

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIA E TECNOLOGIA DE ALIMENTOS
DEPARTAMENTO DE TECNOLOGIA DE ALIMENTOS

TRABALHO DE CONCLUSÃO EM BIOTECNOLOGIA
CLONAGEM MOLECULAR, EXPRESSÃO E CARACTERIZAÇÃO DA
TRANSGLUTAMINASE DE *Paenibacillus* sp. JDR-2 EM *Escherichia coli*

MATHEUS QUINTANA BARRETO

Porto Alegre, 2015

MATHEUS QUINTANA BARRETO

CLONAGEM MOLECULAR, EXPRESSÃO E CARACTERIZAÇÃO DA
TRANSGLUTAMINASE DE *Paenibacillus* sp. JDR-2 EM *Escherichia coli*

Trabalho apresentado como um dos requisitos para a
obtenção do grau de bacharel em Biotecnologia,
ênfase em Biotecnologia Molecular.

Orientador: Marco Antônio Záchia Ayub

Comissão Examinadora:

Prof. Dr. Henrique Bunselmeyer Ferreira

Departamento de Biologia Molecular e
Biotecnologia - UFRGS

Prof. Dr. Richard John Ward

Departamento de Química- FFCLRP - USP

Prof. Dr. Marco Antônio Záchia Ayub

Departamento de Tecnologia de Alimentos– UFRGS

Porto Alegre, 2015

Este trabalho foi realizado no Laboratório de Biotecnologia do Instituto de Ciência e Tecnologia de Alimentos da Universidade Federal do Rio Grande do Sul e no Laboratório de Bioquímica e Biofísica de Proteínas do Departamento de Química da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo. O trabalho contou com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico

(CNPq).

AGRADECIMENTOS

A meus pais pelo incentivo e apoio constantes durante toda minha formação.

Ao professor Marco Antônio Záchia Ayub por aceitar orientar este trabalho e prover todo o necessário para que ele fosse realizado, auxiliando sempre que necessário.

Ao professor Richard John Ward do departamento de química da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo por ter aberto as portas do seu laboratório para que eu realizasse parte deste trabalho.

Ao professor Plinho Francisco Herz do ICTA-UFRGS pela permissão a usar seu laboratório e por sempre se disponibilizar a discutir as dúvidas que surgiam.

Ao professor Arnaldo Zaha e Henrique Bunselmeyer Ferreira do Centro de Biotecnologia da UFRGS por permitir que eu realizasse alguns experimentos de SDS-PAGE em seus laboratórios. A todos os integrantes dos laboratórios 204, 206 e 210 por ajudarem quando precisei usar algum equipamento. Um agradecimento especial à Edi, do laboratório 206, por sempre ter se disponibilizado a ajudar.

Ao professor Eduardo Brandt de Oliveira da FMRP-USP por ceder a glutatona reduzida e ao professor Herenilton Paulino de Oliveira do departamento de química da FFCLRP-USP pela hidroxilamina para realização dos testes de atividade.

Ao professor Arthur Henrique Cavalcante de Oliveira do departamento de química da FFCLRP-USP pela ajuda com os experimentos de espectrometria de massas.

À Micheli Bueno, mestranda do PPGCTA da UFRGS, por me ensinar a fazer o teste de atividade enzimática e ceder a transglutaminase comercial para que eu pudesse realizar os testes de atividade.

Aos amigos do laboratório 212 e 216 do ICTA pelo convívio e ajuda sempre que possível. Às gurias do laboratório 210 do ICTA pela convivência e por ajudar sempre que necessário. Um agradecimento especial à Natália Guilherme Graebin e à Jéssie da Natividade Schöffner por terem preparado e enviado material a Ribeirão Preto quando precisei.

Um grande agradecimento ao pessoal do Laboratório de Bioquímica e Biofísica de Proteínas da FFCLRP da USP pelos quatro meses de convivência e pela recepção no laboratório. À

Lara A. B. C. Carneiro e à Luana de Fátima Alves por toda ajuda durante este tempo, pela ajuda nos experimentos, pelas discussões de resultados e disponibilidade em ajudar. Ao Gustavo Molina, pela convivência. Aos técnicos do laboratório, o André, e ao Leonardo pelo preparo dos materiais sempre que preciso. Obrigado a todos!

Ao CNPq pela bolsa provida e auxílio do projeto.

À *Escherichia coli*, por ter expressado a proteína recombinante.

Aos amigos, que permitiram que esta caminhada fosse mais leve e alegre, oferecendo seu apoio e amizade.

Aos demais, que de alguma forma contribuíram com este trabalho, deixo minha gratidão.

“We ourselves feel that what we are doing is just a drop in the ocean. But the ocean would be less because of that missing drop.”

Mother Theresa

“Às vezes sentimos que aquilo que fazemos é apenas uma gota no oceano. Mas o oceano seria menor se lhe faltasse essa gota. ”

Madre Teresa de Calcutá

RESUMO

Transglutaminases são proteínas que catalisam a formação de uma ligação cruzada entre uma amina primária e o grupamento carboxamida presente em resíduos de glutamina, possuindo um importante papel na re-estruturação de alimentos proteínicos na indústria de alimentos. *Paenibacillus* sp. JDR-2 é uma bactéria Gram-positiva relacionada filogeneticamente com outros microrganismos produtores de transglutaminases, como aqueles do gênero *Bacillus*, sendo, portanto, um possível produtor desta enzima. Este trabalho visou a identificar genes codificadores de transglutaminases presentes no genoma de *Paenibacillus* sp. JDR-2 *in silico* e expressar essas sequências em um sistema heterólogo para sua caracterização. Uma busca por genes codificadores de transglutaminases foi feita através da procura por ORFs que foram anotadas como contendo um domínio de transglutaminase e depositadas no NCBI. Primers específicos foram projetados para amplificar três destas ORFs, que foram então amplificadas por PCR e clonadas no vetor de expressão pET23d+. Os plasmídeos recombinantes gerados foram transformados em três linhagens hospedeiras de *Escherichia coli* para a síntese das proteínas recombinantes, cujos extratos foram analisados por SDS-PAGE e a atividade enzimática verificada pelo teste do hidroxamato. A análise *in silico* mostrou a presença de dez transglutaminases codificadas pelo genoma de *Paenibacillus* sp. JDR-2, sendo nove sequências identificadas como possíveis transglutaminases e uma destas sendo uma sequência truncada que foi, portanto, descartada de futuras análises. As sequências nucleotídicas das ORFs 4765, 4870 e 5011 foram clonadas em vetor pET23d+ e os plasmídeos recombinantes gerados foram transformados em *E. coli*, mas apenas com o vetor contendo a sequência da ORF 4765 transformado em *E. coli* Arctic Express foi observada a expressão da proteína recombinante, apesar de esta estar totalmente na fração insolúvel. Tentativas de resolubilizar a proteína 4765 utilizando condições não-desnaturantes foram eficientes, mas a proteína não mostrou qualquer atividade enzimática de transglutaminase. As próximas etapas deste trabalho envolvem a tentativa de solubilizar a proteína 4765 com um método de desnaturação protéica seguido pelo seu *refolding* para verificar sua atividade, e clonar as sequências 4870 e 5011 em novos vetores para expressão em *E. coli* e também em vetores para expressão em *Pichia pastoris*.

Palavras-chave: transglutaminase microbiana, *Paenibacillus* sp. JDR-2

ABSTRACT

Transglutaminases are proteins that catalyze the formation of a crosslink between a primary amine and the γ -carboxamide group of protein-bound glutamine, having an important role in the food industry in the production of re-structured food. *Paenibacillus* sp. JDR-2 is a Gram-positive microorganism closely related to known transglutaminase-producing organisms, such as *Bacillus* sp.. This study aimed to identify transglutaminase-coding DNA sequences in the genome of *Paenibacillus* sp. JDR-2 *in silico* and to express these sequences in a heterologous system in order to characterize the recombinant enzymes. We searched for predicted genes encoding transglutaminase-domains from *Paenibacillus* sp. JDR-2 that were identified by automated annotation and deposited in the NCBI database. Specific primers were designed to amplify three predicted transglutaminase-coding ORFs from the genomic DNA of *Paenibacillus* sp. JDR-2 by PCR and cloning into the vector pET23d+. Recombinant plasmids were transformed into expression strains of *Escherichia coli* for the synthesis of the recombinant proteins and analyzed by SDS-PAGE, followed by the hydroxamate assay to test for transglutaminase activity. *In silico* analysis indicated the presence of ten putative transglutaminases in *Paenibacillus* sp. JDR-2, with one being a truncated protein, which was therefore not pursued. The predicted DNA sequences of the ORFs 4765, 4870 and 5011 were cloned into the expression vector pET23d+ and transformed into three expression strains of *E. coli*, however only the protein encoded by ORF 4765 was expressed. The protein was only expressed under low temperature conditions in *E. coli* Arctic Express, but it was completely insoluble. Attempts to solubilize this protein using non-denaturing conditions yielded soluble protein with no catalytic activity. The next steps are to clone the ORFs 4870 and 5011 into other expression vectors for expression in *E. coli* and *Pichia pastoris*, and to solubilize the inclusion bodies containing the protein encoded by ORF 4765 using denaturing conditions followed by refolding to obtain a soluble enzyme for tests for transglutaminase activity.

Key words: microbial transglutaminase, *Paenibacillus* sp. JDR-2

LISTA DE ABREVIACOES

mTG	Transglutaminase microbiana (<i>microbial transglutaminase</i>)
aTG	Transglutaminase animal (<i>animal transglutaminase</i>)
ORF	Fase aberta de leitura (<i>open reading frame</i>)
DTT	Ditiotreitol
PMSF	fluoreto de fenilmetanosulfonil (<i>phenylmethanesulfonyl fluoride</i>)
HEPES	Ácido Hidroxi-Etil-Piperazina-Etano-Sulfônico
TEMED	N, N, N',N'-Tetrametiletilenodiamina
pI	Ponto isoelétrico
Mw	Massa molecular (<i>molecular weight</i>)
PCR	Reaço em cadeia da polimerase (<i>polymerase chain reaction</i>)
PBS	Soluço Salina Fosfatada (<i>Phosphate Buffered Saline</i>)
SDS-PAGE	Eletroforese em gel de poliacrilamida com dodecilsulfato de sdio
Z-gln-gly	Carbobenzoxi-glutaminil-glicinil
pH	Potencial hidrogeninico

LISTA DE FIGURAS

Figura 1. Fator XIII de coagulação.....	16
Figura 2. Estrutura típica de mTGs e reação catalisada por esta enzima	17
Figura 3. Vetor pET23d+	26
Figura 4. Mapa dos plasmídeos recombinantes construídos	26
Figura 5. Teste de clonagem da sequência 4765	34
Figura 6. Teste de clonagem da sequência 4870	35
Figura 7. Teste de clonagem da sequência 5011	35
Figura 8. Expressão dos vetores recombinantes em <i>Escherichia coli</i> pLysE	37
Figura 9. Expressão dos plasmídeos recombinantes em <i>Escherichia coli</i> Rosetta	37
Figura 10. Teste de expressão dos plasmídeos recombinantes em <i>Escherichia coli</i> Arctic Express	39
Figura 11. Identificação da proteína recombinante por MS/MS	40
Figura 12. Expressão e Purificação da proteína 4765	41
Figura 13. Síntese da proteína 4765 e sua solubilização a partir de corpos de inclusão	42
Figura 14. Relação de publicações sobre transglutaminases de 1959 a 2016	43

LISTA DE TABELAS

Tabela 1: <i>Primers</i> projetados para a amplificação das sequências 4765, 4870 e 5011 de <i>Paenibacillus</i> sp. JDR-2	24
Tabela 2: ORFs de <i>Paenibacillus</i> sp. JDR-2 anotadas como transglutaminase. Busca realizada em 10 de janeiro de 2015.....	32
Tabela 3: Ponto isoelétrico (pI) e peso molecular estimado da sequência peptídica codificada por cada uma das ORFs identificadas como transglutaminases	33
Tabela 4. Dados sobre os peptídeos obtidos através de espectrometria de massas	39

Sumário

AGRADECIMENTOS	4
RESUMO	7
1. INTRODUÇÃO.....	14
1.1 Enzimas na indústria de alimentos.....	14
1.2 Transglutaminases	14
1.3 Produção de Transglutaminases na Indústria.....	17
1.4 <i>Paenibacillus</i> sp. JDR-2	18
2. JUSTIFICATIVA	20
3. OBJETIVO	21
3.1 Objetivo Geral.....	21
3.2 Objetivos Específicos	21
4. MATERIAIS E MÉTODOS.....	22
4.1 Materiais utilizados.....	22
4.2 Busca por sequências codificadoras de transglutaminase em <i>Paenibacillus</i> sp. JDR-2.....	23
4.3 Extração de DNA genômico	23
4.4 Projeto de iniciadores e amplificação das sequências de interesse.....	23
4.5 Clivagem de sequências amplificadas por PCR e de vetores com enzimas de restrição.....	25
4.6 Clonagem das sequências codificadoras das proteínas putativas 4765, 4870 e 5011 de <i>Paenibacillus</i> sp. JDR-2 em vetor de expressão pET23d+.....	25
4.7 Preparação de células competentes	26
4.8 Transformação de diferentes cepas de <i>Escherichia coli</i>	27
4.9 Seleção, Purificação e Análise de restrição de plasmídeos recombinantes	27
4.10 Expressão heteróloga das proteínas recombinantes de <i>Paenibacillus</i> sp. JDR-2 (4765, 4870 e 5011) em <i>E. coli</i>	28
4.11 Eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE)	28
4.12 Espectrometria de massas	29
4.13 Purificação de proteínas por cromatografia de afinidade	29
4.13 Solubilização dos grânulos de inclusão	30
4.14 Teste de atividade enzimática	30
5 RESULTADOS	32
5.1 Identificação de possíveis sequências codificadoras de transglutaminase utilizando ferramentas de bioinformática	32

5.2 Amplificação e clonagem das sequências 4765, 4870 e 5011 de <i>Paenibacillus</i> sp. JDR-2 em <i>Escherichia coli</i>	33
5.3 Análise da expressão das construções pETG4765, pETG4870 e pETG5011	36
5.3.1 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em <i>E. coli</i> BL21 (DE3) pLysE	36
5.3.2 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em <i>Escherichia coli</i> BL21 (DE3) Rosetta.....	37
5.3.3 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em Arctic Express	38
5.4 Espectrometria de massas	39
5.5 Purificação da proteína recombinante 4765.....	4040
6. DISCUSSÃO.....	43
7. CONCLUSÃO.....	48
8. PERSPECTIVAS	49
9. REFERÊNCIAS	50
APÊNDICE A - Identificação <i>in silico</i> da sequência 0339 como uma transglutaminase	55
APÊNDICE B - Identificação <i>in silico</i> da sequência 0834 como uma transglutaminase	56
APÊNDICE C - Identificação <i>in silico</i> da sequência 0937 como uma transglutaminase:	58
APÊNDICE D - Identificação <i>in silico</i> da sequência 1235 como uma transglutaminase	59
APÊNDICE E - Identificação <i>in silico</i> da sequência 1489 como uma transglutaminase.....	60
APÊNDICE F - Identificação <i>in silico</i> da sequência 3259 como uma transglutaminase.....	62
APÊNDICE G - Identificação <i>in silico</i> da sequência 4024 como uma transglutaminase	63
APÊNDICE J - Identificação <i>in silico</i> da sequência 5011 como uma transglutaminase	67
APÊNDICE K - Predição da presença de peptídeo sinal	68
APÊNDICE L - Teste de amplificação das sequências 4765, 4870 e 5011	69
APÊNDICE M – Comparação entre as proteínas de <i>Paenibacillus</i> sp. JDR-2 clonadas e parentais.....	70
APÊNDICE N – Resultado do sequenciamento dos plasmídeos construídos.....	72

1. INTRODUÇÃO

1.1 Enzimas na indústria de alimentos

A produção biotecnológica de alimentos e, empiricamente, o uso de enzimas pelo homem, teve seu início nos primórdios da história das civilizações, quando os primeiros alimentos produzidos através de processos fermentativos foram relatados. Sabe-se que os sumérios e os babilônios produziam cerveja há 8.000 anos, referências à produção de vinho podem ser encontradas no livro bíblico do Gênesis e os egípcios antigos usavam leveduras para a produção de pão (VASIC-RACKI, 2006).

A produção industrial de enzimas para uso no processamento de alimentos iniciou em 1874 com a extração de quimosina, utilizada na produção de queijos, do estômago de cabras pelo cientista dinamarquês Christian Hansen. No século XX, entre os principais avanços neste campo estão o uso de pectinases para a clarificação de sucos nos anos 30 e o uso de invertase durante a segunda guerra mundial para a produção de açúcar invertido, processo pioneiro no uso de enzimas imobilizadas na indústria. A aplicação de enzimas em larga-escala, no entanto, só se deu a partir de 1960, quando o tratamento ácido do amido foi substituído pelo tratamento enzimático com amilases e amiloglicosidases (FERNANDES, 2010).

Com a descoberta da estrutura do DNA, em 1953, e os avanços da biologia molecular nos anos subsequentes, a produção de enzimas por microrganismos recombinantes se tornou uma alternativa para a produção de enzimas em larga-escala (HEADON; WALSH, 1994). Entre as vantagens oferecidas por essa estratégia estão a maior produção da enzima pelo microrganismo recombinante do que pelo organismo no qual a enzima é naturalmente produzida e a possibilidade de melhorar as características físico-químicas da enzima de interesse através de diferentes estratégias, como, por exemplo, a evolução dirigida (ZHAO; CHOCKALINGAM; CHEN, 2002). A primeira enzima recombinante produzida para uso na indústria de alimentos foi a enzima amilase maltogênica, produzida pela Novozymes, em 1984 (OUTTRUP; NORMAN, 1984), sendo que, atualmente, estima-se que 50% das enzimas comercializadas provenham de organismos geneticamente modificados (OORT, 2010).

1.2 Transglutaminases

Transglutaminase, ou proteína-glutamina γ -glutamil transferase (EC 2.3.2.13), é uma proteína pertencente a uma família de enzimas que catalisam reações de acil-transferência

(figura 1a), deamidações e ligações cruzadas intra- e intermoleculares entre proteínas e peptídeos (FOLK; FINLAYSON, 1977; GRIFFIN; CASADIO; BERGAMINI, 2002; MARGOSIAK et al., 1990). Essa enzima já foi encontrada em uma vasta quantidade de organismos, incluindo bactérias, plantas e animais, (CAMPOS et al., 2013; LE et al., 2013; OTENG-PABI; KEILLOR, 2013) sendo que a transglutaminase eucariótica e a transglutaminase bacteriana (mTG) possuem diferenças estruturais e funcionais significantes.

Transglutaminases foram descobertas em 1957 através da verificação da atividade de incorporação de aminas de baixo peso molecular em proteínas no fígado de porquinho-da-índia (MYCEK et al., 1959). Desde então, esta enzima tem sido encontrada em uma grande quantidade de organismos, exercendo as mais variadas funções, demonstrando, assim, seu caráter pleiotrópico (LORAND; GRAHAM, 2003). Entre as principais transglutaminases encontradas em animais pode-se citar o fator XIII de coagulação sanguínea (LORAND, 2001), e a proteína transglutaminase 2, ubíqua em humanos e cujas funções incluem o acometimento da célula à apoptose (CHEN; MEHTA, 1999) e a inibição da adipogênese (MYNENI; MELINO; KAARTINEN, 2015).

Transglutaminases animais (aTG) são dependentes do íon Ca^{++} , possuem uma massa molecular alta, entre 65 e 360 kDa (CHEN; MEHTA, 1999) e são encontradas como monômeros ou multímeros, como o fator XIII de coagulação, que existe como um dímero ou tetrâmero (MUSZBEK et al., 2011). O fator XIII de coagulação possui estrutura conhecida, compreendendo quatro domínios: um domínio beta-sanduíche na extremidade amino, um domínio catalítico e dois domínios em forma de barril, barris 1 e 2, na extremidade carbóxi da proteína (figura 1) (YEE et al., 1994).

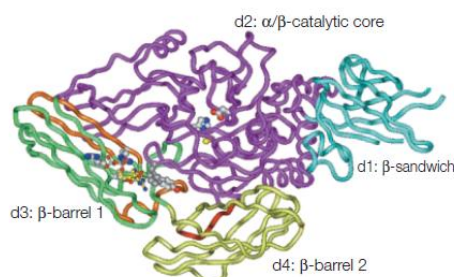


Figura 1. Fator XIII de coagulação: figura mostrando uma transglutaminase eucariótica com estrutura de seus domínios conhecida. Em azul está localizado o domínio β -sanduíche; em roxo está o centro catalítico composto por alfa-hélices e folhas-beta; em verde e amarelo estão os barris-Beta 1 e 2, respectivamente (Fonte: Lorand; Graham, 2003).

A transglutaminase microbiana, por sua vez, foi primeiramente descrita em *Streptoverticillium mobaraense* (ANDO et al., 1989), sendo posteriormente identificada em outras espécies do mesmo gênero e também em espécies do gênero *Bacillus* e *Streptomyces* (NAGY; SZAKACS, 2008; SUZUKI et al., 2000; ZHU et al., 1995). Diferentemente da transglutaminase encontrada em eucariotos, a reação catalisada pela transglutaminase microbiana não depende do íon Ca^{++} (figura 2a), é exclusivamente monomérica e sua tríade catalítica contém resíduos de cisteína, histidina e asparagina ou aspartato (KASHIWAGI et al., 2002). A principal função desta enzima em microrganismos é a de auxiliar na formação do endosporo (KOBAYASHI et al., 1996), embora outras funções, como a de atuar como um fator de virulência (YU et al., 2015), venham sendo descobertas.

A mTG de *S. mobaraense* foi a primeira a ter sua estrutura tridimensional conhecida, sendo composta por oito folhas-beta circundadas por onze alfa-hélices (figura 2b), estando o sítio ativo localizado ao fundo de um sulco localizado a 16 Å da superfície (KASHIWAGI et al., 2002). Esta enzima é expressa como uma pró-enzima cujos 45 primeiros aminoácidos da extremidade amino-terminal formam uma alfa-hélice que bloqueia o sítio catalítico impedindo o acesso do substrato (figura 2c) (YANG et al., 2011a), atuando também como uma chaperona que auxilia no enovelamento correto da proteína (YURIMOTO et al., 2004).

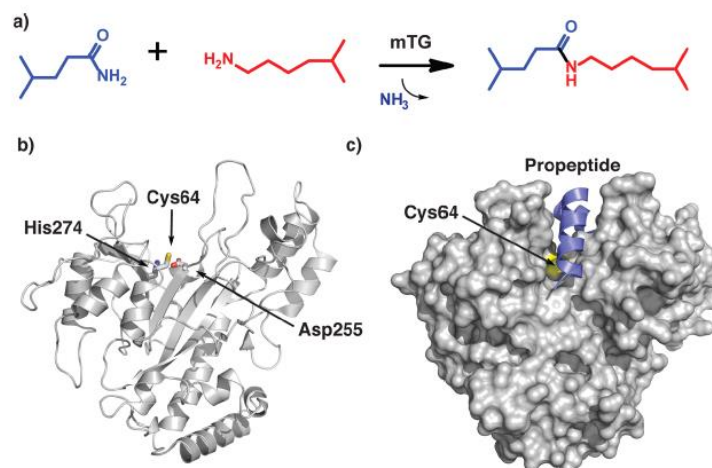


Figura 2. Estrutura típica de mTGs e reação catalisada por esta enzima: em (A) podemos observar a formação de uma ligação ϵ - γ -acil entre as cadeias laterais de resíduos de lisina e glutamina, respectivamente, catalisados por uma mTG. A estrutura de uma pro-transglutaminase (B) e da transglutaminase madura com seu sítio catalítico (A) também podem ser verificadas (Fonte: Strop, 2014).

