



Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Biologia Celular
Programa de pós-graduação em Biologia Molecular
Laboratório de Biologia Molecular

Identificação e análise funcional de sinais de secreção de *Pichia pastoris*

Osmar de Souza Oliveira Neto

Dissertação de Mestrado em Biologia Molecular

Brasília - DF

2012



UnB - IB

Departamento de Biologia Celular

Programa de pós-graduação em Biologia Molecular

Laboratório de Biologia Molecular

Grupo de Biotecnologia de Leveduras

Identificação e análise funcional de sinais de secreção de *Pichia pastoris*

Dissertação de Mestrado apresentada ao Departamento de Biologia Celular do
Instituto de Ciências Biológicas da Universidade de Brasília como requisito
parcial para a obtenção do grau de Mestre em Biologia Molecular

Aluno: Osmar de Souza Oliveira Neto

Matrícula: 10/0067425

Orientador: Fernando Araripe Gonçalves Torres

Co-Orientadora: Lídia Maria Pepe de Moraes

Trabalho realizado no Laboratório de Biologia Molecular, Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, sob a orientação do Professor Fernando Araripe Gonçalves Torres com apoio financeiro do CNPq, CAPES e FAPDF.

Banca Examinadora

Professor Dr. Fernando Araripe G. Torres

Universidade de Brasília

Presidente da Banca

Professor Dr. João Ricardo M. Almeida

Embrapa – Agroenergia

Examinador Externo

Professora Dra. Ildinete Silva Pereira

Universidade de Brasília

Examinadora Interna

Professora Dra. Lídia Maria P. de Moraes

Universidade de Brasília

Membro Suplente

*Dedico este trabalho a todos aqueles
que de alguma forma colaboraram
para a sua conclusão.
Em especial à minha
família.*

Agradecimentos

Em primeiro lugar a toda minha família, por acreditar em mim e me apoiar incondicionalmente. Tios, primos, irmãos... Impossível citar cada um sem escrever outra dissertação. Aos Oliveira, Viana e Petry muito obrigado!

Aos meus avós Osmar e Maria de Lourdes, pela fé e exemplo de vida. Sem eles não chegaria aqui. Amo vocês.

À minha querida mãe Ivana, pela vida e pelas palavras de afeto e sabedoria nos momentos difíceis e de extremo cansaço. Orgulho-me de ser seu filho.

À minha irmã Amanda, pela companhia.

À minha namorada Carol, pelo carinho e ajuda. Companheira fora e dentro do laboratório.

Aos irmãos de escolha Leonardo, Oscar, Marcelo e Pedro, pela companhia, amizade e por tornar meus dias mais alegres.

Aos amigos João, Bruna, Sara e Nolah, pela amizade sem fim. Nem sempre perto, mas sempre juntos.

Aos integrantes da minha banda (The Memories) e parceiros Eliceu e Eduardo, pela oportunidade de trabalhar na minha outra paixão: a música. Minha válvula de escape!

Ao Professor Fernando Araripe, pela oportunidade e pelo voto de confiança em mim e no meu trabalho.

À Professora Lídia, pela experiência em bancada.

A todos os amigos do Grupo de Biotecnologia de Leveduras (Lab.2), pela ajuda e por tornar o meu trabalho mais fácil e agradável.

À Dra. Viviane, pela amizade e co-orientação. Muito importantes para a conclusão deste trabalho.

Às amigas Paula e Fernanda, ex-lab.2, pela ajuda e diversão. Fazem muita falta no laboratório.

A todos os colegas do Laboratório de Biologia Molecular, pela companhia, amizade, ensinamentos e momentos de descontração.

Aos técnicos Ivonildes, Fátima e Thompson, por darem suporte e estrutura para o funcionamento do laboratório.

Ao professor Antônio Teixeira e colegas do Laboratório Multidisciplinar de Pesquisa em Doença de Chagas (Marol, Ciro, Alessandro, Fernando e Estér), pelas primeiras aventuras na pesquisa, ensinamentos, amizade e momentos de descontração. Foram muito importantes para o fundamento da minha carreira acadêmica.

Aos professores da banca, pela disponibilidade e paciência para correção do meu trabalho.

A Deus, pela vida e por me permitir, ao menos um pouco, ver de perto como ele age.

“Deus é a lei e o legislador do Universo.”
“Quero conhecer os pensamentos de Deus... O resto é detalhe.”
Albert Einstein

Enfim, a todos aqueles que são importantes para mim e com certeza contribuíram com esta conquista, direta ou indiretamente, de coração:

MUITO OBRIGADO!

“O futuro tem muitos nomes.
Para os fracos é o inalcançável.
Para os temerosos, o desconhecido.
Para os valentes é a oportunidade.”
Victor Hugo

Índice

Índice de figuras	i
Índice de tabelas	iii
Lista de abreviaturas e símbolos	iv
Resumo	vi
Abstract	vii
1. Introdução	1
1.1 Biotecnologia: expressão heteróloga	1
1.2 A levedura <i>Pichia pastoris</i>	3
1.3 Linhagens de <i>Pichia pastoris</i>	4
1.4 Genoma de <i>Pichia pastoris</i>	5
1.5 Vetores de expressão de <i>Pichia pastoris</i>	5
1.6 Promotores	6
1.7 Marcas de Seleção	7
1.8 Secretoma de <i>Pichia pastoris</i>	8
1.9 Peptídeo-sinal e secreção de proteínas	8
1.10 Gene repórter: α -amilase de <i>Bacillus subtilis</i>	10
2. Justificativa do trabalho	12
3. Objetivos e metas	13
4. Estratégias	14
5. Principais contribuições da proposta	18
6. Resultados e Discussão	19
6.1 Identificação de peptídeos sinais nativos de <i>Pichia pastoris</i>	19
6.2 Amplificação dos peptídeos sinais fusionados ao gene <i>amyE</i>	22
6.3 Amplificação do gene <i>amyE</i> sem PS e do fator- α	22
6.4 Clonagem dos produtos de PCR no vetor TOPO	24
6.5 Sequenciamento dos produtos de PCR	25
6.6 Clonagem dos produtos de PCR no vetor pPIC9	25

6.7	Análise de integração dos cassetes de expressão	27
6.8	Análise e identificação dos melhores sinais de secreção (teste piloto)	28
6.9	Análise do perfil secretório por SDS-PAGE	30
6.10	Análise e identificação dos melhores sinais de secreção (definitivo)	31
6.11	Determinação do número de cópias por Southern Blot	33
6.12	Confirmação do número de cópias por qPCR	34
6.13	<i>Screening</i> de clones 17 com uma cópia por qPCR	35
7.	Conclusões	36
8.	Perspectivas	37
9.	Materiais e Métodos	38
9.1	Materiais	38
9.1.1	Microorganismos	38
9.1.2	Enzimas	38
9.1.3	Marcadores moleculares	39
9.1.4	Kits	39
9.1.5	Ferramentas de bioinformática	40
9.1.6	<i>Primers</i>	41
9.1.7	Vetores	43
9.1.8	Meios de cultura	43
9.1.9	Extração de DNA plasmidial	45
9.1.10	Extração de DNA total	45
9.1.11	Análise de DNA em gel de agarose	46
9.1.12	Transferência e Southern Blotting	46
9.1.13	Análise de proteína em gel SDS-PAGE	47
9.1.14	Reagentes para atividade enzimática	49
9.1.15	Outras soluções	49
9.2	Métodos	50
9.2.1	Identificação de peptídeos sinais e amplificação por PCR	50
9.2.2	Amplificação da amilase e do fator- α	51
9.2.3	Análise em gel de agarose	51

9.2.4	Preparo das células bactérias para choque térmico	51
9.2.5	Ligação dos produtos PCR no vetor Blunt-TOPO	52
9.2.6	Ligação dos produtos de PCR no vetor TOPO-XL	52
9.2.7	Ligação dos fragmentos de DNA (vetor-inserto)	52
9.2.8	Transformação de <i>Escherichia coli</i> por choque térmico	52
9.2.9	Preparação de DNA plasmidial em pequena escala (miniprep) caseira	52
9.2.10	Preparação de DNA plasmidial em pequena escala (miniprep) de kit	53
9.2.11	Preparação de DNA plasmidial em média escala (midiprep) de kit	53
9.2.12	Extração de DNA total	53
9.2.13	Digestão de DNA com enzimas de restrição	54
9.2.14	Purificação e eluição de fragmentos de DNA	54
9.2.15	Precipitação de DNA	54
9.2.16	Transferência e Southern Blot	55
9.2.17	PCR em tempo real (qPCR)	55
9.2.18	Transformação de <i>Pichia pastoris</i> por eletroporação	56
9.2.19	Expressão heteróloga em <i>Pichia pastoris</i>	57
9.2.19.1	<i>Screening</i> em placas de petri	57
9.2.19.2	Indução em frascos (teste piloto)	57
9.2.19.3	Indução em frascos (teste definitivo)	57
9.2.20	Liberação de extrato protéico intracelular	58
9.2.21	Quantificação de proteínas totais (Bradford)	58
9.2.22	Atividade amilolítica quantitativa (dextrinizante)	59
9.2.23	Precipitação de proteínas	60
9.2.24	Análise em gel de SDS-PAGE	60
9.2.25	Coloração com Comassie Blue	60
9.2.26	Análise de sequências de DNA, eletroferogramas e alinhamentos	60
10.	Referências Bibliográficas	62
11.	Anexos	66

Índice de Figuras

Figura 1. Exemplos de peptídeos-sinais de diferentes organismos	8
Figura 2. Mapa físico do vetor pPIC9	14
Figura 3. Estratégia de amplificação do gene repórter <i>amyE</i> fusionado à sequência de nucleotídeos que codifica para os peptídeos-sinais analisados neste trabalho	15
Figura 4. Sequência de parte do vetor pPIC9	15
Figura 5. Amplificação por PCR dos 13 PS fusionados ao gene <i>amyE</i>	22
Figura 6. Comparação entre as sequências primárias do fator- α de <i>P. pastoris</i> e <i>S. cerevisiae</i>	23
Figura 7. Amplificação por PCR do gene <i>amyE</i> sem PS e do fator α <i>P. pastoris</i>	23
Figura 8. Clonagem dos produtos de PCR (PS + <i>amyE</i>) no vetor pCR-Blunt II-TOPO	24
Figura 9. Clonagem dos produtos de PCR (<i>amyE</i> sem PS e fator α de <i>P. pastoris</i>) no vetor TOPO-XL	24
Figura 10. Estratégia de clonagem dos 13 PS fusionados ao gene <i>amyE</i>	25
Figura 11. Clonagem dos cassetes PS + <i>amyE</i> no vetor pPIC9	26
Figura 12. Estratégia de clonagem do gene <i>amyE</i> sem PS e do fator α de <i>P. pastoris</i>	26
Figura 13. Clonagem do gene <i>amyE</i> sem PS e do fator α de <i>P. pastoris</i> no vetor pPIC9	27
Figura 14. Análise fenotípica em placa de transformantes de <i>P. pastoris</i> expressando α -amilase	27
Figura 15. Curva de crescimento dos clones amilolíticos recombinantes	28
Figura 16. Atividade amilolítica específica do sobrenadante de cultura dos clones recombinantes	29
Figura 17. Atividade amilolítica específica do extrato intracelular de cultura dos clones recombinantes	30
Figura 18. Perfil secretório dos clones amilolíticos	31
Figura 19. Densidade óptica celular (OD 600nm) das culturas dos melhores clones amilolíticos em 48 horas de indução com metanol	31

Figura 20. Atividade específica do sobrenadante de cultura dos melhores clones amilolíticos	32
Figura 21. Atividade específica do extrato intracelular de cultura dos melhores clones amilolíticos	32
Figura 22. Esquema do locus <i>HIS4</i> de <i>P. pastoris</i> recombinante	33
Figura 23. Análise de integração por Southern Blotting em clones amilolíticos selecionados	33
Figura 24. Determinação do número de cópias integradas por qPCR dos clones C+, 5 e 17	34
Figura 25. Determinação do número de cópias integradas por qPCR dos clones C+, 17.1, 17.2, 17.8 (1) e 17.8 (2)	35

Índice de Tabelas

Tabela 1. Exemplos de proteínas recombinantes expressas e secretadas em <i>P. pastoris</i> utilizando o fator α de <i>S. cerevisiae</i>	9
Tabela 2. Proteínas identificadas por Mattanovich e colaboradores no secretoma de <i>P. pastoris</i>	19
Tabela 3. Identificação, localização e sequência dos 13 peptídeos sinais selecionados neste trabalho	20
Tabela 4. Enzimas de restrição utilizadas	38
Tabela 5. <i>Primers</i> utilizados para amplificar o gene <i>amyE</i> fusionado em fase com os 13 peptídeos sinais selecionados	41
Tabela 6. <i>Primers</i> utilizados para amplificar o gene <i>amyE</i> sem peptídeo-sinal de <i>Pichia pastoris</i>	42
Tabela 7. <i>Primers</i> utilizados para amplificar o a região “pré-pró” do fator α de <i>Pichia pastoris</i>	42
Tabela 8. <i>Primers</i> utilizados para amplificar fragmentos do gene <i>GAP</i> e <i>amyE</i> na qPCR	43
Tabela 9. Vetores utilizados em <i>E. coli</i> e <i>P. pastoris</i>	43

Lista de Abreviaturas e Símbolos

A	adenina
<i>amyE</i>	gene que codifica para α -amilase
<i>AOX1</i>	gene que codifica para a álcool oxidase 1
C	citosina
CaCl₂	cloreto de cálcio
cm	centímetro
kDa	quilodalton
DNA	ácido desoxirribonucleico
EDTA	ácido etileno-diamino-tetra-acético
G	guanina
g	grama
<i>g</i>	força gravitacional
<i>GAP</i>	gene que codifica para a gliceraldeído-3-fostato desidrogenase
h	hora
HCl	ácido clorídrico
His	histidina
<i>HIS4</i>	gene que codifica para a histidinol desidrogenase
kb	quilo base
KOH	hidróxido de potássio
L	litro
LB	Luria-Bertani
M	molar
mA	miliampère
mg	miligrama
mL	mililitro
mM	milimolar

NaCl	cloreto de sódio
ng	nanograma
nm	nanômetro
OD_{600nm}	densidade celular óptica a 600 nm
pb	par de bases
PCR	reação em cadeia da polimerase
pH	potencial hidrogeniônico
p/v	peso por volume
PS	peptídeo-sinal
RNAse A	ribonuclease A
rpm	rotações por minuto
s	segundo
SDS	dodecil-sulfato de sódio
TAE	tampão tris-acetato-EDTA
TEMED	N, N, N', N' tetra-metil-etileno-diamina
U	unidade
v	volume
V	volts
YNB	base nitrogenada para leveduras
°C	graus Celsius
μF	microfaraday
μg	micrograma
μL	microlitro
μm	micrômetro

Resumo

Quase todos os vetores de expressão de *Pichia pastoris* utilizam a sequência da região pré-pró do fator- α de *Saccharomyces cerevisiae* como sinal de secreção, o qual pode ser processado incorretamente em condições de hiperexpressão. Utilizar um sinal de secreção nativo de *P. pastoris* pode contornar este problema, mantendo a estrutura correta da proteína heteróloga. Utilizando o software Signal P. 3.0, os peptídeos sinais de 13 proteínas nativas do secretoma e a região “pré-pró” do fator α putativo de *P. pastoris* foram identificados e em seguida suas sequências foram amplificadas via PCR. Essas sequências foram clonadas no vetor pPIC9, integradas e expressas nessa levedura, junto com o gene repórter α -amilase de *Bacillus subtilis*. O fator α de *S. cerevisiae* presente nesse vetor foi removido, permitindo que apenas as sequências testadas neste trabalho influenciassem a secreção da enzima heteróloga. Após análise dos halos de hidrólise em placa e ensaios enzimáticos utilizando sobrenadante de cultura, observou-se que pelo menos quatro sinais de secreção testados possuíam níveis de secreção semelhantes ao fator α de *S. cerevisiae*. O sequenciamento do N-terminal da α -amilase será feito por espectrometria de massa, após purificação da enzima em coluna de níquel, e permitirá a análise da eficiência de processamento dos peptídeos sinais testados. Este último teste permitirá a escolha de um novo sinal de secreção nativo de *P. pastoris*, introduzindo novas opções para expressão heteróloga nesta levedura, além das disponíveis comercialmente pela empresa Invitrogen (EUA).

Abstract

Almost all *Pichia pastoris* expression vectors uses *Saccharomyces cerevisiae* α -factor secretion signal, which can be improperly processed in conditions of hiperexpression. Using a native *P. pastoris* secretion signal should overcome this problem, maintaining the right structure of the heterologous protein. Using Signal P. 3.0 software, 13 secretion signals of native secretome proteins and the putative α factor of *P. pastoris* were identified and then their sequences were amplified via PCR. These sequences were cloned in pPIC9 vector, integrated and expressed in this yeast, along with a *Bacillus subtilis* α -amylase gene as a reporter. The *S. cerevisiae* α factor present in this vector was removed, allowing that only the sequences tested in this work had influence on the secretion of the heterologous enzyme. After analyses of hydrolysis halos on plates and enzymatic assays using culture supernatants, it was observed that at least four secretion signals tested had similar secretion levels of the *S. cerevisiae* α factor. Sequencing of the N-terminus of α -amylase will be carried out by mass spectrometry, after purification of the enzyme in a nickel based column, and will allow analysis of processing efficiency of the signal peptides tested. The latter test will enable the selection of a new native *P. pastoris* secretion signal, introducing new options for heterologous expression in this yeast, aside the ones commercially available by Invitrogen Company (USA).

Introdução

Biotecnologia: expressão heteróloga

O uso de sistemas biológicos para produção de proteínas ou biomoléculas de interesse humano não é recente. O primeiro relato de obtenção de uma vacina proteica se deu com Jenner, em 1796, quando este produziu uma vacina contra a varíola; e desde os anos 1900 a indústria biotecnológica tem se aquecido devido à implantação de técnicas de fermentação microbiana em larga escala (Demain & Vaishnav, 2009).

Produtos deste mercado são utilizados nas áreas de medicina, nutrição e alimentícia, têxtil, detergentes, plásticos, dentre outras. Nesse cenário, as enzimas têm um papel importante. Enzimas como proteases, amilases, glicoamilases, xilanases, lipases, celulases, isomerases e outras diversas alcançaram o mercado anual de 2,5 bilhões de dólares atualmente (Demain & Vaishnav, 2009). Entretanto, somente com a técnica de DNA recombinante foi possível alcançar esse patamar. Em 1993, quando o mercado de enzimas era de 140 milhões de dólares por ano, cerca de 50% das enzimas disponíveis eram recombinantes (Hodgson, 1994; Stroh, 1994).

Os sistemas de expressão mais utilizados para a produção de proteínas recombinantes atualmente são baseados em culturas de bactérias, leveduras, fungos filamentosos, células de insetos (via baculovírus) e células de mamíferos ou em organismos transgênicos (plantas e mamíferos) (Demain & Vaishnav, 2009).

A bactéria mais utilizada para este fim é *Escherichia coli*, uma enterobactéria gram-negativa. Seu uso ainda é a primeira opção para investigações laboratoriais e etapas iniciais de desenvolvimento de atividades comerciais envolvendo proteínas recombinantes (Chen, 2011). Trata-se de um versátil sistema de expressão, pois possui genética altamente conhecida e de fácil manipulação, crescimento rápido e de alta densidade celular, alta produção e baixo custo. Suas desvantagens são a falta de processamento pós-traducional (glicosilação, formação de pontes dissulfeto, entre outras), a co-produção de endotoxinas, produção de produtos citotóxicos (acetato) durante a fermentação, baixa secreção do produto recombinante e produção de proteínas inativas na forma de corpos de inclusão (Chen, 2011; Demain & Vaishnav, 2009).

Dentre as leveduras empregadas industrialmente para a produção de proteínas heterólogas destacam-se *Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha* e *Kluyveromyces lactis* e o uso dessa classe de organismos como sistema de expressão vem sendo realizado com sucesso crescente ao longo dos últimos 20 anos (Damasceno *et al.*, 2012; Demain & Vaishnav, 2009). As vantagens no uso de leveduras são a alta produtividade, crescimento rápido e em alta densidade celular, linhagens recombinantes mais estáveis, secreção mais eficiente e processamento pós-traducional mais similar ao de mamíferos (Demain & Vaishnav, 2009). A principal desvantagem é a presença de um padrão de

glicosilação diferente do de humanos (Hinnen *et al.*, 1994; Jung & Williams, 1997; Tate & Grisshammer, 1996).

A levedura *S. cerevisiae* foi o primeiro eucarioto a ser utilizado em técnicas de expressão heteróloga e a ter o genoma sequenciado (Çelik & Çalik, 2011; Goffeau, 1996). Como consequência, é um dos sistemas de expressão mais bem estudados. O conhecimento acerca de sua genética está entre os mais bem desenvolvidos e as técnicas de cultivo e fermentação são bem dominadas. Apesar de realizar várias modificações pós-traducionais importantes e de possibilitar a secreção de proteínas recombinantes, este sistema é geralmente incapaz de atingir os mesmos níveis de expressão daqueles atingidos pelo uso de *E. coli*.

Já no caso de fungos filamentosos, diversas espécies são utilizadas, geralmente do gênero *Aspergillus* (Yin *et al.*, 2007). As que dominam o mercado são *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus nidulans* e *Trichoderma reesei* (Nevalainen *et al.*, 2005). Seu uso está mais relacionado com a produção de proteínas nativas, mas o seu uso para produção de proteínas heterólogas é relativamente difundido. Suas vantagens são a alta capacidade de secreção de proteínas, alta estabilidade do gene heterólogo, alta produção e capacidade de realizar modificações pós-traducionais típicas de eucariotos. Suas desvantagens são a lenta taxa de crescimento, alta produção de proteases e técnicas mais complexas de cultivo (Yin *et al.*, 2007; Demain & Vaishnav, 2009).

O uso de baculovírus para a produção de proteínas heterólogas é dependente de células de insetos para a expressão, devido à especificidade de hospedeiros desse vírus (Yin *et al.*, 2007). Nesse sentido, o baculovírus, geralmente da espécie *Autographa californica*, é utilizado como vetor para infectar e transferir o gene heterólogo para células de insetos, geralmente da espécie *Spodoptera frugiperda* (Demain & Vaishnav, 2009). As células de insetos transfectadas são bons sistemas de expressão e apresentam como vantagens altos níveis de expressão, facilidade na expressão conjunta de múltiplos genes, realização de modificações pós-traducionais de eucariotos superiores e maior similaridade do produto recombinante ao original. As desvantagens são a secreção ineficiente, padrão de glicosilação diferente do de células de mamíferos e fermentação descontínua (Yin *et al.*, 2007; Demain & Vaishnav, 2009). A fermentação descontínua se refere à incapacidade do sistema de produzir a proteína heteróloga por longos períodos, uma vez que as células infectadas morrem, sendo necessárias rodadas sucessivas de adição de células novas e recém-transfectadas (Yin *et al.*, 2007).

A expressão heteróloga em plantas geralmente está mais relacionada com a transgenia e produção de linhagens resistentes do que com sistema de expressão propriamente dito. Entretanto, no cenário atual, as plantas emergiram como uma alternativa às células de procaríotos e eucariotos para a produção de proteínas heterólogas (Desai *et al.*, 2010). Como sistema de expressão suas vantagens são a localização da proteína em órgãos e tempos de crescimento específicos (garantido pela manipulação genética de sequências regulatórias tecido e desenvolvimento-específicas) e alta facilidade de crescimento e escalonamento

(ausência de fermentadores). As desvantagens são o longo tempo de desenvolvimento, baixa eficiência de transformação e expressão e padrão de glicosilação diferente do de células animais (Yin *et al.*, 2007; Demain & Vaishnav, 2009; Desai *et al.*, 2010).

