



**Pró-Reitoria de
Pós-Graduação e Pesquisa
Programa de Pós-Graduação Stricto Sensu em Ciências
Genômicas e Biotecnologia**

**IDENTIFICAÇÃO DE ENZIMAS EXPRESSAS EM CLONES
METAGENÔMICOS POR ESPECTROMETRIA DE MASSA**

**Autora: Flávia Rodrigues Dutra
Orientadora: Beatriz Simas Magalhães , PhD
Co-orientadora: Betania Ferraz Quirino, PhD**

**Brasília - DF
2014**

FLÁVIA RODRIGUES PEREIRA DUTRA

**IDENTIFICAÇÃO DE ENZIMAS EXPRESSAS EM CLONES
METAGENÔMICOS POR ESPECTROMETRIA DE MASSA**

Relatório parcial apresentado ao programa *Strictu Sensu* em Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília, como parte dos requisitos para obtenção do Título de Mestre.

Orientadora: Beatriz Simas
Magalhães, PhD

Co-orientadora: Betania Ferraz
Quirino, PhD

Brasília
2014

D978i Dutra, Flávia Rodrigues Pereira.

Identificação de enzimas expressas em clones metagenômicos por espectrometria de massa. / Flávia Rodrigues Pereira Dutra – 2014.

74 f.; il.: 30 cm

Dissertação (Mestrado) – Universidade Católica de Brasília, 2014.
Orientação: Profa. PhD Beatriz Simas Magalhães

Coorientação: Profa. PhD Betania Ferraz Quirino

Aos meus pais Reinaldo e Eliane e aos meus irmãos Thiago e Bruno. A vocês, que fazem parte de mim, dedico este trabalho. Uma vitória dentre tantas outras que virão.

AGRADECIMENTO

Primeiramente agradeço a Deus, por toda Graça que tem me oferecido;

À professora doutora Beatriz Simas Magalhães, por me proporcionar crescimento profissional e pessoal. A sua calma foi fator relevante na realização deste trabalho;

À professora doutora Betania Ferraz Quirino por me oferecer a oportunidade de trabalharmos em colaboração e pela confiança depositada em mim;

Ao meu primo, amigo e irmão Guilherme Vale;

Aos amigos de longa data Eduardo e Monica Felberg, Rondinelle Muniz, Nelson de Oliveira Júnior, Mirna Freire, Samantha da Silveira e Luis Felipe Schroeder. Sem o apoio de vocês, a tarefa teria sido mais árdua;

Aos amigos que fiz no laboratório quando entrei como iniciação científica e que continuam a fazer parte da minha vida, Lucas Carvalho, Virgílio Hipólito e Jéssica Bergmann. Vocês alegram o nosso ambiente de trabalho;

Aos amigos que me ajudaram diretamente na realização deste trabalho, Stella Maris de Freitas, Ludovico Migliolo e William Porto;

Às amigas recentes Fernanda Araújo e Juliana Nobre, sempre dispostas a ajudar nos momentos de desespero;

À amiga Cláudia Venturelle, que dispôs de seu tempo para me ajudar nos experimentos e nas horas de descontração, mostrou-se de grande valia;

À Universidade Católica de Brasília, pela estrutura;

Ao CNPq, pelo apoio financeiro;

A todos, os meus mais sinceros agradecimentos.

“A adversidade é o primeiro caminho para a verdade”.

George Gordon Byron

RESUMO

Referência: DUTRA, Flávia Rodrigues Pereira. **Identificação de enzimas expressas em clones metagenômicos por espectrometria de massa**. 2014. 74 pp. Dissertação Programa *Strictu sensu* em Ciências Genômicas e Biotecnologia. Universidade Católica de Brasília, Brasília, 2014.

O desenvolvimento de ferramentas biotecnológicas que propiciem a busca por novas enzimas que desempenhem a sua ação catalítica em ambientes não fisiológicos, é um dos focos de aplicações da ciência moderna. A procura por novos produtos gênicos oriundos de bibliotecas metagenômicas tornou-se fundamental para diversas áreas tecnológicas. Com o advento desta técnica, os genomas de organismos não cultivados também se tornaram acessíveis para aplicações biotecnológicas. Assim, a busca de novos genes funcionais com potencial uso comercial e de interesse para a indústria tornou-se rotina. Dentre os produtos gênicos mais procurados, enzimas são amplamente exploradas nestes acervos. Uma vez que um clone com atividade enzimática é identificado, o gargalo está na correlação entre diversos genes presentes no fragmento de DNA e a atividade enzimática de interesse. Análises *in silico* baseiam-se em comparações de sequências de ORFs (open reading frame) depositadas em banco de dados. Quando não há similaridade entre as sequências alinhadas, não é possível definir qual a sequência decodificadora. Estas sequências presentes em bancos de dados representam apenas uma ínfima parte de ORFs existentes, o que é um limitante na busca por novas moléculas. Assim, faz-se necessário um processo que minimize o trabalho e o tempo na busca destes novos produtos enzimáticos a partir de clones metagenômicos. A associação de técnicas de purificação de proteínas à espectrometria de massa, pode se tornar importante ferramenta aplicada a esta busca. O objetivo central deste trabalho foi minimizar o tempo que é gasto na obtenção de sequências gênicas e diminuir as limitações da análise *in silico*, que apenas reconhecem sequências de ORFs que possuam sequências de organismos filogeneticamente próximos depositadas em bancos de dados. Para isto, neste trabalho, utilizou-se uma abordagem proteômica, baseada em técnicas de espectrometria de massa, para tornar mais célere a correlação entre a enzima de interesse e sua respectiva ORF.

Palavras-chave: Metagenômica, expressão proteica, espectrometria de massa.

ABSTRACT

The development of biotechnological tools that facilitate the search for new enzymes perform their catalytic action in non-physiological environments, is one of the focuses of applications of modern science. The search for new gene products derived from metagenomic libraries has become essential for many technological areas. With the advent of this technique, the genomes of uncultivated organisms have also become available for biotechnological applications. Thus, the search for new functional genes with potential commercial use and interest to the industry has become routine. Among the most popular gene products, enzymes are widely exploited in these collections. Once a clone is identified with enzyme activity, the bottleneck is no correlation between various genes present in the DNA fragment of interest and enzymatic activity. In silico analysis based on sequence comparisons of ORFs (open reading frame) deposited in the database. When there is no similarity between the aligned sequences, it is not possible to define which sequence decoder. These sequences present in databases represent only a small part of existing ORFs, which is a limiting factor in the search for new molecules. Thus, it makes necessary a process that minimizes the work and time in the pursuit of these new enzyme products from metagenomic clones. The combination of techniques for protein purification and mass spectrometry, can become an important tool applied to this search. The central objective of this study was to minimize the time that is spent in obtaining gene sequences and decrease the limitations of in silico analysis, which only recognize sequences of ORFs that have sequences phylogenetically close organisms deposited in databases. For this, in this work we used a proteomics based approach em técnicas mass spectrometry to make faster the correlation between the enzyme of interest and its respective ORF.

Keywords: Metagenomics, protein expression, mass spectrometry

LISTA DE ILUSTRAÇÕES

Figura 1- Esquema da parte experimental do trabalho	30
Figura 2- Fluxograma do desenho experimental	31
Figura 3- Mapa do vetor de clonagem PCF430 utilizado na montagem da biblioteca. O DNA clonado foi digerido com a enzima de restrição <i>Pst</i> I	32
Figura 4 - Representação do promotor P _{BAD} e as posições relativas dos elementos genéticos no vetor PCF430	32
Figura 5- Mapa do vetor pCC1FOS	33
Figura 6- Mapa do vetor de expressão pET21a	34
Figura 7- Ensaio enzimático para atividade de endoglucanases realizado com o lisado de células armazenadas a -20 °C	42
Figura 8- Ensaio enzimático para atividade de endoglucanases realizado com o lisado de células armazenadas a 4 °C	42
Figura 9- Ensaio enzimático para atividade enzimática de endoglucanases realizado com o extrato celular que não foi estocado	43
Figura 10- Quantificação do açúcar redutor liberado durante o ensaio enzimático realizado com extratos celulares que não foram armazenados	44
Figura 11- Comparação dos níveis de açúcar redutor liberado entre os extratos celulares que não foram armazenados e os que foram a 4°C	45

Figura 12- Detecção de β -glicosidases com o substrato pNPG nos clones de pequenos insertos	46
Figura 13- Detecção de atividade β -glicosidase com o substrato pNPG nos clones de grandes insertos da biblioteca metagenômica de solo da Amazônia	47
Figura 14- Alinhamento da proteína BGL11 com uma glicosil hidrolase pertencente à família 3.....	48
Figura 15- Sequência de aminoácidos da enzima BGL11	48
Figura 16- Eletroforese em gel de poliacrilamida SDS-PAGE dos extratos celulares provenientes dos clones metagenômicos da biblioteca de pequenos insertos	49
Figura 17- Alinhamento da proteína PL8G1 com uma glicosil hidrolase pertencente à família 31.....	50
Figura 18- Sequência de aminoácidos da enzima PL8G1	51
Figura 19- Alinhamento da proteína B17.1 com uma glicosil hidrolase pertencente à família 3	51
Figura 20- Sequência de aminoácidos da enzima B17.1	52
Figura 21- Alinhamento da proteína B17.2 com uma glicosil hidrolase pertencente à família 2	52
Figura 22- Sequência de aminoácidos da enzima B17.2	53
Figura 23- Alinhamento da proteína B17.2 com uma glicosil hidrolase pertencente à família 1	53
Figura 24- Sequência de aminoácidos da enzima B18	54
Figura 25- Eletroforese em gel de poliacrilamida SDS-PAGE dos extratos celulares provenientes dos clones metagenômicos da biblioteca de grandes insertos, que tiveram atividade enzimática para endoglucanases e β -glicosidases	55
Figura 26- Espectro MS/MS do íon de massa $[M+H^+]^{+2}$ 566, 29	57
Figura 27- Espectro MS/MS do íon de massa $[M+H^+]^{+2}$ 687,74	58
Figura 28- Espectro MS/MS do íon de massa $[M+H^+]^{+2}$ 712, 30	59

Figura 29- Espectro do íon de massa $[M+H]^+2$ 780,46 60

Figura 30- Alinhamento entre os fragmentos de peptídeos obtidos por MS/MS e a enzima BGL11, utilizada como referência 61

LISTA DE TABELAS

Tabela 1- Número de espectros MS/MS obtidos por corrida cromatográfica .. 55

Tabela 2- Lista das massas dos íons obtidos por LC-MS/MS de fragmentos de peptídeos encontrados na sequência da enzima referência 56

LISTA DE ABREVIATURAS

Bar	Pressão
°C	Graus Celsius
C-Terminal	Carboxy terminal
CMC	Carboximetilcelulose
DNSA	Ácido dinitrosalisílico
DTT	Ditiotreitol
EDTA	Ácido etilenodiaminotetracético
ESI	<i>Electrospray</i>
eV	Eletron volt
FNHM	<i>Full-midth half-maximum</i>
g	Força g
h	Hora
HPLC	<i>High performance liquid chromatography</i> (Cromatografia líquida de alta definição)
IPTG	Isopropil-β-D-tiogalactopiranosídeo
kb	Quilobase
kDa	Quilodalton
LB	Luria Bertani
LBA	Ágar Luria Bertani
LV	Baixa viscosidade
µg.mL⁻¹	Microgramas por microlitro
µL	Microlitro
µm	Micrometro
mL	Mililitro
mM	Milimolar

min	Minuto
MS	<i>Mass spectrometry</i> (Espectrometria de massa)
MS/MS	<i>Tandem mass spectrometry</i> (Espectrometria de massa em tandem)
m/z	Massa/carga
nm	Nanômetro
N-terminal	Amino terminal
OD	Densidade Óptica
ORF	<i>Open Reading Frame</i> (Fase de leitura aberta)
PCR	<i>Polimerase chain reaction</i> (Reação da polimerase em cadeia)
pH	Potencial hidrogeniônico
PMSF	Fenilmetilsulfonil
pNPG	p-nitrofenil-alfa-glucopiranosídeo
rpm	Rotação por minuto
s	segundo
STE	Cloreto de sódio Tris e EDTA
TB	<i>Terrific Broth</i>
TFA	Ácido Trifluoroacético
ToF	<i>Time of flight</i> (Tempo de voo)
Tris-HCl	Base Tris- Ácido Clorídrico
UFLC	<i>Ultra fast liquid chromatography</i> (Cromatografia líquida de ultra definição)
Vpp	Tensão pico a pico

SUMÁRIO

1. Introdução	16
2. Revisão de Literatura	18
2.1 Busca por novas enzimas.....	18
2.2 Mircrorganismos como fontes de novas enzimasd	19
2.3 Desenvolvimento da metagenômica	20
2.3.1 Bibliotecas Metagenômicas	21
2.3.2 Triagem de bibliotecas metagenômicas baseada na função de genes.....	22
2.4 Expressão de Enzimas	22
2.5 Expressão de Fenótipos	23
2.5.1 Limitações das análises baseadas em função.....	25
2.6 Busca por novas enzimas usando ferramentas ligadas à proteômica .	26
2.7 Bibliotecas metagenômicas utilizadas	27
3. Justificativa.....	28
4. Objetivos	29
4.1 Objetivo geral.....	29
4.2 Objetivos específicos.....	29
5. Material e Métodos.....	30
5.1 Desenho experimental.....	30
5.2 Bibliotecas metagenômicas estudadas.....	31
5.2.1 Biblioteca de pequenos insertos	31
5.2.2 Biblioteca de grandes insertos	32
5.3 Clones metagenômicos selecionados Busca por novas enzimas.....	34
5.4 Crescimento celular para teste de atividade celulolítica <i>in vitro</i>	35
5.4.1 Seleção de clones de pequenos insertos	35
5.4.2 Clone subclonado no vetor de expressão pET21a	36
5.4.3 Seleção de clones de grandes insertos	37
5.4.4 Busca por fases de leitura relacionadas a enzimas de interesse .	37
5.5 Teste de atividade celulolítica.....	38
5.5.1 β -glicosidase.....	38

5.5.2 Atividade de endoglucanases	38
5.6 Análise estatística	39
5.7 Eletroforese em gel de poliacrilamida SDS-PAGE	39
5.8 Digestão de proteínas para a aquisição de LC-MS/MS	39
5.9 Aquisição de LC MS/MS.....	39
6 Resultados	41
6.1 Ensaio enzimático <i>in vitro</i> para a detecção de atividade de endoglucanases.....	41
6.2 Ensaio enzimático <i>in vitro</i> para detecção de atividade de β -glicosidases	45
6.2.1 Biblioteca de pequenos insertos	45
6.2.2 Biblioteca de grandes insertos	46
6.3 Busca por fases de leitura relacionadas a enzimas de interesse	47
6.4 Aquisição de LC-MS/MS.....	54
7 Discussão.....	62
8 Conclusão	66

1. INTRODUÇÃO

Enzimas são compostos moleculares de natureza comumente protéica que tem função catalisadora, ou seja, que diminuem a energia de ativação necessária para que aconteça uma reação química (BON; FERRARA; CORVO, 2008). A heterogeneidade deste grupo em catalisar diferentes reações químicas, fez com que estas biomoléculas fossem amplamente exploradas por diversos setores da indústria, visando à celeridade na produção de insumos. A busca por novas enzimas, que sejam mais adequadas à produção biotecnológica tornou-se um dos pilares da indústria moderna (JIN et al., 2012).

Com o advento da engenharia genética, a produção de enzimas em larga escala tornou-se possível, permitindo a seleção da biomolécula mais adequada para a atividade em questão (BON et al., 2008). Ferramentas biotecnológicas são utilizadas para desenvolver moléculas que sejam estáveis em condições extremas (DÍAZ-RODRÍGUEZ; DAVIS, 2011).

Neste aspecto, microrganismos tornaram-se alvos na busca por novas enzimas. Seu papel ecológico na degradação de substratos complexos e a infinita quantidade de espécies microbianas que ainda são desconhecidas, fez deste grupo um acervo infindável de biomoléculas (KNIGHT et al., 2012).

De forma a ter acesso a essas espécies microbianas desconhecidas, a metagenômica foi desenvolvida como uma ferramenta inicialmente concebida para classificar microrganismos filogeneticamente de forma mais assertiva. Com a evolução da técnica, tornou-se possível a busca pelas sequências gênicas responsáveis pela síntese protéica em organismos não cultiváveis (SIMON; DANIEL, 2009).

A triagem por enzimas em bibliotecas metagenômicas é realizada através de testes *in vitro*, tendo em vista uma determinada atividade enzimática. A correlação que se faz entre o gene codificante e a enzima de interesse, quase sempre é uma tarefa árdua. Para a busca por ORFs em clones metagenômicos, é necessário que seja feita uma busca em bancos de dados a procura de sequências equivalentes. Com isso, há a inevitável perda de informação de novas ORFs que estão relacionadas à codificação de proteínas, mas por não possuírem nenhuma correlação com os bancos de dados, não são identificadas, e isto acaba se tornando um viés na busca por novas enzimas.

Após adquirir a sequência da ORF relacionada à enzima, é necessária a transferência desta sequência de nucleotídeos para um vetor de expressão para a confirmação da atividade, não sendo garantido que esta confirmação será obtida. A idéia de se identificar a sequência do gene codante através da sequência de aminoácidos proveniente da enzima que está sendo expressa, é nova e está arraigada a desafios.

O uso da cromatografia líquida de alta eficiência (HPLC) acoplado a um espectrômetro de massa do tipo *electrospray* (ESI) vem sendo comumente utilizado para a identificação de peptídeos clivados enzimaticamente. (CHEN; CANALES; NEAL, 2011). A sensibilidade da técnica permite uma busca mais acurada por proteínas em amostras complexas com um alto rendimento.

Este trabalho foi desenvolvido com o intuito de desenvolver uma metodologia por espectrometria de massa, que vise identificar a enzima de interesse com base no gene codificante do clone metagenômico.

2. REVISÃO DE LITERATURA

2.1 BUSCA POR NOVAS ENZIMAS

O uso de enzimas para a fabricação de produtos confunde-se com a história da civilização (BON et al., 2008). Relatos do uso de microrganismos na produção de alimentos mostram que já há milhares de anos, enzimas eram manipuladas por nossos ancestrais de forma indireta, já que o entendimento de que não eram os microrganismos em si, mas sim as enzimas sintetizadas por estes que eram responsáveis pelas reações químicas só foi possível em meados do século XIX (LIESE; SEELBACH; WANDREY, 2008).

Processos industriais são extremamente dependentes de biocatalisadores. Pela sua variabilidade de funções, enzimas são utilizadas em larga escala desde a produção de alimentos e detergentes até a produção de biocombustíveis (AKEROYD et al., 2013). São recursos utilizados também por indústrias químicas, têxteis e farmacêuticas. De maneira geral, biocatalisadores idealizados pela indústria precisam ser ativos *in vitro*, estáveis quando submetidos a altas temperaturas e, altas variações de pH, eficientes e altamente específicos quanto ao substrato (LORENZ; ECK, 2005).

A utilização de reações químicas para a produção de bens de consumo tem se consolidado como um dos pilares da indústria moderna. Como a maior parte destas reações precisam de catalisadores para serem concluídas, a utilização de enzimas tem se mostrado vantajosa (JIN et al., 2012). O fato deste grupo ser extremamente heterogêneo, faz com que estas moléculas sejam responsáveis pela catálise de uma gama de diferentes reações químicas, além de serem biocatalisadores extremamente eficientes, e não causarem danos ao meio ambiente (MA et al., 2013). Todas estas vantagens fazem das enzimas alvo de grande interesse comercial (WANG et al., 2012).

A variação nos preços das ações, o alto preço dos insumos, assim como o aumento da importância da sustentabilidade, exige que a produção biotecnológica seja extremamente eficiente a ponto de garantir uma produção em massa que gere lucro. Com isso, é necessário o investimento maciço no desenvolvimento de metodologias que aperfeiçoem a catálise enzimática (KUPPERS et al., 2014).

A eficiência das enzimas em catalisar reações químicas está relacionada com o tipo de interação que pode ocorrer com o substrato. Características como

estereosseletividade, químiosseletividade e regioosseletividade podem ser atraentes do ponto de vista industrial (CHICA; DOUCET; PELLETIER, 2005; YANG; WONG; LIGHTSTONE, 2014).

Em nível industrial, é necessário que a enzima esteja adequada a condições que não são as ideais para a sua atividade, e sim condições que favoreçam a produção em larga escala. É interessante que a enzima tenha alta seletividade e especificidade a substratos não naturais. É importante também que a enzima não seja inibida por altas concentrações de produto ou de substrato. Manter a atividade elevada em altas temperaturas e condições extremas de pH, também são quesitos necessários (DRAUZ, 2012).

Com o objetivo de desenvolver moléculas cujo desempenho em condições não fisiológicas seja adequado, ferramentas biotecnológicas são utilizadas. Como exemplo pode-se citar a modificação química, onde alguns resíduos de amino ácidos são substituídos, proporcionando maior estabilidade à enzima (DÍAZ-RODRÍGUEZ; DAVIS, 2011; HU et al., 2014).

Outra estratégia utilizada está em explorar a microbiota existente em ambientes extremos. Microrganismos sujeitos a condições extremas fornecem enzimas que são estáveis quando em altas temperaturas ou baixo níveis de pH (DEIVE et al., 2013; NAVARRO; VON BERNATH; JEREZ, 2013).

Existe ainda a possibilidade de unir as duas técnicas, buscando enzimas em organismos extremófilos e a utilização de modificações físico-químicas com o intuito de melhorar a atividade catalítica (COWAN; FERNANDEZ-LAFUENTE, 2011).