1.3 Produção de transglutaminases na Indústria

A descoberta de transglutaminases chamou a atenção das mais variadas indústrias pela sua atividade de catalisar modificações pós-traducionais em proteínas. Até meados da década de 80, a transglutaminase extraída do fígado de porquinho-da-índia remanesceu como a única fonte comercial de transglutaminase, sendo seu custo de aproximadamente US\$80/U devido à dificuldade em obter fígados deste animal para extrair a enzima e ao difícil processo de purificação do material (ZHU; TRAMPER, 2008). A dependência de cálcio desta enzima também levava a outros problemas, como a precipitação de proteínas, o que podia acarretar na perda das características desejáveis do produto (MOTOKI; SEGURO, 1998).

Com o isolamento e caracterização da transglutaminase microbiana de *S. mobaraense*, esta enzima se tornou uma alternativa para a indústria frente à transglutaminase animal devido à sua menor massa molecular, baixo custo de produção e atividade em um amplo espectro de temperaturas e pHs (YOKOYAMA; NIO; KIKUCHI, 2004). Testes de toxicidade mostraram que esta enzima é segura e não possui potencial carcinogênico ou toxicológico, sendo, portanto, capaz de ser utilizada na indústria de alimentos (K. BERNARD, S. TSUBUKU, S. SHIOYA, 1998).

A transglutaminase de origem microbiana é principalmente explorada na indústria alimentícia, atuando na alteração da textura da carne, na melhora da consistência de produtos

derivados de leite e da resiliência de massas a base de farinha (KIELISZEK; MISIEWICZ, 2014). Apesar de poder melhorar o valor nutricional dos alimentos, adicionando aminoácidos necessários para a dieta (IKURA et al., 1985), o principal papel exercido pela mTG nesta indústria está na polimerização de proteínas, causando grandes modificações em sua estrutura e aumentando, conseqüentemente, seu valor comercial (NIELSEN, 1995; YOKOYAMA; NIO; KIKUCHI, 2004).

O potencial das transglutaminases fez com que ela venha sendo estudada para os mais variados propósitos nos últimos anos. Na indústria farmacêutica, a mTG está sendo utilizada para ligar anticorpos a drogas a fim de auxiliar no tratamento de câncer (ANDO et al., 2014; STROP, 2014). A ligação de anticorpos a outras moléculas, como a biotina, vem sendo pesquisada para uso em laboratório (JOSTEN et al., 2000), e a fusão de marcadores moleculares como o peptídeo c-myc facilita a incorporação de moléculas via transglutaminase devido à presença de uma glutamina reativa em sua estrutura (DENNLER et al., 2015). Recentemente, pesquisas têm demonstrado também o potencial uso de mTG na indústria de biopolímeros através da produção de bioplásticos, tanto biodegradáveis quanto comestíveis, utilizando esta enzima (PORTA et al., 2011).

A transglutaminase microbiana é considerada um aditivo eficiente, porém caro, para a indústria. Mesmo após mais de 20 anos de sua descoberta, a transglutaminase de *S. mobaraensis* ainda é a principal fonte desta enzima para a indústria (ARRIZUBIETA, 2007), sendo produzida unicamente pela empresa japonesa Ajinomoto sob vários nomes comerciais. Nos últimos anos, muitos pesquisadores tentaram, sem sucesso, produzir esta enzima em *E. coli* e em outros microrganismos utilizando ferramentas de engenharia genética, apesar da expressão desta enzima ter se mostrado difícil, com baixa atividade e produzindo proteínas insolúveis em corpos de inclusão.

1.4 *Paenibacillus* sp. JDR-2

Paenibacillus sp. JDR-2 é uma bactéria Gram-positiva isolada em troncos da árvore-do-âmbar (*Liquidambar styraciflua*) nos estado da Flórida, Estados Unidos, e capaz de degradar xilanos com extrema eficiência (CHOW et al., 2012). Seu genoma foi sequenciado recentemente, apontando ter cerca de 7 Mb de tamanho e cerca de 6.400 genes, dentre os quais dez foram preditos como contendo domínios de transglutaminase. Análises filogenéticas

feitas por Chow e colaboradores, em 2006, mostraram que *Paenibacillus* sp. JDR-2 é relacionado a organismos produtores de transglutaminase do gênero *Bacillus*, como *B. subtilis*, sendo, portanto, um possível produtor desta enzima.

2. JUSTIFICATIVA

Transglutaminases são enzimas muito utilizadas na indústria de alimentos, atuando na reestruturação de alimentos proteínicos. Ultimamente, esta enzima vem sendo utilizada nas mais variadas indústrias na obtenção de produtos biotecnológicos, com destaque para a indústria farmacêutica, em que transglutaminases estão sendo utilizadas para ligar anticorpos a drogas a fim de tornar o tratamento de certas doenças mais efetivo. Um dos empecilhos para o seu uso está no fato da produção desta enzima ser considerada cara e ineficiente, sendo produzida através do cultivo de *Streptovercillium mobaraense*, que exige substratos caros, o tempo de produção da enzima é demorado e rende 2 U/mL de atividade enzimática. Isso posto, a busca por novos microrganismos produtores de transglutaminases pode ser uma alternativa para encontrar enzimas mais eficientes em determinados processos e que possam ser produzidas em menor tempo e a custos mais baixos, aumentando a eficiência do processo e diminuindo, conseqüentemente, os custos dos produtos derivados deste processo.

3. OBJETIVO

3.1 Objetivo Geral

Identificar genes codificadores de transglutaminase no genoma de *Paenibacillus* sp. JDR-2, expressá-los em um sistema heterólogo e realizar a caracterização físico-química desta enzima.

3.2 Objetivos Específicos

- Identificar sequências putativas codificadoras de transglutaminase microbiana no genoma de *Paenibacillus* sp. JDR-2 por análises *in silico*;
- Amplificar as ORFs encontradas a fim de confirmar sua presença no genoma do microrganismo;
- Clonar estas ORFs em vetor de expressão para *E. coli* pET23d+;
- Expressar as sequências clonadas em *E. coli*;
- Verificação da atividade de transglutaminase pelas sequências expressas e caracterização físico-química das transglutaminases encontradas

4. MATERIAIS E MÉTODOS

4.1 Materiais utilizados

4.1.1 Material biológico

Paenibacillus sp. JDR-2 foi obtido através do Banco Genético de *Bacillus* (Columbus, OH, EUA). As cepas comerciais de *E. coli* foram obtidas de seus respectivos distribuidores: *E. coli* XL1-Blue foi obtida da Stratagene (Santa Clara, California, EUA), *E. coli* BL21 (DE3) Rosetta foi obtida da Novagen (Darmstadt, Alemanha), *E. coli* BL21 (DE3) Arctic Express foi obtida da Agilent Technologies (Santa Clara, CA, EUA) e *E. coli* BL21 (DE3) pLysE (Thermo Scientific, MA, EUA). O vetor pET23d+ foi também obtido da empresa Novagen (Darmstadt, Alemanha).

4.1.2 Kits de purificação

Os kits Purelink® Quick Plasmid Miniprep e Purelink® Quick Gel Extraction and PCR Purification foram obtidos da empresa ThermoFischer Scientific (Waltham, MA, EUA).

4.1.3 Material químico

Lisozima, N-lauroilsarcosina, 1- α -ciano- 4-ácido hidroxicinaminico, ácido-L-glutâmico- γ -hidroxâmico, ampicilina, tetraciclina, cloranfenicol, gentamicina, cloreto de cálcio (CaCl₂), glicerol, cloreto de sódio (NaCl), DTT, ácido trifluoroacético, bicarbonato de amônio acetona, PMSF, fosfato monopotássico (KH₂PO₄), azul de coomassie, tris, Z-gln-gly, glicina e glutatona reduzida foram obtidos da empresa Sigma-Aldrich (Saint Louis, MO, EUA). As enzimas T4 DNA ligase e Taq polimerase Platinum hot-start, o corante SYBR® Safe DNA Gel Stain foi obtidos da empresa ThermoFischer Scientific (Waltham, MA, EUA). O kit Wizard® DNA Genomic Purification e a tripsina foram obtidos da Promega (Madison, WI, EUA). As enzimas de restrição NheI e XhoI foram obtidas do New England Biolabs (Ipswich, MA, EUA). A síntese de primers foi feita pela empresa Invitrogen (CA, EUA). Fosfato de sódio monobásico (Na₂HPO₄), metanol, HEPES, sódio-dodecilsulfato, ácido clorídrico, imidazol, acrilamida e ureia foram compradas da empresa J. T. Baker (Center Valley, PA, EUA). Triton X-100 foi obtido da empresa Dinâmica (Diadema, SP, Brasil). Triptona e extrato de levedura foram obtidos da empresa Neogen (Lansing, MI, EUA). TEMED foi obtido da Acros Organics (New Jersey, EUA). Bis-acrilamida foi obtida da

Fischer Scientific (MA, EUA). Ácido tricloroacético e cloreto de ferro (III) foram comprados da Synth (Diadema, SP, Brasil).

4.2 Busca por sequências codificadoras de transglutaminase em *Paenibacillus* sp. JDR-2

A procura por sequências codificadoras de transglutaminase no genoma de *Paenibacillus* sp. JDR-2 foi feita através de uma busca entre as fases abertas de leitura

(ORFs) anotadas automaticamente a partir da sequência do genoma deste microrganismo depositadas no NCBI. As sequências nucleotídicas obtidas foram traduzidas com a ferramenta de tradução de nucleotídeos do portal de recursos de bioinformática ExpASy (<http://web.expasy.org/translate/>) e a análise dos domínios contidos nessas sequências peptídicas foi feita com auxílio do banco de dados de domínios conservados (CDD) no NCBI. Essas sequências foram analisadas por BLASTp a fim de verificar similaridade com outras transglutaminases descritas e foram verificadas quanto à presença de um peptídeo sinal com o programa SignalP4.1 a fim de verificar a possível secreção destas proteínas pela célula. Foi feita também uma inferência da massa molecular e do ponto isoelétrico (pI) de cada uma das proteínas codificadas utilizando o algoritmo Compute pI/Mw no ExpASy (http://web.expasy.org/compute_pi/).

4.3 Extração de DNA genômico

O DNA de *Paenibacillus* sp. JDR-2 foi extraído utilizando-se o kit Wizard® Genomic DNA Purification Kit, utilizando uma incubação prévia com lisozima, seguindo o protocolo do fabricante para a extração de material genômico de bactérias Gram-positivas.

4.4 Projeto de iniciadores e amplificação das sequências de interesse

As sequências nucleotídicas das fases abertas de leitura (ORFs) 4765, 4870 e 5011, que codificam uma transglutaminase putativa, cada, no genoma de *Paenibacillus* sp. JDR-2, foram amplificadas através da reação em cadeia da polimerase (PCR).

Para tanto, oligonucleotídeos iniciadores (*primers*) foram desenhados, com auxílio do programa SnapGene, de forma que a temperatura de anelamento preditiva dos *primers* com o genoma do microrganismo ficasse entre 50 e 60°C. À extremidade 5' de cada um dos *primers* direto (*forward*) e reverso (*reverse*) desenhados foi adicionada a sequência de restrição para as enzimas NheI e XhoI, respectivamente, bem como a adição de quatro bases aleatórias para a facilitação da clonagem (tabela 1).

Tabela 1: *primers* projetados para a amplificação das sequências 4765, 4870 e 5011 de *Paenibacillus* sp. JDR-2

Primer	Sequência
Pjdr_4765F	5' <i>ccccgctagc</i> AAGCTAAATATTTCCCACGTTACGC
Pjdr_4765R	5' <i>tttctcgag</i> ATCCTCGATTTTACGCACATC
Pjdr_4870F	5' <i>ccccgctagc</i> GAATGGATTCGCGCTGGGAC
Pjdr_4870R	5' <i>tttctcgag</i> TACTGCCCGGATATGAGTCTTG
Pjdr_5011F	5' <i>ccccgctagc</i> CGAAAAACAGGTTGGAAAATAGCG
Pjdr_5011R	5' <i>tttctcgag</i> TTGATAATCCAGCTGAACCTGC

*As sequências em itálico nos primers F correspondem à sequência reconhecida pela enzima NheI, enquanto as sequências em itálico nos primers R correspondem ao sítio de reconhecimento da enzima XhoI.

As condições de amplificação compreenderam: 1) 5 minutos a 95°C para ativação da enzima Taq polimerase Platinum hot-start; 2) 1 minuto a 95°C para desnaturação das fitas de DNA; 3) 1 minuto a 60°C para o anelamento dos primers; 4) 1 minuto e 15 segundos a 72°C para a amplificação dos primers (para a sequência 4765 foi utilizado 1 minuto); 5) repetição dos passos de 2 a 4 por 35 vezes e 6) 10 minutos extras a 72°C para amplificação de qualquer primer que tenha restado. Os produtos de PCR foram verificados por gel de agarose 1,5%, corados com SYBR® Safe DNA Gel Stain e visualizados no sistema de fotodocumentação L-Pix Touch (Loccus Biotecnologia). A ORF 4765 foi amplificada com os *primers* Pjdr_4765F e Pjdr_4765R, a ORF 4870 foi amplificada utilizando os primers Pjdr_4870F e Pjdr_4870R e a ORF 5011 foi amplificada com os primers Pjdr_5011F e Pjdr_5011R.

A purificação dos fragmentos amplificados, clivados com enzimas de restrição ou não, e de vetores clivados com enzimas de restrição foi feita utilizando-se o kit Purelink Quick Gel Extraction and PCR Purification, utilizando-se 100 µL de reação de PCR ou 50 µL de reação de clivagem, seguindo as instruções do fabricante. Os produtos purificados foram analisados por eletroforese em gel de agarose 1,5 %, nas mesmas condições acima citadas e quantificados com espectrofotômetro Nanodrop® 2000 (Thermo Scientific).

4.5 Clivagem de sequências amplificadas por PCR e de vetores com enzimas de restrição

A fim de facilitar a clonagem de sequências amplificadas nos vetores selecionados, tanto o vetor quanto o produto de PCR a ser clonado foram previamente clivados com enzimas de restrição NheI e XhoI para que fossem produzidas extremidades coesivas em cada uma destas moléculas. As extremidades do vetor e do produto de PCR clivadas com a mesma enzima foram posteriormente ligadas através de uma reação de ligação, concluindo, assim, o processo de clonagem.

Para realizar a clivagem das sequências 4765, 4870 e 5011, amplificadas por PCR, incubou-se 43µL de sequências purificadas ou de vetor extraído por miniprep, 1µL da enzima NheI, 1µL da enzima XhoI e 5µL de tampão NEB 2.1, de acordo com as instruções do fabricante. A reação se deu em um período de 3h a 37°C seguidos de 20 minutos a 65°C para desnaturação das enzimas de restrição. Os produtos clivados foram purificados, quantificados e posteriormente armazenados a -20°C.

4.6 Clonagem das sequências codificadoras das proteínas putativas 4765, 4870 e 5011 de *Paenibacillus* sp. JDR-2 em vetor de expressão pET23d+

Para realizar a expressão heteróloga das sequências de *Paenibacillus* sp. JDR-2 em *Escherichia coli*, o vetor pET23d+ foi o vetor escolhido. Este vetor possui 3663 pb, opera sob ação do promotor T7, possui um gene que confere resistência à ampicilina para seleção de transformantes e uma sequência codificadora de uma cauda de hexahistidinas na extremidade carbóxi do polipeptídeo formado, o que facilita a purificação das proteínas expressas (figura 3).

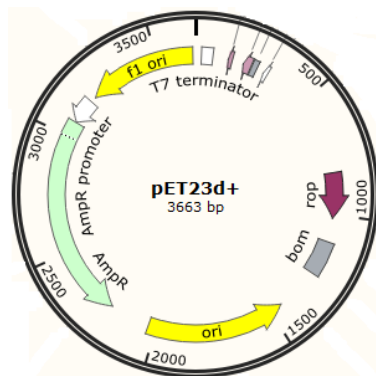


Figura 3. Vetor pET23d+: representação do vetor pET23d+ e os elementos moleculares que o compõem. Em verde é representado o gene codificador de ampicilina e em amarelo a origem de replicação do vetor.

A ligação das sequências 4765, 4870 e 5011 com o vetor pET23d+ foi feita incubando-se 10 ng de vetor clivado em uma razão molar de inserto:vetor de 1:3 utilizando a enzima T4 DNA ligase em uma reação de 20µL, de acordo com o fabricante, produzindo os plasmídeos pETG4765, pETG4870 e pETG5011, respectivamente (figura 4).

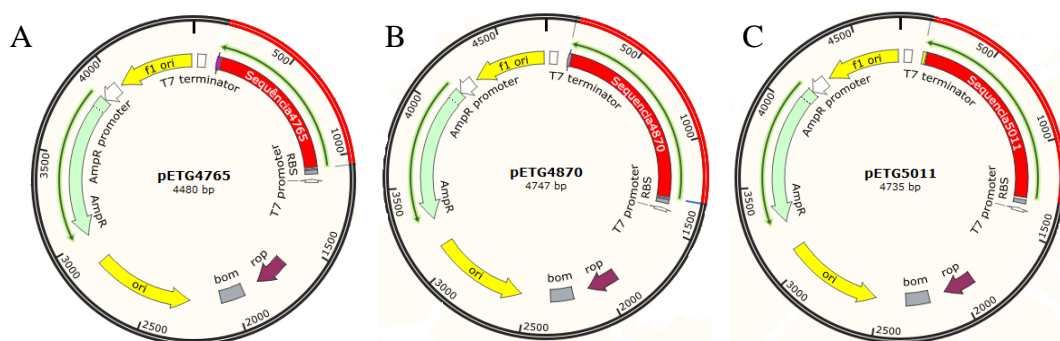


Figura 4. Mapa dos plasmídeos recombinantes construídos: estrutura do vetor pET23d+ após a inserção de cada uma das ORFs entre os sítios de restrição para NheI e XhoI: plasmídeo recombinante pETG4765 (A), pETG 4870 (B) e pETG 5011 (C). Em vermelho é possível ver onde cada ORF foi inserida no vetor com relação aos demais elementos.

4.7 Preparação de células competentes

Todas as cepas de *Escherichia coli* utilizadas neste trabalho foram transformadas utilizando-se o método químico de transformação. *E. coli* XL1 Blue foi a cepa utilizada para a clonagem, ligação e propagação de plasmídeos, enquanto *E. coli* BL21 (DE3) pLysE, *E. coli* BL21 (DE3) Rosetta e *E. coli* BL21 (DE3) Arctic Express foram utilizadas para a expressão de proteínas recombinantes.

Para a produção de células quimiocompetentes, primeiramente fez-se o pré-inóculo de uma colônia isolada da cepa a ser transformada em meio LB-Miller (10 g/L NaCl, 10 g/L triptona e 5 g/L de extrato de levedura) contendo antibióticos de seleção (20 µg/µL de gentamicina para Arctic Express, 12,5 µg/µL de tetraciclina para XL1-Blue e 34 µg/µL de cloranfenicol para Rosetta e pLysE), crescendo-a a 37°C e 180 rpm por 16 h em agitador horizontal. Após esse período, o inóculo foi feito adicionando-se o pré-inóculo em meio LB-Miller na razão de 1:100 (v/v), permanecendo nas mesmas condições até que fosse atingida uma OD₆₀₀ de 0,6. Neste ponto, as células foram centrifugadas por 10 minutos a 5000 rpm em centrífuga Eppendorf® 5810R e o sobrenadante descartado. As células foram resuspendidas em 10 mL de CaCl₂ 100mM gelado e mantidas em gelo por 30 minutos. As células foram então centrifugadas novamente a 5000 rpm por 10 minutos e o sobrenadante foi novamente descartado. As células decantadas foram resuspendidas em 1mL de CaCl₂ 100mM contendo 10% glicerol e armazenadas a -80°C em alíquotas de 50 µL.

4.8 Transformação de diferentes cepas de *E. coli*

As transformações realizadas foram feitas incubando-se 5 µL da reação de ligação ou 1µL de plasmídeo purificado com 50 µL da cepa de *E. coli* a ser utilizadas por 30 minutos em gelo, seguido de 45 segundos (ou 25 se *E. coli* BL21 (DE3) Arctic Express) a 42°C seguidos de mais dois minutos em gelo. Após esse período, adicionou-se 250 µL de LB-Miller às células, que foram incubadas a 37°C com agitação de 180 rpm de 30 a 60 min. Um volume de 100 µL de células transformadas foi plaqueado em placas de Petri contendo meio LB-Miller sólido com 100 µg/µL de ampicilina para seleção das células que receberam o plasmídeo mais o antibiótico de seleção específico para cada cepa, sendo mantidas em estufa a 37°C por 16 horas. As reações de ligação foram transformadas em *E. coli* XL1-Blue, enquanto transformações com plasmídeos purificados para expressão foram feitas em *E. coli* cepas pLysE, Rosetta ou Arctic Express.

4.9 Seleção, purificação e análise de restrição de plasmídeos recombinantes

A fim de verificar se a reação de ligação foi efetiva, colônias isoladas de *E. coli* XL1-Blue transformadas com a reação de ligação foram selecionadas aleatoriamente e inoculadas

em 25mL de meio LB-Miller contendo 100 µg/mL de ampicilina e 12,5 µg/mL de tetraciclina, e cultivadas em agitador horizontal por 16h a 37°C e 200 rpm. Alíquotas de 500 µL de cada cultura foram estocadas em glicerol 10% e armazenadas a -20 e -80°C. Para a extração de plasmídeo foi utilizado o kit PureLink® Quick Plasmid Miniprep, utilizando 3 mL de cultura, de acordo com as instruções do fabricante.

A análise da ligação foi feita clivando-se o plasmídeo extraído com as enzimas NheI e XhoI, como no item 4.5, verificando-se o resultado da clivagem em eletroforese em gel de agarose 1,5 %. Caso a ligação tenha sido bem-sucedida, espera-se ver duas bandas no gel após a reação de clivagem, uma correspondendo ao vetor pET23d+ e outra ao inserto nela contido. A clivagem de plasmídeos que mostraram duas bandas do tamanho esperado no gel de agarose foram enviadas para sequenciamento pela empresa ACTGene no Centro de Biotecnologia da UFRGS.

4.10 Expressão heteróloga das proteínas recombinantes de *Paenibacillus* sp. JDR-2 (4765, 4870 e 5011) em *E. coli*

Para verificar a expressão das proteínas recombinantes em *E. coli*, os plasmídeos recombinantes foram transformados nas cepas de *Escherichia coli* BL21 (DE3) Rosetta, Arctic Express e pLysE. Testes de expressão foram feitos pré-inoculando-se colônias isoladas transformadas com os plasmídeos pETG4765, pETG4870 ou pETG5011 em 25 mL de meio LB-Miller a 37°C, 180 rpm em agitador horizontal por 16h. O inóculo foi feito diluindo-se o pré-inóculo cem vezes a um volume final de 25mL de meio LB-Miller nas mesmas condições acima descritas até que se atingisse uma OD₆₀₀ de 0,6. O restante da cultura foi induzida com uma concentração final de 0,5mM de IPTG e mantida sob as mesmas condições por 24h. Alíquotas de 1mL foram retiradas em no momento da indução e 2, 3 e 24h após a indução para posterior verificação da expressão das proteínas recombinantes.

4.11 Eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE)

Para verificar a expressão de proteínas recombinantes, as alíquotas de 1mL retiradas em pontos específicos, como explicado no item anterior, foram centrifugadas a 5000 rpm por 5 minutos a fim de separar a fração celular, que foi então resuspendida em 100 µL de tampão

fosfato-salino (PBS) (1,37 M NaCl, 27 Mm KCl, 100 mM Na₂HPO₄ e 18 mM KH₂PO₄, pH 7,4). Para realizar a lise das células, os extratos celulares foram incubados com tampão de lise (200 mM Tris-Cl, pH 6.8; 400 mM DTT; 8% SDS; 0,4% azul de coomassie; 40% glicerol) e fervidos a 95°C por 5 minutos. Para a separação de proteínas presentes nos lisados celulares utilizou-se o protocolo de Laemmli (1970), com gel de resolução 12% e gel de empilhamento 5%, com voltagem constante de 200V. Após o término da corrida, os géis foram corados em solução de azul de coomassie (50% metanol, 10% ácido acético, 1 mM azul de Coomassie) por 2h com agitação e descorados em solução descorante (10% ácido acético), sendo agitados até a visualização das bandas.

