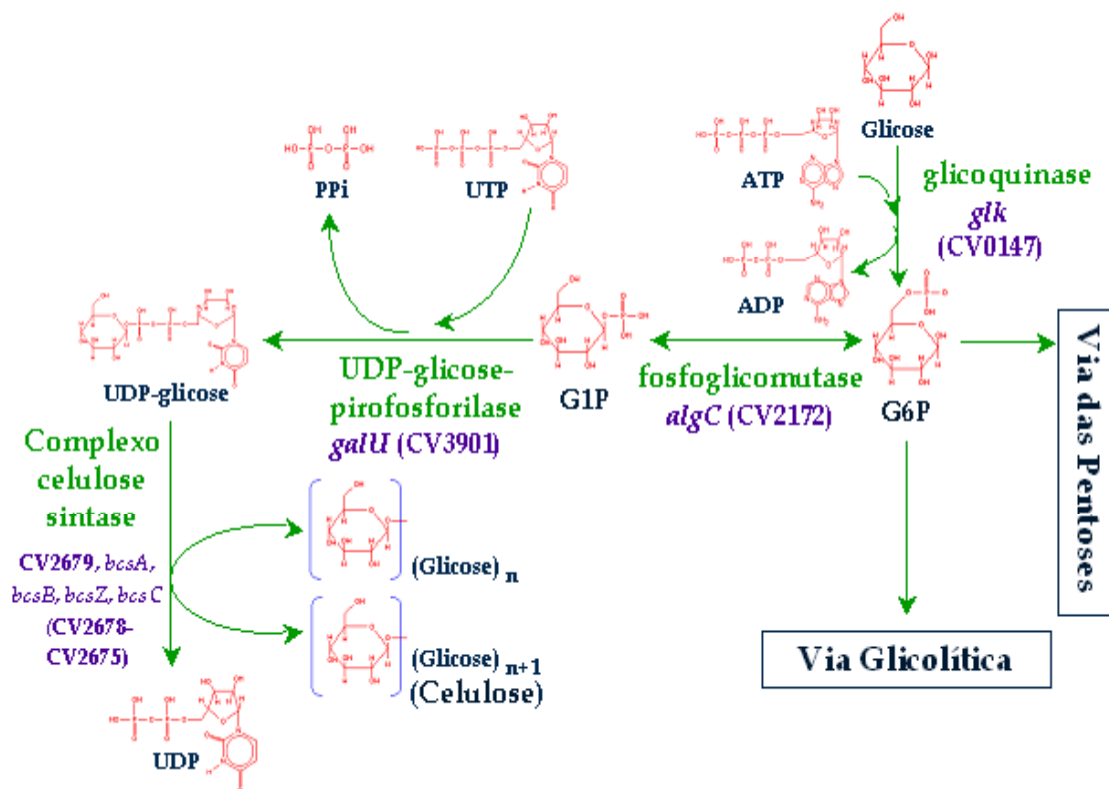


Figura 16 -Via de biossíntese de celulose em *C. violaceum* a partir do substrato glicose.

Os resultados da organização dessas informações indicam que este microrganismo possui a capacidade de sintetizar celulose, embora não existissem até então relatos de evidências experimentais.

4.1.2 Identificação do operon *bcs* de *Chromobacterium violaceum*

O genoma de *C. violaceum* apresenta os genes *bcsC* (NP_902345), *bcsZ* (NP_902346), *bcsB* (NP_902347) e *bcsA* (NP_902348), homólogos aos genes envolvidos na biossíntese de celulose por bactérias já bem caracterizadas como produtoras de celulose. A investigação do genoma da *C. violaceum* nos permitiu identificar não somente um provável operon *bcs* - que reúne os quatro genes anteriormente citados - mas também a existência de um novo gene - ORF CV2679 - número de acesso no GenBank NP_902349, homólogo dos genes *ylhQ*

de *Escherichia coli* K12 (NP_417991) e *Salmonella typhimurium* LT2 (NP_462521), localizado na região a montante do operon, da fita reversa complementar. (Entre parênteses, o número de acesso da proteína no GenBank do NCBI).

A Figura 17 apresenta um esquema geral da estrutura do operon *bcs* de *C. violaceum*. Na região a jusante do operon *bcs* existe uma pequena ORF com 123 aminoácidos, CV2674 (NP_902344), que poderia fazer parte deste operon; porém, esta não foi considerada por não apresentar similaridade com nenhuma proteína depositada no GenBank até o presente momento. As distâncias intergênicas foram obtidas utilizando o software Artemis V5. A distância intergênica encontrada entre o gene *bcsC* e *bcsZ* foi de -4 bases, ou seja, uma sobreposição de 4 bases; entre o gene *bcsZ* e *bcsB* a distância encontrada foi de -1 base, *i.e.*, uma sobreposição de 1 base; entre o gene *bcsB* e *bcsA* foi 32 bases; entre o gene *bcsA* e a ORF CV2679 foi -4 bases.

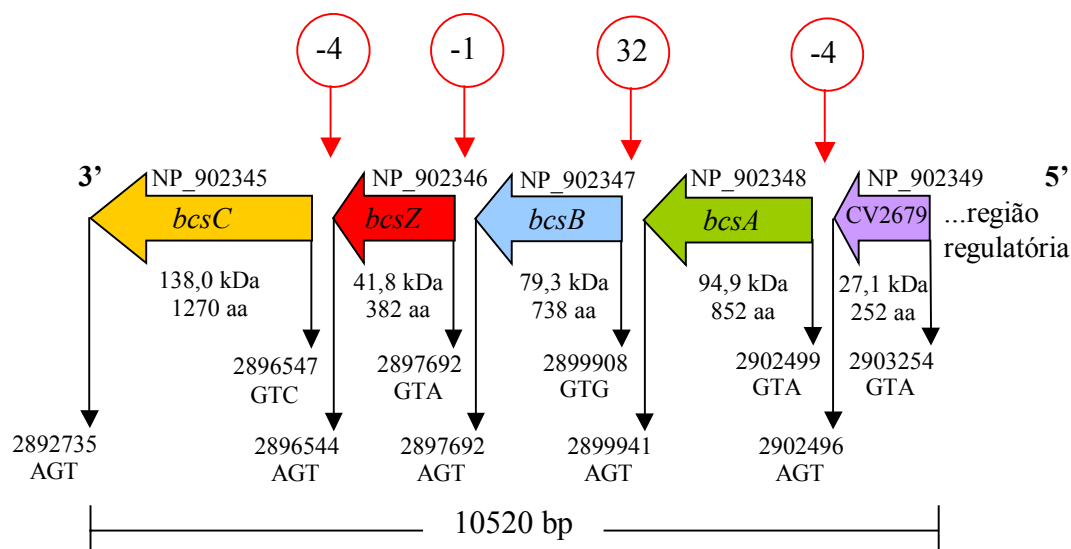


Figura 17 - Esquema geral da estrutura do operon *bcs* de *C. violaceum*. As setas indicam os genes ou ORFs que compõem o operon. O nome dentro da seta corresponde ao gene ou ORF. O número acima da seta corresponde ao número de acesso do gene no GenBank do NCBI. Imediatamente abaixo das setas está indicada a massa molecular de cada polipeptídeo, bem como, o número de aminoácidos que o compõem. Os números dentro dos círculos correspondem às distâncias intergênicas. As anotações sob as setas indicam a posição de início e de término de cada gene/ORF e também o códon de início e de parada. As informações apresentadas neste esquema foram obtidas do BRGene e NCBI.

As seqüências de aminoácidos codificadas pelos genes/ORFs citados acima foram submetidas ao programa BLASTP, e do resultado foram selecionadas oito bactérias, para o estudo comparativo, listadas na Tabela 6. O critério adotado para a seleção foi o percentual de identidade entre as seqüências de aminoácidos das proteínas/enzimas envolvidas na síntese de celulose e a existência de comprovação experimental de produção de celulose pela bactéria correspondente. Foram selecionadas as proteínas/enzimas codificadas pelos genes do operon *bcs* das bactérias *A. xylinum* (M37202) (A), *A. xylinum* (X54676) (B), *A. tumefaciens* (L38609), *R. leguminosarum* (AF121340), *S. typhimurium* (NC_003197), *E. coli* (NC_000913), *P. putida* (NC_002947) e *P. fluorescens* (AY074776). A seqüência de aminoácidos de cada proteína codificada pelos genes do operon *bcs* de *C. violaceum*, juntamente com a seqüência de aminoácidos correspondente às proteínas das bactérias selecionadas, foram submetidas ao programa Clustal W para alinhamento múltiplo. O resultado do alinhamento nos permitiu identificar regiões conservadas entre as espécies que correspondem aos domínios e motivos relatados na literatura (SAXENA *et al.*, 1995; RÖMLING, 2000).

Tabela 6 – Bactérias selecionadas para estudo comparativo. Número de acesso no GenBank de cada proteína/enzima do operon *bcs* das bactérias selecionadas e o nome do gene/ORF destas proteínas/enzimas.

Organismo/ No. Acesso GenBank	No. Acesso GenBank	Gene/ORF
<i>Acetobacter xylinum</i> , M37202, (A)	AAA21884	<i>bcsA</i>
	AAA21885	<i>bcsB</i>
	AAA21886	<i>bcsC</i>
	AAA21887	<i>bcsD</i>
<i>Acetobacter xylinum</i> , X54676, (B)	CAA38487	<i>acsA</i> (<i>bcsA</i>)
	CAA38488	<i>acsB</i> (<i>bcsB</i>)
	CAA38489	<i>acsC</i>
	CAA38490	<i>acsD</i>
<i>Agrobacterium tumefaciens</i> , L38609	AAC41435	<i>cel A</i> (<i>bcsA</i>)
	AAC41434	<i>cel B</i> (<i>bcsB</i>)
	AAC41433	<i>cel C</i> (<i>bcsZ</i>)
	AAC41431	<i>cel D</i>
	AAC41432	<i>cel E</i>
<i>Rhizobium leguminosarum</i> , AF121340	AAD28574	<i>cel A</i> (<i>bcsA</i>)
	AAD28575	<i>cel B</i> (<i>bcsB</i>)
	AAD28576	<i>cel C</i> (<i>bcsZ</i>)
<i>Salmonella typhimurium</i> , LT2 NC_003197	NP_462521	<i>yhjQ</i>
	NP_462520	<i>yhjO</i> (<i>bcsA</i>)
	NP_462519	<i>yhjN</i> (<i>bcsB</i>)
	NP_462518	<i>yhjM</i> (<i>bcsZ</i>)
	NP_462517	<i>yhjL</i> (<i>bcsC</i>)
<i>Escherichia coli</i> , K12 NC_000913	NP_417991	<i>yhjQ</i>
	NP_417990	<i>yhjO</i> (<i>bcsA</i>)
	NP_417989	<i>yhjN</i> (<i>bcsB</i>)
	NP_417988	<i>yhjM</i> (<i>bcsZ</i>)
	YP_026226	<i>yhjL</i> (<i>bcsC</i>)
<i>Pseudomonas putida</i> , KT2440 NC_002947	NP_744778	PP2634
	NP_744779	PP2635 (<i>bcsA</i>)
	NP_744780	PP2636 (<i>bcsB</i>)
	NP_744781	PP2637 (<i>bcsZ</i>)
	NP_744782	PP2638 (<i>bcsC</i>)
<i>Pseudomonas fluorescens</i> , OprQ AY074776	AAL71850	<i>wssA/wssJ</i>
	AAL71842	<i>wssB</i> (<i>bcsA</i>)
	AAL71843	<i>wssC</i> (<i>bcsB</i>)
	AAL71844	<i>wssD</i> (<i>bcsZ</i>)
	AAL71845	<i>wssE</i> (<i>bcsC</i>)

A seguir serão apresentadas as características estruturais de cada proteína/enzima codificada pelos genes que compõem o operon *bcs* de *C. violaceum*: ORF CV2679, *BcsA*, *BcsB*, *BcsZ* e *BcsC*.

4.1.3 Características estruturais de cada proteína/enzima codificada pelos genes que compõem o operon *bcs* de *Chromobacterium violaceum*

4.1.3.1 ORF CV2679

O gene *YhjQ* em *S. typhimurium* LT2 e *E. coli* K12 faz parte do operon *bcs*, localiza-se na região a montante do operon e codifica a proteína de mesmo nome. Estudo realizado por KIM *et al.* (2002) de inativação desse gene teve como resultado o retardo na divisão celular e a produção de células filamentosas de vários tamanhos. Um estudo mais aprofundado da função deste gene poderia fornecer indicações da relação entre a síntese de celulose e o crescimento celular.

A ORF CV2679 de *C. violaceum* encontra-se na região a montante do operon *bcs*. Pesquisa realizada utilizando o programa BLASTP revelou que esta ORF é homóloga do gene *yhjQ* de *S. typhimurium* e *E. coli* com 31%/44% e 27%/46% de identidade/similaridade, respectivamente. A Figura 18 apresenta o resultado do alinhamento Clustal W realizado entre as seqüências de aminoácidos da ORF CV2679 de *C. violaceum* (NP_902349), *S. typhimurium* (NP_462521) e *E. coli* (NP_417991). Podem ser observados nos primeiros 70 resíduos de aminoácidos do alinhamento o seguinte padrão conservado: RGGXGXTX₃AAXAWX₉LX₂DX₂PXNLLRX₂FXVX₆GWX₆GXDW.

Um esquema gráfico dos domínios da proteína *YhjQ* de *S. typhimurium* obtido do PFAM é apresentado na Figura 19. O esquema gráfico apresenta um domínio *YhjQ* que compreende a região dos resíduos 66-250, um domínio PFAM-B na região N-terminal, correspondendo aos resíduos de 1-65 e um motivo de baixa complexidade na região dos resíduos 15-26. O domínio *YhjQ* corresponde a uma família de proteínas *YhjQ*, sendo que a função dessas

proteínas ainda permanece desconhecida (THE PFAM DATABASE OF PROTEIN FAMILIES AND HMMS, 2004). As regiões de baixa complexidade parecem representar regiões funcionalmente importantes em algumas proteínas, porém com poucas informações conhecidas (WAN e WOOTTON, 2000).

