

EFFECT OF SIGNAL SEQUENCES AND PROMOTERS IN
RECOMBINANT EXTRACELLULAR PROTEIN PRODUCTION BY
PICHA PASTORIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ASLAN MASSAHI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

FEBRUARY 2017

Approval of the thesis

**EFFECT OF SIGNAL SEQUENCES AND PROMOTERS IN
RECOMBINANT EXTRACELLULAR PROTEIN PRODUCTION BY
*Pichia Pastoris***

Submitted by **ASLAN MASSAHI** in partial fulfilment of the requirements
for the degree of **Doctor of Philosophy in Biotechnology Department,
Middle East Technical University** by,

Prof. Dr. Gülbin D. Ünver

Dean, Graduate School of **Natural and Applied Sciences**

Assoc. Prof. Dr. Çağdaş D. Son

Head of Department, **Biotechnology**

Prof. Dr. Pınar Çalık

Supervisor, **Chemical Engineering Dept., METU**

Assoc. Prof. Dr. Çağdaş D. Son

Co-supervisor, **Biology Dept., METU**

Examining Committee Members:

Prof. Dr. Tunçer H. Özdamar

Chemical Engineering Dept., Ankara University

Prof. Dr. Pınar Çalık

Chemical Engineering Dept., METU

Prof. Dr. Haluk Hamamcı

Food Engineering Dept., METU

Prof. Dr. Mehmet İnan

Food Engineering Dept., Akdeniz University

Assist. Prof. Dr. Eda Çelik Akdur

Chemical Engineering Dept., Hacettepe University

Date:

2.02.2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ASLAN MASSAHI
Signature :

ABSTRACT

EFFECT OF SIGNAL SEQUENCES AND PROMOTERS IN RECOMBINANT EXTRACELLULAR PROTEIN PRODUCTION BY *Pichia pastoris*

Massahi, Aslan

PhD., Department of Biotechnology

Supervisor: Prof. Dr. Pınar Çalık

Co-Supervisor: Assoc. Prof. Dr. Çağdaş D. Son

February 2017, 576 pages

The objective of the current research was to develop a recombinant *Pichia pastoris* (r-*P. pastoris*) strain having high-performance extracellular recombinant protein production potential. In this context, human growth hormone (rhGH) was considered as model protein, and in the first research programme endogenous secretion signal peptides (SP) for the secretion of rhGH were examined; thereafter, strains having promoters that work under oxygen limitation conditions were developed. At first step, the base plasmid, pGAPZ α A::*hGH*, was developed by two parent plasmids, pGAPZ α A and pPICZ α A::*hGH*, in which the rhGH gene was fused to a fusion partner including 6His-tag and factor Xa protease cleavage site at N-terminus. In base plasmid, the expression of rhGH was driven by glyceraldehyde-3-phosphate dehydrogenase gene promoter (P_{GAP}) and its subsequent secretion

to the extracellular medium was achieved by *Saccharomyces cerevisiae* α -mating factor prepro leader sequence (α -MF).

In the first research programme the endogenous SPs were selected in order to replace the original α -MF in base plasmid. For this purpose, at first step, in-silico analyses were conducted by recruitment of softwares including: SignalP, Phobius, WolfPsort, and ProP by using a secretome dataset of *P. pastoris* as inputs of the analyses in order to predict potential efficient SPs. The SPs with higher D-score, the score quantifying the signal peptide-ness of a given sequence, which is the output of the SignalP were selected in comparison with widely-used α -MF. Based on the obtained results the selected endogenous SPs were SP23 (MKILSALLLLFTLAFA), SP24 (MKVSTTKFLAVFLLVRLVCA), SP26 (MWSLFISGLLIFYPLVLG), and SP34 (MRPVLSLLLLLASSVLA); in addition, SP13 (MLSTILNIFILLLLFIQASLQ) was included in the candidates in order to compare the secretion results with previous studies.

In the second research programme, the potentially-efficient oxygen limitation-induced promoters were selected by analysing a proteome/transcriptome dataset obtained in different oxygenation levels; subsequent comparison of the fold of increase in protein level in oxygen-limitation condition with GAP enzyme lead to the selection of the three promoters including: pyruvate kinase gene promoter (P_{PYRK}), pyruvate decarboxylase gene promoter (P_{PDC}), and pyrimidine precursor biosynthesis enzyme thi3 gene promoter (P_{THI3}). These promoters were intended to replace the original P_{GAP} in base plasmid.

In the subsequent step replacement of the α -MF with five selected endogenous SPs and replacement of the P_{GAP} with selected three oxygen limitation-induced promoters was conducted. The prepared eight new plasmids along with base plasmid were cloned in *E. coli* and were, subsequently, used for transfection of *P. pastoris*. The confirmed transfectants were passed through copy number determination experiments

and single-copy *r-P. pastoris* strains were selected using real-time quantitative polymerase chain reaction (qPCR).

Shake-bioreactor experiments were performed using five *r-P. pastoris* strains of SPs along with *r-P. pastoris* strain of base plasmid to assess the efficiency of SPs in secretion of rhGH compared to α -MF. The results revealed that the highest secretion efficiency of the endogenous SPs belonged to SP23 (D-score=0.883) which was comparable to α -MF (D-score=0.885). For further confirmation, laboratory-scale bioreactor experiments were conducted with two strains of SP23 and α -MF. The cell concentration reached to the maximum amount of 82 g/L and 75 g/L for α -MF and SP23 strains, respectively, at t=15 h. In addition, the secreted rhGH reached to maximum concentration of 70 mg/L and 56 mg/L for α -MF and SP23 strains, respectively, at t=12 h. Intracellular rhGH concentration also approved the better secretion of rhGH under leadership of α -MF. Overall, the efficiency of SP23 was obtained to be about 70-80% of the α -MF. Although the privilege of α -MF could be ascribed partially to major physicochemical properties like isoelectric point (pI), hydrophobicity, and aliphatic index, no clear relationship between the efficiency of the endogenous SPs and their corresponding physicochemical properties was found.

Second-group shake-bioreactor experiments conducted with three *r-P. pastoris* strains of selected promoters along with *r-P. pastoris* strain of base plasmid revealed inability of the P_{THI3} in expression of rhGH; thus, it was discarded. Then, laboratory-scale bioreactor experiments were conducted with the remaining three *r-P. pastoris* strains to investigate the efficiency of selected oxygen limitation-induced promoters in expression of rhGH compared to P_{GAP} . The maximum cell concentration was obtained at t=15 h as 80 g/L, 82 g/L, and 90 g/L for recombinant strains related to P_{PYRK} , P_{PDC} , and P_{GAP} , respectively. In addition, the maximum secreted rhGH was obtained as 101 mg/L, 122 mg/L, and 58 mg/L for P_{PYRK} , P_{PDC} , and P_{GAP} , respectively. Metabolic flux analysis performed by using a mass-balance based stoichiometric model with 102 metabolites and 146 reactions. Using

the elaborate fermentation data the intracellular *P. pastoris* fluxes were calculated which disclosed the relationship between the oxygen limitation and the fluxes through the branched pathways from pyruvate node, the energy content of the cell in the form of ATP, together with the cell synthesis flux. The excreted ethanol during oxygen limitation and the excreted organic acids like pyruvic acid, acetic acid, and lactic acid, all confirmed the fluxes through the branched pathways and revealed the shift towards the fermentative metabolism under oxygen limitation. Moreover, the mRNA level of the rhGH transcribed under three different promoters was in agreement with the extracellular rhGH level until t=9 h. Overall, based on the conducted experiments and measured final rhGH titer, P_{PDC} was selected as the most efficient oxygen limitation-induced promoter regarding the adopted strategy and conditions.

Keywords: *Pichia pastoris*; endogenous signal peptide; secretion; GAP promoter; pyruvate kinase promoter ; pyruvate decarboxylase promoter; recombinant human growth hormone; oxygen limitation.

ÖZ

SİNYAL DİZİNLERİN VE PROMOTORLARIN *Pichia pastoris* İLE HÜCRE-DIŞI REKOMBİNANT PROTEİN ÜRETİMİNE ETKİSİ

Massahi, Aslan

Doktora, Biyoteknoloji Bölümü

Tez Yöneticisi: Prof. Dr. Pınar Çalık

Ortak Tez Yöneticisi: Doç. Dr. Çağdaş D. Son

Şubat 2017, 576 sayfa

Bu doktora tezinin amacı, hücre-dışı rekombinant protein üretim potansiyeli yüksek rekombinant *Pichia pastoris* (r-*P. pastoris*) suşu geliştirmektir. Bu kapsamda, rekombinant insan büyüme hormonu (rhGH) model protein olarak seçilmiştir. Birinci araştırma programında endojen sinyal peptidler (SP) kullanarak rhGH'nin hücre-dışına aktarımı incelenmiş, ikinci araştırma programında oksijen-kısıtlı malzeme koşullarda aktif promotorları içeren suşlar geliştirilmiştir. Bu çerçevede önce, pGAPZαA ve pPICZαA::*hGH* kullanılarak baz plazmid pGAPZαA::*hGH* geliştirilmiştir. rhGH'nin N-terminal ucuna 6 histidine ve faktör Xa proteaz tanıma bölgesi entegre edilmiştir. *P. pastoris* ile gliseraldehit-3-fosfat dehidrojenaz promotoru (P_{GAP}) altında üretilen rhGH'nin hücre-dışına aktarımı *Saccharomyces*

cerevisiae α -çiftleşme faktörü prepro öncü sekansı (α -MF) ile gerçekleştirilmiştir.

İlk araştırma programında, endojen sinyal peptid kodonları baz plazmiddeki α -MF kodonları yerine klonlanmıştır. Potansiyel yüksek SP'lerin belirlenmesi için, *P. pastoris* sekretom verileri ve in-silico analiz SignalP, Phobius, WolfPsort, ve ProP programları kullanılmıştır. SignalP sonuçları temelinde, α -MF'den yüksek D-puanına, sekansın sinyal peptidi olabilme özelliğini gösteren puana sahip SP'ler belirlenmiştir. Elde edilen sonuçlara göre seçilen endojen SP'ler:

SP23 (MKILSALLLLFTLAFA), SP24 (MKVSTTKFLAVLLVRLVCA), SP26 (MWSLFISGLLIFYPLVLG), ve SP34 (MRPVLSLLLLLASSVLA); ek olarak SP13 (MLSTILNIFILLFIQASLQ) de, sonuçları literatürdeki çalışmalarla kıyaslamak için araştırılmıştır.

İkinci araştırma programında, farklı oksijen aktarım koşullarında elde edilen proteom/transkriptom verileri incelenerek oksijen-kısıtlamalı koşullarında indüklenen promotorlar seçilmiştir. Seçilen üç promotor i) pirüvat kinaz geni promotoru (P_{PYRK}), ii) pirüvat dekarboksilaz geni promotoru (P_{PDC}), ve iii) primidin öncüsü biyosentez enzim thi3 geni promotoru (P_{THI3}), rhGH üretimi için baz plazmidde bulunan P_{GAP} 'in yerine klonlanması ve oksijen-kısıtlamalı koşulda rhGH üretimdeki artışın P_{GAP} altında üretilen rhGH ile kıyaslanması planlanmıştır.

α -MF'ün yerine, bulunan beş adet endojen SP, P_{GAP} 'in yerine de seçilen üç adet oksijen-kısıtlamalı koşulda indüklenen promotor ayrı-ayrı klonlanmıştır. Hazırlanan sekiz yeni plazmid, baz plazmid pGAPZ α A::hGH ile birlikte *E. coli*'ye klonlanmış ve sonrasında *P. pastoris* transfekte edilmiştir. Tek-kopya gen içeren r-*P. pastoris* suşları gerçek-zamanlı kantitatif polimeraz zincir reaksiyonu (qPCR) ile seçilmişlerdir. Çalkalamalı-biyoreaktör deneyleri ile beş farklı SP içeren r-*P. pastoris* suşları ve baz plazmidi içeren r-*P. pastoris* suşu ile rhGH üretim deneyleri yapılmıştır. SP'lerin etkinliği, hücre-dışı r-protein üretiminde yaygın kullanılan α -MF (D-puanı=0.885) ile kıyaslanmış ve SP23 (D-puanı=0.883) en yüksek değere

sahip olduğu için laboratuvar-ölçekli biyoreaktör deneyleri SP23 ve α -MF suşları ile gerçekleştirilmiştir. Maksimum hücre derişimi $t = 15$ st'ta α -MF ve SP23 suşları için, sırasıyla, 82 g/L ve 75 g/L olarak bulunmuştur. Ayrıca, en yüksek hücre-dışı rhGH derişimi α -MF ve SP23 suşlarıyla, sırasıyla, 70 mg/L ve 56 mg/L olarak $t = 12$ st'ta ulaşılmıştır. Hücreiçi rhGH derişimleri de α -MF öncü-sekansı ile rhGH'nin daha iyi salgılandığını doğrulamıştır. Sonuçta, SP23'ün etkinliğinin α -MF'nin %70-80'i kadar olduğu bulunmuştur. Her ne kadar α -MF'nin ayrıcalıklı durumu izoelektrik nokta (pI), hidrofobiklik ve alifatik indeks gibi temel fizikokimyasal özelliklere kısmen bağlı olsa da, endojen SP'lerin etkinliği ve fizikokimyasal özellikleri arasında net bir ilişki bulunamamıştır.