2.2 MICRORGANISMOS COMO FONTES DE NOVAS ENZIMAS

Microrganismos presentes em diversos ecossistemas possuem a capacidade de degradar substratos complexos. Esta atividade catalítica é explorada no desenvolvimento de produtos oriundos do processamento de matérias naturais. A quantidade de espécies microbianas existente no planeta ultrapassa a ordem de 10^{30} . A grande maioria das espécies ainda continua desconhecida, assim como a respectiva função que desempenha dentro do sistema (KNIGHT et al., 2012).

Características fenotípicas são influenciadas por fatores ambientais. A escolha do nicho de onde microrganismos serão amostrados é uma forma de pré estabelecer quais as condições físico-químicas em que as biomoléculas produzidas

serão ativas. Estados extremos relacionados a altas temperaturas, variações de pH, pouca disponibilidade de nutrientes e altas concentrações de sais, são condições em que a maioria dos processos industriais ocorrem e por isso as enzimas utilizadas precisam ser estáveis nestas a estas situações.

Sedimentos terrestres estão entre os ambientes mais explorados em busca de novas biomoléculas. Solos são ambientes amplamente diversos em comunidades microbianas, onde o número de células procarióticas pode chegar a mais de 10 bilhões, podendo conter milhares de genomas diferentes (KNIGHT et al., 2012).

A diversidade de nichos terrestres, com diferentes composições físico-químicas tem ampliado o campo de busca por novas biomoléculas. Microrganismos extraídos de sedimentos de ambientes de altas temperaturas, pHs extremos ou com altas concentrações de sais, são explorados em busca de enzimas altamente estáveis e que sejam viáveis em processos industriais (MOE; MCMAHON; THOMAS, 2010).

Além da triagem em ambientes mais abrangentes, em que a diversidade de microrganismos amplia a busca por diferentes biomoléculas, outro tipo de abordagem pode ser utilizado, com a exploração de ambientes que favorecem a expressão de fenótipos específicos. Microrganismos simbiotes presentes no trato intestinal de alguns animais são explorados por serem especializados em degradar materiais complexos (SIMON; DANIEL, 2009).

2.3 DESENVOLVIMENTO DA METAGENÔMICA

A abordagem clássica para a obtenção do genoma de microrganismo, dependente de cultivo, mostrou ser um método limitado, pois apenas um ínfimo número de espécies são cultiváveis (LEE; LEE, 2013).

A metagenômica surgiu com a necessidade de melhorar a classificação filogenética de microrganismos. Técnicas de cultivo *in vitro*, permitem que apenas um ínfimo grupo (1% a 10%) de microrganismos cresça em condições controladas, sendo assim, a informação sobre a maior parte das espécies é perdida (SIERRA-GARCIA et al., 2014).

A metagenômica permitiu o acesso a todo o DNA genômico de microrganismos existentes em um determinado sistema. Este tipo de informação

permitiu além da classificação filogenética, a classificação funcional que a microbiota desempenha em seu ambiente (XU; GUNSOLLEY, 2014).

A técnica resume-se à extração do DNA total de uma amostra ambiental coletada. O material genético extraído é clivado, gerando fragmentos randômicos que por sua vez serão ligados a vetores de clonagem e inseridos em células hospedeiras.

A abordagem funcional da metagenômica está relacionada à busca de novos genes codantes existentes nos fragmentos de DNA que foram inseridos nos vetores de clonagem. Para esta busca, é necessário que sejam realizados ensaios voltados para a expressão de determinada função. Após a confirmação, clones com atividade positiva são sequenciados em busca de possíveis ORFs.

2.3.1 Bibliotecas Metagenômicas

A bioprospecção em bibliotecas metagenômicas tornou-se uma vertente importante na área biotecnológica. A busca por genes que codificam produtos de interesse biotecnológico em bibliotecas metagenômicas vem sendo amplamente explorada à procura de recursos que sejam aplicáveis em setores agrícolas, farmacêuticos e industriais. O desenvolvimento de novas tecnologias proporcionou o acesso a inúmeros produtos oriundos de microrganismos (SIMON; DANIEL, 2009).

Vetores de clonagem podem ser classificados de acordo com o tamanho do fragmento de DNA exógeno que comportam. A escolha do vetor está relacionada ao tipo de informação que se deseja obter.

Plasmídeos são vetores que suportam fragmentos pequenos, que variam de 5 a 10 kb. O uso destes vetores para a construção de bibliotecas de pequenos insertos é comum pela facilidade de clonagem e manipulação (MOE et al., 2010). Plasmídeos podem oferecer até centenas de cópias por célula e continuarem estáveis. Como a maior parte dos genes que codificam enzimas de origem microbiana possuem o tamanho de 1 a 2 kb de comprimento, é comum o uso de plasmídeos para a construção de bibliotecas metagenômicas onde o foco de busca são novas enzimas. Uma das limitações no uso de plasmídeos está no fato de que algumas sequências de ORFs (*Open Reading Frame*) são maiores que a capacidade do vetor e por isso podem ser truncadas (YUN et al., 2004).

A construção de bibliotecas de grandes insertos, que suportam fragmentos de aproximadamente 50 kb, aumenta a probabilidade de fragmentos possuírem ORFs inteiras. Vetores para grandes insertos são comumente utilizados em montagens de genomas (MOE et al., 2010). Vias metabólicas para a biossíntese de metabolitos secundários em bactérias são relacionadas a genes que estão fortemente agrupados. Vetores para grandes insertos também são utilizados na identificação de grupos de genes (VICK et al., 2011).

2.3.2 Triagem de bibliotecas metagenômicas baseada na função de genes

A maioria dos estudos envolvendo metagenomas está relacionadas a um dos três tipos de abordagens a seguir: a procura de fenótipos em bibliotecas metagenômicas e posterior tentativa de determinar a origem filogenética; a classificação filogenética através de busca por âncoras filogenéticas e posterior varredura em busca de genes de interesse e que caracterizem a função do organismo dentro do bioma; e por último, a montagem completa de genomas e a partir disso determinar a função específica de cada gene (VICK et al., 2011).

A caracterização de organismos através de estudos relacionados à função de genes passou a ser uma das principais vertentes de estudos metagenômicos. Além de possibilitar a descrição funcional de espécies em seu habitat, a técnica permitiu o acesso a microrganismos não cultiváveis como fontes de recursos aplicáveis em atividades biotecnológicas (SIMON; DANIEL, 2009).

Dois tipos de abordagem são utilizados para a busca de informação relacionada à função gênica dentro de bibliotecas metagenômicas: em uma, baseada na sequência do fragmento, clones são sequenciados em busca de genes de interesse. Esta técnica é adequada quando há sequências de genes homólogos depositados em bancos de dados. A outra abordagem baseia-se na detecção de um fenótipo através de testes de atividade. Uma das vantagens desta abordagem está na quantidade de clones que podem ser testados em um curto espaço de tempo (MOE et al., 2010).

2.4 EXPRESSÃO DE ENZIMAS

A engenharia genética tornou possível a produção de enzimas em larga escala. A tecnologia do DNA recombinante permitiu que genes ligados a expressão

de enzimas fossem transferidos a outros organismos que se tornam então geneticamente modificados, que expressam a enzima de interesse em condições controláveis, levando à super produção da enzima alvo (BON et al., 2008). Escolher de forma assertiva as ferramentas a serem utilizadas no desenvolvimento do bioprocesso, como a seleção da célula hospedeira e as condições para o desenvolvimento da cepa, a escolha de um vetor de expressão estável e que tenha um promotor de transcrição forte, que permita a transcrição eficaz do gene heterólogo, permitem que a enzima seja obtida em grandes quantidades, tornando o produto compETitivo no mercado (KUPPERS et al., 2014).

Dentre os muitos sistemas de expressão existentes, o mais utilizado é a bactéria Gram negativa *Escherichia coli*. Esta preferência se deve ao rápido crescimento das cepas, o baixo custo para a manutenção do sistema e o constante aperfeiçoamento do sistema, diminuindo vieses como a diferenças relacionadas ao códon utilizado, ineficiência na formação de pontes dissulfeto, dificuldades em expressar proteínas de membrana, assim como a expressão de proteínas secretadas (ASSENBERG et al., 2013; DE MARCO, 2012).

Vetores de expressão em geral são geneticamente arquitetados para obterem elementos que induzam a expressão heteróloga de forma eficaz. São compostos por *replicons*, marcadores de seleção, promotores e terminadores de transcrição.

Promotores de transcrição são projetados para aumentar os níveis de expressão do gene na presença de indutores. A atividade de um promotor fortemente induzido garante que a maior parte da expressão dentro da célula seja do gene de interesse. A superexpressão precoce de proteínas heterólogas pode interferir no crescimento celular. Baixos níveis de expressão na ausência do indutor reduzem níveis de estresse celular e eventual perda do plasmídeo, caso o produto da expressão seja tóxico à célula (BROADWAY, 2013).

2.5 EXPRESSÃO DE FENÓTIPOS

Análises baseadas na função são iniciadas com a identificação de clones que expressam determinado fenótipo, seguido pela caracterização dos clones positivos por análises bioquímicas e sequenciamento. Clones de bibliotecas metagenômicas são submetidos a testes que permitem a detecção da expressão de determinado fenótipo (REYMOND; FLUXA; MAILLARD, 2008). A primeira triagem em busca de

atividade biológica é comumente realizada em placas de ágar, em que a detecção de atividade acontece pela observação de mudanças físicas no meio de cultura (UCHIYAMA; MIYAZAKI, 2009). Após esta primeira triagem, o DNA plasmidial é sequenciado. As sequências obtidas são montadas por sobreposição e confrontadas com bancos de dados com o intuito de serem identificadas possíveis sequências de ORFs que possam ser responsáveis pela atividade funcional do clone (YUN et al., 2004).

A busca por genes é direcionada pela procura de janelas de leitura (*Open Reading Frames*). ORFs são regiões presentes em genes e tem como função, a codificação de proteínas. À montante possuem um códon de iniciação e à jusante são marcados por um códon de terminação. São identificadas pelo alinhamento de sequências de genes ortólogos ou pelo alinhamento de sequências de genes homólogos de uma mesma família (BRENT, 2005). A predição de genes também pode ser realizada pelo uso de sequências não redundantes de aminoácidos (VENTER et al., 2004).

Após a realização da busca por ORFs, aquelas que estão relacionadas à atividade enzimática estudada, tem a sequência reinserida em vetores de superexpressão para a caracterização da nova enzima.

2.5.1 Limitações das análises baseadas em função

A probabilidade de identificar a função de um gene heterólogo é influenciada por fatores como o sistema vetor-hospedeiro, o tamanho do gene de interesse, quão abundante no ambiente é a espécie da qual o gene é oriundo, o tipo de ensaio enzimático e a eficiência na expressão deste gene na célula hospedeira (UCHIYAMA; MIYAZAKI, 2009).

O uso predominante de cepas de *Escherichia coli* como hospedeiras na construção de bibliotecas metagenômicas traz algumas limitações à expressão de alguns genes de DNA metagenômico (MCMAHON et al., 2012).

A triagem funcional em bibliotecas metagenômicas é realizada principalmente através de ensaios enzimáticos em meio sólido. Testes em placas de ágar permitem a verificação de uma mesma atividade em centenas de clones ao mesmo tempo. Porém o método traz algumas limitações, pois mesmo com o uso de ferramentas que possam induzir a expressão ou o aumento da replicação dos vetores, há a

possibilidade da proteína ser pouco secretada, dificultando a análise visual da atividade biológica (MEWIS; TAUPP; HALLAM, 2011).

O uso de cepas de *Escherichia coli* como hospedeiras em processos de expressão heteróloga já é bem fundamentado. A escolha de *E. coli* como sistema modelo está relacionada à alta taxa de crescimento em meios quimicamente definidos, além da facilidade de transformação e manipulação genética ajudaram a consolidar o uso deste organismo. O conhecimento sobre o metabolismo destes organismos permitiu o desenvolvimento de diversas cepas e dezenas de vetores compatíveis (FAKRUDDIN et al., 2013). Apesar de *E. coli* ser um organismo já estabelecido como modelo em expressão de proteínas recombinantes, alguns fatores podem dificultar a recuperação da proteína heteróloga. A recuperação de produtos recombinantes é simplificada quando estas são secretadas para o meio externo ou para o espaço periplasmático. Além disso, proteínas recombinantes quando são secretadas apresentam-se mais estáveis e solúveis, o que leva ao aumento da atividade biológica. As vias de secreção presentes em cepas de *Escherichia coli* podem não ser capazes de transportar determinadas proteínas e o acúmulo destas no meio citoplasmático pode levar à formação de corpos de inclusão, dificultando o processo de recuperação de proteínas ativas. Proteínas que são secretadas para o meio extracelular ou espaço periplasmático são mais estáveis por que estão menos expostas à atividade de proteases, presentes em maior número dentro do citoplasma. Além disso, *E. coli* é um organismo não secreta altas quantidades de proteína naturalmente (SCHLEGEL et al., 2013).

Esta dificuldade em exportar a proteína pode estar relacionada à preferência que a *E. coli* tem de usar outros códons, podendo levar à substituição de aminoácidos presentes no peptídeo sinal, diminuindo a eficiência na exportação (ZALUCKI; BEACHAM; JENNINGS, 2009).

Uma das alternativas para otimizar o processo de produção de proteínas recombinantes é substituição da *E. coli* por cepas de bactérias mais suscetíveis a secretarem proteínas. *Bacillus sp.* possuem algumas vantagens sobre cepas de *E. coli* como hospedeiros em processos de expressão heteróloga. São bactérias gram-positivas que possuem um eficiente sistema de secreção de proteínas, não fazem uso frequente de códons incomuns, além produzirem proteínas de melhor qualidade (VAVROVA; MUCHOVA; BARAK, 2010).

A triagem de bibliotecas metagenômicas em busca de novas biomoléculas é um processo que exige que o fragmento de DNA exógeno seja retransformado e subclonado para que a atividade seja confirmada. Da montagem da biblioteca até a obtenção de uma possível sequência ligada à ORF de gene de interesse, é um processo delongado, devido à incessante necessidade de reconfirmar a atividade do fragmento a cada processo.

A busca por ORFs através do alinhamento de sequências tem como limitante o número de sequências com função conhecida depositadas em bancos de dados. Janelas de leitura que não possuem sequências de genes órtologos ou homólogos depositadas, não podem ter a sua atividade predita.

2.6 BUSCA POR NOVAS ENZIMAS USANDO FERRAMENTAS LIGADAS À PROTEÔMICA

O desenvolvimento de técnicas para identificação de proteínas em larga escala, que envolvem uma abordagem proteômica como ferramentas ligadas à caracterização de proteínas por espectrometria de massa, estão sendo cada vez mais utilizadas na caracterização proteica em amostras complexas (AEBERSOLD; MANN, 2003).

Peptide Mass Fingerprint é um método utilizado em espectrometria de massa para correlacionar sequências de aminoácidos a uma determinada proteína. Proteínas com diferentes sequências de aminoácidos são submetidas à proteólise por proteases que irão gerar fragmentos específicos, apenas encontrados na proteína em questão. Desta forma, se a sequência de aminoácidos estiver depositada em um banco de dados, então é possível correlacionar ambas as sequências e determinar qual é a proteína (AEBERSOLD; GOODLETT, 2001).

Para que este mapeamento baseado na massa de peptídeos ocorra, é necessário que a proteína esteja isolada. É necessário que haja uma separação prévia, ou utilizando técnicas de gel 2DE ou por cromatografia líquida (THIEDE et al., 2005).

Porém, estas técnicas estão sendo aprimoradas, permitindo a busca por enzimas em amostras complexas. Akeroyd e colaboradores (2013) utilizaram fragmentos proteicos adquiridos por espectrometria de massa para a identificação

de proteínas superexpressas provenientes de biblioteca de cDNA. Enzimas secretadas por *B. subtilis* foram usadas após a superexpressão. A proposta deste trabalho é utilizar técnicas de alto rendimento por LC- MS/MS com o intuito agilizar a busca pelas proteínas secretadas dentre os clones da biblioteca. A técnica está relacionada a obtenção da sequência de aminoácidos presentes em peptídeos por espectrometria de massa.

A proposta deste trabalho é apresentar uma metodologia para caracterizar celulasas provenientes de clones de bibliotecas metagenômicas de grandes e pequenos insertos como prova de conceito. Por meio da análise do extrato celular utilizando ferramentas baseadas em proteômica, o sequenciamento de fragmentos obtidos por MS/MS e identificação da proteína por meio da comparação das massas e das sequências dos fragmentos bancos de dados, além de comparar essas sequências proteicas com as proteínas hipotéticas das ORFs identificadas em clones metagenômicos. Dessa forma, a intenção é diminuir a dependência das análises do sequenciamento do fragmento do DNA exógeno em busca de ORFs para a obtenção da identidade da proteína.

2.7 BIBLIOTECAS METAGENÔMICAS UTILIZADAS

A relação de organismos entre si e com o ambiente ao qual pertencem permite a seleção de um nicho a ser amostrado de acordo com o tipo de função a qual se deseja explorar. A busca por glicosil hidrolases tende a ser realizada em ambientes onde há o acúmulo e degradação de matéria orgânica vegetal, como por exemplo, o trato intestinal de alguns animais, como ruminantes (HESS et al., 2011). Microrganismos provenientes de solo também são explorados em busca de enzimas que hidrolisem matéria lignocelulósica (DANIEL, 2011).

Clones para a realização deste trabalho foram obtidos de duas bibliotecas metagenômicas, uma de pequenos insertos, em que rúmen caprino foi a amostra utilizada para a extração de DNA metagenômico (CUNHA et al., 2009), e uma de grandes insertos, onde o DNA foi extraído de amostra proveniente de solo de floresta Amazônica (BERGMANN, et al., 2013). Celulasas são o objeto de busca em ambas as bibliotecas.

3. JUSTIFICATIVA

A abordagem atual para a busca de novas enzimas em clones metagenômicos despende muito tempo entre retransformações e reconfirmações da atividade, além da dificuldade de se obter uma sequência de ORF com homologia em bancos de dados, de forma a se confirmar a identidade da nova enzima expressa. A abordagem proposta neste trabalho tem como objetivo testar a atividade enzimática de clones metagenômicos, e a partir do cultivo deles em meio líquido e análise do extrato celular, tentar identificar a enzima de interesse por meio de técnicas de LC-MS/MS de alto desempenho. Dessa forma, a partir do extrato celular contendo a enzima ativa, pequenas sequências de aminoácidos tripticas obtidas poderiam servir de base para a busca da sequência da ORF relacionada que não possuía semelhança suficiente em bancos de dados. A utilização de técnicas baseadas em proteômica na busca por novas enzimas em bibliotecas metagenômicas é uma abordagem inovadora, que tem como intuito acelerar e auxiliar a busca por novas enzimas.

4. OBJETIVOS

4.1 OBJETIVO GERAL

Utilizar técnicas baseadas em espectrometria de massa para identificar celulasas expressas em clones metagenômicos que possuem ORFs sequenciadas.

4.2 OBJETIVOS ESPECÍFICOS

- Ratificar a atividade enzimática em meio líquido dos clones obtidos em bibliotecas metagenômicas, tendo em vista que os clones foram testados apenas em meio sólido;
- Tripsinizar e fracionar os extratos celulares por cromatografia líquida de alta eficiência;
- Acoplar o fracionamento dos peptídeos trípticos a um espectrômetro de massa do tipo *electrospray* para a obtenção de espectros de MS/MS;
- Obter as sequências de fragmentos tripsinizados por MS/MS;
- Confrontar os peptídeos sequenciados às sequências proteicas hipotéticas das ORFs do clone metagenômico de forma a identificar a ORF expressa;
- Validar a estratégia comparando a sequência peptídica obtida com a sequência do inserto no clone metagenômico de partida;

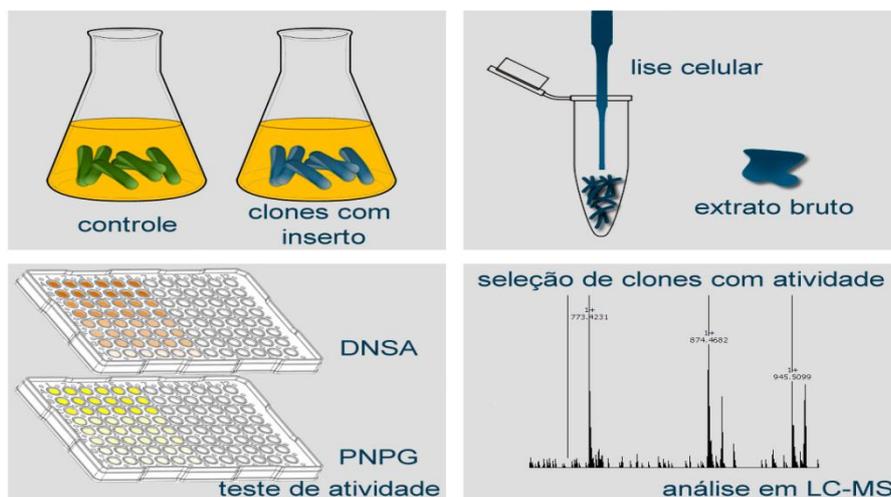
5. MATERIAL E MÉTODOS

5.1 DESENHO EXPERIMENTAL

Para este estudo, foram utilizadas duas bibliotecas, uma de grandes insertos proveniente do DNA extraído de solo Amazônico, de onde foram obtidos inicialmente 15 clones testados para atividade celulolítica e uma de pequenos insertos, proveniente da extração do DNA de microrganismos presentes em rúmen caprino, de onde foi obtido um clone testado para atividade celulolítica. Este último clone está, em trabalho paralelo, tendo sua cinética enzimática caracterizada, portanto foi também objeto de estudo a sua forma superexpressa, onde a sequência da ORF relacionada à expressão da enzima está inserida em um vetor de expressão.