Os sistemas de expressão baseados em células de mamíferos e animais transgênicos são geralmente utilizados quando padrões de glicosilação típico de mamíferos são necessários (Demain & Vaishnav, 2009). Isso significa que o seu uso está normalmente ligado à expressão de biofármacos para uso na indústria médica e farmacêutica. Com esse objetivo, o uso de células de mamíferos já se tornou mais eficiente e eficaz que o uso de microrganismos, com relação à quantidade de biofármaco produzido e recuperado (Wurm, 2004). Suas vantagens são modificações pós-traducionais complexas típicas de células de mamífero e produção moderada a alta. Suas desvantagens são a tecnologia complexa, altíssimos custos de produção, longo período de tempo para obtenção do produto, baixos níveis de secreção e potencial contaminação por vírus (Yin *et al.*, 2007; Demain & Vaishnav, 2009).

Nenhum dos sistemas de expressão citados é perfeito. Cada um possui vantagens e desvantagens inerentes ao organismo utilizado. A escolha de qual sistema é mais eficiente depende de diversos fatores, dentre eles a natureza da proteína a ser produzida (enzima ou biofármaco), além dos objetivos do trabalho e do pesquisador (pesquisa de base ou aplicada; uso em indústria médica, alimentícia, de combustíveis, dentre outras).

A levedura *Pichia pastoris*

Pichia pastoris é uma levedura classificada como pertencente ao Reino *Fungi*, Divisão *Eumycota*, Subdivisão *Ascomycotina*, Classe *Hemoascomycetes*, Ordem *Endomycetales*, Família *Saccharomycetaceae* e Subfamília *Saccharomycetoideae*. Recentemente esta levedura foi reclassificada no novo gênero *Komagataella* e dividida em três espécies: *K. pastoris*, *K. phaffii* e *K. pseudopastoris* (Kurtzman, 2009).

A mais notável característica fisiológica dessa levedura é o fato de ela ser metilotrófica, ou seja, capaz de crescer em meio contendo metanol como única fonte de carbono. Isto ocorre graças à hiperexpressão de uma enzima peroxissomal chamada álcool oxidase, que pode atingir 30% das proteínas intracelulares na presença de metanol (Couderc & Baratti, 1980).

Nos anos 60, *P. pastoris* foi considerada uma promissora fonte de proteínas (*single cell protein*, *SCP*) por ser capaz de crescer em uma fonte de carbono produzida de forma relativamente barata, a partir do metano presente no gás natural (Cregg, 2004; Macauley-Patrick *et al.*, 2005). Porém, com o aumento do cultivo da soja e com a crise petrolífera de 70, o interesse em *P. pastoris* como fonte de *SCP* diminuiu.

A partir dos anos 80, a empresa *Phillips Petroleum* e o *Salk Institute Biotechnology/Industrial Associates Inc.* (SIBIA) desenvolveram *P. pastoris* como sistema de expressão heteróloga. As poucas leveduras capazes de utilizar metanol; como *H. polymorpha* (*Pichia angusta*), *P. pastoris*, *Candida boidinii* e *Pichia methanolica* (Gellissen, 2000); tornaram-se então atraentes sistemas para a produção industrial de proteínas recombinantes. Isso é

justificado, pois elas permitem um rígido controle da expressão com a simples manipulação do meio de cultura.

Do ponto de vista de sistema de expressão, *P. pastoris* apresenta uma série de vantagens sobre *S. cerevisiae* (Cereghino & Cregg, 2000; Cregg *et al.*, 2000; Demain & Vaishnav, 2009), destacando-se as seguintes:

- a) Maior produtividade.
- b) Ausência de hiper-glicosilação.
- c) Crescimento em meio com metanol, que inibe o crescimento de contaminantes.
- d) Sistema barato para estabelecimento e manutenção em escala industrial.
- e) Integração estável de várias cópias do DNA transformante.
- f) Modificações pós-traducionais mais eficientes: processamento de sequências sinal, enovelamento, formação de pontes dissulfeto, certos tipos de adição de lipídeos e glicosilação dos tipos *O* e *N*.
- g) Melhor secreção do produto recombinante.

Além destas vantagens, *P. pastoris* pode crescer até altas densidades celulares (130 g/L) em meios definidos relativamente simples (Wegner, 1990; Cereghino & Cregg, 2000). Embora os relatos de produtividade de proteína heteróloga na faixa de 1 g/L não sejam raros, há um caso excepcional em que foram obtidos 12 g/L de antígeno tetânico (Clare *et al.*, 1991). Geralmente, a mensagem genética externa é integrada ao cromossomo, o que confere sua estabilidade mesmo sem pressão seletiva em fermentadores de grande capacidade (Romanos, 1995). A secreção de poucas proteínas endógenas é outra importante vantagem apresentada por *Pichia*, pois facilita a purificação de proteínas secretadas (Cregg *et al.*, 1993).

Em suma, a relativa facilidade na manipulação genética aliada aos baixos custos de produção fazem desta levedura um versátil sistema de expressão sem precedentes na Biotecnologia moderna.

Linhagens de *Pichia pastoris*

As linhagens de *P. pastoris* mais comumente utilizadas são divididas com base em duas características: linhagens auxotróficas (*ade1*, *ura3*, *arg4* e *his4*) e/ou protease-deficientes (*prb1*, *kex1* e *pep4*) (Cereghino & Cregg, 2000).

A linhagem NRRL Y-11430 (CBS 7435), classificada como *K. phaffii*, é a linhagem parental das linhagens auxotrófica GS-115 (*his4*) e prototrófica (X-33) (Kurtzman, 2009). Já as linhagens protease-deficientes são classificadas como *K. pastoris*. Esta espécie também inclui a linhagem DSMZ 70382 (NRRL Y-1603 ou CBS 704), considerada como linhagem referência (*type strain*) para todas as outras (Mattanovich *et al.*, 2009).

Genoma de *Pichia pastoris*

Recentemente, foram publicadas e disponibilizadas as sequências dos genomas das linhagens GS-115 (De Schutter *et al.*, 2009), DSMZ 70382 (Mattanovich *et al.*, 2009) e CBS 7435 de *P. pastoris* (Küberl *et al.*, 2011).

O genoma nuclear de *P. pastoris* tem um tamanho aproximado de 9.43 (De Schutter *et al.*, 2009), 9.41 (Mattanovich *et al.*, 2009) e 9.35 (Küberl *et al.*, 2011) Mpb distribuído em 4 cromossomos. Foram anotados 5313 (De Schutter *et al.*, 2009), 5450 (Mattanovich *et al.*, 2009) e 5007 (Küberl *et al.*, 2011) genes codificadores de proteínas. Trata-se de um genoma relativamente pequeno quando comparado ao de outras leveduras. *S. cerevisiae*, por exemplo, possui um genoma de aproximadamente 12.16 Mpb, distribuído em 16 cromossomos e com 6275 genes (Goffeau, 1996). Isto é atribuído ao fato de *P. pastoris* apresentar pouca redundância gênica, além de não ter sofrido duplicação genômica durante a evolução como ocorreu com *S. cerevisiae*.

Com a publicação da sequência do genoma de *P. pastoris*, vários aspectos de sua fisiologia podem agora ser investigados e explorados de forma acelerada dentro de um programa de melhoramento genético visando à expressão heteróloga de proteínas de interesse econômico.

Vetores de expressão de *Pichia pastoris*

Os vetores de expressão para *P. pastoris* são geralmente plasmídeos integrativos que contêm, obrigatoriamente:

- 1) Elementos genéticos para manipulação e propagação em *E. coli*;
- 2) Sequências regulatórias (promotores), induzíveis ou constitutivos, que são reconhecidas pela célula hospedeira;
- 3) Sequências terminadoras de transcrição de levedura (*AOX1tt*);
- 4) Marcas de seleção auxotrófica ou dominante;
- 5) Sítios múltiplos de restrição para clonagem.

Alguns vetores ainda podem conter, opcionalmente: sequências *tag* (His6 ou myc), que facilitam a identificação e purificação da proteína heteróloga; sinal de secreção, uma sequência que codifica um peptídeo sinalizador que direciona a exportação da proteína heteróloga para o meio extracelular; e sequências 5' e 3' do gene *AOX1*, que direcionam a integração do plasmídeo para este *locus* (Invitrogen EUA, manual *Pichia Expression Kit*).

Poucos plasmídeos foram desenvolvidos para *P. pastoris* e a maioria destes foi patenteada pela empresa Invitrogen (EUA). Os primeiros vetores para transformação de *Pichia* foram baseados no forte promotor do gene *AOX1* (álcool oxidase 1), que é induzido por metanol. A partir destes vetores (séries pPIC e pPICZ - Invitrogen), foram desenvolvidos métodos de transformação, marcas de seleção e estratégias de fermentação a fim de explorar o potencial deste sistema (Rosenfeld *et al.*, 1999). Estes vetores se tornaram os mais

difundidos sendo que mais de 500 proteínas já foram expressas neste sistema (Cregg, 2004; Macauley-Patrick *et al.*, 2005).

Além de vetores para expressão induzida, outros plasmídeos com promotores constitutivos se tornaram disponíveis nos últimos anos. Alguns destes vetores possuem o promotor *GAP* (série pGAPZ - Invitrogen) um forte promotor constitutivo que direciona a expressão de gliceraldeído-3-fosfato desidrogenase (Waterham *et al.*, 1997).

De forma geral, os vetores de *P. pastoris* são considerados grandes (maiores que 7 kb), por causa da quantidade de informação genética incluída. O tamanho pode muitas vezes dificultar o seu uso durante etapas de clonagens *in vitro* e pode gerar perda de estabilidade dos transformantes (Romanos *et al.*, 1992; Sears *et al.*, 1998). Algumas das sequências não essenciais listadas acima podem ser removidas ou outras de menor tamanho podem ser utilizadas (ex: gene *ble* - resistência à zeocina - 375 pb), com a finalidade de tornar o vetor menor e facilitar sua utilização (Daily & Hearn, 2005; Invitrogen EUA, manual *Pichia Expression Kit*).

Promotores

O promotor do gene *AOX1* apresenta uma força comparável aos promotores dos genes altamente expressos da via glicolítica. Entretanto, ao contrário destes, é firmemente controlado a nível transcricional pela presença do agente indutor, o metanol (Ellis *et al.*, 1985; Cregg *et al.*, 1989). Ao mesmo tempo, ele é fortemente reprimido na presença de glicose ou glicerol, já que são fontes de carbono preferenciais (Cereghino & Cregg, 2000).

A vantagem deste sistema é que a expressão de genes possivelmente deletérios para a célula pode ser controlada até que a célula atinja altas densidades celulares. Embora sempre tenha havido uma preocupação com relação aos possíveis efeitos deletérios causados pela expressão constitutiva de proteínas citotóxicas, este fenômeno tem sido raramente observado em *Pichia* (Macauley-Patrick *et al.*, 2005).

Apesar de ser o promotor mais estudado e, como consequência, mais utilizado, o pAOX1 nem sempre é a melhor opção. Seu uso com produtos utilizados na indústria alimentícia é desaconselhado, uma vez que o metanol é um subproduto tóxico da indústria petrolífera. Seu uso em biorreatores de larga escala também pode ser problemático, já que o metanol é altamente inflamável e o armazenamento de grandes quantidades deste combustível é perigoso. Por isso, outros promotores de *Pichia* (pGAP, pPEX8, pYPT1, pFLD1, entre outros) podem ser utilizados como alternativa (Cereghino & Cregg, 2000).

Quando as células são crescidas em glicose, os níveis de expressão obtidos pelo promotor pGAP são similares àqueles obtidos com o promotor pAOX1 em culturas crescidas com metanol. Há vários relatos em que a utilização do promotor pGAP se mostrou igual ou superior ao pAOX1 (Menendez *et al.*, 2004). A vantagem de se utilizar um sistema de expressão constitutivo é que o metanol não é necessário à expressão, não sendo necessário fazer mudanças no meio de cultura para se obter a produção. Além disto, a fonte de carbono é

mais barata (glicose) podendo também ser utilizado o glicerol, um subproduto da indústria de biodiesel. Uma desvantagem de promotores constitutivos é a sua utilização com produtos que possam ser tóxicos à célula (Cereghino & Cregg, 2000).

O gene *FDL1* está envolvido na utilização do metanol como fonte de carbono e na utilização de aminas metiladas como fonte de nitrogênio. O seu produto é a enzima formaldeído-desidrogenase dependente de glutatona (Shen *et al.*, 1998). O promotor desse gene pode ser utilizado na expressão de proteínas recombinantes utilizando o metanol (indutor) como fonte de carbono e sulfato de amônio como fonte de nitrogênio ou utilizando glicose como fonte de carbono e metilamina (indutor) como fonte de nitrogênio, uma fonte barata e não tóxica de nitrogênio (Cereghino & Cregg, 2000).

Níveis altos de expressão obtidos com os promotores *AOX1*, *GAP* e *FDL1* podem causar sobrecarga na maquinaria celular da levedura *P. pastoris*. Essa sobrecarga pode causar problemas de processamento, dobramento e citolocalização do produto recombinante. Nesses casos, promotores moderados como os dos genes *PEX8* e *YPT1* são interessantes (Cereghino & Cregg, 2000). O gene *PEX8* codifica uma proteína envolvida na biogênese do peroxissomo e o *YPT1* uma GTPase. O *pPEX8* é induzido por metanol e fracamente por glicose, enquanto o *pYPT1* é expresso na presença de glicose, manitol ou metanol (Liu *et al.*, 1995; Sears *et al.*, 1998).

Marcas de seleção

Com relação a marcas de seleção, os vetores de *P. pastoris* geralmente possuem a marca auxotrófica *HIS4* (série pPIC) ou a marca dominante *ble* (série pPICZ e pGAPZ) que confere resistência ao antibiótico zeocina. Menos utilizadas, também existem as marcas dominante *kan* (pPIC9K), que confere resistência ao antibiótico kanamicina (G418), e auxotróficas *ADE1*, *ARG4* e *URA3* (Cereghino e Cregg, 2000). As marcas *ble* e *kan* apresentam a importante vantagem de possibilitar a seleção de múltiplos eventos de integração, e conseqüentemente múltiplas cópias do gene heterólogo, em placas contendo concentrações crescentes de antibiótico (Daily & Hearn, 2005).

A marca auxotrófica *HIS4* é utilizada na maioria das vezes em conjunto com a linhagem GS115 de *P. pastoris* (*his4*). Os vetores da Invitrogen da série pPIC possuem em sua seqüência uma versão funcional do gene da enzima histidinol desidrogenase (*HIS4*), que permite o crescimento de clones transformantes em meio sem o aminoácido histidina.

O locus *his4* também pode ser utilizado para integração dos cassetes de expressão por recombinação homóloga. O direcionamento para a recombinação é dado pela linearização dos vetores da série pPIC com as enzimas *Sall* ou *Stul*, que possuem sítios presentes no meio do gene *HIS4* desse vetor. A linearização nesse ponto gera uma molécula de DNA com extremidades recombinantes homólogas ao gene *his4* da levedura.

Secretoma de *Pichia pastoris*

Sendo utilizada como um sistema de expressão, uma das vantagens da levedura *P. pastoris* sobre outros organismos é a sua capacidade de secretar poucas proteínas endógenas, como dito anteriormente. Direcionar a proteína recombinante para via secretória funciona como uma etapa inicial de purificação, já que ela perfaz a maior parte da fração proteica no sobrenadante (Cereghino & Cregg, 2000).

Ao analisar dados do sequenciamento da linhagem GS115 de *P. pastoris*, De Schutter *et al.* (2009) previram que o secretoma teórico dessa levedura seria composto por 54 proteínas. Porém, ao analisar o secretoma da linhagem DSMZ 70382 de *P. pastoris* crescida em glicose como fonte de carbono, Mattanovich *et al.* (2009) observaram a presença de apenas 20 proteínas nativas, quando tinham previsto 88 por métodos computacionais. Já Huang *et al.* (2011) observaram que, quando a linhagem X-33 dessa levedura é crescida em metanol como fonte de carbono, é possível encontrar 37 proteínas no sobrenadante de cultura.

Esses estudos de secretômica aumentam o conhecimento científico sobre o perfil secretório de *P. pastoris* e abrem portas para o desenvolvimento de linhagens e produtos tecnológicos mais eficientes, principalmente de novos sinais de secreção.

Peptídeo-sinal e secreção de proteínas

Como dito anteriormente, os vetores de expressão podem conter também um sinal de secreção (peptídeo-sinal) que direciona a proteína heteróloga para a via secretória.

O peptídeo-sinal é uma sequência curta composta por 13 a 36 resíduos de aminoácidos, presente no N-terminal das proteínas que são transportadas ao retículo endoplasmático para serem secretadas. Tipicamente, essa sequência possui, respectivamente, do lado N- ao C-terminal, três regiões: 1) uma região com um ou mais resíduos de aminoácidos de carga positiva; 2) outra com cerca de 10 a 15 resíduos hidrofóbicos; 3) e mais uma com relativa polaridade e resíduos de cadeia lateral curta (como alanina), perto do sítio de clivagem (Figura 1). Logo após sua síntese, ao emergir do ribossomo livre, o peptídeo-sinal é reconhecido pela *Signal-Recognition Particle* (SRP), um complexo de seis proteínas e um RNA. A SRP dirige o acoplamento do ribossomo livre à parede do retículo endoplasmático e a transposição da proteína pela membrana dessa organela. Uma enzima, a peptidase sinal, cliva o peptídeo-sinal antes da liberação da proteína no lúmen do retículo (Nelson & Cox, 2004).

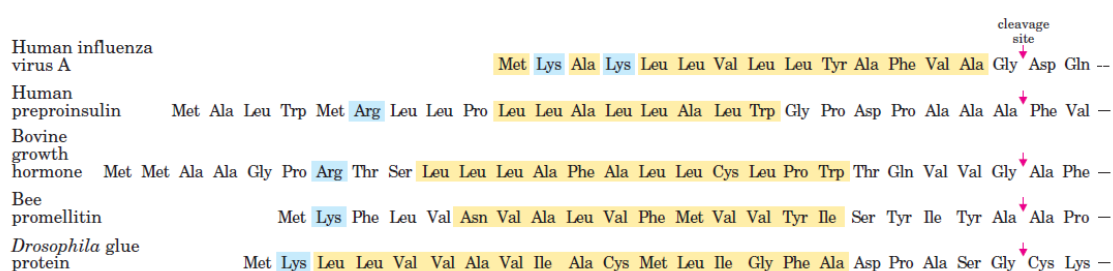


Figura 1. Exemplos de peptídeos-sinais de diferentes organismos. Azul: resíduos de aminoácidos de carga positiva. Amarelo: resíduos de aminoácidos hidrofóbicos. Seta: sítio de clivagem pela peptidase sinal precedido por aminoácidos polares com cadeia lateral curta, como alanina. Removido de Nelson & Cox, 2004.

A maioria dos vetores de secreção de *Pichia* utiliza o sinal de secreção heterólogo correspondente à região “pré-pró” do fator α de *S. cerevisiae*. Isso o torna responsável pela secreção de grande parte das proteínas expressas em *P. pastoris* (Tabela 1). A região “pré” é clivada pela peptidase-sinal enquanto a região “pró” é processada pelas proteinases codificadas pelos genes *KEX2* e *STE13*. A região “pró” necessita ser N-glicosilada em duas posições para uma eficiente clivagem por Kex2p e Ste13p (Cereghino & Cregg, 2000; Daily & Hearn, 2005). Já foi observado que em condições de hiperexpressão, a capacidade de glicosilação e de remoção da região “pró” pode ser comprometida o que leva a um processamento parcial do N-terminal da proteína heteróloga (Kozlov & Yagudin, 2008). Além disto, descobriu-se que a proteinase Kex2p de diferentes fungos, incluindo *P. pastoris*, apresenta diferentes especificidades. Este comportamento foi atribuído a uma adaptação estrutural de cada enzima aos seus substratos (Bader *et al.*, 2008).

Tabela 1. Exemplos de proteínas recombinantes expressas e secretadas em *P. pastoris* utilizando o fator α de *S. cerevisiae*. Modificado de Daily & Hearn, 2005.

Proteínas secretadas

Domínio extracelular do receptor de activina ACTR1IIa/b
 Antígeno SAG1 de *Toxoplasma gondii*
 Proteína anticoagulante e anti-metástase ghlanten
 Cistatina de galinha
 Hormônio ovino estimulante de folículos
 Alfa-lactalbumina
 Protease 7 de mastócitos de rato
 Lewis fucosiltransferase humana
 Alérgeno Ole e 1 de pólen de azeitona
 Moléculas HLA-DR2
 Pró-colágeno tipo III humano
 Beta-lactoglobulina
 Insulina
 1,2-alfa-D-manosidase de *Trichoderma reesei*
 Neuraminidase de *Influenza*
 ADP-ribosil-ciclase de *Aplysia*
 Anticorpo murino
 Fator de crescimento marsupial
 Fosfatase do nódulo radicular de soja
 Colagenase de fibroblasto humana
 Subunidade beta do hormônio bovino estimulante de folículos
 Receptor opioide MU humano
 Subunidade catalítica da enteroquinase bovina
 Alfa-N-acetilgalactosaminidase

Apesar de ser o sinal de secreção mais utilizado e que geralmente apresenta os melhores resultados, o fator- α de *S. cerevisiae* não é perfeito (Damasceno *et al.*, 2011). A empresa Invitrogen (EUA) também disponibiliza em alguns de seus vetores o sinal de secreção da fostatase ácida (*PHO1*) de *P. pastoris* ou da invertase (*SUC2*) de *S. cerevisiae* (Cereghino & Cregg, 2000; Daily & Hearn, 2005). Estes peptídeos sinais geralmente são utilizados quando a secreção com o fator α de *S. cerevisiae* não foi satisfatória. Há relatos do uso com sucesso do sinal de secreção *PHO1* na expressão da proteína G2 do envelope de *Hantaan virus* (Ha *et al.*, 2001) e da enzima levansucrase (LsdA) de *Acetobacter diazotrophicus* (Trujillo *et al.*, 2001). Já com o sinal de secreção *SUC2*, há relatos do uso com sucesso na expressão de uma α -amilase bacteriana (Paifer *et al.*, 1994) e da antitrombina III humana (Mochizuki *et al.*, 2001).

Em último caso, quando os sinais de secreção disponíveis comercialmente falham, a sequência sinal nativa da proteína heteróloga ou outra associada a altos níveis secretórios também podem ser utilizadas. Um exemplo é o caso da utilização do peptídeo sinal da fitohemaglutinina de *Phaseolus vulgaris* na expressão e secreção da própria fitohemaglutinina, lectina (aglutinina de *Galanthus nivalis*, GNA) e GFP (*green fluorescent protein*) (Raemaekers *et al.*, 1999). Neste estudo, a secreção dessas proteínas com o sinal de secreção da fitohemaglutinina se deu no mesmo nível da com o fator- α de *S. cerevisiae*, porém o processamento do N-terminal foi mais correto e a proteína heteróloga resultante era mais fiel. Outro exemplo é a utilização do peptídeo sinal de duas hidrofobinas de *T. reesei* (HFBI e HFBI) para a secreção de eGFP (Kottmeier *et al.*, 2009). Esse estudo demonstrou, usando essas sequências sinal, a eficiente secreção dessa proteína heteróloga em igual ou maior nível que com o fator α de *S. cerevisiae*, usado como controle.

Gene repórter: α -amilase de *Bacillus subtilis*

Gene repórter é qualquer gene utilizado em técnicas de biologia molecular que codifique uma proteína de fácil detecção. Quando aliado a outro gene ou sequência regulatória, desconhecido(a) ou pouco(a) estudado(a), o gene repórter indica, ou seja reporta, a sua atividade ou localização. Exemplos clássicos são os genes da GFP (*green fluorescent protein*) de *Aequorea victoria*, da β -galactosidase de *E. coli*, entre outros (Wood, 1995). Nesse sentido, as enzimas amilases também podem ser utilizadas como repórteres, pois possuem atividade enzimática detectada de forma rápida, fácil e barata e seus genes são facilmente manipuláveis.

Amilases, ou enzimas conversoras de amido, são enzimas que hidrolisam o amido (um polímero de moléculas de glicose) em oligossacarídeos de menor tamanho (Moraes, 2004). Alpha-amilases, ou α -1,4-glucano-4-glucanohidrolases (EC 3.2.1.1), são endoenzimas da família das amilases que clivam polissacarídeos em ligações internas α -glicosídicas (α -1,4) (Prakash & Jaiswal, 2010).