Üç promotor geni içeren *r-P. pastoris* suşları ile birlikte baz plazmidi içeren *r-P. pastoris* suşu ile gerçekleştirilen çalkalamalı-biyoreaktör üretim deneylerinde P_{THI3} 'ün rhGH ekspresyonunu gerçekleştiremediği görülmüş ve suş elenmiştir. Laboratuvar-ölçekli biyoreaktör deneyleri, üç oksijen-kısıtlamalı koşullarda indüklenen promotorlar ile gerçekleştirilmiş ve rhGH ekspresyon seviyeleri P_{GAP} ile kıyaslanmıştır. En yüksek hücre derişimi $t = 15$ st'ta 80 g/L, 82 g/L ve 90 g/L olarak, sırasıyla, P_{PYRK} , P_{PDC} , ve P_{GAP} içeren suşlar ile elde edilmiştir. Bunun yanında, en yüksek hücre-dışı rhGH derişimi 101 mg/L, 122 mg/L ve 58 mg/L olarak, sırasıyla, P_{PYRK} , P_{PDC} , ve P_{GAP} ile elde edilmiştir. Metabolik akı analizleri 102 metabolit ve 146 tepkimeden oluşan stokiyometrik kütle-korunum temelli *P. pastoris* matematik modeli ile yapılmıştır. Hesaplanan tepkime akıları, oksijen-kısıtlamalı koşulda üretilen etanol ve, pirüvik, asetik ve laktik asit gibi, organik asitlerin yolizi akılarını doğrulamış; prosesin oksijen-kısıtlamalı fermentatif metabolizmaya kaydığını göstermiştir. Üç promotor altında mRNA düzeyleri, $t = 9$ st'ta kadar hücre-dışı rhGH üretimi ile paralel sonuç vermiştir. Sonuç olarak, deney sonuçlarına ve rhGH derişimleri temelinde, uygulanan strateji ve koşullar altında P_{PDC} oksijen-kısıtlamalı koşullarda en etkili indüklenen promotor olarak seçilmiştir.

Anahtar kelimeler: *Pichia pastoris*; endojen sinyal peptidi; sekresyon; GAP promotoru; pirüvat kinaz promotoru; pirüvat dekarboksilaz promotoru; rekombinant insan büyüme hormonu; oksijen limitasyonu.

To
My Family
and
My Beloved Wife, SAHAR

In memory of my passed away grandmother, uncle, and cousin

ACKNOWLEDGEMENTS

It is with immense gratitude that I acknowledge the support, motivation, and trouble-shooting of my supervisor Prof. Dr. Pınar Çalık during conducting current research. I must also express my sincere gratitude to my co-supervisor, Assoc. Prof. Dr. Çağdaş D. Son , because of his helpful comments and advices during thesis accomplishment.

I would like to thank Prof. Dr. Tunçer H. Özdamar and Prof. Dr. Mehmet İnan for their great contribution in conducting research. This thesis would not have been possible without their pertinent recommendations.

This thesis would not have been possible unless in a nice place full of kind friends. I cannot find words to express my gratitude to my labmates in industrial biotechnology laboratory. I should appreciate Özge Ata Akyol, Erdem Boy, Burcu Gündüz, Duygu Yalçınkaya, Damla Hücçetoğulları, Sibel Öztürk, Abdullah Keskin, Yiğit Akgun, Hande Güneş, Bebeta Hoxha, Onur Ersoy, Özge Kalender, Begüm Akcan for their supports and helps during hard times of research and sharing their valuable experiences with me.

I am indebted to my friends Reza Moonesi Rad, Yashar Tavakkoli Oskouei, Kaveh Dehghanian, Negar Taghavi Pourianazar, Maryam Parsian, Kaveh Gharabaghi, Hamed Asadi, Mohammad Maadi, Salar Koushan, and Ali Farshkaran because of their supports gave me during thesis accomplishment.

I am also grateful to all academic, administrative, and technical staffs of chemical engineering and biology department as well as all the security staffs of Middle East Technical University for providing secure and convenient atmosphere for conducting this research.

Financial support of Scientific and Technological Research Council of Turkey (TÜBİTAK) via project number of 114R091 and Middle East Technical University fund should also be gratefully acknowledged.

My gratitude is extended to Naseem Yekita for revising article published during thesis work.

I must express my endless and sincere gratitude to Nader Massahi, Mina Siminghalam, Asal Massahi, and Ashkan Massahi as my family members for their ever-present love, support, and motivation during whole time of my studies.

Heartfelt thanks should be given to my wife, Sahar Saeb, for her love, patience, help, and kindness during conducting my PhD thesis.

Finally, I owe my deepest gratitude to whole Turkey for hosting me during past six years.

TABLE OF CONTENTS

ABSTRACT.....	v
ÖZ.....	ix
ACKNOWLEDGEMENTS.....	xiv
TABLE OF CONTENTS.....	xvi
LIST OF TABLES.....	xxvii
LIST OF FIGURES.....	xxxv
NOMENCLATURE.....	xliv

CHAPTERS

1. INTRODUCTION.....	1
2. LITERATURE SURVEY.....	19
2.1. Expressing hosts.....	19
2.1.1. Prokaryotes.....	22
2.1.2. Eukaryotes.....	24
2.1.2.1. Non-yeast expression platforms.....	25
2.1.2.1.1. Mammalian cell lines.....	25
2.1.2.1.2. Transgenic organisms.....	27
2.1.2.1.2.1. Transgenic animals.....	27
2.1.2.1.2.2. Transgenic plants.....	28
2.1.2.1.2.3. Insect cell lines.....	29
2.1.2.1.3. Filamentous fungi (molds).....	31
2.1.2.2. Yeast expression platforms.....	31
2.1.2.2.1. <i>Saccharomyces cerevisiae</i>	35
2.1.2.2.2. <i>Schizosaccharomyces pombe</i>	37

2.2.1.1.2.4. 3-phosphoglycerate kinase gene promoter.....	72
2.2.1.1.3. Novel promoters.....	73
2.2.1.2. Secretion signal peptides.....	78
2.2.1.2.1. Importance of secretion.....	78
2.2.1.2.2. Signal peptides.....	79
2.2.1.2.3. Features of the secretion signal peptides.....	81
2.2.1.2.4. Mechanisms of secretion in eukaryotes.....	82
2.2.1.2.5. Secretion signal peptides in <i>Pichia</i> -based systems.....	87
2.2.1.3. Codon optimization.....	90
2.2.1.4. Gene copy number (Gene dosage).....	92
2.2.2. Process parameters.....	94
2.2.2.1. Medium composition (Nutrient consideration).....	95
2.2.2.2. pH.....	98
2.2.2.3. Temperature.....	99
2.2.2.4. Dissolved oxygen.....	100
2.2.2.4.1. Importance of oxygen provision.....	100
2.2.2.4.2. Mathematical modeling of oxygen transfer.....	102
2.2.2.4.3. Oxygen uptake rate and relation with OTR.....	111
2.2.2.4.4. Factors affecting oxygen transfer rate.....	116
2.2.2.4.4.1. Operational conditions.....	116
2.2.2.4.4.2. Presence of the cells and solutes in the medium....	117
2.2.2.4.4.3. Surfactants.....	118
2.2.2.4.4.4. Antifoaming substances.....	119
2.2.2.4.4.5. Oxygen uptake rate.....	119
2.2.2.4.5. Oxygen-limited situations in bioprocesses.....	120
2.3. Further considerations.....	121
2.3.1. Glycosylation.....	122
2.3.2. Endoplasmic reticulum under stress.....	126
2.3.3. Proteolysis (Product stability).....	133
2.4. Bioprocess kinetics in fed-batch operation.....	139
2.4.1. Cell growth correlations.....	140

2.4.2. Correlations for substrate utilization.....	143
2.4.3. Correlations for product formation.....	147
2.4.4. Yield coefficients.....	148
2.5. Mathematical modelling of the metabolic network.....	150
2.5.1. Kinetic (dynamic) modelling.....	151
2.5.2. Stoichiometric modelling and metabolic flux analysis.....	153
2.6. Recombinant hGH production.....	160
3. MATERIALS AND METHODS.....	171
3.1. Substances and Stuffs.....	171
3.1.1. Enzymes, kits, and markers.....	171
3.1.2. Chemicals.....	171
3.1.3. Antibiotics.....	171
3.1.4. Immunoassay compounds.....	172
3.1.5. Buffers and stock solutions.....	172
3.2. Strains and plasmids.....	172
3.3. Media.....	174
3.3.1. Growth media.....	174
3.3.2. rhGH production media.....	174
3.3.2.1. Pre-cultivation medium.....	175
3.3.2.2. Production medium.....	175
3.3.2.2.1. Shake flask air-filtered bioreactor.....	176
3.3.2.2.2. Laboratory-scale bioreactor.....	178
3.4. Finding the nucleotide sequences of the desired biological elements.....	179
3.4.1. Secretion signal peptides.....	179
3.4.2. Promoters.....	180
3.5. In-silico analyses.....	182
3.6. Genetic engineering techniques.....	182
3.6.1. Synthesizing of the determined nucleotide sequences.....	182
3.6.2. <i>E. coli</i> transformation.....	183

3.6.3. Plasmid DNA isolation from transformed <i>E. coli</i> cells.....	185
3.6.4. Confirmation of the DNA fragments by Agarose gel electrophoresis.....	187
3.6.5. Amplification of the target DNA.....	188
3.6.6. Purification of the PCR products.....	190
3.6.7. Digestion of the plasmid DNA.....	191
3.6.8. Extraction of the DNA fragments from the agarose gel.....	194
3.6.9. Ligation of the DNA fragments.....	194
3.6.10. Sequencing of the isolated plasmids from <i>E. coli</i> cells.....	195
3.6.11. Transfection of the <i>Pichia pastoris</i> X-33 cells.....	197
3.6.12. Genomic DNA isolation from <i>Pichia pastoris</i> cells.....	199
3.6.13. Verification of the genomic integration in isolated genomes.	200
3.6.14. Determination of the hGH copy number in <i>Pichia</i> transfectants.....	201
3.6.14.1. Pre-screening of the <i>Pichia</i> transfectants.....	201
3.6.14.2. qPCR with genomes of the pre-screened <i>Pichia</i> transfectants.....	202
3.6.14.2.1. Provision of the standard curve.....	204
3.6.14.2.1.1. Preparation of the standard DNA sample for <i>ARG4</i> gene.....	204
3.6.14.2.1.2. Preparation of the standard DNA sample for <i>hGH</i> gene.....	205
3.6.14.2.2. Inner primers for qPCR.....	205
3.6.14.2.2.1. Preliminary PCR with inner primers designed for <i>hGH</i>	206
3.6.14.2.2.2. Preliminary PCR with inner primers designed for <i>ARG4</i>	206
3.6.14.2.3. Rendering qPCR experiments.....	207
3.7. rhGH production with selected r- <i>P.pastoris</i> strains.....	209
3.7.1. Shake flask air-filtered bioreactor experiments.....	209
3.7.2. Laboratory-scale bioreactor experiments.....	211

3.7.2.1. Bioreactor experiments with r- <i>P. pastoris</i> strains of SPs...	212
3.7.2.2. Bioreactor experiments with r- <i>P. pastoris</i> strains of promoters.....	213
3.7.2.2.1. Cell cycle synchronization.....	214
3.7.2.2.2. Total RNA isolation.....	215
3.7.2.2.3. Complementary DNA synthesis.....	217
3.8. Analysis.....	219
3.8.1. Measurement of the cell concentration.....	219
3.8.2. SDS-PAGE.....	219
3.8.3. Dot-blot.....	220
3.8.4. Total protein measurement.....	222
3.8.5. Measurement of the rhGH concentration.....	223
3.8.5.1. Measurement of the intracellular rhGH.....	223
3.8.5.2. Measurement of the extracellular rhGH.....	224
3.8.6. Western blotting.....	224
3.8.7. Measurement of the residual glucose.....	226
3.8.8. Measurement of the produced ethanol.....	227
3.8.9. Measurement of the organic acids concentrations.....	229
3.8.10. Measurement of the protease activity.....	231
3.8.11. Metabolic flux analysis.....	232
4. RESULTS AND DISCUSSION.....	233
4.1. Development of the base plasmid.....	233
4.1.1. Isolation and validation of the parent plasmids.....	234
4.1.2. Double-digestion of the parent plasmids with <i>EcoRI</i> and <i>XbaI</i>	236
4.1.3. Ligation of the double-digested fragments.....	237
4.1.4. Transformation of the <i>E. coli</i> DH5 α cells with base plasmid...	238
4.2. In-silico determination of promising endogenous SPs of <i>P. pastoris</i>	242
4.3. Candidate promoters.....	253

4.3.1. Deciding on the promising oxygen limitation-induced promoters.....	253
4.3.2. Finding the nucleotide sequence of the selected promoters....	257
4.3.2.1. Pyruvate kinase promoter (P_{PYRK}).....	257
4.3.2.2. Pyruvate decarboxylase promoter (P_{PDC}).....	259
4.3.2.3. Pyrimidine precursor biosynthesis enzyme thi3 promoter (P_{THI3}).....	261
4.3.3. In-silico appraisal of the selected putative promoter regions	265
4.4. Provision of the restriction sites for endogenous promoters and SPs.....	266
4.4.1. Secretion signal peptides constructs and related primers.....	266
4.4.2. Promoters constructs and related primers.....	269
4.5. Generation of r- <i>E. coli</i> strains with developed plasmid constructs	270
4.5.1. Synthesizing and preparation of the designed constructs.....	270
4.5.1.1. Desired promoter constructs.....	272
4.5.1.2. Desired signal peptide constructs.....	273
4.5.2. Development of the new plasmids with selected promoters....	278
4.5.2.1. Preparation of the desired promoter-free backbone from BP.....	278
4.5.2.2. Preparation of the selected promoters.....	280
4.5.2.3. Ligation of the backbone and selected promoters constructs.....	280
4.5.2.4. Transformation of the <i>E. coli</i> cells with developed plasmids.....	280
4.5.3. Development of the new plasmids with selected endogenous SPs.....	289
4.5.3.1. Preparation of the desired SP-less backbone from BP.....	289
4.5.3.2. Preparation of the selected endogenous signal peptides....	289
4.5.3.3. Ligation of the SP-less backbone and selected SPs.....	289
4.5.3.4. Transformation of the <i>E. coli</i> cells with developed plasmids.....	291