4.12 Espectrometria de massas

As bandas proteicas do gel SDS-PAGE foram recortadas e digeridas com adição de 0,5µg de tripsina + 17 µL de tampão bicarbonato de amônio 0,1M pH 8,0. Depois da digestão os peptídeos foram aplicados em coluna de fase reversa Poros 50 R2 (PerSeptive Biosystems). Os peptídeos purificados foram hidratados em 6 µL de solução matriz (5mg mL⁻¹ α-ciano-4-ácido hidroxicinaminico em 50% de acetronitrila e 0,1% de ácido trifluoracetico (v/v). Desta hidratação 2 µL de cada amostra foram aplicados na placa de MALDI-TOF/TOF (Axima Performace, Kratos-Shimadzu, Manchester, UK). Os MS/MS obtidos de cada spot digerido foi analisado utilizando o software MASCOT (Matrix Science, London, UK) e o banco de dados do NCBIInr.

4.13 Purificação de proteínas por cromatografia de afinidade

A purificação da proteína recombinante obtida foi feita por cromatografia de afinidade utilizando a resina HisLink™ Protein Purification (Promega). As culturas expressas segundo o item 4.10 foram centrifugadas a 6.000 rpm por 30min em centrífuga Sorvall RC-5B Plus a 4°C e o precipitado resuspendido em 10mL de tampão A (20 mM HEPES, 500 mM NaCl, pH 8) ou tampão B (20 mM Hepes, 150 mM KCl, pH 8); acrescidos de PMSF e Triton X-100 a um volume final de 1 mM e 1%, respectivamente). As amostras foram sonicadas por 12 minutos em ciclos de 10 segundos a uma potência de 50% com intervalos de 10 segundos a cada pulso em sonicador Sonic Dismembrator (Fischer Scientific). Após a sonicação, as amostras foram centrifugadas a 10.000 rpm por 30 minutos. Os sobrenadantes foram submetidos à cromatografia de afinidade por níquel.

Procedeu-se à purificação por meio de lavagens com concentrações crescentes de imidazol: tampão de lavagem 1 (tampão A ou B acrescido de 20 mM imidazol) e tampão de lavagem 2 (tampão A ou B acrescido de 40 mM imidazol). Quando o tampão B foi utilizado, foi testada uma terceira etapa de lavagem, contendo tampão B acrescido de 40 mM de imidazol e 2M de ureia, pH 8, seguido de uma quarta etapa de lavagem, idêntica ao tampão de lavagem 2 para remoção da ureia. Finalmente as proteínas foram eluídas com tampão de eluição (tampão A ou B acrescido de 250 mM imidazol). Todas as frações foram coletadas e as amostras foram submetidas à SDS-PAGE para análise.

Após a purificação, as frações desejadas foram dialisadas contra 100 mL do tampão utilizado (tampão A ou tampão B) por pelo menos três vezes em membrana Spectra/Por (MWCO 3500) a 4° C com agitação.

4.13 Solubilização dos corpos de inclusão

A fim de tentar solubilizar os corpos de inclusão, modificou-se a metodologia de Peternel e colaboradores (PETERNEL et al., 2008), na qual um detergente fraco é adicionado aos grânulos para que aquelas proteínas que possuam um enovelamento adequado sejam desprendidas e testadas quanto à sua atividade. A estratégia consiste na diluição dos corpos de inclusão em tampão B na proporção de 1:40 (m/v) acrescidos de N-lauroilsarcosina a uma concentração final de 0.2%. Essa solução foi mantida a 20°C sob agitação de 140 rpm em agitador horizontal por 24 horas, sendo centrifugada a 10000 rpm por 30 minutos para separar a fração solubilizada da insolúvel, que foram checadas em SDS-PAGE e quanto à atividade da fração solúvel.

4.14 Teste de atividade enzimática

A atividade enzimática foi determinada pelo método de formação de hidroxamato desenvolvido por Grossowicz e colaboradores (GROSSOWICZ et al., 1950). Em suma, 400 µL de amostra foram incubados com 350 µL do reagente A (tampão tris-HCl 200 mM, 100 mM de hidroxilamina, 50 mM de CaCl₂, 10 mM de glutatona reduzida e 30 mM de Z-Gln-Gly) por 10 minutos a 37°C. A reação foi parada com adição de 400 µL de reagente B (solução 1:1:1: (v/v/v) de HCl 3N, ácido tricloroacético 12% e FeCl₃.6H₂O 5% dissolvido em HCl 0,1N), sendo então lida sua absorbância a 525 nm. A curva padrão foi feita com ácido L-

glutâmico- γ -monohidroxâmico e uma unidade de atividade enzimática (U) foi definida como a quantidade de enzima capaz de formar 1,0 μmol de hidroxamato por minuto.

5 RESULTADOS

5.1 Identificação de possíveis sequências codificadoras de transglutaminase utilizando ferramentas de bioinformática

O primeiro passo para a identificação de transglutaminases codificadas pelo genoma de *Paenibacillus* sp. JDR-2 foi feito a partir de uma busca entre as ORFs de *Paenibacillus* sp. JDR-2 anotadas automaticamente e depositadas no NCBI utilizando-se as palavras "Transglutaminase" e "*Paenibacillus* sp. JDR-2". O resultado da busca identificou dez ORFs: 0339, 0834, 0937, 1245, 1489, 3259, 4024, 4765, 4870 e 5011 como sendo transglutaminases, como mostra a tabela 2.

Tabela 2: ORFs de *Paenibacillus* sp. JDR-2 anotadas como transglutaminase. Busca realizada em 10 de janeiro de 2015.

Nome do gene/ ID	Descrição	Localização no genoma
Pjdr2_0339/ ID: 8127845	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (404157..404921)
Pjdr2_0834/ ID: 8125814	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (980768..983239)
Pjdr2_0937/ ID: 8129903	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (1115571..1116179, complement)
Pjdr2_1235/ ID: 8126175	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (1479396..1479839)
Pjdr2_1489/ ID: 8128228	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (1781222..1783861)
Pjdr2_3259/ ID: 8128471	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (3725799..3726722)
Pjdr2_4024/ ID: 8127942	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (4565733..4568003, complement)
Pjdr2_4765 / ID: 8126076	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (5482158..5483045, complement)
Pjdr2_4870/ ID: 8127408	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (5592903..5594060, complement)
Pjdr2_5011/ ID: 8129746	Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	NC_012914.1 (5732770..5733912, complement)

A partir da identificação das possíveis transglutaminases codificadas pelo genoma deste microrganismo, as sequências peptídicas foram obtidas e submetidas a análises de

domínios conservados e de predição de similaridade com outras transglutaminases. Os resultados indicam que as possíveis transglutaminases codificadas pelo genoma de *Paenibacillus* sp. JDR-2 não mostram similaridade com nenhuma outra já caracterizada e depositada em bancos de dados, sendo similares a outras transglutaminases preditas em organismos do mesmo gênero, somente. Todas as sequências possuem ao menos um domínio de transglutaminase predito (Apêndice A-J).

Das sequências analisadas, apenas duas mostraram possuir um peptídeo sinal da extremidade amino terminal (Apêndice K). A predição do ponto isoelétrico e do peso molecular das transglutaminases codificadas por cada ORF encontrada revelou que a maioria possui pI entre 5 e 6, sendo a transglutaminase 3259 a única que apresentou pI básico. A maioria dos peptídeos também apresentou peso molecular em torno de 30 e 40 kDa, embora três peptídeos tenham seu peso molecular predito acima de 80 kDa (tabela 3). A sequência 0937 é uma sequência truncada cuja sequência que codifica o domínio de transglutaminase não possui um aminoácido iniciador, e foi, portanto, descartada das predições seguintes.

Tabela 3: ponto isoelétrico (pI) e peso molecular estimado da sequência peptídica codificada por cada uma das ORFs identificadas como transglutaminases

ORF	pI predito	Peso molecular estimado (Daltons)
0339	4.98	29204.58
0834	5.36	87554.38
0937	-	-
1235	10.36	16501.68
1489	5.04	99339.89
3259	9.62	34260.21
4024	5.37	85895.47
4765	5.34	33359.30
4870	5.82	42099.33
5011	5.68	42723.22

5.2 Amplificação e clonagem das sequências 4765, 4870 e 5011 de *Paenibacillus* sp. JDR-2 em *Escherichia coli*

A amplificação das sequências 4765, 4870 e 5011 e sua purificação para posterior clonagem foram realizadas com sucesso utilizando a metodologia previamente descrita. Os

primers desenhados para as sequências 4765 e 5011 amplificaram uma banda robusta e única entre 48 e 60°C no tamanho esperado quando o DNA genômico de *Paenibacillus* sp. JDR-2 foi utilizado como DNA molde (Apêndice L). Os *primers* desenhados para a sequência 4870 também amplificaram uma banda única, mas somente entre 49 e 60°C. Os sítios de inserção das ORFs 4765, 4870 e 5011 fazem com que o peptídeo sintetizado possua dois aminoácidos a mais na extremidade amino do que o peptídeo parental (Apêndice M).

A clonagem das sequências amplificadas no vetor pET23d+ foi efetuada com sucesso em *E. coli* XL1-Blue (figuras 5-7), sendo um grande número de colônias transformantes obtidas após a transformação com a reação de ligação. Os plasmídeos recombinantes foram sequenciados e os insertos foram confirmados (Apêndice N).

Foram observadas muitas colônias transformadas nas placas dos controles que haviam sido transformados unicamente com o vetor pET23d+ clivado com as enzimas *NheI* e *XhoI* em presença da enzima T4 ligase, o que indica a clivagem parcial do vetor com apenas uma das enzimas utilizadas. Isso pode ser devido ao fato de o tempo de incubação das enzimas de restrição com o vetor intacto ter sido de apenas 3 horas; porém, como plasmídeos recombinantes foram obtidos na primeira tentativa de clonagem, outras condições não foram testadas. Corroborando essa hipótese está o fato de que os plasmídeos recombinantes clivados com ambas as enzimas mostraram uma banda fraca na altura do inserto. A eficiência das transformações não foi verificada.

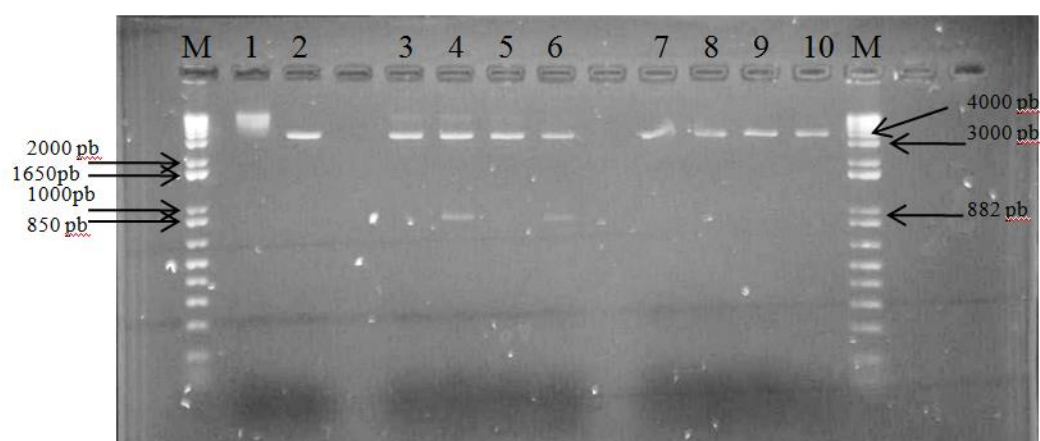


Figura 5. Teste de clonagem da sequência 4765. Análise feita em gel de agarose 1,5%. Quatro colônias transformadas com a reação de ligação foram selecionadas ao acaso e tiveram seus plasmídeos extraídos para verificação da clonagem. O vetor pET23d+ clivado (canaleta 2) e não clivado (canaleta 1) foram usados para comparação. Canaletas 3-6 mostram os quatro possíveis construtos pETG4765 clivados com as enzimas *NheI* e *XhoI*, sendo que apenas aqueles das canaletas 4 e 6 apresentaram a banda do inserto em 882 pb, indicando que haviam sido ligados com a ORF 4765. Canaletas 7 e 8 mostram o plasmídeo da canaleta 4 clivado

com a enzima NheI e XhoI, respectivamente; enquanto canaletas 9 e 10 mostram o plasmídeo da canaleta 6 clivado com as mesmas enzimas.

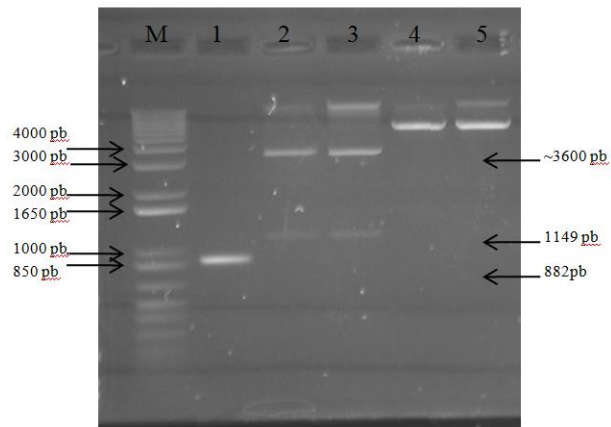


Figura 6. Teste de clonagem da sequência 4870. Análise feita em gel de agarose 1,5%. duas colônias transformadas com a reação de ligação foram selecionadas ao acaso e tiveram seus plasmídeos extraídos para verificação da clonagem (canaletas 4 e 5). Estes plasmídeos foram clivados com as enzimas NheI e XhoI (canaletas 2 e 3). A sequência 4765 foi utilizada para comparação de sequências (canaleta 1). M= marcador molecular.

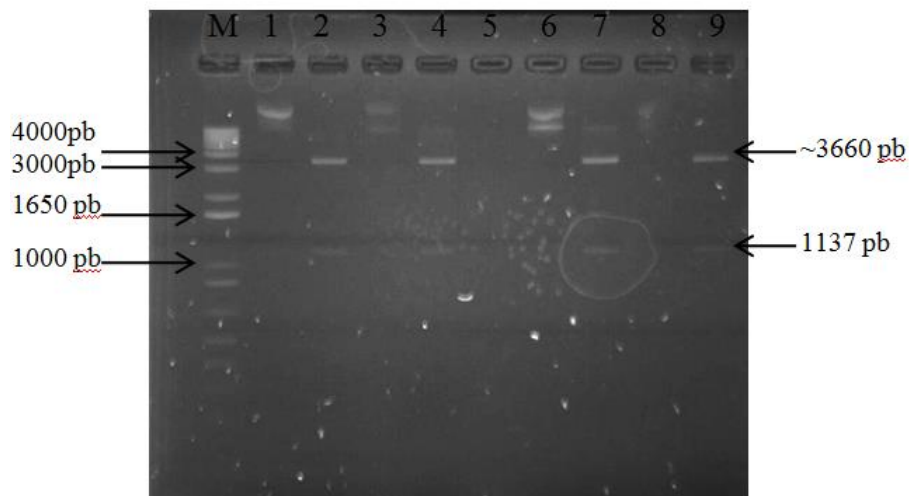
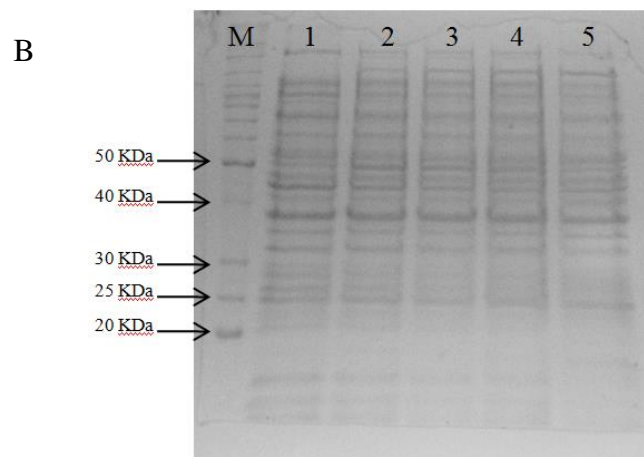
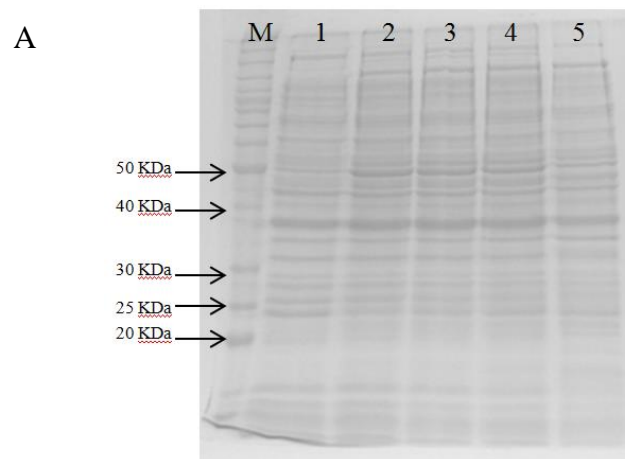


Figura 7. Teste de clonagem da sequência 5011. Análise feita em gel de agarose 1,5%. Quatro colônias transformadas tiveram seus plasmídeos extraídos e clivados para verificação do inserto. Canaletas 1, 3, 6 e 8 mostram cada um destes plasmídeos, enquanto as canaletas 2, 4, 7 e 9 mostram estes plasmídeos após clivagem com as enzimas NheI e XhoI para verificação da clonagem. Pode-se ver a banda correspondente à sequência 5011 na altura de 1137 pb e a do vetor pET23d+ entre 3000 e 4000 pb. M corresponde ao marcador de peso molecular.

5.3 Análise da expressão das construções pETG4765, pETG4870 e pETG5011

5.3.1 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em *E. coli* BL21 (DE3) pLysE

A transformação dos plasmídeos recombinantes em *E. coli* pLysE foi bem-sucedida para os três plasmídeos recombinantes em questão bem como para o vetor pET23d+, usado como controle. Os testes de indução, no entanto, não mostraram a expressão das proteínas recombinantes codificadas pelos três plasmídeos recombinantes construídos (Figura 8 A-C).



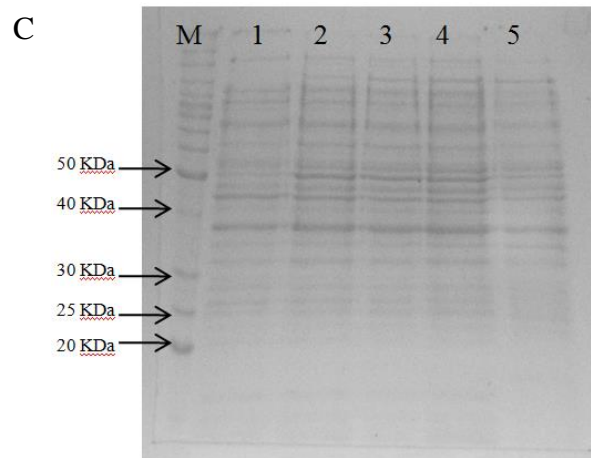


Figura 8. Expressão dos vetores recombinantes em *Escherichia coli* pLysE. Análise feita por SDS-PAGE 12%. Verificação da expressão dos plasmídeos recombinantes pETG4765 (A), pETG4870 (B) e pETG5011 (C) em *E. coli* pLysE utilizando 0,5 mM IPTG. Os resultados mostram nenhuma banda induzida em torno do tamanho esperado (cerca de 34 kDa para a proteína 4765 e 42 kDa para as proteínas 4870 e 5011). (M corresponde ao marcador de peso molecular, 1 corresponde ao controle negativo (pET23d+ vazio), 2 corresponde ao lisado celular antes do momento da indução, 3 ao lisado 2h após a indução, 4 ao lisado celular 3h após a indução e 5 ao lisado celular 24h após a indução)

5.3.2 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em *Escherichia coli* BL21 (DE3) Rosetta

A transformação dos plasmídeos recombinantes foi bem-sucedida nesta cepa, apresentando muitos transformantes em cada transformação. Tentativas de expressar as proteínas recombinantes codificadas por cada vetor recombinante falharam, não sendo possível observar nenhuma banda induzida em comparação ao controle (Figura 9).

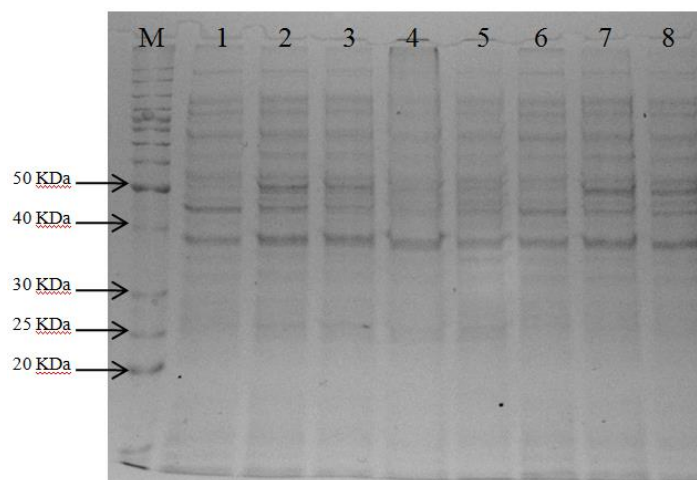


Figura 9. Expressão dos plasmídeos recombinantes em *Escherichia coli* Rosetta. Análise feita por SDS-PAGE 12%. Verificação da expressão dos plasmídeos recombinantes

pETG4765 (canaletas 3 e 4), pETG4870 (canaletas 5 e 6) e pETG5011 (canaletas 7 e 8) em *E. coli* Rosetta. O vetor pET23d+ foi utilizado como controle (canaletas 1 e 2). Os resultados mostram nenhuma banda induzida em torno do tamanho esperado (. Canaletas 1, 3, 5 e 7 correspondem às culturas não induzidas, ao passo que as canaletas 2, 4, 6 e 8 correspondem àquelas induzidas com 0,5 mM de IPTG após 24h de indução. (M corresponde ao marcador de peso molecular).

5.3.3 Transformação e expressão dos vetores pETG4765, pETG4870 e pETG5011 em Arctic Express

A transformação dos plasmídeos recombinantes em *Escherichia coli* BL21 (DE3) Arctic Express seguindo as instruções do desenvolvedor foi eficiente para a transformação dos plasmídeos pETG4765 e pETG5011, somente, sendo que a transformação do plasmídeo pETG4870 não mostrou qualquer colônia. Tentou-se aumentar o tempo do choque térmico a 42°C de 25 para 45 segundos, sem sucesso. Sendo assim, a expressão da sequência 4870 em Arctic Express não pôde ser testada.

Os testes de expressão em Arctic Express foram surpreendentes quanto aos seus resultados, uma vez que esta cepa foi a única em que, após indução, foi verificada uma banda de proteína recombinante com o vetor pETG4765 (Figura 10B). Análise da expressão mostra uma banda proteica entre 30 e 40 kDa, o tamanho esperado da proteína recombinante 4765. Colônias transformadas com o vetor pETG5011 não mostraram expressão da proteína recombinante quando comparadas com a mesma cepa transformada com o vetor pET23d+ (Figura 10B).

A banda induzida da proteína 4765 foi verificada desde o momento da indução, o que pode indicar uma expressão basal alta desta proteína clonada no vetor pET23d+. A verificação da solubilidade desta proteína mostrou que ela está na forma de corpos de inclusão em sua maior parte, sendo que esta condição não foi alterada com testes de indução com menos IPTG (0,1 e 0,2 mM) nem com alterações na agitação da cultura (125 ou 250 rpm). Em todas as condições testadas foi possível verificar a presença da chaperona de 56 kDa codificada pela cepa de *E. coli* utilizada.