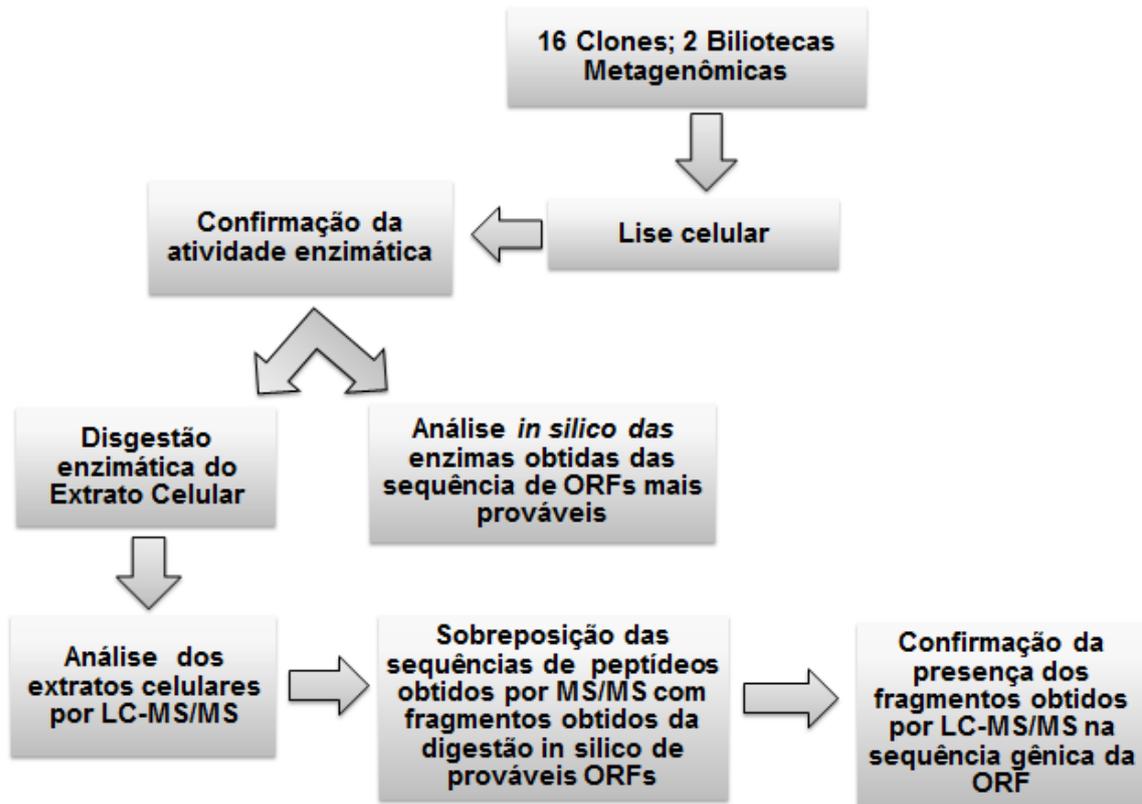
Após a confirmação da atividade enzimática, constituída por triplicatas técnicas e biológicas, os resultados obtidos foram validados por teste estatístico apropriado. Os extratos celulares provenientes dos clones com atividade confirmada foram tripsinizados para a obtenção de espectros por LC-MS/MS. Sequências de aminoácidos foram obtidas por MS/MS e sobrepostas com as sequências de aminoácidos das ORFs obtidas nas sequências gênicas e que estavam relacionadas à atividade enzimática de interesse (Figura 1). Fragmentos trípticos das sequências gênicas foram obtidos *in silico*.

Figura 1. Esquema da estratégia experimental utilizada. Da esquerda para direita, tem-se o crescimento das células; a obtenção do extrato bruto; o ensaio enzimático; a obtenção de fragmentos trípticos pro LC-MS/MS.



In silico, foi realizada a a localização dos fragmentos obtidos por LC-MS/MS no vetor de clonagem. O fluxograma geral do trabalho pode ser observado na figura 2.

Figura 2. Fluxograma do desenho experimental.



5.2 BIBLIOTECAS METAGENÔMICAS ESTUDADAS

5.2.1 Bibliotecas de pequenos insertos

A biblioteca de pequenos insertos, construída pela mestre Isabel Cunha (CUNHA, 2010), é resultante da clonagem de DNA extraído e isolado da microbiota de rúmen caprino no vetor de clonagem PCF430 (NEWMAN; FUQUA, 1999). O DNA total extraído foi digerido com a enzima de restrição PstI (Figura 3). Fragmentos de 2 a 8 kB foram inseridos no vetor e clonados em cepas de *Escherichia coli* EPI300. O vetor PCF430 possui promotor P_{BAD} , fortemente induzido por L-arabinose, o que permite o controle acurado da expressão em cepas de *E. coli* (Figura 4). A regulação ocorre com a proteína AraC, que atua tanto no feedback positivo quanto no negativo. Na ausência de L-arabinose, esta proteína se liga à montante do promotor P_{BAD} , inibindo a expressão.

Figura 3– Mapa do vetor de clonagem PCF430 utilizado na montagem da biblioteca. O DNA clonado foi digerido com a enzima de restrição *PstI* (NEWMAN; FUQUA, 1999).

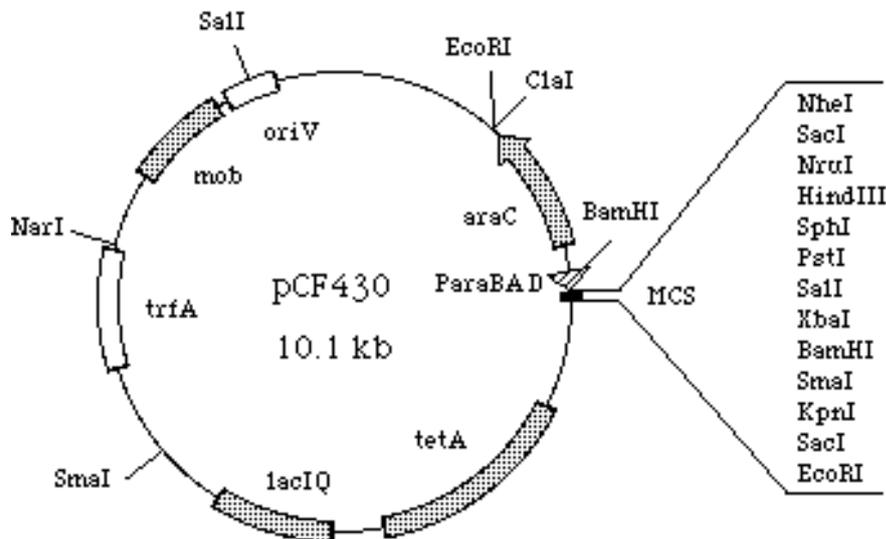
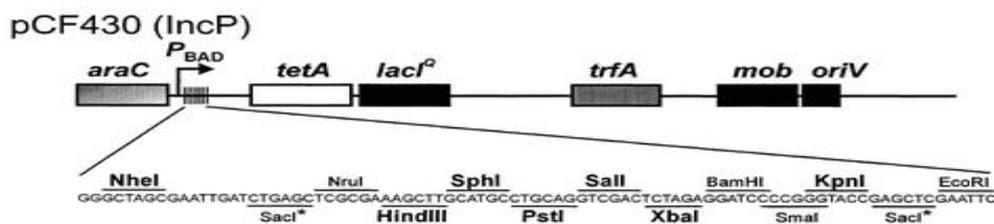


Figura 4 – Representação do promotor P_{BAD} e as posições relativas dos elementos genéticos no vetor PCF430. À jusante do promotor, o gene *AraC* que regula a atividade do P_{BAD} na presença de L-arabinose. O gene *tetA* confere resistência à tetraciclina. O operon *lac*. O gene *trfA* ligado à iniciação de replicação. *OriV* representa a origem de replicação (NEWMAN; FUQUA, 1999).

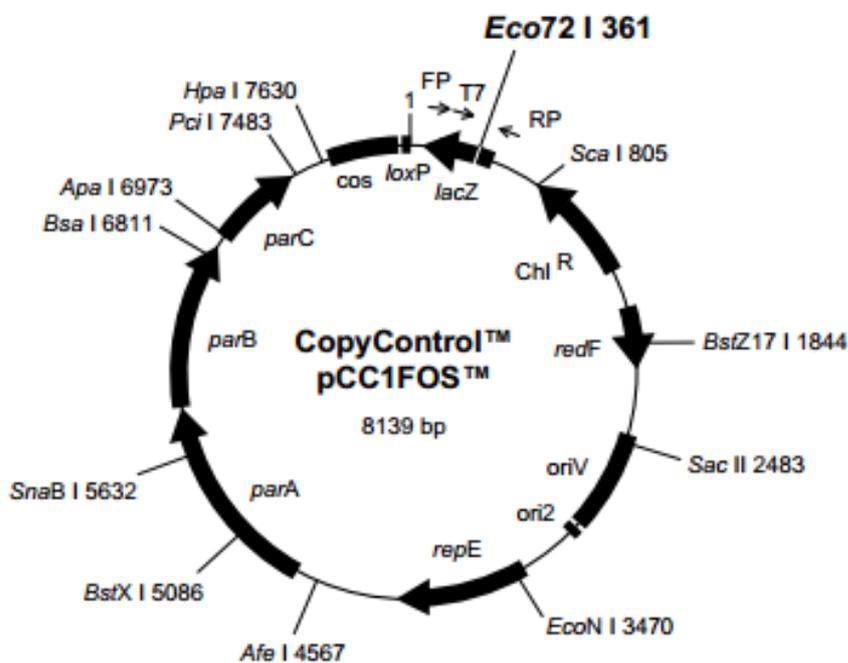


5.2.2 Biblioteca de grandes insertos

O DNA isolado oriundo de solo de floresta Amazônica foi clonado no vetor pCC1FOS (Epicentre). O DNA purificado foi ligado ao vetor, usando *Copy Control Fosmid Library Production Kit* (Epicentre) de acordo com as instruções do fabricante. Fragmentos de aproximadamente 40 kB foram obtidos após purificação em gel de agarose e clonados no fósideo pCC1FOS. Este vetor contém tanto uma origem de replicação para uma única cópia quanto uma origem de replicação para alto número de cópias (*oriV*). Este sítio de replicação para várias cópias precisa do produto do gene *trfA*. A cepa EPI300 de *E. coli* resistente ao fago T1, possui um gene *trfA* mutante que fornece o produto necessário para a iniciação da replicação pelo *oriV*

(Figura 5). Depois de realizada a ligação, foi feito o empacotamento do DNA, utilizando o *MaxPlax Lambda Packing* (Epicentre) de acordo com as instruções do fabricante. O DNA empacotado foi incubado em células de *Escherichia coli* EPI300-T1, fornecidas pelo kit *Copy Control Fosmid Library*. Após a transfecção, as células foram espalhadas em placas contendo *Ágar Luria Bertani* (LBA) e $12,5 \mu\text{g.mL}^{-1}$ de cloranfenicol. Esta biblioteca foi construída pela doutora Jessica Carvalho Bergmann (BERGMANN, 2013).

Figura 5 – Mapa do vetor pCC1FOS (Epicentre).

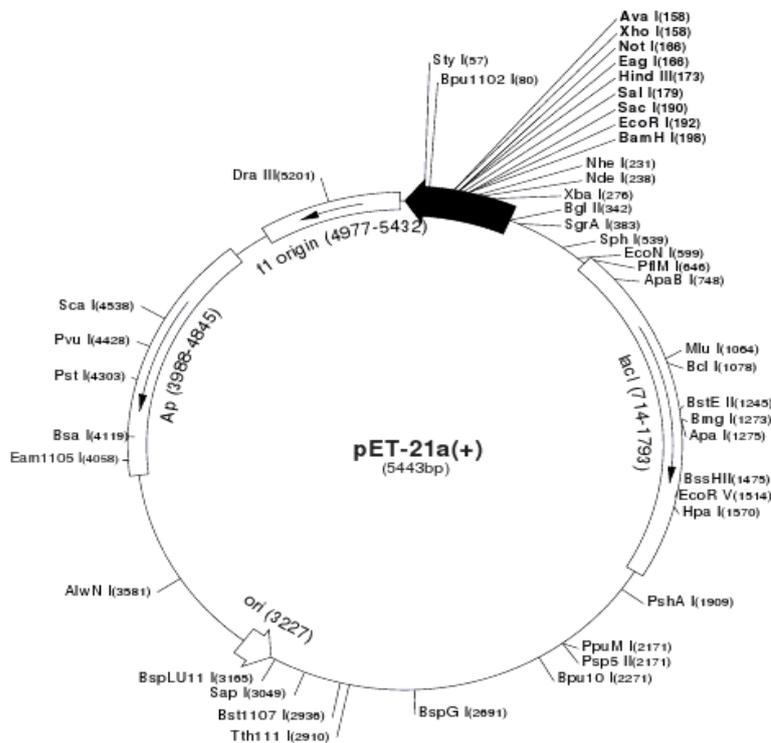


5.3 CLONES METAGENÔMICOS SELECIONADOS

Foram selecionados clones das duas bibliotecas metagenômicas descritas. Da biblioteca de pequenos insertos, construída a partir da microbiota obtida em rúmen caprino Desta biblioteca foi selecionado um clone, que teve o gene de interesse denominado BGL11 por ser positivo para atividade de β -glicosidase. O fragmento obtido deste clone tem aproximadamente 3 kb de tamanho e apenas uma ORF. Foi escolhido pelo fato de já ter sido sequenciado e estarem sendo realizados, em trabalho paralelo, testes a fim de determinar sua cinética enzimática (Souto B., comunicação pessoal). Além de ser utilizado no vetor pCF430, após o sequenciamento e comparação com o banco de dados, o fragmento de interesse foi

inserido no vetor expressão pET21a, clonados em cepas de *Escherichia coli* BL21 DE3. O vetor pET21a possui em sua porção N-terminal o gene para RNA polimerase do bacteriófago T7. Graças à alta seletividade da RNA polimerase de T7, a célula hospedeira é incitada a converter quase todos os seus recursos em prol da expressão do gene alvo. Além disso, possui um promotor induzido por IPTG (isopropil β -D-1-tiogalactopiranosídeo) (Figura 6).

Figura 6 – Mapa do vetor de expressão pET21a (Novagen)



Em sua porção C-terminal uma cauda de histidina opcional, que visa à purificação por cromatografia líquida por afinidade, em colunas de níquel. É interessante a análise destas cepas pelo fato do vetor ter inserido em si apenas o gene alvo, que é expresso em alta quantidade.

Da biblioteca de grandes insertos, obtida a partir da microbiota de solo amazônico, foram selecionados um total de quinze clones, que apresentaram atividade celulotítica de acordo com o ensaio realizado em placas contendo LB ágar e 0,1 % de carboximetilcelulose de baixa viscosidade. Após incubação de 16 dias à 37 °C, as placas foram coradas com vermelho congo, de acordo com Teather e Wood, (1982) com modificações.

Para diminuir o número de candidatos, os clones que tiveram sua atividade enzimática confirmada pela a degradação de CMC *low viscosity* foram também testados para atividade de β -glicosidase. Apenas os clones que foram positivos para as duas atividades, foram utilizados neste estudo. Fragmentos obtidos nos clones desta biblioteca possuem o tamanho de aproximadamente, 50 kb. Como cada ORF possui tamanhos que variam de 1 a 2 kb, é possível que em cada clone desta biblioteca existam de 30 a 50 ORFs.

Nada se sabe sobre as enzimas provenientes da biblioteca de grandes insertos. Por isso é necessário comprovar a efetividade da técnica. A abordagem na biblioteca de pequenos insertos visa à obtenção de um validador para a técnica. Já é conhecida a sequência do gene que codifica a proteínas BGL11, sendo que apenas esta sequência foi ligada ao vetor, sendo assim apenas uma ORF está sendo super expressa, o que poderá facilitar a sua identificação.

5.4 CRESCIMENTO CELULAR PARA TESTE DE ATIVIDADE CELULOLÍTICA *IN VITRO*

5.4.1 Seleção de clones de pequenos insertos

De forma a detectar a atividade celulolítica *in vitro* do clone selecionado, as cepas com o vetor pCF430 vazio (cepa utilizada como controle negativo) e com o vetor pCF430 contendo o inserto do gene BGL11, foram plaqueadas em meio LB ágar e tetraciclina 20 $\mu\text{g}\cdot\text{mL}^{-1}$ e incubadas em estufa por aproximadamente 16 horas a 37°C. Para o pré-inóculo, as células foram raspadas da placa com o auxílio de palitos estéreis e incubadas em 5 mL de meio Terrific Broth (TARTOFF, 1987), contendo tetraciclina 20 $\mu\text{g}\cdot\text{mL}^{-1}$ por 16 horas a 37°C sob 200 rpm de agitação. As cepas foram crescidas em triplicatas. O pré-inóculo foi adicionado a 95 mL de meio TB (TARTOFF, 1987) e tetraciclina 20 $\mu\text{g}\cdot\text{mL}^{-1}$. As células permaneceram sob 200 rpm de agitação, a 37°C até atingirem a OD₆₀₀ entre 0,8 e 1,0, o que ocorreu normalmente após 2,5 h de crescimento. Após as células alcançarem a densidade óptica desejada, foi adicionada às culturas 0,02% de arabinose e estas continuaram incubadas por mais 2,5 horas a 30°C e 200 rpm. Depois do período de crescimento, as células foram centrifugadas por 15 min, a aproximadamente 3640 g, a uma temperatura de 4 °C, sendo o sobrenadante descartado, preservando-se o *pellet* celular. Para remoção de meio de cultura residual, o pellet foi lavado com 2 mL de

solução de STE (Tris HCl pH 8,0 10 mM; NaCl 100 mM; EDTA pH 8,0 1mM) e centrifugado novamente por 15 min, por 3640 g, a 4 °C. O pellet foi ressuscitado em 1 mL tampão Tris-HCl 1 M pH 8,0. Para a lise celular, foi utilizado um sonicador de ponta *Sonics Vibra-Cell VCX 130*, a uma amplitude de 50% em 3 ciclos de 30 s, com intervalos de 10 s. O lisado celular foi centrifugado a aproximadamente 18900 g durante 15 min, a 4 °C. O sobrenadante foi transferido para novo tubo. Para preservar o lisado de degradação por proteases, foi adicionado 100 mM de fluoreto de fenilmetilsulfonil (PMSF).

5.4.2 Clone subclonado no vetor de expressão pET21a

As cepas contendo o vetor de expressão pET21a vazio e com o gene de interesse foram plaqueadas em meio LB e ampicilina 100 $\mu\text{g.mL}^{-1}$. Com o auxílio de palitos de dente, as células foram inoculadas em triplicatas por aproximadamente 16 horas a 37 °C e 200 rpm, em 5 mL de meio TB e ampicilina 100 $\mu\text{g.mL}^{-1}$. Após este período, foi feito o inóculo em 95 mL de meio TB contendo ampicilina 100 $\mu\text{g.mL}^{-1}$. Ao atingirem a OD_{600} no intervalo 0,8 – 1,0, foi adicionado o IPTG a uma concentração de 1 mM. Após o acréscimo do IPTG, as células foram incubadas por 16 horas, a 30 °C sob agitação de 200 rpm. Para a recuperação da massa celular, o volume foi centrifugado a 3640 g por 15 min. As células foram lavadas em 2 mL de solução de STE e novamente centrifugadas por 15 min, a 3640 g. As células foram ressuscitadas em 1 mL de tampão Tris-HCl 50 mM pH 8,0 e lisadas via sonicador de ponta *Sonics Vibra-Cell VCX 130*, a uma amplitude de 50 % em 3 ciclos de 30 s, seguidos de intervalos de 10 s. O lisado foi centrifugado a 18900 g por 15 min a 4 °C. O sobrenadante foi recuperado em novo tubo e nele foi adicionado PMSF 100 mM. Os lisados obtidos durante esta fase foram armazenados a 4°C por aproximadamente 16 horas, pois observou-se melhor conservação e prolongação da atividade hidrolítica quando o armazenamento ocorreu a esta temperatura, por determinado período, para posteriores ensaios de atividade *in vitro*, descritos na seção 5.6. As cepas com o vetor vazio foram utilizadas como controles negativos durante o teste de atividade enzimática.

5.4.3 Seleção de clones de grandes insertos

As células foram retiradas do estoque em -80°C e plaqueadas em meio LB contendo $12,5\ \mu\text{g/mL}$ de cloranfenicol e então incubadas a $37\ ^{\circ}\text{C}$ por 16 h. Foram preparados em triplicatas, pré-inóculos com os clones da biblioteca de grandes insertos, crescidos por 16 horas, a $37\ ^{\circ}\text{C}$ em 5 mL de meio Luria-Bertani (LB) e cloranfenicol $12,5\ \mu\text{g.mL}^{-1}$. Em seguida foram inoculados em 95 mL de meio LB e cloranfenicol $12,5\ \mu\text{g.mL}^{-1}$ e crescidos sob agitação à $37\ ^{\circ}\text{C}$. Após 2,5 h de crescimento foram acrescentados à cultura 0,02% de arabinose. Permaneceram sob agitação por mais 2,5h. As células foram centrifugadas a $3640\ g$ por 15 min. O pellet foi lavado 1 mL de tampão fosfato 50 mM pH 6,0 e centrifugado a $18900\ g$ por 10 min. As células foram ressuspensas em 1 mL de tampão fosfato 50 mM pH 6,0 e lisadas mecanicamente com sonicador de ponta *Sonics Vibra-Cell VCX 130* a uma amplitude de 50% em três ciclos de 30 s com intervalos de 10 s. Após a lise, o extrato celular foi centrifugado por 10 min a $18900\ g$. Aos sobrenadantes, foi acrescentado PMSF a uma concentração final de 2 mM. Para servir de controle negativo, foi utilizada a cepa onde não havia DNA exógeno ligado ao fósforo. Neste caso em específico, o lisado obtido foi imediatamente utilizado no respectivo teste enzimático.