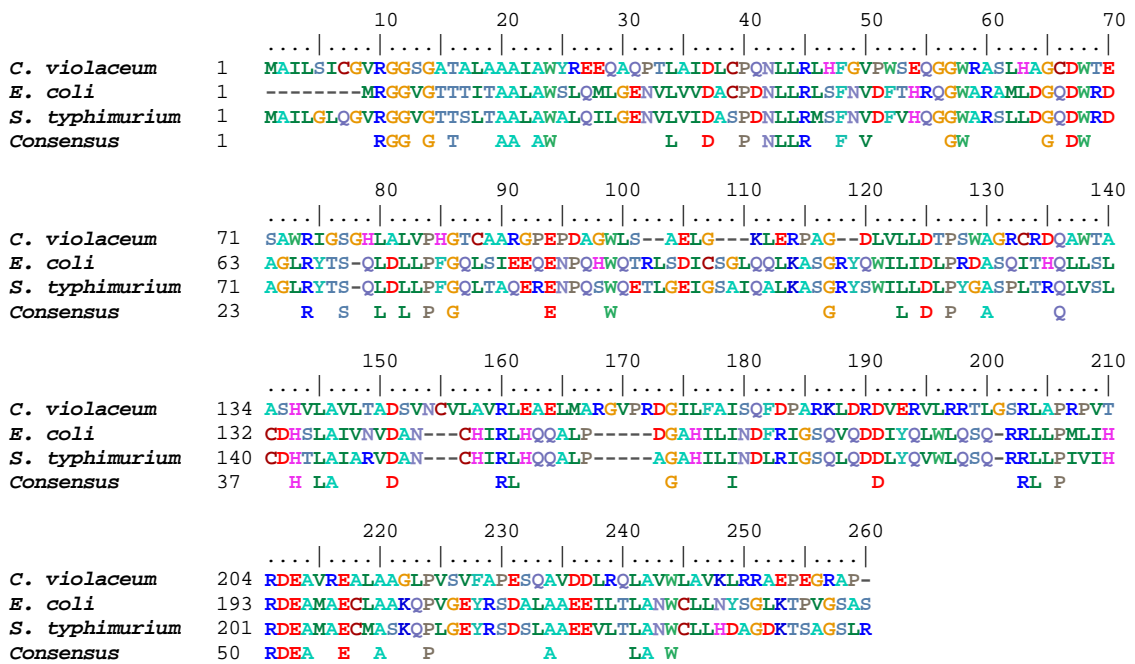


Figura 18 - Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da ORF CV2679 de *C. violaceum* (NP_902349), *S. typhimurium* (NP_462521) e *E. coli* (NP_417991).

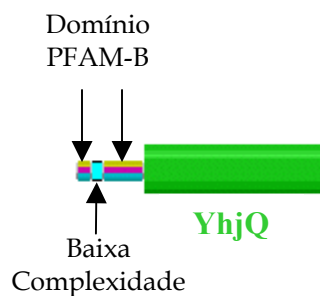


Figura 19 - Esquema gráfico dos domínios da proteína YhjQ de *S. typhimurium* obtido do PFAM. O número de acesso no SwissPfam é YHJQ_SALTY.

A análise baseada na identidade entre as seqüências sugere que a ORF CV2679 faz parte do operon *bcs* da *C. violaceum*, sendo similar ao gene *yhjQ* de *S. typhimurium* e *E. coli*, embora tenha sido anotada como uma proteína hipotética conservada. Esta análise sugere fortemente que este é o gene *yhjQ* cuja função está relacionada à diferenciação celular.

4.1.3.2 Proteína BcsA: celulose sintase

A proteína BcsA, celulose sintase é a subunidade catalítica da celulose sintase pertencente à família 2 das β -glicosiltransferases processivas. A Figura 20 indica que a celulose sintase de *C. violaceum*, BcsA(Cv), enquadra-se perfeitamente no modelo apresentado por BROWN *et al.* (1996), onde estão definidos dois domínios, A e B, e determinados o número de resíduos de aminoácidos entre os resíduos catalíticos e o motivo conservado QXXRW.

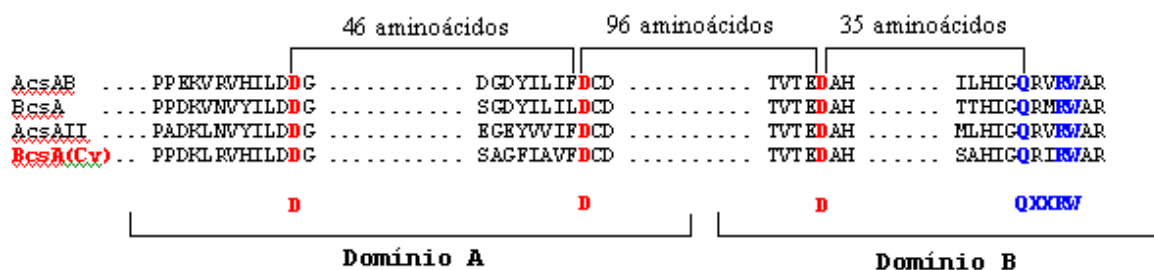


Figura 20 – Modelo da região catalítica da celulose sintase apresentado por BROWN *et al.* (1996), a partir da seqüência de aminoácidos de proteínas BcsA de três linhagens de *A. xylinum*, mostrando os domínios A e B, conforme descrito no texto. Ao modelo foi inserida a proteína BcsA de *C. violaceum*, (que como pode ser observado na Figura 21 apresentada a seguir), está de acordo com o mesmo número de resíduos entre os resíduos catalíticos D e motivo conservado QXXRW.

A proteína BcsA de *C. violaceum* possui 852 aminoácidos, uma massa molecular de 94,9 kDa, conforme mostrado na Figura 17. O resultado do

programa BLASTP mostrou identidade/similaridade da seqüência de aminoácidos da proteína BcsA de *C. violaceum* de 43%/59% com *A. xylinum* (A) (AAA21884), 43%/61% com *A. xylinum* (B) (CAA38487), 34%/50% com *A. tumefaciens* (AAC41435), 32%/48% com *R. leguminosarum* (AAD28574), 45%/61% com *S. typhimurium* LT2 (NP_462520), 46%/62% com *E. coli* K12 (NP_417990), 48%/61% com *P. putida* KT2440 (NP_744779), e 43%/64% com *P. fluorescens* OprQ (AAL71842).

Do alinhamento obtido através do programa Clustal W para as proteínas BcsA deste estudo, mostrado na Figura 21, podemos observar que a estrutura da celulose sintase de *C. violaceum* está de acordo com a arquitetura multidomínio apresentada por SAXENA e colaboradores (1995). O motivo D,D,D35QRXRW que prediz uma celulose sintase está presente na celulose sintase de *C. violaceum*, inclusive com o mesmo número de resíduos entre os resíduos catalíticos D,D,D e motivo conservado QXXRW.

Além dos domínios e motivos citados acima, que caracterizam a proteína como uma celulose sintase, outras seqüências conservadas estão descritas na literatura e são consideradas relevantes para a identificação da função da proteína. Como já citado na sessão 2.3.3.2, RÖMLING (2002) relata cinco regiões altamente conservadas entre as β -glicosiltransferases processivas, contendo os seguintes motivos: PVDPYE, HAKAGN, QTP, FFCGS e RFLPL. Constatou-se ainda que a curdlana sintase possui os motivos D,D,D35Q(R,Q),XRW, PVDPYE, HAKAGN e QTP, mas não possui os motivos FFCGS e RFLPL, o que sugere que os motivos FFCGS e RFLPL são responsáveis pela determinação da especificidade das ligações $\beta(1\rightarrow4)$ glicana. Como pode ser observada na Figura 21, a celulose sintase de *C. violaceum* possui todos os motivos citados.

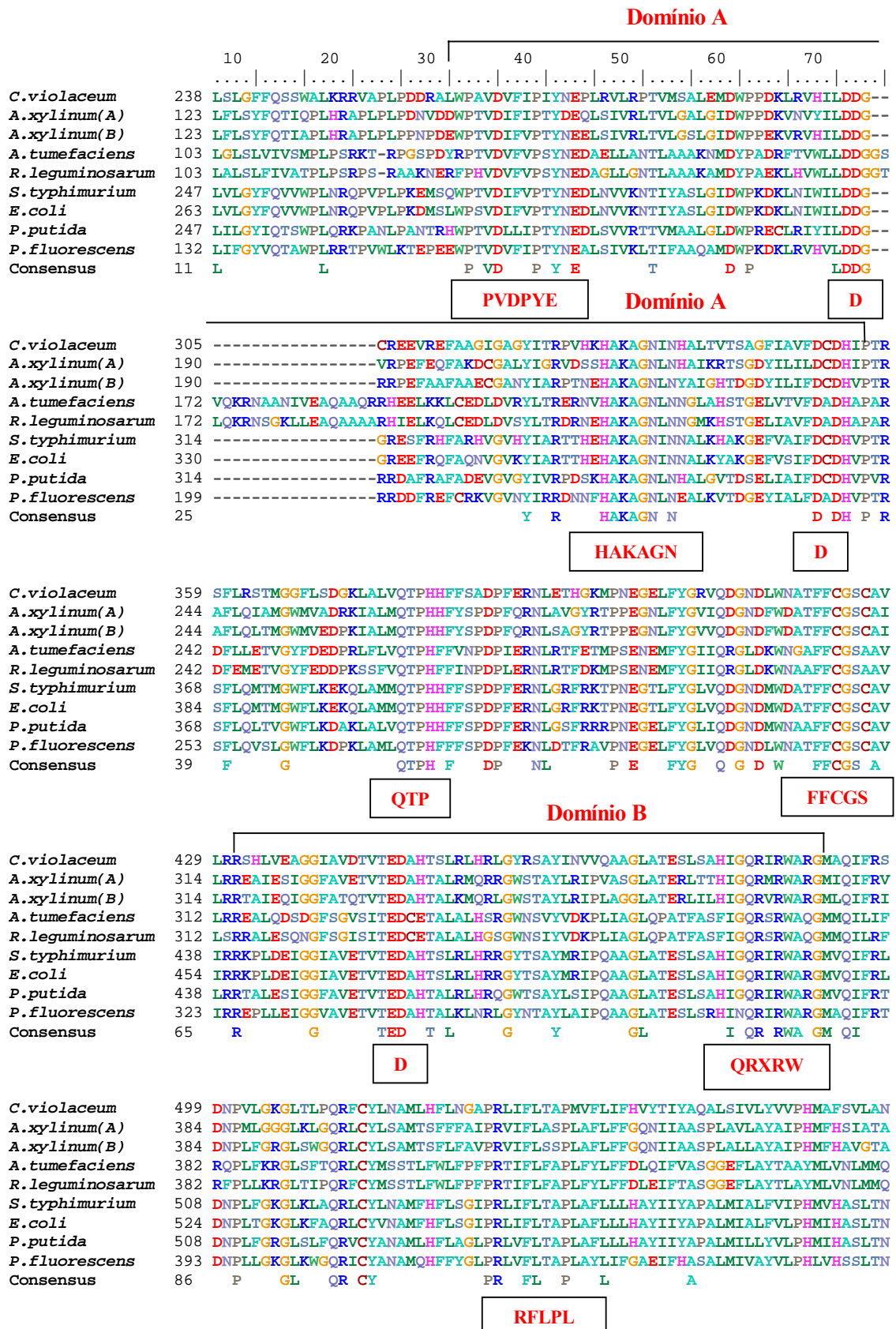


Figura 21 - Regiões conservadas obtidas do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsA de *C. violaceum* (NP_902348), *A. xylinum* (A) (AAA21884), *A. xylinum* (B) (CAA38487), *A. tumefaciens* (AAC41435), *R. leguminosarum* (AAD28574), *S. typhimurium* (NP_462520), *E. coli* (NP_417990), *P. putida* (NP_744779), e *P. fluorescens* (AAL71842).

A Figura 22 apresenta um esquema gráfico dos domínios da proteína BcsA de *C. violaceum* obtido do PFAM. O esquema gráfico apresenta um domínio Glycos_transf_2, referente à proteína da família glicosiltransferase_2, e também nove domínios transmembrana, um domínio peptídeo sinal e dois domínios de baixa complexidade.

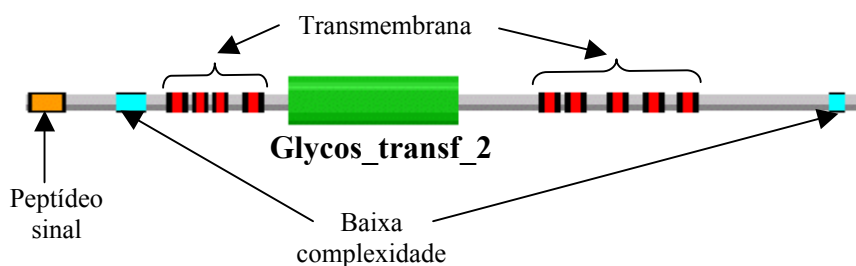


Figura 22 – Esquema gráfico dos domínios da proteína BcsA de *C. violaceum* obtido do PFAM. Q7NUL9 é o número de acesso no SwissPfam.

O domínio Glycos_transf_2 é encontrado em diversas famílias de glicosiltransferases que transferem o açúcar da molécula de UDP-glicose, de UDP-N-acetil-galactosamina, GDP-manose ou CDP-abequose, a uma variedade de substratos incluindo a celulose (CAMPBELL *et al.*, 1997).

O domínio transmembrana indica que a proteína está localizada na membrana citoplasmática e pode ser predito considerando-se a polaridade, carga e posição dos aminoácidos (KROGH *et al.*, 2001). Segundo RÖMLING (2002), as celulosas sintases possuem de oito a dez domínios transmembrana. A celulose sintase de *C. violaceum*, como podemos observar na Figura 22, possui nove domínios transmembrana.

Cada proteína carrega em sua estrutura a informação necessária específica para sua localização na célula. Seqüências específicas de aminoácidos (peptídeo sinal) determinam se a proteína passará através da membrana para o interior de uma organela em particular, se será integrada à membrana, ou se será exportada da célula (BLOBEL, 1999; NIELSEN *et al.*, 1997). O domínio

peptídeo sinal presente em celulose sintase, neste caso, sinaliza a proteína deve se integrar na membrana.

4.1.3.3 Proteína BcsB: proteína de ligação de c-di-GMP

A proteína BcsB e a proteína BcsA estão presentes em todos os operon *bcs* analisados neste trabalho. A proteína BcsB é considerada a subunidade de maior interação com a subunidade catalítica da celulose sintase, e também é considerada a subunidade de regulação da síntese de celulose por ligar-se ao ativador alostérico da celulose sintase, c-di-GMP. Porém, ela é a menos caracterizada do operon *bcs*, é a única que não possui EC, e também não possui registro no banco de dados COG.