A α -amilase de *Bacillus subtilis*, codificada pelo gene *amyE*, é uma enzima termoestável com atividade ótima obtida em 60°C e pH 6,0 – 6,5 (Liu *et al.*, 2010). Uma versão truncada desse gene é utilizada como repórter pelo grupo de Biotecnologia de Leveduras, no

Laboratório de Biologia Molecular da Universidade de Brasília. Essa versão truncada não possui a região que codifica para 33 resíduos de aminoácidos C-terminais, referente à região de ligação da enzima ao substrato (Moraes *et al.*, 1995; Valência, 1990).

Justificativa do trabalho

Para uma eficiente secreção de proteínas heterólogas, seria desejável, pois, que peptídeos-sinais nativos de *P. pastoris* fossem utilizados no desenvolvimento de vetores de expressão para esta levedura.

De Schutter *et al.* (2009) haviam previsto que 9% das proteínas identificadas no genoma de *P. pastoris* possuíam um peptídeo-sinal canônico. Todavia, Mattanovich *et al.* (2009) ao estudarem o secretoma de *P. pastoris* identificaram apenas 20 proteínas no sobrenadante de culturas crescidas em glicose, sendo que a maioria representava enzimas envolvidas com a síntese da parede celular. A partir das estruturas primárias destas proteínas secretadas, é possível propor a sequência do peptídeo-sinal utilizando ferramentas computacionais disponíveis (Emanuelsson *et al.*, 2007).

O uso de peptídeos-sinais nativos poderá contribuir para o correto processamento do N-terminal de proteínas heterólogas expressas em *P. pastoris*, o que é altamente desejável considerando a importância desta levedura para a produção de proteínas recombinantes, sobretudo biofármacos.

Objetivos e Metas

O objetivo principal deste projeto foi identificar e analisar possíveis peptídeos-sinais de proteínas secretadas de *P. pastoris* com a finalidade de desenvolver um novo vetor de expressão para este organismo. Para atingir este objetivo, foram propostas as seguintes metas:

- 1) Identificar os peptídeos-sinais (PS) das proteínas secretadas em *P. pastoris*;
- 2) Amplificar as sequências dos PS fusionados ao gene repórter α -amilase (*amyE*) de *Bacillus subtilis* por PCR;
- 3) Clonar os produtos de PCR (PS + *amyE*) em um vetor de expressão de *P. pastoris*;
- 4) Obter e identificar clones transformantes e secretores de α -amilase em placa;
- 5) Avaliar a eficiência de secreção dos PS com a determinação de atividade enzimática em culturas crescidas em meio líquido.

Estratégias

1) Identificação de peptídeos-sinais (PS)

A partir dos dados do secretoma de *P. pastoris* (Mattanovich *et al.*, 2009) foram escolhidas algumas proteínas para análise dos seus PS. Foi dada preferência às enzimas remodeladoras da parede celular pois algumas das proteínas secretadas são enzimas glicolíticas, também encontradas intracelularmente, cujos sinais de secreção não são óbvios. A predição do PS foi feita com base em softwares disponíveis (Emanuelsson *et al.*, 2007). A sequência da região pré-pró do fator- α de *P. pastoris* proposta por Küberl *et al.* (2011) também foi estudada neste projeto.

2) Clonagem dos PS em vetor para *P. pastoris*

Todos os PS propostos neste projeto foram clonados no vetor pPIC9, disponibilizado comercialmente pela empresa Invitrogen, EUA (Figura 2).

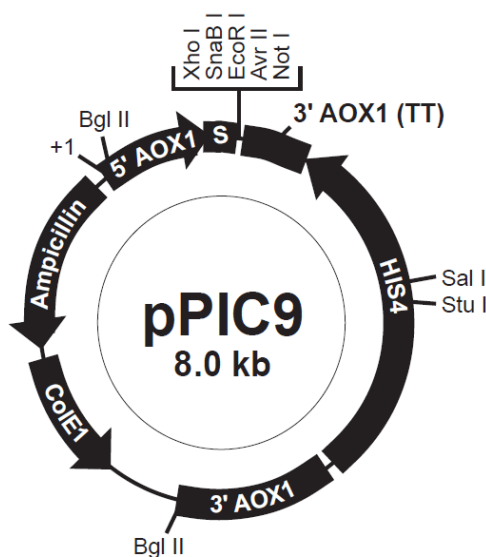


Figura 2. Mapa físico do vetor pPIC9. 5' AOX1: promotor do gene AOX1 de *P. pastoris*; S: peptídeo-sinal do fator α de *S. cerevisiae*; 3' AOX1-TT: terminador de transcrição do gene AOX1 de *P. pastoris*; ColE1: origem de replicação de *E. coli*; Ampicillin: gene de resistência a ampicilina. Mapa disponibilizado pela empresa Invitrogen (EUA).

As sequências de PS propostas neste projeto fizeram parte da sequência de primers 5' sintetizados para uso em reações de PCR. O primer 5' possui região de anelamento na porção N-terminal da versão truncada do gene da α -amilase (*amyE*) de *B. subtilis* (Moraes *et al.*, 1995),

além dos PS fusionados e o sítio para enzima de restrição *Bam*HI. O primer 3' se anela na porção C-terminal deste gene, além de possuir informação para tradução de uma cauda de histidina (*his-tag*) e sítio para enzima de restrição *Eco*RI (Figura 3).

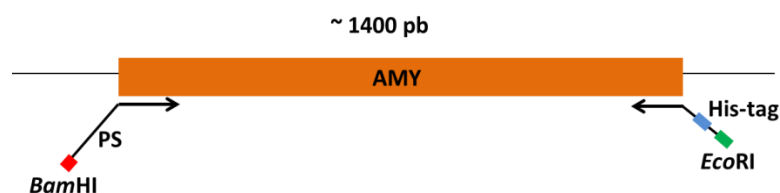


Figura 3. Estratégia de amplificação do gene repórter *amyE* fusionado à sequência de nucleotídeos que codifica para os peptídeos-sinais analisados neste trabalho. Setas: região de anelamento dos primers. PS: sequência de DNA que codifica para a sequência protéica de um peptídeo-sinal. Vermelho: sítio de *Bam*HI. Verde: sítio de *Eco*RI. Azul: sequência que codifica para cauda de histidina.

Os sítios para as enzimas de restrição *Bam*HI e *Eco*RI nas extremidades 5' e 3' do amplicom, respectivamente, permitiu a subclonagem de forma orientada no vetor pPIC9 (vide Figura 2). A digestão do vetor pPIC9 com as enzimas de restrição *Bam*HI e *Eco*RI remove o PS presente neste vetor (fator α de *S. cerevisiae*), permitindo que apenas os PS propostos neste trabalho influenciassem a secreção da α -amilase repórter (Figura 4).

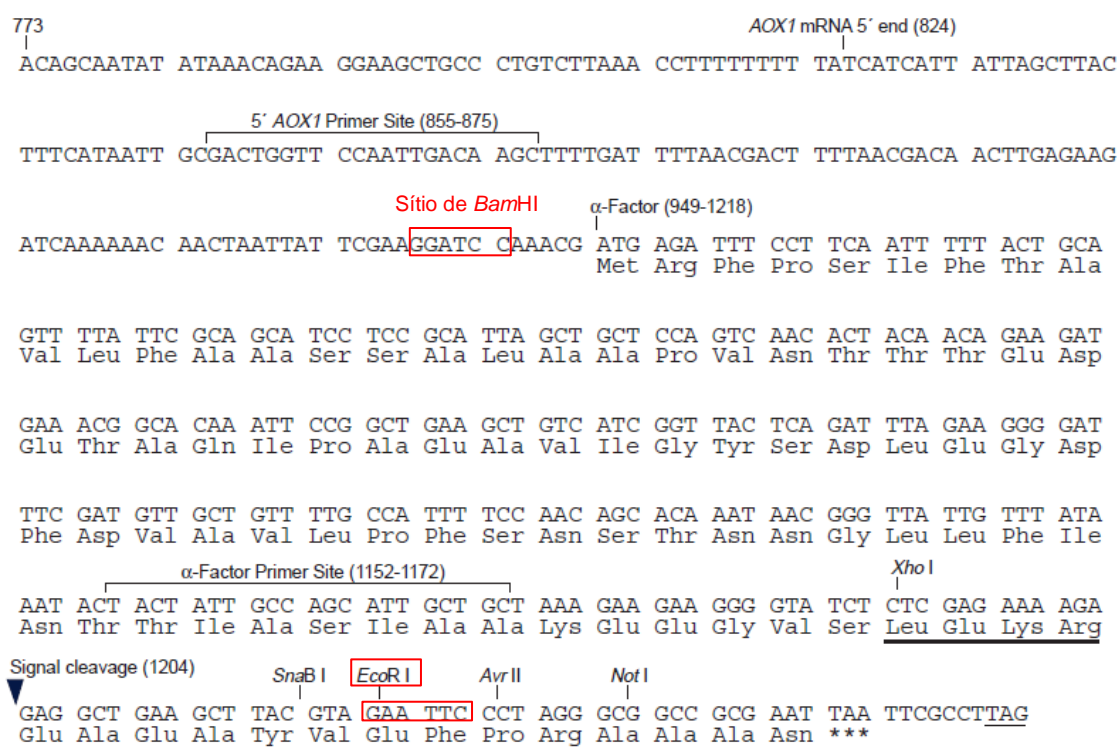


Figura 4. Sequência de parte do vetor pPIC9. 5' AOX1: promotor do gene *AOX1* de *P. pastoris*. α -factor: peptídeo-sinal do fator α de *S. cerevisiae*. Sítios de *Bam*HI e *Eco*RI destacados em vermelho. Mapa disponibilizado pela empresa Invitrogen (EUA), com modificações.

Como a enzima é eficientemente secretada a partir do PS presente no vetor pPIC9, o gene *amyE* sem peptídeo-sinal foi utilizado como controle positivo para estudos qualitativos e quantitativos de secreção enzimática em comparação com outros diferentes PS. Para este fim, o gene foi amplificado sem peptídeo-sinal e com os sítios das enzimas *XhoI* e *NotI*, cuja digestão do vetor pPIC9 não remove o sinal de secreção presente (Figura 2).

3) Transformação de *P. pastoris* e seleção de clones

Após amplificação e clonagem, os plasmídios contendo os diferentes PS foram linearizados com a enzima de restrição *SalI* para direcionar a integração do vetor para a região 5' do gene *HIS4* por recombinação homóloga. A linhagem de laboratório GS115 de *P. pastoris* (*his4*) foi transformada por eletroporação e os clones transformantes foram selecionados em placas de meio mínimo com glicose 2% (MD) sem histidina. Clones transformantes foram transferidos para uma nova placa com meio mínimo com glicose 2% (MD) sem histidina para o isolamento de clones individuais. Em seguida, estes clones foram transferidos para placas com meio mínimo contendo metanol 0,5% e amido solúvel (Sigma) 1%. A identificação dos halos de hidrólise de amido foi feita após coloração das placas com vapor de iodo. A razão do diâmetro da colônia pelo diâmetro do halo será determinada, como um dos métodos de identificação dos clones que secretam α -amilase com mais eficiência. Como controle, foi usado o vetor pPIC9 contendo o gene *amyE* clonado sob controle de secreção do peptídeo sinal do fator α de *S. cerevisiae*. Além do ensaio em placa, foram realizados ensaios enzimáticos para atividade amilásica (teste de Fuwa) a partir de sobrenadantes de culturas crescidas em meio líquido (Moraes *et al.*, 1995). Para efeito de normalização de resultados, trabalhamos com clones transformantes que só continham 1 cópia integrada do vetor, o que foi determinado por Southern Blotting e PCR em tempo real. Estudos anteriores mostraram que isto ocorre frequentemente com o vetor pPIC9 quando a integração é feita no *locus HIS4*. Ainda assim, este evento foi verificado por Southern Blotting usando-se a sequência do gene *AOX1* e *HIS4* como sonda.

4) Determinação da sequência N-terminal

Para determinar o correto processamento do PS pela peptidase sinal o N-terminal da amilase recombinante secretada será determinado. Clones de *P. pastoris* secretores de α -amilase serão crescidos em meio líquido MD (glicose 2%) até uma OD₆₀₀ de 0,7. As células serão coletadas e o sobrenadante será diretamente aplicado em uma coluna de afinidade de níquel que permite a purificação de proteínas com *his-tag* em uma única etapa cromatográfica após eluição com imidazol. A enzima eluída será dializada e concentrada por liofilização. A

determinação do N-terminal será feita pelo método de espectrometria de massas no Lab. de Espectrometria de Massas da Embrapa – Gado de Corte (MS) com colaboração do Dr. Newton Valério Verbisck.

5) Validação do peptídeo-sinal escolhido

Para testar a eficiência do peptídeo-sinal selecionado, um biofármaco humano será expresso em *P. pastoris*. Para tanto, expressaremos o fator de crescimento epidermal humano (EGF) já clonado em nosso laboratório e que já foi expresso com sucesso em *P. pastoris*. O peptídeo-sinal selecionado será fusionado às sequências 5' correspondentes à extremidade N-terminal do EGF maduro. Esta fusão será feita por PCR. O gene resultante será clonado em vetor de expressão seguindo-se transformação de *P. pastoris*. A expressão será feita em meio líquido e, após concentração do sobrenadante, a amostra será dializada, liofilizada e aplicada em coluna de troca iônica (DEAE-Sepharose) para purificação de EGF. A proteína purificada terá seu N-terminal determinado.

Principais contribuições da proposta

A principal contribuição tecnológica desta dissertação é a seleção de peptídeos-sinais nativos de *P. pastoris* que serão empregados em novos vetores de expressão para esta levedura, que estão sendo desenvolvidos no Laboratório de Biologia Molecular da UnB. A construção destes vetores é parte integrante do projeto “Plataforma Tecnológica para a Produção de Proteínas Recombinantes em *Pichia pastoris*” (Edital Pronex FAPDF, 2009).

O estudo do processamento dos peptídeos-sinais também levará à identificação de sequências sinal que permitam a produção de proteínas recombinantes mais autênticas, o que é de suma importância no caso de biofármacos. Além disto, este trabalho contribui para as análises pós-genômicas em *P. pastoris* pela caracterização dos sinais moleculares nativos que direcionam a secreção de proteínas neste organismo.

Como resultado dessa proposta, propomos a publicação de um artigo em revista indexada internacional tratando da identificação e caracterização dos peptídeos sinais identificados neste trabalho e da expressão de outro gene repórter com a sequência sinal escolhida.

Resultados e Discussão

Identificação de peptídeos sinais nativos de *Pichia pastoris*

De acordo com o descrito por Mattanovich *et al.* (2009), a análise do secretoma de culturas de *P. pastoris* crescidas com glicose como fonte de carbono revelou apenas 20 proteínas em sua composição, sendo que grande parte dessas proteínas está envolvida com o remodelamento da estrutura da parede celular da levedura (Tabela 2).

Tabela 2. Proteínas identificadas por Mattanovich e colaboradores no secretoma de *P. pastoris*. Informações sobre pI, massa molecular, função e localização. Modificado de Mattanovich *et al.* (2009).

PIPA ID	Função predita	pI / MM (kDa) teóricos	Localização predita
1) PIPA 00211	Função desconhecida. Proteína de parede celular ligada covalentemente	5.01 / 45.73	Secretada
2) PIPA 00246	Hexoquinase fúngica hipotética	5.98 / 24.92	Sem PS
3) PIPA 00436	Proteína de parede celular relacionada com glucanases	4.83 / 36.07	Secretada
4) PIPA 00545	Proteína de parede celular relacionada com glucanases	4.33 / 45.02	Secretada
5) PIPA 00748	Proteína O-glicosilada necessária à estabilização da parede celular	4.22 / 31.86	Secretada
6) PIPA 00934	Função desconhecida. Proteína extracelular com domínio SCP	5.55 / 31.72	Secretada
7) PIPA 00956	L18A proteína ribossômica 60S	9.92 / 21.82	Sem PS
8) PIPA 01008	GAS1: beta-1,3-glucanosiltransferase	3.98 / 57.20	Secretada
9) PIPA 01010	GAS1: beta-1,3-glucanosiltransferase	3.99 / 58.37	Secretada
10) PIPA 01223	Glucanase de parede celular potencial	4.34 / 49.39	Secretada

11) PIPA01958	Endo-beta-1,3- glucanase	4.03 / 33.76	Secretada
12) PIPA02332	Sem similaridade	6.01 / 23.64	Sem PS
13) PIPA02510	Gliceraldeído-3-fosfato desidrogenase	6.24 / 35.74	Sem PS claro
14) PIPA02524	glucana 1,3-beta- glucosidase similar a EXG1 de <i>S. cerevisiae</i> (YLR300W)	4.51/46.22	Secretada
15) PIPA02544	aldeído desidrogenase, Adh2p [<i>S.</i> <i>cerevisiae</i>]	6.00 / 36.86	Sem PS
16) PIPA03955	endo-1,3-beta- glucanase [<i>P. stipitis</i> CBS 6054], Dse4p [<i>S.</i> <i>cerevisiae</i>]	4.70 / 109.45	Secretada
17) PIPA04722	Proteína de parede cellular semelhante a glucanases	5.18 / 32.95	Secretada
18) PIPA05357	Sem similaridade	4.25 / 66.46	Sem PS / 2 TM
19) PIPA05673	Proteína YLR286Cp-like [<i>S. cerevisiae</i>], endoquitinase	4.05 / 71.87	Sem PS claro
20) PIPA05771	Quitina desacetilase, Cda2p [<i>S.</i> <i>cerevisiae</i>]	5.25 / 34.66	Secretada, baixa probabilidade

MM = massa molecular. PS = peptídeo sinal. TM = domínio transmembrânico.

Utilizando esse trabalho como base, uma análise de bioinformática (utilizando o software SignalP 3.0) foi feita na sequência dessas 20 proteínas visando a identificação e localização de peptídeos sinais em sua estrutura primária. Como esperado a partir dos resultados de Mattanovich *et al.* (2009), a análise das sequências primárias mostrou que das 20 proteínas 13 apresentaram regiões típicas de um peptídeo-sinal em suas sequências e nas outras 7 não houve identificação de uma região de PS. As sequências de tais peptídeos sinais estão discriminadas na tabela 3.

Tabela 3. Identificação, localização e sequência dos 13 peptídeos sinais selecionados neste trabalho.

Nome	Locus ID	Sequência PS + Contexto de Kozak (3 pb)
PS1 PIPA 00211	chr2- 2_0271	cat M N L L S L T L L L F T T E A <u>L A</u> L T G E
		ATGAACTTATTAAGTCTAACCTTACTATTGTTCACTACTGAAGCACTCGCCCTCACCGGTGAA

PS3 PIPA 00436	chr4_004 6	M K S Q L I F M A L A S L V A S A P L E H aat ATGAAATCTCAACTTATCTTTATGGCTCTTGCCTCTCTGGTGGCCTCCGCTCCTTTAGAACAC
PS4 PIPA 00545	chr2- 2_0064	M K I S A L T A C A V T L A G L A I A A P A P cca ATGAAGATATCCGCTCTTACAGCCTGCGCTGTTACTCTAGCTGGTCTTGCAATTGCAGCACCAGCTCCA
PS5 PIPA 00748	chr4_030 5	M K L A A L S T I A L T I L P V A L A G Y A P att ATGAAGCTCGCTGCACTCTCCACTATTGCATTAACATATTTTACCCGTTGCCTTGGCTGGCTATGCTCCA
PS6 PIPA 00934	chr3_007 6	M K L S T N L I L A I A A A S A V V S A A P V A cca ATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTCCGCGTTGTCTCAGCTGCCTCCAGTTGCT
PS8 PIPA 01008	chr1- 3_0226	M F K S L C M L I G S C L L S S V L A A D F aat ATGTTTAAATCTCTGTGCATGTTAATAGGATCCTGCCTATTGAGTTCAGTCTTGGCAGCAGATTTTC
PS9 PIPA 01010	chr1- 3_0227	M L S I L S A L T L L G L S C A S D L T P P gca ATGTTGTCCATTTTAAAGTGCATTAACCTCTGCTGGCCTGTCTTGTGCTTCAGATCTGACCCACCT
PS10 PIPA 01223	chr2- 1_0052	M L S T I L N I F I L L L F I Q A S L Q A P I P aat ATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGCATCCCTACAGGCTCCAATACCT
PS11 PIPA 01958	chr1- 4_0426	M I F N L K T L A A V A I S I S Q V S A V S S L G F aag ATGATCTTTAATCTTAAAACACTGGCTGCGGTTGCAATCTCCATTTACAAAGTGTCTGCAGTTTCTCTCTGGGTTTT
PS14 PIPA 02524	chr2- 1_0454	M N L Y L I T L L F A S L C S A I T L P K R D ggg ATGAACTTGTACCTAATTACATTACTATTTCGCCAGTCTATGCAGCGCAATTACTCTCCCAAAGAGAGAT
PS16 PIPA 03955	chr1- 1_0130	M S F S S N V P Q L F L L L V L L T N I V S G A V I S V W S caa ATGTCAATCTCTCCAACGTGCCACAACCTTTCTTGTGTTGTTCTGTTGACCAATATAGTCAGTGGAGCAGTCATTTCAATGATGGTCC
PS17 PIPA 04722	chr1- 3_0229	M Q V K S I V N L L L A C S L A V A R P L E H acc ATGCAAGTTAAATCTATCGTTAACCTACTGTTGGCATGTTTCGTTGGCCGTTGGCCAGACCTTTGGAGCAC
PS20 PIPA 05771	chr1- 1_0103	M T Q K T Q Y I I I L L T L I V S A F S E S S G Q N A taa ATGACTCAAAGACACAGTATATAATCATTTTGTGACTCTGATAGTTTCAGCTTTCTCCGAAAGCTCGGGTCAAATGCT

Obs: O provável sítio principal de ação da peptidase sinal é entre os aminoácidos assinalados em negrito e sublinhados.

Uma vez que o SignalP 3.0 utiliza diversos algoritmos para a predição de peptídeos-sinais em sequências primárias de proteínas, dependendo do algoritmo utilizado, múltiplos sítios prováveis de clivagem pela peptidase sinal foram gerados ao lado do sítio principal. Por isso foi decidido estender em alguns aminoácidos as sequências que seriam utilizadas, garantindo um processamento correto dos peptídeos sinais analisados.

Como o objetivo principal deste trabalho foi a identificação de sinais de secreção nativos de *Pichia* que pudessem ser utilizados em novos vetores de expressão, não havia interesse nas 7 proteínas sem peptídeo sinal reconhecido pelo programa empregado; já que a secreção de tais proteínas deva acontecer por uma via alternativa ainda não esclarecida. Por isso, o trabalho prosseguiu baseado nos 13 peptídeos sinais identificados.

Amplificação das sequências dos peptídeos sinais fusionados ao gene repórter *amyE*

Com base na sequência do gene *amyE* de *B. subtilis*, *primers* foram desenhados para a amplificação desse gene e a sequência dos 13 peptídeos sinais selecionados foram incluídos no *primer forward*, originando 13 *primers forward* diferentes (página 41, materiais e métodos).

As PCRs foram realizadas e o gene *amyE* foi amplificado com sucesso fusionado à sequência dos 13 peptídeos-sinais (Figura 5).