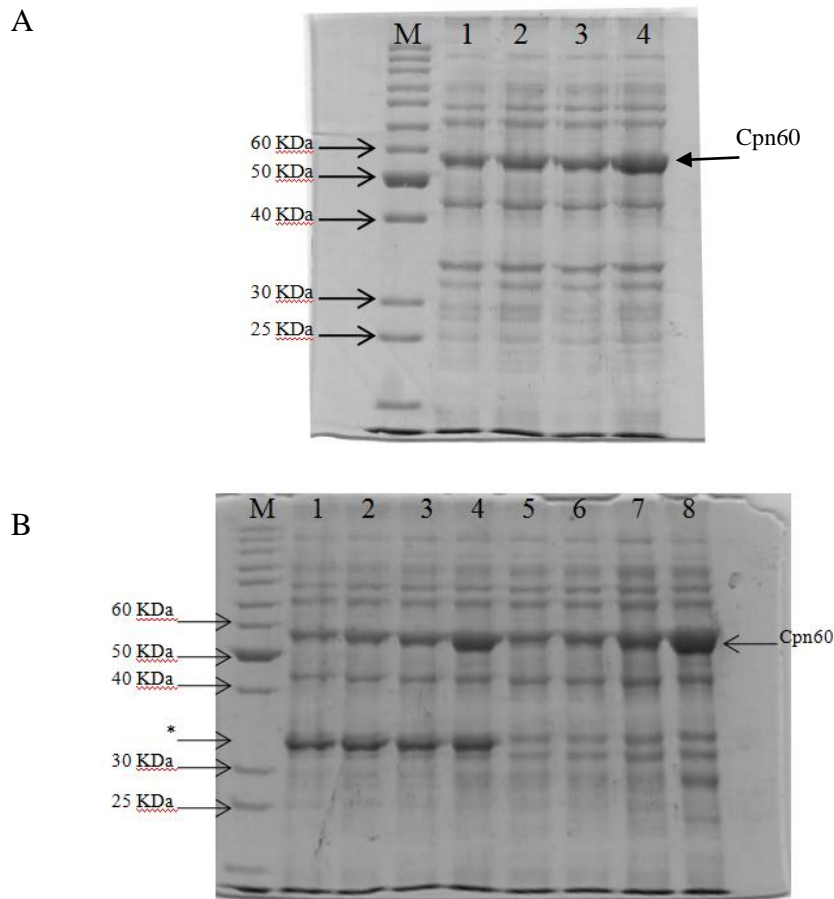


Figura 5. Teste de expressão dos plasmídeos recombinantes em *Escherichia coli* Arctic Express. Análise feita por SDS-PAGE 12%. Verificação da expressão dos plasmídeos recombinantes pETG4765 (B, 1-4) e pETG5011 (B, 5-8) em *E. coli* Arctic Express induzidos com IPTG 0,5 mM. Figura A corresponde à célula transformada com o vetor pET23d+, utilizado como controle. Canaletas 1-A e 1 e 5-B correspondem ao extrato celular não induzido; canaletas 2-A e 2 e 6-B ao extrato 2h após a indução; canaletas 3-A e 3 e 7-B ao extrato 3h após a indução, enquanto as canaletas 4-A e 4 e 8-B correspondem ao extrato 24h após a indução. É possível verificar a presença da chaperona 60 expressa por esta cepa de *E. coli* e da proteína recombinante (*).

5.4 Espectrometria de massas

A banda de proteína induzida obtida na cultura de *E. coli* Arctic Express foi excisada do gel e submetida à espectrometria de massas para análise. Os resultados mostraram que apenas dois picos, correspondendo a dois peptídeos tripsinizados (Tabela 4).

Tabela 4. Dados sobre os peptídeos obtidos através de espectrometria de massas

Início – Fim	Massa esperada	Massa observada	Score	Peptídeo
102 – 115	1704.8033	1705.8107	84	R.EAWTWLETEEAANR.F
160 - 176	1985.1947	1985.051	92	K.IRTEFIYDPEATTVTTK.A

Análises dos peptídeos obtidos feitas pelo *software* Mascot indicam que estes correspondem à proteína 4765, cobrindo 10% de sua sequência (Figura 11).

Protein View: gi 506326564					
transglutaminase [Paenibacillus sp. JDR-2]					
Database:	NCBIInr				
Score:	206				
Nominal mass (M_r):	33359				
Calculated pI:	5.35				
Taxonomy:	Paenibacillus sp. JDR-2				
Protein sequence coverage: 10%					
Matched peptides shown in bold red .					
1	MKLNISHVTQ	YDYEESATDS	VNEIRLTPST	NERQSCYQQA	IFIEPNASLF
51	SYEDFFGNRV	HSFSVNPSHK	RLTIRTQMTV	VTRHAPSPEE	QAALLNGKLA
101	REAWTWLETE	EAANRFAEFL	LPTSADVTA	DVERYAESVD	AREGEGISVL
151	SWLISLSSKI	RTEFIYDPEA	TTVTTKASDL	IERKRGVCQD	FAHLMIACCR
201	FHQIPARYVS	GYHFVGDLOQ	GSADFEQASH	AWVEAYVPQL	GWCSFDPTNE
251	APVGERYVKL	GHGRDYKDIV	PVKGVYRGS	LQQLKVTVDV	RKIED

Figura 6. Identificação da proteína recombinante por MS/MS: os dados do *software* Mascot contendo o MS-FIT (“ProteinProspector”) identificados por meio dos íons obtidos pelo espectro de massas da proteína transglutaminase4765 de *Paenibacillus* sp. JDR-2. A sequência de aminoácidos mostrando a localização dos peptídeos no mapa de cobertura em comparação à transglutaminase 4765 deste organismo (em vermelho é possível verificar os peptídeos tripsinizados).

5.5 Purificação da proteína recombinante 4765

Após a verificação da expressão da proteína 4765 em Arctic Express, as mesmas condições de cultivo foram utilizadas para a produção desta proteína em 300 mL e 1L de meio LB-Miller a fim de tentar escalonar o processo e obter uma maior quantidade de proteína recombinante. Frações de todas as etapas de lavagem e eluição foram coletadas a fim de observar a melhor condição de purificação desta proteína, sendo que em nenhuma condição testada foi possível obter uma única banda purificada (Figura 12 A-B).

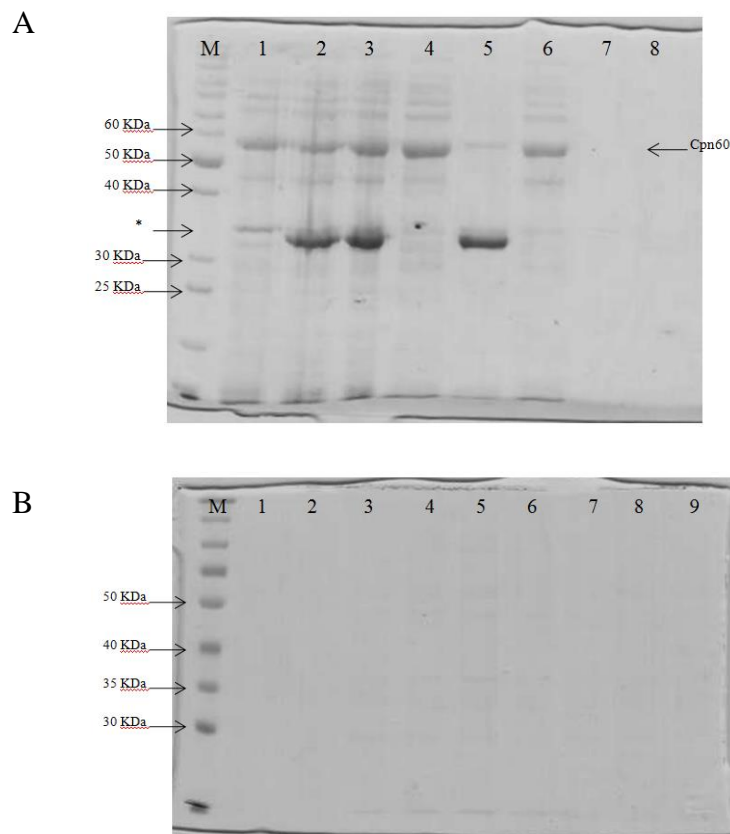


Figura 7. Expressão e Purificação da proteína 4765. Análise feita por SDS-PAGE 12%. Tentativa de purificação da proteína 4765 por cromatografia de afinidade com resina de níquel. No gel mostrado em A podemos ver o controle negativo (1), células transformadas com o plasmídeo pETG4765 (2) no momento da indução e 20h após a indução (3), a fração solúvel (4) e insolúvel (5) da fração celular após o fim do cultivo. A fração não ligada à resina de níquel (6), eluída após lavagem com 20mM de imidazol (7) e com 40mM de imidazol (8) também são mostradas. Na figura B podemos ver frações de 1mL do eluído com 250mM (1-9), em que nenhuma banda é observada.

Como praticamente toda a proteína recombinante é expressa na forma de corpos de inclusão, tentou-se solubilizar as proteínas presentes nestes grânulos utilizando-se uma estratégia não-desnaturante, na qual uma concentração baixa de um detergente fraco, N-lauroilsarcosina, foi incubada com os grânulos insolúveis a fim de solubilizá-los. Após 20h de incubação a 20°C e 140 rpm de agitação, podemos ver que parte da proteína foi solubilizada, ainda que uma grande quantidade permanecesse na fração insolúvel (Figura 13A).

Tentativas de purificação da proteína resolubilizada foram feitas utilizando cromatografia de afinidade com resina de níquel, utilizando-se o protocolo com 150mM de KCl com 2M de ureia. Apesar de uma fração considerável da proteína ter eliminada nas lavagens prévias à eluição final, uma fração da mesma foi eluída com o tampão B com 250 mM de imidazol ao final da purificação (Figura 13B). Esta fração foi dialisada para

verificação da atividade enzimática, porém, não foi verificada qualquer atividade característica de transglutaminases.

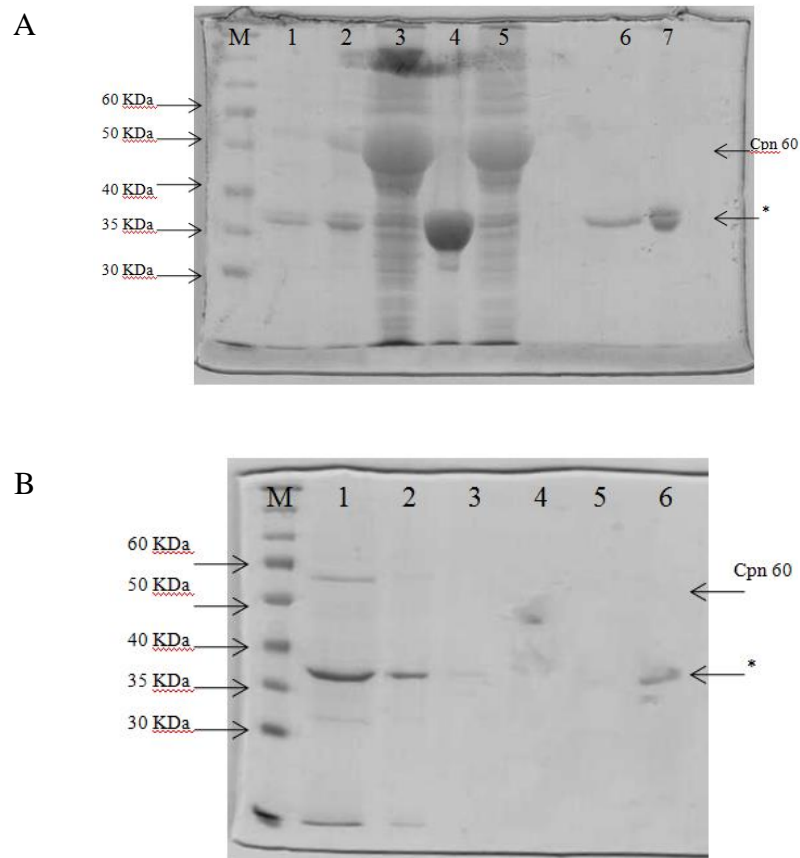


Figura 8. Síntese da proteína 4765 e sua solubilização a partir de corpos de inclusão. Análise feita por SDS-PAGE 12%. Em A é possível ver a cultura antes da indução (1) e 20h após a indução (2) com 0,5 mM de IPTG. A fração celular solúvel (3) não-solúvel (4) e a fração solúvel não ligada à resina de níquel (5) também são mostradas. Após a incubação com N-lauroilsarcosina é possível ver a fração solubilizada (6) e a que se manteve insolúvel (7). A figura B mostra a tentativa de purificação das proteínas solubilizadas: fração não ligada à resina de níquel (1), após lavagem com 20mM de imidazol (2), 40mM de imidazol (3), 40mM de imidazol mais 2M de ureia (4), 40mM de imidazol (5) e fração eluída com 250mM de imidazol (6). A proteína recombinante é indicada com um * e a chaperona 60 pela sua sigla (Cpn60).

6. DISCUSSÃO

Desde sua descoberta, na década de 50, transglutaminases vêm sendo estudadas, sendo que um grande número de funções atribuídas a esta enzima já foram identificadas, que vão desde participar da coagulação sanguínea à estruturação de tecidos, em eucariotos (GREENBERG; BIRCKBICHLER; RICE, 1991), até a reticulação de proteínas do endosporo (KOBAYASHI et al., 1996) e atuação como um fator de virulência (YU et al., 2015), em procariotos. Dada sua atividade de modificação nas cadeias laterais de proteínas, essa enzima possui, além do seu importante papel fisiológico, um grande potencial biotecnológico, o que corroborou para o crescente interesse no estudo dessa enzima nos anos posteriores à sua descoberta (RACHEL; PELLETIER, 2013) (figura 14).

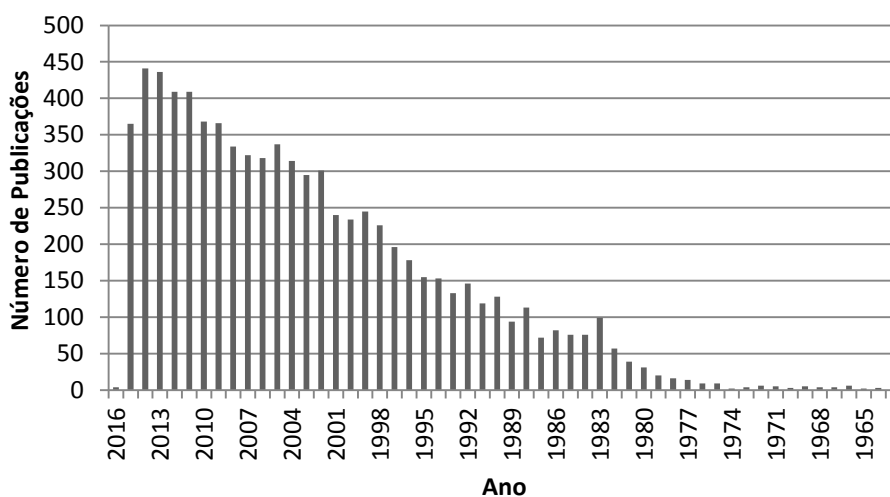


Figura 9. Relação de publicações sobre transglutaminases de 1959 a 2016: é possível verificar o crescente número de publicações ao longo dos últimos anos, sendo que mais de 400 artigos sobre transglutaminases foram depositados em cada um dos últimos cinco anos. (Fonte: National Center for Biotechnology Information)

O presente trabalho visou a identificar e caracterizar novas transglutaminases codificadas por *Paenibacillus* sp. JDR-2, uma espécie recentemente descoberta cujo genoma (CHOW et al., 2012) e transcriptoma (SAWHNEY et al., 2015) já estão disponíveis para acesso. Apesar dessas informações disponíveis sobre esta espécie, os únicos trabalhos publicados sobre ela até o presente momento dizem respeito aos seus genes codificadores de enzimas que degradam xilanos (CHOW; NONG; PRESTON, 2007; SAWHNEY; PRESTON, 2014; STJOHN; RICE; PRESTON, 2006), o que torna *Paenibacillus* sp. JDR-2 um potencial microrganismo para identificação de enzimas de interesse biotecnológico.

Por ser uma bactéria Gram-positiva formadora de endósporo relacionada a outras já descritas como codificando uma transglutaminase, como aquelas do gênero *Bacillus*, *Paenibacillus* sp. JDR-2 é um possível produtor desta enzima. Uma análise *in silico* das possíveis transglutaminases codificadas pelo seu genoma mostrou a presença de dez ORFs que codificam polipeptídeos contendo domínios de transglutaminase, um número elevado quando se leva em consideração que nenhum outro microrganismo possui mais de uma transglutaminase descrita até hoje. Suzuki e colaboradores investigaram a presença de duas transglutaminases em *Bacillus subtilis* após a verificação de duas frações com atividade de transglutaminase que eram eluídas por cromatografia de interação hidrofóbica. Extensivas análises de caracterização, no entanto, levaram à conclusão de que estas duas frações eram referentes à mesma proteína e que *B. subtilis*, portanto, possuía apenas uma mTG (SUZUKI et al., 2000).

Considerando as transglutaminases putativas encontradas em *Paenibacillus* sp. JDR-2, observa-se uma certa discrepância entre elas referente ao seu peso molecular estimado, ao ponto isoelétrico de cada e presença de um peptídeo sinal. Das nove ORFs que codificam um peptídeo encontradas, sete possuem seu ponto isoelétrico entre pH 5 e 6, enquanto três possuem pI básico. Transglutaminases com pontos isoelétricos em faixas de pH ácido têm sido relatadas em eucariotos, como a transglutaminase do camarão branco chinês, que possui pI predito de 6.1, enquanto as mTGs descritas possuem pI básicos, como a de *S. mobaraense*, que possui pI de 8.9 (ANDO et al., 1989).

Das mTGs conhecidas até hoje, sabe-se que elas possuem um peso molecular entre 30 de 40 kDa, podendo sofrer processamento (KIELISZEK; MISIEWICZ, 2014). Das possíveis transglutaminases encontradas em *Paenibacillus* sp. JDR-2, aquelas codificadas pelas ORFs 0339, 3259, 4765, 4870 e 5011 possuem peso molecular predito em torno do esperado para uma mTG, entre 30 e 40 kDa, enquanto aquelas codificadas pelas ORFs 0834, 1489 e 4024 possuem massa predita acima de 80 kDa, peso molecular encontrado nas transglutaminases de eucariotos (LORAND; GRAHAM, 2003). A mTG codificada pela ORF 1235 apresentou tamanho incomum, de apenas 16.5 kDa, o que é muito menor do que o observado para qualquer transglutaminase conhecida. Observou-se, também, que apenas duas sequências possuem um peptídeo sinal predito, codificadas pelas ORFs 3259 e 5011, enquanto as demais não indicaram possuir peptídeo sinal na sua extremidade amino terminal. Com base nesses resultados não é possível indicar qualquer uma das proteínas hipotéticas como tendo maio

probabilidade de ser uma transglutaminase, pois apresentam características de transglutaminase eucarióticas e microbianas.

Três das nove ORFs encontradas, as ORFs 4765, 4870 e 5011 foram clonadas em vetor de pET23d+ para expressão da sequência peptídica codificada por elas a fim de verificar se seu produto peptídico possui atividade de transglutaminase. Estas ORFs foram escolhidas com base no peptídeo codificado por elas possuir massa similar às mTG descritas.

Tentativas de expressão de transglutaminases vêm sendo feitas a fim de buscar por um método mais eficiente de obtenção desta enzima e diminuição do seu custo de produção, que ainda hoje é feito através de *S. mobaraense*, o primeiro microrganismo no qual ela foi encontrada (ARRIZUBIETA, 2007). Todos os trabalhos, no entanto, em que se tentou expressar uma transglutaminase, tanto eucariótica quanto microbiana, em *E. coli*, mostraram que praticamente toda a proteína recombinante é expressa na forma insolúvel. Tentativas da otimização dos códons para melhor expressão em *E. coli* (YOKOYAMA et al., 2000) e de expressar a enzima sem seu pró-peptídeo (LIU et al., 2011) renderam proteínas insolúveis na forma de corpos de inclusão, o que também foi observado com a expressão de transglutaminase de milho (CARVAJAL et al., 2011). O recente trabalho de Rickert e colaboradores mostrou que modificações no pró-peptídeo da mTG de *S. mobaraense* e a inserção de um sítio de clivagem entre este e a enzima madura foram capazes de expressar a enzima de forma solúvel e ativa em *E. coli*, demonstrando a importância do pró-peptídeo para a estruturação da proteína madura (RICKERT et al., 2015).

A tentativa de expressão das ORFs 4765, 4870 e 5011 em *E. coli* pLysE não foi eficiente quanto à expressão das proteínas recombinantes. Esta cepa de *E. coli* reduz a expressão basal do gene recombinante ao produzir uma lisozima que especificamente degrada a polimerase que reconhece o promotor do fago T7, inibindo o vazamento da expressão por este promotor e a consequente expressão da proteína recombinante antes da indução. Esta cepa foi utilizada com o objetivo de evitar uma possível toxicidade causada pela transglutaminase devido à sua atividade de ligação de diferentes peptídeos, o que poderia acabar ligando proteínas importantes para o funcionamento do microrganismo. Em nenhuma tentativa de expressão em *E. coli* BL21 (DE3) pLysE foi observada a expressão de qualquer proteína recombinante.

Uma segunda tentativa de expressar os mesmos clones em *E. coli* BL21 (DE3) Rosetta foi feita a fim de verificar a possibilidade do uso de diferentes códons entre este microrganismo e *Paenibacillus* sp. JDR-2, do qual os genes foram amplificados, ser um impasse na expressão destes clones. Infelizmente, a suplementação de códons raros nesta cepa de *E. coli* também não foi suficiente para verificação da expressão de proteínas recombinantes por nenhum dos clones.

Como as cepas antes utilizadas não foram capazes de expressar as proteínas de interesse, uma terceira alternativa, *E. coli* BL21 (DE3) Arctic Express, foi utilizada para expressão dos três clones obtidos. Esta cepa é modificada geneticamente para conter duas chaperonas que são ativas a baixas temperaturas em seu genoma, e, portanto, facilitaria o enovelamento e solubilidade de proteínas expressas em temperaturas tão baixas quanto 13°C (FERRER et al., 2003). Com isto, poderíamos verificar se a expressão dos clones seria obtida pelo auxílio no enovelamento das cadeias polipeptídicas formadas. Apesar do plasmídeo pETG4870 não ter transformado nesta cepa e do plasmídeo recombinante pETG5011 não ter mostrado nenhuma expressão da proteína recombinante, o plasmídeo recombinante pETG4765 mostrou uma banda de indução no tamanho esperado, o que indica que provavelmente esta proteína precisa de auxílio de chaperonas para seu enovelamento correto.

A expressão da proteína 4765 foi eficiente, mas análises realizadas mostraram que toda a proteína expressa estava na fração insolúvel do lisado celular, um problema recorrente na expressão desta enzima. Tentativas de purificação utilizando cromatografia de afinidade com resina de níquel não renderam nenhuma proteína purificada, nem mesmo com tentativas de escalonamento da cultura para 1L de meio de cultivo foram utilizada para tentar purificar quantidade minutas de proteína solúvel. Tentou-se modificar o protocolo de purificação desenvolvido por Belval e colaboradores para verificar se a chaperona poderia estar agregada à proteína recombinante, levando-a à insolubilidade, mas não foi obtido sucesso.

Altos níveis de expressão de proteínas recombinantes em *Escherichia coli* podem levar à degradação ou à agregação destas no citosol do microrganismo em estruturas denominadas corpos de inclusão (BANEYX; MUJACIC, 2004). Estes corpos são estruturas densas, largas e relativamente esféricas que possuem de 0.2 a 1.2 µm, sendo compostos de 80 a 95% da proteína recombinante produzida. Apesar de a maioria dos corpos de inclusão relatados apresentar proteínas cuja estrutura secundária não corresponde àquela encontrada na

proteína funcional, corpos de inclusão não convencionais, contendo uma grande quantidade de proteínas ativas têm sido relatados (GARCÍA-FRUITÓS et al., 2005).