4.6. Developing recombinant <i>P. pastoris</i> cells.....	297
4.6.1. Single-digestion of the developed plasmids.....	298
4.6.2. Transfection of <i>P. pastoris</i> X-33 cells with linearized new plasmids.....	299
4.7. Pre-screening of the selected <i>P. pastoris</i> transfectants.....	304
4.8. <i>hGH</i> copy number determination in selected verified strains.....	313
4.8.1. Standard DNA sample preparation for <i>ARG4</i> gene.....	313
4.8.2. Standard DNA sample preparation for <i>hGH</i> gene.....	316
4.8.3. qPCR experiments.....	319
4.9. Experiments for selection of the most efficient SP.....	328
4.9.1. Shake flask air-filtered bioreactor experiments.....	328
4.9.2. Laboratory-scale bioreactor experiments.....	332
4.9.3. Correlating SP efficiency with its physicochemical properties	339
4.10. Experiments for selection of the best oxygen limitation-induced promoter.....	351
4.10.1. Preliminary shake flask air-filtered bioreactor experiments..	351
4.10.2. Laboratory-scale bioreactor experiments.....	353
4.10.2.1. mRNA level of <i>hGH</i> in fed-batch fermentations.....	361
4.10.2.2. Proteases secretion in fed-batch fermentations.....	365
4.10.2.3. Metabolic flux analysis.....	366
4.10.2.3.1. Metabolic flux analysis for <i>rhGH</i> expression under P_{PYRK}	367
4.10.2.3.2. Metabolic flux analysis for <i>rhGH</i> expression under P_{PDC}	374
4.10.2.3.3. Metabolic flux analysis for <i>rhGH</i> expression under P_{GAP}	377
4.10.2.3.4. Comparative analysis of the metabolic fluxes distributions.....	379
4.10.2.3.5. Metabolic flux distribution at important nodes...	385
4.10.2.4. Deciding on the most efficient promoter.....	389

5. CONCLUSION	397
REFERENCES	403
APPENDICES	
A. Details on immature hGH nucleotide sequence.....	455
B. Media, buffers, and stock solutions	458
C. The synthesized plasmid (pUC57:: <i>PRSP</i>) containing selected oxygen limitation-induced promoters and endogenous secretion signal peptides.....	468
D. Primers recruited for PCRs and DNA sequencing.....	469
E. Reaction condition for single-digestion/double-digestion with utilized restriction enzymes in current research.....	470
E.1. <i>EcoRI</i> restriction enzyme.....	470
E.2. <i>XbaI</i> restriction enzyme.....	470
E.3. <i>Bsp119I</i> restriction enzyme.....	471
E.4. <i>NsiI</i> (<i>Mph1103I</i>) restriction enzyme.....	471
E.5. Double-digestion with <i>EcoRI</i> & <i>XbaI</i>	472
E.6. Double-digestion with <i>EcoRI</i> & <i>Bsp119I</i>	473
E.7. Double-digestion with <i>EcoRI</i> & <i>NsiI</i>	474
F. Location of the pUC-ori forward and hGH reverse primers used in pre-screening PCRs.....	475
G. The linearized constructs (with <i>NsiI</i> restriction enzyme).....	476
H. Location of the designed primers for preparation of standard solutions and rendering qPCR experiments.....	477
H.1. <i>hGH</i> gene specific primers.....	477
H.2. <i>ARG4</i> gene specific primers.....	477
I. Nucleotide sequence of the pGAPZ α A:: <i>hGH</i>	479
J. The standard curves used during experiments.....	480
K. Verification of the sequences of typical developed plasmids by BLAST analysis.....	487

K.1. pGAPZ α A:: <i>hGH</i> (Base plasmid).....	487
K.2. pPDCZ α A:: <i>hGH</i>	493
K.3. pGAPZ23A:: <i>hGH</i>	507
L. Amino acid and nucleotide sequence of the corresponding proteins of utilized SPs.....	514
L.1. Corresponding protein of SP13.....	514
L.2. Corresponding protein of SP23.....	514
L.3. Corresponding protein of SP24.....	515
L.4. Corresponding protein of SP26.....	515
L.5. Corresponding protein of SP34.....	516
L.6. Corresponding protein of <i>S. cerevisiae</i> α -MF.....	517
M. Nucleotide sequence of the utilized promoters.....	518
M.1. PYRK promoter.....	518
M.2. PDC promoter.....	519
M.3. THI3 promoter.....	521
M.4. GAP promoter.....	522
M.5. SDH promoter.....	522
N. Nucleotide sequence of the constructs of selected promoters.....	523
P. Typical results of single-digestion (with <i>Nsi</i> I restriction enzyme) prior to <i>Pichia</i> transfection.....	524
Q. Tabulated results of the qPCRs related with copy number determination.....	525
R. Selected qPCR reports.....	528
R.1. Selected run for promoter group.....	528
R.2. Selected run for signal peptide group.....	536
R.3. Selected run for mRNA level measurement.....	544
S. Complementary SDS-PAGE results related with shake flask bioreactor experiments.....	549
T. A glimpse on selected experimental techniques.....	551
T.1. Agarose gel electrophoresis.....	551

T.2. Polymerase chain reaction.....	552
T.3. Real-time polymerase chain reaction.....	554
T.4. SDS-PAGE.....	556
T.5. Dot-blot.....	558
T.6. Cell cycle synchronization.....	559
T.7. High pressure liquid chromatography.....	561
U. Metabolic reactions for <i>P. pastoris</i> used in metabolic flux analysis.....	565
V. Utilized DNA markers and protein marker.....	572
CURRICULUM VITAE.....	575

LIST OF TABLES

Table	Page
Table 1.1. Important features of hGH from UNIPROT.....	12
Table 1.2. Four isoforms of hGH as the products of the <i>GH-N</i> gene...	13
Table 2.1. A comparison of pharmaceutical expression systems.....	32
Table 2.2. Non-methylotrophic and methylotrophic yeast species recruited for r-protein production.....	34
Table 2.3. Advantages and disadvantages of <i>P. pastoris</i> expression system.....	47
Table 2.4. Common <i>P. pastoris</i> expression vectors.....	48
Table 2.5. Constitutive and inducible promoters used in <i>P. pastoris</i> for expression of r-proteins.....	55
Table 2.6. Advantages and disadvantages of P _{AOXI} –based expression system.....	60
Table 2.7. Activities of the various promoters in <i>P. pastoris</i> cells grown on different carbon sources based on (intracellular) GUS (β-glucuronidase) activity.....	71
Table 2.8. Comparison of the strength of P _{TEFI} and P _{GAP} in carbon-limited fed-batch culture.....	72
Table 2.9. <i>Pichia pastoris</i> codon usage bias.....	93
Table 2.10. Oxygen solubility in the pure water at an oxygen pressure of 1 atm.....	101
Table 2.11. Non-cytosolic proteases of yeasts.....	136
Table 2.12. Summary of the selected researches conducted related with rhGH production using <i>P. pastoris</i>	163
Table 3.1. LSLB agar medium composition for <i>E. coli</i> growth.....	175

Table 3.2. YPD agar medium composition for <i>P. pastoris</i> growth.....	175
Table 3.3. BMGY medium for pre-cultivation in shake flask and laboratory-scale bioreactor experiments.....	176
Table 3.4. Production medium for shake flask bioreactor experiments.	177
Table 3.5. PTM1 solution including trace salts for <i>P.pastoris</i> cultivation.....	177
Table 3.6. BSM for laboratory-scale bioreactor experiments.....	178
Table 3.7. Ingredients of the glucose solution in fed-batch phase of the laboratory-scale bioreactor experiments.....	179
Table 3.8. Typical amounts of PCR constituents.....	189
Table 3.9. Typical thermal profile for PCR.....	190
Table 3.10. Typical reaction mixture for single-digestion with <i>EcoRI</i> restriction enzyme.....	191
Table 3.11. Typical reaction mixture for double-digestion with <i>EcoRI</i> and <i>XbaI</i> restriction enzymes.....	192
Table 3.12. Typical reaction mixture for double-digestion with <i>EcoRI</i> and <i>Bsp119I</i> restriction enzymes.....	193
Table 3.13. Typical reaction mixture for double-digestion with <i>EcoRI</i> and <i>NsiI</i> restriction enzymes.....	193
Table 3.14. Typical ligation reaction mixture.....	195
Table 3.15. The recruited primers for sequencing of the isolated plasmids from <i>E. coli</i> cells for their verification prior to <i>Pichia</i> transfection.....	196
Table 3.16. Typical reaction mixture for single-digestion with <i>NsiI</i> restriction enzyme.....	197
Table 3.17. The primers used in PCR for verification of the genomic integration.....	201
Table 3.18. Utilized forward and reverse primers in pre-screening PCR.....	202
Table 3.19. Outer primers for preparation of the standard DNA for <i>ARG4</i> gene.....	205

Table 3.20. Outer primers for preparation of the standard DNA for <i>hGH</i> gene.....	205
Table 3.21. Inner primers for conducting qPCR experiments.....	206
Table 3.22. qPCR mixture used in copy number determination experiments.....	208
Table 3.23. qPCR experiments thermal profile.....	209
Table 3.24. Specifications of the laboratory-scale bioreactor used in current research.....	211
Table 3.25. Order and amount of the components to prepare template-primer mixture.....	218
Table 3.26. Order and amounts of RT mix component.....	218
Table 3.27. Sequential procedure of the silver nitrate staining.....	220
Table 3.28. The amount of the ingredients in glucose concentration measurement.....	227
Table 3.29. The amount of the ingredients in ethanol concentration measurement.....	229
Table 3.30. Specifications and conditions of the HPLC analysis.....	231
Table 4.1. PCR mixture for verification of BP with <i>GAP</i> forward and <i>AOX</i> reverse primers.....	241
Table 4.2. PCR thermal profile for verification of the BP with <i>GAP</i> forward and <i>AOX</i> reverse primers.....	241
Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used in <i>P. pastoris</i> expression system.....	244
Table 4.4. List of the potentially suitable endogenous SPs based on their D-scores in different combination.....	251
Table 4.5. Amino acid sequence and nucleotide sequence of the selected endogenous SPs along with the chromosomal location of the corresponding gene.....	252
Table 4.6. List of the proteins with increased expression in shift from normal to oxygen limitation conditions in both wild type and recombinant <i>P. pastoris</i> X-33 strain.....	254

Table 4.7. List of the remained proteins with increased expression in shift from normal to oxygen limitation conditions in both wild type and recombinant <i>P. pastoris</i> X-33 strain after first round of the screening.....	255
Table 4.8. Final selected proteins which their corresponding promoters were regarded as the promising oxygen limitation-induced promoters can be utilized in <i>Pichia</i> -based expressions.....	257
Table 4.9. The prediction results of the in-silico analyses conducted with <i>putative</i> promoter regions along with two positive controls.....	265
Table 4.10. SP constructs after inclusion of the required nucleotides in both ends.....	268
Table 4.11. The final form of the designed primers for SPs.....	268
Table 4.12. The final form of the designed primers for promoters.....	270
Table 4.13. PCR mixture for signal peptides amplification.....	274
Table 4.14. PCR thermal profile for SP13 amplification.....	275
Table 4.15. PCR thermal profile for SP23 and SP34 amplification.....	275
Table 4.16. PCR thermal profile for SP24 amplification.....	276
Table 4.17. PCR thermal profile for SP26 amplification.....	276
Table 4.18. Expected size of the developed plasmids harboring desired promoters.....	280
Table 4.19. Expected length of the bands in verification PCR of the putative transformants related with promoter constructs.....	284
Table 4.20. 1 st run PCR thermal profile for verification of pPDCZ α A::hGH transformants.....	284
Table 4.21. 1 st run PCR thermal profile for verification of pPYRKZ α A::hGH and pTHI3Z α A::hGH transformants.....	284
Table 4.22. 2 nd run PCR thermal profile for verification of pPDCZ α A::hGH, pPYRKZ α A::hGH, and pTHI3Z α A::hGH plasmids as template utilizing specific forward primer of each promoter along with AOX reverse primer.....	287

Table 4.23. Ligation mixture for development of the plasmids harboring desired SPs.....	290
Table 4.24. Expected size of the developed plasmids harboring selected endogenous SPs.....	291
Table 4.25. Double-digestion mixture for verification of the SP-harboring plasmids.....	294
Table 4.26. Digestion mixture for plasmids harboring desired endogenous SPs.....	298
Table 4.27. Digestion mixture for BP and plasmids harboring selected promoters.....	298
Table 4.28. PCR mixture for verification of the genomic integration of the plasmids possessing selected endogenous SPs.....	300
Table 4.29. PCR mixture for verification of the genomic integration of the plasmids possessing selected promoters.....	302
Table 4.30. Thermal profile of the PCR for verifying the insertion of the plasmids harboring selected promoters in <i>Pichia</i> genome.....	302
Table 4.31. PCR mixture for the pre-screening experiments.....	305
Table 4.32. PCR thermal profile for the pre-screening experiments.....	305
Table 4.33. Final pre-screening results of the verified <i>P. pastoris</i> transfectants.....	312
Table 4.34. Selected three probable single-copy r- <i>P. pastoris</i> strains for each of the developed plasmids.....	313
Table 4.35. Optimum PCR mixture for preparation of the standard DNA for ARG4 using <i>P. pastoris</i> genomic DNA as template.....	314
Table 4.36. PCR thermal profile for preparation of the ARG4 standard DNA.....	314
Table 4.37. Optimum PCR mixture for preparation of the standard DNA for hGH gene using base plasmid.....	316