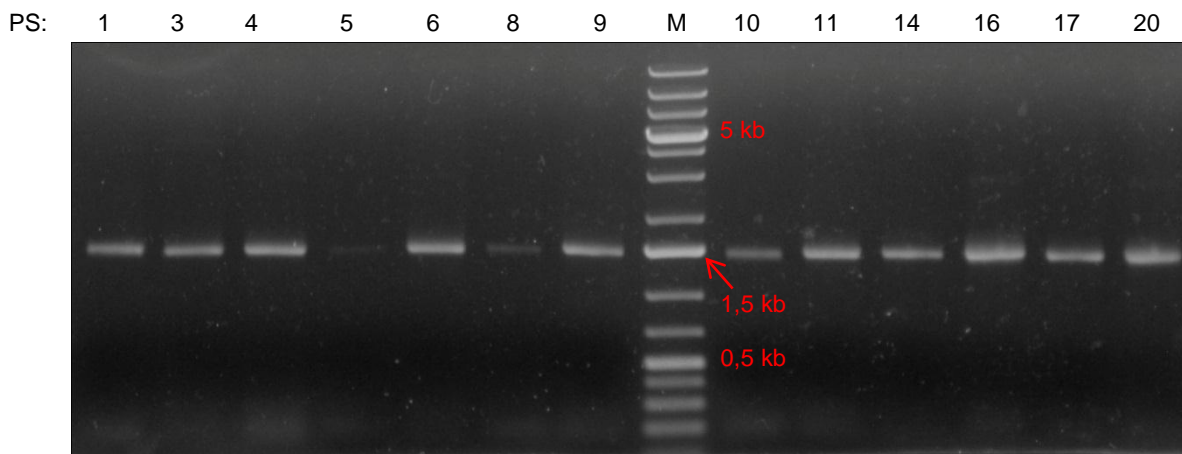


Figura 5. Amplificação por PCR dos 13 PS fusionados ao gene *amyE*. Análise de massa molecular indicando o tamanho aproximado de 1,5 kb dos treze produtos de PCR (PS + *amyE*) realizada por eletroforese em gel de agarose 1%, corado com 0,5 µg/mL de brometo de etídeo. M: Marcador O'Gene Ruler 1 kb plus (Fermentas).

Amplificação do gene *amyE* (sem peptídeo-sinal) e da sequência da região pré-pró do fator- α de *Pichia pastoris*

Com base na sequência do gene *amyE* de *B. subtilis*, *primers* foram desenhados para a amplificação desse gene sem peptídeo sinal, para uso como controle positivo (página 42, materiais e métodos). Com base da sequência do fator- α de *P. pastoris* disponibilizada por Küberl *et al.* (2011), *primers* foram desenhados para a amplificação da sequência relativa ao peptídeo-sinal (região “pré”) e à região “pró” desse gene (Figura 6 e página 42, materiais e métodos).

Clonagem dos produtos de PCR no vetor de clonagem TOPO

Em seguida, os 15 produtos de PCR foram clonados no vetor de clonagem pCR-Blunt II-TOPO ou TOPO-XL (Invitrogen). A finalidade foi purificar os fragmentos amplificados digeridos com as enzimas de restrição específicas de cada um, cujos sítios foram incorporados às sequências dos *primers*. Etapa necessária para posterior clonagem no vetor de expressão pPIC9 (Invitrogen) (Figuras 8 e 9).

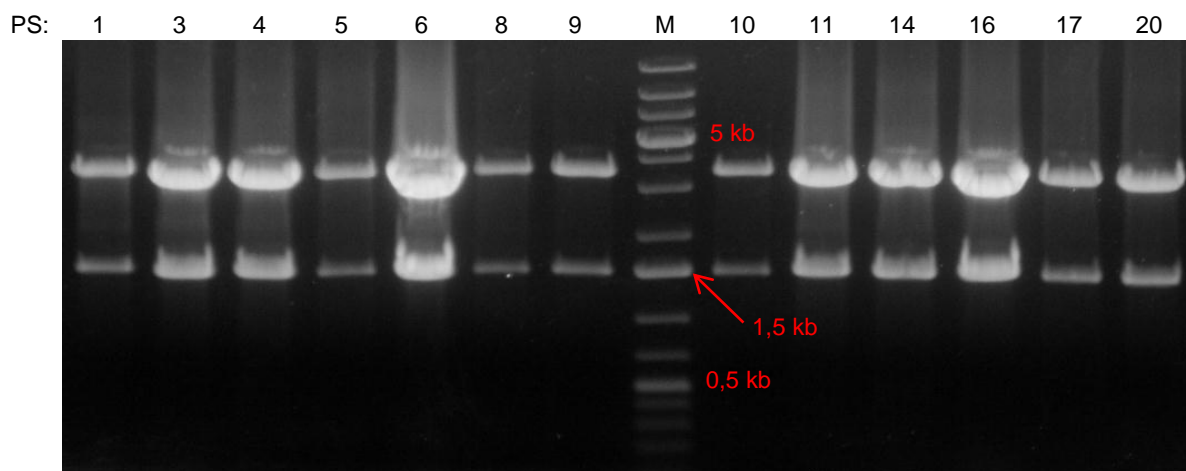


Figura 8. Clonagem dos produtos de PCR (PS + *amyE*) no vetor pCR-Blunt II-TOPO. Análise de restrição do vetor pCR-Blunt II-TOPO por eletroforese em gel de agarose 1%, corado com 0,5 µg/mL de brometo de etídeo, indicando os treze produtos de PCR clonados no vetor. Vetores digeridos com *Bam*HI e *Eco*RI, liberando o fragmento do inserto (~ 1,5 kb) e do vetor (~ 3,5 kb). M: Marcador O'Gene Ruler 1 kb plus (Fermentas).

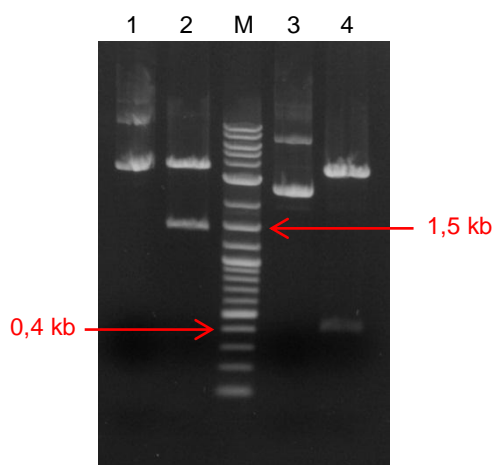


Figura 9. Clonagem dos produtos de PCR (*amyE* sem PS e fator α de *P. pastoris*) no vetor TOPO-XL. Análise de restrição do vetor TOPO-XL por eletroforese em gel de agarose 1%, corado com 0,5 µg/mL de brometo de etídeo, indicando os dois produtos de PCR clonados no vetor. 1: TOPO-XL-AMY intacto. 2: TOPO-XL-AMY digerido com *Xho*I e *Not*I (liberando o fragmento *amyE* de *B. subtilis* sem peptídeo sinal de tamanho aproximado de 1,45 kb). 3: TOPO-XL-Fa intacto. 4: TOPO-XL-Fa digerido com *Bam*HI e *Eco*RI (liberando o fragmento fator α de *P. pastoris* de tamanho aproximado de 0,4 kb). M: Marcador 2-log (New England Biolabs).

Sequenciamento dos produtos de PCR

Após confirmação da clonagem no vetor TOPO, os produtos de PCR foram sequenciados para confirmar a fidelidade das sequências relativas aos peptídeos sinais após processo de amplificação. Todos os produtos apresentaram 100% de identidade com a sequência alvo, menos o PS6 e o PS14, que apresentaram uma deleção na região do *primer* (anexo I).

No caso do PS6 a deleção ocorreu na região de anelamento ao gene *amyE* e no caso do PS14 a deleção ocorreu na sequência relativa ao peptídeo-sinal incorporada ao *primer*. As deleções provavelmente foram causadas por erro de síntese pela empresa IDT Technologies, o que pode ter acontecido devido ao grande tamanho dos *primers*.

Era esperado que essas deleções causassem mudança na fase de leitura (*frameshift*) do processo de transcrição, gerando um produto não funcional ou mesmo nenhum produto após o processo de tradução. Entretanto, o trabalho prosseguiu com todos os PSs, enquanto novos *primers* eram resintetizados para que o PS6 e o PS14 fossem amplificados sem mutações junto com o gene *amyE*.

Clonagem dos produtos de PCR no vetor de expressão pPIC9

Após serem digeridos com enzimas de restrição específicas e serem purificados, os 15 produtos de PCR obtidos foram clonados no vetor de expressão pPIC9 (Invitrogen) para posterior transformação de *Pichia*. Na digestão com *Bam*HI e *Eco*RI, o vetor pPIC9 perde seu peptídeo-sinal original (fator α de *S. cerevisiae*) que é substituído por cada um dos 13 PS fusionados com o gene *amyE* (Figuras 10 e 11).

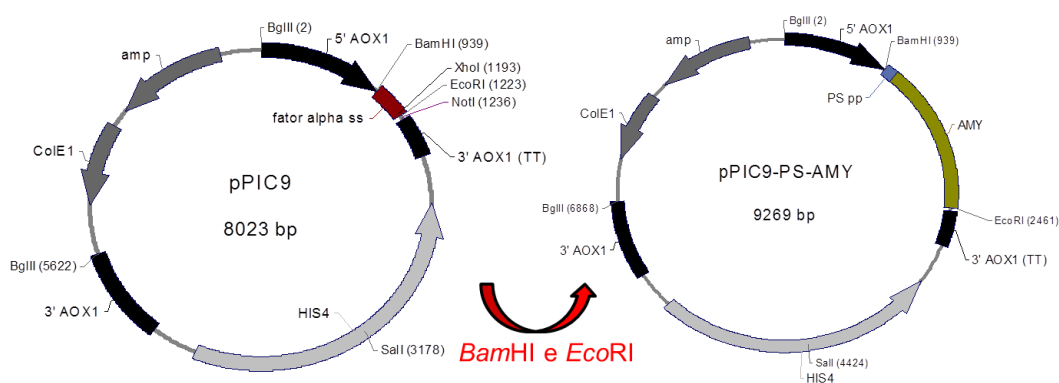


Figura 10. Estratégia de clonagem dos 13 PS fusionados ao gene *amyE*. PSpp: peptídeo-sinal de *P. pastoris*. Fator alpha ss: sequência relativa ao sinal de secreção presente no vetor pPIC9 (fator α de *S. cerevisiae*).

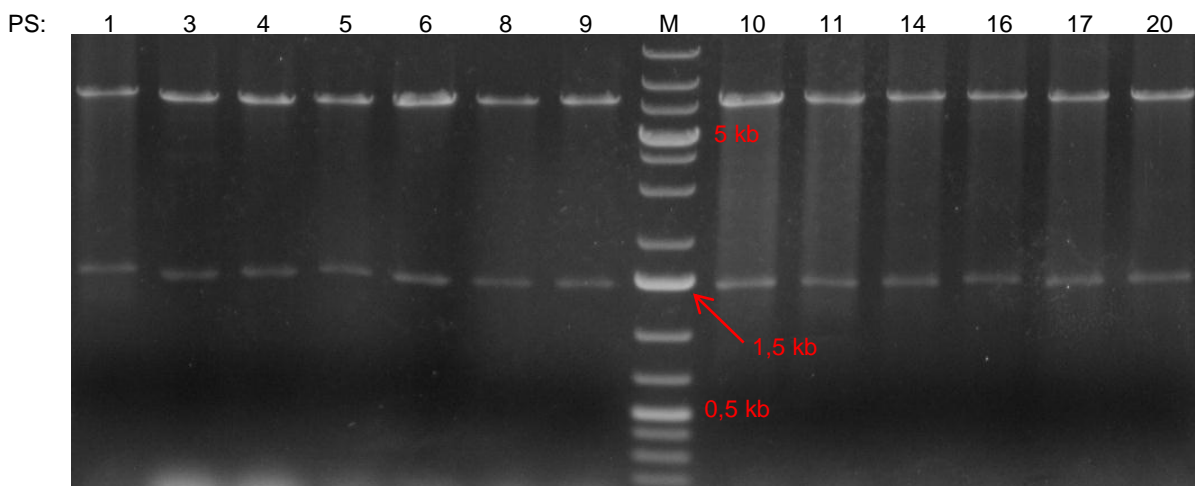


Figura 11. Clonagem dos cassetes PS + *amyE* no vetor pPIC9. Análise de restrição do vetor pPIC9 por eletroforese em gel de agarose 1%, corado com 0,5 µg/mL de brometo de etídeo, indicando os treze produtos de PCR (PS + *amyE*) clonados no vetor. Vetores digeridos com *Bam*HI e *Eco*RI, liberando o fragmento do inserto (~ 1,5 kb = PS + *amyE*) e do vetor (~ 8,0 kb). Marcador: O'Gene Ruler 1 kb plus (Fermentas).

Na digestão com *Xho*I e *Not*I, o vetor pPIC9 não perde o seu sinal de secreção original e o gene *amyE* sem PS pode ser clonado com essas enzimas. O vetor resultante pôde, então, ser utilizado como controle positivo e os PS utilizados neste trabalho puderam ser comparados à sequência relativa ao fator α de *S. cerevisiae*, com relação ao nível de secreção. Para testar a sequência relativa ao fator α de *P. pastoris*, o vetor utilizado como controle positivo foi digerido com *Bam*HI e *Eco*RI para a troca do sinal de secreção presente no pPIC9 pelo fator α de *P. pastoris* (Figuras 12 e 13).

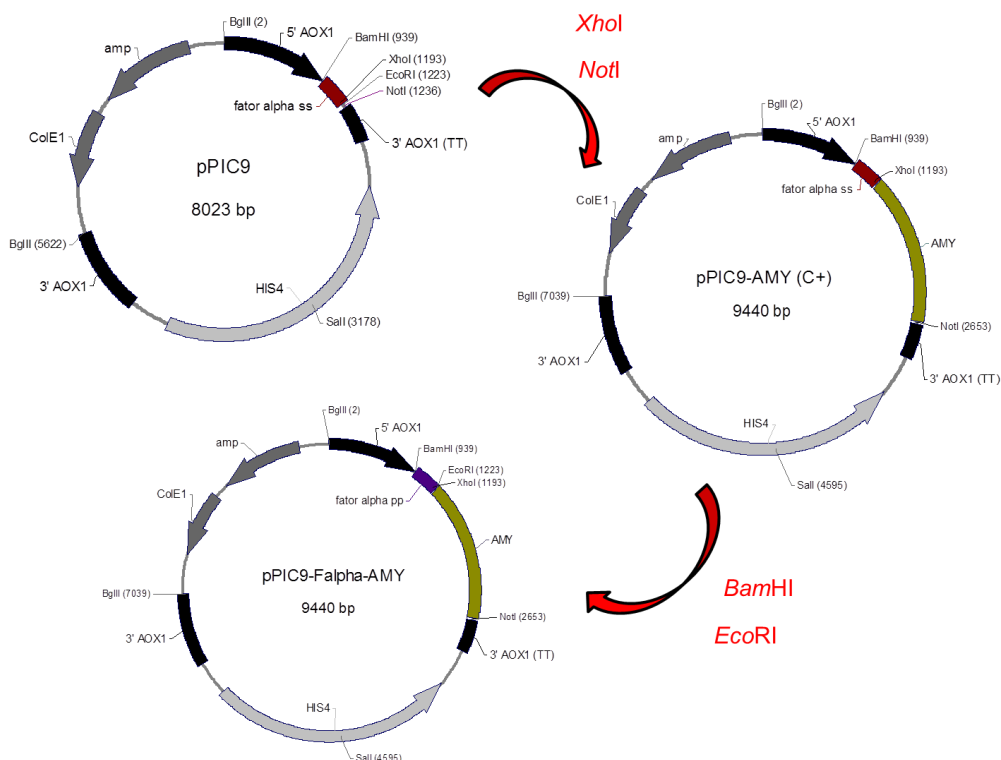


Figura 12. Estratégia de clonagem do gene *amyE* sem PS e do fator α de *P. pastoris*. Fator alpha ss: sequência relativa ao sinal de secreção presente no vetor pPIC9 (fator α de *S. cerevisiae*). Fator alpha pp: sequência relativa ao fator α de *P. pastoris*.

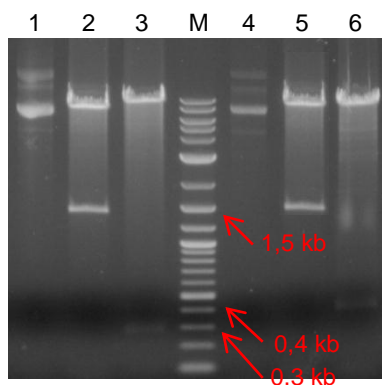


Figura 13. Clonagem do gene *amyE* sem PS e do fator α de *P. pastoris* no vetor pPIC9. Análise de restrição do vetor pPIC9 por eletroforese em gel de agarose 1%, corado com 0,5 $\mu\text{g/mL}$ de brometo de etídeo, indicando os dois produtos de PCR (fator α e *amyE* sem PS) clonados no vetor. 1: pPIC9-AMY intacto. 2: pPIC9-AMY digerido com *XhoI* e *NotI* (liberando o fragmento *amyE* de *B. subtilis* sem peptídeo sinal de tamanho aproximado de 1,45 kb). 3: pPIC9-AMY digerido com *BamHI* e *EcoRI* (liberando o fragmento da sequência do fator α de *S. cerevisiae* de tamanho aproximado de 0,3 kb). 4: pPIC9-AMY-F α intacto. 5: pPIC9-AMY-F α digerido com *XhoI* e *NotI* (liberando o fragmento *amyE*). 6: pPIC9-AMY-F α digerido com *BamHI* e *EcoRI* (liberando o fragmento da sequência do fator α de *P. pastoris* de tamanho aproximado de 0,4 kb). M: Marcador 2-log (New England Biolabs).

Análise de integração do cassete de expressão contendo a sequência dos peptídeos sinal selecionados

Depois de concluída as clonagens, os vetores PIC1 (pPIC9 + PS1 + *amyE*), PIC3, PIC4, PIC5, PIC6, PIC8, PIC9, PIC10, PIC11, PIC14, PIC16, PIC17, PIC20, PIC-alpha (pPIC9 + fator α de *P. pastoris* + *amyE*) e C+ (pPIC9 + *amyE*) foram utilizados para transformação de *P. pastoris*, linhagem GS-115, após serem linearizados com a enzima *SalI*.

Os cassetes foram integrados com sucesso e a confirmação foi realizada por *screening* de atividade amilolítica em placa. Após o *screening* dois clones transformantes de cada vetor foram escolhidos para estoque e ensaios enzimáticos (Figura 14).

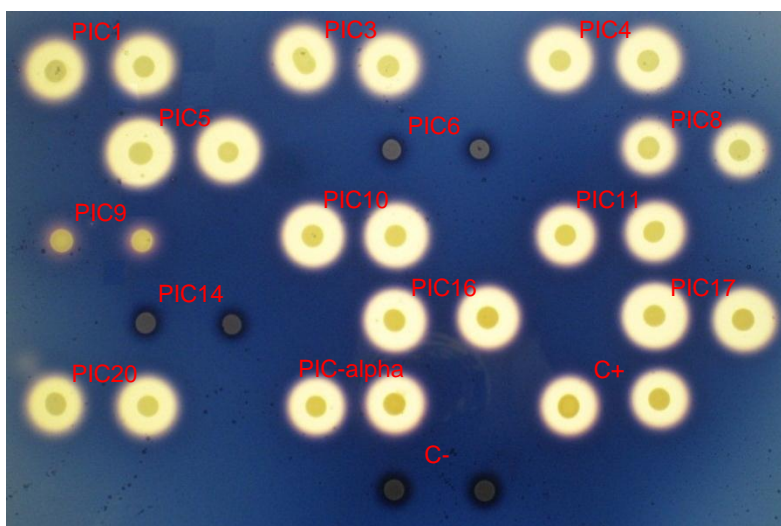


Figura 14. Análise fenotípica em placa de transformantes de *P. pastoris* expressando α -amilase. Placa de meio mínimo com amido 1%, corada com vapor de iodo, mostrando os halos de hidrólise dos clones selecionados. Controle negativo (C-): clone *P. pastoris* transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Como esperado, todos os clones exibiram halo de hidrólise com exceção dos transformados com o PIC6 e o PIC14. Estes não apresentaram atividade amilolítica, pois o erro na síntese do *primer* induziu mutação do tipo *frameshift*. O que impossibilitou a produção de uma amilase funcional.

Qualitativamente, é possível notar que no ensaio em placa já é possível identificar que alguns sinais de secreção se destacam mais que outros. O peptídeo-sinal 9 foi o que aparentemente apresentou menor eficiência na secreção de amilase.

Análise e identificação dos melhores sinais de secreção (teste piloto)

Após a análise qualitativa em placa da secreção, uma cinética de crescimento e de produção enzimática dos clones amilolíticos selecionados foi realizada em meio líquido (BMMY – metanol 0,5%), a fim de identificar quais peptídeos sinais possuíam maior potencial de secreção da amilase.

Segundo a cinética de crescimento, observou-se que os diferentes clones cresceram semelhantemente. Isso corrobora a ideia de que não há influência significativa dos peptídeos-sinais nem das taxas de secreção de amilase no crescimento da levedura. As OD₆₀₀ médias atingidas por esses clones foram: 17,5 em 24 horas; 21 em 48 horas e 20 em 72 horas (Figura 15).

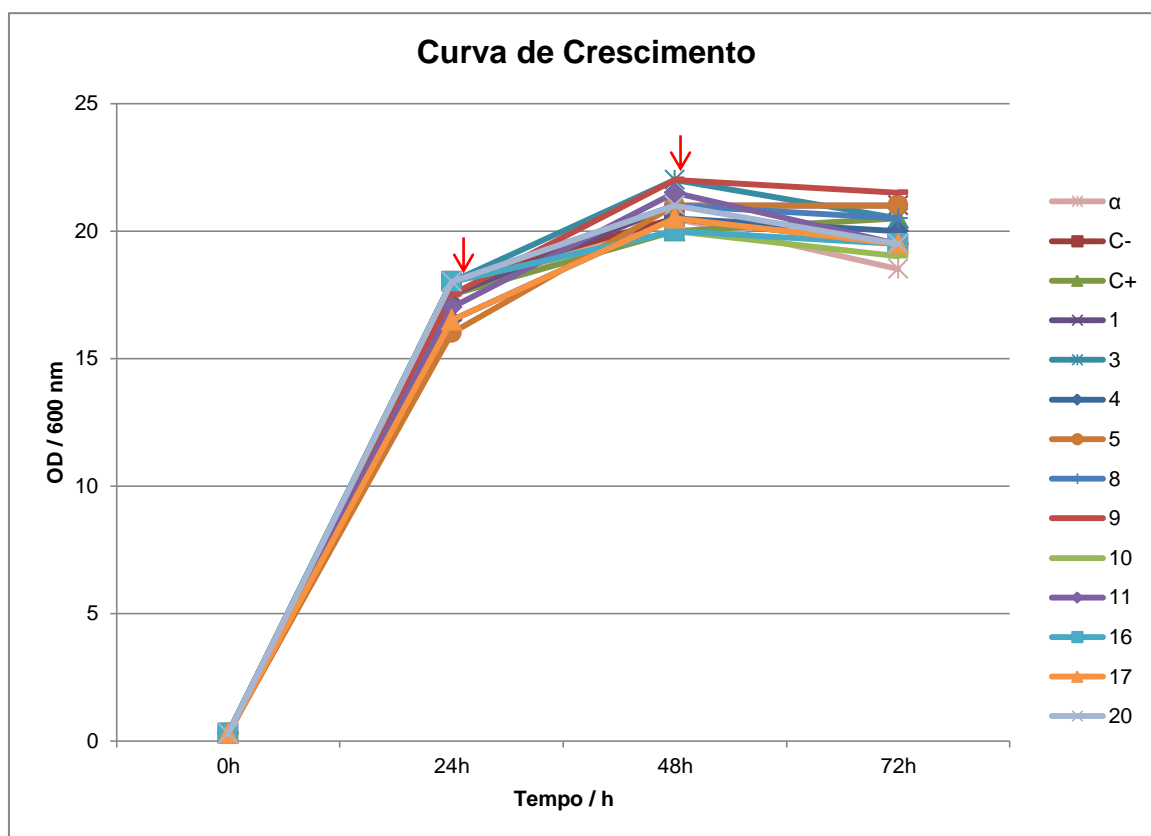


Figura 15. Curva de crescimento dos clones amilolíticos recombinantes. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção do fator α de *S. cerevisiae*, presente no vetor pPIC9. Todos os clones começaram a cinética com OD₆₀₀ de 0,3. Setas vermelhas indicam o ponto de adição de 0,5% de metanol.