A proteína BcsB de *C. violaceum* possui 738 aminoácidos e uma massa molecular de 79,3 kDa, conforme mostrado na Figura 17. O programa BLASTP apresentou uma identidade/similaridade da seqüência de aminoácidos da proteína BcsB de *C. violaceum* de 29%/48% com *A. xylinum* (A) (AAA21885), 35%/54% com *S. typhimurium* LT2 (NP_462519), 35%/55% com *E. coli* K12 (NP_417989), 38%/54% com *P. putida* KT2440 (NP_744780), e 37%/54% com *P. fluorescens* OprQ (AAL71843). As demais proteínas BcsB das espécies selecionadas para este estudo não possuem identidade com a proteína BcsB de *C. violaceum*. A Figura 23 apresenta a região mais conservada obtida do resultado do alinhamento múltiplo realizado entre as seqüências de aminoácidos das proteínas BcsB de *S. typhimurium* LT2, *E. coli* K12, *P. putida* KT2440, e *P. fluorescens* OprQ, que possuem maior identidade com a proteína BcsB de *C. violaceum*.

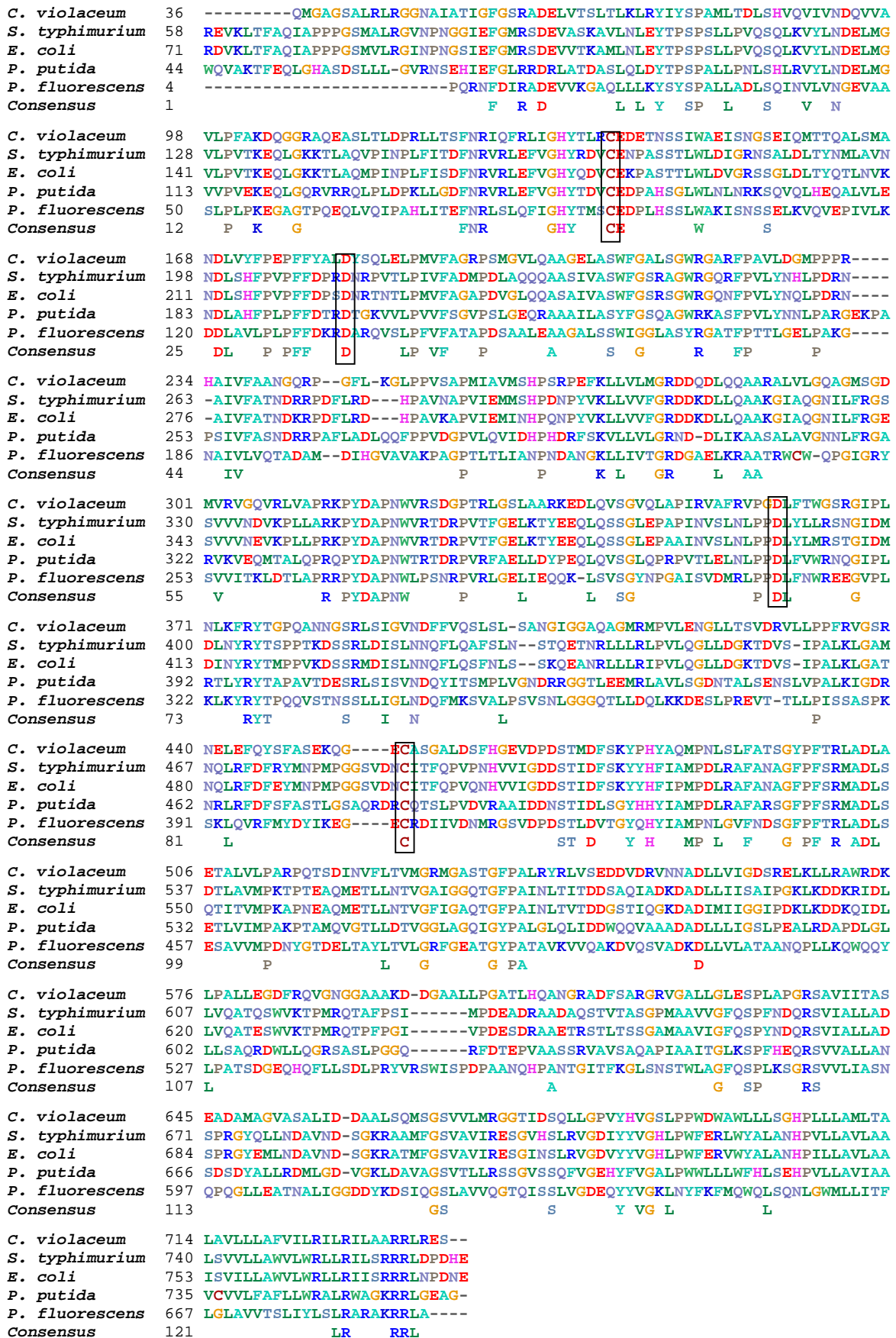


Figura 23 - Regiões conservadas resultantes do alinhamento Clustal W da sequência de aminoácidos da proteína BcsB de *C. violaceum* (NP_902347), *S. typhimurium* LT2 (NP_462519), *E. coli* (NP_417989), *P. putida* (NP_744780), e *P. fluorescens* (AAL71843). Os retângulos salientam os resíduos de cisteína e aspartato conservados entre as espécies.

RÖMLING (2002) realizou uma análise comparativa entre a proteína BcsB de *S. typhimurium* (CAC44016.1) e as proteínas BcsB (CelB) de *A. tumefaciens* e *R. leguminosarum* bv. *trifolii*. Na Figura 24 apresenta-se uma síntese de seus resultados. A região N-terminal é definida como domínio PA, e ao longo da seqüência são salientados os resíduos de cisteína e aspartato conservados entre as espécies nesta região.

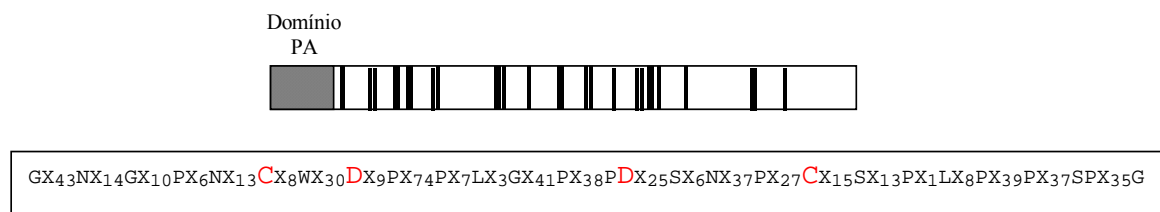


Figura 24 - Resíduos de aminoácidos conservados entre as proteínas BcsB (CelB) de *A. tumefaciens*, *R. leguminosarum* bv. *trifolii* e *S. typhimurium* (CAC44016.1) obtido do estudo realizado por RÖMLING (2002), sendo X_n é qualquer resíduo de aminoácido (o índice n representa o número de repetições do resíduo X).

Através do alinhamento das seqüências de diversas BcsB e a de *C. violaceum* mostrado na Figura 25 observa-se uma região N-terminal rica em alanina, descrita por RÖMLING (2002) como domínio PA. Porém em *P. fluorescens* a região N-terminal região não apresenta o domínio PA. De acordo com RAWLINGS e BARRETT (1993) o domínio PA está associado à atividade de protease. Os resíduos conservados de cisteína e aspartato ressaltados por RÖMLING (2002) também estão presentes nas bactérias deste estudo, porém o primeiro resíduo de aspartato encontra-se deslocado duas posições, conforme pode ser observado na Figura 23 apresentada anteriormente.



Figura 25 - Região N-terminal obtida do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsB de *C. violaceum* (NP_902347), *S. typhimurium* LT2 (NP_462519), *E. coli* (NP_417989), *P. putida* (NP_744780), e *P. fluorescens* (AAL71843).

Um esquema gráfico dos domínios da proteína BcsB de *C.violaceum* obtido do PFAM é apresentado na Figura 26. O esquema gráfico apresenta um domínio BcsB, um domínio transmembrana na região C-terminal, um domínio peptídeo sinal na região N-terminal, e dois domínios de baixa complexidade, um em cada extremidade. O domínio PFAM BcsB na proteína BcsB de *C. violaceum* localiza-se na região entre os resíduos 24 e 642.

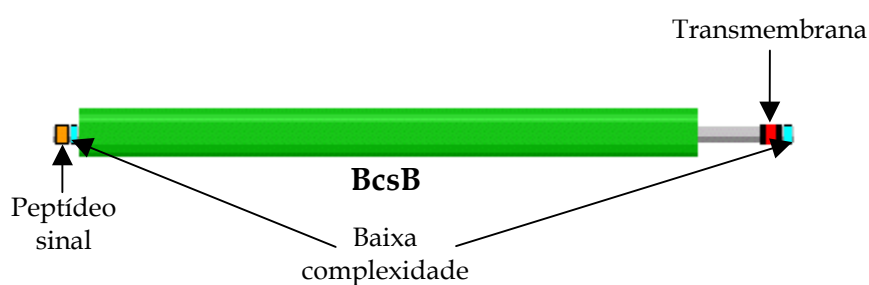


Figura 26 – Esquema gráfico dos domínios da proteína BcsB de *C. violaceum* obtido do PFAM. O número de acesso no SwissPfam é Q7NUM0.

4.1.3.4 Proteína BcsZ: celulase

A proteína BcsZ é uma endoglicanase, mais comumente conhecida como celulase. Existem várias hipóteses para seu papel na síntese de celulose, tanto em vegetais quanto em bactérias, porém sua função permanece pouco esclarecida. A proteína BcsZ (celulase) de *C. violaceum* possui 382 aminoácidos, uma massa molecular de 41,8 kDa, conforme mostrado na Figura 17. O resultado do programa BLASTP apresentou identidade/similaridade da seqüência de aminoácidos da proteína BcsZ de *C. violaceum* (NP_902346) de 26%/42% com *A. tumefaciens* (AAC41433), 28%/41% com *R. leguminosarum* (AAD28576), 47%/58% com *S. typhimurium* LT2 (NP_462518), 47%/58% com *E. coli* K12 (NP_417988), 46%/56% com *P. putida* KT2440 (NP_744781), e 35%/48% com *P. fluorescens* OprQ (AAL71844).

A Figura 27 apresenta a região mais conservada obtida do resultado do alinhamento múltiplo realizado entre as seqüências de aminoácidos das proteínas BcsZ de *A. tumefaciens*, *R. leguminosarum*, *S. typhimurium* LT2, *E. coli* K12, *P. putida* KT2440, e *P. fluorescens* OprQ, que possuem maior identidade com a proteína BcsZ de *C. violaceum*.

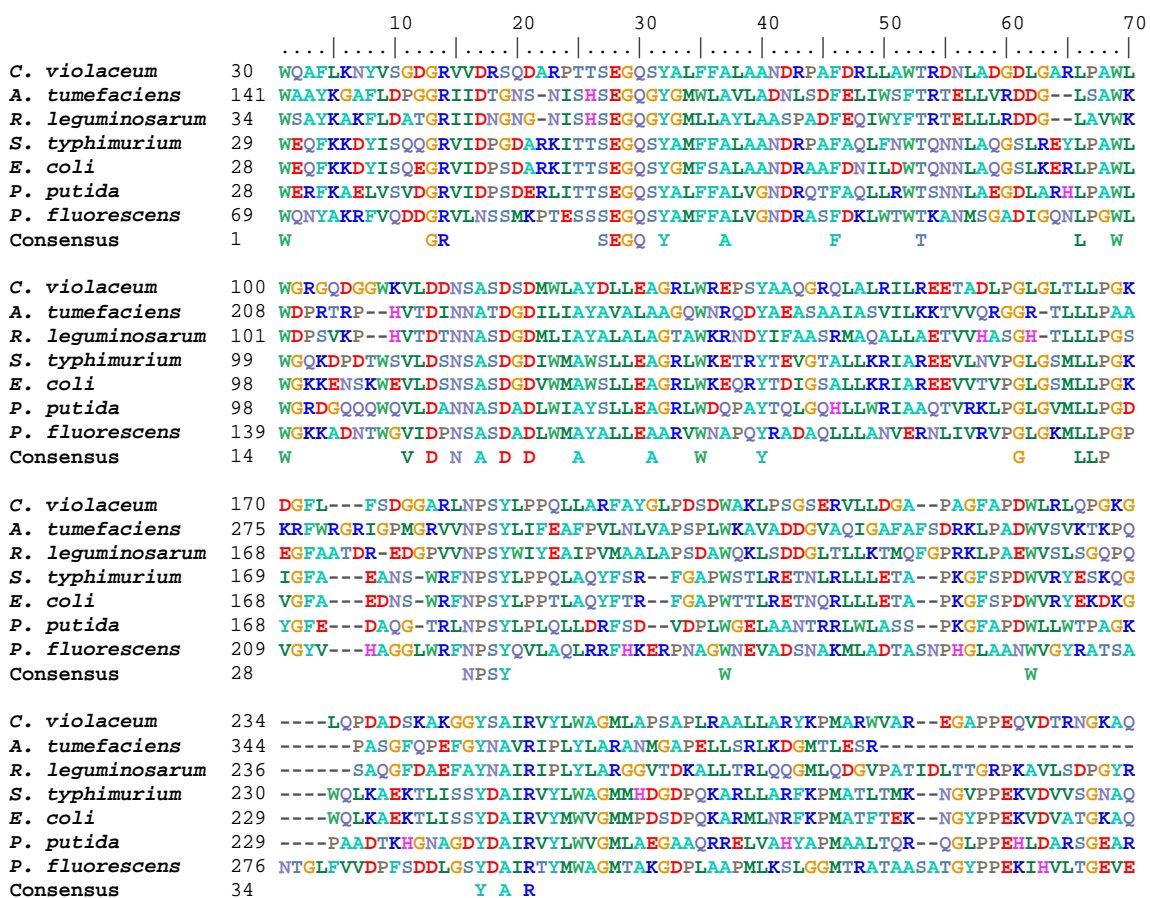


Figura 27 - Regiões conservadas resultantes do alinhamento Clustal W da seqüência de aminoácidos da proteína BcsZ de *C. violaceum* (NP_902347), *A. tumefaciens* (AAC41433), *R. leguminosarum* (AAD28576), *S. typhimurium* LT2 (NP_462518), *E. coli* (NP_417988), *P. putida* (NP_744781), e *P. fluorescens* (AAL71844).