A proteína 4765 foi expressa a 13°C, mostrando-se completamente insolúvel na forma de corpos de inclusão. Uma vez que foi verificado que todas as estratégias para purificá-la utilizando cromatografia de afinidade com resina de níquel não funcionaram pela falta de proteína solúvel, tentou-se solubilizar estes grânulos utilizando um detergente fraco para verificar se a expressão em baixa temperatura poderia reter proteínas ativas nos corpos de inclusão. Os resultados mostraram que uma parte do grânulo foi solubilizada, mas sem qualquer atividade enzimática, corroborando parcialmente com os resultados de Peternel e colaboradores, que mostraram que o uso de uma concentração baixa de N-lauroilsarcosina foi suficiente para remover proteínas corretamente enoveladas dos corpos de inclusão e com atividade (PETERNEL et al., 2008). Resultados de Groot e Ventura, e de Jevševar e colaboradores mostraram, também, que a temperatura afeta negativamente a atividade das proteínas que formam o corpo de inclusão, sendo que este possui mais proteínas ativas quando formado a temperaturas mais baixas (DE GROOT; VENTURA, 2006; JEVŠEVAR et al., 2005). Analisando os dados obtidos por estas pesquisas, podemos inferir que a proteína solubilizada dos corpos de inclusão pode estar enovelada corretamente, visto sua solubilização, e que sua atividade pode não ter sido aferida devido a esta não possuir atividade de transglutaminase, o que é possível por se tratar de uma proteína anotada automaticamente, ou que as condições nas quais ela foi mantida acarretaram na perda de sua atividade, devendo novas condições serem testadas para que se comprove esta hipótese.

7. CONCLUSÃO

Até o presente momento foi possível:

- Identificar 9 possíveis ORFs codificadoras de transglutaminase no genoma de *Paenibacillus* sp. JDR-2, um número ainda não descrito em nenhum outro microrganismo;
- Amplificar eficientemente três ORFs de *Paenibacillus* sp. JDR-2 anotadas como transglutaminase e depositadas no NCBI, confirmando sua presença no genoma deste microrganismo;
- Clonar estas ORFs em vetor de expressão pET23d+ para expressão em *E. coli*;
- Expressar a proteína codificada pela ORF 4765 de *Paenibacillus* sp. JDR-2 e verificar sua insolubilidade em *E. coli*, corroborando com outros dados obtidos para a expressão de transglutaminase em *E. coli*;
- Solubilizar a proteína presente nos corpos de inclusão utilizando uma estratégia não-desnaturante.

8. PERSPECTIVAS

Como sequência a este trabalho, serão realizadas:

- Solubilização da proteína 4765 através de sua desnaturação e subsequente *refolding*, seguindo o protocolo de Yang e colaboradores (YANG et al., 2011b) a fim de retirá-la dos corpos de inclusão e torná-la corretamente enovelada para verificar sua atividade enzimática;
- Clonagem das sequências 4870 e 5011 em novos vetores para expressão em *E. coli* ;
- Clonagem das sequências 4870 e 5011 nos vetores pPICZ α e pGAPZ α para expressão em *Picchia pastoris*

9. REFERÊNCIAS

- ANDO, H. *et al.* Purification and characteristics of a novel transglutaminase derived from microorganisms. **Agricultural and Biological Chemistry**, v. 53, n. February 2015, p. 2613–2617, 1989.
- ANDO, H. *et al.* Crystal structure of microbial transglutaminase from *Streptovercillium mobaraense*. **Journal of Biological Chemistry**, v. 25, n. 2, p. 1–8, maio 2014.
- BANEYX, F.; MUJACIC, M. Recombinant protein folding and misfolding in *Escherichia coli*. **Nature Biotechnology**, v. 22, n. 11, p. 1399–1408, 2004.
- BERNARD, B. K.; TSUBUKU, S.; SHIOYA, S. Acute toxicity and genotoxicity studies of a microbial transglutaminase. **International journal of toxicology**, v. 17, n. 6, p. 703–721, 1998.
- CAMPOS, N. *et al.* Rice transglutaminase gene: Identification, protein expression, functionality, light dependence and specific cell location. **Plant science : an international journal of experimental plant biology**, v. 205-206, p. 97–110, maio 2013.
- CARVAJAL, P. *et al.* Activity of maize transglutaminase overexpressed in *Escherichia coli* inclusion bodies: an alternative to protein refolding. **Biotechnology progress**, v. 27, n. 1, p. 232–40, jan. .
- CHEN, J. S.; MEHTA, K. Tissue transglutaminase: an enzyme with a split personality. **The international journal of biochemistry & cell biology**, v. 31, n. 8, p. 817–36, ago. 1999.
- CHOW, V. *et al.* Complete genome sequence of *Paenibacillus* sp. strain JDR-2. **Standards in Genomic Sciences**, v. 6, n. 1, p. 1–10, 2012.
- CHOW, V.; NONG, G.; PRESTON, J. F. Structure, function, and regulation of the aldouronate utilization gene cluster from *Paenibacillus* sp. strain JDR-2. **Journal of bacteriology**, v. 189, n. 24, p. 8863–70, dez. 2007.
- DE GROOT, N. S.; VENTURA, S. Effect of temperature on protein quality in bacterial inclusion bodies. **FEBS Letters**, v. 580, n. 27, p. 6471–6476, 2006.
- DENNLER, P. *et al.* Microbial transglutaminase and c-myc-tag: a strong couple for the functionalization of antibody-like protein scaffolds from discovery platforms. **Chembiochem : a European journal of chemical biology**, v. 16, n. 5, p. 861–7, 23 mar. 2015.
- FERNANDES, P. Enzymes in Food Processing: A Condensed Overview on Strategies for Better Biocatalysts. **Enzyme Research**, v. 2010, p. 1–19, 2010.
- FERRER, M. *et al.* Chaperonins govern growth of *Escherichia coli* at low temperatures. **Nature biotechnology**, v. 21, n. 11, p. 1266–7, nov. 2003.
- FOLK, J. E.; FINLAYSON, J. S. The epsilon-(gamma-glutamyl)lysine crosslink and the catalytic role of transglutaminases. **Advances in protein chemistry**, v. 31, p. 1–133, jan.

1977.

GARCÍA-FRUITÓS, E. *et al.* Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. **Microbial cell factories**, v. 4, p. 27, 2005.

GREENBERG, C. S.; BIRCKBICHLER, P. J.; RICE, R. H. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. **FASEB journal: official publication of the Federation of American Societies for Experimental Biology**, v. 5, n. 15, p. 3071–7, dez. 1991.

GRIFFIN, M.; CASADIO, R.; BERGAMINI, C. M. Transglutaminases: nature's biological glues. **The Biochemical journal**, v. 368, n. Pt 2, p. 377–96, 1 dez. 2002.

GROSSOWICZ, N. *et al.* THE ENZYMATIC FORMATION OF HYDROXAMIC ACIDS FROM GLUTAMINE AND ASPARAGINE. **J. Biol. Chem.**, v. 187, n. 1, p. 111–125, 1 nov. 1950.

HEADON, D. R.; WALSH, G. The industrial production of enzymes. **Biotechnology Advances**, v. 12, n. 4, p. 635–646, 1994.

IKURA, K. *et al.* Incorporation of lysyldipeptides into food protein by transglutaminase. **Agricultural and Biological Chemistry (Japan)**, 1985.

JEVŠEVAR, S. *et al.* Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. **Biotechnology Progress**, v. 21, n. 2, p. 632–639, 2005.

JOSTEN, A. *et al.* Use of microbial transglutaminase for the enzymatic biotinylation of antibodies. **Journal of Immunological Methods**, v. 240, p. 47–54, 2000.

KASHIWAGI, T. *et al.* Crystal structure of microbial transglutaminase from *Streptovercillium mobaraense*. **Journal of Biological Chemistry**, v. 277, n. 46, p. 44252–44260, 2002.

KIELISZEK, M.; MISIEWICZ, A. Microbial transglutaminase and its application in the food industry. A review. **Folia microbiologica**, v. 59, n. 3, p. 241–50, maio 2014.

KOBAYASHI, K. *et al.* ϵ -(γ -Glutamyl)lysine cross-links of spore coat proteins and transglutaminase activity in *Bacillus subtilis*. **FEMS Microbiology Letters**, v. 144, p. 157–160, 1996.

LE, B. V. *et al.* Characterization of *Anopheles gambiae* transglutaminase 3 (AgTG3) and its native substrate Plugin. **The Journal of biological chemistry**, v. 288, n. 7, p. 4844–53, 15 fev. 2013.

LIU, S. *et al.* The order of expression is a key factor in the production of active transglutaminase in *Escherichia coli* by co-expression with its pro-peptide. **Microbial cell factories**, v. 10, p. 112, jan. 2011.

LORAND, L. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. **Annals of the New York Academy of Sciences**, v. 936, p. 291–311, jan. 2001.

- LORAND, L.; GRAHAM, R. M. Transglutaminases: crosslinking enzymes with pleiotropic functions. **Nature Reviews Molecular Cell Biology**, v. 4, n. 2, p. 140–156, 2003.
- MARGOSIAK, S. A. *et al.* Identification of the Large Subunit of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase as a Substrate for Transglutaminase in *Medicago sativa* L. (Alfalfa). **Plant physiology**, v. 92, n. 1, p. 88–96, jan. 1990.
- MOTOKI, M.; SEGURO, K. Transglutaminase and its use for food processing. **Trends in Food Science & Technology**, v. 9, n. 5, p. 204–210, maio 1998.
- MUSZBEK, L. *et al.* Factor XIII: a coagulation factor with multiple plasmatic and cellular functions. **Physiological reviews**, v. 91, n. 3, p. 931–72, jul. 2011.
- MYCEK, M. J. *et al.* Amine incorporation into insulin as catalyzed by transglutaminase. **Archives of biochemistry and biophysics**, v. 84, p. 528–40, out. 1959.
- MYNENI, V. D.; MELINO, G.; KAARTINEN, M. T. Transglutaminase 2--a novel inhibitor of adipogenesis. **Cell death & disease**, v. 6, p. e1868, jan. 2015.
- NAGY, V.; SZAKACS, G. Production of transglutaminase by *Streptomyces* isolates in solid-state fermentation. **Letters in Applied Microbiology**, v. 47, p. 122–127, 2008.
- NIELSEN, P. M. Reactions and potential industrial applications of transglutaminase. Review of literature and patents. **Food biotechnology (New York, N.Y.) (USA)**, 1995.
- OORT, M. VAN. Enzymes in food technology – introduction Maarten van Oort. **Food Technology**, 2010.
- OTENG-PABI, S. K.; KEILLOR, J. W. Continuous enzyme-coupled assay for microbial transglutaminase activity. **Analytical Biochemistry**, v. 441, n. 2, p. 169–173, 2013.
- OUTTRUP, H. properties and application of a Thermostable Maltogenic Amylase Produced by A Strain of *Bacillus* Modified by Recombinant-DNA Techniques. p. 405–411, 1984.
- PETERNEL, S. *et al.* Engineering inclusion bodies for non denaturing extraction of functional proteins. **Microbial cell factories**, v. 7, p. 34, 2008.
- PORTA, R. *et al.* Promising Perspectives for Transglutaminase In “Bioplastics” Production. **Journal of Biotechnology & Biomaterials**, v. 01, n. 03, p. 1–4, 2011.
- RACHEL, N. M.; PELLETIER, J. N. Biotechnological applications of transglutaminases. **Biomolecules**, v. 3, n. 4, p. 870–88, 22 jan. 2013.
- RICKERT, M. *et al.* Production of soluble and active microbial transglutaminase in *Escherichia coli* for site-specific antibody drug conjugation. **Protein science : a publication of the Protein Society**, 20 out. 2015.
- SAWHNEY, N. *et al.* Transcriptomic analysis of xylan utilization systems in *Paenibacillus* sp. strain JDR-2. **Applied and environmental microbiology**, v. 81, n. 4, p. 1490–501, fev. 2015.
- SAWHNEY, N.; PRESTON, J. F. GH51 arabinofuranosidase and its role in the

methylglucuronarabinoxylan utilization system in *Paenibacillus* sp. strain JDR-2. **Applied and environmental microbiology**, v. 80, n. 19, p. 6114–25, out. 2014.

STJOHN, F. J.; RICE, J. D.; PRESTON, J. F. *Paenibacillus* sp. strain JDR-2 and XynA1: a novel system for methylglucuronoxylan utilization. **Applied and environmental microbiology**, v. 72, n. 2, p. 1496–506, fev. 2006.

STROP, P. Versatility of microbial transglutaminase. **Bioconjugate Chemistry**, v. 25, p. 855–862, 2014.

SUZUKI, S. *et al.* Purification and characterization of novel transglutaminase from *Bacillus subtilis* spores. **Biosci. Biotechnol. Biochem.**, 2000.

VASIC-RACKI, D. **History of industrial biotransformations. Dreams and realities.** [s.l: s.n.].

YANG, M. TE *et al.* Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. **Journal of Biological Chemistry**, v. 286, n. 9, p. 7301–7307, 2011a.

YANG, Z. *et al.* Highly Efficient Production of Soluble Proteins from Insoluble Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. **PLoS ONE**, v. 6, n. 7, p. e22981, 29 jul. 2011b.

YEE, V. C. *et al.* Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 15, p. 7296–300, 19 jul. 1994.

YOKOYAMA, K. I. *et al.* Overproduction of microbial transglutaminase in *Escherichia coli*, in vitro refolding, and characterization of the refolded form. **Bioscience, biotechnology, and biochemistry**, v. 64, n. 6, p. 1263–70, 22 jun. 2000.

YOKOYAMA, K.; NIO, N.; KIKUCHI, Y. Properties and applications of microbial transglutaminase. **Applied microbiology and biotechnology**, v. 64, n. 4, p. 447–54, maio 2004.

YU, J. *et al.* Functional and Structural Characterization of the Antiphagocytic Properties of a Novel Transglutaminase from *Streptococcus suis*. **Journal of Biological Chemistry**, v. 290, n. 31, p. 19081–19092, 2015.

YURIMOTO, H. *et al.* The pro-peptide of *Streptomyces mobaraensis* transglutaminase functions in cis and in trans to mediate efficient secretion of active enzyme from methylotrophic yeasts. **Bioscience, biotechnology, and biochemistry**, v. 68, n. 10, p. 2058–69, out. 2004.

ZHAO, H.; CHOCKALINGAM, K.; CHEN, Z. Directed evolution of enzymes and pathways for industrial biocatalysis. **Current Opinion in Biotechnology**, v. 13, n. 2, p. 104–110, 2002.

ZHU, Y. *et al.* Microbial transglutaminase—a review of its production and application in food

processing. **Applied Microbiology and Biotechnology**, v. 44, n. 3-4, p. 277–282, dez. 1995.

ZHU, Y.; TRAMPER, J. Novel applications for microbial transglutaminase beyond food processing. **Trends in biotechnology**, v. 26, n. 10, p. 559–65, out. 2008.

APÊNDICE A - Identificação *in silico* da sequência 0339 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 0339 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

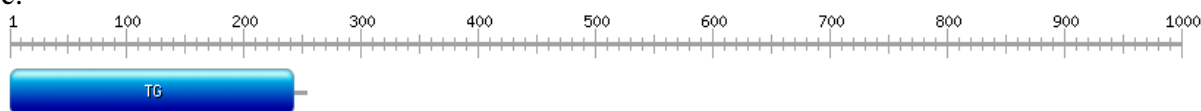
a.

ATGCTTCAAATTTCCGGGCACTGACGCCGGGGATATTAACCCGTTTGCCTAACGCCGGTCGAGCGGGACGTGCTGCAAGGCA
 AGCAGAACAGTCCCGTTGTCTACCAATATGATTCAATGGACTCGCTCCGGTTCGAGCTGCGCCTCCGGGACCGTACGGTGAAT
 GCGGCCAGAATGCTGCAGGACAGCCGGGCGGACTTCGCGGTATTCCGTAATCCAGATGCAATCCGCGTTATTGGACGCGGA
 CGGAGAACGGGGCTTCCAGATGAATCCGGATGTCCAGCCTTCGGATGCGATTAACGATATTTTTCAGAATGGCGAGAATTA
 CGGGTTCGAATGCGCGACCGCGATTGTAATCGTGTGTACAAAGCGGTGCTGGATAACAATCGGGGAGCAAGCGTTCAACAAG
 CTGTTTAATCCGCTGTATCTGCGGGACTGGAACCATGACAGCGACTTGTGGTTTATTACGACACATAACAAGAACGAATCGTA
 CCCGGGCGACGTGATGTACTCCGCAATCCCGATTTCGACCGGGAGACGCCGGAATGGCAAGGGGAGAACGTCGTGAAGCTG
 TCGGATAATCTGTTCTTCGGCCACGGCATCGGTGTTGAGACGGGTGAACAGATCGTCAGCGCATTAAATAGAATGCGCGAGC
 CGTTCAGCACGGAATCGGCTTATCTCGACGATTTGGTCGTTTCATCCGAACTTCCAGCATTGGAAGCAGATCGCGGCTCGGGGC
 AGCCGCTTTGCGGTAAGACTGTAA

b.

MLQISGTDAGDINPFALTPVERDVLQGKQNSPVVYQYDSMDSLRFELRLRDRTVNAARMLQDSRADFAVFRNSRCNPRYWTRTEN
 GGFQMNPDVQPSDAINDIFQNGENYGFECATAIVIVLYKAVLDTIGEAFNKLFNPLYLRDWNHDSDLWFITTHNKNESYPGDVM
 YFRNPDFDRETPEWQGENVVKLSNLFHGIGVETGEQIVSALNRMREPFSTESAYLDDLVVHPNFQHLKQIAARGSRFAVRL

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Proteína-glutamina-gama-glutamiltransferase (<i>Paenibacillus</i> sp. JDR-2)	530	530	100%	0.0	100%
Proteína hipotética (<i>Paenibacillus</i> sp. OSY-SE)	296	296	99%	1e-96	54%
Proteína-glutamina-gama-glutamiltransferase (<i>Paenibacillus alvei</i>)	295	295	95%	3e-06	54%
Proteína-glutamina-gama-glutamiltransferase (<i>Paenibacillus</i> sp. JDR-2)	292	292	95%	4e-95	54%
Transglutaminase (<i>Paenibacillus alvei</i>)	291	291	95%	9e-95	54%

APÊNDICE B - Identificação *in silico* da sequência 0834 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 0834 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

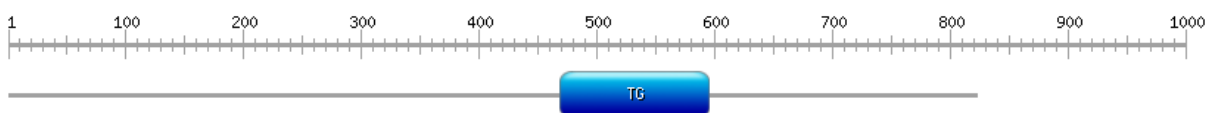
a.

ATGCGAGACTCCACTCCTGATAACGGGTCGATAGGCTTCCGGGCACTGACGACCATTGTTATTGCGCTGTTGGCCGAATG
 GCTGCTGCCATGGCTCCATTACAAGGAATGGGCGGATGTATACCAGATTGGCCCGTTGCTGACGGTGGTCGGCTGCGTGATGG
 TTATGGGGCTTTTTGCCCCGGCTTGGTATATGTCGCTCATCATGAACGGCTTCTTATGCGTCATGACGATCATGCTTCTATTTA
 AAGGAAATGACGAAAGCGCTGCATTGGCTTGCGAACACCCGGGCATGCTGGCCGGAGATATTA AAAAGCTGTTCACTTA
 CGGCATGTGGACGATGACCAGCGAGCTGAGGACGCTGCTGCTGTTCTGAGGCTGGGCTATGCTGGCTCCTGCGCTGCAGGCG
 CTTGTATGGATGAGGCAAGTTGCCCTTGGACTGACGGCAATCACGGTCTTACCTGTTGATTATTCATACCTGGGTTGGAATC
 GACGTGTTAACGGTTTTGCTTCGCACAACCGCGGAGGGCCTGCTGTTAGGTGCCGTCGTAACGATGGCCAGAGTAAGGCGGA
 TGCAGGGCGGCGGACGGAATTCGACCGCCAGCTCCATCCGAAATGGCTGGCTGCCGGCGAATTTCTGGTGTCTTTATGCGTC
 GGCGTCGGAATGCTGTTATCCAGCGGGCAGCCTAGGCAAACGGAGCCTTCGGCATGGACCCTAAACGCGGCAACCGGCTCG
 AACAGGCGATGGCGGAGCTTGGCAAGGAGAATTATTCGGCAGCCATGCAGGCGGCAATTTGGTACGGCGGGGCAAGGCCGAG
 GCATGACCGGATATGGCTTCGACGACCGGGAGCTTGGGCTTCCCTATCCAGGACGATACGATAGTGTTCAGCGGGTTTTCC
 CCGGTTACAGACCTACTGGAGAGGTGAATCCAAGGCAGTCTATAACGGACGAGGCTGGGACAGCAGCTGGAGCGGGAAGGTG
 CTGCATCCGGTGAATACGGAGATAAGCTGTCCGAGCTTGGCGTCCGTAACCGGGGATAACCGGACCATCCGGCAGACGG
 TAGTGATGACCCGAGCTTCCGGCTTCGACGGGATTGCGGCTCTTTTCGCGGGGAGCGGGGACGTGTGCTTAGCCTCTCGGG
 ATTGATCCCAATCGGCAATTAAGCACGTATGTGGAGGATGAGCTGGCGGGCCTCTGTATTCGGCGGATAAGGAACCGCAGG
 TGGAGCGTTATACCGTAGAGAGCCGGCTTCCCTAACGGACTTCCCTATTGCGGGGAGACTGACAGCTCCCGCTTTGCGGGA
 GAAACCGCGGGCGTTACGCAAACAGAGGAGACGGATCTGTCCGATTACCTGCAGCTGCCGGATTCTTTGCCGGACCGTGTGA
 AGACGCTGGCGGACAAGATCGTTGGAGATGAACGGACCGACCTTACGATCAGGTGAAAGCGGTAGAGCAATATTTGAAGA
 ACAGCTATACGTATTCGCTGGACAGCGCGGTTCCGCCAAAGGGAGCGGACTTTGTCGACTACTTCTGTTTGGAGCAGGAGCG
 GGTATTGCGGTGCATTTTCCAGCGCCATGGTTGTTCTGCTCCGCGAGGAAGGGATTCCGGCGAGATGGGTA AAAAGGCTTCGT
 ATCCGGCACCGGATCAACGGCGGGCGGGACGTTACCAGCACAGCCGGGACCCTCTACAACGTGCGCGGAAGGATGCG
 CACGCTTGGGTAGAGGTATACTTCCCGGGTGTGGCTGGGTGCCGTTTACCCGACGCGCGGGCTTTGCCGGCGGGCTTGGGGC
 AGCGGGTGTGACGCGGGGGATCGGCTCGGACGGGTTGTATCGGCCGGTGGAGCTGCCGATGATGCGGGGGCGCCAGA
 GCATGCCGATGCCGGAACGGACGGCATGCTGGCGCAGTTTGAAGCGGGCGGTGGACCACGCCCGCTCCCGCTAGCGCGAGG
 GCGGATGCCCTCGCGAGGCGGGCGGAGCGCCGCCATGGCGGGCGGGGCTTCGCCCGCCGCGCTGGCCGCCGAGGC
 GCGGCCGCGTGTCTTTCGCGCCGGCCTTGGCGTGCGGCGCAGCGCAGGCGGCTGCGCTC GCGCTC GCGCTGCGCCGGT
 ACGCCGCGGCGCATGGCGCAGGCCGCGCCGCGCTGCGCAAGAGCAGTTCCGCCGGTATCCGCCGCGGGCTGGGTCTTTGCT
 CGGCCCGGCATGATGGCGCGGCCCGCATACACCGCGCAGTACGCGGACGCCGCTGTCGCGTCTGATGCCCGCGC
 CGCGCCGAAGCGCTTCGGCGGTTATCGCTGGGACGATGCCGCCGGTTCGGGCAGCGGGGCGAATGGAGCGCCCGCCTC
 CGGAGCAGCTTCCCGCCGATGAATGCGCTGCTCGCAAGCCGGCAAAGCCGTTTCGGCGGCCACCGTGAAGCCGCCA
 ATAA

b.