Segundo a cinética de produção enzimática no sobrenadante, observou-se que diferentes clones contendo diferentes sequências de peptídeos sinais exibem diferentes perfis secretórios. Entre quais se destacaram os dos clones com as sequências do PS5, PS10, PS16, PS17 e PS α (Figura 16).

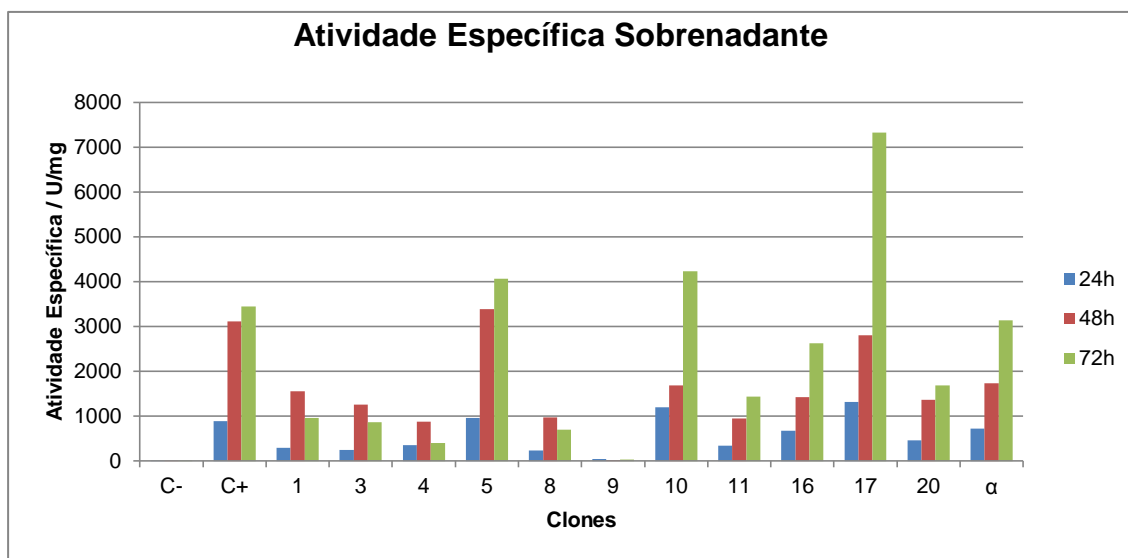


Figura 16. Atividade amilolítica específica do sobrenadante de cultura dos clones recombinantes. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção do fator α de *S. cerevisiae*, presente no vetor pPIC9.

As taxas de secreção de amilase podem ter sido influenciadas por expressão diferenciada do gene heterólogo induzida pelos diferentes peptídeos-sinais. Entretanto a integração de todos os cassetes de expressão no mesmo *locus* (*HIS4*) e, conseqüentemente, no mesmo contexto genético visou minimizar essa influência, tendo em vista que não há relatos na literatura de que sinais de secreção influenciem os níveis de expressão de um gene.

Outro ponto importante era avaliar o quanto da produção da amilase ainda se encontrava na forma intracelular. Todos os clones exibiram baixa atividade amilolítica intracelular, sendo que os que mais se destacaram, ou seja, possuíam menos atividade, foram os clones com a amilase sob controle do sinal de secreção do fator α de *S. cerevisiae* e do sinal secreção do fator α de *P. pastoris* (Figura 17).

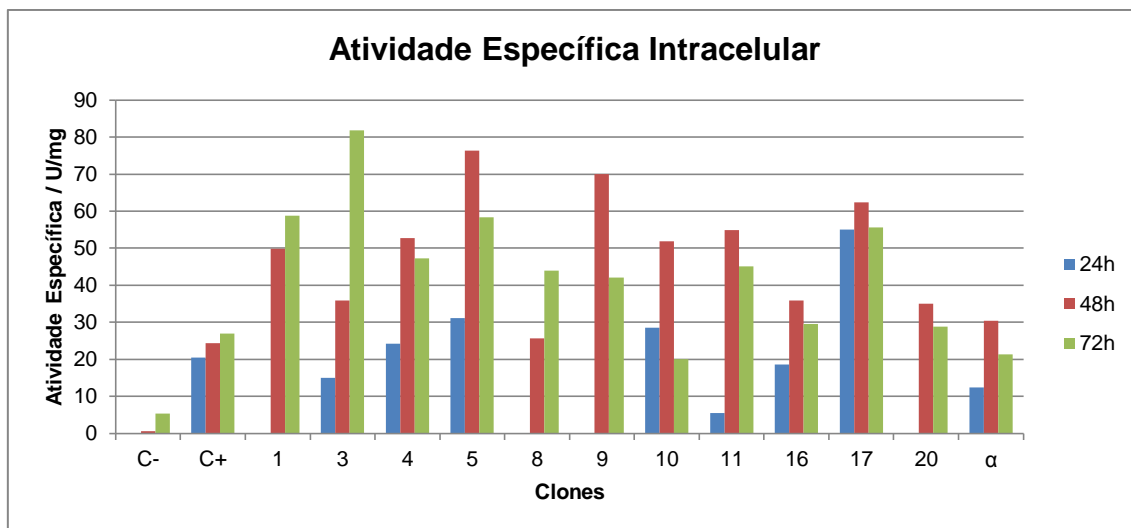


Figura 17. Atividade amilolítica específica do extrato intracelular de cultura dos clones recombinantes. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Os valores mais altos de atividade amilolítica intracelular exibidos por alguns clones podem ser explicado por: 1) baixa capacidade de secreção da amilase, no caso dos clones que secretavam pouca amilase; ou 2) por sobrecarga da maquinaria de secreção, no caso dos clones de secretavam muita amilase (ex: clones transformados com as sequências do PS5, PS10, PS16, PS17 e PS α).

Por terem exibidos níveis secretórios semelhantes ou maiores que o controle positivo, os clones transformados com as sequências do PS5, PS10, PS16, PS17 e PS α foram selecionados para continuar o trabalho.

Análise do perfil secretório por SDS-PAGE

Para se analisar o perfil de secreção dos clones, uma amostra de 250 μ L do sobrenadante de cultura no ponto de 72 horas foi precipitada e aplicada em gel de poliacrilamida desnaturante.

O perfil de bandas corrobora o que foi observado nos ensaios enzimáticos, sendo que os clones com as sequências do PS5, PS10, PS16, PS17 e PS α exibiram as bandas mais intensas de amilase e novamente se mostraram os mais secretores dessa enzima (Figura 18).

Por terem exibidos níveis secretórios semelhantes ou maiores que o controle positivo, os clones com os PS5, PS10, PS16, PS17 e PS α foram selecionados para continuar o trabalho.

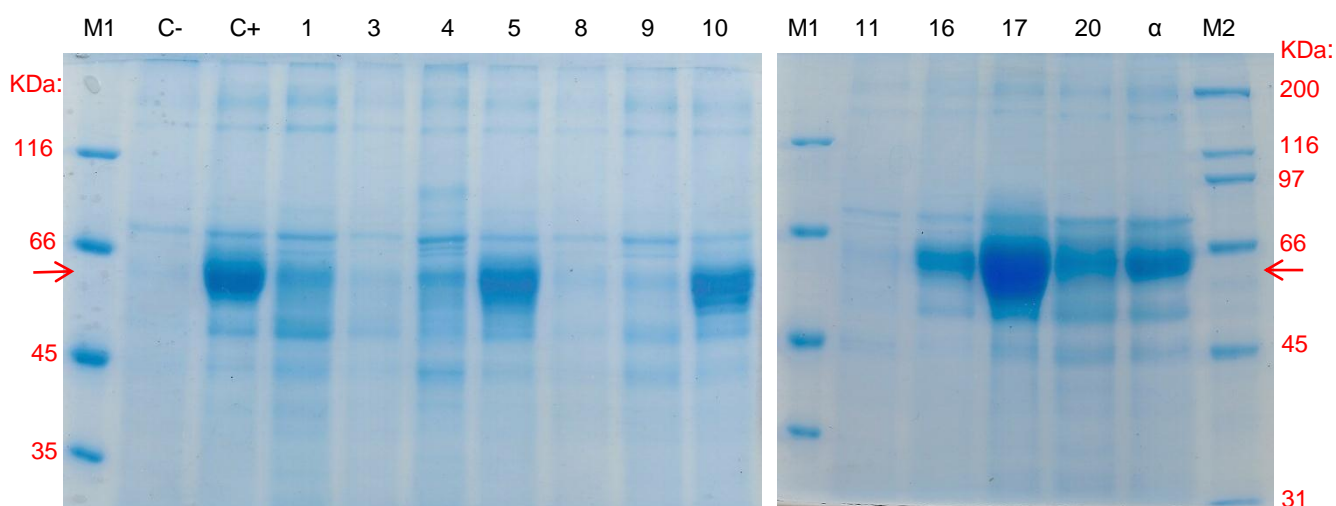


Figura 18. Perfil secretório dos clones amilolíticos. Análise feita por SDS-PAGE 12%, corado com Comassie Blue. M1: Marcador Unstained Molecular Weight (Fermentas). M2: Marcador Broad Range (Biorad). Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9. A banda correspondente à amilase (~ 60 kDa) está indicada por uma seta vermelha.

Análise e identificação dos melhores sinais de secreção (teste definitivo)

Os clones transformados com as sequências do PS5, PS10, PS16, PS17 e PS α foram induzidos novamente, desta vez em triplicata, com a finalidade de se caracterizar melhor o crescimento e a eficiência de secreção de cada um deles em referência ao controle positivo. Os ensaios foram realizados após 48 horas de crescimento, tendo em vista que após este tempo, o crescimento das células da cultura já atingia um platô (Figura 15).

Segundo a biomassa final, não houve diferença significativa no crescimento dos clones em 48 horas, indicando novamente que as diferentes sequências de peptídeos-sinais não influenciam o crescimento da levedura (Figura 19). A OD_{600nm} média atingida neste ponto foi 20,7.

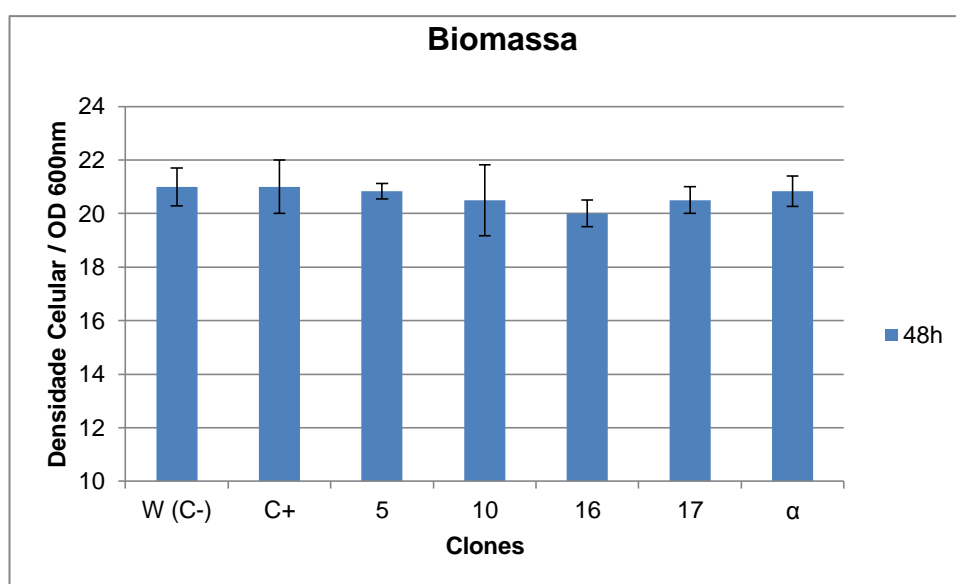


Figura 19. Densidade óptica celular (OD_{600nm}) das culturas dos melhores clones amilolíticos em 48 horas de indução com metanol. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Segundo a eficiência de secreção, todos os clones exibiram níveis semelhantes. Com exceção do clone com PS17 que apresentou cerca de duas vezes mais atividade quando comparado ao controle positivo (Figura 20).

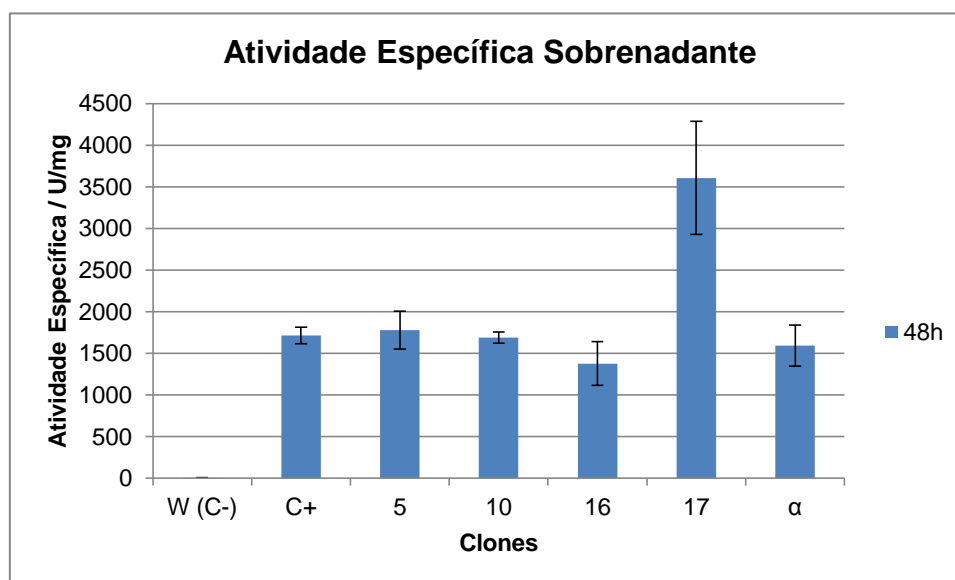


Figura 20. Atividade específica do sobrenadante de cultura dos melhores clones amilolíticos. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Quanto à atividade de amilase intracelular, os que mais se aproximaram do controle positivo foram os clones com PS16 e α (Figura 21).

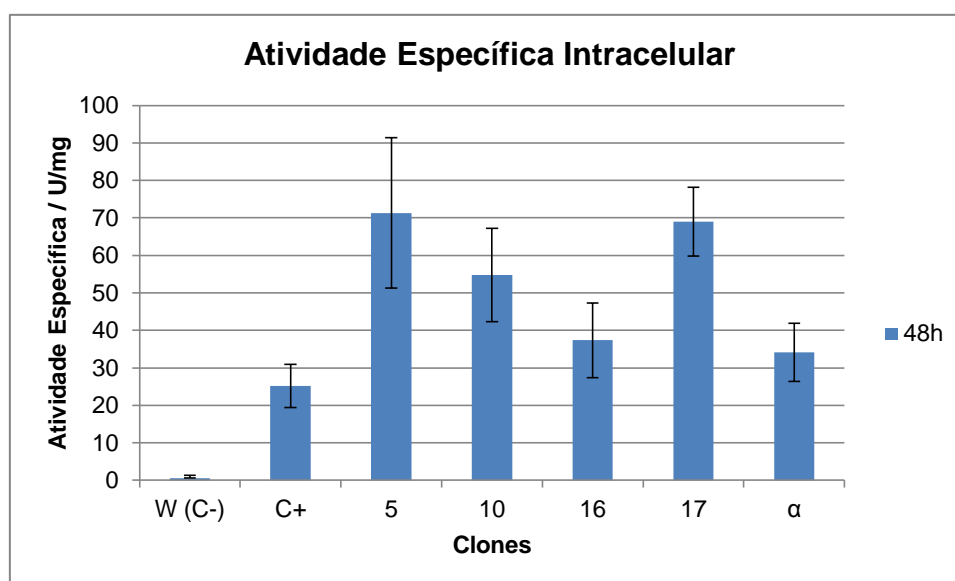


Figura 21. Atividade específica do extrato intracelular de cultura dos melhores clones amilolíticos. Controle negativo (C-): clone *P. pastoris* GS115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Determinação do número de cópias integradas do cassete de expressão por Southern Blotting

A técnica de Southern Blotting foi aplicada para quantificar o número de cópias integradas dos cassetes de expressão nos clones transformados com as sequências do PS5, PS10, PS16, PS17 e PS α . Assim foi possível normalizar as atividades enzimáticas observadas nesses clones. Segundo o padrão esperado (Figura 22) verificou-se que todos os clones exibiam uma cópia integrada, menos o com PS17 que possuía duas versões do gene da amilase em seu genoma (Figura 23).

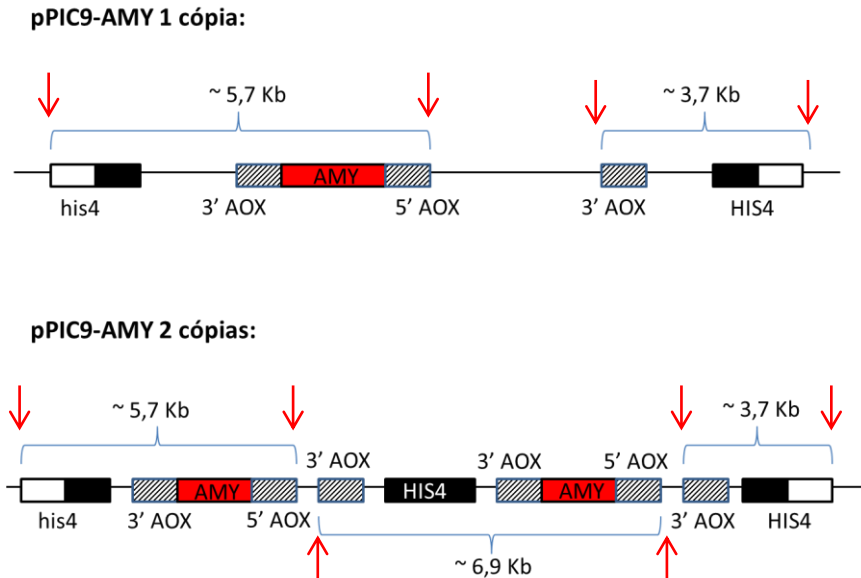


Figura 22. Esquema do locus *HIS4* de *P. pastoris* recombinante. Locus *HIS4* nativo (branco) após integração por recombinação homóloga do vetor pPIC9-*amyE* em uma ou duas cópias. Sítios de *Bgl*II indicados por setas vermelhas.

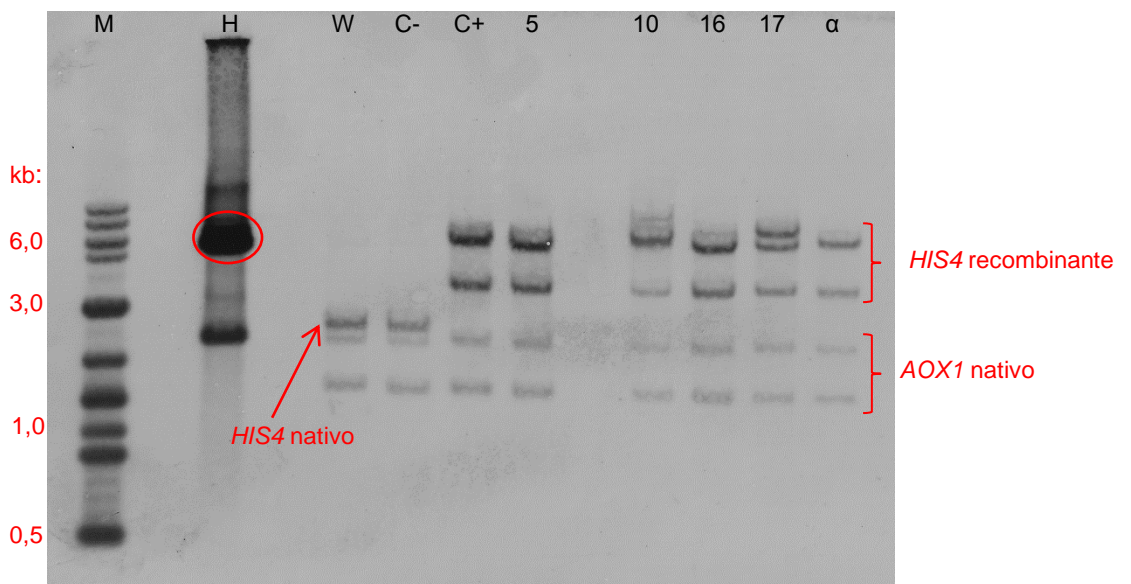


Figura 23. Análise de integração por Southern Blotting em clones amilolíticos selecionados. M: Marcador 2 log (New England Biolabs). H: pPIC9 digerido com *Bgl*II usado como controle interno do experimento, fragmento utilizado como sonda circulado em vermelho. W: *P. pastoris* GS-115 "selvagem". Controle negativo (C-): clone *P. pastoris* GS-115 transformado com pPIC9 sem *amyE*. Controle positivo (C+): clone com a amilase sob controle do sinal de secreção fator α de *S. cerevisiae* nativo do vetor pPIC9.

Como esperado, a maioria dos clones apresentou apenas uma cópia integrada como é comum no *locus HIS4*. O fato do clone com PS17 ter exibido duas cópias integradas dificulta a sua análise como sinal de secreção.

Clones de *P. pastoris* costumam exibir níveis crescentes de produção de proteína heteróloga quando se aumentam o número de cópias de um gene em seu genoma (Cereghino & Cregg, 2000; Daily & Hearn, 2005). Entretanto essa relação não é necessariamente linear e positiva. Algumas vezes o aumento do número de cópias não influencia ou até diminui a produção da proteína de interesse e clones com mais de um determinado número de cópias podem ter a eficiência de produção prejudicada (Inan *et al.* 2006; Zhu *et al.* 2009).

Com base nisso, não é possível prever o comportamento das duas cópias integradas no clone com PS17. Elas podem estar aumentando ou diminuindo a produção da amilase e com isso a determinação da eficiência de secreção deste clone fica prejudicada. Para avaliar corretamente o efeito desse sinal de secreção sobre a amilase foi necessário confirmar o número de cópias presente nesse clone e encontrar outro com apenas uma cópia desse gene.

Confirmação do número de cópias integradas do cassete de expressão por PCR em tempo real (qPCR)

A fim de validar o resultado obtido no Southern Blotting, o número de cópias do clone com PS17 foi confirmado via qPCR (Figura 24).

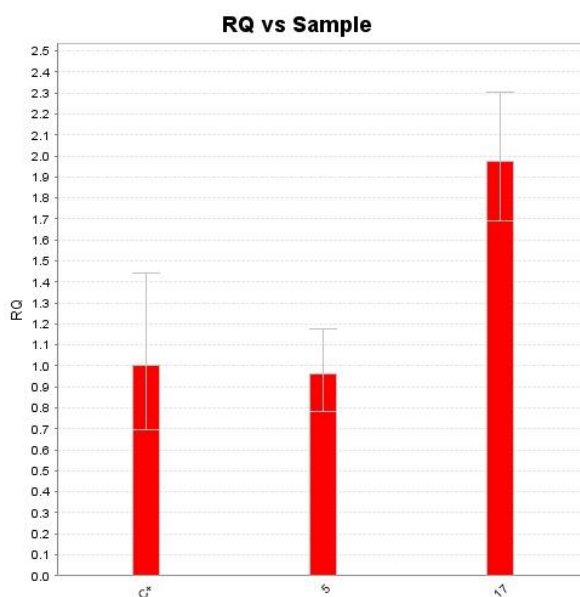


Figura 24. Determinação do número de cópias integradas por qPCR dos clones C+, 5 e 17. Referências internas para uma cópia: gene GAP de *P. pastoris* e gene *amyE* do clone C+.

Screening de clones com PS17 com uma cópia integrada do cassete por PCR em tempo real (qPCR)

Por ser necessário ao trabalho, foi realizado uma procura por clones transformados com o vetor PIC17 que tivessem uma cópia do cassete de expressão integrado.

Como dito anteriormente, após a transformação com os vetores, foram estocados dois clones de cada transformação (Figura 14). O clone 17.1 foi utilizado nas induções e nele foram verificadas duas cópias integradas. Para o *screening* de clones transformados com o vetor PIC17 foram utilizados o clone 17.2 (estocado) e o clone 17.8 recuperado da placa da transformação armazenada a 4°C (Figura 25). A estricagem do clone 17.8 originou duas colônias, sub-denominadas 17.8(1) e 17.8(2).