Table 4.38. Optimum amount of the ingredients for conducting qPCR using hGH standard (1/100 and 1/1000 dilution) and r- <i>P. pastoris</i> genomic DNA (1/100 and 1/1000 dilutions) as template.....	318
Table 4.39. Optimum amount of the ingredients for conducting qPCR with <i>ARG4</i> -F and <i>ARG4</i> -R primers, and <i>ARG4</i> standard (1/1000, 1/10000, and 1/100000 dilutions), and r- <i>P. pastoris</i> genomic DNA (1/100 and 1/1000 dilution) as template.....	319
Table 4.40. Preliminary PCR conditions for obtaining optimum concentration of the inner primers utilized in qPCR.....	319
Table 4.41. Copy number of the <i>ARG4</i> and <i>hGH</i> genes in serial dilutions of the corresponding standard DNA samples.....	320
Table 4.42. Configuration of the samples in 1 st set of the qPCR experiments.....	321
Table 4.43. Typical configuration of the prepared samples for 2 nd set of the qPCR experiments.....	322
Table 4.44. Summarized results of 1 st and 2 nd sets of the qPCR experiments.....	323
Table 4.45. Samples configuration in 3 rd set of the qPCR experiments related to promoter.....	325
Table 4.46. Samples configuration in 3 rd set of the qPCR experiments related to SP group.....	326
Table 4.47. Results of 3 rd set of the qPCR experiments including promoter and SP groups separately.....	327
Table 4.48. The evaluated average copy number results deduced from triplicate experiments presented in Table 4.47.....	327
Table 4.49. Calculated physicochemical properties of hGH (mature and immature), its SP, and the N-terminus motif.....	342
Table 4.50. In-silico calculation of pI, GRAVY score, and aliphatic index related with utilized SPs.....	343

Table 4.51. In-silico calculation of pI, GRAVY score, and aliphatic index related with SPs utilized in Liang et al. study (2013).....	344
Table 4.52. Calculation of the D-scores of utilized SPs in combination with two extra proteins (CALB and Xyl) in presence of the N-terminus motif in order to check the effect of this motif on D-score.....	349
Table 4.53. D-scores of the SP23 with three proteins having different amino acids after N-terminus motif.....	350
Table 4.54. The comparison of three bioreactor experiments related with promoter constructs based on productivity (mg/L.h) and specific production rate of rhGH (q_{rhGH})	358
Table 4.55. Concentration of the organic acids in the fermentation medium related with production of rhGH with <i>P. pastoris</i> using three different promoters.....	360
Table 4.56. Configuration of the samples for qPCR experiments related to hGH mRNA level measurement	362
Table 4.57. The mRNA level of hGH during rhGH expression under P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5) at selected times of glucose fed-batch phase of the fermentation	363
Table 4.58. The summarized results of the conducted measurements related with extracellular rhGH level and its corresponding mRNA level	380
Table 4.59. Normalized flux distributions (with respect to glucose uptake rate) of rhGH production by <i>P. pastoris</i> with different promoters in low oxygen availability.....	369
Table 4.60. Variation in generated ATP throughout the bioprocess.....	373
Table 4.61. The amino acid composition of hGH expressed in current research by inclusion of its N-terminus fusion tag.....	385

Table Q.1. The processed data related with SP group in order to calculate average copy number.....	525
Table Q.2. The processed data related with promoter group in order to calculate average copy number.....	527

LIST OF FIGURES

Figure	Page
Figure 1.1. Members of the human endocrine system and their respective roles.....	6
Figure 1.2. The anatomic location of the pituitary gland and its anterior and posterior parts.....	8
Figure 1.3. Gene cluster of five closely-related genes coding for human growth hormones and placental lactogen hormones.....	9
Figure 1.4. 3-D structure of hGH exposing four available helices.....	11
Figure 1.5. Nucleotide sequence and amino acid sequence (FASTA format) of the pre-mature hGH.....	11
Figure 1.6. Illustration of the recombinant DNA technology.....	14
Figure 1.7. Transformation of the host cells with recombinant DNA (plasmid) and selection of the right (transformed) colonies.....	15
Figure 1.8. Recombinant plasmid isolation from the cloning host.....	15
Figure 2.1. Relative application of mammalian versus non-mammalian-based expression systems in the production of approved biopharmaceuticals.....	21
Figure 2.2. Biopharmaceutical products approvals, cumulative (1982-2014) and for the period of the study (2010 – July 2014).....	21
Figure 2.3. Linearized vector and its integration to the <i>P. pastoris</i> chromosome.....	51
Figure 2.4. Simple representation and explanation of the promoter region of a gene with the general structure of the gene under control....	54
Figure 2.5. The methanol utilization pathway in <i>P. pastoris</i>	58
Figure 2.6. The role of intermediate formaldehyde.....	63

Figure 2.7. Map of the pFLD.....	63
Figure 2.8. Central carbon metabolism of <i>P. pastoris</i>	67
Figure 2.9. P _{PGK1} and P _{AOX1} activities in directing the expression of the <i>B. subtilis</i> α -amylase	73
Figure 2.10. Induction of the endo-1,4- β -xylanase expression in <i>P. pastoris</i> host under the control of P _{ADH2} of <i>P. stipitis</i>	75
Figure 2.11. Expression of the bacterial hemoglobin (VHb) in <i>P. pastoris</i> under the control of P _{ADH2} of <i>P. stipitis</i>	76
Figure 2.12. Tripartite structure of a secretion SP.....	83
Figure 2.13. Two proposed mechanisms for protein secretion in eukaryotes.....	84
Figure 2.14. Available steps in oxygen transfer from a gas bubble to a single cell.....	106
Figure 2.15. Two-film theory as the simplest model for explanation of the mass transfer from gas phase to liquid phase.....	107
Figure 2.16. The schematization of dynamic method for determination of OUR and OTR in presence of microorganisms.....	113
Figure 2.17. Common pattern of O-glycosylation in yeasts.....	124
Figure 2.18. Structure of the core Asn-linked oligosaccharide.....	126
Figure 2.19. Representation of the protein folding, quality control, degradation, and secretion in yeasts.....	129
Figure 2.20. UPR as an intracellular signaling pathway from ER to nucleus.....	131
Figure 2.21. Simple representation of ERAD including its four well-defined steps.....	132
Figure 3.1. Schematic representation of pGAPZ α A.....	173
Figure 3.2. Schematic representation of pPICZ α A::hGH.....	173
Figure 3.3. The sequence of events which were followed in order to obtain the nucleotide sequence of the endogenous and exogenous secretion SPs	180

Figure 3.4. The sequence of events which were followed in order to obtain the nucleotide sequence of the desired promoters.....	181
Figure 3.5. Schematic representation of the typical single- or double-copy insertion in <i>Pichia</i> genome and location of the <i>hGH</i> reverse (red arrow) and <i>pUC</i> ori forward (blue) primers.....	203
Figure 3.6. Schematic representation of the steps in shake flask bioreactor experiments.....	210
Figure 3.7. Schematic representation of the steps in bioreactor experiments related with SPs.....	213
Figure 3.8. Schematic representation of the steps in bioreactor experiments related with promoters.....	215
Figure 3.9. Order of the components of the transfer stack in Western blot analysis.....	225
Figure 4.1. Nucleotide sequence of the <i>hGH</i> gene utilized in current research.....	233
Figure 4.2. Agarose gel electrophoresis result of the isolated plasmids of two different colonies of pGAPZ α A and four different colonies of pPICZ α A:: <i>hGH</i>	234
Figure 4.3. Further verification of pGAPZ α A and pPICZ α A:: <i>hGH</i> plasmids after single digestion of putative corresponding isolated plasmids with <i>Eco</i> RI restriction enzyme.....	235
Figure 4.4. Double-digested (with <i>Eco</i> RI and <i>Xba</i> I) pGAPZ α A and pPICZ α A:: <i>hGH</i> plasmids.....	236
Figure 4.5. Control of the purified double-digested pGAPZ α A and pPICZ α A:: <i>hGH</i>	237
Figure 4.6. Schematic representation of the base plasmid (BP).....	238
Figure 4.7. Isolated plasmids from 6 colonies related with putative transformants of BP.....	239
Figure 4.8. Double-digestion of five putative transformants' isolated plasmids by <i>Eco</i> RI and <i>Xba</i> I.....	240

Figure 4.9. PCR results with isolated plasmids related with two putative transformant colonies of BP.....	242
Figure 4.10. Simple representation of the pyruvate node in yeasts.....	256
Figure 4.11. The chromosomal location of the pyruvate kinase gene and its neighboring genes.....	258
Figure 4.12. Nucleotide sequence of the <i>putative</i> promoter region peculiar to pyruvate kinase gene with 345 bp length in <i>P. pastoris</i> GS115 strain.....	259
Figure 4.13. The chromosomal location of the pyruvate decarboxylase gene and its neighboring genes.....	260
Figure 4.14. Nucleotide sequence of the <i>putative</i> promoter region peculiar to pyruvate decarboxylase gene with 546 bp length in <i>P. pastoris</i> GS115 strain.....	261
Figure 4.15. Schematic structure of thiamine molecule.....	262
Figure 4.16. The chromosomal location of the <i>THI3</i> enzyme gene and its neighboring genes.....	263
Figure 4.17. Nucleotide sequence of the <i>putative</i> promoter region peculiar to pyrimidine precursor biosynthesis enzyme gene with 458 bp length in <i>P. pastoris</i> GS115 strain.....	263
Figure 4.18. The BLAST analysis result between the determined putative promoter region in <i>P. pastoris</i> GS115 strain and the whole genome of the <i>P. pastoris</i> DSMZ strain.....	264
Figure 4.19. pUC57:: <i>PRSP</i> plasmid verification.....	271
Figure 4.20. Double-digestion of the pUC57:: <i>PRSP</i> plasmid with <i>Nsi</i> I and <i>Eco</i> RI restriction enzymes.....	272
Figure 4.21. Verification of the eluted promoter constructs.....	273
Figure 4.22. PCR results for amplification of SP13 (45 minutes run on 1% agarose gel).....	277
Figure 4.23. PCR results for amplification of SP23 and SP34 (45 minutes run on 1.5 % agarose gel).....	277

Figure 4.24. PCR results for amplification of SP24 and SP26 (45 minutes run on 1.5 % agarose gel).....	278
Figure 4.25. Double-digestion of the BP with <i>EcoRI</i> and <i>NsiI</i> restriction enzymes.....	279
Figure 4.26. Developed promoter-free backbone after double-digestion of BP with <i>EcoRI</i> and <i>NsiI</i> restriction enzymes and subsequent gel extraction.....	279
Figure 4.27. Verification of the isolated plasmids of putative transformants related with selected promoters.....	281
Figure 4.28. Verification of the isolated plasmids of putative transformants related with selected promoters.....	282
Figure 4.29. Double-digested (with <i>EcoRI</i> and <i>XbaI</i>) plasmids of putative transformants related with selected promoters (P_{PDC} , P_{PYRK} , and P_{THI3}).....	283
Figure 4.30. 1 st run of the verification PCR specific to P_{PDC} -containing transformants.....	285
Figure 4.31. 1 st run of the verification PCR specific to P_{PYRK} -containing and P_{THI3} -containing transformants.....	286
Figure 4.32. Comparison of the results of the 1 st run of the verification PCR related with three promoter constructs.....	287
Figure 4.33. 2 nd run PCR for promoter constructs verification.....	288
Figure 4.34. Double-digestion of the BP with <i>Bsp119I</i> and <i>EcoRI</i> restriction enzymes.....	290
Figure 4.35 Isolated circular plasmids of twelve putative transformants of pGAPZ13A:: <i>hGH</i>	292
Figure 4.36. Isolated circular plasmids of six putative transformants of pGAPZ23A:: <i>hGH</i> along with six putative transformants of pGAPZ24A:: <i>hGH</i>	293
Figure 4.37. Isolated circular plasmids of six putative transformants of pGAPZ26A:: <i>hGH</i> along with six putative transformants of pGAPZ34A:: <i>hGH</i>	293

Figure 4.38. Double-digested (with <i>Bsp119I</i> & <i>XbaI</i>) plasmids of selected putative transformants (8 colonies) related to pGAPZ13:: <i>hGH</i>	294
Figure 4.39. Double-digested (with <i>Bsp119I</i> & <i>XbaI</i>) plasmids of selected putative transformants related to pGAPZ23:: <i>hGH</i> , pGAPZ24:: <i>hGH</i> , pGAPZ26:: <i>hGH</i> , and pGAPZ34:: <i>hGH</i>	295
Figure 4.40. Verification PCR related to putative transformants of pGAPZ13A:: <i>hGH</i>	296
Figure 4.41. Verification PCR related to putative transformants of pGAPZ23:: <i>hGH</i> , pGAPZ24:: <i>hGH</i> , pGAPZ26:: <i>hGH</i> , and pGAPZ34:: <i>hGH</i>	297
Figure 4.42. Genomic DNA isolates of 6 putative transfectants of pGAPZ13A:: <i>hGH</i>	300
Figure 4.43. Results of the verification PCR to confirm genomic integration related with pGAPZ13:: <i>hGH</i> transfectants.....	301
Figure 4.44. Results of the verification PCR to confirm genomic integration related with pGAPZ α A:: <i>hGH</i> and pTHI3Z α A:: <i>hGH</i> transfectants.....	303
Figure 4.45. Results of the verification PCR to confirm genomic integration related with pPDCZ α A:: <i>hGH</i> and pPYRKZ α A:: <i>hGH</i> transfectants.....	304
Figure 4.46. Pre-screening results of the verified transfectants of pGAPZ13A:: <i>hGH</i>	306
Figure 4.47. Pre-screening results of the verified transfectants of pGAPZ23A:: <i>hGH</i> and pGAPZ24A:: <i>hGH</i>	306
Figure 4.48. Pre-screening results of the verified transfectants of pGAPZ26A:: <i>hGH</i>	307
Figure 4.49. Single-digestion (by <i>EcoRI</i> restriction enzyme) results of pGAPZ23A:: <i>hGH</i> (3477 bp) and pGAPZ34A:: <i>hGH</i> (3480 bp).....	308
Figure 4.50. Pre-screening results of the verified transfectants of pGAPZ34A:: <i>hGH</i>	308