5.4.4 Busca por fases de leitura relacionadas a enzimas de interesse

Para a obtenção das sequências gênicas, foi utilizado o programa IGM/M ER (*Integrated Microbial Genomes/Metagenome "Expert review"*) (MARKOWITZ et al., 2008), que faz uso de diferentes bancos de dados para a obtenção das anotações gênicas de possíveis janelas de leitura (ORFs). A tradução das anotações gênicas para a obtenção das sequências de aminoácidos foi realizada através da ferramenta *ORF FINDER (Open Reading Frame Finder)* (WHEELER et al., 2003), que permite a análise gráfica da sequência, fornecendo o comprimento e em qual dos seis possíveis quadros de leitura a proteína é traduzida. A análise funcional da sequência de aminoácidos foi realizada através da ferramenta *Inter-Pro Scan (Protein sequence analysis & classification)* (JONES, P. et al., 2014), que conduz a identificação de domínios e famílias.

5.5 TESTE DE ATIVIDADE CELULOLÍTICA

5.5.1 β -glicosidase

Para a realização do ensaio enzimático de β -glicosidase, o substrato p-nitrofenil- β -D-glicopiranosídeo (pNPG), cromogênico que muda sua cor para amarelo quando clivado, foi utilizado. Para o teste foram utilizados pNPG 1 mM; Tampão Universal (CAMMACK; ATTWOOD, 2006) 700 mM, pH 5.6; 20 μ L de extrato bruto lisado, totalizando o volume de 100 μ L. A reação teve um tempo de incubação de 30 min a 50°C. Para interromper a reação, foi adicionado ao ensaio, 500 mM de carbonato de sódio. A leitura de absorbância foi realizada em espectrofotômetro de microplaca EON (Biotek), em um comprimento de onda de 405 nm.

5.5.2 Atividade de endoglucanases

Para o teste de atividade enzimática foram utilizados 200 μ L de extrato celular e CMC LV (Carboximetilcelulose Baixa viscosidade) 0,5 %. O CMC LV foi a fonte de carbono utilizada como substrato, de acordo com Ilmberger and Streit, 2010 – com modificações. A reação foi incubada em banho seco a 40 °C durante 24 horas. Após o período de incubação, foram adicionados às amostras 600 μ L de 3,5- ácido dinitrosalisílico (DNSA). Este reagente interage com o açúcar reduzido liberado do CMC. Esta reação faz com que o DNSA mude de cor, passando de um alaranjado para um marrom bem escuro. Após acrescentar o DNSA, a reação foi fervida por 10 min a 95 °C. A leitura do ensaio foi realizada em um espectrofotômetro de microplaca EON (Biotek), em um comprimento de onda de 540 nm.

5.6 ANÁLISE ESTATÍSTICA

Os dados foram apresentados como média e desvio padrão (média \pm DP) dos resultados das réplicas técnicas e biológicas. O teste de hipótese utilizado foi o teste t-Student bicaudal, considerando-se $p < 0,05$. Para as análises estatísticas foi utilizado o programa GraphPad InStat 5.1.

5.7 ELETROFORESE EM GEL DE POLIACRILAMIDA (SDS-PAGE)

A análise do perfil proteico dos extratos celulares com resultado positivo nos ensaios enzimáticos foi realizada por eletroforese em gel de poliacrilamida SDS-

PAGE (LAEMMLI, 1970). Os extratos foram quantificados por fluorimetria (Qubit-Ivitrogen) e aproximadamente 50 µg de amostra foram aplicados no gel. A malha do gel de empacotamento foi de 5 % e a malha do gel de separação foi de 12 %. O gel foi corado por aproximadamente 16 h sob agitação com o corante *comassie blue g-250*. O marcador molecular utilizado foi o *Broad Range Protein Molecular Weight Marker* (Promega).

5.8 DIGESTÃO DE PROTEÍNAS PARA A AQUISIÇÃO DE LC-MS/MS

Para o processo de redução, 100 µL da solução Ditioneitol (DTT) (GE Healthcare) 50 mM, diluído em bicarbonato de amônio (VETEC) 50 mM, foi adicionado a 60 µL de extrato obtido da lise celular. Esta amostra foi mantida a 70°C por 1 hora.

Para o processo de alquilação, foi adicionada à amostra 100 µL da solução iodoacetamida (GE Healthcare) 50 mM diluída em bicarbonato de amônio (VETEC) 50 mM. Após a adição da iodoacetamida, a amostra foi mantida sob aquecimento a 37°C e por 1 hora, em local escuro.

Para a digestão enzimática do extrato, 20 µL da tripsina imobilizada (Trypsin and Immobilized TPCK – Thermo Scientific) foi utilizada, de acordo com as instruções do fabricante, em 60 µL do extrato tratado (alquilado e reduzido). Esta solução permaneceu sob 200 rpm de agitação a uma temperatura de 37 °C por aproximadamente 12 h. Após a digestão, a amostra foi centrifugada a uma rotação de 18900 g por 15 min, para a remoção da tripsina imobilizada. O sobrenadante foi transferido para um novo tubo de microcentrífuga e armazenadas a -80 °C.

5.9 AQUISIÇÃO DE LC-MS/MS

A separação dos fragmentos trípticos foi executada em sistema cromatográfico proeminente UFLC (Shimadzu), acoplado a um espectrômetro de massa do tipo *electrospray* micrOTOF-Q III, (Bruker Daltonics). Para a fase cromatográfica, foi utilizada a coluna ZORBAX 300SB-C18 MicroBore RR 3.5 µm, 1.0 x 150 mm (Agilent), com fluxo de 0,1 mL.min⁻¹. As amostras foram eluídas em um gradiente linear de 5 a 95 % de acetonitrila e ácido fórmico 0,1 %. A eluição foi monitorada nos comprimentos de onda a 216 e 280 nm, Antes de serem aplicadas

na coluna, as amostras foram centrifugadas a 18900 *g*. As corridas foram realizadas em três tempos de eluição diferentes, de 45, 65 e 95 min.

Os peptídeos trípticos eluídos foram analisados por espectrometria de massa. Em todas as análises, o espectrômetro foi operado em modo positivo, sendo as amostras ionizadas por *electrospray* e com resolução de no mínimo 3000 FWHM (*full-width half-maximum*). O analisador foi calibrado externamente com o calibration Tune Mix ESI (Sigma) com *m/z* variando entre 118,08 até 2721,89. O método utilizado foi *Simple tryptic digest nitrogen collision* (fornecido pelo fabricante do espectrômetro de massa), com modificações. A massa do íon referência foi mostrado com intervalos de 1 min. No modo MS/MS, massas foram selecionadas em um intervalo de 50 a 3000 *m/z*. Para a aquisição dos espectros contínuos (modo MS) a voltagem de colisão utilizada foi de 800 Vpp. Para aquisição no modo MS/MS, a voltagem de colisão foi de 100 Vpp. Tanto no modo MS, quanto no modo MS/MS, alguns parâmetros não foram modificados, como a voltagem do cone 1, que foi de 300 Vpp, a voltagem do cone 2, de 400 Vpp, a energia do hexapolo, de 400 Vpp. A energia do quadrupolo *trap* foi 5 eV e a energia da célula de colisão foi de 10 eV. Em relação aos parâmetros relacionados à fonte, a pressão de nebulização foi de 1,0 Bar e a temperatura foi de 180 °C. O único parâmetro modificado foi o volume do gás nitrogênio, que de 9 l/min foi diminuído para 7 l/min.

Os dados obtidos no espectrômetro de massa foram processados no programa Data analysis 4.1 (Bruker Daltonics). Após a deconvolução dos espectros, foi obtida uma lista de íons trípticos fragmentados por MS/MS. Após a obtenção desta lista, o programa Biotools (Bruker Daltonics), foi utilizado para comparar os íons fragmentados com a sequência de aminoácidos da enzima referência, proveniente da ORF. Para que fosse possível a comparação, uma lista teórica com as massas dos fragmentos trípticos que poderiam ser obtidos foi gerada utilizando a sequência de aminoácidos da enzima referência como molde. Para tal, a ferramenta do Biotools, Sequence editor foi utilizada.

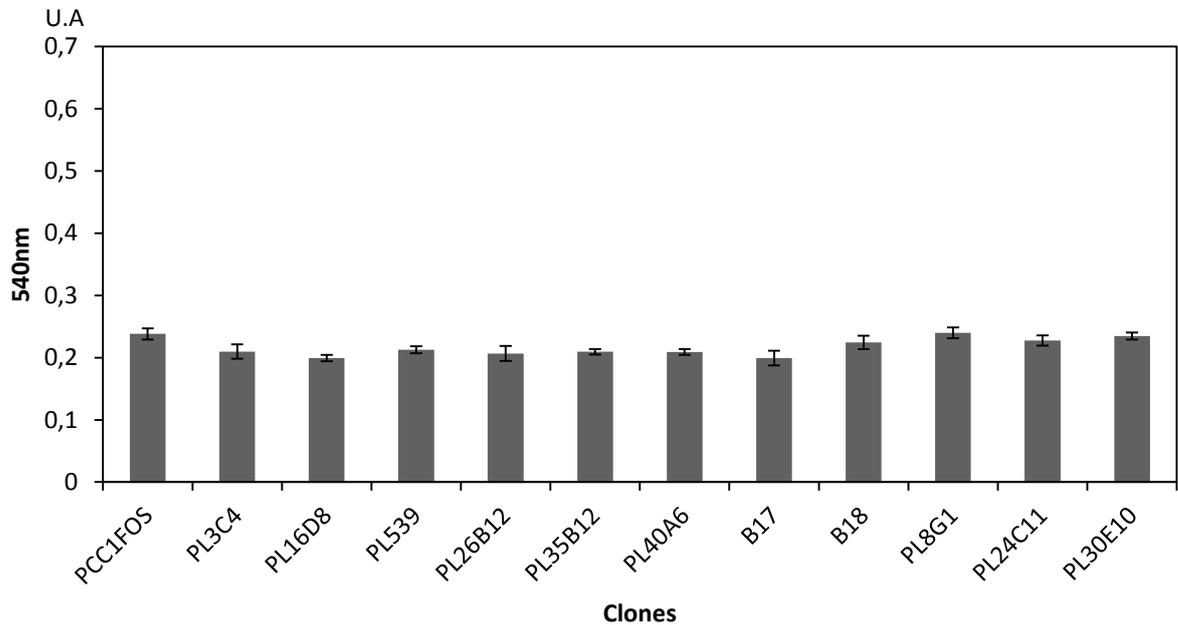
6 RESULTADOS

6.1 ENSAIO ENZIMÁTICO *IN VITRO* PARA A DETECÇÃO DE ATIVIDADE DE ENDOGLUCANASES

Foram utilizados neste teste enzimático, quinze clones da biblioteca metagenômica de grandes insertos de solo amazônico previamente selecionados em ensaio com *congo red*, além de *E. coli* com o fósforo vazio, denominada pCCFOS, utilizada como controle negativo. Ensaio enzimáticos em meio líquido foram realizados com o intuito de comparar quantitativamente os níveis de açúcares reduzidos entre o clone em que o fósforo encontra-se vazio com os clones em que o inserto está no fósforo. O ensaio enzimático foi realizado com extrato celular proveniente dos clones. Consideraram-se clones com atividade celulolítica, aqueles em que o valor da absorbância foi maior do que o da cepa que possuía o vetor vazio. O ensaio foi realizado com triplicatas técnicas e biológicas. Para avaliar a perda de atividade, os ensaios foram realizados com extratos celulares armazenados a -20 °C por aproximadamente 16 h, armazenados a 4 °C por aproximadamente 16 h e extratos que não foram armazenados, onde o ensaio enzimático foi realizado logo após a lise celular. Foram comparados os ensaios realizados com os extratos celulares armazenado a 4 °C e aqueles que não foram armazenados para saber se a estocagem influenciava de forma significativa a atividade enzimática. A validação estatística foi realizada pelo teste *student t*, não pareado bicaudal, com $\alpha < 0,05$.

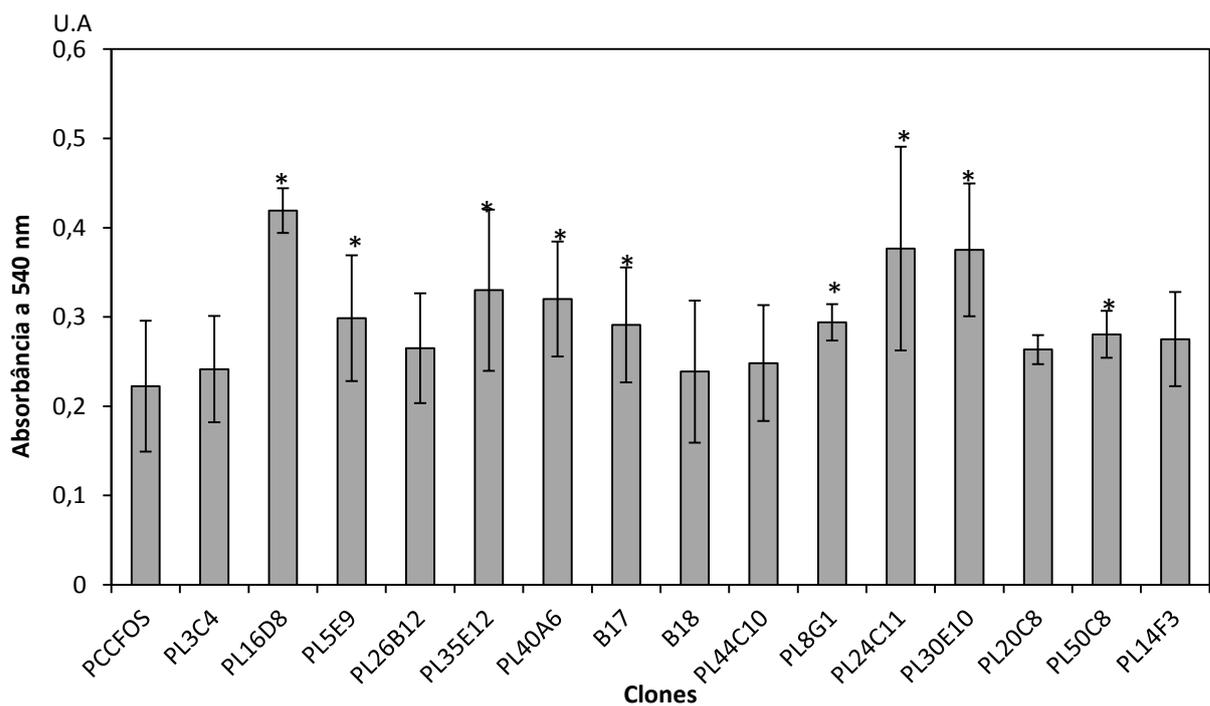
O ensaio enzimático realizado com os extratos celulares estocados a -20 °C, não resultou em atividade significativa para nenhum dos clones testados (Figura 7).

Figura 7- Ensaio enzimático para atividade de endoglucanases realizado com o lisado de células armazenadas a -20 °C. As leituras foram realizadas em uma absorbância de 540 nm.



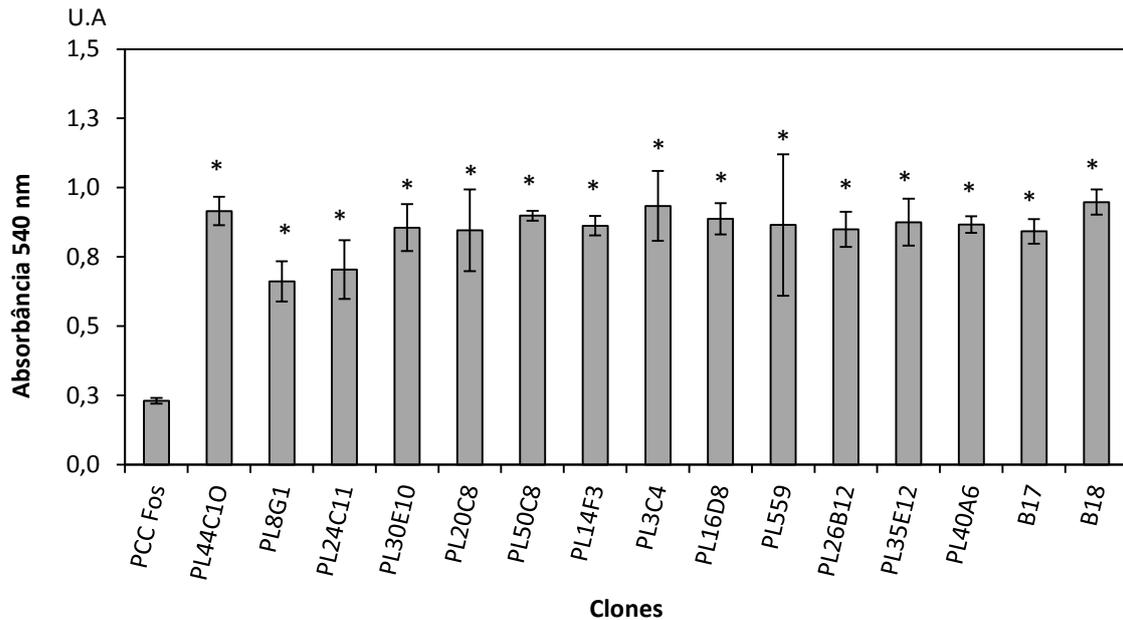
No ensaio enzimático realizado com os extratos celulares que foram armazenados a 4 °C, houve atividade enzimática significativa em 9 dos 15 clones testados (Figura 8).

Figura 8- Ensaio enzimático para atividade de endoglucanases realizado com o lisado de células armazenadas a 4 °C. As leituras foram realizadas em uma absorbância de 540 nm. * indica valores de valores de $p < 0,05$



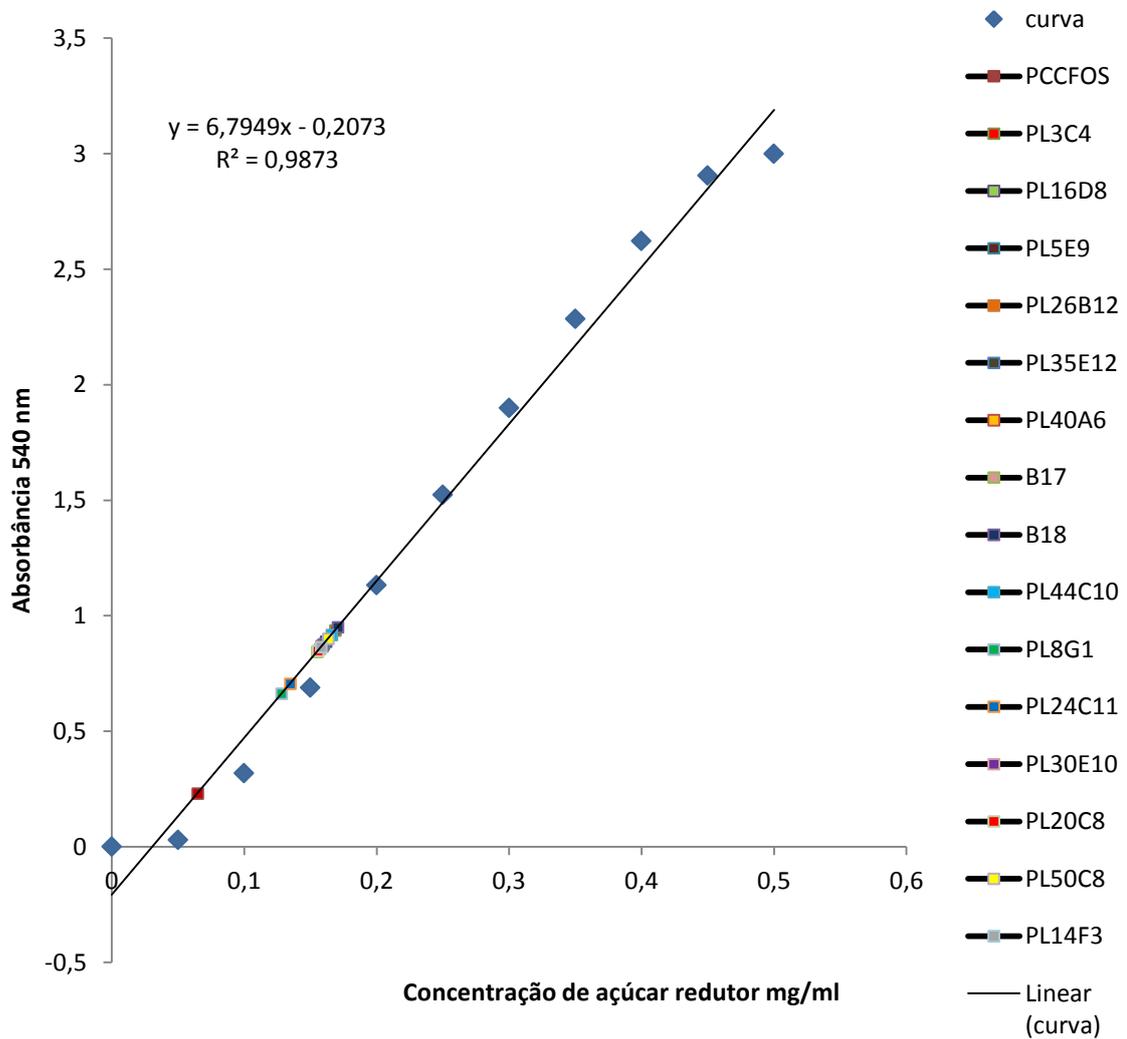
No ensaio enzimático onde os extratos celulares foram utilizados imediatamente após a lise celular, todos os clones apresentaram valores significativos para o teste de endoglucanases (Figura 9).

Figura 9- Ensaio enzimático para atividade enzimática de endoglucanases realizado com o extrato celular que não foi estocado. De acordo com o teste student t, todos os clones obtiveram valores de $p < 0,05$.