Um esquema gráfico dos domínios da proteína BcsZ de *C. violaceum*, obtido do PFAM é apresentado na Figura 28. O esquema gráfico apresenta um domínio Glyco_hydro_8 na região dos resíduos 4-356, e um domínio peptídeo sinal na região N-terminal. O domínio Glyco_hydro_8 compreende as enzimas

da família 8 de glicosil hidrolases que hidrolisam ligações glicosídicas, sendo a mais conhecida entre elas as endogliconases (celulases).



Figura 28 - Esquema gráfico dos domínios da proteína BcsZ de *C. violaceum* obtido do PFAM. O número de acesso no SwissPfam é Q7NUM1.

4.1.3.5 Proteína BcsC

A proteína BcsC está localizada na região terminal a jusante de vários operons bacterianos responsáveis pela biossíntese da celulose. Sua função ainda permanece desconhecida. Uma possível função desta proteína no operon *bcs*, considerando as características estruturais dos motivos TPR presentes, poderia ser a de manter unida a associação de proteínas do operon *bcs*. Conforme discutido na sessão 2.3.3.5.

Um esquema gráfico dos domínios da proteína BcsC de *C. violaceum* obtido do PFAM é apresentado na Figura 29. O esquema gráfico apresenta um domínio BCSC_C na região C-terminal correspondendo aos resíduos de aminoácidos 910-1253, sete motivos TPR localizados nas regiões dos resíduos de aminoácidos 55-88, 300-333, 377-410, 454-487, 600-625, 713-746 e 816-849, e nove motivos de baixa complexidade.

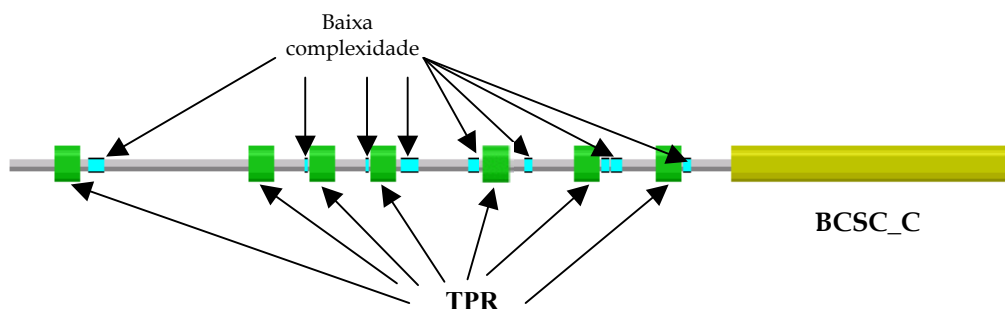


Figura 29 - Esquema gráfico dos domínios da proteína BcsC de *C. violaceum* obtido do PFAM. O número de acesso no SwissPfam é Q7NUM2.

A proteína BcsC de *C. violaceum* possui 1270 resíduos, uma massa molecular de 138,0 kDa, conforme mostrado na Figura 17. Esta proteína possui o maior número de resíduos de aminoácidos entre as proteínas que compõe o operon *bcs*. O resultado do programa BLASTP apresentou uma identidade/similaridade da seqüência de aminoácidos da proteína BcsC de *C. violaceum* (NP_902345) de 29%/50% com *A. xylinum* (A) (AAA21886), 30%/49% com *A. xylinum* (B) (CAA38489), 35%/56% com *S. typhimurium* LT2 (NP_462517), 33%/54% com *E. coli* K12 (YP_026226), 34%/48% com *P. putida* KT2440 (NP_744782), e 32%/50% com *P. fluorescens* OprQ (AAL71845). A Figura 30 apresenta o alinhamento múltiplo realizado entre as seqüências de aminoácidos das proteínas BcsC de *C. violaceum*, *A. xylinum* (A), *A. xylinum* (B), *S. typhimurium* LT2, *E. coli* K12, *P. putida* KT2440, e *P. fluorescens* OprQ, somente a região que correspondente ao domínio BCSC_C. Como pode ser observado é uma região com grande número de glicinas conservadas.

4.1.4 Predição de região regulatória de transcrição do operon *bcs* de *Chromobacterium violaceum*

As prováveis regiões regulatórias foram obtidas da anotação automática do BRGene. A Figura 31 mostra as possíveis regiões Box -35 e Box -10, bem como a possível região Shine-Dalgarno(SD).

			10	20	30	40	50	60	70						
<i>C. violaceum</i>	901		RGPS	DGTPGASQ	MTLWQTPL	LLLEKAA	PGNGHY	FFRGEAG	SVSAGSL	DLGPGND	YTLNRFG	SVAACAP	SAS		
<i>A. xylinum</i> (A)	920		FRSRS	GGHGMGR	LTLEAN	IPIVGR	RLPLQ	-AGASAL	TFSITPT	MIWSGN	LNTG	SVYDV	VPVRYG	TMMGVQ	AYNQ
<i>A. xylinum</i> (B)	934		FRVRS	GGTPGM	GALTEA	SVPIVGR	RIPLQ	-AGTSAL	TFTATPT	FLTS	GHLPO	-TGYD	IPRFG	TNLFAL	ERNL
<i>S. typhimurium</i>	816		WG-	SSGTGG	YSDLKA	HITMLQ	VDAPWS	-DGRAFF	RITDMV	NMDVGR	FSTDA	DGKYDN	-NWGTC	-----	
<i>E. coli</i>	776		WG-	SSGTGG	YSDLKA	HITMLQ	VDAPYS	-DGRMFF	RSDFV	NMNVSF	STNADG	KWDD	-NWGTC	-----	
<i>P. putida</i>	796		WREDD	GTPGTS	ELSTQ	STLLH	VDTPWQ	-EGTAF	ARVERI	GMDAGS	FEPND	QGSYTP	-DFGSCQ	FVVGQ	TAD
<i>P. fluorescens</i>	924		VRSNN	GKGLS	KITDVE	APFEAR	MPVG	-DNTVAL	LRVTPV	HLSAGS	VKAESL	SRFGK	-----		
Consensus	37			G	G										
<i>C. violaceum</i>	971		PQACA	AAAYGA	QQRARG	-----	VALGGG	YESDD	WRFD	IGVTP	LGFP	VSNI	VGGARR	TGDL	GP--
<i>A. xylinum</i> (A)	989		YDSYTN	NAGRD	QQRIA	AGTA	EAGFAP	DVQF	NSWVR	ADVGAS	PIGFP	ITNVL	GGVEF	SPRV	GP--
<i>A. xylinum</i> (B)	1002		QNQNN	SAEHR	INTD	TIG-RE	AGVAP	DVRFAN	NWVSAD	VGAS	PLGFT	LPNV	IGGVE	FAPRV	GP--
<i>S. typhimurium</i>	874		--TLEK	CSGHRS	QADTG	-----	ASVAVG	WQNET	WRWD	IGTTP	MGFN	VVDV	VGGV	SYSD	DIGP
<i>E. coli</i>	834		--TLQD	CSGNRS	QSDSG	-----	ASVAVG	WRND	VWSWD	IGTTP	MGFN	VVDV	VGGV	SYSD	DIGP
<i>P. putida</i>	864		GKTL	PACTGG	-SQTANG	-----	SLALG	WQGS	RAFDL	GTT-Q	GYAIN	NWLGG	ATVNG	DLGQ	--
<i>P. fluorescens</i>	978		--GGT	EPA	SQSDSG	-----	LAVAF	ENPQ	GLKAD	VGVSP	LGFLY	NLTV	GGVSV	SRPFE	ANSN
Consensus	39								D	G		G		GG	
<i>C. violaceum</i>	1031		ALARR	PLTSS	LLSYAG	VRPYAG	-----	QAWGG	VVATGL	GGGLY	DKGGS	FVWVS			
<i>A. xylinum</i> (A)	1056		SAERR	SITNS	VLSYGG	LRDTN	YNSAL	GRYAR	QVYV	QALS	SKWGS	EWGG	VVNH	PHGQ	VEATL
<i>A. xylinum</i> (B)	1068		SGERR	SITNS	VLSYGG	MTDALT	-----	GKKW	GGVTV	NHFG	QVEAT	LGNIT	IVYG		
<i>S. typhimurium</i>	935		NAHRR	PIS	SSLLAF	GGQK	DASSN	-----	TGTK	WGGV	RANG	GGVSL	SYDK	GEAN	VWA
<i>E. coli</i>	895		NAHRR	PIS	SSLLAF	GGQK	DSPSN	-----	TGKK	WGGV	RADG	VLSL	SYDK	GEAN	VWA
<i>P. putida</i>	924		TASRR	PMSN	LLSFAG	ARDP	-----	TGVR	WGGV	TANG	ATLGL	SWDQ	GGDN	GVWA	
<i>P. fluorescens</i>	1042		NISRR	PVTDS	VTSFAG	SEDGAGN	-----	KWGG	VTANG	RRGEL	SYDN	NQ-KL	GVYG		
Consensus	44		RR		S	G	D						W	G	G
<i>C. violaceum</i>	1081		NFAYQ	QLSG	ENVDS	NRKLTAM	GGVY	VRVDE	PDRL	LITAG	VNTV	NFWYR	KNLGG	FTFG	QGGY
<i>A. xylinum</i> (A)	1125		GGGYA	IQTG	KNVQR	NSEERE	EAGIGAN	TLVVH	NANML	VLRIG	VSLTY	FYGYAK	NEDFY	TYGQ	GGYF
<i>A. xylinum</i> (B)	1117		GGGYA	IQTG	HHVQS	NTEVE	GGLGAN	TLVVH	NANML	VLRIG	VSLTY	FYGYAK	NEDFY	TYGQ	GGYF
<i>S. typhimurium</i>	987		SLSGD	QLSG	KNVED	NWRV	WTGGY	YKVIN	ENNR	RVTV	GLNN	MHWYD	KLDS	GSYSL	GQGGY
<i>E. coli</i>	947		SLSGD	QLSG	KNVED	NWRV	WTGGY	YKVIN	ENNR	RVTV	GLNN	MHWYD	KLDS	GSYSL	GQGGY
<i>P. putida</i>	974		SLGH	HWLYG	ENVAD	NQRT	RAMAG	FYHR	VVEKA	DERV	RV	VTLM	HW	HKDL	GGYSL
<i>P. fluorescens</i>	1090		YASL	HELLI	GNVED	NTRLEL	SGSI	YWYLR	NNPR	DTL	TLG	ISGS	SAM	TFK	ENQDF
Consensus	53			G	V	N			G						G
<i>C. violaceum</i>	1151		SLPL	R	YAARN	ERWSY	FIRGA	ASFSS	AREDA	SPFYP	-----	TRPDL	QARAG	NPF	FAGS
<i>A. xylinum</i> (A)	1195		TVPV	RYAG	QH	KRLD	WDVT	GSVGY	QVFHE	HSSAP	FFP--	TSSLL	QSGAN	TIAS	SNYS
<i>A. xylinum</i> (B)	1187		TVPV	RYAG	QH	KRLD	WDVT	GSVGY	QVFHE	HSSAP	FFP--	TSSLL	QSGAN	TIAS	SNYS
<i>S. typhimurium</i>	1057		AVPVM	WRQ	TENWS	WELGGS	VSVS	WHSR	NR	TM	PRYP	---LM	NLIP	ADYQ	EDARD
<i>E. coli</i>	1017		AIPVM	WRQ	TENWS	WELGGS	VSVS	WHSR	NR	TM	PRYP	---LM	NLIP	ADYQ	EDARD
<i>P. putida</i>	1044		GVVPS	YAWR	NYD	WSLL	LEGS	VSVS	QAH	GSSR	LYPDA	HA	INR	KVLAN	YGV
<i>P. fluorescens</i>	1160		GVPI	RWAQ	SFDR	F	SYQV	KSSV	GLQH	IAQD	GADY	F	---	GDS	TLQ
Consensus	65														
<i>C. violaceum</i>	1213		GAF	EYQL	SPG	VVLGG	MLDL	QRSQ	YQPS	RLLF	YLR	YQPDG	-APRT	LPFP	VEPL
<i>A. xylinum</i> (A)	1262		SAYYP	GDSI	AGLT	GGFN	ARVGY	RFRN	VR	LDLS	GRYQ	KAGN	WTES	GAMI	SAHYL
<i>A. xylinum</i> (B)	1245		SARYP	GDDV	GS	LVGG	F	DGRV	G	YRV	SHSL	R	LDLS	GRFQ	KAGN
<i>S. typhimurium</i>	1123		ALI	ERRV	TAN	WFVGT	AVDI	QAKD	YTP	SH	LLLV	VRYS	AAG-W	QGD	MDL
<i>E. coli</i>	1083		ALL	ERRV	TAN	WFVGT	AVDI	QAKD	YTP	SH	LLLV	VRYS	AAG-W	QGD	MDL
<i>P. putida</i>	1114		GLF	FERR	LD	QWV	LG	GGFD	WQHS	DDY	AP	SH	GMLY	LR	YLF
<i>P. fluorescens</i>	1222		AAA	EYRL	SSR	FYLG	G	GEI	GLD	NAQ	DYR	QYAG	NA	YL	R
Consensus	67						G								R

Figura 30 – Regiões conservadas resultantes do alinhamento Clustal W da sequência de aminoácidos da proteína BcsC de *C. violaceum* (NP_902347), *A. xylinum*(A) (AAA21886), *A. xylinum*(B) (CAA38489), *S. typhimurium* (NP_462517), *E. coli* (YP_026226), *P. putida* (NP_744782), e *P. fluorescens* (AAL71845), que abrangem o domínio BCSC_C.

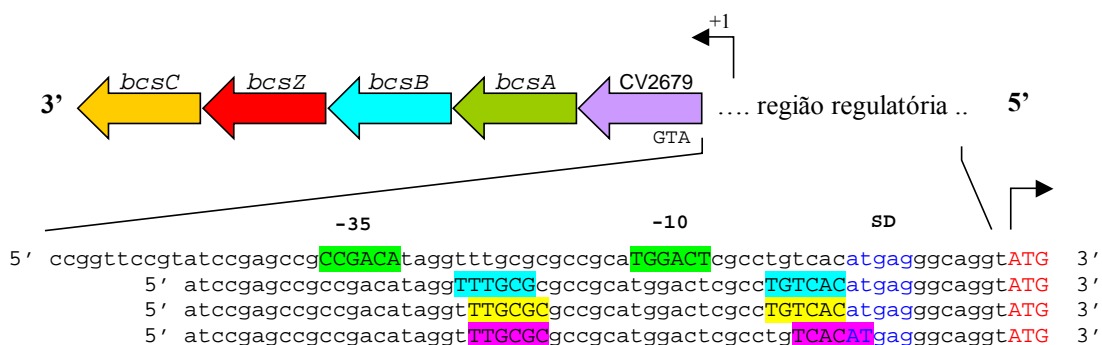


Figura 31 - Possíveis regiões regulatórias de transcrição do operon *bcs* de *C. violaceum*. Em destaque as prováveis Box -35 e -10 e a região SD.