Figure 4.51. Pre-screening results of the verified transfectants of pGAPZ α A:: <i>hGH</i> (BP) and pPYRKZ α A:: <i>hGH</i>	309
Figure 4.52. Pre-screening results of the verified transfectants of pTHI3Z α A:: <i>hGH</i>	310
Figure 4.53. Pre-screening results of the verified transfectants of pPDCZ α A:: <i>hGH</i>	311
Figure 4.54. Preparation of the <i>ARG4</i> gene standard sample.....	315
Figure 4.55 Purified sample of the <i>ARG4</i> gene standard for qPCR experiments.....	315
Figure 4.56. Preparation of the <i>hGH</i> gene standard sample.....	317
Figure 4.57. Purified sample of the <i>hGH</i> gene standard for qPCR experiments.....	317
Figure 4.58. The average cell concentration of the cultivated single-copy <i>r-P. pastoris</i> strains related to SPs at t=24 h of production in triplicate shake flask bioreactor experiments.....	329
Figure 4.59. SDS-PAGE results of the 1 st shaker experiment at t=24 h of production.....	330
Figure 4.60. Dot-blot analysis for t=24 h samples of three shake flask bioreactor experiments.....	330
Figure 4.61. Relative amount of the normalized (by cell amount) secreted r-hGH by endogenous SPs as a percentage of the secretion directed by α -MF.....	332
Figure 4.62. Cell concentration and residual glucose concentration during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of α -MF (BR-1) and SP23 (BR-2).....	333
Figure 4.63 SDS-PAGE results related with BR-1 (<i>r-P. pastoris</i> pGAPZ α A:: <i>hGH</i> strain).....	334
Figure 4.64. SDS-PAGE results related with BR-2 (<i>r-P. pastoris</i> pGAPZ23A:: <i>hGH</i> strain).....	335

Figure 4.65. SDS-PAGE results related with <i>r-P. pastoris</i> pGAPZ α A:: <i>hGH</i> strain (BR-1) and <i>r-P. pastoris</i> pGAPZ23A:: <i>hGH</i> strain (BR-2) samples loaded in one gel.....	335
Figure 4.66. Western blot analysis result related with <i>r-P. pastoris</i> pGAPZ α A:: <i>hGH</i> strain (BR-1) and <i>r-P.pastoris</i> pGAPZ23A:: <i>hGH</i> strain (BR-2).....	336
Figure 4.67. Cconcentration of the secreted rhGH during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of α -MF (BR-1) and SP23 (BR-2).....	337
Figure 4.68. The normalized (by cell amount) secreted r-hGH during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of α -MF (BR-1) and SP23 (BR-2).....	338
Figure 4.69. Relative intracellular rhGH concentration during glucose fed-batch phase of the fermentation with corresponding <i>r-P.pastoris</i> strains of α -MF (BR-1) and SP23 (BR-2).....	339
Figure 4.70. Schematization of the N-terminus motif in expressed rhGH.....	342
Figure 4.71. SDS-PAGE results related with preliminary shake flask bioreactor experiment for validation of promoter activity of the selected <i>putative</i> promoter regions (t=24 h).....	352
Figure 4.72. Cell concentration during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	354
Figure 4.73. Residual glucose concentration during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	354
Figure 4.74. Produced ethanol concentration during glucose fed-batch phase of the fermentation in BR-3, BR-4, and BR-5.....	356
Figure 4.75. SDS-PAGE results related with BR-3, BR-4, and BR-5...	356

Figure 4.76. Western-blot results related with t=15 h samples of triple bioreactor experiments of <i>r-P. pastoris</i> strains corresponding to promoter constructs.....	357
Figure 4.77. Concentration of the secreted rhGH during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	358
Figure 4.78. hGH mRNA level in selected samples of bioreactor runs.	364
Figure 4.79. Protease activity during glucose fed-batch phase of the fermentation with corresponding <i>r-P. pastoris</i> strains of P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	365
Figure 4.80. The simplified metabolic map of <i>P. pastoris</i> for metabolic flux analysis.....	368
Figure 4.81. The normalized fluxes around G3P node in two periods (I, II) during rhGH expression under three different promoters P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	387
Figure 4.82. The normalized fluxes around Pyr node in two periods (I, II) during rhGH expression under three different promoters P _{PYRK} (BR-3), P _{PDC} (BR-4), and P _{GAP} (BR-5).....	388
Figure J.1. Glucose standard curve used in residual glucose concentration measurement in laboratory-scale bioreactor experiments	480
Figure J.2. Bovine serum albumin (BSA) standard curve used in total protein measurement.....	480
Figure J.3. Gluconic acid standard curve used in HPLC analysis.....	481
Figure J.4. Formic acid standard curve used in HPLC analysis.....	481
Figure J.5. Malic acid standard curve used in HPLC analysis.....	482
Figure J.6. Lactic acid standard curve used in HPLC analysis.....	482
Figure J.7. Acetic acid standard curve used in HPLC analysis.....	483
Figure J.8. Maleic acid standard curve used in HPLC analysis.....	483
Figure J.9. Citric acid standard curve used in HPLC analysis.....	484
Figure J.10. Fumaric acid standard curve used in HPLC analysis.....	484
Figure J.11. Succinic acid standard curve used in HPLC analysis.....	485

Figure J.12. Glutaric acid standard curve used in HPLC analysis.....	485
Figure J.13. Pyruvic acid standard curve used in HPLC analysis.....	486
Figure J.14. Oxalic acid standard curve used in HPLC analysis.....	486
Figure P.1. Single-digestion results of pGAPZ23A:: <i>hGH</i> and pGAPZ24A:: <i>hGH</i> plasmids.....	524
Figure P.2. Single-digestion results of pGAPZ α A:: <i>hGH</i> and pPDCZ α A:: <i>hGH</i>	524
Figure S.1. SDS-PAGE results of the 1 st shaker experiment at t= 32 h of production.....	549
Figure S.2. SDS-PAGE results of the 1 st shaker experiment at t=48 h of production.....	549
Figure S.3. SDS-PAGE results of the 2 nd shaker experiment at t=24 h of production.....	550
Figure S.4. SDS-PAGE results of the 3 rd shaker experiment at t=24 h of production.....	550
Figure T.1. SDS effect on protein and simple illustration of SDS-PAGE.....	558
Figure T.2. Schematic representation of the dot-blot analysis.....	559
Figure T.3. Schematic of the cell cycle.....	561
Figure T.4. Simple representation of an HPLC experiment components and their interaction.....	564

NOMENCLATURE

a	Area available for mass transfer	m^2
A	Stoichiometric matrix	
ATP_{op}	ATP obtained from oxidative phosphorylation	
$A_{Protease}$	Protease activity	
ATP_{sp}	ATP obtained from substrate-level phosphorylation	
C	Concentration in the medium	$g L^{-1}$ or $mol m^{-3}$
C_O^*	Saturated dissolved oxygen concentration	$mol m^{-3}$
C_A^*	Saturation concentration of component A	$mol m^{-3}$
CV	Coefficient of variation	%
D_{AB}	Diffusion coefficient of component A in B	$m^2 sec^{-1}$
DCW	Dry cell weight	g
DO	Dissolved oxygen	%
E	Biological enhancement factor	
H	Henry's constant	$L.atm.mol^{-1}$
k	Kinetic rate constant	
K_{La}	Overall liquid phase mass transfer coefficient	s^{-1}
L	Thickness of stagnant fluid film	m
K_s	Saturation constant	$g L^{-1}$
m	Maintenance coefficient	$g g^{-1} h^{-1}$
min.	Minute	
N	Agitation rate	min^{-1}
N_A	Mass transfer rate of component A	$mol m^{-3} sec^{-1}$
OUR	Oxygen uptake rate	$mol m^{-3} sec^{-1}$
OTR	Oxygen transfer rate	$mol m^{-3} sec^{-1}$
OD	Oxygen demand	$mol m^{-3} sec^{-1}$

OD ₆₀₀	Optical density at 600 nm	
P	Product amount (mass)	g
P	Pressure	Nm ⁻²
Q	Volumetric flow rate of inlet feed	L h ⁻¹
q	Specific formation or consumption rate	g g ⁻¹ h ⁻¹
r	Formation or consumption rate	g L ⁻¹ h ⁻¹
R	Reaction	
R _m	Mass transfer resistance	
Re	Reynolds number	
SD	Standard deviation	
Sc	Schmidt number	
Sec.	Second	
Sh	Sherwood number	
t	Cultivation time	h
t _i	Intended time	h
T	Bioreaction medium temperature	°C
u	Fluid velocity	ms ⁻¹
U	One unit of an enzyme	
V	Volume of the bioreactor	L
V _s	Superficial gas velocity in bioreactor	ms ⁻¹
X	Cell amount (mass)	g
Y	Yield	g g ⁻¹
\bar{Y}	Overall yield	g g ⁻¹

Greek Letters

α-MF	<i>Saccharomyces cerevisiae</i> mating factor	
ρ	Density	g L ⁻¹
μ	Specific growth rate	h ⁻¹
μ ₀	Pre-determined specific growth rate	h ⁻¹

Subscripts

0	Refers to initial condition
a	Refers to apparent variable
A	Refers to component A in a mixture/solution
c	Refers to cell
Cr.	Refers to critical
d	Refers to dynamic method
de	Refers to death
l	Refers to liquid
max	Refers to maximum amount of a parameter
O ₂	Refers to oxygen
P/s	Refers to product over substrate
P/x	Refers to product over biomass
x/s	Refers to biomass over substrate

General Abbreviations

ACTH	Adrenocorticotrophic hormone
ADH	Alcohol dehydrogenase
AGE	Agarose gel electrophoresis
AOD	Alternative oxidase
AOX	Alcohol oxidase
ARG4	Argininosuccinate lyase
BHK	Baby Hamster kidney cells
BLAST	Basic local alignment search tool
BMGY	Buffered minimal glycerol complex medium
bp	Base pair
BP	Base plasmid
BR	Bioreactor
BSA	Bovine serum albumin
BSM	Basal salt medium
CAI	Codon adaptation index

cAMP	Cyclic adenosine 5'-monophosphate
cDNA	Complementary Deoxyribonucleic acid
CHO	Chinese Hamster ovary
CJD	Creutzfeld-Jacob disease
CNX	Calnexin (chaperone)
CRT	Calreticulin (chaperone)
CT /C _t	Threshold cycle
DAB	3,3'-Diaminobenzidine
DHAS	Dehydroxyacetone synthase
DNA	Deoxyribonucleic acid
DP	Docking protein
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-aided degradation
ERQC	Endoplasmic reticulum quality control
FDA	Food and drug administration
FLD	Formaldehyde dehydrogenase
FMD	Formate dehydrogenase
GAP/GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GH	Growth hormone
GRAS	Generally regarded as safe
GRAVY	Grand average of hydropathy
GSM	Genome-scale model
GTF	General transcription factor
HEK	Human embryonic kidney
hGH	Human growth hormone
HIV	Human immune deficiency virus
HPLC	High pressure liquid chromatography
H-region	Hydrophobic region of a signal peptide
HSA	Human serum albumin

HSP	Heat shock protein
IB	Inclusion body
LB	Luria broth (lysogeny broth)
LSLB	Low-salt Luria broth
MCS	Multiple cloning site
MFA	Metabolic flux analysis
MOX	Methanol oxidase
mRNA	Messenger ribonucleic acid
MUT	Methanol utilization pathway
NCBI	National center for biotechnology information
NTC	Non-template control
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDC	Pyruvate decarboxylase
PDI	Protein disulfide isomerase
PGH	Placental growth hormone
PGK	Phosphoglycerate kinase
pI	Isoelectric point
PMSF	Phenyl methyl sulfonyl fluoride
PPP	Pentose phosphate pathway
PRL	Prolactin
PTM	Post-translational modification
PTM1	Pichia trace minerals medium
PVDF	Polyvinylidene difluoride
PYRK	Pyruvate kinase
RBS	Ribosome binding site
rDNA	Recombinant DNA
rhGH	Recombinant human growth hormone
RNA	Ribonucleic acid
ROS	Reactive oxygen species

RSCU	Relative synonymous codon usage
RT	Room temperature
SCP	Single cell protein
SD	Standard deviation
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis
SP	Signal peptide
SRP	Signal recognition particle
TBP	TATA-binding protein
TBS	Tris buffered saline
TCA	Tricarboxylic acid
TE	Tris EDTA
TEF	Translation elongation factor
TF	Transcription factor
TFBS	Transcription factor binding site(s)
THI3	Pyrimidine precursor biosynthesis enzyme
TM	Transmembrane
TPP	Thiamine pyrophosphate
TSH	Thyroid-stimulating hormone
TT	Transcriptional termination
tRNA	Transfer ribonucleic acid
UPR	Unfolded protein response
UTR	Untranslated region
YPD	Yeast extract peptone dextrose
YEP	Yeast extract peptone
Z	Objective function