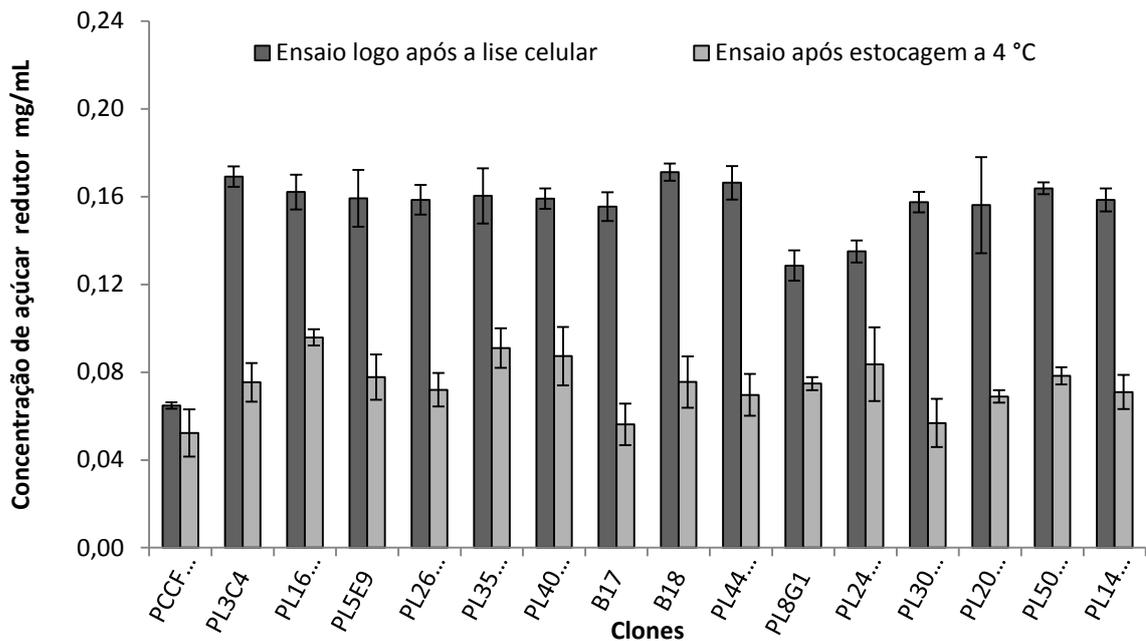
A quantificação de açúcar redutor liberado durante o ensaio enzimático realizado com os extratos celulares que não foram armazenados, foi realizada utilizando-se diluições de glicose 0,2 % para a montagem de uma curva padrão. Os resultados podem ser observados na Figura 10.

Figura 10- Quantificação da quantidade de açúcar redutor liberado pelos extratos celulares dos clones de grandes insertos durante o ensaio enzimático. A curva padrão é representada pelo pontos azuis



Os níveis de açúcar redutor obtidos nos ensaios enzimáticos realizados com os extratos celulares que não foram armazenados e àqueles que foram estocados a 4 °C, foram comparados, mostrando que as enzimas presentes nos extratos celulares possuem sua atividade afetada quando armazenadas (Figura 11).

Figura 11- Comparação dos os níveis de açúcar redutor liberado entre os extratos celulares que não foram armazenados e os que foram a 4°C. As leituras foram realizadas em uma absorbância de 540 nm.



6.2 ENSAIO ENZIMÁTICO *IN VITRO* PARA DETECÇÃO DE ATIVIDADE DE β -GLICOSIDASES

6.2.1 Biblioteca de pequenos insertos

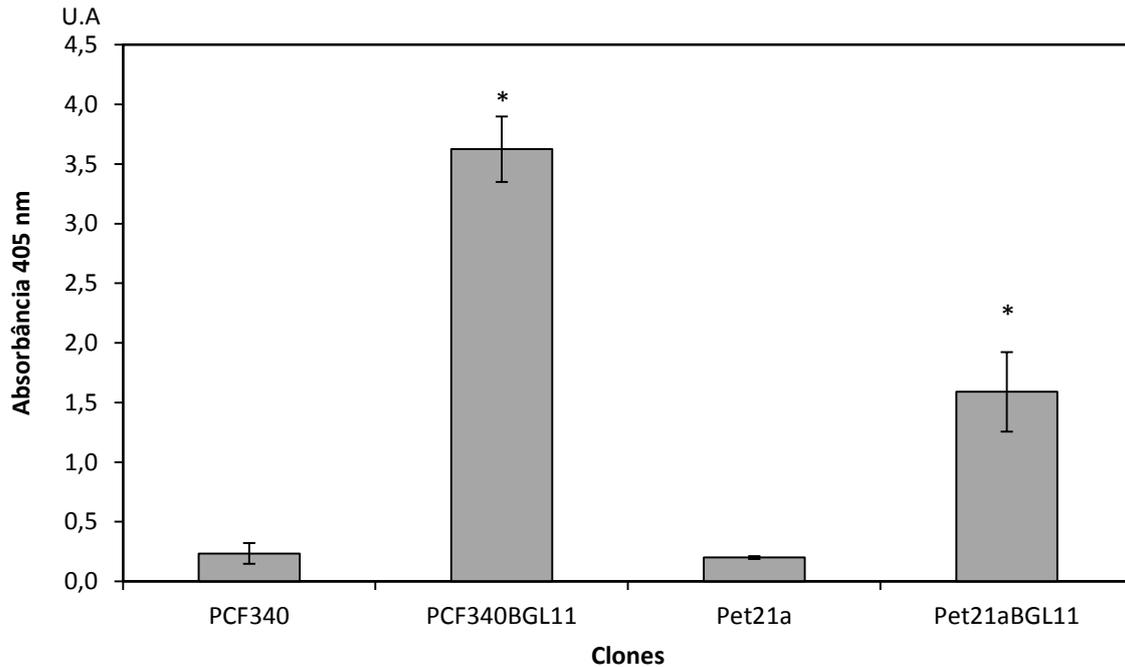
Para averiguar a presença de atividade enzimática e induzir a expressão das enzimas alvo, foi realizado com os clones provindos da biblioteca metagenômica de pequenos insertos, teste enzimático direcionado para atividade de β -glicosidases. Neste ensaio, foram testados os clones com os vetores de clonagem pCF430 e de expressão pET21a . A sequência de ORF presente nos dois vetores é a mesma e está relacionada à enzima denominada BGL11.

Para a validação do teste, as atividades dos clones que possuem o fragmento exógeno inserido nos vetores foram comparadas às atividades dos clones com os vetores vazios, utilizados como controle negativo, denominados de PCF430 e PET21a, respectivamente. Este ensaio permitiu observar que a atividade enzimática

obtida no extrato celular do clone PCF430BGL11 é maior do que no extrato celular do clone que possui o vetor de superexpressão pET21a (Figura 12)

Para a análise estatística, foi utilizado o teste *student t*, não pareado bicaudal com $\alpha = 5\%$ (Figura 12).

Figura 12- Detecção de β -glicosidasas com o substrato pNPG nos clones de pequenos insertos. * indica valores de valores de $p < 0,05$.



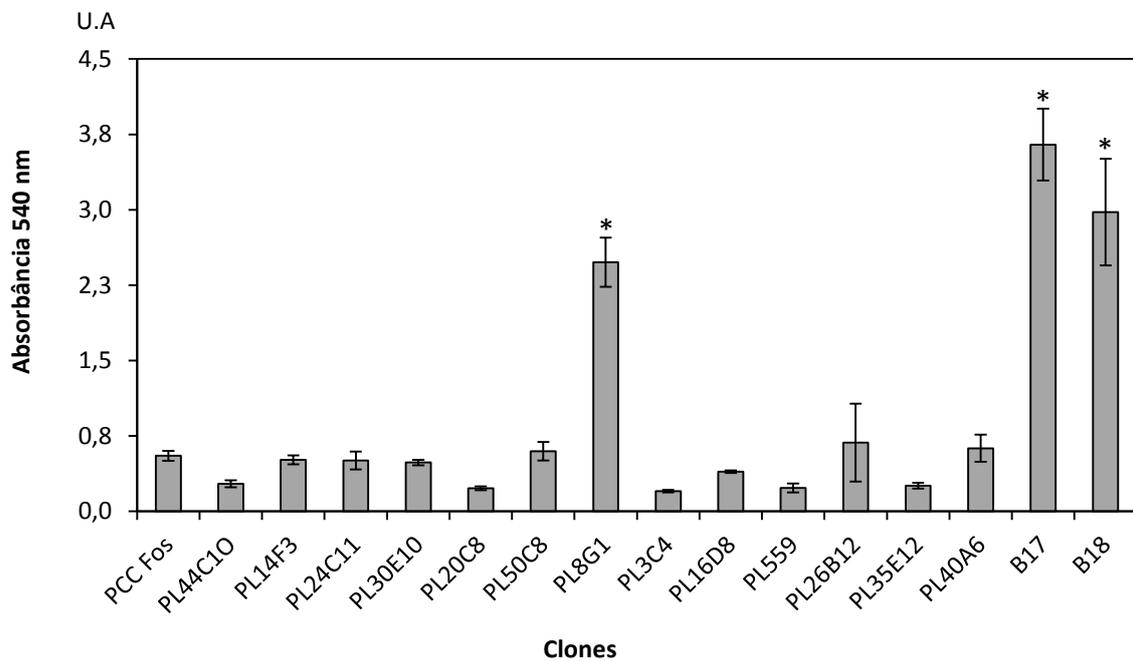
6.2.2 Biblioteca de grandes insertos

Para averiguação da presença de atividade enzimática de β -glicosidasas, foi realizado teste de atividade em meio líquido com os clones da biblioteca de grandes insertos. Neste ensaio foram testados os 16 clones com o substrato sintético pNPG (β -Glucuronidase). Como parâmetro, a cepa que não possuía o inserto, denominada pCCFOS, foi utilizada como controle negativo.

Neste teste, apenas os clones denominados PL8G1, B17 e B18 tiveram resultados significativos (Figura 13).

Para a análise estatística, foi utilizado o teste *student t*, não pareado bicaudal com $\alpha = 5\%$

Figura 13- Detecção de atividade β -glicosidase com o substrato pNPG nos clones de grandes insertos da biblioteca metagenômica de solo da Amazônia. * indica valores de valores de $p < 0,05$.



6.3 BUSCA POR FASES DE LEITURA RELACIONADAS A ENZIMAS DE INTERESSE

Após os ensaios enzimáticos, as sequências gênicas dos clones com atividade enzimática foram analisadas em busca de ORFs que pudessem estar relacionadas à síntese da enzima expressa. Para saber o grau de similaridade, as sequências de aminoácidos das enzimas de interesse foram alinhadas com enzimas afins através do programa CLUSTAL W2. No clone proveniente da biblioteca de pequenos insertos foi encontrada uma ORF situada no quadro de leitura +1, sendo formada por 2427 nucleotídeos, resultando em uma proteína de 808 aminoácidos. Domínios presentes na sequência de aminoácidos, permitiram a identificação da enzima como uma glicosil hidrolase pertencente à família 3 (Figura 14). Esta família compreende, entre outras, enzimas com atividades de β -glicosidases (EC: 3.2.21).

Figura 14- Alinhamento da proteína BGL11 com uma glicosil hidrolase pertencente à família 3.

BGL11	-----MTALLAGCQSVDPVNVEQQVDALYQKMPQEERIAQLKSMYMEELF	45
glicosil	MKKELKTIGSWAMMALAALITSCNTATDM--EQQIDELYQKMPQEERIAQLKSMYMEELF	58
BGL11	DEQGQLDTAKCRELIPYGIGHFSQYAFQKPRDPNVVRDMVAAVQDWMHHTPNGIPALFH	105
glicosil	DADGKLDTAKKELIPYGIGHFSQFCMQQPRDPNELRRAVAIQDWMHHTPNGIPALLH	118
BGL11	EEVLSGINTQGATIYPQQIGQACSFNPELAELKTRQTSTQMRMGVLSLSPMVDVVRNP	165
glicosil	EEVLSGINTLGSITIYPQQIGQAGSFNPELAELKTRQTSTQLRKMGGIFALSPMVDVCRTP	178
BGL11	SFNRLEESYGEDAYLSAVMGTAFAVQGLQQGDLRKGVGACSKHYLYGGGGDADEKELMEE	225
glicosil	SFNRLEESYGEDGYLSAVMGTAFAVNGLQQGDLRKGVGACSKHYLYGGGGDADEKEMMEE	238
BGL11	ILLPHEAMIRVAGSKAVMPGYHAMKSAPVTCVADSFILKDVLRDYLGFDMVSDYTAID	285
glicosil	ILLPHETMIRLAGSVAVMPGYHEVHGT--KCVANSEILQDILRGYVGFDMVSDYTAID	296
BGL11	QLAPSNLPQKGEATVAPEVMKAAMAINGGNDVDFPFGANYKYLQEAIDSGLVKPETLERA	345
glicosil	QIP-----NIDSAVEKAAAAINGGNDVDFPFGANYQLLQEAIDKGLVKPETLERA	346
BGL11	VKNVLRYSKFRAGLFDKPKYLYSTKIVLDTPEERQTAYGIAVESVLLQNDNLPLIQP	405
glicosil	VKNVLRISKFRMGLFDKPKYLYSKEDIKLDTPPEERKTAYDIATQSVVLLNNG-VLPLKE-	404
BGL11	TANSQQPTAKKILVTGPNANSMWAMCGDYSFPAMTFFWKMEEDLDHPHVITLLEGMQNH	465
glicosil	-----VKNVLLTGPANSIWAMCGDYTPAMSYFWKKVDYKSENPHIVKLEGMQNH	456
BGL11	KPDSVDILYSRGCWTEEIEYKSHVGDKRSWEYPLLHRKVNSGETANKAEALAMADSAD	525
glicosil	KPEGVNMYSRGCWTEEIEYKSHVGDARAWEYELLHRKVDAGEKADKAEALQMAKNAD	516
BGL11	VIIAAMGENVMLCGENRDRQGLRPLPGKQEQEYVEALLATGKPVVLLVFGGRAQVIGDLAER	585
glicosil	VI IAAVGENVMLCGENRDRQGLRPLPGKQEQFVEELIQTGKPVVLLVFMFGGRAQVISGLAEK	576
BGL11	CAAVIQAWYPGEEGGHAVADILYGLVSPSAKLSVSYPKKEIYEPICYNYPLSQEPTANSQ	645
glicosil	CAAVIQAWYPGEEGGNAVDILYGLVSPSAKLSVSYPNVEINEPICYNY-----AAEKD	630
BGL11	QPIQWPFYGLTYTTFEYKHLMDQ SARVSADYVELSFEVKNTGKMAATEIAQVYLSVSP	705
glicosil	ARVAWPFYGLSYTTFEYANLEATTEVTTKDESFNLTDFVKNTGKVADEVAQIYLSPAD	690
BGL11	KANSQELKAIKLGQFARVALEPGQTKRVMKLYTDQFGYNTDGGQRQWNIQPGTYTVKVG	765
glicosil	KS--RQIRPIQLQGFARVSLKPGESKKISVKMYTEQLGYTHDGGQRQWNIQPGKYIIVKVG	748
BGL11	ASSADIRLEQDVTLKGERVTKLLRDHYFSEVTITNLE	802
glicosil	ASSQDIKLQQEINLSGEAVKMLPKRYFSEAKQI---	782

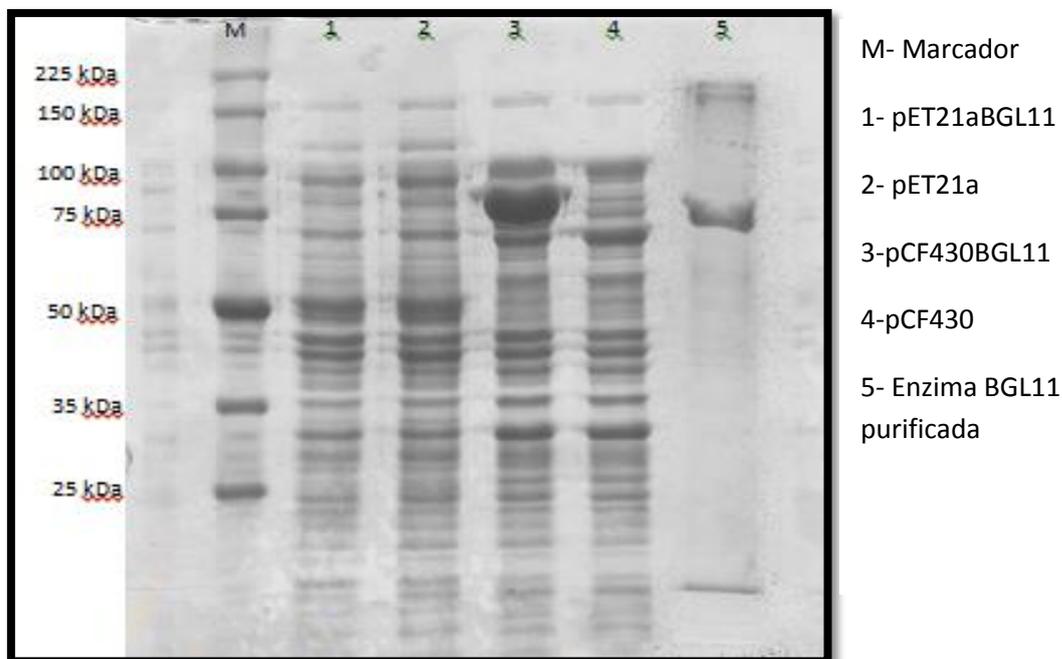
A enzima denominada BGL11, possui um tamanho de 89, 78 kDa. Sua sequência pode ser visualizada na figura 15.

Figura 15- Sequência de aminoácidos da enzima BGL11.

MTALLAGCQSVDPVNVEQQVDALYQKMPQEERIAQLKSMYMEELFDEQGQLDTAKCRELIPYGIGHFSQYAFQKPRDPNVVR
 DMVAAVQDWMHHTPNGIPALFH
 EEVLSGINTQGATIYPQQIGQACSFNPELAELKTRQTSTQMRMGVLSLSPMVDVVRNP
 PSFNRLEESYGEDAYLSAVMGTAFAVQGLQQGDLRKGVGACSKHYLYGGGGDADEKELMEEILLPHEAMIRVAGSKAVMPGY
 HAMKSAPVTCVADSFILKDVLRDYLGFDMVSDYTAIDQLAPSNLPQKGEATVAPEVMKAAMAINGGNDVDFPFGANYKYL
 QEAIDSGLVKPETLERAVKNVLRYSKFRAGLFDKPKYLYSTKIVLDTPEERQTAYGIAVESVLLQNDNLPLIQPTANSQ
 QPTAKKILVTGPNANSMWAMCGDYSFPAMTFFWKMEEDLDHPHVITLLEGMQNHKPDSDVDILYSRGCWTEEIEYKSHVGD
 KRSWEYPLLHRKVNSGETANKAEALAMADSADVI IAAAMGENVMLCGENRDRQGLRPLPGKQEQEYVEALLATGKPVVLLVFGG
 RAQVIGDLAERCAAVIQAWYPGEEGGHAVADILYGLVSPSAKLSVSYPKKEIYEPICYNYPLSQEPTANSQQPIQWPFYGL
 TYTTFEYKHLMDQ SARVSADYVELSFEVKNTGKMAATEIAQVYLSVSPKANSQELKAIKLGQFARVALEPGQTKRVMKLY
 TDQFGYNTDGGQRQWNIQPGTYTVKVGASSADIRLEQDVTLKGERVTKLLRDHYFSEVTITNLE

Os perfis proteicos dos extratos celulares provenientes dos clones da biblioteca de pequenos insertos com atividade enzimática positiva foram obtidos por SDS-PAGE. A enzima purificada foi utilizada como marcador. A amostra foi cedida pela pesquisadora Betulia Souto. A figura 16 mostra que grande quantidade de proteína expressa no clone PCF430BGL11 na faixa de tamanho onde se encontra a enzima de interesse.

Figura 16- Eletroforese em gel de poliacrilamida SDS-PAGE dos extratos celulares provenientes dos clones metagenômicos da biblioteca de pequenos insertos.



Nos clones selecionados da biblioteca de grandes insertos, quatro prováveis ORFs foram obtidas dos clones que tiveram atividade enzimática confirmada tanto para a endoglucanases como para β -glicosidases.

Do clone denominado PL8G1, foi obtida uma sequência de nucleotídeos de uma fase de leitura localizada no quadro de leitura +1, sendo formada por 2208 nucleotídeos e resultando em uma sequência de 735 aminoácidos. A análise funcional da enzima a caracterizou como uma glicosil hidrolase pertencente à família 31 (Figura 17). Nesta família estão compreendidas enzimas que hidrolisam ligações glicosídicas entre dois ou mais carboidratos ou entre um carboidrato e um não carboidrato. Dentre as enzimas pertencentes a este grupo, estão as α -glicosidases (EC: 3.2.1.20)

Figura 17- Alinhamento da proteína PL8G1 com uma glicosil hidrolase pertencente à família 31.