4.1.5 Organização dos genes de biossíntese de celulose em diferentes grupos de bactérias Gram-negativas

A partir da investigação computacional é possível prever a presença de um operon com base na comparação em outros operons presentes e conservados entre genomas já seqüenciados. A aplicação desta técnica tornou-se possível devido ao rápido aumento da disponibilidade de dados de seqüências de procariotos (OVERBEEK *et al.*, 1999). Em todas as bactérias mencionadas neste estudo, os genes centrais necessários para a biossíntese de celulose, incluindo-se os da celulose sintase, formam um operon. Dentre estes, dois genes estão sempre presentes: o que codifica a proteína BcsA (celulose sintase - subunidade catalítica) e o que codifica a proteína BcsB (proteína de ligação do c-di-GMP) (SAXENA *et al.*, 1994; CHAO *et al.*, 2001; RÖMLING, 2002).

A Figura 32 ilustra a homologia entre os operons das diversas espécies aqui estudadas. Entre elas também se encontra o provável operon *bcs* de *Burkholderia fungorum* LB400 (NZ_AAAJ000000000) mais um representante das β -proteobactérias, recentemente seqüenciada. Esta bactéria não foi selecionada

para o estudo comparativo, embora seja a que apresenta maior identidade com as seqüências de aminoácidos das proteínas do operon *bcs* de *C. violaceum*, pelo fato de não existir citações na literatura de estudos experimentais da produção de celulose por este organismo e só recentemente seus dados de seqüenciamento foram publicados.

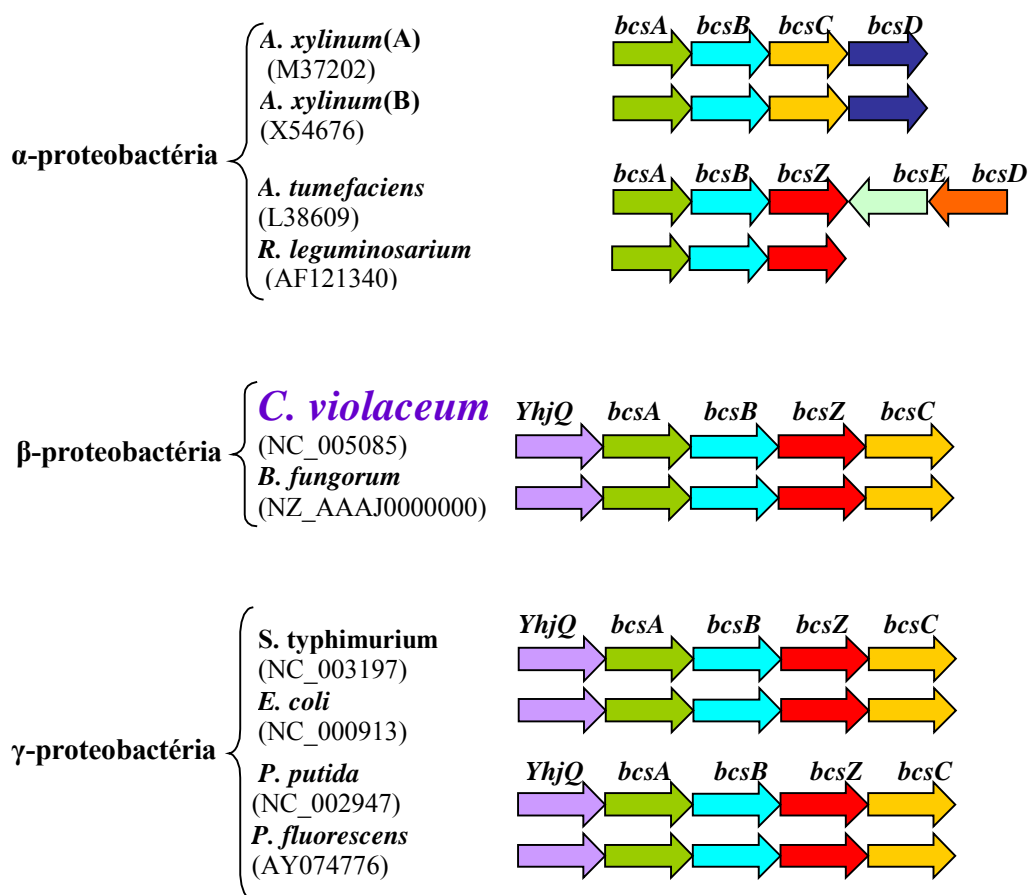


Figura 32 - Organização dos genes de biossíntese de celulose em diferentes grupos de bactérias Gram-Negativas.

Estes resultados sugerem que a celulose pode ser produzida em *C. violaceum*, e reforçam a idéia de que a celulose é um componente do biofilme produzido por esta bactéria. A literatura não descreve a produção experimental de celulose por *C. violaceum*, mas visto que os genes necessários para a sua

biossíntese de celulose estão presentes, e organizados em um operon *bcs*, isto, por si só, constitui uma forte evidência de sua produção.

4.1.6 Proteínas com motivo GGDEF em *Chromobacterium violaceum*

A análise *in silico* do genoma da *C. violaceum* revelou 43 ORFs, listadas na Tabela 7, com o domínio GGDEF. Os domínios GGDEF e EAL estão presentes em proteínas envolvidas na síntese e degradação de c-di-GMP. Uma função conhecida, como já citado anteriormente, da molécula c-di-GMP é a ativação da síntese de celulose em *A. xylinum*. Interessantemente, os domínios GGDEF e EAL estão presentes em proteínas de diversas espécies bacterianas, nem todas produtoras de celulose. Várias proteínas que possuem domínio GGDEF estão envolvidas em sistemas de transdução de sinal. A *C. violaceum* é uma das bactérias que atualmente possuem o maior número de proteínas com domínio GGDEF até agora seqüenciadas, só perdendo para *A. tumefaciens* e algumas *Vibrionaceae*, de acordo com informações obtidas em nossa pesquisa realizada no PFAM. A Figura 33 apresenta o alinhamento das representações gráficas da arquitetura dos domínios de proteínas com motivo GGDEF encontradas no genoma de *C. violaceum*. (As representações gráficas da arquitetura de domínios foram obtidas do PFAM).

Tabela 7 - Proteínas com domínio GGDEF em *C. violaceum*.

Swiss-Prot/TrEMBL	ORF	Accession	Definição
Q7NPW3	CV4379	NP_904049	proteína hipotética conservada
Q7NQA3	CV4237	NP_903907	provável proteína da família GGDEF
Q7NQQ2	CV4135	NP_903805	proteína hipotética conservada
Q7NQQ0	CV4087	NP_903757	proteína hipotética conservada
Q7NQU6	CV4041	NP_903711	proteína hipotética conservada
Q7NRP2	CV3738	NP_903408	proteína hipotética conservada
Q7NRP3	CV3737	NP_903407	proteína hipotética conservada
Q7NSC2	CV3504	NP_903174	provável regulação de resposta
Q7NSG5	CV3461	NP_903131	proteína hipotética conservada
Q7NSJ7	CV3427	NP_903097	proteína hipotética conservada
Q7NT05	CV3260	NP_902930	proteína hipotética conservada
Q7NT13	CV3252	NP_902922	proteína hipotética conservada
Q7NT69	CV3192	NP_902862	provável sensor de dois componentes
Q7NTH3	CV3082	NP_902752	provável proteína da família GGDEF
Q7NTV0	CV2953	NP_902623	proteína hipotética conservada
Q7NUX4	CV2573	NP_902243	provável proteína transmembrana
Q7NUX9	CV2568	NP_902238	proteína hipotética conservada
Q7NVN4	CV2308	NP_901978	proteína hipotética conservada
Q7NWX8	CV1851	NP_901521	proteína hipotética conservada
Q7NXF9	CV1667	NP_901337	proteína hipotética conservada
Q7NXX6	CV1620	NP_901290	proteína hipotética conservada
Q7NXX9	CV1587	NP_901257	provável proteína da família GGDEF
Q7NXP4	CV1582	NP_901252	provável proteína da família GGDEF
Q7NXP8	CV1578	NP_901248	provável proteína da família GGDEF
Q7NYA0	CV1374	NP_901044	proteína hipotética conservada
Q7NYE2	CV1332	NP_901002	proteína hipotética conservada
Q7NYX4	CV1148	NP_900818	proteína hipotética conservada
Q7NYX9	CV1143	NP_900813	proteína hipotética conservada
Q7NZ88	CV1034	NP_900704	proteína hipotética conservada
Q7NZ89	CV1033	NP_900703	proteína hipotética conservada
Q7NZF8	CV0984	NP_900634	proteína hipotética conservada
Q7NZT0	CV0841	NP_900511	proteína hipotética conservada
Q7NZV8	CV0812	NP_900482	proteína hipotética conservada
Q7P0M4	CV0542	NP_900212	proteína hipotética conservada
Q7P0N5	CV0531	NP_900201	proteína hipotética conservada
Q7P127	CV0387	NP_900057	proteína hipotética conservada
Q7P128	CV0386	NP_900056	proteína hipotética conservada
Q7P1A3	CV0310	NP_899980	proteína hipotética conservada
Q7P1H4	CV0239	NP_899909	proteína hipotética conservada
Q7P1K5	CV0208	NP_899878	proteína hipotética conservada
Q7P1H3	CV0204	NP_899874	proteína hipotética
Q7P1X1	CV0091	NP_899761	provável regulação de resposta
Q7P215	CV0046	NP_899716	proteína PleD



Figura 33 – Proteínas com motivo GGDEF em *C. violaceum*. Arquitetura de domínios obtidos do PFAM.

4.2 Produção de Biofilme Celulósico por *Chromobacterium violaceum*

Já há algum tempo tem sido relatada a capacidade da *C. violaceum* produzir EPS (MARTIN e RICHARDS, 1963; CORPE, 1964). Recentemente, estudo realizado por MARTINELLI *et al.* (2002) mostram que a produção de EPS por esta bactéria ocorre inclusive com a participação de processos de comunicação intercelular conhecidos como *quorum sensing*. Entretanto, nenhum estudo foi realizado até o momento para caracterizar o tipo de EPS produzido. A análise do genoma de *C. violaceum*, realizada nesta pesquisa, mostrou a presença de enzimas da via metabólica responsável pela biossíntese de celulose, evidenciando-se genes homólogos aos genes do organismo modelo de produção de celulose, a bactéria *A. xylinum*. Estes novos resultados caracterizam a *C. violaceum* como um microrganismo com capacidade potencial de produzir celulose.

4.2.1 Resultados do cultivo de *Chromobacterium violaceum*

Na Figura 34 (A) são mostradas triplicatas do cultivo estático após 1 dia de cultivo. A presença de um filme sobrenadante em todos os frascos comprova a reprodutibilidade da produção de biofilme em cultura estática. As Figura 34 (A, B, C, D e E) mostram a evolução da produção do biofilme, com seu espessamento ao longo dos dias. É notório que no sexto dia de cultivo a massa de biofilme sobrenadante é grande o suficiente para provocar sua precipitação para o fundo do frasco. Na Figura 34 (F) é mostrado o cultivo após 6 dias, porém em cultura agitada. Neste caso não se observa aparentemente a formação de biofilme. Este resultado é esperado, uma vez que a agitação impede a associação das bactérias numa matriz visível de EPS. Na Figura 34 (G) é mostrado que embora sob agitação não se note a formação do biofilme este de fato está presente, uma vez que a adição de etanol à cultura leva à agregação do biofilme antes disperso.

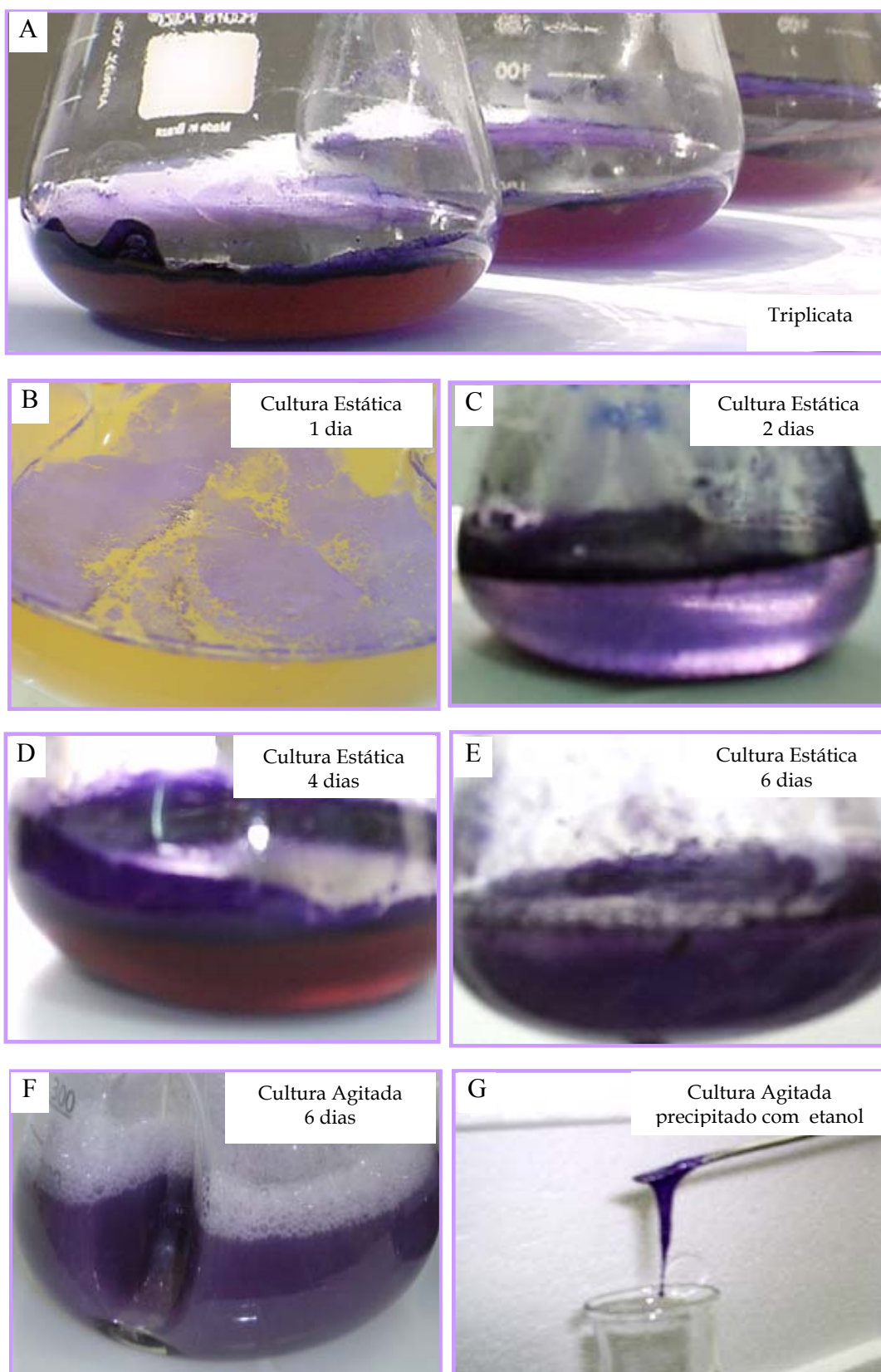


Figura 34 - Culturas estática e agitada realizadas com *C. violaceum* em meio LB. Coleta com 6 dias de cultivo e precipitação do meio líquido com etanol.