Abbreviations Used in Metabolic Flux Analysis

Ac	Acetate
AcCoA	Acetyl coenzyme A
AcCoAm	Acetyl coenzyme A, mitochondrial
Acet	Acetaldehyde

ADP	Adenosine 5'-diphosphate
aKG	α -ketoglutarate, mitochondrial
Ala	L-Alanine
AMP	Adenosine 5'-monophosphate
Arg	L-Arginine
Asn	L-Asparagine
Asp	L-Aspartate
ATP	Adenosine 5'-triphosphate
CaP	Carbamoyl-phosphate
CARBH	Carbohydrates
CDP	Cytidine 5'-diphosphate
Chor	Chorismate
Cit	Citrate, mitochondrial
CMP	Cytidine 5'-monophosphate
CO ₂	Carbondioxide
CTP	Cytidine 5'-triphosphate
Cys	L-Cysteine
dATP	2'-Deoxy-ATP
dCTP	2'-Deoxy-CTP
dGTP	2'-Deoxy-GTP
DHF	Dihydrofolate
dTTP	2'-Deoxy-TTP
E4P	Erythrose 4-phosphate
EPO	Erythropoietin
EtOH	Ethanol
F10THF	N ¹⁰ -Formyl-THF
F6P	Fructose 6-phosphate
FA	Fatty acids
FADH ₂	Flavine adenine dinucleotide, reduced
For	Formate
Form	Formaldehyde

Fruc	D-Fructose
Fum	Fumarate, mitochondrial
G3P	Glyceraldehyde 3-phosphate
G6P	Glucose 6-phosphate
GDP	Guanosine 5'-diphosphate
Glc	Glucose
Gln	L-Glutamine
Glu	L-Glutamate
Gly	L-Glycine
GMP	Guanosine 5'-monophosphate
GTP	Guanosine 5'-triphosphate
H ₂ S	Hydrogen sulfide
His	L-Histidine
HSer	Homoserine
ICit	Isocitrate, mitochondrial
Ile	L-Isoleucine
IMP	Inosinemonophosphate
Kval	Ketovaline
Lac	Lactic acid
Leu	L-Leucine
Lys	L-Lysine
MaCoA	Malonyl-CoA
Mal	Malate, mitochondrial
MeOH	Methanol
Met	L-Methionine
MeTHF	N ⁵ - N ¹⁰ -methenyl-THF
MetTHF	N ⁵ - N ¹⁰ -methylene-THF
MTHF	N ⁵ - methyl-THF
NADH	Nicotinamide-adeninedinucleotide (reduced)
NADHm	Nicotinamide-adeninedinucleotide, mitochondrial (reduced)
NADPH	Nicotinamide-adeninedinucleotide phosphate (reduced)

NH ₄	Ammonia
OA	Oxaloacetate
OLE	Oleic acid
PA	Phosphatidate
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PEP	Phosphoenolpyruvate
3PG	3-Phospho-D-glycerate
Phe	L-Phenylalanine
Pi	Inorganic orthophosphate
PINS	1-Phosphatidyl-D-myo-inositol
PLLM	Palmitoleic acid
PLM	Palmitic acid
PPi	Inorganic pyrophosphate
PRAIC	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide
Pro	L-Proline
PRPP	5-Phospho-D-ribosylpyrophosphate
PS	Phosphatidylserine
Pyr	Pyruvate
R5P	Ribulose 5-phosphate
Rib5P	Ribose 5-phosphate
S7P	Sedoheptulose-7-phosphate
Ser	L-Serine
Sorb	Sorbitol
STE	Stearate
Suc	Succinate, mitochondrial
SucCoA	Succinate coenzyme A, mitochondrial
THF	Tetrahydrofolate
Thr	L-Threonine
Trp	L-Tryptophan
Tyr	L-Tyrosine

UDP	Uridine 5'-diphosphate
UMP	Uridine 5'-monophosphate
UTP	Uridine 5'-triphosphate
Val	L-Valine
Xyl5P	Xylulose 5-phosphate

CHAPTER 1

INTRODUCTION

Biotechnology as an assembly of engineering fields like reaction engineering and separation technology, microbiology, and enzyme technology, has harnessed the capabilities of living cells in order to improve and accomplish manufacturing novel products spanning from animal feedstock to pharmaceuticals (Nielsen et al., 2002). By introduction of modern biotechnology after emergence of new molecular biology techniques, like recombinant DNA technology (rDNA), biotechnology field was revolutionized by feasibility of direct manipulation of the genome of the cells. In consequence, along with increasing the knowledge about the fermentation techniques and “microbial metabolic systems” rational modification and alteration of the available cells recruited in industrial processes was changed to a common method in order to conduct processes in an economically-efficient and environmentally-sustainable way (Nielsen et al., 2003; Erickson et al., 2012).

After novelties came into existence in medical and agricultural sector, industrial biotechnology, known as “white biotechnology”, as the third wave of the new biotechnology era (Erickson et al., 2012) has introduced itself as a reliable alternative in implementation of chemical processes, especially by confronting the declining reservoirs of fossil fuels, global warming, and price of feedstocks. Industrial biotechnology is the recruitment of living cells or their corresponding enzymes in order to sustainable production of bio-based products from renewable sources (such as agricultural crops) by means of modern biotechnology (Tang and Zhao, 2009). As biotechnology is based on the living

systems, their ability in efficiently managing of their chemistry in order to produce the waste in mostly biodegradable and recyclable form can be exploited in enhancing energy efficiency and “environmental profile” of chemical reactions (Erickson et al., 2012). Therefore, industrial biotechnology mainly accompanies decreased amount of generated waste, evolved greenhouse gas, and utilized energy. In addition, the lower operating and capital cost of these processes make their difference from chemical counterparts very conspicuous (Tang and Zhao, 2009). Furthermore, higher reaction rate and increased efficiency are among the major performance benefits of industrial biotechnology-based reactions (Soetaert and Vandamme, 2004). It had been predicted that the share of industrial biotechnology in chemical industry will become 10% in 2010 (Tang and Zhao, 2009).

Development of the technologies and processes related with microorganisms, enzymes, and their genetic engineering for commercialization has occupied the spotlight (Erickson et al., 2012). Protein engineering, metabolic engineering, synthetic biology, systems biology, and downstream processing are among the “biological toolsets” which their advances have profound effect on industrial biotechnology; protein engineering has paved the way for obtaining tailored enzymes, metabolic engineering by interfering in cellular functions has provided us with methods for fine-tuning (desired) gene expression, synthetic biology has developed synthetic gene expression circuits which can be a prelude for de novo genome synthesis in synthetic microorganisms for intended industrial purposes, systems biology has led to consider the cellular network as a whole in order to optimize or improve any industrial strain, and finally, recent advances in purification and separation of biological products has conducted to decrease in the downstream processing cost and more favorable industrial processes (Tang and Zhao, 2009). The outcome of the utilization of such life-science techniques will be finding novel enzymes or improving the state of the present industrial strains (Erickson et al., 2012).

In the realm of pharmacology, "pharmaceutical drug", also called a medication or medicine, is any (chemical) substance intended for use in prevention, diagnosis, cure, treatment, or mitigation of a disease or to

enhance wellbeing by causing a physiological change in body (Web 1). Pharmaceutical products can be either produced in biological systems or synthesized chemically. Traditional vaccines and proteins like blood factors and immunoglobulins are the examples of the medical products obtained from biological sources before renaissance in methods of molecular biology. By emergence of the novel molecular biology techniques, the term biopharmaceutical was attributed to the biological medical products generated or produced via recruiting modern molecular biology techniques in order to distinguish them from traditional biological products. However, generally, biosynthesized products including traditional and molecular biology-derived entities, all are referred as “biotech products” or biopharmaceuticals (Walsh, 2014). Biopharmaceuticals are inalienable part of the modern medicine (Vogl et al., 2013) especially by considering the 25% share of this category in commercial pharmaceuticals (Martinez et al., 2012). The total number of biopharmaceuticals which have been licensed in United States and European Union have been brought to 246; however, withdrawal of 34 products from the market (after approval) has decreased the number of the approved and marketed products to 212 (Walsh, 2014). The amount of manufactured biopharmaceuticals has been reported to be around 26400 kg (of pure protein) in 2010 (Walsh, 2014). Additionally, the global biologics market totaled \$200.6 billion in 2013 and reached to nearly \$234 billion in 2014. The total market is expected to grow to \$386.7 billions by the end of 2019 (Web 2).

In a narrow sense recombinant therapeutic proteins and nucleic acid-based products fall under the category of biopharmaceuticals which can be extended to all engineered cells and tissue-based products (Vogl et al., 2013). Protein-based therapeutics are considered as the largest class of new chemical entities (Nel et al., 2009) and the major intend in biotechnology industry has been therapeutic proteins especially those engaged with human health (Ascacio-Martinez et al., 2004). Proteins in either form of native or recombinant can benefit different sectors including pharmaceutical, enzyme,

or agricultural industry (Demain and Vaishnav, 2009). As the basic elements of the life and having the most diverse and dynamic role in the body, proteins can play role as enzymes (biocatalysts), hormones, cell signaling elements, channels in cell membrane, cell cycle directors, cell adhesion mediators, or structural components of the living assemblies (Leader et al., 2008). Increasing demand for therapeutic and industrial proteins (enzymes) as a consequence of growing human population and, thus, the related issues such as hygiene, health, and environment and also realizing the mechanisms of diseases, are driving forces for researches targeting discovery of novel, easily-manipulated, and economic methods to meet the global requirements for production of desired proteinaceous products. It should be emphasized that, in addition to the applications of industrial enzymes in different industries, many enzymes are used also as therapeutic agents in treating diseases such as cancer or cystic fibrosis (Demain and Vaishnav, 2009). The global market for industrial enzymes was worth nearly \$4.5 billions in 2012 and nearly \$4.8 billions in 2013; the market is expected to reach around \$7.1 billions by 2018 (Web 2). The market of the recombinant proteins along with industrial enzymes was supposed to be around \$169 billions in 2014 (Martinez et al., 2012) and, furthermore, the global protein therapeutics market reached nearly \$174.7 billion in 2015 and should reach nearly \$248.7 billion by 2020 (Web 2). On the other hand, another research estimated the global therapeutic proteins market to be valued at \$93 billions in 2010. The market is forecasted \$141.5 billions in 2017 (Web 3). Among the protein therapeutics, monoclonal antibodies (mAb) have been pioneers in market in 2010 by sales exceeding \$18 billions which was followed by hormones with \$11 billions (Martinez et al., 2012). In the period between 2010 and 2014 also 54 new biologics have been introduced to the market where 17 of them are mAbs and 9 of them are hormones (Walsh, 2014); conducted researches also confirms that the high demand for monoclonal antibodies will drive the market (Web 3).

Therapeutic proteins are more lucrative candidates compared to chemically-synthesized (small-molecule) drugs by taking into consideration that: i) they are highly specific which leads to less interference with normal biological functions and, thus, less adverse effects, ii) they can afford a variety of biological roles which can not be simply devised in a small chemical entity, iii) they are also a biological agent which are less prone to provoke immune response in the body, and iv) they can replace gene therapies in the cases that a deleted or mutated gene is the origin of a disease (Leader et al., 2008).

Hormones, as one of the sub-groups of therapeutic proteins, have regulatory activities and are used in the case of endocrine disorders (hormone deficiencies). Therapeutic hormones are utilized as medicine in order to compensate deficiency or abnormality of the corresponding hormones (Leader et al., 2008). In order to maintain homeostasis the different parts of the multi-cellular organisms need to communicate with each other (Hiller-Sturmhöfel and Bartke, 1998). This communication is further required for proper responding to the environmental changes. Therefore, necessity of message transport within the organism has led to development of two systems: nervous system and endocrine system (Kleine and Rossmannith, 2016) (i.e., neuroendocrine). Hormones are peptides/proteins, lipids or amino acid derivatives that are produced by endocrine glands (Figure 1.1), upon sensing a change (stimuli) by sensor cells (releasing cells), and are released to the bloodstream (or interstitial fluid) and are carried by blood toward their target cells. Upon receiving to the target cells, the hormone interacts with its high affinity receptors either on the cell surface or inside the cell and this, in turn, initializes series of biochemical reactions inside the cell which ends with regulated function of the cell which is the response of the target cell to the change (stimuli) (Hiller-Sturmhöfel and Bartke, 1998). In brief, hormones are chemical messengers that carry information to the organ (Schauer, 1972).

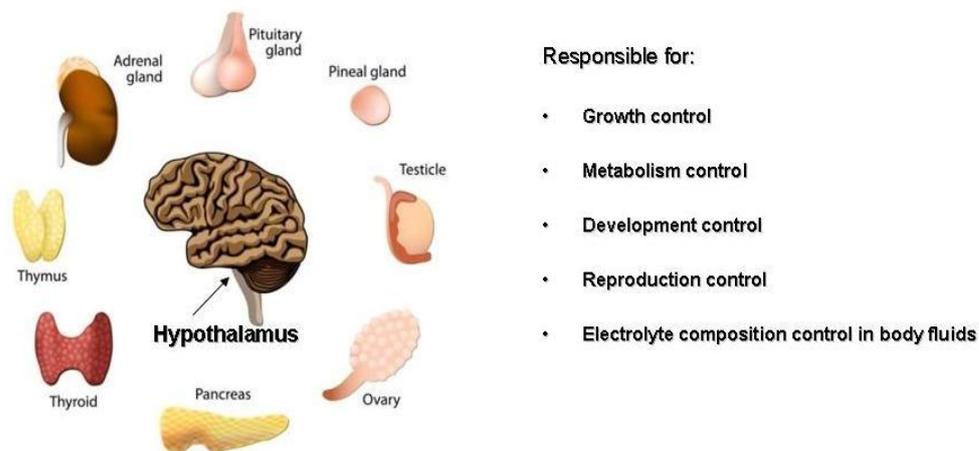


Figure 1.1. Members of the human endocrine system and their respective roles; reproduced from (Web 4).