PL8G1	-----	
glicosil	MRSRESFLLKTSLTLKTSLTLILVTEALSAYGQQFKPTGHGVEASADGTV	50
PL8G1	-----MAFRTS	6
glicosil	LQISALREDVLRVRMVKAGKESEDASWAVLPKARKSSVHVVAEPRGFHTQ	100
PL8G1	SLQVTCDPTRLRLTVTDLNGRILQQDARPIQWQQSFRITYKQKFTDSHFFG	56
glicosil	QLRVITIGTDLTLTVADLNGRVLQKDAEPVTWQGDRFRVRKDRSSEDSFFG	150
PL8G1	LGDKPGPLDRAGETFTMWNTDSFGWQES TDPIYKSIPYFIDYRQGRALGI	106
glicosil	LGDKPGPLDRSQAFAMWNTDTFGWQETDPIYKSIPFFIDFKNGRALGV	200
PL8G1	LFDNTWRTFFDFGREVPDQYSFGAPDGPVDYLIYGPDPKHVVETYAWLT	156
glicosil	LVDNTWRSFFDFGRENPNQISFGAPNGPIDYLLYGPPEPKQVVSAYAWLT	250
PL8G1	GPTPLPPLWSLGFQQSRYSYYPEARVMEIADRLRADRIPADALYLDIDYQ	206
glicosil	GPTPLPPLWSLGFQQSRYSYYPASQLMEVAKRLRDKIPSDALYLDIDFQ	300
PL8G1	DHNRPFVDTTKFPHFSQMIHELAQTQFHVVAITDLHIAKLPDANYAPYD	256
glicosil	NKNRPFVDDDEHYPHFSEMVKQLAQDHFHLVITDLHIAHLPNADYAPYD	350
PL8G1	TGIAGDHFLKNPDGSVYTGTVWPGPSVFPDFTRQITRQWWGTLYKDFLND	306
glicosil	SGTAGDQFVKNPDGSTYVGEVWPGPSVFPDFTRQATTRAWWGTLYKDFLND	400
PL8G1	GVAGFWMNDNEPAIFNVPTKTMDDIKHRIAEPGFQERTADHLEIHNVYG	356
glicosil	GVAGFWMNDNEPAVFGVPSKTMDDNIQHRIHEPGFAERTATHAEIHNVYG	450
PL8G1	MENSRYATDGLLALKPNERPVLTRASYAGGQRYAATWSGDNSSTWNHLR	406
glicosil	MENSRGTYEGLLALQPNVRPFVLTRASYAGGQRYAATWTGDNSSTWNHLR	500
PL8G1	MTVPQLVNLGLSGFALSGADVGGFAGSPSPDLLTRWIQLAAFQPIDRDHA	456
glicosil	MTVPQIVNLGLSGFMSGADVGGFAGSPSPDLLTKWIEVAAFQPIDRDHA	550
PL8G1	AKGTRDHEPWVDGPEHEAIRRHFIERYRLMPYLYTTAEETSRTGLPIMR	506
glicosil	AKGTRPHEVWADGPVHEAIRKRYIEERYRLMPYLYTTAEETSRTGLPIMQ	600
PL8G1	PLFLEFPNATTDGHPIDLDSGN-EFLFGPRLLIAPSPSPPEEVGQYEVNLP	555
glicosil	PLFLEFPNANGRAMDLDTGGGEFLFGQHILVAPNPSPEEGAAYEVHLP	650
PL8G1	PGVWFDYWTGERLDRRAQTAARDLEQRDAKQPNKPLLI TPKLAE LPVYVR	605
glicosil	PGTWYDYWSGERYERNQMKTLDEQRDKLTQEKPLMVTPTLEALPVYVR	700
PL8G1	EGSILPIAPLIQSTDETPVGPLTLRIYVGGN----CSGDIYQDDGKSFD	651
glicosil	GGSIIPMAPLTQSVDETPQGPLTVRIYLASDAKEPCAGEVYTDGKSFDY	750
PL8G1	RSQQFLRRHFSCELAQDGS�TVHLDARQGSFAPWVKQVRLETVGWTPQTK	701
Glicosil	RKQYARVHFACAAQADGSLQISVLPQEGKYTPWWSYRFEVLGWTPEK	800
PL8G1	QVTSASGNYYLEQSGNVWTTTIPESSAATDLILH	735
glicosil	KAEVSGRSTSLDTGATAKVTSEAKRNGLIILR	834

A enzima proveniente de uma ORF presente na sequência gênica do clone metagenômico PL8G1 tem o tamanho de 83,33 kDa e a sua sequência de aminoácidos pode ser visualizada na figura 18.

Figura 18- Sequência de aminoácidos da enzima PL8G1.

MAFRTSSLQVTCDPTRLRLTVTDLNGRI LQQDARPIQWQGSFRIYKQKFTDShFFGLGDKPGPLDRAGETFTMWNTDSFGWQEST
 DPIYKSIPIFYIDYRQGRALGILFDNTWRFFDFGREVPDQYSFGAPDGPVDYLIYGPDPKHVVETIYAWLTGPTPLPPLWSLGFQ
 QSRYSYYPEARVMEIADRLRADRI PADALYLDIDYQDHNRPFTVDTTKFPHFSSQMIHELAQTQFHVVAITDLHIAKLPDANYAPY
 DTGIAGDHF LKNPDGSVYTGTVWPGPSVFPDFTRQITRQWWTLYKDFLNDGVAGFWNDMNEPAIFNVPTKMPDDIKHRIAEPG
 FQERTADHLEIHNVYGMENSRATYDGLLALKPNERPFVLRASYAGGQRYAATWSGDNSSTWNHLRMTVPQLVNLGLSGFALS
 DVGGFAGSPSPDLLTRWIQLAAFPIDRDHAAKGRDHEPWVDPGPEHEAIRRHFIERYRLMPYLYTTAEETSRTGLPIMRPLFL
 EFPNATTDGHPIDLDGNEFLFGPRLLIAPSPSPPEEVGQYEVNLPFGVWFDYWTGERLDRRAQTAARDLEQRDAKQPNKPLITP
 KLAELPVYVREGSILPIAPLIQSTDETPVGPLTLRIYVGGNCSGDIYQDDGKSFDFRSGQFLRRHFSCELAQDGLSLTVHLDARQ
 SFAPWWKQVRLTVGWTVPQTKQVTSASGNVYTLQSGNVVWTTTIPESAAATDLILH

Do clone denominado B17, duas possíveis fases de leituras foram encontradas na sequência gênica. A primeira ORF encontrada tem 2253 nucleotídeos e está situada no quadro de leitura +1. Sua sequência de aminoácidos é composta por 750 resíduos. A enzima foi identificada como uma glicosil hidrolase pertencente à família 3 e denominada como B17.1 (Figura 19).

Figura 19- Alinhamento da proteína B17.1 com uma glicosil hidrolase pertencente à família 3.

B17.1	MVEATKSLIEQLTLEEKAAALCVGITAWDTPPIERLGIPTVTVSDGPHGLRRVPDMAG---	57
glicosil	-----MTLEEKAAALITGASAWTTTPIERLGIPELLMSDGPVRRVDPVHALGA	49
B17.1	DSLPTACFPPTASALAASWDVDLIAEQGRALADECIAIGVDVLLGPGINMKRTPLCGRNFE	117
glicosil	PSLPATCFPTAALLASTWDRELLYAMQALAEAAIALGVGVLLGPGVNMKRTPLCGRNFE	109
B17.1	YFSEDPFLAGELAASFINAVQDKGVGTSLKHFVANNQETHRMSISADVDARTLHEIYLAG	177
glicosil	YFSEDPFLAGELAASLVQGIQSKGVGTSLKHYAANNQETQRMRIDAVVDERALREIYLPA	169
B17.1	FERAVKKANPWTVMCGYNRINGIHCSENRWLLTDVLKREWGVDGVAVSDWGAVHDRVKGV	237
glicosil	FERVVKQARPWTVMCAYNKVNGTFCSEHHQLLTQILKDEWGFEGVVVSDWGAVHDRVQAL	229
B17.1	AAGLDLEMPGPQPMRVQRIVEAVQNGILDEAQLDSDVNRLTLTIQRASAVPKGGGSVDVD	297
glicosil	QAGLDLEMPGPKPHRTQAVVDVAVKSGRLNEAVLDEAVRRVLQIVLRAAETPKGG-AFDVD	288
B17.1	KHHALARKVAAETIVLLKNEGDILPLGPADVIAVVGASAQQPHFQGGSSHINATRV DIP	357
glicosil	GHHALARRVAAEGMVLKNGN-ILPLTNPQRIAVIGRAAKEPHFQGGSSHINPTRVDS	347
B17.1	MDELROQAPDAEFVYAPGYTMEEGFDQAMIDEAVKIATGADVALLYIALPPFKESEGYDR	417
glicosil	FDALQARAGGAVLTYCPGYPADDNFDQMLIDEAVTAAQAAEVALLYIALPSYKESEGYDR	407
B17.1	PDIDLSDQQVALIKAVSAVQPRIVILNHGSAIAMSDWIEDVPAVLEAWMMGQAGAGAIT	477
glicosil	ADLDLTRQQVALIQAVSTVQPRCVVILNHGSAVAMNEWIDGVGAVLEAWMMGQAGAGAI	467
B17.1	DVLFGRVNPGRVAETFPLRLSDTPAYLNFPGERGHVRYGEGFLIGYRYYDAKQQEVLF	537
glicosil	DILYGVVNPSPGKLAETFPRRLVDPAYLNFPGENGVRYGEGLYIGYRYYDAKQVAPLFP	527
B17.1	FGFGLSYTTFEMRNHLHLSKSTIREGEPVTATVEVTNTGVSAGKETVQVYVRDVSAVVRP	597
glicosil	FGHGLSYTTFAYSNVRVVSASPFRDVDGVTVSVVDTNTGERAGKEVVQVYVRDRRCSLQRP	587
B17.1	IKELKGFQKAVSLQPGETKTVTIALEDRAFMYDPRKRWIAESGEFEILIGKSAEIVLR	657
glicosil	PKELKGFQKAVSLQPGETKTVSIPDLDFRAFAFWHPGHKQWVTEDEGFDLLIGASSADIRAQ	647
B17.1	QSMMLSHYTERADLSRFSTYGEWLADPRGSAVLRPQLTALFHTVNTPDG---HTVAIVE	714
glicosil	VTVTLQSTLQLPSLLNRESTVREWLEDPRGRAVFEFFQQMMQMRAMGGGEEEGGIGM	707
B17.1	TIP-WLLDVPLDAVLGFFGRNERLTSDEVVANLLKQL--	750
glicosil	DMTGFIMETPLVSILGFQEATLPAPAEVIVADLLQKVHG	746

A enzima obtida no clone B17, denominada B17.1, possui 81,71 kDa de tamanho e sua sequência pode ser observada na figura 20.

Figura 20- Sequência de aminoácidos da enzima B17.1

MVEATKSLIEQLTLEEKAAALCVGITAWDTVPIERLGIPTVTVSDGPHGLRRVPMAGDSLTPATCFPTASALAASWDVDLIAEQ
 GRALADECIAIGVDVLLGPGINMKRTPLCGRNFEYFSEDPFLAGELAASFNAVQDKGVGTSKHFVAVNNQETHRMSISADVD
 ARTLHEIYLAGFERAVKKANPWTVMCGYNRINGIHCSENRLTDLVKREWGVDGVAVSDWGAHVDRVKVAAGLDLEMPGPQ
 PMRVQRIVEAVQNGILDEAQLDDSVNRLTLIQRASAVPKGGGSDVDKHHALARKVAAETIVLLKNEGDILPLGPAVIAV
 GASAQQPHFQGGSSHINATRVDPMDLRRQAPDAEFVYAPGYTMEEGFDQAMIDEAVKIATGADVALLYIALPPFKESEGY
 DRPDI DLSDQQVALIKAVSAVQPRSIIVILNHGSAIAMS DWIEDVPAVLEAWMMGQAGAGAITDVLFGRVNPSGRVAETFPRL
 SDTPAYLNFPGERGHVRYGEGFLFIGYRYDAKQQEVLFPFGFGLSYTTFEMRNLHLSKSTIREGEPVTATVEVTNTGVSAGKE
 TVQVYVRDVSASVVRPIKELKGFVSLQPGETKTVTIALEDRAFMYDPDRKTWIAESGEFEILIGKSAAEIVLRQSMMLLES
 HYTERADLSRFSTYGEWLADPRGSAVLRPQLTALFHTVNTPDGHTVAIVETIPWLLDVLDAVLGFFGRNERLTSDEVVANLL
 KQL

A segunda ORF encontrada no clone B17 é composta por 2007 nucleotídeos e resultou em uma sequência de 668 aminoácidos. A análise funcional identificou a enzima, denominada como B17.2, como uma glicosil hidrolase da família 2 (Figura 21). Dentre as atividades enzimáticas compreendidas neste grupo, encontram-se as β -galactosidases (EC 3.2.1.23).

Figura 21- Alinhamento da proteína B17.2 com uma glicosil hidrolase pertencente à família 2.

B17.2	MLYPQQNSVRQVLP LSGLWEFQ PD DPDQVGEAQA WYDHLPA PRMIAVPA SW	50
glicosil	MLYPQQNLARNKFDLSGIWDFRIDPDLVGEQQGWQNGVNNRPIAVPGSW	50
B17.2	NDQFPDLFNYFGAA WYRCECYV PRAW TDER IMLRGGS ANYAAE VWLN QR	100
glicosil	NEQYEDIYNYLGLAWYIKSTYIPQSWRGQRVFLRVGSACYFGTVYINGVK	100
B17.2	VGDHEGGHLP FE FD ITD IVVWN QPNRLA IRVENELK PTRV PPGNL PPSGP	150
glicosil	MGSHEGGHLPFAFEITDQIRWETENIVAISVENELKPTRVPSGNMSTPLL	150
B17.2	AISGYPNSNF DF FF AGLHR PV MLY TVPK HH I DDV S VI TQ I ADGMS T VEV	200
glicosil	PVASFPHTTYDFPFAGIHRLLVLYSVQTYIDDVTVVTDIDGVNGIIVK	200
B17.2	TVSG-S HR GS GL VR LS GN SSE IE AK F D L R D H V A S A V I T V P N A R L W S P D	249
glicosil	NVRLNSAVSAQGDQLKGGDATIVAKLNFLDGLAEMQLTVANAKFWSDKM	250
B17.2	P F L Y Q L D V E L T E R R R L I D R Y T L H I G I R T V E V Q D H L L N G E P I T L R G F G R	299
glicosil	PYLYDLVIQAG-----QDHYTLKVGIRTVRVDNQILLNGKPKVKNLYGR	295
B17.2	H E D F Y V S G R Q N L P V M I R D Y H L M K W I G A N S F R T S H P Y S E E E M L M A D R M G	349
glicosil	HEDFIASGKGLNPLLVKDYQLMRWTGANSYRTSHYPYSEEEMQLADREG	345
B17.2	FL V I D E T P A V G L T F D D S R E N I A V R L R Q A R Q Q T H E L I A R D K N H P S V I L W S I	399
glicosil	FLIIDEIPAVSLQFEDD-DNIAIRQRICLQQVDELIARDKNHPSVIMWCV	394
B17.2	A N E P A P P K M I E R L I G L E D T P I P E V A T Q S L K S L I D L A H Q L D P T R P A T F A A V	449
glicosil	ANEPMTPRNLNTASEILTEPAVEKKGKFLERLMHRARELDPTRLVTLVTL	444
B17.2	M G G F V E W L T L G D V A C I N R Y W G W V Q P G Q A D V G A Q M L D Q E L D S L Y E A I N K P	499
Glicosil	GGTPSSWVEQCQDVICMNRWGWYVLGGEMDKALAALEQELDAAWDTWRKP	494
B17.2	I M I T E F G A D T L A G M H R Q P P V M W S E E Y Q R E L L R G Y L D A A D S K P F V I G T H I W	549
glicosil	VILTEFGADTLAGMHGHPNVMWTEEYQAEYIRGHLRVAGEKDYIAGMQVW	544
B17.2	N F A D F A A V Q S V L R V G A N L K G V F T R D R Q P K L A A H Y L Q E R W A G R L S D A P R L	599
glicosil	NFADFAAVQSIMRVGGINMKGVFTRGRTPKMAAHVLRFEWAR-----	586
B17.2	P L A E M A Q P V D E T T S P T G V E V A P S I D P N A P F E A I I A A A A R L D G K Y P D L N R	649
glicosil	-----	
B17.2	R L R F Q L G D E G V Y D L I F E A G	669
Glicosil	-----	

A enzima correspondente a uma das ORFs encontradas no clone B17, enzima esta denominada de B17.2, possui o tamanho de 75,26 kDa e sua sequência pode ser observada na figura 22.

Figura 22- Sequência de aminoácidos da enzima B17.2.

```

MLYPQQNSVRQVLPLSGLWEFQPPDPDQVGEAQAWYDHLPAPRMIAVPASWNDQFPDLFNYFGAAWYRCECYVPRAWT
DERIMLRGGSANYAAEVWLNQQRVGDHEGGHLPFEFDITDIVVWNQPNRLAIRVENELKPTRVPPGNLPPSGPAISG
YPNSNFDFPFAGLHRPVMLYTVPKHHIDDVSVITQIADGMSTVEVTVSGSHPRGSGLVRLSGNSSEIEAKFDLRDH
VASAVITVPNARLWSPDDPFYQLDVELTERRRLIDRYTLHIGIRTVEVQGDHLLLNGEPIITLRGFGRHEDFYVSGR
GQNLPMIRDYHLMKWIGANSFRTSHYPYSEEEMLMADRMGFLVIDETPAVGLTFDDSRENIAVRLRQARQOTHELI
ARDKNHPSVILWSIANEPAPPKMIERLIGLEDTPPEVATQSLKSLIDLAHQLDPTRPATFAAVMGGPVVEWLTGLGDV
ACINRYWGWYVQPGQADVGAQMLDQELDSLIEAINKPIMITEFGADTLAGMHRQPPVMWSEEYQRELLRGYLDAADS
KPFVIGTHIWNFADFQAVQSVLRVYGAYNLKGVFTRDRQPKLAAHYLQERWAGRLSDAPRLPLAEMAQPVDETTSPG
VEVAPSIDPNAPFEAIIAAAAARLDGKYPDLNRRRLRFQLGDEGVYDLIFEAGG

```

No clone denominado B18, foi encontrada uma ORF relacionada à síntese de glicosil hidrolase. A busca resultou em uma sequência de 1404 nucleotídeos e 467 aminoácidos. Está situada no quadro de leitura +1. A análise funcional da enzima, denominada B18, resultou em uma glicosil hidrolase da família 1 (Figura 23), que também compreende β -glicosidases.

Figura 23- Alinhamento da proteína B18 com uma glicosil hidrolase pertencente à família 1.

B18	MLRFPDGLFWGAATSSYQIEGATTADGRGESIWDRAARDGTIADGSDGVSACDHYHRWP	60
glicosil	MRQFPNDFLWGVATSSYQIEGATHADGRGESIWDRAATPGKISDGS DGK VACDHYHRWR	60
B18	RDIELMKWLG LGAYRFSIAWPRVFPPTGRWEVNARGLDFYDRLTDGLLEAGIEPFVVTLYHW	120
glicosil	EDVELMRWLG VKSYRFSVAWPRVLP TGRGAVNAAGLDFYSRLVDGLLEAGIEPFVVTLYHW	120
B18	DLPQPLEDAGGWANRETVDAFVEYVDVVSDDLGDVRVRFITHNEPWCISVLGYAEG LHAP	180
glicosil	DLPQILQDQGWPSRDTGNAFVEYADVMSRKLGDVRKRWITHNEPWCISVLGYNGEHAP	180
B18	GRRDWPAAALAAHLLLSHGLATAVIRGNAPDAEVGITLNLVPAEPASPSAHDRDACRAF	240
glicosil	GHKNGGEMLAHLLLVSHGQAVPVIRGNVKAQVGITLNLSPAEPASPSAEDAECRRH	240
B18	DGAFNRWFLDPLYGRGYPEDVIADFGVSGRLPSLELPFVADGDLAVISAPTDFLGINYY S	300
glicosil	DGSFNRWFLDPLYGRGYPKDVVEDYVRDGHLSATLPFVRDGDLETIAVPTDFLGINYY S	300
B18	RAVVRSDRIA EADNAPRAVHLS DERTDIGWEVWPDGLRQLLVRVQRDYAPSAVYITENGA	360
glicosil	RAIMRSSRIPESQNA PRTVHPVEEHTDMGWEVYAPALTRLLVHLHTHYQPGPIHITENGC	360
B18	AYDTAPDEHGRVHDLERQRF LHTHLAAALDAMRDGVPLAGYFVWSLLDNYEQEGYRKR F	420
glicosil	AYATGPSEDGKVHDEKRVAYLRSHLEASLEAIRQGVPLAGYFAWSLLDNFEWAFGYQKR F	420
B18	GIVVVDYTTQERVLKESAAQYRAIIAANGLPAAEPHQDHFQRRILRR	467
glicosil	GIVYVDYDSQRRLPKDSAHLYKALIARNGLDVELAA-----	456

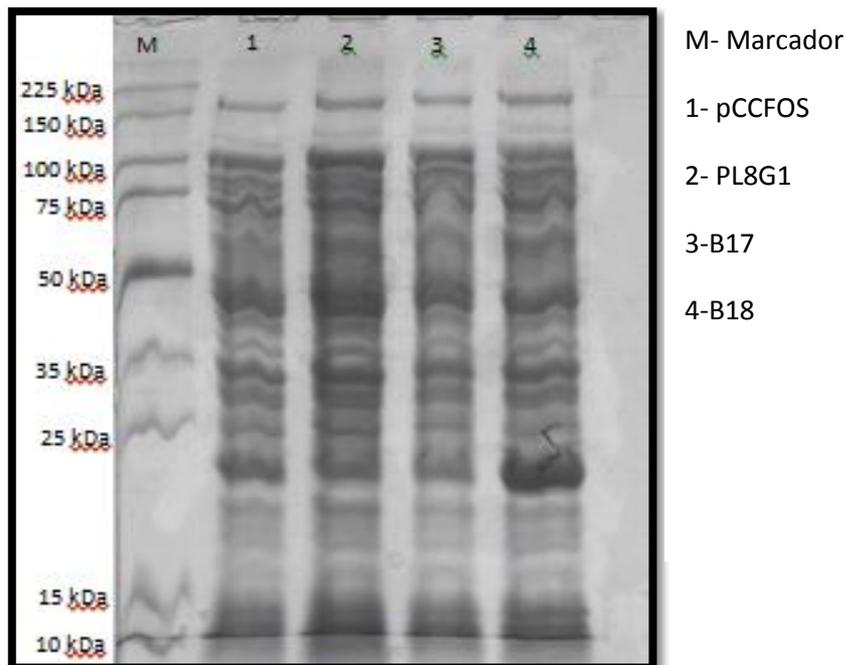
A enzima proveniente do clone B18, possui o tamanho de 52,37 kDa e sua sequência de aminoácidos pode ser observada na figura 24.