MRDSTPDNGSIGFRALTTVLLFGLLAEWLLPWLHYKEWADVYQIGPLLTVVCVMVMGLFRPAWYMSLIMNGFLCVMTIMLLFK
 GNDESALHWLANYPGMLAGDIKLFYGMWMTSELRTLLLFGWAMLAPALQALVWMRQVALGLTAITVSYLLIHTWVGID
 VFNGLLRRTAEGLLLGAVVTMARVRRMQGGSEFDRQLHPKWLAAGEFLVLLCVGVGMLLSSGQPRQTEPSAWTVNAANRLEQ
 AMAELGKENYSAAMQAAIGTAGQGRGMTGYGFDDRELGASLSQDDTIVFSGFSPVQTYWRGESKAVYNGRGWDSWSGKVLHP
 VEYGDKLELAVGTSGDNRITRQTVVMTEPSASQGLPLFFAGSGGRVLSLSAIDPNRQLSTYVEDELALYSDKEA QVER YTVE
 SRLPLDTSLLRETDSSAFAGETAALQTETE DLSDYLQLPDSLPRVKTLADKIVGDERTDRYDQKAVEQYLKNSYTYSLDSAV
 PPKGADFVDYFLFEQERGYCVHFSSAMVLLREEGIPARWVKGFVSGTPVSTAGADVTSTAGTLYNVRAKDAHAHWEVYFPGAG
 WVPFDPTPGFAGGLPAGADAGGSASDGLSSAGGAADDAGAPEHADAGTDGMLAQFEAAVDHAASAVARGADALAAARSAA
 HGA AAASPAALAAAGAAALLFAAGLAAAAQRRRLRLALALRRYAAAHGAGRAAAAQEQA AVSAAAWVLLGRRMMARPPHH
 TAREYADAVSAMPGARAEALRRFIAWDDAARFGRGEWSAPPPEQLAAAMNALLGKPGKAVSAATVQAEE

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	1655	1655	100%	0.0	100%
Proteína hipotética (<i>Paenibacillus harenae</i>)	605	605	71%	0.0	49%
Proteína hipotética (<i>Paenibacillus pinihumi</i>)	361	361	93%	2e-107	36%
Proteína hipotética (<i>Paenibacillus terrigena</i>)	341	341	76%	2e-100	36%
Proteína hipotética (<i>Paenibacillus sanguinis</i>)	334	334	96%	3e-97	33%

APÊNDICE C - Identificação *in silico* da sequência 0937 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 0937 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

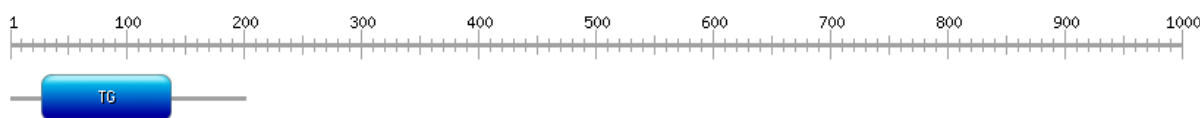
a.

TTGCCTTTCCTCCGCAACTCCCCGTCACCTGACCGACTATCTCGCCGACTCCGATATTGCGGATTTTCATCATCCCCGCTATTTCAG
GCTCTTGGTGCAGAGCTGCGTCAGCAGGCGCCCGACGAGACAACGTTTATCCGGCTTGCTTTCGAATACGTCCGGGATCAAAT
CTCCCATTCCTGGGATATCCAAAGCTCCCGCATTACTTGCAGCGCTTCCGATGTCCTCGCCATAAAGAAGGCATCTGTTACG
CCAAGTGAACCTGTTATGCGCCATTCTGCGCAGTCAGGGAATACCCGCGGGATATTGCTATCAGCGCTTGACGCTTGGGGAT
ACACCGGATACCGTTATGTCGTCCATGCTCTGAACGCCGTATATCTGCCTGCTGCGGAGCGGTGGATCAGAGTAGATGCGCG
CGGCAACAAGCCGGGATCGATGCTCAATTCTCGTTAGACCATGAACAGCTGGCCTTTCCAATCCGGAAGGAATATGACGAG
GTCGATTATCCGATCATTACTCCGAGCCTCATCCAAGCACCATCCAGGTTTTGCGGCAGCATAACGATTGCAAAGTCATGTA
TCTGCACGGTCTCCAGCTGAGCTTAA

b.

LPFLRNPSPLTDYLADSDIADFHHPAIQALGAELRQQAPDETTFIRLAFEYVRDQISHSWDIQSSRITCSASDVLAHKEGICYAKSNLL
CAILRSQGIPAGYCYQRLTLGDPDTGYVVHALNAVYLPAAERWIRVDARGNKPGIDAQFSLDHEQLAFPIRKEYDEVDPYPIYSEP
HPSTIQVLRQHNDCKVMYLHGLPAEL

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	418	418	99%	2e-107	99%
Transglutaminase (<i>Paenibacillus curdlanolyticus</i>)	294	294	97%	9e-98	67%
Transglutaminase (<i>Paenibacillus</i> sp. TCA20)	289	289	99%	8e-96	65%
Transglutaminase (<i>Paenibacillus durus</i>)	282	282	96%	4e-93	65%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus sabiniae</i> T27)	280	280	96%	2e-92	65%

APÊNDICE D - Identificação *in silico* da sequência 1235 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 1235 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

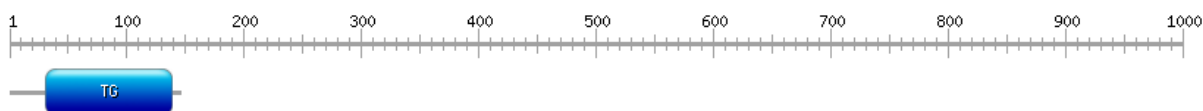
a.

ATGGGACGGTTGCGAAAATTTTGGCTTCCGATAGAGCGTTTAAGCTGCTGCTGCTGGAAGCTTTTGTGAACCTGGGTAAGGC
 CCGTATTATGAAGGCGGTTCTTTTGGCGAGAACGGCTCCGAAACTTGGTGTTGCGATGAAAGAAACCTCCTTCTCGGATTCCG
 GGGAGGAGACGGAGCTCATCCGTAATATTGCCCGGGCTGTGCGGATTATGAGCCGCTATACCTTGTGGGAGAGCAAATGTCT
 CGTGATGGCAATTGCCTGTATGAACATGCTGGCGAAGCGCCGGATCGAGAGCACATTATATCTTGGAAACAATGAAAAATGAT
 AAAGGATCCATGAGCGCACATGCCTGGCTCCGGACGGACCGCATGTGGTGACCGGTGCAGAAATTATGGATCGTTACACCG
 TCGTTGGCATTTTTGGAAAAGTGCTGAGGTGA

b.

MGRLRKFLASDRAFKLLLEAFVNLGKARIMKAVPFARTAPKLGVAMKETSFSDSGEETELIRNIARAVRIMSRYTLWESKCLVMA
 IACMNLAKRRIESTLYLGTMTKNDKGSMSAHAWLRTGPHVVTGAEIMDRYTVVGFVKVLR

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Proteína hipotética (<i>Paenibacillus</i> sp. JDR-2)	300	300	100%	6e-102	100%
Proteína do estágio V de esporulação (<i>Bacillus</i> sp. 123MFChir2)	175	175	96%	1e-52	59%
Proteína do estágio V de esporulação Multiespécies (<i>Bacillus</i>)	172	172	97%	2e-51	57%
Proteína hipotética (<i>Paenisporosarcina</i> sp. TG-14)	171	171	96%	3e-51	54%
Proteína do estágio V de esporulação (<i>Bacillus cereus</i>)	170	170	97%	6e-51	57%

APÊNDICE E - Identificação *in silico* da sequência 1489 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 1489 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

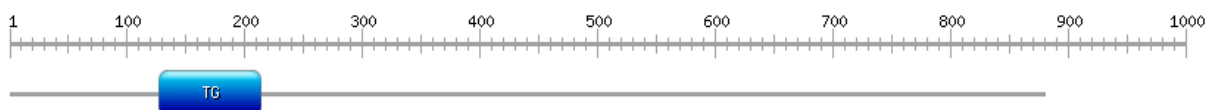
a.

ATGACGAATTGGACTTCGCCGTTTTTCGCTGAGCCCCGGCGGAGCGGGATTTGATCGAGAGCAAATTCCTTAGAAAACGCCAGC
 TGGCGCAGGCCAGAGAACACGAGCTGTTTCGATGTGTTCCGGGAGGAAGACTGACGGAGGAAGAGAGCTGGGCGCTCAAATATTT
 GTTTGCCATATGCCCGTGAACGATCTCGCCGATTATGACGGCGTCTGTTCTTGAACCATGTGCGGAGGACGCTGGAGATCC
 GCCGGCTTGTCCCTTGGGGCGACCGGGTGCCTGTTCTTGCATTATGTCCTCCCTACCGGGTCAATACGGAAAAT
 ATCGAGGATCACCGCGGCGTTTTGTTCAATGAGCTGGCCTATCGCACAAAGTCGCTGACGATGGAAGAGGCTATTCTGGAGA
 CGAATTATTGGTGTACAGAGAAGGCTGCCTATATCGGCAGCGATTTGAGAACGATGTCGCCGCTTACGATGATCAAGAGCGC
 AAGGGCCGTCGGCGAGGAGTGCACGTTAGCGGTTGCGCGCAGCATTGGCATTCCCGCCGTCAGGTATACACG
 CCCAGGTGGGGCATTGCGATGATAATCATGCTGGGTAGAAAGCTGGGGCGATGGGAAATGGTCTTTCTCGGCGCTGCG
 AGCCGGAAGCGCGGCTCGACCAAGGGTGGTTCGGTCCGCGGCCCGCGGGCGATGCTTCTGAATACGCGCATCTGGCCGA
 CTATCCCGTCCGGAAGAGATTACGGTAGCCAATGAATGGTTCACGGAAATTAATCTGCTCGAAAACCTACGCCTCGACGCGT
 ACCGTTGACATTCTGGTTAAAGACATTGCCGAAAACCTGTCCCGGGCGCCGAAGTGCCTTCGAGTTGTACAATATGGCGGA
 ATTTTATCCGATTGCGACGCTGCTCTCGGACGACCGGGGAGAAATCAGCTTCAAGACCGGATATGGCGATCTGCTGGTTCCG
 GCGCTTGGAGCGGTCGATGGGGCGAGGTTAAGCTGGCTGCCTATGGGAGCAACAGGCTCGAGCTTGTCTGGATCGGACGAA
 GCAGCCGGAAGGGATGTTTCGACCTCGATATGGTTCCTCCCGCGGAAAGCGAAGGCGAACCGGCAGAACTCTGTCCGGCGAG
 CGGGAGAGCAGGCATGCTCACCGGCTTGAAGAAGGGGCGAGGCTTCGTGCCGCTATGAAAGCACGTTTCGCAACCAAGCAA
 GATGAGACGACATTGGCTGCTTCGCTAGGGCTTCCGCGGAAAGAGTGTGGAACATTCTAAGCCGCTCGCGCGGAAACTTCC
 GCGAACTCTCGTCTTCTGCAAGAGCGGACGGCCGAATATGGACAGTGGCCGCTGCGGCTTTTGGAGAGCCTTCGCGAGAA
 AGATTGGATCGATGCGCGAAAGGATACGCTGGACGATCATCTGATCGGTTTCGCTTGCCTGCGGCGACCTGCCCGATAAC
 GAGTTCTGTTTCTTACGTGCTTTGTCGCGCGTGTCTAACGAGATGATCGTGCCGTACCGTCAAGTTATCCAAAACGGCGTTTACG
 GAGCCGGAGCGGAAACCTTCCGGTCCGACCCCGTGTGCGCTGTTGCGTTGGCTGAACGACAGACATGGTGAGTGGGAGGACC
 TTCCGAATCTGAAGGGGAAAGGAAATCCGGTCCGAAACGTTTAAAGCTGCAATGGGGAGACCAGGACTCGCTGAACATTTCTT
 CGTAGCCGCTGTCGCAAGTTTAAAGCATCCCGGCGCGGCTGCATCCGCATGAGCTAAACCGCAATACAAGGCGGGCAGCGAA
 TGGCGGGACGCGATGCAAGGCGCGAATTCCTCAGTCTTCGATTGGGGGCAAGGGCAGCGTTTGGCTGCAACGCGATCCCG
 AAGCAACGGAGAATGCGCCAGAGCCGTCGATGCCGAGAATTTTACCTGGCTCGACTGGAGGAAACCGGATTCTACAAAAC
 GTTAATCTATCCGGATAAAATCCGGTGACGCTATGACAAGATGCTAGAGGTAGAGCCGGGTTTCGTAACCGGCTTACGACGGG
 ATGCGGCTTACGACGGTACGGTATTGGCTAGGCTTTCCTATTTCCGCGTACATGCAGGGCAGAAATACGGATGTATCGCTTAC
 CTATCGTGAAGCCTCCCGGGATCTGCCTGTGCTAGGCGTTATGGACCGGGCCAGACCCTCGAAGGGCTTGACGGTTCTGCTC
 TGACGATGGGCGAACAGATCATGGAGGATGGTGCNAATCGTGCCTGGATCGAGCCGGAGCGTGAGCCTACGAAGCATTGCTC
 TCCGAGTTGGGCGAGTTGGCTGTTTCTGCGGCAATCTCGGGGAGCAGTCCCTGCTCATGATCGGCGATGCGGAATGGACCT
 CGTCTTTGATCCCGCACTATATCCGCAATTCGCGTACAGAGGTCAGATTTCGCGAGGGATCGCTCACCGGCTCTCGTCCGCTCA
 GCGGCGGTCCGCGAGGCGGGCTACCCGCATCTGTTGTGCTGGATGGCAGAGACCGTATTCGCTATGCCGCATCGGGCTATAA
 AATCGGAATGGGTAAGAGGCATGGCAAGCGCTAACTGCCGCAAGGCAATCCAACGAATCATCGCCAACGAAAGGGAGCGA
 ATCTTAA

b.

MTNWTSPFSLSPAERDLIESKFLRKRQLAQAREHELFDVFREELTEESWALKYLFAYMPVNDLADYDGVFLNHNRRRLEIRRLV
 PWGDRVPDSLFLHYVLPYRVNTENIEDHRGVLFNELAYRTKSLTMEEAILETNYWCHEKAAAYIGSDLRTMSPLTMIKSARGRCGEE
 STLAVAAALRSIGIPARQVYTPRWAHCDNDHAWVEAWADGKWCFLGACEPEARLDQWFGPPARRAMLLNTRILADYPGPEEITV
 ANEWFTENLLENYASTRTVDILVKDIAGNPVPGAIEVRFELYNMAEFYPIATLLSDDRGEISFKTYGDDLVRGAWSGRWGEVKLA
 AYGSRNLELVLDRTKQPEGMFDLDMVPPPESEGEPAESLSGERESRHAHRLEEGARLRAVYESTFATKQDAETLAASLGLPAERV
 WNILSRSRGNFRELSSFLQERTAEYQWPLRLLLESLREKDWIDARKDTLDDHLIGSLAACGDLDPNEFVSYVLCPRVLNEMIVPYR
 QLFQTAFTEPERETFRSDPSALLRWLNDRHGEWEDLPNLKKGKGNPVGTFKLQWGDQDSLNLIFVAVCRSLGIPARLHPHELNPQYK
 AGSEWRDAMQGANSPQSSIGGKGSVWLQRDPEATENAPEPSYAENFTLARLEENGFYKTLIYDKFGDVYDKMLEVEPGSYRLTT
 GMRLQDGTVLARLSYFRVHAGENTDVSLTYREASRDLPVLGVMDRARPLEGLDGSALTMGEQIMEDGAIVAWIEPEREPTKHLR
 ELGELAVSAANLGGAVLLMIGDAEWTSFDPALYPQLPSEVRFARDRSPVSLPSAAVREAGYPHFLVLDGRDRIRYAASGYKIGMG
 KEAWQALTAARQSNESPTKGS

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	1804	1804	100%	0.0	100%
Proteína hipotética JCM16418_2580 (<i>Paenibacillus pini</i> JCM 16418)	1152	1152	98%	0.0	64%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus</i> sp. FSL R5-808)	1112	1112	99%	0.0	63%
Transglutaminase (<i>Paenibacillus harenae</i>)	1101	1101	98%	0.0	63%
Transglutaminase (<i>Paenibacillus vortex</i>)	1099	1099	99%	0.0	62%

APÊNDICE F - Identificação *in silico* da sequência 3259 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 3259 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

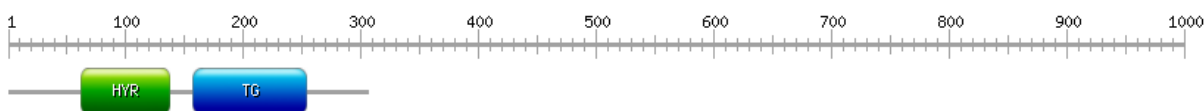
a.

ATGCGCACTAACGAACTAGACGCGCTAGACGTAACAAGGCCATCTTAGGATTTTCTTCATTTTTATTGCCGCATTGGTTATA
TGGACCTCGATCGACAGCATATCGGCGGATACCAATTCCGCAGCCGAGCCCACTGCCGCACCGACTGCCACGTCGGCTGCCA
CGCCGGACACCAAGGCTGATACGGAGCCGCCAAAATTACGGGAGTCAAAAATTAACCGTCTTATCGGAGACCGCGTTTC
CTACAAAACCGGCGTAACCGTTAAGGATAATCACGATACGAACGTCGCTTTGCTTGTAGACACGGGTGGCGTCAACCTAAAG
AAAGCCGGCGTTTACAACATTACGTATATGGCCAAGGATGCCGCAGGCAACGTTGCCCAAGCGGAGGCAGCCGTAACGGTTA
AGGTAAGCCTGCGAGCCCCCTCAACGATGACGAACTGAATAAACTGGCGGATACCATACTTGCCAAGATTGTGACGAAAAGA
CATGACGGATGTCGACAAATTATGGGCAATCTTCAAATGGACGAAATACAACATCGGCTATGTCGACAGTCCAGCCACACC
GATTGGCGGAAAGGCGCGCTTCAAGGGTTTAAACAAGGTACGGGCGACTGCTTCGTCTACTATGCCACCGCTAGAATGCTGA
TGGAGAGAAGCGGCTTCGAAACGCAAACCGTGATCCGGGACAAAGGCAGAAAAACGAAACATTACTGGAGCCTAGTCAAAG
TCGAAGGCGCCTGGTATCACTTCGATACCACTCCCACGATCAAAAAGAATACGACTTTCCTTCTAACGGACGCAGAACTGGCG
TTATTCAATAAGCAATTAAGAACTACTACCTATTCGATAAAAACCAACATCCGGCAACGCCGACCAACCGCTCAAAGAAA
GAAGCAATTATCTTAAATAA

b.

MRTKRTRRRARRNKAILGFFFIIFIAALVIWTSIDSISADTNSAAEPTAAPTATSAATPDTKADTEPPKITGVKNLTVFIGDRVSYKTGVT
VKDNHDTNVALLDVTGGVNLKKAGVYNITYMAKDAAGNVAQAEAAVTVKVKPASPLNDDELNKLADTILAKIVTKDMTDVVK
LWAIFKWTKYNIGYVDSSTHTDWRKQALQGFQGTGDCFVYYATARMLMERSGFETQTVIRDKGRKTKHYWSLVKVEGAWYHF
DTPPTIKNNTFLLDLDAELALFNKQFKNYLFDKTKHPATPTKPLKERSNYLK

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	628	628	100%	0.0	100%
Proteína da família das transglutaminases (<i>Eubacterium rectale</i> CAG36)	222	222	80%	5e-65	41%
Proteína contendo domínio de transglutaminase (<i>Firmicutes bacterium</i> CAG424))	221	221	76%	5e-64	44%
Proteína da família das transglutaminases (<i>Clostridium stercorarium</i>)	204	204	77%	2e-57	43%
Proteína hipotética (<i>Catonella morbi</i>)	1800	1800	78%	1e-51	40%

APÊNDICE G - Identificação *in silico* da sequência 4024 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 4024 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

a.

```

ATGAGCCTCTTGCCGCGTTTGTACTTTCTGCCGTACGGTGATCATCAAGTTACGCCGAAGTGGTATATGAAGCTCTCGCAA
TTATTTATTGCTCTTATGCTTATTCATACGATCCGGATCTTCGAGGACTTCTGGTGGAGTGAGACCTTCTACGCGGCCAGCTTC
ACGCTGCTCTTCGCTGTTGCTGCCGACTGGCTGCTGCCGCGAAGCATGCGGGTGTGCGGATTGTTGTTTCAGCTATTGCTTGCG
GTAGGCGCGACGGCATATTTTGCCGATTTGAAAAAAGTCGTCTTAACCGTTTCCGATCAAGAGAACC GGCTGGGCGATTGTT
CCGTACGTTGTGGGGGTATGCCGAGCAGCTGTATCCTTTTATATGGATCAGCGGTGCTCTGCTCGTTTTGTATATCTTGTGTTGA
TATGTGGATTACAACAAGGATGCGATTATTCGGATTTATAAGCGTCAATATTCTCGTTTGCACCATGCGGATTCTTTTACTCC
AATCTGGCTGTGGGATGAGGTGCGCTGGATTGTCTTCTGGGCTGCTCTGGCTGGTAGCCAACCATCTGCATAAAGTGGCAA
GCCATCATCCGAACAGCTGGAAGGAGCTTACGGAATATCCGCTTCAGTTGCTGATGCCCATCGTTATTGTTTTGTCCGTGCTG
ATGACAGCCGGCATTCTGATGCCGACCGTATCGCCGATTATCAAGATCCTTATACGATGTGGAAGGAATCCAAGGGCGAGA
CGGTTAACGTATTCCCTTGAGAGACAAAGGAATCGCGGCTCCCGTTGACGTGAGCAGCAGCAATACCAAGTCCGGCTATGGACG
CGATGATTCGGAGCTTGGCGCGGTTTTGATTTGACTATTGCCGGTTATGACGATTACAACCAACCATCGGAGCTACTGGA
GAGGCGAGACGAAGACGGAGTACACCGGAGCGGGCTGGACGGTGAACGACGAGACGGAGCGGAACCGGTTAGCCTCGGTG
GGGAAAGACTGGGAGCTTCCGCTGATCGAAGGACGCGAGAAGGCCGAGACGATGAAGGTGGACCAACAGGTCCGGATGTGCG
AACATAATATTTATCCGGTTTTGTTCCCGCGTCTCCAATTACAGCTGTTTCATTGGGTGGATGATCTGATGCTTCCGTATTG
CCGAAAAGCTTAAAGTGGGCGGTTAACGAGCAGGAGTTGATCTGGTCTGACAAGCCGCTGCGCAGATTCCCGCAAAGCTATG
CGGTAACCTCGGAAGTAACCATATTAGATGAAGAAGACTTCGCAAGACAGAAGCCGAATTGCCGGATCCGGCCGAGACGA
AGCAATATTTGAGCTTCCGCTTCTTCCGCGGCGCGTAAGGTGCGTTGCCGTATCGTAACGGCTTCGGCAGTGAATGAT
TACGATAAGGCCAAGCTGATTGAACTTATCTCCGCCAGAACTTCATGTATACGAATGAACCCGACAAATCAAGCTGACCG
GAGACAGCAAGGATTCGTCGACCAGTTCCTCTTTGAACTGAAGGAAGGCTATTGCGATTATTTCTCGACGGCTATGGCGGTA
ATGTCCCGTTCTATCGGATTGCCTACCCGCTGGGTAAGGCTTCGCTCCTGGCGTACTCCCTGCCTCGGAATATGTGCCGTCT
GCCCGCGGAATTTCAAGCGAAGCGGAATTAATCCGGACGGAGCTGGCACGTATACCGTCCGCAACTCGGATGCGCATTCCT
GGTAGAGATTTATTCGACGGGTATGGCTGGATTCCGTTTCGAGCTACGTCAGGCTTTACCTCCCTACTCGGTGACGGAG
GTGCAGCAGGAAGCATTGCCGACGCCTACCGCAGCGGTTGATACCGATCCGGTACAGGAGGAAGCGACCGAAAAGCACGGGA
GGCAACCATTCCGCTATCATCTGGGCGAGTGTATCGGCAGCCGTTCTTGTGCTGCTTGTGCTTGTCTTGTCTGTGCTACT
CCATGGATCCGATCCAGATACCGGGCTTATCCGGCAATGACCGGATTGTATGGGAGACGGAGAAGCTGCTGCGGCTTTGCC
GGCGCAAAGGTTATGAACGCGGGGAGCATGAGACGCTTCGCGAAGCGATTACAGAGATGGTCGCCAAGCCACAAACGGCTGA
ACAATCTTTCCACGAGGTGTGGGCGAGTTCGAACGCGCGAAATACAGTTCCTCAGGCAACAAGCTTGGAGACCGACCC
GTTTATTCTAAAGATCAAGCGATTAAAGAGGAATTGTAG