In the cases that the hormone receptors reside on the cell surface while span the cell membrane, the hormonal signaling, as an example of signal transduction, is mediated by second messengers (e.g., cAMP) where this secondary signal can also be amplified through signaling cascades; water soluble peptide and amine hormones act by this mechanism. On the other hand, where the hormones enter the cell, hormone receptor locates at the nucleus or cytoplasm; in this situation the message transport (signal transduction) is conducted by hormone-receptor protein complex. Water-insoluble hormones like steroid hormones because of their hydrophobic nature can pass through the cell membrane and reach their corresponding receptors within the cell (Nelson and Cox, 2005).

Regulation of the hormones release is achieved by a complex of neuronal and hormonal signals where the hypothalamus locates at the center of the endocrine regulatory system. Upon receiving messages from the nervous system, hypothalamus in response to them produces releasing factors which are, indeed, regulatory hormones that are passed to the pituitary gland (Figure 1.2) through special blood vessels. Pituitary gland can

be regarded as the “command post” for regulation of the hormones in the body (Schauer, 1972). Between the two functionally distinct parts of the pituitary gland, i.e., posterior and anterior, the anterior pituitary gland is the site which replies to the hypothalamic hormones carried in the blood and, thus, makes specific hormones to travel within circulatory system and to transduce the signals to the other endocrine glands. Consequently, the initial electrical signal to the hypothalamus is amplified for at least a million-fold and changes into a secreted hormone in an amount of milligram (Nelson and Cox, 2005).

It is obvious that there should be a tight control over the hormonal release in order to prohibit any kind of disequilibrium and maintaining body’s homeostasis (Hiller-Sturmhöfel and Bartke, 1998; Schauer, 1972). There is always a consistent feedback to the hypothalamus and pituitary gland from the target gland in order to make sure that the hormone activity is in the appropriate limits (Hiller-Sturmhöfel and Bartke, 1998). The mechanism of the control in hormone secretion is, generally, feedback inhibition or negative feedback. Therefore, any unnecessarily excess amount of a specific hormone leads to cease of its secretion by inhibition of earlier steps in the secretion cascade (Nelson and Cox, 2005). However, there are positive feedback controls over some of the hormone systems where the target gland hormone affects hypothalamus and pituitary glands and, thus, its secretion is increased (Hiller-Sturmhöfel and Bartke, 1998).

The anterior pituitary gland secretes several hormones which not only can directly influence the target cell but also can stimulate the target gland to produce target hormone. These hormones include: adrenocorticotrophic hormone (ACTH), gonadotropins, thyrotropin or thyroid-stimulating hormone (TSH), growth hormone (GH), and prolactin, amongst, GH and prolactin directly act on their target organ. GH is the most abundant hormone among pituitary-derived hormones (Hiller-Sturmhöfel and Bartke, 1998).

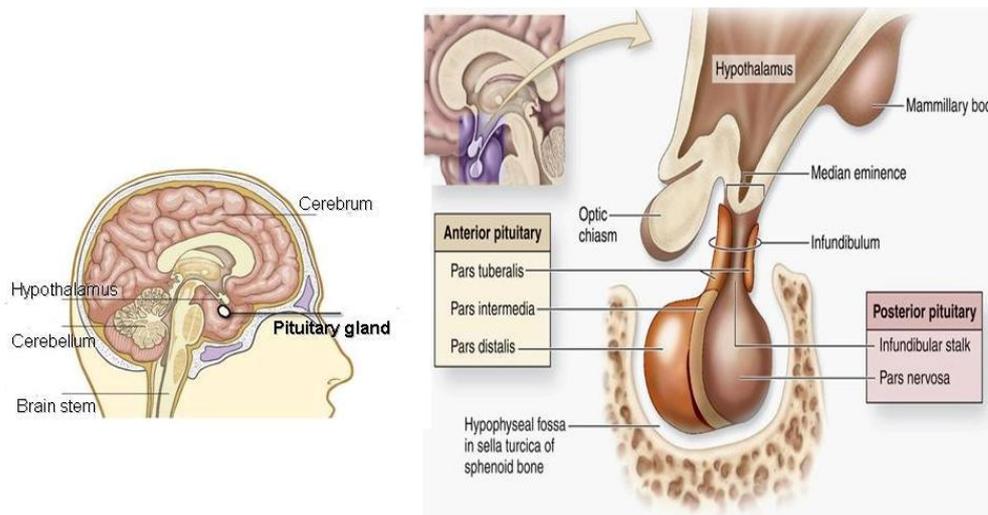


Figure 1.2. The anatomic location of the pituitary gland and its anterior and posterior parts (Web 5; Web 6).

Pituitary growth hormone and prolactin (PRL), together with the homologous placental hormones, comprise a family of related protein hormones (Wallis, 1981). There is a 46.83-kb gene cluster (Baumann, 2009) on chromosome 17 (q22-q24) which contains five closely-related genes (Figure 1.3); they are in the same transcriptional orientation and the order of the genes from 5' to 3' is *GH-N* (growth hormone normal or pituitary growth hormone), *PL-L* (placental lactogen-like hormone), *PL-A* (placental lactogen gene A), *GH-V* (growth hormone variant), and *PL-B* (placental lactogen gene B). The *PL* genes exclusively and *GH-V* gene, primarily if not exclusively, are expressed in the placenta, and *GH-N* is expressed exclusively in the pituitary (Zhan et al., 2005; Baumann, 2009). High homology in 5' flanking and coding sequences but divergence in 3' flanking region has been confirmed for these two groups of genes (Barsh et al., 1983).

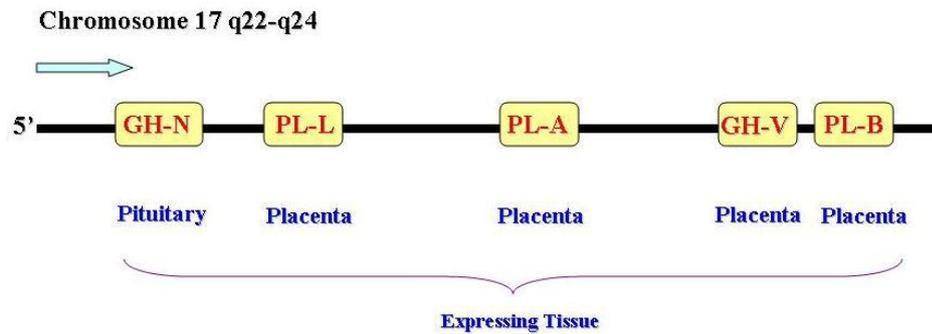


Figure 1.3. Gene cluster of five closely-related genes coding for human growth hormones and placental lactogen hormones.

Placental growth hormone (PGH) is the product of the *GH-V* gene and during pregnancy *GH-N* expression in the mother is suppressed and *GH-V* becomes the predominant GH in the mother (Handwerger and Freemark, 2000; Baumann, 2009). It is structurally similar to *GH-N* gene and its product is single-chain and possesses two disulfide bridges; however, there is a 13-amino acid difference and a consensus sequence as glycosylation site (Baumann, 2009).

The GH which is known as human growth hormone or normal hGH is the product of the *GH-N* gene, namely the prototype pituitary GH (Zhan et al., 2005; Baumann, 2009). Human growth hormone known as somatotropin is a 191-amino acid, anionic, non-glycosylated (Kim et al., 2013), single-chain polypeptide with two disulfide bonds for internal cross-linking and stabilization of its structure (Zhan et al., 2005; Baumann, 2009; Rezaei et al., 2013). Approximately 50% of the hGH residues hold α -helical conformation (Catzel et al., 2003) in a 4-helical twisted bundle form (Figure 1.4) where the bundle is necessary for functional interaction with the GH receptor (Baumann, 2009). hGH is regarded as one of the most vital hormones of the

body because of its indispensable wide-range roles in biological functions (Kim et al., 2013) such as development and growth promotion, metabolism of carbohydrates, fats, and proteins and immunity (Hiller-Sturmhöfel and Bartke, 1998). As a therapeutic agent it benefits the treatment of hypopituitary dwarfism, growth hormone deficiency, Turner's syndrome, chronic renal failure and human immune deficiency virus (HIV) syndrome. It is also used in the treatment of injuries, bone fractures, bleeding ulcers, burns and maintaining health in the elderly (Ecamilla-Treviño et al., 2000). The nucleotide sequence and amino acid sequence of the hGH has been represented in Figure 1.5 and related data are available in Appendix A. It should be emphasized that, the growth promoting function of hGH is mainly related to its N-terminal segment and any manipulation of N-terminal would result in decreased function. However, the residual amino acids in C-terminal would not impact the bioactivity of the hGH (Xu and Kieliszewski, 2011).

Like most of the proteins, hGH also undergoes several post translational modifications (PTMs) such as disulfide bond formation and phosphorylation; although it is not glycosylated (Table 1.1). Its phosphorylation seems to be very important in its signal transduction via phosphorylation and dephosphorylation mechanism (Zhan et al., 2005). However, the biological activity of hGH is supposed to be independent of its disulfide bonds (Kim et al., 2013).

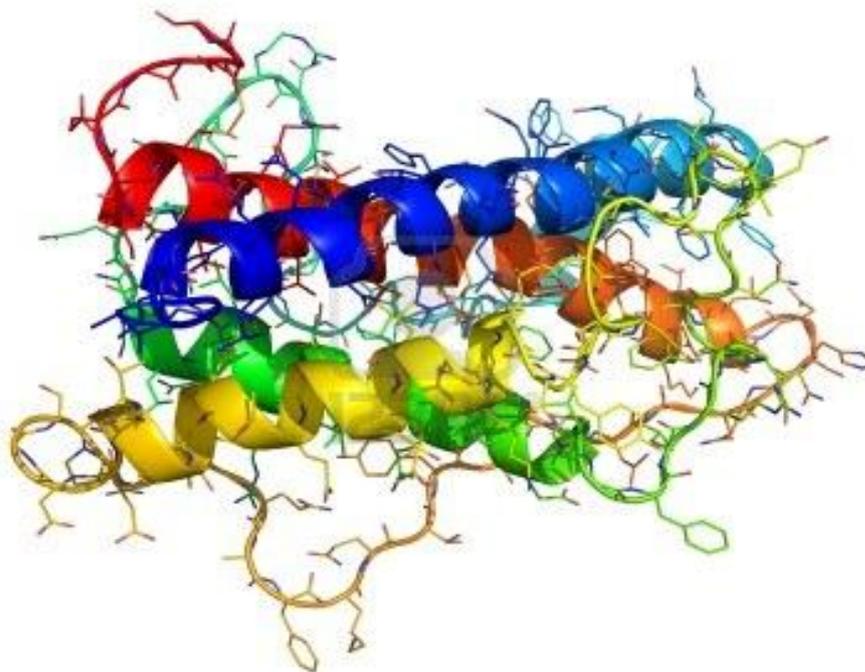


Figure 1.4. 3-D structure of hGH exposing four available helices (Web 7).

```

ATG GCT ACA GGC TCC CGG ACG TCC CTG CTC CTG GCT TTT GGC
CTG CTC TGC CTG CCC TGG CTT CAA GAG GGC AGT GCC TTC CCA
ACT ATA CCA CTA TCT CGT CTA TTC GAT AAC GCT ATG CTT CGT
GCT CAT CGT CTT CAT CAG CTG GCC TTT GAC ACC TAC CAG GAG
TTT GAA GAA GCC TAT ATC CCA AAG GAA CAG AAG TAT TCA TTC
CTG CAG AAC CCC CAG ACC TCC CTC TGT TTC TCA GAG TCT ATT
CCG ACA CCC TCC AAC AGG GAG GAA ACA CAA CAG AAA TCC AAC
CTA GAG CTG CTC CGC ATC TCC CTG CTG CTC ATC CAG TCG TGG
CTG GAG CCC GTG CAG TTC CTC AGG AGT GTC TTC GCC AAC AGC
CTG GTG TAC GGC GCC TCT GAC AGC AAC GTC TAT GAC CTC CTA
AAG GAC CTA GAG GAA GGC ATC CAA ACG CTG ATG GGG AGG CTG
GAA GAT GGC AGC CCC CGG ACT GGG CAG ATC TTC AAG CAG ACC
TAC AGC AAG TTC GAC ACA AAC TCA CAC AAC GAT GAC GCA CTA
CTC AAG AAC TAC GGG CTG CTC TAC TGC TTC AGG AAG GAC ATG
GAC AAG GTC GAG ACA TTC CTG CGC ATC GTG CAG TGC CGC TCT
GTG GAG GGC AGC TGT GGC TTC

```

```

MATGSRTSLLLAFLGLLCLPWLQEGSAFPTIPLSR LFDNAMLRHRLHQ LAFD TYQE
FEEAYIPKEQKYSFLQNPQTS LCFSESIPTPSNREETQ QKSNLELLRIS LLLIQSW
LEPVQFLRSV FANSLVYGASDSNVYDLLKDLEEGIQ TLMGRLEDGSPRTGQIFKQT
YSKFDTNSHNDDALLKNYGLLYCFRKMDK VETF LRI VQCRSVEGSCGF

```

Figure 1.5. Nucleotide sequence and amino acid sequence (FASTA format) of the premature hGH. The “blue” region is the secretion signal peptide and the “red” region is the mature hGH (Web 8).

Table 1.1. Important features of hGH from UNIPROT. Attributed residue numbers are for pre-mature hGH.