4.2.2 Concentração de glicose em amostras de exopolissacarídeo celulósico produzido por *Chromobacterium violaceum*

Supostamente se celulose fosse o polissacarídeo produzido, a ação da celulase sobre este levaria a liberação de unidades de glicose e esta poderia ser determinada no produto de hidrólise. Objetivando identificar a presença de celulose nos biofilmes produzidos por *C. violaceum*, estes foram isolados e tratados com celulase. Para a determinação da concentração de glicose foi realizada a hidrólise enzimática do material produzido por *C. violaceum*, em culturas estáticas e culturas agitadas, utilizando a enzima Celluclast 1,5 L (celulase). O procedimento adotado para a realização da hidrólise enzimática está descrito com detalhes em “Materiais e Métodos”.

Amostras do produto da hidrólise enzimática foram coletadas, e analisadas através do método enzimático colorimétrico do kit de glicose oxidase, objetivando a determinação da concentração de glicose. A Figura 35 mostra o resultado obtido deste ensaio. O mesmo resultado é apresentado na Tabela 8, na qual se apresenta a concentração de glicose das amostras.

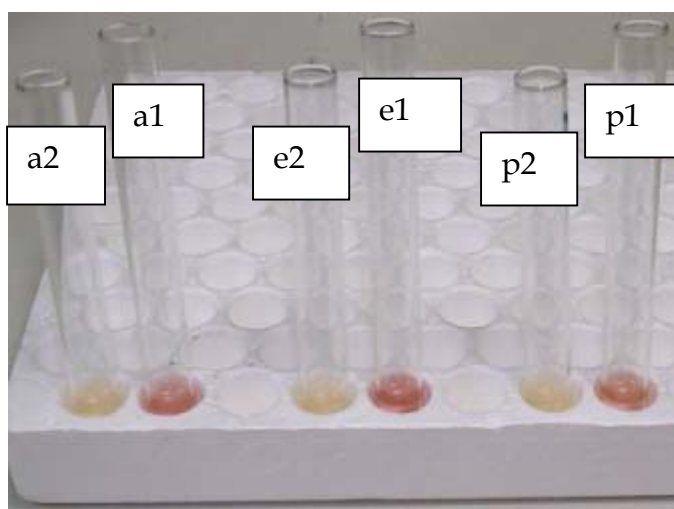


Figura 35 – Determinação da concentração de glicose em amostras do produto da hidrólise enzimática, utilizando o método enzimático colorimétrico do kit de glicose enzimática. a1, e1 e p1 são amostras tratadas com celulase, enquanto a2, e2, p2 são as mesmas amostras submetidas às condições de hidrólise na ausência de celulase.

A partir da leitura da absorbância das amostras e de uma solução padrão de glicose, cuja absorbância lida foi de 0,1918, foi possível calcular a concentração de glicose. A grande diferença de concentração de glicose apresentada entre as amostras que foram tratadas com celulase e as amostras onde a celulase não foi adicionada demonstra atividade da celulase sobre o substrato presente nas amostras. Esta é uma evidência clara de que no substrato existem ligações $\beta(1\rightarrow4)$ glicana, características do polímero celulose.

Tabela 8 - Concentração de glicose em amostras de exopolissacarídeos celulósico produzido por *C. violaceum*, em cultura estática e cultura agitada.

Amostra	Concentração de glicose (g/L) (amostras com enzima)	Concentração de glicose (g/L) (amostras sem enzima)
Película	0,4432	0,050
Precipitado da cultura estática	0,4505	0,051
Precipitado da cultura agitada	0,4406	0,050

A Figura 36 apresenta a concentração de glicose presente nas amostras do produto hidrolisado da cultura estática e cultura agitada de *C. violaceum*. Observe que as quantidades de glicose presente nas amostras tratadas com enzima celulase são praticamente as mesmas.

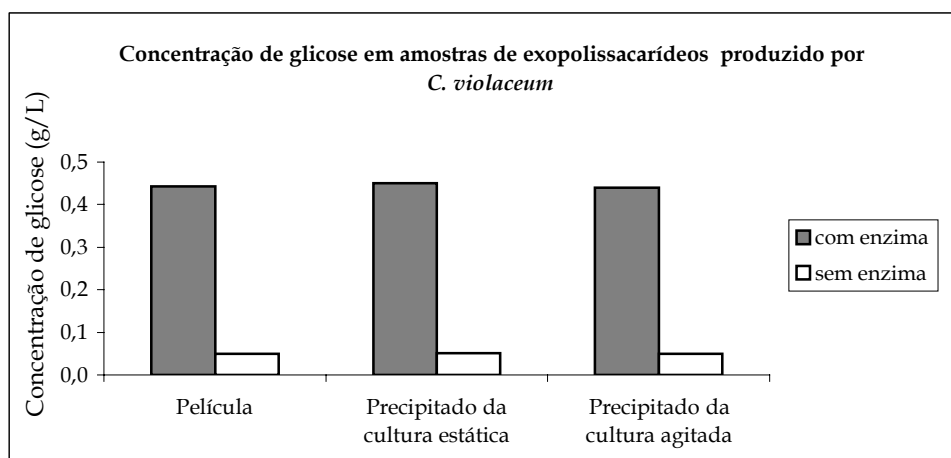


Figura 36 - Concentração de glicose em amostras de exopolissacarídeos celulósico produzido por *C. violaceum*, em cultura estática e cultura agitada.

Os resultados apresentados indicam que de fato celulose é constituinte do biofilme produzido por *C. violaceum*, uma vez que a atividade de celulase é específica o suficiente para garantir a presença neste EPS de ligações $\beta(1\rightarrow4)$ entre unidades de glicose. Deu-se preferência à utilização de um método de determinação específico para glicose, ao invés de um para determinação de açúcares redutores, o que garante a detecção da liberação especificamente de glicose a partir da hidrólise por celulase do EPS. Observa-se que a massa de biofilme produzida é sempre maior nas culturas estáticas, como era de se esperar (dados não mostrados).

Embora não seja o enfoque deste trabalho, observou-se que a biossíntese de celulose e violaceína ocorrem simultaneamente. A biossíntese de violaceína em *C. violaceum* está comprovado experimentalmente que é dependente da indução por homoserina lactona, molécula sinal. Pode-se, então, inferir que também a biossíntese de celulose em *C. violaceum* pode estar sob o mesmo controle. Assim, a biossíntese de celulose deve ocorrer como a de violaceína, em resposta, às condições ambientais de crescimento, relacionada à limitação de algum nutriente.

Em resumo, a investigação computacional realizada neste estudo, ao identificar por homologia o operon *bcs*, responsável pela biossíntese de celulose, e a estrutura das enzimas nele codificadas, fornecem um conjunto de evidências da produção de celulose por *C. violaceum*, nunca reportada experimentalmente. Além disso, os ensaios laboratoriais realizados mostraram a presença de glicose após a hidrólise com a enzima celulase do produto da cultura. Estes resultados sugerem fortemente a presença de celulose na matriz extracelular produzida por *C. violaceum* e incluem a *C. violaceum* como uma bactéria do grupo de bactérias produtoras de celulose.

CAPÍTULO V

Conclusões e Sugestões

5.1 Conclusões

A produção de celulose bacteriana está amplamente distribuída entre várias espécies de bactérias, incluindo pelo menos os grupos de proteobactérias Gram-negativas alfa, beta e gama. A produção de celulose, que se pensava anteriormente ser restrita ao organismo modelo *Acetobacter xylinum*, é evidenciada pelos diversos operons expressos em várias das bactérias estudadas. Esse grupo inclui a *Chromobacterium violaceum* e a *Burkholderia fungorum*, do grupo das β -proteobactérias.

Neste trabalho, a investigação computacional, realizada para identificar por homologia o operon *bcs* (*b*acterial *c*ellulose *s*ynthesis) responsável pela biossíntese de celulose e a estrutura das proteínas por ele codificadas, fornece um conjunto de evidências em favor da produção de celulose bacteriana por *C. violaceum*, nunca antes reportada experimentalmente.

A análise do genoma da *C. violaceum* revelou ainda o papel da ORF CV2679 como sendo um dos genes que formam o operon *bcs*. Esta ORF era até então reportada como uma proteína hipotética conservada, cuja função não era claramente conhecida. Esta ORF CV2679 é homóloga ao gene *yhjQ*, encontrado em *Salmonella typhimurium* e *Escherichia coli*, normalmente associado ao operon *bcs*. Estudo da inativação desse gene em *E. coli*, realizado por KIM *et al.* (2002), apresentou efeitos na morfologia e crescimento celular o que demonstra que a biossíntese de celulose pode estar relacionada ao crescimento celular. Estudos futuros envolvendo esse gene poderão ajudar a esclarecer a papel biológico da celulose e a sua função no crescimento bacteriano.

O estudo do operon *bcs* em *C. violaceum* revelou a existência de cinco genes em sua estrutura: (*yhjQ*, *bcsA*, *bcsB*, *bcsZ* e *bcsC*), que codificam na fita reversa complementar, respectivamente as proteínas YhjQ, BcsA–celulose sintase, BcsB–subunidade de regulação da celulose sintase que liga ao ativador diguanilmonofosfato cíclico (c-di-GMP), BcsZ–celulase, e BcsC–proteína com motivos *tetratrico peptide repeat* (TPR). Portanto, o operon *bcs* de *C. violaceum* é composto por pelo menos cinco genes, compreendendo as ORFs CV2675-2679.

A seqüência a montante do operon *bcs* em *C. violaceum* apresenta possíveis regiões regulatórias de transcrição, *Box -35* e *Box -10*, e a existência de uma provável região Shine-Dalgarno.

A análise do genoma da *C. violaceum* apresentou 43 ORFs com o domínio GGDEF. Os domínios GGDEF e EAL estão presentes em proteínas envolvidas na síntese e degradação de c-di-GMP. Em *A. xylinum* a molécula c-di-GMP ativa alostericamente a biossíntese de celulose. Uma vez que estudo realizado por AUSMEES *et al.* (2001) apresenta evidência de que proteínas que contêm o domínio GGDEF possuem atividade diguanilato ciclase, sugere-se que ORFs com domínio GGDEF aqui identificadas possuam também a capacidade de sintetizar c-di-GMP, participando desta forma da regulação da biossíntese de celulose por *C. violaceum*.

A identificação no genoma de *C. violaceum* de genes/ORFs que codificam as enzimas envolvidas nas reações da via de biossíntese de celulose em *A. xylinum*, a partir do substrato glicose, indicam que a mesma via biossintética pode estar presente em *C. violaceum*.

Os ensaios realizados para produção de biofilme por *C. violaceum* demonstraram forte evidência da presença de celulose na matriz extracelular produzida por este microrganismo, uma vez que ocorreu após a atividade de celulase sobre o substrato presente nas amostras, aumento da concentração de

glicose, o que indica que no substrato existem ligações $\beta(1\rightarrow4)$ glicana, características do polímero celulose.

Estudos de determinação de compostos da matriz extracelular produzida por microrganismo e a elucidação das vias de regulação auxiliam a compreensão do mecanismo molecular da biossíntese de celulose. Isto pode então ser utilizado no desenvolvimento de estratégias, tanto para prevenir quanto para otimizar a sua formação.

É importante salientar que a disponibilidade desses dados direciona ensaios laboratoriais para otimização da produção de celulose por *C. violaceum*, bem como representa um grande avanço para futuros trabalhos de pesquisa e desenvolvimento de novos produtos envolvendo celulose produzida por este microrganismo.

5.2 Sugestões para Trabalhos Futuros

Este trabalho abre muitas oportunidades para futuros trabalhos envolvendo celulose produzida por *Chromobacterium violaceum*, uma vez que não tem sido reportado até a presente data qualquer outra pesquisa abordando este tema. Como sugestões para futuras pesquisas podemos citar:

- Caracterização do EPS celulósico produzido pela *C. violaceum*
- Otimização da produção de celulose por *C. violaceum*
- Modelagem estrutural das enzimas envolvidas na via de biossíntese de celulose
- Função de compostos sinalizadores na formação de biofilme
- Regulação da via de biossíntese de celulose
- Relação entre a biossíntese de celulose com o crescimento celular
- Papel de cada proteína/enzima do operon *bcs* na biossíntese de celulose.

CAPÍTULO VI

Referências Bibliográficas

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic Local Alignment Search Tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.

ANDRIGHETTI-FROHNER, C. R.; ANTONIO, R. V.; CRECZYNSKI-PASA, T. B.; BARARDI, C. R.; SIMOES, C. M. Cytotoxicity and potential antiviral evaluation of violacein produced by *Chromobacterium violaceum*. **Mem Inst Oswaldo Cruz**, v. 98, n. 6, p. 843-848, 2003.

ANTÔNIO, R. V. **Biossíntese de violaceína pelo *Chromobacterium violaceum*: síntese e atividades biológicas de um provável intermediário**. 1994. Tese (Doutorado em Biologia Funcional e Molecular), UNICAMP, Campinas.

APWEILER, R.; BAIROCH, A.; WU, C. H.; BARKER, W. C.; BOECKMANN, B.; FERRO, S.; GASTEIGER, E.; HUANG, H.; LOPEZ, R.; MAGRANE, M.; MARTIN, M. J.; NATALE, D. A.; O'DONOVAN, C.; REDASCHI, N.; YEH, L. S. UniProt: the Universal Protein knowledgebase. **Nucleic Acids Res**, v. 32 Database issue, n., p. D115-119, 2004.