```

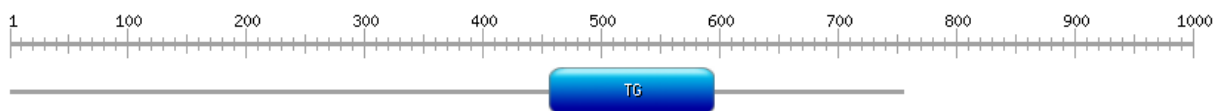
b.

```

MSLLPRLFTFCRTVIIVKTPNWMKLSQLFIALMLIHTIRIFEDFWWSETFYAASFLLFAVAADWLLPRSMRVLRIIVVQLLLAVGAT
AYFADLKKVVLTVSDQENRDLGDLFRTLWGYAEQLYPIWISGALLVLYILFDMWITRMLRFGFISVNILVLTIA DSFTPIWLWDEV
AWIVFLGLLWLVANHLHKLASHHPNSWKELTEYPLQLLMPDIVLSVLMTAGILMPTVSPIIQDPYTMWKESKGETVNVFLGDKGI
AAPVDVSSSNTKSGYGRDSELGGGDFDYSVPMITITNHRYSWRGETKTEYTGAGWTVNDETERNRLASVKGDWELPLIEGREK
AETMKVDQQRMSNNNIYPVLFASPITAVHWVDDPDASVLPKSLKWAVNEQELIWSDKPLRRFPQSYAVTSEVTILDEEGLRKT
EAE LDPDAETKQYLQLPSSLPGRVRS L AVLTA S AVNDYDKAKLIETYLRQNFMYTNEPDKSRLTGDSKDFVDQFLFELKEGYCDY
FSTAMAVMSRSIGLPTRVWKGFAPGVLPASEYVPSAGGIPSEAE LNP DGAGTYTVRNSDAH SWVEIYFDGYGWIPFEPTS GFTFPYS
VTEVQQEALPTPTAAVDTPVQEEATESTGGNHS AIIWGSVSA AVL VLV LVL VLLRTPWIRSR YRAYSGNDRIVWETEKLLRLCR
RKG YERGEHETLREAIQRWSPSHKRLNNLFHEVLGEFERAKYSSVQATSLETDRFILKIKAIKEEL

```

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	1548	1548	100%	0.0	100%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus curdlanolyticus</i>)	591	591	98%	0.0	42%
Transglutaminase (<i>Thermobacillus composti</i>)	575	575	97%	0.0	42%
Transglutaminase (<i>Paenibacillus</i> sp.Br)	564	564	98%	0.0	42%
Transglutaminase (<i>Paenibacillus</i> sp.Br)	564	564	98%	0.0	42%

APÊNDICE H - Identificação *in silico* da sequência 4765 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 4024 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

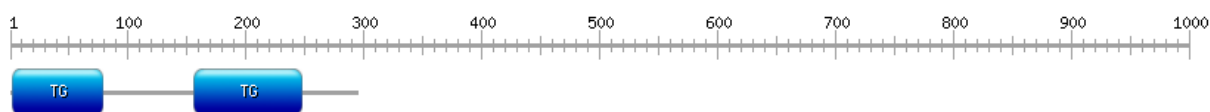
a.

ATGAAGCTAAATATTTCCACGTTACGCAATACGATTACGAGGAGTCGGCGACGGACAGTGTCAATGAGATCCGGCTTACGC
CAAGCACAAATGAACGTCAGTCCCTGCTATCAGCAGGCTATCTTCATAGAGCCGAATGCCTCGCTCTCAGCTATGAGGATTTT
TTCGGCAACCGGGTGCATTCCTTCTCGGTGAACCCGTCGCATAAGCGCTTGACGATCCGAACGCAGATGACGGTTGTAACAAG
GCATGCGCCTAGTCCGGAGGAGCAGGCTGCTCTATTGAACGGCAAGCTTGCAAGGGAGGCGTGGACATGGCTGGAGACGGA
AGAAGCGGCTAACCGTTTCGCAGAGTTTCTGTGCTGCCGACCAGTTATGCGGATGTAACGGCTGATGTAGAGCGTTATGCGGAG
AGCGTTGACCGGAGGGAAGGCGAAGGGATAAGCGTATTGAGCTGGCTCATTTCTTGTCTAGCAAGATCAGAACC GAATTTA
TTTATGATCCGGAAGCTACGACGGTTACAACGAAAGCCAGCGATCTGATCGAGCGCAAACGCGGAGTATGCCAGGATTTTCGC
CCATCTCATGATTGCCTGCTGCCGGTTTCATCAGATTCTGCCCGTTACGTAAGCGGCTATCATTGTAGGGGATCTTCAGGG
CGGCAGCGCCGATTCGAGCAGGCCTCTCATGCGTGGGTGGAGGCTTATGTGCCTCAGCTTGGTTGGTGCAGCTTTGATCCGA
CGAATGAAGCGCCGTTGGCGAACGGTATGTCAAGCTGGGCGATGGCCGTGACTACAAAGATATTGTGCCGGTAAAAGGGGT
GTACCGCGGAAGCGGACTGCAACAGCTGAAGGTAACGGTAGATGTGCGTAAAATCGAGGATTAA

b.

MKLNISHVTQYDYEESATDSVNEIRLTPSTNERQSCYQQAIFIEPNASLFSYEDFFGNRVHSFSVNP SHKRLTIRTQMTVVTRHAPSPE
EQAALLNGKLAREAWTWLET EEAANRFAEFLPTS YADVTADVERYAESVDAREGEGISVLSWLISLSSKIRTEFIYDPEATTVTK
ASDLIERKRGVCQDFAHLMIACCR FHQIPARYVSGYHFVGD LQGG SADF EQASHAWVEAYVPQLGWCSFDPTNEAPVGER YVKL
GHGRDYKDIVPVKGVYRGSGLQQLKVTVDVRKIED

c.



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	618	618	100%	0.0	100%
Transglutaminase (<i>Paenibacillus ginsengihumi</i>)	462	462	100%	5e-161	75%
Transglutaminase (<i>Paenibacillus harenae</i>)	458	458	100%	2e-159	74%
Transglutaminase (<i>Paenibacillus curdlanolyticus</i>)	407	407	100%	5e-139	64%
Proteína hipotética (<i>Paenibacillus daejeonensis</i>)	363	363	99%	4e-122	62%

APÊNDICE I - Identificação *in silico* da sequência 4870 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 4870 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

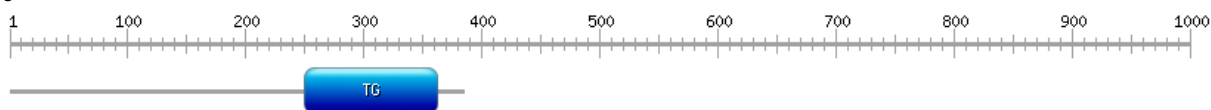
a

ATGATGGAATGGATTTCGCGCTGGGACGGTATGGGAGCCGGTCTCGCTTATCGTTCCTCTCTGCTTGTTTCGCTGCTGCAA
 GGAATGAGAAGAGGGCGCTCGGGTTCAGCGAGGCGCCTTTTCTTTTTGTATGGGACGGTATTCTGTTAATCGTTTGTCTTGTG
 CTGCGGGAAAATTGCGGGCAGCTTGTCCTCCATTGGCAGCGGATTGGCTGAGCGGGCATGTAACCTGCCGAGAAGGAAA
 TGGGTGCAACCTCCAGGCTTGGTATACGTTTATAACAAGCCTGCGCGATTTTGCCTGCTCCGGTACGGCATCTGTTTTGC
 TCTCGTATTGCTGCTTCGGCTGGCTGCTTCTTCTTGGCCTCTGGTGAAACGCTGTTTGCCAGCAGACAGCAACGCGGCA
 GACATGAGAGCATGGAGCGTCTTCCGGGAGGAGGGCGGCAAGCCGCGCAGCCGGAGCGGTTTTCCGGCACGGTTCATGGTGC
 CGGACGGGCATTTGTATTTGTAGCCGCTTGTTCATTTATGTCTCGCTGGCGCCAAACGGCCTCTGGTCCGGATTGATTCAGCA
 ATCTCCGTTATACATGAAGGCATCGGATGTTCTTGAACCGGTTGCCGGAGATATGCTGGCCGGACGTGGCCCCGTTCTGGCTG
 AAGCCGTGCAGGAAGAGTTCAGAATGTGCTGCAGCGCAAATACGAGATTATCGATTACAATATCCCCTCGGATATTGAGCA
 GGCAGCCCTGCAGGTTACGAAGGACAAGAATAGCGAAGAAGACAAGGCAAGGGCTTTGTATGACTGGTTAGGCAGCCGGAT
 CGCCTATGACTGGGACAAGGCTGATAATTATGTAGACAACGGAGTATGGAAGGAACAGAACCCGGAAGAGACGTTTGCTACC
 CGAAAAGGCGTATGTATCGACGTTGCTAGGCTGTACGCCGTAATGGCGCGCGCTCAGGCTGAGTGAAGGTCGTTACCG
 GCACAGGACTTGACGGGCGCGGAGGCTCCGGGCCGATGCTGGAACGAAGTGAAGGTAGGCGGAGACGGCGGTACCTGGA
 TTCGCTCGATGCGACATGGGCATCCTCCGGCGATTGGTTCAACCCGGTTGATTTCGACAAGACTCATATCCGGGCAGTATAA

b.

MMEWIRAGTVWEPVSLIVLLLLLVSLLQGMRRGASGSARRLFFVWDGILLIVCLVLAGKIAGSLSPLAADWLSGHVNVQPKEMG
 ATSQA WYTFITSLRDFALLRYGILFLLSYLLRLAASFLAPLVETLFASRQQGRHESMERLPGGGAASRAAGAVFGTVHGAGRAF
 VFVAALFIYVSLAPNGLWSGLIQSPLYMKASDVLEPVAGDMLAGRGPVLAEA VQEEFQNVLRKYEIIDYNIPSDIEQAALQVTK
 DKNSEEDKARALYDWLGSRIAYDWDKADNYVDNGVWKEQNPEETFATRKGVCIDVARLYAVMARASGLDVKVVVTGTGLDGRG
 GSGPHAWNEVKVGGDGGTWIPLDATWASSGDWFNPVDFDKTHIRAV

c



d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	778	778	100%	0.0	100%
Transglutaminase (<i>Paenibacillus harenae</i>)	496	496	99%	2e-171	67%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus</i> sp. P22)	427	427	97%	3e-144	60%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus</i> sp. P22)	427	427	97%	4e-144	60%
Proteína hipotética (<i>Paenibacillus pinihumi</i>)	402	402	99%	1e-134	57%

APÊNDICE J - Identificação *in silico* da sequência 5011 como uma transglutaminase: os dados mostram a sequência nucleotídica da ORF 5011 (a), o peptídeo codificado por ela (b), a predição dos domínios presentes neste peptídeo (c) e resultado da comparação com outras proteínas obtido por BLASTp (d).

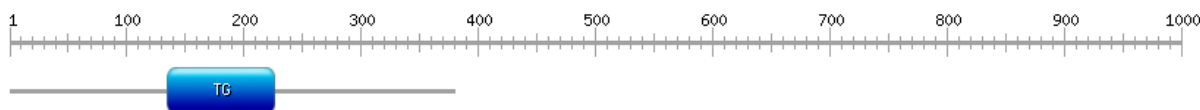
a.

ATGCGAAAAACAGGTTGGAAAAATAGCGACAGTAGCGGCCTTGCTTGCCGTTGCCATTGGTTTAAATATGAAATCCGATCTGTTCTCTGTTTTTGGCTCTGCGTCAAGCGGCGGAGGGCGGTACGATCGATCAGCTTGAGGATGAGCTGGCATCGCAATTTCAAAGCAGATCCGAGCATTTTACGCTTCAATATAAAGGAGACAAGCACAAGCTTTCCGACGGCTTGCAAAATACGATTAATCTGGCGCTTGCGCAGGATGATTACACCGCGTATATCCTTGACTCCTATCTTTATACAATCCGAGCTGGGGTAATCTGTCCACCATAAAGCTGGAGCCAGATACCGGAAACGACGGACCAGACCGCTGCCGTTGATGCAATGGTGAGCAAAGCGATCAAGCAGATTATTACCCCAGATGAATGAGCATGAGAAGGTGAAGGCGATCCACGACTGGATTGTTAATAGACTTCAGTATGACGAGACCCTAAAGCATTACACCGCTTACGCTCTCACACCGGAACCGCTGTATGTCAAGGTTATTCGCTACTTGCTTATAAAATGCTAAAAGCTGTCGGAAATTCCTGTTCTTATTGCCGAGGGCAAAGTTAAGACCCGGCAGCATGCCCTGGAATATGGTGCAAGCTCGGCGGCAATTGGTATCATCTCGATATGACTGGGATGATCCGTTGTGAAGCAAGCCGGCACGCAGCAGACCGAGTCCCCTCAGCTATAATATTATTGAAGACGGATGACGAGATGCGCGTCGATCATCAATGGATCAAAACCTATCCAACGGCGAACCATTCTACGGCAGCGAGCTCAGCAAGCTGAAGAGCACGGACAGCGGAAAGCGGATGTTTACGCTAAGCTCATCACCGATCTTGGCCTGCATTGGACGGAGCCGGAAAAATACGGTAGCAAGCCTGGTAGAGCTGGAGCAGGAGATCCGCGAGGCTGTAATGGAGCAGCAGACGGGCATGAAGTCCGTTATCTGCAAGGTTCCGAGCTTGGCAGCGATCTGCCGACGGCCTTAAGGTAGCGCATGTATCCGTTCCGCTATTCGGCTCGTTACGAGCCATACGGCAATGACGGTTCGATGCTTGTGCAGGTTACGCTGGATTATCAATAA

b.

MRKTGWKIATVAALLAVAIGLNMKSDLFSVFASASSGGGGTIDQLEDELASQFQSRSEHFTLQYKGDHKLSDGLQNTINLALAQDDYTAYILDSYLYTIRSWGNLSTIKLEARYRETTDQTAADVAMVSKAIKQIITPQMNEHEKVKAIHDWIVNRLQYDETLKHYTAYDALTTGTAVCQGYSLAYKMLKAVGIPVLIAEGKVKTDHAWNMVQLGNGWYHLDMTWDDPVVKQAGTQQTESRLSYNYLKTDDEMRVDHQWIKTYPTANHSYGSLSKLDSDSGKADVYAKLITDLGLHWTEPENTVASLVELEQEIREAVMEQQTGMKFRYLQGSQATLDLPTAFKVAHVSVRYSARYEPYGNDSMLVQVQLDYQ

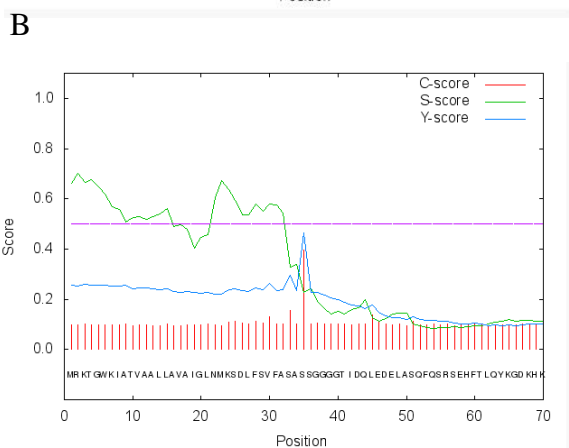
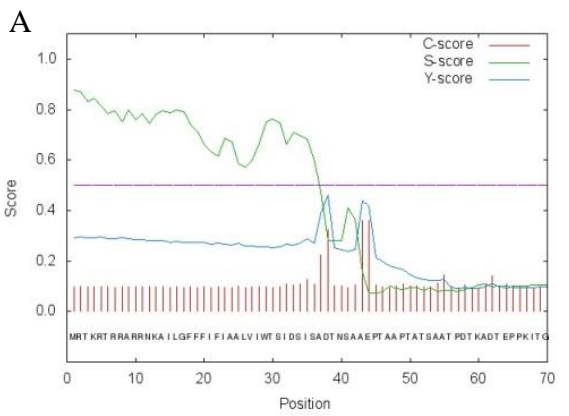
c.



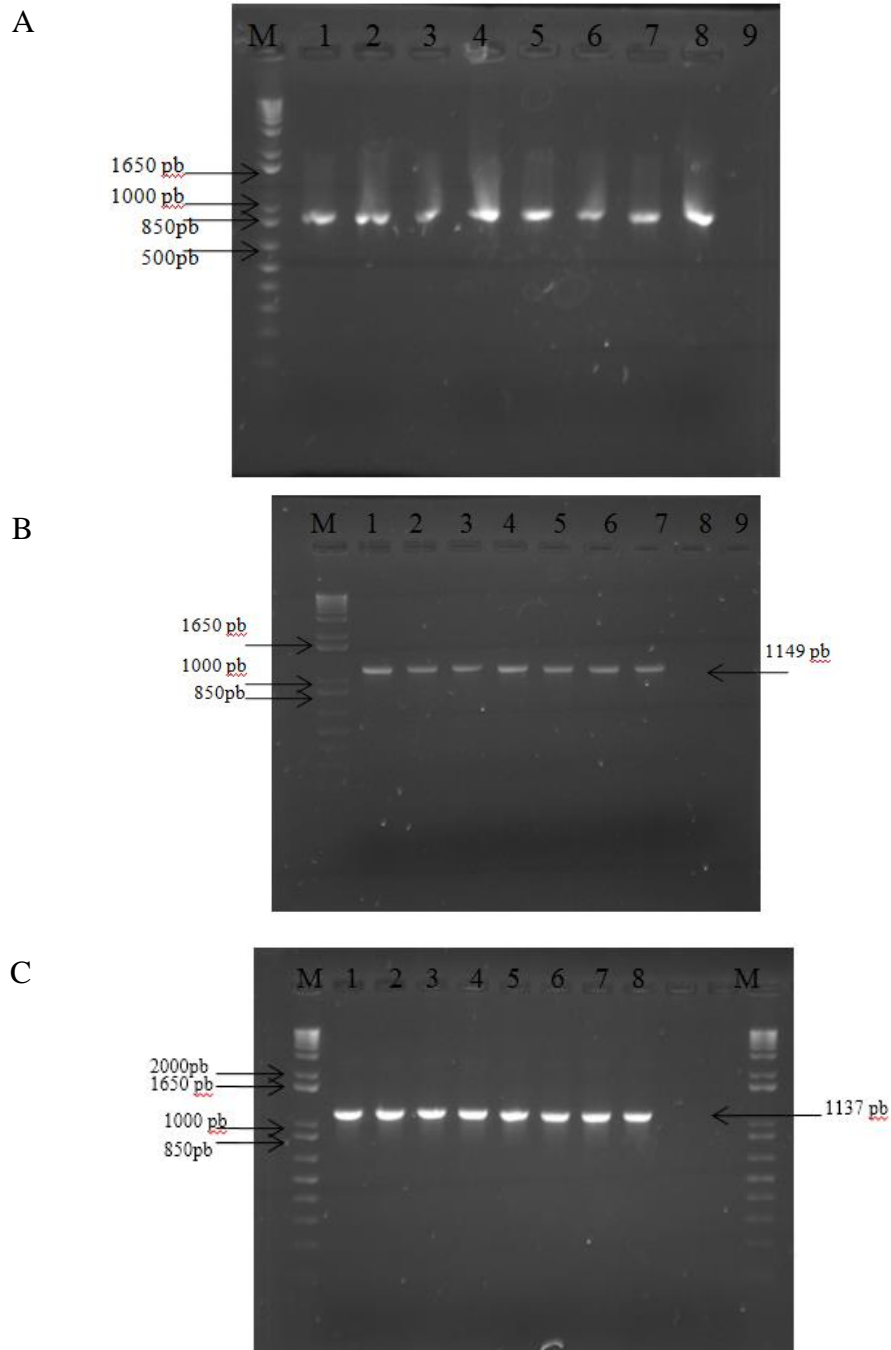
d.

Descrição	Escore máximo	Escore total	Cobertura da sequência	Valor-E	Identidade
Transglutaminase (<i>Paenibacillus</i> sp. JDR-2)	785	785	100%	0.0	100%
Proteína hipotética (<i>Paenibacillus harenae</i>)	417	417	99%	1e-140	51%
Transglutaminase (<i>Thermobacillus composti</i>)	325	325	87%	9e-105	48%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus curdolanolyticus</i>)	275	275	85%	6e-85	43%
Proteína contendo domínio de transglutaminase (<i>Paenibacillus</i> sp. P22)	252	252	60%	3e-77	50%

APÊNDICE K - Predição da presença de peptídeo sinal: predição da presença de peptídeo sinal das proteínas codificadas pelas ORFs 3259 (A) e 5011 (B). Os resultados indicam que a proteína 3259 possui um peptídeo sinal, que é clivado entre os aminoácidos 37 e 38, enquanto a proteína 500 é clivada entre os aminoácidos 34 e 35. *Score-C*, probabilidade de ocorrência da clivagem pela enzima sinal peptidase; *Score-S* corresponde à probabilidade da sequência estar presente em um peptídeo sinal, distinguindo entre este e a enzima madura; *Score-Y* é a média geométrica do score-c e da inclinação do score-y, resultando em uma melhor predição dos sítios de clivagem.



APÊNDICE L - Teste de amplificação das sequências 4765, 4870 e 5011: verificação da melhor temperatura de amplificação das ORFs 4765 (A), 4870 e 5011. As temperaturas testadas foram 60°C (1), 59,1°C (2), 57,8°C (3), 55,8°C (4) 53,2°C (5), 51,1°C (6), 49,7°C (7), 49°C (8). A canaleta 9 corresponde ao controle negativo e M ao marcador de peso molecular. Os três pares de *primers* amplificaram bem em todas as temperaturas, com exceção dos *primers* usados para a sequência 4870, que não amplificou em 49°C.