Figura 24- Sequência de aminoácidos da enzima B18.

```
MLRFPDGFLWGAATSSYQIEGATTADGRGESIWDRFAARDGTIADGSDGSVACDHYHRWPRDIELMKW
LGLGAYRFSIAWPRVFPTGRWEVNARGLDYDRLTDGLLEAGIEPFVTLYHWDLPQPLEDAGGWANRE
TVDAFVEYVDVVSDDLGDVRRFITHNEPWCISVLGYAEGHLHAPGRRDWPAAALAAHLLLSHGLATA
VIRGNAPDAEVGITLNLVPAEPASPSAHDRDACRAFDGAFNRWFLDPLYGRGYPEDVIADDFVGSGRLP
SLELPFVADGDLAVISAPTDFLGINYYSRVVRSRDAEADNAPRAVHLSDERTDIGWEVWPDGLRQL
LVRVQRDYAPSAVYITENGAAYDTAPDEHGRVHDLERQRFHLTHLAAALDAMRDGVPLAGYFVWSLLD
NYEWQEGYRKRFGIVWVDYTTQERVLKESAAQYRAIIAANGLPAAEPHQDHFQRRILRR
```

Dos clones provenientes da biblioteca metagenômica de grandes insertos, os perfis das amostras que foram positivas tanto para o ensaio enzimático de endoglucanases, quanto para o ensaio enzimático para β -glicosidase foram obtidos. Não houve diferença entre os perfis dos extratos celulares nas faixas de tamanho onde se encontram as referidas enzimas (figura 25).

Figura 25- Eletroforese em gel de poliacrilamida SDS-PAGE dos extratos celulares provenientes dos clones metagenômicos da biblioteca de grandes insertos, que tiveram atividade enzimática para endoglucanases e β -glicosidases.



6.4 AQUISIÇÃO DE LC-MS/MS

Extratos celulares tripsinizados dos clones com atividade enzimática confirmada foram analisados em um espectrômetro de massa do tipo *electrospray* acoplado a um cromatógrafo líquido. Após a obtenção dos espectros MS/MS, os

dados foram comparados com as sequências de aminoácidos traduzidas de ORFs existentes nas sequências gênicas dos clones metagenômicos. Células que continham os vetores sem insertos foram utilizadas como controle negativo. Dados MS/MS foram adquiridos dos extratos celulares dos clones da biblioteca de pequenos insertos, pCF430BGL11, pCF430 (controle negativo), dos clones com vetor de superexpressão pET21aBGL11, PET21a (controle negativo), e dos extratos celulares dos clones da biblioteca de grandes insertos, PL8G1, B17, B18 e pCCFOS (controle negativo). As corridas foram realizadas em triplicatas técnicas. Após a obtenção dos espectros, listas de massas dos íons que foram quebrados no modo MS/MS foram geradas pelo programa *Data analysis*. As listas geradas foram transferidas ao programa *Biotoools*, que através do programa *Sequence editor* correlacionou as sequências de aminoácidos obtidas nas listas de MS/MS com as sequências de aminoácidos provenientes das enzimas traduzidas das ORFs encontradas nas sequências gênicas.

O número de espectros MS/MS obtidos durante a análise nos diferentes tempos de corrida são descritos na tabela 1, assim como o número de espectros em que as massas dos fragmentos obtidos corresponderam a massas da lista teórica e o número de espectros em que foram obtidas séries y e b, principalmente.

Tabela 1- Número de espectros MS/MS obtidos por corrida cromatográfica.

Amostra	Tempo de corrida	Nº de espectros MS/MS	Nº de espectros com massa correspondente na lista teórica	Número de espectros com série y e b
pCF430BGL11	45 min	290	14	1
pCF430BGL11	65 min	529	15	3
pCF430BGL11	95 min	612	18	2
pCF430	45 min	300	0	0
pCF430	65 min	586	18	0
pET21aBGL11	65 min	437	5	0
pET21aBGL11	95 min	548	8	0
pET21a	65 min	477	7	0
pET21a	95 min	550	6	0
PL8G1	45 min	229	0	0

B17.1	45 min	272	3	0
B17.2	45 min	272	9	0
B18	45 min	246	3	0
pCCFOS	45 min	257	13	0

A correlação entre a lista teórica de fragmentos trípticos e a lista de fragmentos obtidos por espectrometria foi obtida apenas com o extrato celular do clone PCF430BGL11. Em nenhum dos outros clones, incluindo o clone com o vetor de superexpressão pET21aBGL11 foi obtida a correlação entre as listas.

A lista de massas, seus respectivos fragmentos e se as massas correspondentes foram também encontradas no controle negativo, podem ser observados na tabela 2.

Tabela 2- Lista das massas dos íons obtidos por LC-MS/MS de fragmentos de peptídeos encontrados na sequência da enzima referência

Massas dos íons [M+H ⁺]	Sequências dos aminoácidos	Número de espectros de mesma massa	Massas correspondentes encontradas nos controles negativos
1131,57	GEATVAPEVMK	2	NÃO
2061,086	YLQEAIDSGLVKPETLER	8	NÃO
2134,93	SMYMEELFDEQQQLDTAK	5	NÃO
1559,82	MGGVLSLPMVDVVR	6	NÃO

Os espectros obtidos representam a sequência de aminoácidos distribuídos nas séries a, y e b (Figuras 26 a 29). O alinhamento dos fragmentos de peptídeos com a enzima referência foi realizado com a ferramenta *clustalW*.(Figura 30).

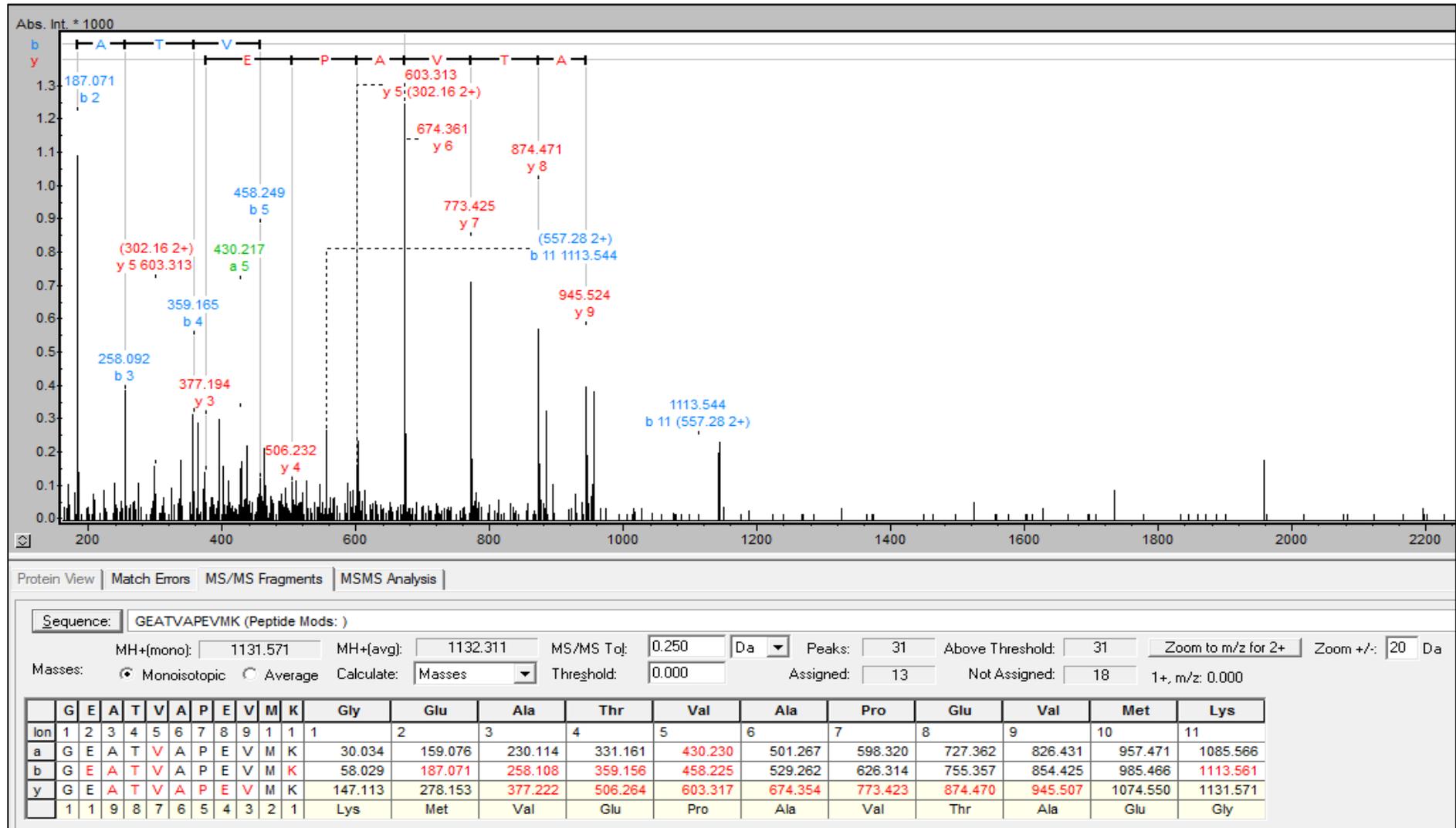
Figura 26- Espectro MS/MS do íon de massa $[M+H]^{+2}$ 566, 29.

Figura 27- Espectro MS/MS do íon de massa $[M+H]^+2$ 687,74.

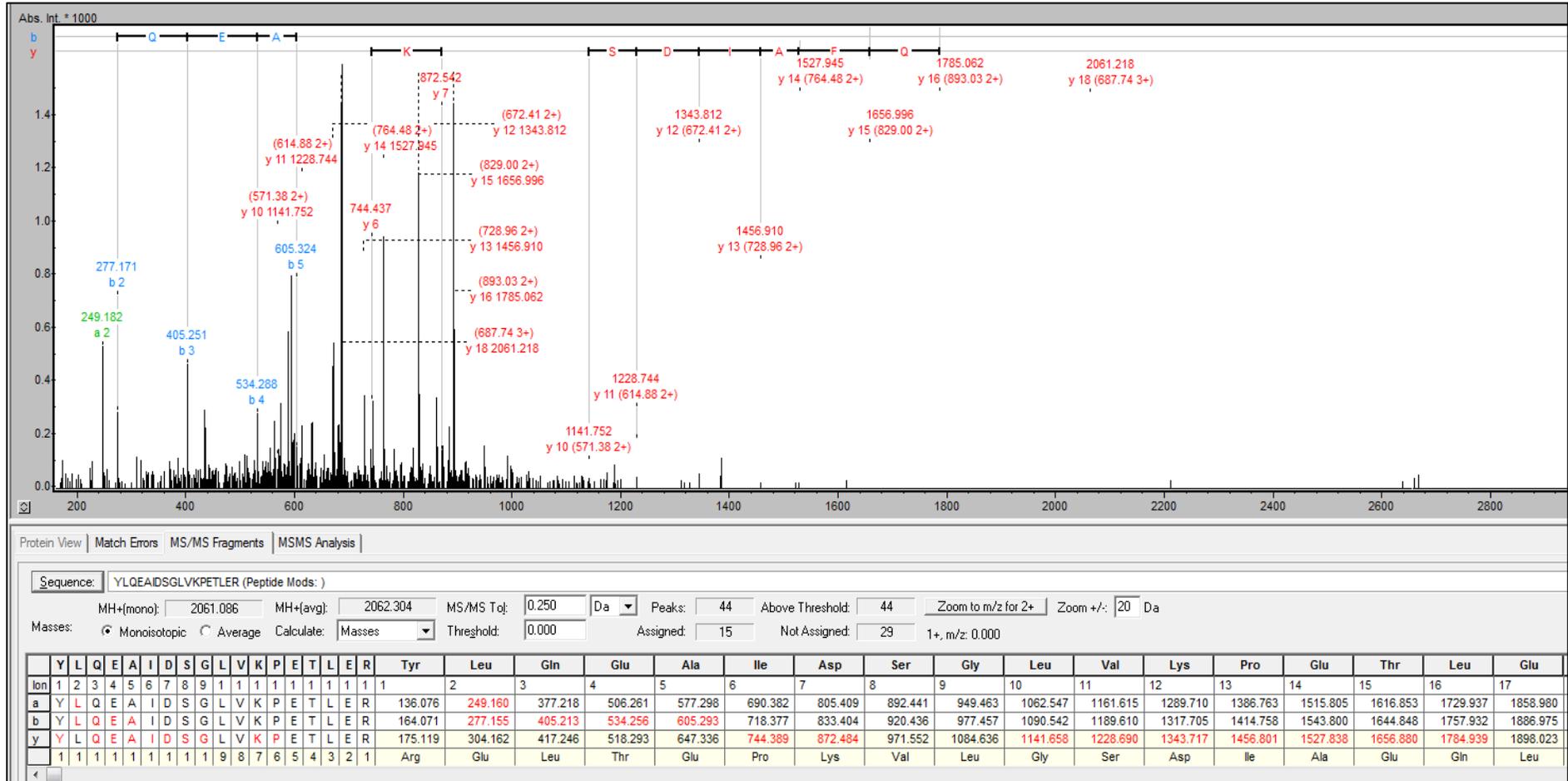


Figura 28- Espectro MS/MS do íon de massa $[M+H]^+2$ 712, 30

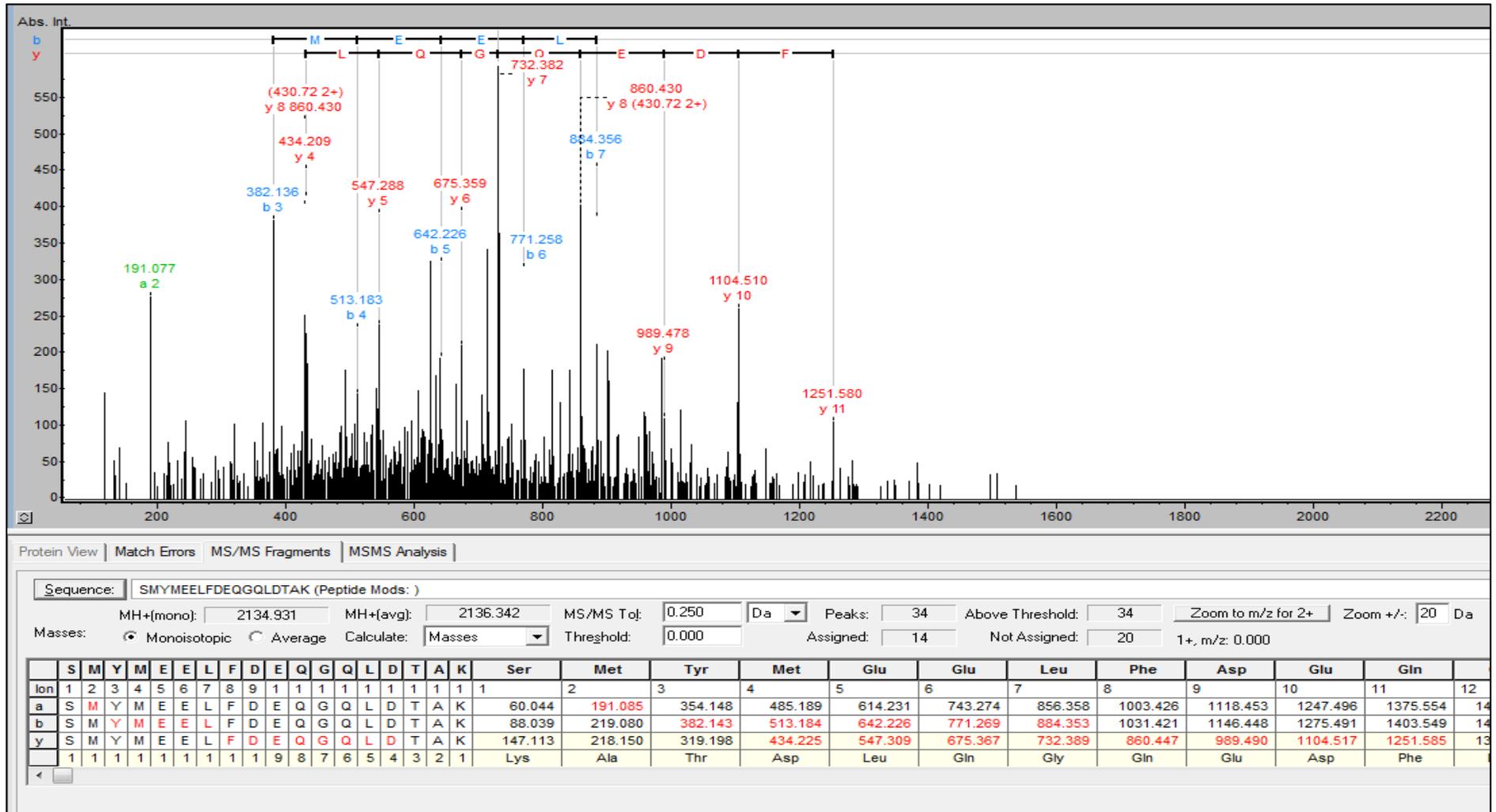


Figura 29- Espectro do íon de massa $[M+H]^{+2}$ 780,46

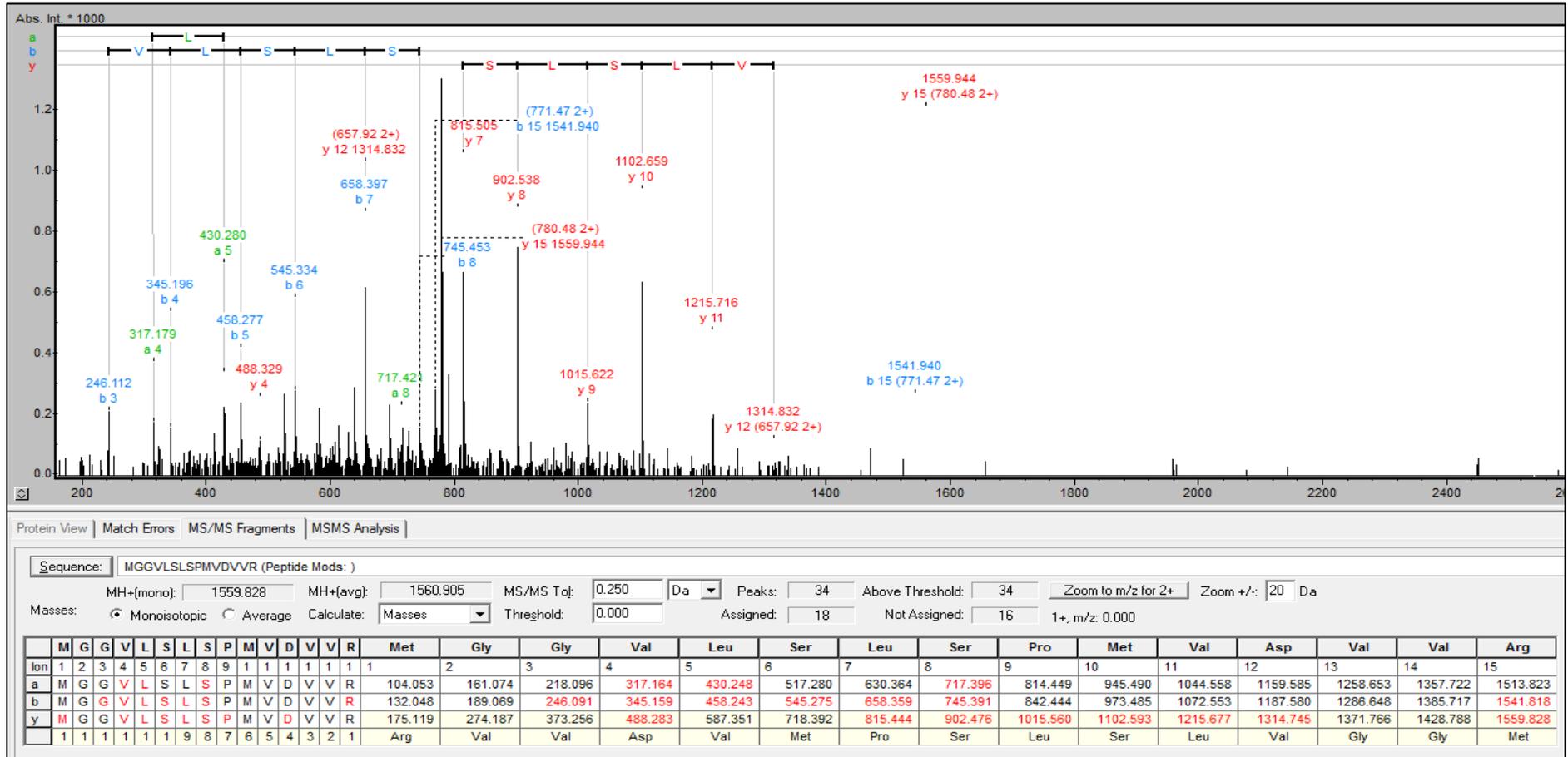


Figura 30- Alinhamento entre os fragmentos de peptídeos obtidos por MS/MS e a enzima BGL11, utilizada como referência

	10	20	30	40	50	60	70	80	90	100
BGL11										
Fragmentos	MRGFIMTCMTALLAGCQSVDPVNVEQQVDALYQK	MPQEER	IAQLKSMYMEELFDEQGQLDTAK	CRELIPYGIGHFSQYAFQKPRDPNVVDRDMVA	AAVQDWL					

			*****		*****					
	110	120	130	140	150	160	170	180	190	200
BGL11										
Fragmentos	MHHTPNGIPALFHEEVLSGINTQGATIYPQQIQACSFNPELAEKTRQTSTQMR	MGGVLSLSPMVDVVR	NPSFNRLEESYGEDAYLSAVMGTA	FVQGL						

	210	220	230	240	250	260	270	280	290	300
BGL11										
Fragmentos	QQGDLRKGVGACSKHYLGYGGGGDADEKELMEEILLPHEAMIRVAGSKAVMPGYHAMKSAPVTCVADSFILKDVLRDYLGF	DGMVVS	DYTAIDQLAPSNL							

	310	320	330	340	350	360	370	380	390	400
BGL11										
Fragmentos	PQKGEATVAPEVMKAAMAINGGNDVDFPFGANYKYLQE	AIDSLV	KPETLER	AVKNVLR	YKFRAGL	FDEKPY	LYSTEKIV	LDTPEER	QTAYGI	AVESVVL

	*****		*****							

7. DISCUSSÃO

Além das limitações relacionadas ao ensaio de atividade funcional, a busca enzimas em bibliotecas metagenômicas também tem como fator limitante a busca por novas ORFs, que por não possuírem similaridade com as sequências depositadas em banco de dados, não são reconhecidas e conseqüentemente, diminui a possibilidade de serem encontradas novas enzimas.