AUSMEES, N.; MAYER, R.; WEINHOUSE, H.; VOLMAN, G.; AMIKAM, D.; BENZIMAN, M.; LINDBERG, M. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. **FEMS Microbiol Lett**, v. 204, n. 1, p. 163-167, 2001.

BATEMAN, A.; COIN, L.; DURBIN, R.; FINN, R. D.; HOLLICH, V.; GRIFFITHS-JONES, S.; KHANNA, A.; MARSHALL, M.; MOXON, S.; SONNHAMMER, E. L.; STUDHOLME, D. J.; YEATS, C.; EDDY, S. R. The Pfam protein families database. **Nucleic Acids Res**, v. 32 Database issue, n., p. D138-141, 2004.

BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; RAPP, B. A.; WHEELER, D. L. GenBank. **Nucleic Acids Res**, v. 28, n. 1, p. 15-18, 2000.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The Protein Data Bank. **Nucl. Acids. Res.**, v. 28, n. 1, p. 235-242, 2000.

BIOFILL. Página na internet da empresa BioFill. <http://www.biofill.com.br>, acesso em 20 set. 2004.

BLOBEL, G. Proteins have intrinsic signals that govern their transport and localization in the cell. Press Release: The 1999 Nobel Prize in Physiology or Medicine. **Nobel Assembly at Karolinska Institutet**. 1999.

BOBBIO, P. A.; BOBBIO, F. O. Carboidratos. Química do processamento de Alimentos. **Livraria Varela**, 3º ed., p. 60-62, 2001.

BOECKMANN, B.; BAIROCH, A.; APWEILER, R.; BLATTER, M. C.; ESTREICHER, A.; GASTEIGER, E.; MARTIN, M. J.; MICHOD, K.; O'DONOVAN, C.; PHAN, I.; PILBOUT, S.; SCHNEIDER, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Res**, v. 31, n. 1, p. 365-370, 2003.

BONATO, M. C. M. Moldes, Módulos e Forma: do DNA às Proteínas. <http://www.biologianaweb.com/Livro2/Moldes.htm>, acesso em 30 set. 2004.

BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. **Proc Natl Acad Sci U S A**, v. 100, n. 20, p. 11660-11665, 2003.

BRAZILIAN NATIONAL GENOME PROJECT CONSORTIUM.
<http://www.brgene.lncc.br/>, acesso em 20 jun. 2004.

BROWN JR., M. R.; SAXENA, I. M.; KUDLICKA, K. Cellulose biosynthesis in higher plants. **Trends in Plants Science**, v. 1, n. 5, p. 149-156, 1996.

BROWN JR., M. R. POSITION PAPER Microbial Cellulose: A New Resource for Wood, Paper, Textiles, Food and Specialty Products," The University of Texas at Austin Botany Department.
<http://www.botany.utexas.edu/facstaff/facpages/mbrown/position1.htm>,
acesso em 30 set. 2004.

BROWN, T. A. Genomes 2nd ed. Oxford: **BIOS Scientific Publishers Ltd.**, 520 p., 2002.

CAMPBELL, J. A.; DAVIES, G. J.; BULONE, V.; HENRISSAT, B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. **Biochemical Journal**, v. 326, p. 929- 939, 1997.

CHAMPE, P. C.; HARVEY, R. A. Bioquímica Ilustrada. Porto Alegre: **ArTMed Editora**, 445 p., 2002.

CHANG, A. L.; TUCKERMAN, J. R.; GONZALEZ, G.; MAYER, R.; WEINHOUSE, H.; VOLMAN, G.; AMIKAM, D.; BENZIMAN, M.; GILLES-GONZALEZ, M. A. Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. **Biochemistry**, v. 40, n. 12, p. 3420-3426, 2001.

CHAO, Y.; MITARAI, M.; SUGANO, Y.; SHODA, M. Effect of addition of water-soluble polysaccharides on bacterial cellulose production in a 50-L airlift reactor. **Biotechnol Prog**, v. 17, n. 4, p. 781-785, 2001.

CHEN, X.; SU, Z.; DAM, P.; PALENIK, B.; XU, Y.; JIANG, T. Operon prediction by comparative genomics: an application to the *Synechococcus* sp WH8102 genome. **Nucleic Acids Research**, v. 32, n. 7, p. 2147-2157, 2004.

CHENG, H. P.; WANG, P. M.; CHEN, J. W.; WU, W. T. Cultivation of *Acetobacter xylinum* for bacterial cellulose production in a modified airlift reactor. **Biotechnol Appl Biochem**, v. 35, n. Pt 2, p. 125-132, 2002.

CORPE, W. A. Factors Influencing Growth and Polysaccharide Formation by Strains of *Chromobacterium violaceum*. **J Bacteriol**, v. 88, n., p. 1433-1441, 1964.

COSTERTON, J. W.; STEWART, P. S.; GREENBERG, E. P. Bacterial biofilms: a common cause of persistent infections. **Science**, v. 284, n. 5418, p. 1318-1322, 1999.

DAS, A. K.; COHEN, P. W.; BARFORD, D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. **Embo J**, v. 17, n. 5, p. 1192-1199, 1998.

DE SOUZA, A. O.; HEMERLY, F. P.; BUSOLLO, A. C.; MELO, P. S.; MACHADO, G. M. C.; MIRANDA, C. C.; SANTA-RITA, R. M.; HAUN, M.; LEON, L. L.; SATO, D. N.; DE CASTRO, S. L.; DURAN, N. 3-[4'-Bromo-(1,1'-biphenyl)-4-yl]-N,N-dimethyl-3-(2-thienyl)-2-propen-1-amine: synthesis, cytotoxicity, and leishmanicidal, trypanocidal and antimycobacterial activities. **J. Antimicrob. Chemother.**, v. 50, n. 5, p. 629-637, 2002.

DELMER, D. P. Cellulose Biosynthesis: Exciting Times for A Difficult Field of Study. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 50, n. 1, p. 245-276, 1999.

DOBLIN, M. S.; KUREK, I.; JACOB-WILK, D.; DELMER, D. P. Cellulose biosynthesis in plants: from genes to rosettes. **Plant Cell Physiol**, v. 43, n. 12, p. 1407-1420, 2002.

DURÁN, N.; ANTONIO, R. V.; HAUN, M.; PILLI, R. A. Biosynthesis of a Trypanocide by *Chromobacterium violaceum*. **World J. Microbiol. Biotechnol.** v.10, p. 686-690, 1994.

DURÁN, N.; MENCK, C. F. *Chromobacterium violaceum*: a review of pharmacological and industrial perspectives. **Crit Rev Microbiol**, v. 27, n. 3, p. 201-222, 2001.

EBERHARD, A.; BURLINGAME, A. L.; EBERHARD, C.; KENYON, G. L.; NEALSON, K. H.; OPPENHEIMER, N. J. Structural identification of autoinducer of *Photobacterium fischeri* luciferase. **Biochemistry**, v. 20, n. 9, p. 2444-2449, 1981.

ERENO D. Celulose na pele. Bactéria produz substância usada para tratar queimaduras e no revestimento de coletes à prova de balas. www.revistapesquisa.fapesp.br. v.101, acesso em 26/09/2004.

ERMOLAEVA, M. D.; WHITE, O.; SALZBERG, S. L. Prediction of operons in microbial genomes. **Nucleic Acids Res**, v. 29, n. 5, p. 1216-1221, 2001.

FUQUA, W. C.; WINANS, S. C.; GREENBERG, E. P. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. **J. Bacteriol**, v. 176, n. 2, p. 269-275, 1994.

FUQUA, C.; WINANS, S. C.; GREENBERG, E. P. Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. **Annu Rev Microbiol**, v. 50, n., p. 727-751, 1996.

GALPERIN, M. Y.; KOONIN, E. V. Who's your neighbor? New computational approaches for functional genomics. **Nat Biotechnol**, v. 18, n. 6, p. 609-613, 2000.

GALPERIN, M. Y.; NIKOLSKAYA, A. N.; KOONIN, E. V. Novel domains of the prokaryotic two-component signal transduction systems. **Fems Microbiology Letters**, v. 203, n. 1, p. 11-21, 2001.

GALPERIN, M. Y. Bacterial signal transduction network in a genomic perspective. **Environ Microbiol**, v. 6, n. 6, p. 552-567, 2004.

GASTEIGER, E.; GATTIKER, A.; HOOGLAND, C.; IVANYI, I.; APPEL, R. D.; BAIROCH, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Res**, v. 31, n. 13, p. 3784-3788, 2003.

GENEVAUX, P.; MULLER, S.; BAUDA, P. A rapid screening procedure to identify mini-Tn10 insertion mutants of *Escherichia coli* K-12 with altered adhesion properties. **FEMS Microbiol Lett**, v. 142, n. 1, p. 27-30, 1996.

GLASER, L. The synthesis of cellulose in cell-free extracts of *Acetobacter xylinum*. **J. Biol. Chem.**, v. 232, n. 2, p. 627-636, 1958.

GLESSNER, A.; SMITH, R. S.; IGLEWSKI, B. H.; ROBINSON, J. B. Roles of *Pseudomonas aeruginosa* las and rhl Quorum-Sensing Systems in Control of Twitching Motility. **J. Bacteriol.**, v. 181, n. 5, p. 1623-1629, 1999.

HECHT, G. B.; NEWTON, A. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. **J Bacteriol**, v. 177, n. 21, p. 6223-6229, 1995.

HENRISSAT, B.; CLAEYSSSENS, M.; TOMME, P.; LEMESLE, L.; MORNON, J. P. Cellulase families revealed by hydrophobic cluster analysis. **Gene**, v. 81, n. 1, p. 83-95, 1989.

HENRISSAT, B.; DAVIES, G. J. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. **Plant Physiology**, v. 124, n. 4, p. 1515-1519, 2000.

HJEMMESIDE, KENN MADENS. Página na internet.
<http://www.kennmadsen.dk/>, acesso em 10 out. 2004.

JONAS, R. e FARAH, L. F. Production and application of microbial cellulose. **Polymer Degradation and Stability**, v. 59, n. 1-3, p. 101-106, 1998.

KANEHISA, M. **Post-Genome Informatics**. Institute for Chemical Research, Kyoto University, Japan: **Oxford University Press**. 148 p., 2000.

KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic Acids Res**, v. 28, n. 1, p. 27-30, 2000.

KIM, M. K.; PARK, S. R.; CHO, S. J.; LIM, W. J.; RYU, S. K.; AN, C. L.; HONG, S. Y.; PARK, Y. W.; KAHNG, G. G.; KIM, J. H.; KIM, H.; YUN, H. D. The effect of a disrupted yhjQ gene on cellular morphology and cell growth in *Escherichia coli*. **Applied Microbiology and Biotechnology**, v. 60, n. 1-2, p. 134-138, 2002.

KOONIN, E. V.; GALPERIN, M. Y. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. **Curr Opin Genet Dev**, v. 7, n. 6, p. 757-763, 1997.

KROGH, A.; LARSSON, B.; VON HEIJNE, G.; SONNHAMMER, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. **J Mol Biol**, v. 305, n. 3, p. 567-580, 2001.

KRYSTYNOWICZ, A.; CZAJA, W.; WIKTOROWSKA-JEZIERSKA, A.; GONCALVES-MISKIEWICZ, M.; TURKIEWICZ, M.; BIELECKI, S. Factors affecting the yield and properties of bacterial cellulose. **J Ind Microbiol Biotechnol**, v. 29, n. 4, p. 189-195, 2002.

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. Princípios de Bioquímica. São Paulo: **Sarvier Editora de Livros Médicos**. 839 p., 2000.

LEON, L. L.; MIRANDA, C. C.; DE SOUZA, A. O.; DURAN, N. Antileishmanial activity of the violacein extracted from *Chromobacterium violaceum*. **J Antimicrob Chemother**, v. 48, n. 3, p. 449-450, 2001.

LEWIN, B. Genes V. **Oxford University Press**. 1296 p., 1995.

LIN, F. C. AND R. M. BROWN, JR. Purification of cellulose synthase from *Acetobacter xylinum*. Cellulose and Wood - Chemistry and Technology, Ed. C. Schuerch. John Wiley and Sons, Inc. N. Y., p. 473-492, 1989.

LOPEZ, R.; ROBINSON, S.; KIBRIA, A.; HARTE, N.; PATEL, G.; HARPER, R.; QUEVILLON, E.; SILVENTOINEN, V.; KALLIO, K.; JOKINEN, P. The European Bioinformatics Institute web site: a new view. **Bioinformatics**, v. 19, n. 4, p. 546-547, 2003.

MARKETON, M. M.; GLENN, S. A.; EBERHARD, A.; GONZALEZ, J. E. Quorum sensing controls exopolysaccharide production in *Sinorhizobium meliloti*. **J Bacteriol**, v. 185, n. 1, p. 325-331, 2003.

MARTIN, J. P.; RICHARDS, S. J. Decomposition and Binding Action of a Polysaccharide from *Chromobacterium violaceum* in Soil. **J Bacteriol**, v. 85, n., p. 1288-1294, 1963.

MARTINELLI, D.; BACHOFEN, R.; BRANDL, H. Effect of medium composition, flow rate, and signaling compounds on the formation of soluble extracellular materials by biofilms of *Chromobacterium violaceum*. **Applied Microbiology and Biotechnology**, v. 59, Numbers 2-3, n., p. 278 - 283, 2002.

MARTINELLI, D.; GROSSMANN, G.; SEQUIN, U.; BRANDL, H.; BACHOFEN, R. Effects of natural and chemically synthesized furanones on quorum sensing in *Chromobacterium violaceum*. **BMC Microbiol**, v. 4, n. 1, p. 25, 2004.

MATTHYSSE, A. G.; THOMAS, D. L.; WHITE, A. R. Mechanism of Cellulose Synthesis in *Agrobacterium-Tumefaciens*. **Journal of Bacteriology**, v. 177, n. 4, p. 1076-1081, 1995.

MORENO-HAGELSIEB, G.; COLLADO-VIDES, J. A powerful non-homology method for the prediction of operons in prokaryotes. **Bioinformatics**, v. 18, n. 90001, p. 329S-336, 2002a.

MORENO-HAGELSIEB, G.; COLLADO-VIDES, J. Operon conservation from the point of view of *Escherichia coli*, and inference of functional interdependence of gene products from genome context. **In Silico Biol**, v. 2, n. 2, p. 87-95, 2002b.