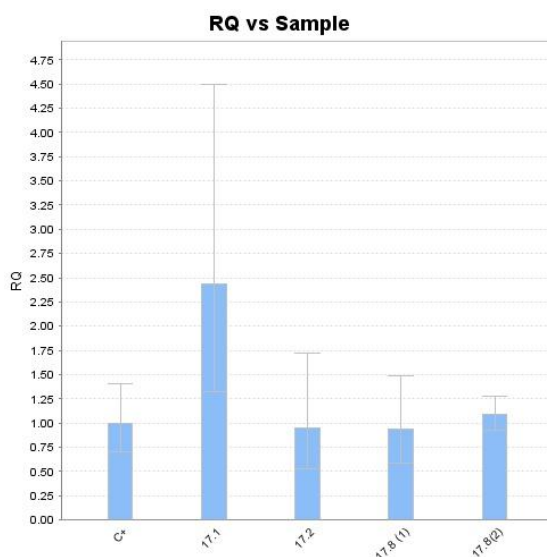


Figura 25. Determinação do número de cópias integradas por qPCR dos clones C+, 17.1, 17.2, 17.8 (1) e 17.8 (2). Referências internas para uma cópia: gene GAP de *P. pastoris* e gene *amyE* do clone C+.

Como esperado, os clones utilizados nesta análise (17.2 e 17.8) exibiram uma cópia do cassete de expressão integrado.

Conclusões

A partir dos dados parciais obtidos, conclui-se que o secretoma de *P. pastoris* crescida em glicose apresenta pelo menos três proteínas com peptídeos-sinais que exibem eficiência secretória próxima ou semelhante ao usual fator- α de *S. cerevisiae*. Essas sequências sinal podem ser utilizadas como sinais de secreção em um novo vetor para expressão de proteínas heterólogas, permitindo novas opções além das sequências disponíveis no mercado e controladas pela empresa Invitrogen (EUA).

O fator- α de *P. pastoris* proposto por Küberl *et al.* (2011) e testado no presente trabalho também apresentou níveis de secreção semelhantes ao fator α de *S. cerevisiae*. Entretanto, é possível que o seu uso, assim como o das outras sequências mostradas neste trabalho, traga maiores benefícios quanto ao processamento do N-terminal da proteína heteróloga, por se tratar de uma sequência sinal nativa de *Pichia*. O correto processamento do N-terminal permitirá a produção de produtos heterólogos mais fiéis, o que é importante no caso de biofármacos.

A atuação do peptídeo sinal 17 com o clone com duas cópias do cassete de expressão não pode ser comparada com os demais clones, pois este clone possui duas cópias integradas do gene repórter. Portanto, para se averiguar corretamente a eficiência deste sinal de secreção é necessário que se obtenham resultados de atividade em meio líquido com um clone com apenas uma cópia integrada.

Os peptídeos-sinais 6 e 14 não puderam ser analisados devido a deleções, geradas durante a síntese dos *primers*, que causaram *frameshift*.

O presente trabalho é pioneiro na utilização de dados de secretoma de *P. pastoris* para o desenvolvimento de sistemas de expressão baseados em novos sinais de secreção. Além de novos produtos biotecnológicos, ele disponibiliza mais informações acerca de genômica e proteômica de *Pichia*. Essas informações podem auxiliar novas descobertas acerca desta levedura e permitir maior eficácia e eficiência para produção biotecnológica, usando *P. pastoris* como sistema de expressão.

Perspectivas

As metas principais agora são reamplificar o PS6 e PS14 junto com o gene *amyE*, sem mutações, com *primers* novos e concluir as etapas de clonagem e transformação com os novos vetores construídos. Além de finalizar as atividades em meio líquido com o clone 17 com apenas uma cópia do cassete de expressão.

Confirmado os melhores sinais de secreção, as amilases com essas sequências sinal serão purificadas em coluna de níquel. Em seguida, o sequenciamento do N-terminal dessas proteínas (por espectrometria de massas) irá propiciar a observação do processamento desses sinais de secreção. O peptídeo sinal escolhido terá a sua sequência otimizada e finalizada para a sua utilização na expressão de um biofármaco (EGF humano) ou outro gene repórter.

Se a expressão, secreção e processamento dessa outra proteína forem bem sucedidas, o sinal de secreção escolhido será adicionado a um novo vetor de *Pichia pastoris*, em construção no Laboratório de Biologia Molecular da UnB pelo Grupo de Biotecnologia de Leveduras, contendo um novo promotor (pPGK - Almeida *et al.*, 2005), novas marcas de seleção (*LEU2*, entre outras) e é claro o sinal de secreção.

Finalizado o projeto, o objetivo também é a publicação de um artigo em revista internacional de impacto.

Materiais e Métodos

1. Materiais

1.1 Microorganismos

A) Bactérias

Escherichia coli XL10-GOLD (Stratagene): *endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]*

B) Levedura

Pichia pastoris GS115 (Invitrogen ®): Apresenta mutações no gene da histidinol desidrogenase (*HIS4*), que permite a seleção auxotrófica com histidina do vetor de expressão contendo *HIS4* após a transformação.

1.2 Enzimas

A) Enzimas de restrição

Tabela 4. Enzimas de restrição utilizadas.

Enzima	Sítio de clivagem	Tampão	Temperatura de incubação	Procedência
<i>Bam</i> HI	G [^] GATCC	NEB 3	37°C	New England Biolabs
<i>Bgl</i> II	A [^] GATC	NEB 3	37°C	New England Biolabs
<i>Eco</i> RI	G [^] AATTC	NEB Buffer EcoRI	37°C	New England Biolabs
<i>Not</i> I	GC [^] GGCCGC	NEB3/BSA	37°C	New England Biolabs
<i>Sal</i> I	G [^] TCGAC	NEB 3	37°C	New England Biolabs
<i>Xho</i> I	C [^] TCGAG	NEB4/BSA	37°C	New England Biolabs

B) Enzima T4 DNA ligase (New England Biolabs): utilizada na concentração de 1U/ μ L.

C) RNase A:

RNase A (*DNase-free*) 10 mg/mL

Acetato de sódio pH4,8 50 mM

Fervida em banho-maria por 10 minutos.

D) Phusion® High-Fidelity DNA Polymerase (New England Biolabs): Utilizada na concentração de 1 U/reacção

1.3 Marcadores moleculares

A) Marcadores de DNA

- Marcador O' Gene Ruler 1kb Ladder, ready-to-use, 75 - 20.000 pb (Fermentas).
- Marcador 2-log DNA Ladder, 100 - 10.000 pb (New England Biolabs).

B) Marcadores de proteína

- Marcador Unstained Molecular Weight; 14,4 - 116 kDa (Fermentas).
- Marcador Broad Range; 6,5 - 200 kDa (Biorad).

1.4 Kits

A) QIAquick PCR purification (Qiagen): Utilizado para purificar produtos de PCR.

B) Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen): Utilizado para clonar produtos de PCR blunt-end.

C) TOPO® XL PCR Cloning Kit (Invitrogen): Utilizado para clonar produtos de PCR com cauda de adenina.

D) QIAquick Gel Extraction (Qiagen): Utilizado para purificar e eluir fragmentos de DNA em gel de agarose.

E) Wizard® plus SV Minipreps DNA purification (Promega): Utilizado para da purificação do DNA plasmidial.

- F) Qiagen Plasmid Midi (Qiagen):** Utilizado para purificação do DNA plasmidial em maior escala.
- G) DIG DNA Labelling and Detection (Roche):** Utilizado para marcação, com digoxigenina-dUTP, da sonda usada no Southern Blot e para detecção dos híbridos por imunoenensaio enzimático.
- H) AlkPhos Direct™ DNA Labelling (Amershan):** Utilizado nas etapas de pré-hibridização e hibridização do Southern Blotting.
- I) Fast SYBR® Green PCR Master Mix (Applied Biosystems):** Utilizado para a reação de PCR em tempo real (qPCR).
- J) Protein Assay (Biorad):** Utilizado para quantificação de proteínas totais com o corante Coomassie ® Brilliant Blue G-250.

1.5 Ferramentas de bioinformática

A) Busca de sequências

GenBank (NCBI) – <http://www.ncbi.nlm.nih.gov>

B) Alinhamento de sequências

BLAST (NCBI) - <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

C) Manipulação de sequências

Molecular Toolkit - <http://www.vivo.colostate.edu/molkit/index.html>

D) Análises de restrição

NEBcutter V2.0 – <http://tools.neb.com/NEBcutter2/index.php>

E) Predição de sítios de clivagem de peptídeos sinal

SignalP 3.0 Server - <http://www.cbs.dtu.dk/services/SignalP/>

F) Plataforma genômica de *Pichia pastoris*

Bioinformatics Online Genome Annotation System (BOGAS) - <http://bioinformatics.psb.ugent.be/webtools/bogas/>

G) Análise de sequências e eletroferogramas

Ferramenta *PHPH* - <http://helix.biomol.unb.br/phph/index.html>

H) Análise de eletroferogramas

Software Chromas Lite versão 2.01, Technelysium Pty Ltd.

1.6 Primers

Os *primers* desenhados com base na sequência do gene da α -amilase de *Bacillus subtilis* (depositada no GenBank - nº de acesso AB015592.1) foram sintetizados pela IDT (Integrated DNA Technologies, Inc. - EUA) e apresentam sítio de *Bam*HI e *Eco*RI em suas extremidades. O *primer forward* possui 13 versões, sendo cada um fusionado com um peptídeo-sinal diferente (tabela 5).

Tabela 5. Primers utilizados para amplificar o gene *amyE* fusionado em fase com os 13 peptídeos sinais selecionados.

Primer	Tamanho (nt)	Sequência 5'-3'
PPPS211 (PS1)	101	CGGATCC CATATGAACTTATTAAGTCTAACCTTACTATTGTTCACTACTGAAGCACTCGCCCTCACC GGTGAAGCT GAAACTGCAAACAAATCGAATAAGG
PPPS436 (PS3)	101	CGGATCC AATATGAAATCTCAACTTATCTTTATGGCTCTTGCCTCTCTGGTGGCCTCCGCTCCTTTA GAACAC GCTGAAACTGCAAACAAATCGAATAAGG
PPPS545 (PS4)	107	CGGATCC CCAATGAAGATATCCGCTCTTACAGCCTGCGCTGTTACTCTAGCTGGTCTTGCAATTGC AGCACCAGCTCCAG CTGAAACTGCAAACAAATCGAATAAGG
PPPS748 (PS5)	107	CGGATCC AATTATGAAGCTCGCTGCACTCTCCACTATTGCATTAACTATTTTACCCGTTGCCTTGGCT GGCTATGCTCCAG CTGAAACTGCAAACAAATCGAATAAGG
PPPS934 (PS6)	110	CGGATCC CCAATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTC AGCTGCTCCAGTTGCT GCTGAAACTGCAAACAAATCGAATAAGG
PPPS1008 (PS8)	104	CGGATCC CAATATGTTTAAATCTCTGTGCATGTTAATAGTTTCTGCCTATTGAGTTCAGTCTTGGCA GCAGATTT CTGCTGAAACTGCAAACAAATCGAATAAGG
PPPS1010 (PS9)	104	CGGATCC CGCAATGTTGTCCATTTTAAAGTGCATTAACTCTGCTGGCCTGTCTTGTGCTTCAGATCTG ACCCCACCT GCTGAAACTGCAAACAAATCGAATAAGG
PPPS1223 (PS10)	110	CGGATCC AATATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGCATCCCTAC AGGCTCCAATACCT GCTGAAACTGCAAACAAATCGAATAAGG
PPPS1958 (PS11)	116	CGGATCC AAGATGATCTTTAATCTTAAACACTGGCTGCGGTTGCAATCTCCATTTACAAGTGTCT GCAGTTTCTCTCTGGGTTTT GCTGAAACTGCAAACAAATCGAATAAGG
PPPS2524 (PS14)	107	CGGATCC GGGATGAACTTGTACCTAATTACATACTATTGCCAGTCTATGCAGCGCAATTACTCTC CAAAGAGAGAT GCTGAAACTGCAAACAAATCGAATAAGG
PPPS3955 (PS16)	128	CGGATCC CAATGTCATTCTTCCAACGTGCCACAACCTTTTCTTGTGTTGTTCTGTTGACCAAT ATAGTCAGTGGAGCAGTCATTTAGTATGGTCC GCTGAAACTGCAAACAAATCGAATAAGG
PPPS4722 (PS17)	107	CGGATCC ACCATGCAAGTTAAATCTATCGTTAACCTACTGTTGGCATGTTGTTGGCCGTGGCCAG ACCTTTGGAGCAC GCTGAAACTGCAAACAAATCGAATAAGG
PPPS5771 (PS20)	119	CGGATCC TAAATGACTCAAAAGACACAGTATATAATCATTTTGTGACTCTGATAGTTTCAGCTTTCT CCGAAAGCTCGGGTCAAAATGCT GCTGAAACTGCAAACAAATCGAATAAGG
AMYECO-R	51	T GAATTC TTAATGATGATGATGATGATGATT GAAAGAATGGACTGCCCCCGT

Obs: Todos os *primers* são *forward*, menos o *primer* AMYECO-R que é *reverse* para todos os outros. Negrito: região de anelamento na α -amilase. Vermelho: sítio de *Bam*HI. Verde: sítio de *Eco*RI. Azul: His-Tag.

Os *primers* desenhados com base na sequência do gene *amyE* de *B. subtilis* (depositada no GenBank - nº de acesso AB015592.1) foram sintetizados pela IDT (Integrated DNA Technologies, Inc. - EUA) e apresentam sítio de *XhoI* e *NotI* em suas extremidades (tabela 6).

Tabela 6. Primers utilizados para amplificar o gene da α -amilase sem peptídeo-sinal de *P. pastoris*.

Primer	Tamanho (nt)	Sequência 5'-3'
AMY-XHO	59	CCTCGAG AAAAGAGAGGCTGAAGCT GAATTC GCTGAAACTGCAAACAAATCGAATAAGG
AMY-NOT	54	T GCGGCCGC TTAATGATGATGATGATGATGATT GAAAGAATGGACTGCCCCCGT

Obs: AMY.XHO é o *primer forward* e o AMY.NOT é o *primer reverse*. Negrito: região de anelamento na α -amilase. Vermelho: sítio de *XhoI*. Verde: sítio de *EcoRI*. Azul: sítio de *NotI*.

Os *primers* desenhados com base na sequência do gene do fator α de *P. pastoris* (depositada no GenBank - nº de acesso FR839629.1, protein ID: CCA38209.1) foram sintetizados pela IDT (Integrated DNA Technologies, Inc. - EUA) e apresentam sítio de *BamHI* e *EcoRI* em suas extremidades (tabela 7).

Tabela 7. Primers utilizados para amplificar o a região “pré-pró” do fator- α de *P. pastoris*.

Primer	Tamanho (nt)	Sequência 5'-3'
PP0522-F	37	GGGATCC AAGCAATGAAATCACTCATTGAAACATCA
PP0522-R	40	G GAAATTC GCTTCTGCTTCTCTT GCCAAATAGCGGTT

Obs: PP0522-F é o *primer forward* e o PP0522-R é o *primer reverse*. Negrito: região de anelamento no fator- α . Vermelho: sítio de *BamHI*. Verde: sítio de *EcoRI*. Azul: contexto de Kozak.

Os *primers* para PCR em tempo real (qPCR), desenhados com base nas sequências do gene da *amyE* de *Bacillus subtilis* (depositada no GenBank - nº de acesso AB015592.1) e do gene da gliceraldeído-3-fosfato desidrogenase (GAP) de *Pichia pastoris* (depositada no

GenBank - nº de acesso U62648.1), foram sintetizados pela Eurofins (MWG – Operon - Alemanha). Vide tabela 8.

Tabela 8. Primers utilizados para amplificar fragmentos do gene GAP e amyE na qPCR.

Primer	Tamanho (nt)	Sequência 5´-3´
qGAP-F	21	GTCGGTGTCAACGAGGAGAAA
qGAP-R	22	TGGAGTGGACGGTGGTCATCAA
qAMY-F	20	TATTCGTGATGCGGGCTATG
qAMY-R	20	TAACGGTTGCCGATTTGGTA

1.7 Vetores

Tabela 9. Vetores utilizados em *E. coli* e *P. pastoris*.

Vetor	Tipo de vetor	Tamanho	Característica principal	Origem
pCR Blunt II TOPO	Vetor de clonagem	3,5 kb	Vetor de clonagem de produto de PCR blunt	Invitrogen
pCR XL TOPO	Vetor de clonagem	3,5 kb	Vetor de clonagem de produto de PCR com <i>a-tailing</i>	Invitrogen
pPIC9	Vetor de integração	8 kb	Gene de seleção dominante (ampicilina) e auxotrófica (<i>HIS4</i>)	Invitrogen

1.8 Meios de cultura

A) Meios de cultivo para bactéria

Meio LB (Luria-Bertani)

Peptona	1% (p/v)
Extrato de levedura	0,5% (p/v)
NaCl	1% (p/v)
pH 7,2	

Meio LB-Ágar

Meio LB adicionado de ágar 2% (p/v)

B) Meios de cultivo para levedura

Meio YPD (Meio extrato de levedura peptona glicose)

Extrato de levedura	1% (p/v)
Peptona	2% (p/v)
Glicose	2% (p/v)

Meio YPD-ágar

Meio YPD adicionado de ágar 2% (p/v)

MD-ágar (Meio mínimo com glicose e ágar)

Biotina	$4 \times 10^{-5}\%$
YNB	1,34%
Glicose	2% (p/v)
Ágar	2% (p/v)

MM-Ágar (Meio mínimo com metanol e ágar)

Biotina	$4 \times 10^{-5}\%$
YNB	1,34%
Metanol	0,5% a cada 24h
Ágar	2% (p/v)

MM-ágar amido 1%

Meio mínimo adicionado com 1% de Amido (Merck) tamponado com tampão Asp-Glu, pH 6,8.

BMGY (Meio complexo tamponado com glicerol)

Extrato de levedura	1% (p/v)
Peptona	2% (p/v)
Tampão Asp-Glu pH 6,8	5% (v/v)
YNB	1,34% (v/v)
Biotina	$4 \times 10^{-5}\%$ (v/v)
Glicerol	1% (v/v)

BMMY (Meio complexo tamponado com metanol)

Extrato de levedura	1% (p/v)
Peptona	2% (p/v)
Tampão Asp-Glu pH 6,8	5% (v/v)
YNB	1,34% (v/v)
Biotina	4x10 ⁻⁵ % (v/v)
Metanol	0,5% (v/v)

1.9 Extração de DNA plasmidial (bactéria)**A) Solução I (TE)**

Tris-HCl pH 8,0	10 mM
-----------------	-------

EDTA	1 mM
------	------

Diluídos em água destilada.

B) Solução II

NaOH	0,2 M
------	-------

SDS	1% (p/v)
-----	----------

Dissolvidos em água destilada.

Solução preparada para uso imediato.

C) Solução III

Acetato de sódio	3 M
------------------	-----

Acido acético	2 M
---------------	-----

pH 4,8

Dissolvidos em água destilada.

1.10 Extração de DNA total (levedura)**A) Solução SE**

Sorbitol	0,9 M
----------	-------

EDTA	100 mM
------	--------

pH 7,5

B) Solução de liticase

Liticase dissolvida em água destilada em concentração final de 20 mg/mL.

C) Tampão TE20

Tris-HCl (pH 7,5) 10 mM

EDTA 20 mM

Diluídos em água destilada.

D) Dodecil Sulfato de Sódio (SDS) 10%

Dodecil Sulfato de Sódio (SDS) dissolvido em água destilada em concentração de 10% (p/v).

E) Acetato de potássio (KOAc) 5M pH 8,9

Acetato de potássio dissolvido em água destilada em concentração final 5M e pH 8,9.

1.11 Análise de DNA em gel de agarose

A) Tampão Tris-acetato EDTA (TAE) 10X

Tris-Acetato 40 mM

EDTA 1 mM

pH 8,0

Dissolvidos em água destilada.

B) Tampão de amostra para gel de agarose 10X

TAE 20X 50% (v/v)

Glicerol 30% (v/v)

Azul de bromofenol 0,25% (p/v)

Dissolvidos em água MilliQ.

1.12 Transferência e Southern Blotting

A) Solução de Depuração

Ácido Clorídrico (HCl) diluído em água destilada em concentração de 0,25 M.

B) Solução de Desnaturação

NaOH 0,5 M

NaCl 1,5 M

Dissolvidos em água destilada.

C) Solução de Neutralização

Acetato de amônio (NH ₄ OAc)	1 M
NaOH	0,02 M

Dissolvidos em água destilada

D) Membrana de Nylon (Hybond™ N+)

Membrana de Nylon carregada positivamente (Amershan) cortada do tamanho do gel de agarose a ser transferido.

E) Papel de filtro

Papel de filtro cortado do tamanho do gel de agarose a ser transferido.

F) SSC 20X

NaCl	5M
Citrato de Sódio	1M

Dissolvidos em água destilada.

1.13 Análise de proteína em gel de poliacrilamida desnaturante (SDS-PAGE)**A) Tampão de amostra para gel de poliacrilamida 2X**

Tris-HCl 1M pH 6,8	200mM
SDS	4% (p/v)
β-mercaptoetanol	4% (p/v)
Glicerol	20% (p/v)
Azul de bromofenol	0,1% (p/v)

B) Tampão de corrida – Tris-Glicina 5X

Trizma base	125mM
Glicina	0,96M
SDS	0,5%

C) Ácido tricloroacético (TCA)

Em um frasco fechado contendo 100g de TCA foi adicionado 45,4 mL de água.

A solução resultante corresponde a TCA 100% (p/v), estocado a 4°C.

D) Deoxicolato de Sódio 2%

Em um Eppendorf (1,5 mL) foi diluído 0,02g de deoxicolato de sódio em 1 mL de água destilada. A solução resultante corresponde a deoxicolato de sódio 2% (p/v), estocado a temperatura ambiente.

E) Gel concentrador 5%

Acrilamida:Bis-acrilamida (29,2:0,8)	2,5mL
Tris-HCl 1M pH 6,8	5mL
Água destilada	12,15mL
SDS 10% (p/v)	0,2mL
APS 10% (p/v)	0,150mL
TEMED (Gibco BRL)	0,015mL

F) Gel separador 12,5%

Acrilamida:Bis-acrilamida (29,2:0,8)	12,5mL
Tris-HCl 1,5M pH 8,8	3,75mL
Água destilada	13,21mL
SDS 10% (p/v)	0,3mL
APS 10% (p/v)	0,225mL
TEMED (Gibco BRL)	0,015mL

G) Solução corante

Comassie Brilliant Blue G-250	0,25% (p/v)
Metanol	30% (v/v)
Ácido acético	7% (v/v)

H) Solução descorante

Metanol	30% (v/v)
Ácido acético	7% (v/v)

1.14 Reagentes para atividade enzimática

A) Tampão Asp-Glu

Ácido Aspártico 8% (p/v)

Ácido Glutâmico 8% (p/v)

Dissolvido em água e esterilizada por filtração.

B) Iodo sólido

Iodo granulado (Sigma-Aldrich)

C) Tampão Acetato de Sódio 0,5 M pH 6,0

Acetato de sódio dissolvido em água e pH ajustado para 6,0 com ácido acético.

Concentração final 0,5 M.

D) Solução de Amido 0,5%

Amido (Merck) dissolvido em água destilada em concentração de 0,5% (p/v)

E) Solução de Iodo 1%

Iodo granulado (Sigma-Aldrich) dissolvido em etanol ultrapuro em concentração de 1% (p/v)

F) Solução de Iodeto de Potássio (KI) 10%

Iodeto de Potássio (KI) dissolvido em água destilada em concentração de 10% (p/v).

G) Solução de FUWA [1:1:3 v/v/v]

Solução de Iodo 1% 1 parte

Solução de KI 10% 1 parte

Água destilada 3 partes

Solução preparada para uso imediato.

H) Ácido Acético 1M

Ácido Acético diluído em água destilada em concentração de 1 M.