Com isso, o desenvolvimento de técnicas baseadas em proteômica para identificação de proteínas em larga escala, utilizando a espectrometria de massa para a caracterização de proteínas em amostras complexas, vem sendo cada vez mais utilizadas (AEBERSOLD; MANN, 2003).

O escopo deste trabalho é desenvolver uma técnica para a busca de sequências gênicas de possíveis ORFs em clones metagenômicos por meio de sequências de aminoácidos de peptídeos referentes à enzima de interesse, obtidos por LC-MS/MS.

A ideia desenvolvida neste projeto é inovadora: a utilização da espectrometria de massa do tipo *electrospray*, para a obtenção de dados MS/MS provenientes de extratos celulares tripsinizados de clones metagenômicos com o intuito de adquirir as sequências de aminoácidos que permitissem a busca pela ORF responsável pela expressão da enzima, tornando mais célere e pontual o processo de análise das sequências gênicas. Com isso, o que é apresentado é uma prova de conceito.

O trabalho foi realizado basicamente em quatro fases: escolha dos clones metagenômicos a serem testados; confirmação das atividades enzimáticas; análises dos extratos por LC-MS/MS e análise *in silico* tanto das sequências de ORFs quanto das sequências de aminoácidos obtidas.

Testes enzimáticos em meio líquido são comumente utilizados em clones previamente selecionados. A triagem inicial é realizada através de testes enzimáticos produzidos em meio sólido porque permite que mais de um clone seja testado por placa. Ensaio enzimático realizado em meio sólido é uma abordagem principalmente qualitativa, em que através reações químicas são observadas mudanças no meio sólido e a distinção entre clones que possuem a atividade biológica testada dos que a não possuem (SHUANGQI et al., 2011).

Dois ensaios em meio líquido, para a detecção de atividades enzimáticas foram testados nos clones selecionados. Como controle negativo, cepas com o respectivo vetor vazio, foram utilizadas como parâmetro de não atividade. O ensaio para detecção de atividade de endoglucanases foi realizado apenas nos clones de grandes insertos. Após o crescimento do pré-inóculo por aproximadamente 16 h e o crescimento do inóculo por aproximadamente 5 h, os extratos celulares foram lisados e utilizados nos ensaio como descrito no item 6.1. Com o intuito de observar se o tempo de armazenamento interferiria na atividade enzimática, o teste enzimático foi realizado com extratos logo após a lise celular, extratos que foram armazenados a -20 °C por aproximadamente 16 h (Figura 7) e extratos que foram armazenados a 4 °C por aproximadamente 16 h (Figura 8). Utilizando os extratos dos lisados celulares que não foram armazenados, todos os 15 clones tiveram atividade enzimática significativa (Figura 9).

O nível de açúcar redutor dos extratos celulares que não foram armazenados após a lise celular foi realizado (Figura 10). As análises quantitativas entre os ensaios com os extratos que foram armazenados a 4 °C e àqueles que foram utilizados logo após a lise celular, revelaram que o armazenamento influencia de forma negativa na performance da proteína durante o ensaio enzimático, sendo observada a diminuição do açúcar redutor liberado pelas enzimas quando o extrato celular é armazenado (Figura 11).

O outro ensaio enzimático, direcionado para a busca de β -glicosidases, foi utilizado o substrato pNPG. Este ensaio foi realizado com todos os clones, tanto aqueles provenientes da biblioteca de grandes insertos, quanto com o clone selecionado da biblioteca de pequenos inserto e o seu respectivo subclonado. O clone da biblioteca de pequenos insertos, denominado PCF430BGL11 e o seu subclonado, denominado PET21aBGL11, tiveram a sua atividade confirmada (Figura 12), o que era esperado já que era de conhecimento que a enzima é uma β -glicosidase.

Já nos clones selecionados na biblioteca de grandes insertos, apenas três possuíram atividade enzimática significativa neste teste. Então, como o intuito de diminuir o número amostral e de aumentar a probabilidade de se encontrar a enzima de interesse, desta biblioteca foram selecionados apenas os clones que foram positivos para ambos os testes enzimáticos, no caso os clones denominados

PL8G1, B17 e B18 (Figura 13). Dessa forma, com o resultado dos ensaios em meio líquido, tem-se a certeza de que a enzima está presente nos extratos celulares.

A busca por ORFs relacionadas às atividades testadas limitaram a quantidade de enzimas a serem utilizadas como referências. A enzima β -glicosidase encontrada no clone da biblioteca de pequenos insertos, possui um tamanho de 89,78 kDa (Figuras 14 e 15). Com o tamanho da molécula, foi possível a análise do perfil proteico, realizado por eletroforese em gel de poliacrilamida SDS-PAGE. O gel permitiu a visualização do perfil da amostra PCF430BGL11, que estava expressando em grandes quantidades, proteínas na faixa entre 75 e 100 kDa, faixa em que se encontra a enzima de interesse. Estranhamente, não foram visualizados sinais de que a proteína estava sendo superexpressa na amostra PET21aBgl11, que é um vetor próprio de superexpressão. Este dado corroborou com o ensaio enzimático, onde a atividade enzimática do clone PCF430BGL11 foi superior à atividade do clone PET21aBGL11 (Figura 16).

As ORFs obtidas nos clones oriundos da biblioteca de grandes insertos resultou em glicosil hidrolases de diferentes famílias (Figuras 17 a 24). Os tamanhos das quatro enzimas encontradas variaram de 50 a 85 kDa, aproximadamente. Em relação às amostras provenientes das células selecionadas da biblioteca de grandes insertos, não foram observadas modificações significativas relacionadas à alta expressão de proteínas nas faixas de tamanhos de interesse por eletroforese em gel de poliacrilamida SDS-PAGE (Figura 25), resultado já esperado, já que o vetor de clonagem em questão não possui um promotor desenhado para ser induzido, não permitindo a expressão em maior escala de uma enzima pré definida.

Os extratos celulares dos clones metagenômicos selecionados foram analisados por espectrometria de massa do tipo *electrospray*, acoplado a um cromatógrafo líquido para a obtenção de dados MS/MS. O número de espectros obtidos variou de acordo com o tempo de corrida cromatográfica. A diferença entre o número de espectros obtidos nas corridas com 65 e 95 min foi relativamente pequena, levando a conclusão de que o tempo mais adequado é o menor, com o intuito de otimizar a técnica (Tabela 1). Após a obtenção dos dados, apenas fragmentos obtidos do clone PCF430BGL11 tiveram similaridade com os fragmentos da enzima referência (Figuras 26 a 29).

As sequências de peptídeos obtidas por LC-MS/MS foram alinhadas à sequência referência, onde foi possível a localização dos fragmentos obtidos dentro da enzima (Figura 30).

Os dados obtidos durante o desenvolvimento deste trabalho demonstraram que é possível o uso de técnicas de LC-MS/MS para a busca por enzimas em bibliotecas metagenômicas, mas são necessárias algumas ressalvas.

Fragmentos foram obtidos em um clone que expressou uma grande quantidade da enzima de interesse, o que proporcionou a sua detecção. Nos clones onde a expressão da enzima alvo não foi alta, até mesmo no clone onde o gene estava inserido em um vetor de superexpressão, não foi possível a detecção de fragmentos que estivessem relacionados à enzima referência.

Outro parâmetro a ser questionado é o uso de *E. coli* como sistema hospedeiro. O sistema selecionado para a montagem das bibliotecas utilizadas não tem a capacidade de secretar de forma otimizada, fazendo com que a célula tenha que ser lisada para que a enzima de interesse seja liberada. Porém a lise da célula libera inúmeras outras proteínas existentes no citoplasma celular, tornando a amostra muito mais complexa, dificultando a detecção da enzima alvo. Akeroyd e colaboradores (2013) utilizaram método semelhante ao detectar enzimas secretadas por clones provenientes de bibliotecas de cDNA, por meio de espectrometria de massa do tipo MALDI. Neste trabalho, o sistema hospedeiro escolhido foi *B. subtilis*, que possui mecanismos para secretar a enzima de interesse.

O uso de sistemas alternativos para a montagem de bibliotecas metagenômicas, que secretem a enzima de forma otimizada pode ser uma alternativa para que se obtenha melhores resultados com a técnica.

Para que esta técnica seja utilizada, é necessário o sequenciamento do inserto de DNA, para que seja feita a busca por prováveis ORFs dentro da sequência.

8 CONCLUSÃO

O uso de técnicas baseadas em espectrometria de massa ainda não foram amplamente exploradas para detecção de enzimas em clones metagenômicos, mas se mostra promissora por permitir que ORFs sejam detectadas sem a necessidade de uma busca prévia em bancos de dados. O uso da ferramenta permitiu que enzimas expressas em clones metagenômicos fossem encontradas e a correlacionadas com enzima obtidas de sequências de ORFs, permitindo a ratificação desta ferramenta como um método de alto rendimento na busca por novas enzimas em bibliotecas metagenômicas.

REFERÊNCIAS

AEBERSOLD, R.; GOODLETT, D. R. Mass spectrometry in proteomics. **Chem Rev**, v. 101, n. 2, p. 269-95, Feb 2001.

AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. **Nature**, v. 422, n. 6928, p. 198-207, Mar 13 2003.

AKERROYD, M. et al. Searching for microbial protein over-expression in a complex matrix using automated high throughput MS-based proteomics tools. **J Biotechnol**, v. 164, n. 1, p. 112-20, Mar 10 2013.

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res**, v. 25, n. 17, p. 3389-402, Sep 1 1997.

ASSENBERG, R. et al. Advances in recombinant protein expression for use in pharmaceutical research. **Curr Opin Struct Biol**, v. 23, n. 3, p. 393-402, Jun 2013.

BERGMANN, J. C. **Utilização de Estratégias Metagenômicas para aplicações Biotecnológicas no Setor** 2013. (Doutorado). Programa de Pós Graduação Stricto Sensu em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília

BERGMANN, J. C. et al. Biodiesel production in Brazil and alternative biomass feedstocks. **Renewable and Sustainable Energy Reviews**, v. 21, n. 0, p. 411-420, 2013.

BON, E.; FERRARA, M.; CORVO, M. Enzimas em biotecnologia. **Rio de Janeiro, Editora Interciência LTDA**, 2008.

BRENT, M. R. Genome annotation past, present, and future: how to define an ORF at each locus. **Genome Res**, v. 15, n. 12, p. 1777-86, Dec 2005.

BROADWAY, N. Recombinant Protein Expression: Vector-Host Systems. **Materials and Methods**, 2013.

CAMMACK, R.; ATTWOOD, T. K. **Oxford dictionary of biochemistry and molecular biology**. 2nd. Oxford: Oxford University Press, 2006. xv, 720 p. ISBN 9780198529171

0198529171.

CHEN, J.; CANALES, L.; NEAL, R. E. Multi-Segment Direct Inject nano-ESI-LTQ-FT-ICR-MS/MS For Protein Identification. **Proteome Sci**, v. 9, p. 38, 2011.

CHICA, R. A.; DOUCET, N.; PELLETIER, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. **Curr Opin Biotechnol**, v. 16, n. 4, p. 378-84, Aug 2005.

COWAN, D. A.; FERNANDEZ-LAFUENTE, R. Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization. **Enzyme Microb Technol**, v. 49, n. 4, p. 326-46, Sep 10 2011.

CUNHA, I. **Caracterização e análise funcional da Comunidade bacteriana ruminal de caprinos da caatinga brasileira utilizando DNA metagenômico**. 2010. 153 Programa de Pós Graduação Stricto Sensu em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília

CUNHA, I. S. KRUEGER., R. H; QUIRINO, B. F. **Construção de uma biblioteca metagenômica de expressão da microbiota de rúmen de caprinos**. Brasília, DF: Emprapa Agroenergia, 2009. AGROENERGIA, E. 002: 6 p. 2009.

DANIEL, R. Soil-Based Metagenomics. In: (Ed.). **Handbook of Molecular Microbial Ecology II**: John Wiley & Sons, Inc., 2011. p.81-92. ISBN 9781118010549.

DE MARCO, A. Recent contributions in the field of the recombinant expression of disulfide bonded proteins in bacteria. **Microb Cell Fact**, v. 11, p. 129, 2012.

DEIVE, F. J. et al. North Western Spain hot springs are a source of lipolytic enzyme-producing thermophilic microorganisms. **Bioprocess Biosyst Eng**, v. 36, n. 2, p. 239-50, Feb 2013.

DÍAZ-RODRÍGUEZ, A.; DAVIS, B. G. Chemical modification in the creation of novel biocatalysts. **Current Opinion in Chemical Biology**, v. 15, n. 2, p. 211-219, 2011.

DRAUZ, K. **Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook**. Wiley-VCH, 2012. ISBN 9783527325474.

FAKRUDDIN, M. et al. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. **ISRN Biotechnology**, v. 2013, p. 7, 2013.

HESS, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. **Science**, v. 331, n. 6016, p. 463-7, Jan 28 2011.

HU, Y. et al. Chemical modification with functionalized ionic liquids: a novel method to improve the enzymatic properties of *Candida rugosa* lipase. **Bioprocess Biosyst Eng**, Feb 4 2014.

JIN, P. et al. Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. **Bioresour Technol**, v. 116, n. 0, p. 234-240, 2012.

JONES, A. et al. Directed evolution of a maltogenic alpha-amylase from *Bacillus* sp. TS-25. **J Biotechnol**, v. 134, n. 3-4, p. 325-33, Apr 30 2008.

JONES, P. et al. InterProScan 5: genome-scale protein function classification. **Bioinformatics**, v. 30, n. 9, p. 1236-40, May 1 2014.

KARIGAR, C. S.; RAO, S. S. Role of microbial enzymes in the bioremediation of pollutants: a review. **Enzyme Res**, v. 2011, p. 805187, 2011.

KIRK, O.; BORCHERT, T. V.; FUGLSANG, C. C. Industrial enzyme applications. **Curr Opin Biotechnol**, v. 13, n. 4, p. 345-51, Aug 2002.

KNIGHT, R. et al. Unlocking the potential of metagenomics through replicated experimental design. **Nat Biotechnol**, v. 30, n. 6, p. 513-20, Jun 2012.

KUPPERS, T. et al. Developing a new production host from a blueprint: *Bacillus pumilus* as an industrial enzyme producer. **Microb Cell Fact**, v. 13, n. 1, p. 46, 2014.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680-5, Aug 15 1970.

LEE, M. H.; LEE, S. W. Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities. **Genomics Inform**, v. 11, n. 3, p. 114-20, Sep 2013.

LIESE, A.; SEELBACH, K.; WANDREY, C. **Industrial Biotransformations**. Wiley, 2008. ISBN 9783527614172.

LORENZ, P.; ECK, J. Metagenomics and industrial applications. **Nat Rev Microbiol**, v. 3, n. 6, p. 510-6, Jun 2005.

MA, B.-D. et al. A thermostable and organic-solvent tolerant esterase from *Pseudomonas putida* ECU1011: Catalytic properties and performance in kinetic resolution of α -hydroxy acids. **Bioresour Technol**, v. 133, n. 0, p. 354-360, 2013.

MARKOWITZ, V. M. et al. IMG/M: a data management and analysis system for metagenomes. **Nucleic Acids Res**, v. 36, n. Database issue, p. D534-8, Jan 2008.

MCMAHON, M. D. et al. Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. **Appl Environ Microbiol**, v. 78, n. 10, p. 3622-9, May 2012.

MEWIS, K.; TAUPP, M.; HALLAM, S. J. A high throughput screen for biomining cellulase activity from metagenomic libraries. **J Vis Exp**, n. 48, 2011.

MILLER, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. **Analytical Chemistry**, v. 31, n. 3, p. 426-428, 1959/03/01 1959.

MOE, L. A.; MCMAHON, M. D.; THOMAS, M. G. Functional Metagenomics as a Technique for the Discovery of Novel Enzymes and Natural Products. In: (Ed.). **Enzyme Technologies**: John Wiley & Sons, Inc., 2010. p.1-39. ISBN 9780470627303.

MURATA, M. et al. Identification of milk proteins enhancing the antimicrobial activity of lactoferrin and lactoferricin. **J Dairy Sci**, v. 96, n. 8, p. 4891-8, Aug 2013.

NAVARRO, C. A.; VON BERNATH, D.; JEREZ, C. A. Heavy metal resistance strategies of acidophilic bacteria and their acquisition: importance for biomining and bioremediation. **Biol Res**, v. 46, n. 4, p. 363-71, 2013.

NEWMAN, J. R.; FUQUA, C. Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. **Gene**, v. 227, n. 2, p. 197-203, Feb 18 1999.

REYMOND, J.-L.; FLUXA, V. S.; MAILLARD, N. Enzyme assays. **Chemical Communications**, n. 1, p. 46, 2008.

SCHLEGEL, S. et al. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. **Microb Cell Fact**, v. 12, p. 24, 2013.

SHUANGQI, T. et al. Determination methods of cellulase activity. **African Journal of Biotechnology**, v. 10, n. 37, p. 7122-7125, 2011.

SIERRA-GARCIA, I. N. et al. New hydrocarbon degradation pathways in the microbial metagenome from brazilian pETroleum reservoirs. **PLoS One**, v. 9, n. 2, p. e90087, 2014.

SIMON, C.; DANIEL, R. Achievements and new knowledge unraveled by metagenomic approaches. **Appl Microbiol Biotechnol**, v. 85, n. 2, p. 265-76, Nov 2009.

TARTOFF, K. D., HOBBS, C. A. . Improved media for growing plasmid and cosmid clones. **Bethesda Research Laboratories Focus**, v. 9, n. 12, 1987.

THIEDE, B. et al. Peptide mass fingerprinting. **Methods**, v. 35, n. 3, p. 237-47, Mar 2005.

UCHIYAMA, T.; MIYAZAKI, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. **Curr Opin Biotechnol**, v. 20, n. 6, p. 616-22, Dec 2009.

VAVROVA, L.; MUCHOVA, K.; BARAK, I. Comparison of different Bacillus subtilis expression systems. **Res Microbiol**, v. 161, n. 9, p. 791-7, Nov 2010.

VENTER, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. **Science**, v. 304, n. 5667, p. 66-74, Apr 2 2004.

VICK, J. E. et al. Optimized compatible set of BioBrick vectors for metabolic pathway engineering. **Appl Microbiol Biotechnol**, v. 92, n. 6, p. 1275-86, Dec 2011.

WAGHMARE, P. R. et al. Enzymatic hydrolysis and characterization of waste lignocellulosic biomass produced after dye bioremediation under solid state fermentation. **Bioresour Technol**, Mar 4 2014.

WANG, Z. et al. Polymersome Colloidosomes for Enzyme Catalysis in a Biphasic System. **Angewandte Chemie International Edition**, v. 51, n. 43, p. 10746-10750, 2012.

WECKWERTH, W. Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. **J Proteomics**, v. 75, n. 1, p. 284-305, Dec 10 2011.

WHEELER, D. L. et al. Database resources of the National Center for Biotechnology. **Nucleic Acids Res**, v. 31, n. 1, p. 28-33, Jan 1 2003.

XU, P.; GUNSOLLEY, J. Application of metagenomics in understanding oral health and disease. **Virulence**, v. 5, n. 3, Mar 18 2014.

YANG, Y.; WONG, S. E.; LIGHTSTONE, F. C. Understanding a Substrate's Product Regioselectivity in a Family of Enzymes: A Case Study of Acetaminophen Binding in Cytochrome P450s. **PLoS One**, v. 9, n. 2, p. e87058, 2014.

YUN, J. et al. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. **Appl Environ Microbiol**, v. 70, n. 12, p. 7229-35, Dec 2004.

ZALUCKI, Y. M.; BEACHAM, I. R.; JENNINGS, M. P. Biased codon usage in signal peptides: a role in protein export. **Trends Microbiol**, v. 17, n. 4, p. 146-50, Apr 2009.

ZHANG, L. et al. Isolation and Characterization of a Novel Endoglucanase from a *Bursaphelenchus xylophilus* Metagenomic Library. **PLoS One**, v. 8, n. 12, p. e82437, 2013.

ZHANG, Y. H.; HONG, J.; YE, X. Cellulase Assays #. In: (Ed.). **T Biofuels**, v.581, 2009. p.213-231. (Methods in Molecular Biology).