MULDER, N. J.; APWEILER, R.; ATTWOOD, T. K.; BAIROCH, A.; BATEMAN, A.; BINNS, D.; BISWAS, M.; BRADLEY, P.; BORK, P.; BUCHER, P.; COPLEY, R.; COURCELLE, E.; DURBIN, R.; FALQUET, L.; FLEISCHMANN, W.; GOUZY, J.; GRIFFITH-JONES, S.; HAFT, D.; HERMJAKOB, H.; HULO, N.; KAHN, D.; KANAPIN, A.; KRESTYANINOVA, M.; LOPEZ, R.; LETUNIC, I.; ORCHARD, S.; PAGNI, M.; PEYRUC, D.; PONTING, C. P.; SERVANT, F.; SIGRIST, C. J. InterPro: an integrated documentation resource for protein families, domains and functional sites. **Brief Bioinform**, v. 3, n. 3, p. 225-235, 2002.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION.

<http://www.ncbi.nlm.nih.gov/>, acesso em 15 jun. 2004.

NIELSEN, H.; ENGELBRECHT, J.; BRUNAK, S.; VON HEIJNE, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. **Protein Eng**, v. 10, n. 1, p. 1-6, 1997.

O'TOOLE, G.; KAPLAN, H. B.; KOLTER, R. Biofilm Formation as Microbial Development. **Annual Review of Microbiology**, v. 54, n. 1, p. 49-79, 2000.

O'TOOLE, G. A.; KOLTER, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. **Mol Microbiol**, v. 30, n. 2, p. 295-304, 1998.

OVERBEEK, R.; FONSTEIN, M.; D'SOUZA, M.; PUSCH, G. D.; MALTSEV, N. The use of gene clusters to infer functional coupling. **Proc Natl Acad Sci U S A**, v. 96, n. 6, p. 2896-2901, 1999.

PARSEK, M. R.; GREENBERG, E. P. Acyl-homoserine lactone quorum sensing in Gram-negative bacteria: a signaling mechanism involved in associations with higher organisms. **Proc Natl Acad Sci U S A**, v. 97, n. 16, p. 8789-8793, 2000.

PAUL, R.; WEISER, S.; AMIOT, N. C.; CHAN, C.; SCHIRMER, T.; GIESE, B.; JENAL, U. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. **Genes Dev**, v. 18, n. 6, p. 715-727, 2004.

PEI, J. M.; GRISHIN, N. V. GGDEF domain is homologous to adenylyl cyclase. **Proteins-Structure Function and Genetics**, v. 42, n. 2, p. 210-216, 2001.

PENG, L.; KAWAGOE, Y.; HOGAN, P.; DELMER, D. Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. **Science**, v. 295, n. 5552, p. 147-150, 2002.

PIEMOLINI, L. T. **Modelagem Estrutural da PHA Sintase de *Chromobacterium violaceum* para Estudos de Mutação Sítio-Dirigida**. 2003. 90f. Dissertação (Mestrado em Engenharia Química) – Programa de Pós Graduação em Engenharia Química, UFSC, Floarionópolis.

PROSDOCIMI, F.; CERQUEIRA, G.C.; BINNECK, E.; SILVA, A. F.; REIS, A. N.; JUNQUEIRA, A. C. M.; SANTOS A. C. F.; NHANI JUNIOR, A.; WUST C. I. ; CAMARGO FILHO, F;; KESSEDJIAN, J. L.; PETRETSKI, J. H.; CAMARGO L. P.; FERREIRA, R. G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. S.; FOLGUERAS-FLATSCHART, A. V. **Bioinformática: manual do usuário. Biotecnologia ciência & desenvolvimento.** , v.29, n.Janeiro, p.18 - 33, 2003.

PUSKAS, A.; GREENBERG, E. P.; KAPLAN, S.; SCHAEFER, A. L. A quorum-sensing system in the free-living photosynthetic bacterium *Rhodobacter sphaeroides*. **J Bacteriol**, v. 179, n. 23, p. 7530-7537, 1997.

RAVEN, P. H, EVERT, R. F, EICHHORN, S. E. **Biologia vegetal**. 6. ed. Rio de Janeiro: **Guanabara Koogan**, 2001. 906 p.

RAWLINGS, N. D.; BARRETT, A. J. Evolutionary families of peptidases. **Biochem J**, v. 290 (Pt 1), n., p. 205-218, 1993.

REDASCHI, N.; STERK, P.; STOEHR, P.; TULI, M. A. The EMBL nucleotide sequence database. **Nucleic Acids Res**, v. 29, n. 1, p. 17-21, 2001.

RICHMOND, T. Higher plant cellulose synthases. **Genome Biol**, v. 1, n. 4, p. REVIEWS3001, 2000.

RÖMLING, U. Molecular biology of cellulose production in bacteria. **Research in Microbiology**, v. 153, n. 4, p. 205-212, 2002.

RÖMLING, U.; ROHDE, M.; OLSEN, A.; NORMARK, S.; REINKOSTER, J. AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. **Mol Microbiol**, v. 36, n. 1, p. 10-23, 2000.

ROSS, P.; MAYER, R.; WEINHOUSE, H.; AMIKAM, D.; HUGGIRAT, Y.; BENZIMAN, M.; DEVROOM, E.; FIDDER, A.; DEPAUS, P.; SLIEDREGT, L. A. J. M.; VANDERMAREL, G. A.; VANBOOM, J. H. The Cyclic Diguanylic Acid Regulatory System of Cellulose Synthesis in *Acetobacter-Xylinum* - Chemical Synthesis and Biological-Activity of Cyclic-Nucleotide Dimer, Trimer, and Phosphothioate Derivatives. **Journal of Biological Chemistry**, v. 265, n. 31, p. 18933-18943, 1990.

ROSS, P.; WEINHOUSE, H.; ALONI, Y.; MICHAELI, D.; WEINBERGEROHANA, P.; MAYER, R.; BRAUN, S.; DEVROOM, E.; VANDERMAREL, G. A.; VANBOOM, J. H.; BENZIMAN, M. Regulation of Cellulose Synthesis in *Acetobacter xylinum* by Cyclic Diguanylic Acid. **Nature**, v. 325, n. 6101, p. 279-281, 1987.

ROSS, P.; MAYER, R.; BENZIMAN, M. Cellulose Biosynthesis and Function in Bacteria. **Microbiological Reviews**, v. 55, n. 1, p. 35-58, 1991.

RUTHERFORD, K.; PARKHILL, J.; CROOK, J.; HORSNELL, T.; RICE, P.; RAJANDREAM, M. A.; BARRELL, B. Artemis: sequence visualization and annotation. **Bioinformatics**, v. 16, n. 10, p. 944-945, 2000.

SALGADO, H.; MORENO-HAGELSIEB, G.; SMITH, T. F.; COLLADO-VIDES, J. Operons in *Escherichia coli*: genomic analyses and predictions. **Proc Natl Acad Sci U S A**, v. 97, n. 12, p. 6652-6657, 2000.

SAXENA, I. M.; LIN, F. C.; BROWN, R. M. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. **Plant Molecular Biology**, v. 16, n. 6, p. 947-954, 1990.

SAXENA, I. M.; LIN, F. C.; BROWN, R. M. Identification of a New Gene in an Operon for Cellulose Biosynthesis in *Acetobacter xylinum*. **Plant Molecular Biology**, v. 16, n. 6, p. 947-954, 1991.

SAXENA, I. M.; KUDLICKA, K.; OKUDA, K.; BROWN, R. M., JR. Characterization of genes in the cellulose-synthesizing operon (acs operon) of *Acetobacter xylinum*: implications for cellulose crystallization. **J Bacteriol**, v. 176, n. 18, p. 5735-5752, 1994.

SAXENA, I. M., BROWN, R. M.; FEVRE, M.; GEREMIA, R. A.; HENRISSAT, B. Multidomain architecture of β -glycosyl transferases: implications for mechanism of action. **J. Bacteriol.** v. 77, p.1419-1424, 1995.

SAXENA, I. M.; BROWN, R. M., JR.; DANDEKAR, T. Structure-function characterization of cellulose synthase: relationship to other glycosyltransferases. **Phytochemistry**, v. 57, n. 7, p. 1135-1148, 2001.

SCHULTZ, J.; COPLEY, R. R.; DOERKS, T.; PONTING, C. P.; BORK, P. SMART: a web-based tool for the study of genetically mobile domains. **Nucleic Acids Res**, v. 28, n. 1, p. 231-234, 2000.

SERVANT, F.; BRU, C.; CARRERE, S.; COURCELLE, E.; GOUZY, J.; PEYRUC, D.; KAHN, D. ProDom: automated clustering of homologous domains. **Brief Bioinform**, v. 3, n. 3, p. 246-251, 2002.

SHODA & ANO Laboratory. <http://www.res.titech.ac.jp/~junkan/>, acesso em 05 set. 2004.

SKINNER, P. O.; CANNON, R. E. *Acetobacter xylinum*: An inquiry into cellulose biosynthesis. **American Biology Teacher**, v. 62, n. 6, p. 442-444, 2000.

SPIERS, A. J.; BOHANNON, J.; GEHRIG, S. M.; RAINEY, P. B. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. **Molecular Microbiology**, v. 50, n. 1, p. 15-27, 2003.

STASINOPOULOS, S.; FISHER, P.; STONE, B.; STANISICH, V. Detection of two loci involved in (1-->3)-beta-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene. **Glycobiology**, v. 9, n. 1, p. 31-41, 1999.

STEINBÜCHEL A., DEBZI E. M., MARCHESSAULT R. H., TIMM A. Synthesis and production of poly (3-hydroxyvaleric acid) homopolymer by *Chromobacterium violaceum*. **App. Microbiol. Biotechnol.**, v.39, p. 443-449, 1993.

STEPHENS, DAVE. Página na internet. <http://niko.unl.edu/bs101/index.html>, acesso em 19 out. 2004.

STOESSER, G.; BAKER, W.; VAN DEN BROEK, A.; CAMON, E.; GARCIA-PASTOR, M.; KANZ, C.; KULIKOVA, T.; LOMBARD, V.; LOPEZ, R.; PARKINSON, H.; REDASCHI, N.; STERK, P.; STOEHR, P.; TULLI, M. A. The EMBL nucleotide sequence database. **Nucleic Acids Res**, v. 29, n. 1, p. 17-21, 2001.

SUTHERLAND, I. W. The biofilm matrix - an immobilized but dynamic microbial environment. **Trends in Microbiology**, v. 9, n. 5, p. 222-227, 2001.

TATENO, Y.; IMANISHI, T.; MIYAZAKI, S.; FUKAMI-KOBAYASHI, K.; SAITOU, N.; SUGAWARA, H.; GOJOBORI, T. DNA Data Bank of Japan (DDBJ) for genome scale research in life science. **Nucleic Acids Res**, v. 30, n. 1, p. 27-30, 2002.

TATUSOV, R. L.; GALPERIN, M. Y.; NATALE, D. A.; KOONIN, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic Acids Res**, v. 28, n. 1, p. 33-36, 2000.

THE EUROPEAN MOLECULAR BIOLOGY LABORATORY.
<http://www.embl.org/>, acesso em 20 jul. 2004.

THE PFAM DATABASE OF PROTEIN FAMILIES AND HMMS.
<http://pfam.wustl.edu/index.html>, acesso em 20 jul. 2004.

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

TIPPMANN, H. F. Analysis for free: comparing programs for sequence analysis. **Brief Bioinform**, v. 5, n. 1, p. 82-87, 2004.

UCHIYAMA, I. MBGD: microbial genome database for comparative analysis. **Nucleic Acids Res**, v. 31, n. 1, p. 58-62, 2003.

VANDAMME, E. J.; DE BAETS, S.; VANBAELEN, A.; JORIS, K.; DE WULF, P. Improved production of bacterial cellulose and its application potential. **Polymer Degradation and Stability**, v. 59, n. 1-3, p. 93-99, 1998.

VILLELA, G. G.; BACILA, M.; TASTALDI, H. Bioquímica. Rio de Janeiro: **Guanabara Koogan**. 780 p., 1978.

VON BODMAN, S. B.; MAJERCZAK, D. R.; COPLIN, D. L. A negative regulator mediates quorum-sensing control of exopolysaccharide production in *Pantoea stewartii* subsp. *stewartii*. **Proc Natl Acad Sci U S A**, v. 95, n. 13, p. 7687-7692, 1998.

WAN, H.; WOOTTON, J. C. A global compositional complexity measure for biological sequences: AT-rich and GC-rich genomes encode less complex proteins. **Comput Chem**, v. 24, n. 1, p. 71-94, 2000.

WATNICK, P. I.; KOLTER, R. Biofilm, City of Microbes. **J. Bacteriol.**, v. 182, n. 10, p. 2675-2679, 2000.

WONG, H. C.; FEAR, A. L.; CALHOON, R. D.; EICHINGER, G. H.; MAYER, R.; AMIKAM, D.; BENZIMAN, M.; GELFAND, D. H.; MEADE, J. H.; EMERICK, A. W.; BRUNER, R.; BEN-BASSAT, A.; TAL, R. Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. **Proc. Natl. Acad. Sci. USA.**, v. 87, p. 8130-8134, 1990.

WU, C. H.; YE, L. S.; HUANG, H.; ARMINSKI, L.; CASTRO-ALVEAR, J.; CHEN, Y.; HU, Z.; KOURTESIS, P.; LEDLEY, R. S.; SUZEK, B. E.; VINAYAKA, C. R.; ZHANG, J.; BARKER, W. C. The Protein Information Resource. **Nucleic Acids Res**, v. 31, n. 1, p. 345-347, 2003.

WULFF, Nelson Arno. **Caracterização Enzimática das Celulases Xf-810, Xf-818 E Xf-2708 de *Xylella Fastidiosa* e Purificação da Proteína XF-818, expressa em *Escherichia coli***. 2002. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", USP, São Paulo.

XAVIER, J. B.; PICIOREANU, C.; ALMEIDA, J. S.; VAN LOOSDRECHT, M. C. M. Monitorização e modelação da estrutura de biofilmes. **Boletim de Biotecnologia. Sociedade Portuguesa de Biotecnologia**, v. 76, n., p. 2-13, 2003.

ZHENG, Y.; SZUSTAKOWSKI, J. D.; FORTNOW, L.; ROBERTS, R. J.; KASIF, S. Computational identification of operons in microbial genomes. **Genome Res**, v. 12, n. 8, p. 1221-1230, 2002.

ZOGAJ, X.; NIMTZ, M.; ROHDE, M.; BOKRANZ, W.; RÖMLING, U. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. **Molecular Microbiology**, v. 39, n. 6, p. 1452-1463, 2001.