1.15 Outras soluções

F) Ampicilina 500X

Ampicilina dissolvida em água MilliQ, em concentração de 50 mg/mL, e esterilizada por filtração.

G) Kanamicina 500X

Kanamicina dissolvida em água MilliQ, em concentração de 30 mg/mL, e esterilizada por filtração.

H) Brometo de Etídeo (EtBr)

Brometo de Etídeo (EtBr) dissolvido em água destilada em concentração de 10 mg/mL.

I) Clorofane [25:24:1 v/v/v]

Fenol 25 partes

Clorofórmio 24 partes

Álcool isoamílico 1 parte

Engarrafado em frasco âmbar.

2. Métodos**2.1 Identificação de peptídeos sinais nativos de *Pichia pastoris* e sua amplificação fusionados com a α -amilase de *Bacillus subtilis***

A partir dos dados do secretoma de *P. pastoris* (Mattanovich *et al.*, 2009), as 20 proteínas secretadas encontradas neste estudo tiveram suas sequências submetidas a uma análise de predição e localização de sítios de clivagem de peptídeos sinal utilizando o software SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>).

As sequências dos peptídeos encontradas foram incorporadas no primer forward para a amplificação do gene *amyE* de *B. subtilis* (vide tabelas 1 e 3). As reações de PCR foram constituídas de: 1X de tampão de reação; 0,2 mM de dNTP; 0,5 μ M de *primer forward* e *reverse* (tabela 3); 10 ng de DNA template (vetor Δ 3-amy contendo amilase clonada, linearizado com *SacI*) e 1 unidade da Phusion High Fidelity DNA Polymerase (New England Biolabs). As condições da reação foram: 1 ciclo de 98°C por 30 segundos (desnaturação inicial); 30 ciclos de 98°C por 10 segundos (desnaturação) seguido de 30 segundos a 65°C (anelamento) seguido de 40s a 72°C (extensão); 1 ciclo de 72°C por 10 minutos (extensão final). Os ciclos foram montados seguindo as sugestões do manual da enzima Phusion.

2.2 Amplificação da amilase (*Bacillus Subtilis*) e do fator- α (*Pichia pastoris*)

Para a amplificação do gene *amyE* de *B. subtilis* sem peptídeo sinal foram utilizados os *primers* da tabela 6. As condições da reação de PCR foram as mesmas descritas no tópico anterior.

Para amplificação da região “pré-pró” do fator α *Pichia pastoris* foram utilizados os *primers* da tabela 7. As condições utilizadas foram as mesmas descritas no tópico anterior, alterando-se apenas o DNA template para DNA genômico de *Pichia pastoris* linhagem GS115, a temperatura de anelamento para 60°C e o tempo de extensão para 30 segundos.

2.3 Análise em gel de agarose (adaptado de Sambrook, Fritsch & Maniatis, 1989)

A análise de fragmentos DNA foi feita em eletroforese em gel de agarose. O gel foi preparado na concentração de 0,8% (p/v) de agarose em tampão de corrida TAE 1X e adicionado brometo de etídeo (análogo de base intercalante na molécula de DNA, que emite fluorescência quando excitado com luz ultravioleta) na concentração final de 0,5 $\mu\text{g/mL}$. O tampão de amostra foi diluído para a concentração final de 1X na amostra analisada. As amostras e marcadores foram aplicados ao gel e submetidos à eletroforese. A análise das amostras foi feita com a exposição do gel a luz ultravioleta.

2.4 Preparação das células bacterianas competentes para choque térmico

O preparo das células de *E.coli* competentes para choque térmico foi realizado como descrito por Cohen (1972) com algumas modificações. As células da linhagem de *E.coli* desejada foram crescidas em 5mL de meio LB e incubadas a 37°C durante a noite sob agitação de 250 rpm. Em seguida 1mL do pré-inóculo foi adicionado a 30 mL de meio LB e essa cultura foi incubada a 37°C sob agitação (250 rpm) até atingir a OD₆₀₀ de 0,2 a 0,3. As células foram coletadas por centrifugação a 3000 x g por 10 minutos a 4°C e ressuspensas em 10 mL de CaCl₂ 100mM estéril e gelado. Em seguida as células foram submetidas à nova centrifugação sob as mesmas condições. Finalmente as células ressuspensas em 1,6 mL de CaCl₂ 100mM mais 0,4 mL de Glicerol estéril e foram aliqüotadas em tubos Eppendorff (50 ou 100 μL por tubo). O armazenamento se deu a -80°C.

2.5 Ligação de fragmentos de PCR no vetor pCR-Blunt II-TOPO

A ligação dos produtos de PCR no vetor pCR-Blunt II-TOPO foi realizada utilizando-se o kit Zero Blunt® TOPO® PCR Cloning (Invitrogen), de acordo com as orientações do fabricante.

2.6 Ligação de fragmentos de PCR no vetor pCR-TOPO-XL

A ligação dos produtos de PCR no vetor pCR-TOPO-XL foi realizada utilizando-se o kit TOPO® XL PCR Cloning Kit (Invitrogen), de acordo com as orientações do fabricante.

2.7 Ligação de fragmentos de DNA (vetor-inserto)

Os sistemas de ligação foram feitos de modo que a razão molar entre vetor e inserto ficasse entre 1:3 e 1:10. A enzima T4 DNA Ligase (New England Biolabs) foi utilizada com o tampão fornecido pelo fabricante os sistemas foram incubados a 16°C por pelo menos 16 horas antes de serem utilizadas na transformação de bactérias.

2.8 Transformação de *E. coli* por choque térmico

Uma alíquota de célula previamente preparada foi utilizada para cada sistema de ligação. As células foram retiradas do freezer -80°C e deixadas no gelo até que descongelassem. Então, foi adicionado 10 µL do sistema de ligação às células, que foram novamente incubadas no gelo por 30 minutos. Após esse período as células foram submetidas ao choque térmico a 37°C por 5 minutos e em seguida incubadas novamente no gelo por 2 minutos. Foram adicionados 1 mL de meio LB ao sistema que foi incubado a 37°C por 1 hora. O plaqueamento foi feito em 2 placas com meio LB adicionado do antibiótico adequado. As placas foram incubadas durante a noite a 37°C.

2.9 Preparação de DNA plasmidial em pequena escala (miniprep) (adaptado de Sambrook, Fritsch & Maniatis, 1989)

Uma colônia de bactéria foi inoculada em 3 mL de meio LB com o antibiótico adequado e cultivada a 37°C por 16 horas. Estas células foram coletadas através de centrifugação (12000 g por 2 minutos a 4°C) e o sobrenadante foi descartado. O sedimento foi ressuspenso em

200 µL de solução I, em seguida foi adicionado 360 µL de solução II. A amostra foi homogeneizada por inversão do tubo e incubada por 5 minutos à temperatura ambiente. Foram adicionados 300 µL de solução III gelada e a amostra foi novamente homogeneizada por inversão do tubo. A mistura foi incubada por 5 minutos no gelo e centrifugada a 12000 g por 5 minutos. O sobrenadante foi descartado e o pellet ressuscitado em 200 µL tampão TE (solução I). Foram adicionados 110 µL de acetato de amônio 7,5 M e após forte agitação foi centrifugado a 12000 g por 10 minutos. O sobrenadante foi recuperado em um tubo limpo, onde foi adicionado 750 µL de etanol 100% gelado. A mistura foi centrifugada a 12000 g por 5 minutos. O precipitado foi lavado (sem ressuspensão) com 1 mL de etanol 70% gelado. Após a secagem por exposição ao ar, o material foi ressuscitado em 50 µL de água MilliQ com RNase A (1 µL de RNase 10mg/mL para 100 µL de água). O DNA plasmidial foi estocado a -20°C até o uso.

2.10 Preparação de DNA plasmidial em pequena escala (miniprep)

O DNA plasmidial foi extraído utilizando o kit Wizard® plus SV Minipreps DNA purification (Promega), segundo as orientações do fabricante.

2.11 Preparação de DNA plasmidial em média escala (midiprep)

O DNA plasmidial foi extraído utilizando o kit Qiagen Plasmid Midi (Qiagen), de acordo com as orientações do fabricante.

2.12 Extração de DNA total (levedura) (adaptado de Burke, Dawson & Stearns, 2000)

Uma colônia de levedura foi inoculada em 3 mL de meio YPD e cultivada a 28°C por cerca de 24 horas. Um por cento desse pré-inóculo foi utilizado para inocular 40 mL de meio YPD em frasco tipo Erlenmeyer de 250 mL e em seguida foi cultivado a 28°C por 24 horas. Depois do crescimento, as células foram coletadas por centrifugação (3.000 g por 5 minutos à temperatura ambiente) e ressuscitadas em 3 mL de tampão SE. Adicionaram-se 100 µL de solução de liticase (20 mg/mL) e a mistura foi incubada a 37°C por 60 minutos. Centrifugou-se a suspensão (3.000 g por 5 minutos à temperatura ambiente) e ressuscitaram-se as células em 3 mL de tampão TE20. Adicionou-se 0,5 mL de SDS 10% e a mistura foi incubada a 65°C

por 30 minutos. Adicionou-se 1,5 mL de acetato de potássio 5 M (pH 8,9) e a mistura foi incubada no gelo por 30 minutos. Centrifugou-se a suspensão (10.000 g por 10 minutos a 4°C) e o sobrenadante foi transferido para um tubo limpo (Falcon 50 mL) e o pellet descartado. Ao sobrenadante foi adicionado um volume de clorofane e, após homogeneização gentil, a mistura foi centrifugada a 3.000 g por 10 minutos. A fase aquosa (menos densa) foi transferida a um novo tubo e adicionaram-se 2,5 volumes de etanol 100%. Após ser incubada a temperatura ambiente por 5 minutos, a mistura foi centrifugada a 10.000 g por 15 minutos a 4°C. Depois da centrifugação, o pellet foi lavado uma vez com etanol 70% e, após ser secado por exposição ao ar, foi ressuscitado em 0,5 mL de tampão TE mais 5 µL de RNase A (10 mg/mL). O DNA total de levedura foi guardado a 4°C para uso ou a -20°C para estoque.

2.13 Digestão de DNA com enzimas de restrição

As digestões foram realizadas de acordo com a instrução dos fabricantes das enzimas. O tempo de incubação, tampões, temperaturas de reação, concentração do DNA e enzima variavam de acordo com o material a ser digerido e a eficácia da enzima.

2.14 Purificação e eluição de fragmentos de DNA

Após as digestões do plasmídeos com enzimas de restrição apropriadas, os fragmentos de DNA contendo a sequência de interesse foram submetidos à eletroforese em gel de agarose. Os fragmentos de interesse foram recortados do gel com lâmina de bisturi de aço carbono estéril (Wiltex plus) e purificados utilizando-se o kit QIAquick Gel Extraction (Qiagen), segundo as recomendações do fabricante.

2.15 Precipitação de DNA

Foram adicionados 0,3 M de acetato de sódio e 2,5 volumes de etanol 100% gelado ao DNA. Após a incubação a -20°C durante a noite, a amostra foi submetida a centrifugação a 12000 g por 30 minutos a 4°C. O sobrenadante foi descartado e foi adicionado 1 mL de etanol 70% e novamente o material foi centrifugado a 12000 g por 15 minutos a 4°C, o sobrenadante foi descartado. O precipitado foi seco, ressuscitado em 10 µL de água Milli-Q e armazenado a -20°C.

2.16 Determinação do número de cópias: Transferência e Southern Blot

As amostras de DNA (10 µg) a serem utilizadas no Southern blot foram digeridas com *Bgl*II e corridas em um gel de agarose 1%. Após a corrida, o gel foi analisado e fotografado em transluminador UV (312 - 365 nm). Em seguida o gel foi: 1) tratado com solução de depurinação por 15 minutos, depois lavado com água; 2) tratado com solução de desnaturação por 15 minutos; 3) tratado com solução de neutralização por 30 minutos. A transferência por capilaridade foi realizada *overnight* montando o sanduíche nesta ordem: placa de vidro, gel, membrana de nylon, 3 folhas de papel filtro, bastante papel toalha, mais uma placa de vidro e um peso de aproximadamente 0,5 kg por cima. Após a transferência, a membrana foi lavada duas vezes com SSC 2X por dois minutos e secada por exposição ao ar por uma hora. Em seguida foi feito o cross-link por exposição à luz UV por 45 segundos duas vezes de cada lado da membrana, com intervalos de 2 minutos entre cada exposição. Seguiu-se a pré-hibridização da membrana (30 minutos a 55°C) e a sua hibridização com a sonda (*overnight* a 60°C), utilizando as especificações do kit AlkPhos Direct™ DNA Labelling (Amershan). A sonda utilizada na etapa de hibridização foi o fragmento de aproximadamente 6,0 kb do vetor pPIC9 (Invitrogen ®) digerido com *Bgl*II e marcado segundo as especificações do kit DIG DNA Labelling and Detection (Roche). A sonda foi previamente fervida por 5 minutos antes da hibridização com a membrana. Depois de realizada esta etapa, foram feitas as seguintes lavagens de estringência na membrana: 1) duas vezes com SSC 1X mais 0,1% de SDS por 10 minutos a 25°C; 2) uma vez com SSC 0,5X mais 0,1% de SDS por 10 minutos a 65°C; 3) uma vez com SSC 0,5X mais 0,1% de SDS por 30 minutos a 65°C. Em sequência foi realizada a reação de revelação dos híbridos segundo as especificações do kit DIG DNA Labelling (Roche).

2.17 Validação do número de cópias: PCR em tempo real (qPCR)

As reações para qPCR, feitas em triplicata, foram constituídas de: 1X Fast SYBR® Green PCR Master Mix (Applied Biosystems), 4 µM de *primer forward* e *reverse* (tabela 6), 30 ng do DNA template (DNA total extraído do clones amilolíticos) e água MilliQ q.s.p. 10 µL. As condições da reação foram: 95°C por 20 s, 40 ciclos de 95°C por 3 min e 60°C por 20 s. Tanto

as reações quanto as condições foram recomendadas pelo fabricante. O equipamento utilizado foi o 7500 Fast Real Time PCR System (Applied Biosystems), disponível em nosso laboratório.

Foi utilizado o método de quantificação relativa ($\Delta\Delta Ct$) (Livak & Schmittgen, 2001) com o gene *GAP* de *P. pastoris* e o gene *amyE* do clone controle positivo como padrões internos para 1 cópia de gene. A especificidade do produto da reação foi verificada por meio da curva de dissociação; segundo o programa: 95°C por 15 segundos, 60°C por 1 min, 95°C por 15 s e 60°C por 15 s; logo após o término da reação.

2.18 Transformação de *Pichia pastoris* GS115 por eletroporação

Uma colônia da levedura *P. pastoris* foi inoculada em 5 mL de meio YPD em *Erlenmeyer* de 125 mL e incubada a 28°C sob agitação (200rpm) por 24 horas. Um mililitro do pré-inóculo foi adicionado a 100 mL de meio YPD. As células cresceram durante 16 horas, até uma OD_{600} entre 1,3 e 1,5. Em seguida as células foram centrifugadas a 1500 g por 5 minutos a 4°C. O sobrenadante foi descartado e as células foram ressuspensas em 100 mL de água estéril e gelada. As células foram centrifugadas nas mesmas condições e novamente o sobrenadante foi descartado e as células foram ressuspensas em 50 mL de água estéril e gelada. Após a repetição desse último procedimento as células foram ressuspensas em 4 mL de sorbitol 1 M gelado. Depois de uma nova centrifugação nas mesmas condições anteriores, as células foram ressuspensas em 200 μ L de sorbitol 1 M gelado e mantidas no gelo até o momento do uso.

Para a eletroporação, 80 μ L das células foram misturadas a 5-10 μ g de DNA (previamente linearizado com *SaI*) e 320 μ L de sorbitol 1 M gelado, e então transferidas para a cubeta de transformação de 0,2 cm gelada. As células foram eletroporadas nas seguintes condições do eletroporador da BioRad: 7,5 KV/cm, 25 μ F e 400 Ω . Imediatamente após o choque, foi adicionado 600 μ L de sorbitol 1M gelado na cubeta. O conteúdo foi transferido para um tubo estéril. As células foram plaqueadas e meio MD-ágar sem histidina para seleção dos transformantes com vetor contendo a marca auxotrófica *HIS4*. As placas foram incubadas a 28°C até o aparecimento de colônias.

2.19 Expressão heteróloga em *P. pastoris*

2.19.1 Screening em placas de petri

Os clones que cresceram em meio MD-ágar sem histidina foram selecionados para esse teste. Os clones foram repicados em placas de MD e MM-ágar amido com rótulo identificador numerado e foram incubadas por 48 horas a 28°C. As placas contendo amido foram utilizadas para análise dos halos enquanto as de MD foram guardadas para manutenção dos clones (placa mestre). Para a observação dos halos as placas foram coradas com vapor de iodo. As regiões que ficam descoradas são os halos formados pela degradação do amido pela α -amilase expressa e secretada.

2.19.2 Análise e identificação dos melhores sinais de secreção: Indução de *P. pastoris* em frascos (teste piloto)

Uma colônia isolada foi inoculada em 25 mL de meio BMGY em Erlenmeyer de 250 mL e incubada a 28°C sob agitação (200 rpm) até atingir uma OD₆₀₀ de 2 a 6 (em torno de 16 a 18 horas), as células foram coletadas por centrifugação a 1500 g por 5 minutos a 4°C, o sobrenadante foi desprezado. As células foram ressuspensas em 50 mL de meio BMMY em Erlenmeyer 500 mL, obtendo uma OD₆₀₀ 0,3 e incubadas novamente sob as mesmas condições. Metanol 100% foi adicionado a cada 24 horas a uma concentração final de 0,5%, para manter a indução dos transformantes. Durante 72 horas de indução foram coletados 4 pontos (0h, 24h, 48h e 72h). Em cada ponto foram coletados 2 mL de cultura em dois Eppendorfs (1,5 mL), sendo 1 mL de cultura para cada Eppendorf. Depois de centrifugado, o sobrenadante foi utilizado para atividade enzimática e perfil protéico em gel SDS-PAGE e o pellet de células foi estourado para determinação de atividade enzimática intracelular.

2.19.3 Análise e identificação dos melhores sinais de secreção: Indução de *P. pastoris* em frascos (teste definitivo)

As induções definitivas foram realizadas da mesma forma que o teste piloto descrito anteriormente. A diferença é que foi coletado apenas um ponto após 48 horas de indução. Foram coletados 2 mL de cultura em dois Eppendorfs (1,5 mL), sendo 1 mL de cultura para cada Eppendorf. Depois de centrifugado, o sobrenadante foi utilizado para atividade enzimática

e perfil protéico em gel SDS-PAGE e o pellet de células foi estourado para determinação de atividade enzimática intracelular.

2.20 Liberação de extrato protéico intracelular

Um mL do sobrenadante de cultura dos clones amilolíticos, após indução até o tempo escolhido (0, 24, 48 e/ou 72 horas), foi centrifugado (13.000 rpm por 2 minutos) para separar o sobrenadante do pellet celular. O sobrenadante foi guardado em outro Eppendorf para ensaio enzimático e o *pellet* foi ressuspensionado em 200 μ L de tampão acetato de sódio 0,5 M pH 6,0. Em seguida, essas células foram novamente sedimentadas por centrifugação a 13.000 rpm por 2 minutos. O tampão foi descartado e as células foram novamente ressuspensionadas em 1 mL do mesmo tampão junto com 1 mM de PMSF. Vinte μ L de esferas de vidro (diâmetro 0,45 X 0,5 mm) foram adicionados à mistura. Para romper as células seguiu-se a agitação do Eppendorf em 4 ciclos de 45 segundos em vortex e 2 minutos no gelo. O material lisado foi sedimentado a 10.000 g por 20 minutos e o sobrenadante foi utilizado para ensaios enzimáticos.

2.21 Quantificação de proteínas totais (método de Bradford)

A quantificação de proteínas totais foi realizada utilizando o método de Bradford descrito no kit Protein Assay (Bio Rad), com o reagente Coomassie® Brilliant Blue G-250.

Para o sobrenadante, as reações foram: 100 μ L de sobrenadante, 700 μ L de água destilada e 200 μ L do reagente de Bradford. Para o extrato intracelular, as reações foram: 200 μ L de extrato, 600 μ L de água destilada e 200 μ L do reagente de Bradford. As reações foram incubadas à temperatura ambiente por 10 minutos e depois 200 μ L da reação foram utilizados para quantificação em espectrofotômetro (595 nm), a partir da curva padrão de soro-albumina bovina (BSA). Esta curva foi preparada com concentrações crescentes de BSA: 0, 2, 4, 6, 8, 10, 12, 14 e 16 μ g/ml.

2.22 Atividade amilolítica quantitativa (método dextrinizante)

A atividade de amilase foi quantificada por ensaio de detecção de amido restante após hidrólise do substrato amido (Merck), conforme descrito por Moraes *et al* (1995). A solução de FUWA (1:1:3 de Iodo1%:KI10%:Água) foi utilizada para detecção do amido restante, pois, por conter iodo (agente intercalante dos anéis de glicose do amido), permite quantificação colorimétrica. Os ensaios foram realizados da seguinte maneira: 60 µL do sobrenadante de cultura ou do extrato intracelular (contendo amilase), 100 µL de solução de amido 0,5%, 40 µL de tampão acetato de sódio 0,5 M pH 6,0. A reação foi mantida a 40°C por 30 minutos e, então, interrompida pela adição de 200 µL de ácido acético 1 M. Depois foram adicionados 200 µL de solução de FUWA para coloração e 4,4 mL de água destilada para completar o volume. Os 5 mL resultantes foram homogeneizados e 200 µL foram utilizados para quantificação de amido restante em espectrofotômetro (660 nm), a partir da curva padrão de amido. Esta curva foi preparada com concentração crescentes de amido: 0,1; 0,2; 0,3; 0,4 e 0,5 mg; nos mesmos padrões do ensaio enzimático, trocando-se os 60 µL de enzima por água destilada. Uma unidade de atividade enzimática foi definida como a quantidade de enzima necessária para hidrolisar 0,1 mg de amido por minuto, sob as condições de ensaio citadas.

$$X = \frac{\text{Absorbância (0,5 mg amido)} - \text{Absorbância (ensaio)}}{\text{Absorbância (0,1 mg amido)}}$$

$$\text{Unidade (U)} = \frac{X}{\text{Tempo de incubação (minutos)}}$$

$$\text{Atividade absoluta} \left(\frac{U}{mL} \right) = \frac{\text{Unidade}}{\text{Volume do sobrenadante (amilase)}}$$

$$\text{Atividade específica} \left(\frac{U}{mg} \right) = \frac{\text{Atividade absoluta}}{\text{Quantidade de proteínas totais} \left(\frac{mg}{mL} \right)}$$

2.23 Precipitação de proteínas com TCA, deoxicolato de sódio e acetona

Durante o processo de fermentação da *P. pastoris* transformante foi retirado em 48 horas 2 mL de cultura. Após a centrifugação e descarte das células, 1 mL de sobrenadante foi precipitado. Para a precipitação foi adicionado 250 µL de TCA (100%) e 8 µL de deoxicolato de sódio (2%) ao sobrenadante e incubado overnight a -20°C. Em seguida foi centrifugado a 8000 g por 30 minutos a 4°C, o sobrenadante foi descartado e o precipitado resultante foi lavado com acetona gelada e novamente centrifugado nas mesmas condições. O precipitado foi seco a temperatura ambiente e ressuspensão em 20 µL de tampão de amostra 2X. Nas amostras com pH ácido foi adicionado 1 µL de tampão acetato de sódio 3 M pH 6. O estoque das amostras foi feito a -20°C até a análise em SDS-PAGE.

2.24 Análise em gel de poliacrilamida desnaturante (SDS-PAGE)

A eletroforese de proteínas foi conduzida em gel de poliacrilamida desnaturante segundo a descrição de Laemmli (1970). Antes da aplicação no gel as amostras foram fervidas por 5 minutos para desnaturação das proteínas. A corrida foi conduzida em tampão de corrida 1X a uma voltagem constante de 100V.