APÊNDICE M – Comparação entre as proteínas de *Paenibacillus* sp. JDR-2 clonadas e parentais

Tabela 5: Comparação entre a cadeia polipeptídica codificada pela ORF 4765 e a cadeia sequência expressa pelo vetor pETG4765

Proteína_4765_clonada	MASKLNISHVTQDYDEESATDSVNEIRLTPSTNERQSCYQQAIFIEPNASLFSYEDFFGN
Proteína_4765_parental	M--KLNISHVTQDYDEESATDSVNEIRLTPSTNERQSCYQQAIFIEPNASLFSYEDFFGN * *****
Proteína_4765_clonada	RVHSFSVNPESHKRLTIRTQMTVVTRHAPSPEEQAALLNGKLAREAWTWLETEEAANRFAE
Proteína_4765_parental	RVHSFSVNPESHKRLTIRTQMTVVTRHAPSPEEQAALLNGKLAREAWTWLETEEAANRFAE *****
Proteína_4765_clonada	FLLPTSADVTADVERYAESVDAREGEGISVLSWLISLSSKIRTEFIYDPEATVTTKAS
Proteína_4765_parental	FLLPTSADVTADVERYAESVDAREGEGISVLSWLISLSSKIRTEFIYDPEATVTTKAS *****
Proteína_4765_clonada	DLIERKRGVCQDFAHLMIAACRFHQIPARYVSGYHFVGDLOGGSADFEQASHAWVEAYVP
Proteína_4765_parental	DLIERKRGVCQDFAHLMIAACRFHQIPARYVSGYHFVGDLOGGSADFEQASHAWVEAYVP *****
Proteína_4765_clonada	QLGWCSFDPTNEAPVGERYVKLGHGRDYKDIPVKGVYRGSGLQQLKVTVDVRKIEDLEH
Proteína_4765_parental	QLGWCSFDPTNEAPVGERYVKLGHGRDYKDIPVKGVYRGSGLQQLKVTVDVRKIED--- *****
Proteína_4765_clonada	HHHHH*
Proteína_4765_parental	-----

Tabela 6: Comparação entre a cadeia polipeptídica codificada pela ORF 4870 e a cadeia sequência expressa pelo vetor pETG4870

Proteína_4870_clonada	MASEWIRAGTVWEPVSLIVLLLLLVSLIQMRRGASGSARRLFFVWDGILLIVCLVLAG
Proteína_4870_parental	M-MEWIRAGTVWEPVSLIVLLLLLVSLIQMRRGASGSARRLFFVWDGILLIVCLVLAG * *****
Proteína_4870_clonada	KIAGSLSPLAADWLSGHVNVPOKEMGATSQAWYTFITSLRDFALLRYGILFLLSYLLLR
Proteína_4870_parental	KIAGSLSPLAADWLSGHVNVPOKEMGATSQAWYTFITSLRDFALLRYGILFLLSYLLLR *****
Proteína_4870_clonada	AASFLAPLVETLFASRQQRGRHESMERLPGGGAASRAAGAVFGTVHGAGRAVFVAALFI
Proteína_4870_parental	AASFLAPLVETLFASRQQRGRHESMERLPGGGAASRAAGAVFGTVHGAGRAVFVAALFI *****
Proteína_4870_clonada	YVSLAPNGLWSGLIQOSPPLYMKASDVLEPVAGDMLAGRGPVLAEAVQEEFQNVLQRYE
Proteína_4870_parental	YVSLAPNGLWSGLIQOSPPLYMKASDVLEPVAGDMLAGRGPVLAEAVQEEFQNVLQRYE *****
Proteína_4870_clonada	IDYNIPSDIEQAALQVTKDKNSEEDKARALYDWLGSRIAYDWDKADNYVDNGVWKEQNPE
Proteína_4870_parental	IDYNIPSDIEQAALQVTKDKNSEEDKARALYDWLGSRIAYDWDKADNYVDNGVWKEQNPE *****
Proteína_4870_clonada	ETFATRKGVCIDVARLYAVMARASGLDVKVVVTGTGLDGRGGSGPHAWNEVKVGGDGGT
Proteína_4870_parental	ETFATRKGVCIDVARLYAVMARASGLDVKVVVTGTGLDGRGGSGPHAWNEVKVGGDGGT *****
Proteína_4870_clonada	PLDATWASSGDWFPVDFDKTHIRAVLEHHHHH*
Proteína_4870_parental	PLDATWASSGDWFPVDFDKTHIRAV----- *****

Tabela 7: Comparação entre a cadeia polipeptídica codificada pela ORF 5011 e a cadeia sequência expressa pelo vetor pETG5011

Proteína_5011_clonada	MASRKTGWKIATVAALLLAVAIGLNMKSDLFSVFASASSGGGGTIDQLEDELASQFQSRSE
Proteína_5011_Parental	M--RKTGWKIATVAALLLAVAIGLNMKSDLFSVFASASSGGGGTIDQLEDELASQFQSRSE * *****
Proteína_5011_clonada	HFTLQYKGDHKHLSLQNTINLALAQDDYTAYILDSYLYTIRSWGNLSTIKLEARYRET
Proteína_5011_Parental	HFTLQYKGDHKHLSLQNTINLALAQDDYTAYILDSYLYTIRSWGNLSTIKLEARYRET *****

Proteina_5011_clonada	TDQTAAVDAMVSKAIKQIITPQMNEHEKVKAIHDWIVNRLQYDETLKHYTAYDALTTGTA
Proteina_5011_Parental	TDQTAAVDAMVSKAIKQIITPQMNEHEKVKAIHDWIVNRLQYDETLKHYTAYDALTTGTA

Proteina_5011_clonada	VCQGYSLLAYKMLKAVGIPVLI AEGKVKTDHAWNMVQLGGNWHLDMTWDDPVVKQAGT
Proteina_5011_Parental	VCQGYSLLAYKMLKAVGIPVLI AEGKVKTDHAWNMVQLGGNWHLDMTWDDPVVKQAGT

Proteina_5011_clonada	QQTESRLSYNYLKTDDMRVDHQWIKTYPTANHSYGSLSKSTDSGKADVYAKLITD
Proteina_5011_Parental	QQTESRLSYNYLKTDDMRVDHQWIKTYPTANHSYGSLSKSTDSGKADVYAKLITD

Proteina_5011_clonada	LGLHWTEPENTVASLVELEQEIREAVMEQQTGMKFRYLQGSQLATDLP TAFKVAHVSRY
Proteina_5011_Parental	LGLHWTEPENTVASLVELEQEIREAVMEQQTGMKFRYLQGSQLATDLP TAFKVAHVSRY

Proteina_5011_clonada	SARYEPYGNDGSMLVQVQLDYQLEHHHHH*
Proteina_5011_Parental	SARYEPYGNDGSMLVQVQLDYQ-----

APÊNDICE N – Resultado do sequenciamento dos plasmídeos construídos

Tabela 8: Comparação entre a sequência da ORF 4870 clonada o seu sequenciamento a fim de verificar a sequência clonada no vetor de expressão pET23d+

Query: Sequence_4870 Length: 1185 Subject: sequencing			
Score: 2189 bits (1185), Expect: 0.0, Identities: 1185/1185(100%), Gaps: 0/1185(0%), Strand: Plus/Plus			
Query	1	ATGGCTAGCGAATGGATTTCGCGCTGGGACGGTATGGGAGCCGGTCTCGCTTATCGTTCTT	60
Sbjct	1	ATGGCTAGCGAATGGATTTCGCGCTGGGACGGTATGGGAGCCGGTCTCGCTTATCGTTCTT	60
Query	61	CTCCTTCTGCTTGTTTCGCTGCTGCAAGGAATGAGAAGAGGCGCGTTCGGGTTACGCGAGG	120
Sbjct	61	CTCCTTCTGCTTGTTTCGCTGCTGCAAGGAATGAGAAGAGGCGCGTTCGGGTTACGCGAGG	120
Query	121	CGCCTTTTCTTTTTTGTATGGGACGGTATTCTGTTAATCGTTTGTCTTGTGCTTGCCGGG	180
Sbjct	121	CGCCTTTTCTTTTTTGTATGGGACGGTATTCTGTTAATCGTTTGTCTTGTGCTTGCCGGG	180
Query	181	AAAATTGCGGGCAGCTTGTCCCATTTGGCAGCGGATTGGCTGAGCGGGCATGTAAACGTG	240
Sbjct	181	AAAATTGCGGGCAGCTTGTCCCATTTGGCAGCGGATTGGCTGAGCGGGCATGTAAACGTG	240
Query	241	CCGCAGAAGGAAATGGGTGCAACCTCCAGGCTTGGTATACGTTTATAACAGCCGCGC	300
Sbjct	241	CCGCAGAAGGAAATGGGTGCAACCTCCAGGCTTGGTATACGTTTATAACAGCCGCGC	300
Query	301	GATTTTGCCTTGCTCCGGTACGGCATCCTGTTTTGCTCTCGTATTTGCTGCTTCGGCTG	360
Sbjct	301	GATTTTGCCTTGCTCCGGTACGGCATCCTGTTTTGCTCTCGTATTTGCTGCTTCGGCTG	360
Query	361	GCTGCTTCCTTCTTTCGCGCTCTGGTGGAAACGCTGTTTGCAGCAGACAGCAACGCGGC	420
Sbjct	361	GCTGCTTCCTTCTTTCGCGCTCTGGTGGAAACGCTGTTTGCAGCAGACAGCAACGCGGC	420
Query	421	AGACATGAGAGCATGGAGCGTCTTCCGGGAGGAGGGCGCAAGCCGCGCAGCCGAGCG	480
Sbjct	421	AGACATGAGAGCATGGAGCGTCTTCCGGGAGGAGGGCGCAAGCCGCGCAGCCGAGCG	480
Query	481	GTTTTCGGCACGGTTCATGGTGCCGGACGGGCATTTGTATTTGTAGCCGCTTTGTTTATT	540
Sbjct	481	GTTTTCGGCACGGTTCATGGTGCCGGACGGGCATTTGTATTTGTAGCCGCTTTGTTTATT	540
Query	541	TATGTCGCTGGCGCCAACGGCCTCTGGTCCGGATTGATTCAGCAATCTCCGTTATAC	600
Sbjct	541	TATGTCGCTGGCGCCAACGGCCTCTGGTCCGGATTGATTCAGCAATCTCCGTTATAC	600
Query	601	ATGAAGGCATCGGATGTTCTTGAACCGGTTGCCGGAGATATGCTGGCCGGACGTGGCCCC	660
Sbjct	601	ATGAAGGCATCGGATGTTCTTGAACCGGTTGCCGGAGATATGCTGGCCGGACGTGGCCCC	660
Query	661	GTTCTGGCTGAAGCCGTGCAGGAAGAGTCCAGAATGTGCTGCAGCGCAAATACGAGATT	720
Sbjct	661	GTTCTGGCTGAAGCCGTGCAGGAAGAGTCCAGAATGTGCTGCAGCGCAAATACGAGATT	720
Query	721	ATCGATTACAATATCCCGTCGGATATTGAGCAGGCAGCCCTGCAGGTACGAAGGACAAG	780
Sbjct	721	ATCGATTACAATATCCCGTCGGATATTGAGCAGGCAGCCCTGCAGGTACGAAGGACAAG	780
Query	781	AATAGCGAAGAAGACAAGGCAAGGGCTTTGTATGACTGGTTAGGCAGCCGGATCGCCTAT	840
Sbjct	781	AATAGCGAAGAAGACAAGGCAAGGGCTTTGTATGACTGGTTAGGCAGCCGGATCGCCTAT	840
Query	841	GACTGGGACAAGGCTGATAATTATGTAGACAACGGAGTATGGAAGGAACAGAACC GGAA	900
Sbjct	841	GACTGGGACAAGGCTGATAATTATGTAGACAACGGAGTATGGAAGGAACAGAACC GGAA	900
Query	901	GAGACGTTTGTACCCGAAAAGGCGTATGTATCGACGTTGCTAGGCTGTACGCCGTAATG	960
Sbjct	901	GAGACGTTTGTACCCGAAAAGGCGTATGTATCGACGTTGCTAGGCTGTACGCCGTAATG	960
Query	961	GCGCGCGCCTCAGGTCTGGATGTGAAGGTCGTTACCGGCACAGGACTTGACGGGCGCGGA	1020
Sbjct	961	GCGCGCGCCTCAGGTCTGGATGTGAAGGTCGTTACCGGCACAGGACTTGACGGGCGCGGA	1020

Query	1021	GGCTCCGGGCCGCATGCCTGGAACGAAGTGAAGGTAGGCGGAGACGGCGGTACCTGGATT	1080
Sbjct	1021	GGCTCCGGGCCGCATGCCTGGAACGAAGTGAAGGTAGGCGGAGACGGCGGTACCTGGATT	1080
Query	1081	CCGCTCGATGCGACATGGGCATCCTCCGGCGATTGGTTCAACCCGGTTGATTTGACAAG	1140
Sbjct	1081	CCGCTCGATGCGACATGGGCATCCTCCGGCGATTGGTTCAACCCGGTTGATTTGACAAG	1140
Query	1141	ACTCATATCCGGGCAGTACTCGAGcaccaccaccaccaccTGA	1185
Sbjct	1141	ACTCATATCCGGGCAGTACTCGAGcaccaccaccaccaccTGA	1185

Tabela 9: Comparação entre a sequência da ORF 4765 clonada o seu sequenciamento a fim de verificar a sequência clonada no vetor de expressão pET23d+

Query: Sequence_4765 Length: 918 Subject: sequencing

Score: 1696 bits(918), Expect: 0.0, Identities: 918/918(100%), Gaps: 0/918 (0%), Strand: Plus/Plus

Query	1	ATGGCTAGCAAGCTAAATATTTCCACGTTACGCAATACGATTACGAGGAGTCGGCGACG	60
Sbjct	1	ATGGCTAGCAAGCTAAATATTTCCACGTTACGCAATACGATTACGAGGAGTCGGCGACG	60
Query	61	GACAGTGTCAATGAGATCCGGCTTACGCCAAGCACAAATGAACGTAGTCTGTATCAG	120
Sbjct	61	GACAGTGTCAATGAGATCCGGCTTACGCCAAGCACAAATGAACGTAGTCTGTATCAG	120
Query	121	CAGGCTATCTTCATAGAGCCGAATGCCTCGCTCTTCAGCTATGAGGATTTCTTCGGCAAC	180
Sbjct	121	CAGGCTATCTTCATAGAGCCGAATGCCTCGCTCTTCAGCTATGAGGATTTCTTCGGCAAC	180
Query	181	CGGGTGCATTCTTCTCGGTGAACCCGTCGCATAAGCGCTTGACGATCCGAACGCAGATG	240
Sbjct	181	CGGGTGCATTCTTCTCGGTGAACCCGTCGCATAAGCGCTTGACGATCCGAACGCAGATG	240
Query	241	ACGGTTGTAACAAGGCATGCGCCTAGTCCGGAGGAGCAGGCTGCTCTATTGAACGGCAAG	300
Sbjct	241	ACGGTTGTAACAAGGCATGCGCCTAGTCCGGAGGAGCAGGCTGCTCTATTGAACGGCAAG	300
Query	301	CTTGCAAGGGAGGCGTGGACATGGCTGGAGACGGAAGAAGCGGCTAACCGTTTCGCAGAG	360
Sbjct	301	CTTGCAAGGGAGGCGTGGACATGGCTGGAGACGGAAGAAGCGGCTAACCGTTTCGCAGAG	360
Query	361	TTTCTGCTGCCGACCAGTTATGCGGATGTAACGGCTGATGTAGAGCGTTATGCCGAGAGC	420
Sbjct	361	TTTCTGCTGCCGACCAGTTATGCGGATGTAACGGCTGATGTAGAGCGTTATGCCGAGAGC	420
Query	421	GTTGACGCGAGGGAAGGCGAAGGGATAAAGCGTATTGAGCTGGCTCATTTCCTTGCTTAGC	480
Sbjct	421	GTTGACGCGAGGGAAGGCGAAGGGATAAAGCGTATTGAGCTGGCTCATTTCCTTGCTTAGC	480
Query	481	AAGATCAGAACCGAATTTATTTATGATCCGGAAGCTACGACGGTTACAACGAAAGCCAGC	540
Sbjct	481	AAGATCAGAACCGAATTTATTTATGATCCGGAAGCTACGACGGTTACAACGAAAGCCAGC	540
Query	541	GATCTGATCGAGCGCAAACGCGGAGTATGCCAGGATTTCCGCCATCTCATGATTGCCTGC	600
Sbjct	541	GATCTGATCGAGCGCAAACGCGGAGTATGCCAGGATTTCCGCCATCTCATGATTGCCTGC	600
Query	601	TGCCGGTTTCATCAGATTCCTGCCCGTTACGTAAGCGGCTATCATTTTGTAGGGATCTT	660
Sbjct	601	TGCCGGTTTCATCAGATTCCTGCCCGTTACGTAAGCGGCTATCATTTTGTAGGGATCTT	660
Query	661	CAGGGCGGCAGCGCCGATTTTCGAGCAGGCCTCTCATGCGTGGGTGGAGGCTTATGTGCCT	720
Sbjct	661	CAGGGCGGCAGCGCCGATTTTCGAGCAGGCCTCTCATGCGTGGGTGGAGGCTTATGTGCCT	720
Query	721	CAGCTTGGTTGGTGCAGCTTTGATCCGACGAATGAAGCGCCGGTTGGCGAACGGTATGTC	780
Sbjct	721	CAGCTTGGTTGGTGCAGCTTTGATCCGACGAATGAAGCGCCGGTTGGCGAACGGTATGTC	780
Query	781	AAGCTGGGGCATGGCCGTGACTACAAAGATATTGTCCCGGTAAAAGGGGTGTACCGCGGA	840
Sbjct	781	AAGCTGGGGCATGGCCGTGACTACAAAGATATTGTCCCGGTAAAAGGGGTGTACCGCGGA	840
Query	841	AGCGGACTGCAACAGCTGAAGGTAACGGTAGATGTGCGTAAAATCGAGGATCTCGAGcacc	900

Sbjct	841		AGCGGACTGCAACAGCTGAAGGTAACGGTAGATGTGCGTAAAATCGAGGATCTCGAGCAC	900
Query	901	caccaccaccaccacTGA	918	
Sbjct	901		CACCACCACCACCTGA	918

Tabela 10: Comparação entre a sequência da ORF 5011 clonada o seu sequenciamento a fim de verificar a sequência clonada no vetor de expressão pET23d+

Query: Sequence_5011 Length: 1173 Subject: sequencing

Score: 2167 bits (1173), Expect: 0.0, Identities: 1173/1173(100%), Gaps: 0/1173 (0%), Strand: Plus/Plus

Query	1	ATGGCTAGCCGAAAAACAGGTTGGAAAATAGCGACAGTAGCGGCCTTGCTTGCCGTTGCC	60
Sbjct	1	ATGGCTAGCCGAAAAACAGGTTGGAAAATAGCGACAGTAGCGGCCTTGCTTGCCGTTGCC	60
Query	61	ATTGGTTTAAATATGAAATCCGATCTGTTCTCTGTTTTTGCCTCTGCGTCAAGCGGCGGA	120
Sbjct	61	ATTGGTTTAAATATGAAATCCGATCTGTTCTCTGTTTTTGCCTCTGCGTCAAGCGGCGGA	120
Query	121	GGCGGTACGATCGATCAGCTTGAGGATGAGCTGGCATCGCAATTTCAAAGCAGATCCGAG	180
Sbjct	121	GGCGGTACGATCGATCAGCTTGAGGATGAGCTGGCATCGCAATTTCAAAGCAGATCCGAG	180
Query	181	CATTTTACGCTTCAATATAAAGGAGACAAGCACAAGCTTCCGACGGCTTGCAAAATACG	240
Sbjct	181	CATTTTACGCTTCAATATAAAGGAGACAAGCACAAGCTTCCGACGGCTTGCAAAATACG	240
Query	241	ATTAATCTGGCGCTTGCGCAGGATGATTACACCGGTATATCCTTGACTCCTATCTTTAT	300
Sbjct	241	ATTAATCTGGCGCTTGCGCAGGATGATTACACCGGTATATCCTTGACTCCTATCTTTAT	300
Query	301	ACAATCCGACAGCTGGGGTAATCTGTCCACCATAAAGCTGGAGGCCAGATACCGCAAACG	360
Sbjct	301	ACAATCCGACAGCTGGGGTAATCTGTCCACCATAAAGCTGGAGGCCAGATACCGCAAACG	360
Query	361	ACGGACCAGACCGCTGCCGTTGATGCAATGGTGAGCAAAGCGATCAAGCAGATTATTACC	420
Sbjct	361	ACGGACCAGACCGCTGCCGTTGATGCAATGGTGAGCAAAGCGATCAAGCAGATTATTACC	420
Query	421	CCGCAGATGAATGAGCATGAGAAGGTGAAGGCGATCCACGACTGGATTGTTAATAGACTT	480
Sbjct	421	CCGCAGATGAATGAGCATGAGAAGGTGAAGGCGATCCACGACTGGATTGTTAATAGACTT	480
Query	481	CAGTATGACGAGACCCTAAAGCATTACACCGCTTACGATGCTCTCACAAACGGGAACCGCT	540
Sbjct	481	CAGTATGACGAGACCCTAAAGCATTACACCGCTTACGATGCTCTCACAAACGGGAACCGCT	540
Query	541	GTATGTCAAGGTTATTTCGCTACTTGCTTATAAAAATGCTAAAAGCTGTTCGGAATTCCTGTT	600
Sbjct	541	GTATGTCAAGGTTATTTCGCTACTTGCTTATAAAAATGCTAAAAGCTGTTCGGAATTCCTGTT	600
Query	601	CTTATTGCCGAGGGCAAAGTTAAGACCGGCGACCATGCCTGGAATATGGTGCAGCTCGGC	660
Sbjct	601	CTTATTGCCGAGGGCAAAGTTAAGACCGGCGACCATGCCTGGAATATGGTGCAGCTCGGC	660
Query	661	GGCAATGGTATCATCTCGATATGACCTGGGATGATCCCCTGTGAAGCAAGCCGGCAGC	720
Sbjct	661	GGCAATGGTATCATCTCGATATGACCTGGGATGATCCCCTGTGAAGCAAGCCGGCAGC	720
Query	721	CAGCAGACCGAGTCCCGTCTCAGCTATAAATTATTATTTGAAGACGGATGACGAGATGCCG	780
Sbjct	721	CAGCAGACCGAGTCCCGTCTCAGCTATAAATTATTATTTGAAGACGGATGACGAGATGCCG	780
Query	781	GTCGATCATCAATGGATCAAACCTATCCAACGGCGAACCATTCCTACGGCAGCGAGCTC	840
Sbjct	781	GTCGATCATCAATGGATCAAACCTATCCAACGGCGAACCATTCCTACGGCAGCGAGCTC	840
Query	841	AGCAAGCTGAAGAGCACGGACAGCGGGAAGCGGATGTTTACGCTAAGCTCATCACCAGAT	900
Sbjct	841	AGCAAGCTGAAGAGCACGGACAGCGGGAAGCGGATGTTTACGCTAAGCTCATCACCAGAT	900
Query	901	CTTGGCCTGCATTGGACGGAGCCGAAAATACGGTAGCAAGCCTGGTAGAGCTGGAGCAC	960

Sbjct	901	 CTTGGCCTGCATTGGACGGAGCCGAAAATACGGTAGCAAGCCTGGTAGAGCTGGAGCAG	960
Query	961	GAGATCCGCGAGGCTGTAATGGAGCAGCAGACGGGCATGAAGTTCCGTTATCTGCAAGGT	1020
Sbjct	961	 GAGATCCGCGAGGCTGTAATGGAGCAGCAGACGGGCATGAAGTTCCGTTATCTGCAAGGT	1020
Query	1021	TCGCAGCTTGCACCGATCTGCCGACGGCCTTTAAGGTAGCGCATGTATCCGTTTCGCTAT	1080
Sbjct	1021	 TCGCAGCTTGCACCGATCTGCCGACGGCCTTTAAGGTAGCGCATGTATCCGTTTCGCTAT	1080
Query	1081	TCGGCTCGTTACGAGCCATACGGCAATGACGGTTCGATGCTTGTGCAGGTTTCAGCTGGAT	1140
Sbjct	1081	 TCGGCTCGTTACGAGCCATACGGCAATGACGGTTCGATGCTTGTGCAGGTTTCAGCTGGAT	1140
Query	1141	TATCAACTCGAGcaccaccaccaccaccactGA	1173
Sbjct	1141	 TATCAACTCGAGCACCACCACCACCACCACTGA	1173