2.25 Coloração com Comassie Blue

As bandas proteicas foram visualizadas após a incubação do gel na solução corante Comassie Blue durante algumas horas sob agitação, seguida da descoloração, após a incubação em solução descorante sob agitação, até o aparecimento das bandas.

2.26 Análise de seqüências de DNA, eletroferogramas e alinhamentos

Após sequenciamento no sequenciador automático disponível no Laboratório de Biologia Molecular (UnB), as seqüências e eletroferogramas obtidos foram analisados utilizando o software Chromas Lite (versão 2.01, Technelysium Pty Ltd.) e com a ferramenta *PHPH* disponível online no site do próprio laboratório (<http://helix.biomol.unb.br/phph/index.html>). Essas ferramentas foram utilizadas para averiguar

a qualidade do sequenciamento e confirmar mutações em regiões de interesse. Os alinhamentos foram realizados utilizando o BLAST (NCBI - <http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Referências Bibliográficas

Almeida JRM, Moraes LMP, Torres FAG, 2005. Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*. *Yeast* 22: 725-737.

Bader O, Krauke Y, Hube B, 2008. Processing of predicted substrates of fungal Kex2 proteinases from *Candida albicans*, *C. glabrata*, *Saccharomyces cerevisiae* and *Pichia pastoris*. *BMC Microbiol* 8: 116-132.

Burke D, Dawson D, Stearns T, 2000. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. Edição ano 2000, Cold Spring Harbor Laboratory Press, Nova Iorque, 205p.

Cereghino JL & Cregg JM, 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24: 45-66.

Chen R, 2011. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol. Adv.* [Epub ahead of print].

Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA, 1991. High level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* 9:455-460.

Couderc R, Baratti J, 1980. Oxidation of methanol by the yeast *Pichia pastoris*: purification and properties of alcohol oxidase. *Agric Biol Chem* 44: 2279-2289.

Cregg JM, 2004. Heterologous proteins expressed in *Pichia pastoris*. [<http://faculty.kgi.edu/cregg>].

Cregg JM, Cereghino L, Shi J, Higgins DR, 2000. Recombinant protein expression in *Pichia pastoris*. *Mol Biotech* 16: 23-52.

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA, 1989. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Mol Cell Biol* 9: 1316–1323

Cregg JM, Vedvick TS, Raschke WC, 1993. Recent advances in the expression of foreign genes in *Pichia pastoris*. *Bio/Technology* 11: 905-910.

Çelik E & Çalik P, 2011. Production of recombinant proteins by yeast cells. *Biotechnol Adv* [Epub ahead of print].

Daily R & Hearn MTW, 2005. Expression of heterologous protein in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* 18: 119–138.

Damasceno LM, Huang CJ, Batt CA, 2012. Protein secretion in *Pichia pastoris* and advances in protein production. *Applied Microbiology and Biotechnology* 93: 31–39

Demain L, Vaishnav P, 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 27: 297-306.

Desai PN, Shrivastava N, Padh H, 2010. Production of heterologous proteins in plants: strategies for optimal expression. *Biotechnol Adv.* (4): 427-35.

- De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N, 2009. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* 24. [Epub ahead of print]
- Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR, 1985. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol Cell Biol* 5: 1111–1121.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H, 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953-971.
- Gellissen G, 2000. Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* 54: 741–750.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG, 1996. Life with 6000 genes. *Science* 272: 546-567.
- Ha SH, Park JJ, Kim JW, Jeong JW, Noh KS, Jeon YJ, Kin HS, Kim HB, 2001. Molecular cloning and high-level expression of G2 protein of *Hantaan* (HTN) virus 76–118 strain in the yeast *Pichia pastoris* KM71. *Virus Genes* 22: 167–173.
- Hinnen A, Buxton F, Chaudhuri B, Heim J, Hottiger T, Meyhack B, Pohlig G, 1994. In: Smith, A. (Ed.), *Gene Expression in Recombinant Microorganisms*. Marcel Dekker, New York, pp. 121–193.
- Hodgson J, 1994. The changing bulk biocatalyst market. *Bio/Technology* 12: 789–90.
- Huang CJ, Damasceno LM, Anderson KA, Zhang S, Old LJ, Batt CA, 2011. A proteomic analysis of the *Pichia pastoris* secretome in methanol-induced cultures. *Appl Microbiol Biotechnol* 90(1): 235-247.
- Invitrogen EUA, manual *Pichia Expression Kit*.
- Inan M, Aryasomayajula D, Sinha J, Meagher MM, 2006. Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnol. Bioeng.* 93:771–778.
- Jung E, Williams KL, 1997. The production of recombinant glycoproteins with special reference to simple eukaryotes including *Dictyostelium discoideum*. *Biotechnol. Appl. Biochem.* 25, 3–8.
- Kottmeier K, Ostermann K, Bley T, Rödel G, 2009. Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in *P. pastoris*. *Appl Microbiol Biotechnol* 91(1): 133-41.
- Kozlov DG & Yagudin TA, 2008. Antibody fragments may be incorrectly processed in the yeast *Pichia pastoris*. *Biotechnol Lett* 30: 1661-1663.
- Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A, Pichler H, 2011. High-quality genome sequence of *Pichia pastoris* CBS7435. *J Biotechnol.* 154(4) :312-20.
- Kurtzman CP, 2009. Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol. Biotechnol.* 36: 1435–1438.
- Liu H, Tan X, Russell KA, Veenhuis M, Cregg JM, 1995. PER3, a gene required for peroxisome biogenesis in *Pichia pastoris*, encodes a peroxisomal membrane protein involved in protein import. *J. Biol. Chem.* 270: 10940-10951.

- Liu Y, Lu F, Chen G, Snyder CL, Sun J, Li Y, Wang J, Xiao J, 2010. High-level expression, purification and characterization of a recombinant medium-temperature alpha-amylase from *Bacillus subtilis*. *Biotechnol Lett.* 32(1): 119-24.
- Livak KJ & Schmittgen TD, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 25(4): 402-8.
- Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM, 2005. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22: 249-270.
- Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B, 2009. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microb Cell Fact* 8: 29.
- Menendez J, Hernandez L, Banguela A, Pais J, 2004. Functional production and secretion of the *Gluconoacetobacter diazotrophicus* fructose-releasing exo-levanase (LsdB) in *Pichia pastoris*. *Enz Microb Technol* 34: 446-452.
- Mochizuki S, Hamato N, Hirose M, Miyano K, Ohtani W, Kameyama A, Kuwae S, Tokuyama T, Ohi H, 2001. Expression and characterization of recombinant human antithrombin III in *Pichia pastoris*. *Protein Expr Purif* 23(1): 55-65.
- Moraes LMP, Astolfi-Filho S, Oliver SO, 1995. Development of yeast strains for the efficient utilization of starch: evaluation of constructs that express α -amylase and glucoamylase separately or a functional fusion protein. *Appl Microbiol Biotechnol* 43: 1067-1076.
- Moraes LMP, 2004. Amilases. In: Said S. & Pietro R.C.L.R. *Enzimas como Agentes Biotecnológicos*. Ribeirão preto, Editora Legis Summa. Capítulo 13, páginas 223 a 242.
- Nelson DL & Cox MM, 2004. *Lehninger Principles of Biochemistry*. 4^a ed. Nova Iorque. Editora W. H. Freeman. Páginas 1068 e 1069.
- Nevalainen KM, Te'o VS, Bergquist PL, 2005. Heterologous protein expression in filamentous fungi. *Trends Biotechnol.* 9: 468-74.
- Paifer E, Margolle E, Cremata J, Montesino R, Herrera L, Delgado JM, 1994. Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences. *Yeast* 10(11):1415-9.
- Prakash O & Jaiswal N, 2010. Alpha-Amylase: an ideal representative of thermostable enzymes. *Appl. Biochem. Biotechnol.* 160(8): 2401-14.
- Raemaekers RJ, de Muro L, Gatehouse JA, Fordham-Skelton AP, 1999. Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*: correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur J Biochem* 265: 394–403.
- Romanos M, Scorer CA, Clare JJ, 1992. Foreign gene expression in yeast: a review. *Yeast* 8: 423-488.
- Romanos M, 1995. Advances in the use of *Pichia pastoris* for high-level gene expression. *Curr Opin Biotechnol* 6: 527-533.
- Rosenfeld SA, Nadeau D, Tirado J, Hollis GF, Knabb RM, Jia S, 1999. Production and purification of recombinant hirudin expressed in the methylotrophic yeast *Pichia pastoris*. *Methods Enzymol* 306: 154-169.

- Sambrook J, Fritsch EF, Maniatis T, 1989. Molecular cloning: a laboratory manual. 2. ed., 3. vol., Cold Spring Harbor Laboratory Press, Nova Iorque, 253p.
- Sears IB, O'Connor J, Rossanese OW, Glick BS, 1998. A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. *Yeast* 14: 783-790.
- Shen S, Sulter G, Jeffries TW, Cregg JM, 1998. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. *Gene* 216: 93-102.
- Stroh WH, 1994. Trends in the use of industrial bioprocessing enzymes for the 21st century. *Gen Eng News* 14(16): 10-2.
- Tate CG & Grishammer R, 1996. Heterologous expression of G protein-coupled receptors. *Trends Biotechnol.* 14, 426-430.
- Trujillo LE, Arrieta JG, Dafnis F, Garcia J, Valdes J, Tambara Y, Perez M, Hernandez L, 2001. Fructo-oligosaccharides production by the *Gluconacetobacter diazotrophicus* levansucrase expressed in the methylotrophic yeast *Pichia pastoris*. *Enzyme Microb Technol* 28: 139-144.
- Valência FF, 1990. Sequenciamento do gene de e-amylase de *Bacillus subtilis*. Dissertação de Mestrado. Universidade de Brasília, Brasil.
- Waterham HR, Digan ME, Koutz PJ, Lair SL, Cregg JM, 1997. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186: 37-44.
- Wegner G, 1990. Emerging applications of the methylotrophic yeasts. *FEMS Microbiol Rev* 7: 279-283.
- Wood KV, 1995. Marker proteins for gene expression. *Current Opinion in Biotechnology* 6: 50-58.
- Wurm FM, 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* (11): 1393-8.
- Yin J, Li G, Ren X, Herrler G, 2007. Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J. Biotechnol.* 127(3): 335-47.
- Zhu T, Guo M, Tang Z, Zhang M, Zhuang Y, Chu J, Zhang S, 2009. Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. *J. Appl. Microbiol.* 107(3):954-63.

Anexos

ANEXO I – Sequenciamento e alinhamento dos peptídeos sinais.

Obs: *query* é a sequência original e *subject* é a sequência clonada e sequenciada

Peptídeo sinal 1 (PS1 - PIPA 00211):

Score = 187 bits (101), Expect = 4e-52
Identities = 101/101 (100%), Gaps = 0/101 (0%)
Strand=Plus/Plus

```
Query 1 CGGATCCCATATGAACTTATTAAGTCTAACCTTACTATGTTCCTACTGAAGCACTCGC 60
|||||
Sbjct 103 CGGATCCCATATGAACTTATTAAGTCTAACCTTACTATGTTCCTACTGAAGCACTCGC 162

Query 61 CCTCACCGGTGAAGCTGAAACTGCAAACAAATCGAATAAGG 101
|||||
Sbjct 163 CCTCACCGGTGAAGCTGAAACTGCAAACAAATCGAATAAGG 203
```

Peptídeo sinal 3 (PS3 - PIPA 00436):

Score = 187 bits (101), Expect = 4e-52
Identities = 101/101 (100%), Gaps = 0/101 (0%)
Strand=Plus/Plus

```
Query 1 CGGATCCAATATGAAATCTCAACTTATCTTTATGGCTCTTGCCTCTCTGGTGGCCTCCGC 60
|||||
Sbjct 83 CGGATCCAATATGAAATCTCAACTTATCTTTATGGCTCTTGCCTCTCTGGTGGCCTCCGC 142

Query 61 TCCTTTAGAACACGCTGAAACTGCAAACAAATCGAATAAGG 101
|||||
Sbjct 143 TCCTTTAGAACACGCTGAAACTGCAAACAAATCGAATAAGG 183
```

Peptídeo sinal 4 (PS4 - PIPA 00545):

Score = 198 bits (107), Expect = 2e-55
Identities = 107/107 (100%), Gaps = 0/107 (0%)
Strand=Plus/Plus

```
Query 1 CGGATCCCAATGAAGATATCCGCTCTTACAGCCTGCGCTGTTACTCTAGCTGGTCTTGC 60
|||||
Sbjct 101 CGGATCCCAATGAAGATATCCGCTCTTACAGCCTGCGCTGTTACTCTAGCTGGTCTTGC 160

Query 61 AATTGCAGCACCAGCTCCAGCTGAAACTGCAAACAAATCGAATAAGG 107
|||||
Sbjct 161 AATTGCAGCACCAGCTCCAGCTGAAACTGCAAACAAATCGAATAAGG 207
```

Peptideo sinal 5 (PS5 - PIPA 00748):

Score = 196 bits (106), Expect = 6e-55
Identities = 106/106 (100%), Gaps = 0/106 (0%)
Strand=Plus/Plus

```

Query 2   GGATCCATTATGAAGCTCGCTGCACTCTCCACTATTGCATTAACATATTTACCCGTTGCC 61
          |||
Sbjct 70   GGATCCATTATGAAGCTCGCTGCACTCTCCACTATTGCATTAACATATTTACCCGTTGCC 129

Query 62   TTGGCTGGCTATGCTCCAGCTGAAACTGCAAACAAATCGAATAAGG 107
          |||
Sbjct 130  TTGGCTGGCTATGCTCCAGCTGAAACTGCAAACAAATCGAATAAGG 175

```

Peptídeo sinal 6 (PS6 - PIPA 00934):

Score = 191 bits (103), Expect = 3e-53
 Identities = 108/110 (98%), Gaps = 1/110 (1%)
 Strand=Plus/Plus

```

Query 1   CGGATCCCCAATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGT 60
          |||
Sbjct 83   CGGATCCCCAATGAAGCTCTCCACCAATTTCATTCTAGCTATTGCAGCAGCTTCCGCCGT 142

Query 61   TGTCTCAGCTGCTCCAGTTGCTGCTGAAACTGCAAACAAATCGAATAAGG 110
          |||
Sbjct 143  TGTCTCAGCTGCTCCAGTTGCTGCTGAAACTGCAAACAAA-CGAATAAGG 191

```

Software Chromas Lite: Na sequência do eletroferograma não aparece a mutação citosina (113), o nucleotídeo é guanina como esperado, além de haver um pico de guanina evidente nessa região. Na sequência do eletroferograma houve a deleção da timina (182). Esse local pertence à região de anelamento no gene da amilase.

Ferramenta *PHPH*: Na sequência do eletroferograma aparece a mutação citosina (113), porém há um pico de guanina mais expressivo nessa região, que demonstrou qualidade de sequenciamento ruim (entre 10 e 20). Na sequência do eletroferograma houve a deleção da timina (182) e a qualidade do sequenciamento na região é ótima (maior que 31). Esse local pertence à região de anelamento no gene da amilase.

Peptídeo sinal 8 (PS8 - PIPA 01008):

Score = 187 bits (101), Expect = 4e-52
 Identities = 103/104 (99%), Gaps = 0/104 (0%)
 Strand=Plus/Plus

```

Query 1   CGGATCCAATATGTTTAAATCTCTGTGCATGTTAATAGGTTCCCTGCCTATTGAGTTCAGT 60
          |||
Sbjct 65   CGGATCCAATATGTTTAAATCTCTGTGCATGTTAATAGGTTCCCTGCCTATTGAGTTCAGT 124

Query 61   CTTGGCAGCAGATTTGCTGAAACTGCAAACAAATCGAATAAGG 104
          |||
Sbjct 125  CTTGGCAGCAGATTTGCTGAAACTGCAAACAAATCGAAAAAGG 168

```

Software Chromas Lite: Realmente existe a mutação adenina (164) na sequência do eletroferograma, mas é possível detectar um pico leve de timina nessa região, que demonstrou qualidade de sequenciamento ruim (eletroferograma desorganizado). Esse local pertence à região de anelamento no gene da amilase.

Ferramenta *PHPH*: Realmente existe a mutação adenina (164) na sequência do eletroferograma, mas é possível detectar um pico leve de timina na região, que demonstrou

qualidade de sequenciamento ruim (entre 10 e 20). Esse local pertence à região de anelamento no gene da amilase.

Peptídeo sinal 9 (PS9 - PIPA 01010):

Score = 193 bits (104), Expect = 8e-54
Identities = 104/104 (100%), Gaps = 0/104 (0%)
Strand=Plus/Plus

```
Query 1   CGGATCCGCAATGTTGTCCATTTTAAGTGCATTAAGTCTGCTGGGCTGTCTTGTGCTTC 60
          |||
Sbjct 51   CGGATCCGCAATGTTGTCCATTTTAAGTGCATTAAGTCTGCTGGGCTGTCTTGTGCTTC 110

Query 61   AGATCTGACCCACCTGCTGAACTGCAAACAAATCGAATAAGG 104
          |||
Sbjct 111  AGATCTGACCCACCTGCTGAACTGCAAACAAATCGAATAAGG 154
```

Peptídeo sinal 10 (PS10 - PIPA 01223):

Score = 204 bits (110), Expect = 4e-57
Identities = 110/110 (100%), Gaps = 0/110 (0%)
Strand=Plus/Plus

```
Query 1   CGGATCCAATATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGC 60
          |||
Sbjct 80   CGGATCCAATATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGC 139

Query 61   ATCCCTACAGGCTCCAATACCTGCTGAACTGCAAACAAATCGAATAAGG 110
          |||
Sbjct 140  ATCCCTACAGGCTCCAATACCTGCTGAACTGCAAACAAATCGAATAAGG 189
```

Peptídeo sinal 11 (PS11 - PIPA 01958):

Score = 215 bits (116), Expect = 2e-60
Identities = 116/116 (100%), Gaps = 0/116 (0%)
Strand=Plus/Plus

```
Query 1   CGGATCCAAGATGATCTTTAATCTTAAACACTGGCTGCGGTTGCAATCTCCATTTTACA 60
          |||
Sbjct 103  CGGATCCAAGATGATCTTTAATCTTAAACACTGGCTGCGGTTGCAATCTCCATTTTACA 162

Query 61   AGTGTCTGCAGTTTCTCTCTGGGTTTTGCTGAACTGCAAACAAATCGAATAAGG 116
          |||
Sbjct 163  AGTGTCTGCAGTTTCTCTCTGGGTTTTGCTGAACTGCAAACAAATCGAATAAGG 218
```

Peptídeo sinal 14 (PS14 - PIPA 02524):

Score = 191 bits (103), Expect = 3e-53
Identities = 106/107 (99%), Gaps = 1/107 (1%)
Strand=Plus/Plus

```
Query 1   CGGATCCGGGATGAACTTGTACCTAATTACATTACTATTCGCCAGTCTATGCAGCGCAAT 60
          |||
Sbjct 82   CGGATCCGGGATGAACTTGTACCTAATTACATTACTATTCGCCAGTCTATGCAGCGCAAT 141

Query 61   TACTCTCCAAAGAGAGATGCTGAACTGCAAACAAATCGAATAAGG 107
          ||
Sbjct 142  TA-TCTCCAAAGAGAGATGCTGAACTGCAAACAAATCGAATAAGG 187
```

Software Chromas Lite: Realmente existe a deleção citosina (144) na sequência do eletroferograma, mas é possível detectar um pico leve de citosina nessa região. Esse local pertence à sequência do peptídeo sinal 14.

Ferramenta *PHPH*: Realmente existe a deleção citosina (144) na sequência do eletroferograma, mas é possível detectar um pico leve de citosina nessa região, que demonstrou qualidade de sequenciamento ótima (acima de 31). Esse local pertence à sequência do peptídeo sinal 14.

Peptídeo sinal 16 (PS16 - PIPA 03955):

```
Score = 237 bits (128), Expect = 5e-67
Identities = 128/128 (100%), Gaps = 0/128 (0%)
Strand=Plus/Plus

Query 1   CGGATCCCAAATGTCATTCTCTTCCAACGTGCCACAACCTTTTCTGTTGTTGGTTCTGTT 60
          |||
Sbjct 83   CGGATCCCAAATGTCATTCTCTTCCAACGTGCCACAACCTTTTCTGTTGTTGGTTCTGTT 142

Query 61   GACCAATATAGTCAGTGGAGCAGTCATTTTCAGTATGGTCCGCTGAAACTGCAAACAAATC 120
          |||
Sbjct 143   GACCAATATAGTCAGTGGAGCAGTCATTTTCAGTATGGTCCGCTGAAACTGCAAACAAATC 202

Query 121  GAATAAGG 128
          |||
Sbjct 203  GAATAAGG 210
```

Peptídeo sinal 17 (PS17 - PIPA 04722):

```
Score = 198 bits (107), Expect = 2e-55
Identities = 107/107 (100%), Gaps = 0/107 (0%)
Strand=Plus/Plus

Query 1   CGGATCCACCATGCAAGTTAAATCTATCGTTAACCTACTGTTGGCATGTTTCGTTGGCCGT 60
          |||
Sbjct 60   CGGATCCACCATGCAAGTTAAATCTATCGTTAACCTACTGTTGGCATGTTTCGTTGGCCGT 119

Query 61   GGCCAGACCTTTGGAGCACGCTGAAACTGCAAACAAATCGAATAAGG 107
          |||
Sbjct 120  GGCCAGACCTTTGGAGCACGCTGAAACTGCAAACAAATCGAATAAGG 166
```

Peptídeo sinal 20 (PS20 - PIPA 05771):

Aguardando sequenciamento...

Região pré-pró do fator α de *P. pastoris* (PS α):

```
Score = 675 bits (365), Expect = 0.0
Identities = 368/369 (99%), Gaps = 1/369 (0%)
Strand=Plus/Plus

Query 1   ATGAAATCACTCATTTTGAACATCATTTTCAGTAACTTTAGCTATCACATCAACTGCGGCC 60
          |||
Sbjct 61   ATGAAATCACTCATTTTGAACATCATTTTCAGTAACTTTAGCTATCACATCAACTGCGGCC 120

Query 61   AGTGCGCCAGTGGAAAGCATTTTGGCTAACCAACCTGATTCATCACTCACTGATACTAAT 120
          |||
Sbjct 121  AGTGCGCCAGTGGAAAGCATTTTGGCTAACCAACCTGATTCATCACTCACTGATACTAAT 180
```

```

Query 121  GATGGTGTCTGGCGTTGGCATGTCTACAATCAAAGAAGAAGATTTGGCAAACATTTTGT 180
          |||
Sbjct 181  GATGGTGTCTGGCGTTGGCATGTCTACAATCAAAGAAGAAGATTTGGCAAACATTTTGT 240

Query 181  GAAAACCAAATTCCTTGATGAGGCCGTAATCATGTCATTGAAGTTAAGAAAGGGAGTAAAC 240
          |||
Sbjct 241  GAAAACCAAATTCCTTGATGAGGCCGTAATCATGTCATTGAAGTTAAGAAAGGGAGTAAAC 300

Query 241  TTGTTTTTTCTAGATGACATCGGATTAGCTACCGAGCTTATAGGTAACAAGATAGCACAG 300
          |||
Sbjct 301  TTGTTTTTTCTAGATGACATCGGATTAGCTACCGAGCTTATAGGTAACAAGATAGCACAG 360

Query 301  ATTGAGGCTATTGATTTGTCAGAAAGACTGGCACAAAGTTGGACAAACATCAGGAAGAAC 360
          |||
Sbjct 361  ATTGAGGCTATTGATTTGTCAGAAAGACTGGCACAAAGTTGGACAAACATCAGGAAGAAC 420

Query 361  CGCCTATTT 369
          |||
Sbjct 421  CGCC-ATTT 428

```

Software Chromas Lite: Realmente existe a deleção da timina (426) na sequência do eletroferograma, mas é possível detectar um pico de timina de mesma intensidade do de adenina (427) nessa região, que apresentou qualidade no sequenciamento ruim (eletroferograma desorganizado).

Ferramenta *PHPH*: não foi possível fazer análise deste sequenciamento com essa ferramenta, pois não foi feito o *upload* do material para o servidor corretamente.