Feature	Explanation
Length of Signal sequence	26 amino acids
Disulfide bonds	79 th →191 st and 208 th →215 th
Metal-binding sites (Zinc binding)	44 th residue and 200 th residue
Phosphorylation	77 th , 132 nd , 176 th serine residues
Deamidation	163 rd residue (glutamine) 178 th residue (asparagine)

It has been revealed that because of the alternative splicing, different PTMs and proteolytic processing, hGH (coded by *GH-N* in GH gene cluster) has extensive molecular heterogeneity (Baumann, 2009). There have been identified 4 isoforms for the *GH-N* gene product (Zhan et al., 2005) which can be seen in Table 1.2. All the isoforms are obtained from a 217-amino acid precursor by theoretical molecular weight (M_w) of 24.847 kDa and after cleavage of 26-amino acid secretion signal peptide (SP). The major and the most common one is the 22-kDa isoform which is known as normal hGH or 22K-GH. Expression proportions of the splicing-resulted isoforms are 87.5%, 8%, 3.3%, and 1.1% for isoforms 1 to 4, respectively (Zhan et al., 2005). The isoform 2 is referred to as 20K-GH which has a similar structure to normal hGH and is obtained by alternative splicing in one of the exons and is expressed at a level around 5-10% of the normal hGH (Baumann, 2009). Although the structure of the GH-V protein is similar to GH-N, its inclination for alternative splicing is less than the GH-N and, therefore, 20K isoform of GH-V has not yet been identified (Baumann, 2009).

Table 1.2. Four isoforms of hGH as the products of the *GH-N* gene (Zhan et al., 2005).

Isoform	Theoretical M_w (kDa)	No. of amino acid	Deleted segments
1	22.129	191	1-26
2	20.274	176	1-26 and 58-72
3	17.843	152	1-26 and 111-148
4	17.083	145	1-26 and 117-162

Many proteins available in the human body have pharmaceutical importance; however, obtaining those proteins from natural sources are cumbersome. The solution has been found in the late 70's by introducing recombinant DNA (rDNA) technology, producing foreign proteins in new non-producing cells, using *Escherichia coli* as biological framework (Ferrer-Miralles et al., 2009). This method of high titer production of proteins using engineered living entities (rDNA technology) provides an alternative to protein extraction from natural sources (Porro et al., 2005). All the therapeutic agents by rDNA technology can be produced in a large quantity and also more economically as this method has a relatively inexpensive series of procedures (Ferrer-Miralles et al, 2009) and, therefore, nowadays rDNA technology along with the hybridoma cell technology are the major methods of producing marketed protein drugs (Corchero et al., 2012). The pioneer products in the realm of biopharmaceuticals were recombinant human insulin and recombinant human growth hormone (rhGH) entered the market in the early 1980s which only were obtained by tissue extraction before (Graumann et al., 2006). Then, therapeutic agents such as hormones, cytokines, growth factors, antibiotics, vaccines, blood products like albumin, thrombolytics, fibrynolytics, clotting factors such as factor VII, factor IX, tissue plasminogen activator and many more have entered the market. Most of the protein therapeutics currently on the market are recombinant products and, further, hundreds of them are in the clinical trials for therapy of cancers, immune disorders, infections, and other diseases (Dimitrov, 2012).

Briefly, in rDNA technology, through genetic engineering techniques by recruiting suitable restriction endonuclease enzymes, the gene of interest which codes for our protein of interest is isolated from the source DNA (e.g., human DNA) and is ligated to a suitable vector, generally a plasmid, in order to make recombinant/chimeric DNA molecule (Figure 1.6). Afterwards, by a suitable method (e.g., CaCl_2 or electroporation), the recombinant DNA molecule is introduced to a host which naturally does not possess the gene of interest and, then, the transformed host cells are selected properly in selective medium by the aid of a selectable marker (most often, an antibiotic gene) of the plasmid which is naturally available in the plasmid or has been devised in it (Figure 1.7). The selected colonies can be used as expressing platforms. In order to validate the recombination procedure the new plasmid should be isolated from the host and analyzed (Figure 1.8).

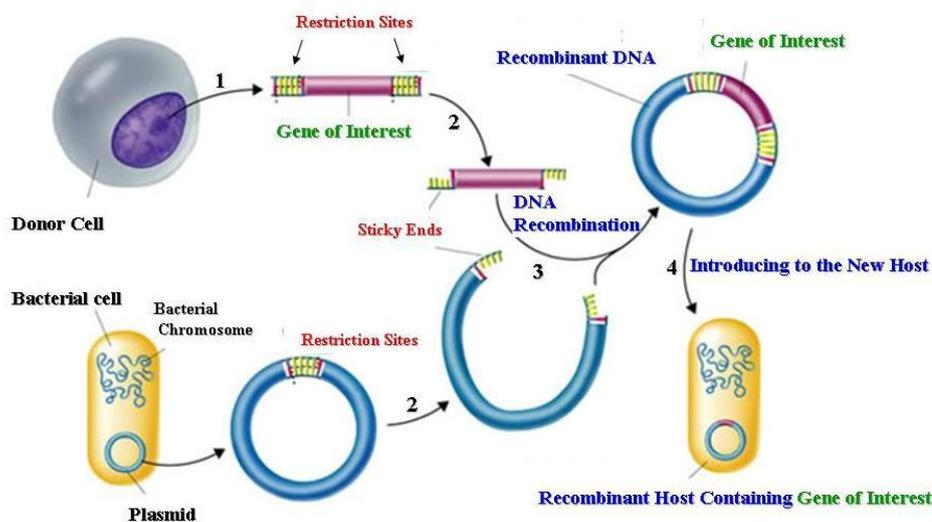


Figure 1.6. Illustration of the recombinant DNA technology; reproduced from (Web 9).

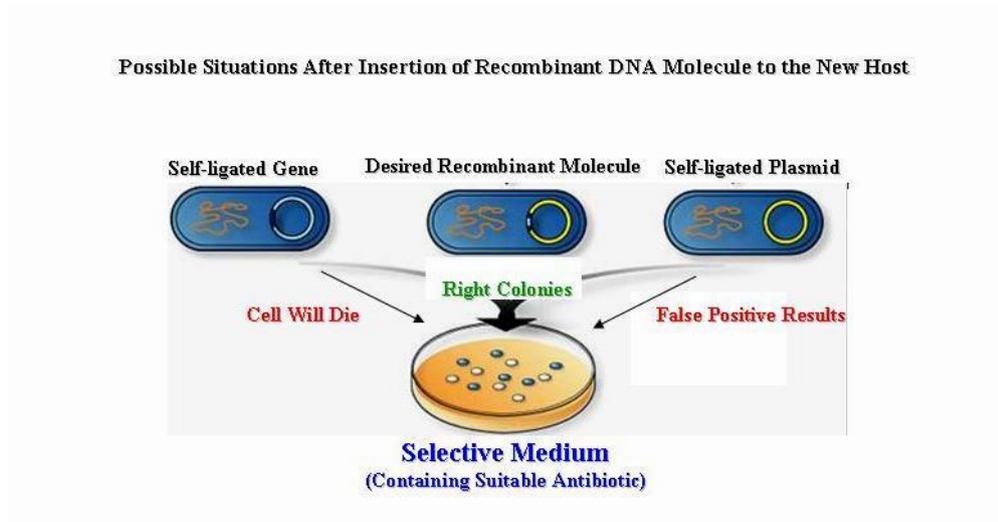


Figure 1.7. Transformation of the host cells with recombinant DNA (plasmid) and selection of the right (transformed) colonies.

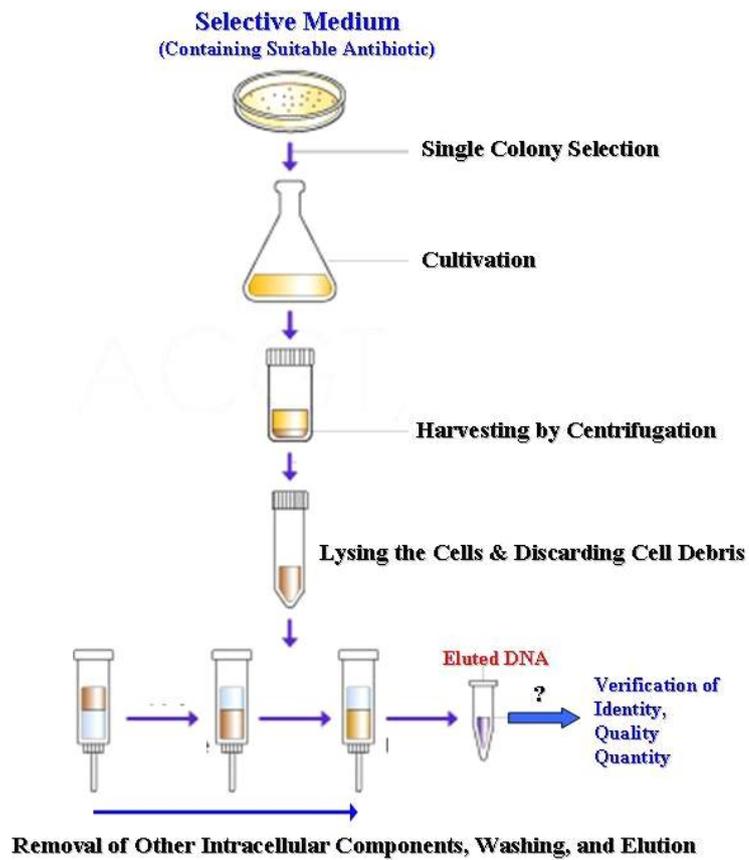


Figure 1.8. Recombinant plasmid isolation from the cloning host; reproduced from (Web 10).

On the other hand, if the primary host, generally a bacterial cell, has been used just for cloning purpose the new plasmid should be isolated and introduced to the second host, i.e., the major producing host. After transformation of the new host the selected (right) expressing colonies are used for production of desired recombinant product in defined bioprocess conditions.

During rDNA procedure special attention should be paid to the nucleotide sequence of the newly-designed plasmid and, thus, verifying the sequence of the new plasmid will be one of the validation steps after plasmid isolation (Figure 1.8) in order to prevent any undesired frame shift or mutation which can change the amino acid sequence of the transcript. An important issue in the recombinant therapeutics expression is the preservation of the structure integrity in the product. Any change in amino acid sequence or any alteration and substitution may lead to inefficiency, immune responses, or adverse effects.

The first usage of the exogenous hGH (or somatotropin), extracted from human pituitary glands, as therapeutic agent for growth hormone deficiency has been reported in 1962 (Raben, 1962) and until the mid-1980s extracting from human cadaver tissue was the only way of obtaining hGH (Staley et al., 2012). However, due to the realization of the transmission risk of the Creutzfeld-Jakob disease (CJD) from pituitary-derived hGH preparations, its supply was discontinued in 1985 (Houdebine, 2009; Cunha et al., 2011). In 1985, biosynthetic human growth hormone replaced pituitary-derived growth hormone (pit-HGH) for therapeutic use in the US and elsewhere. Biosynthetic human growth hormone, also referred to as recombinant growth hormone, is also called somatropin (British: somatrophin) and abbreviated as rhGH (Web 11). The first recombinant expression of the hGH was conducted in *E. coli* in 1979 (Goeddel et al., 1979). At present, rhGH is produced by bacterial platforms. However, the therapy is still very costly and annually may exceed \$ 15,000-20,000 (Cunha et al., 2011). The main reason is the presence of over-expressed protein as

aggregates (or IBs) which makes the subsequent costly denaturation and refolding inevitable. In addition, careful control of the quality to assure absence of endotoxins (from *E.coli*) along with the required sterile conditions for bacterium is the other causative agents of high cost (Cunha et al., 2011).

Regarding the numerous factors that take part in biological processes, in order to optimize any production process and make it more economical, parameters either in genetic level or process level can be taken into consideration. By recruitment of genetic engineering techniques and by considering systems biology approach rhGH production in the prominent host *Pichia pastoris* was followed in current research; the efficiency of the selected endogenous secretion signal peptides (obtained from secretome analysis) and oxygen limitation-induced promoters of *P. pastoris* (obtained from a transcriptome/proteome analysis) in production of rhGH was analysed by comparing to *Saccharomyces cerevisiae* α -mating factor and glyceraldehyde-3-phosphate dehydrogenase promoter, respectively. Finding novel secretion signal peptides and promoters as biological modules with parallel utilization of “omics” studies in order to enhance the production/secretion ability of an industrial microorganism is indeed important for industrial biotechnology.

CHAPTER 2

LITERATURE SURVEY

This chapter entails the major elements play role in expression of a recombinant protein (r-protein). After a glimpse to different available host organisms, selecting the right expressing platform by considering its unique capabilities and features, surveying the parameters that affect the expression of the desired protein that inevitably contain operational parameters in bioreactor, will be comprised. In this context, secretion signal peptides and promoters will be discussed in more detail. Derivation of the equations related with the fed-batch fermentation and metabolic flux analysis along with previous researches done related with human growth hormone (hGH) will complete literature survey.

2.1. Expressing hosts

In order to design an optimized recombinant expression system, a suitable host selection is the first step (Macauley-Patrick et al., 2005). The selection of the most appropriate expression platforms for a certain protein is usually a complex process and takes into account factors ranging from the intended use of the protein to the physiological consequences of its production in the host (Alvaro-Benito et al., 2013). No single expression system is the optimum one for all proteins (Böer et al., 2007) and the most efficient host should be determined for each foreign gene which is going to be expressed by taking into consideration the genetic and fermentation parameters (Nel et al., 2009; Çelik and Çalık, 2011). The product quality, bioactivity (functionality), productivity, and yield are the most critical parameters that

should be considered in the selection of the host (Demain and Vaishnav, 2009). In order to reach high yield and high quality in recombinant protein production, expressing host physiology and stress responses should be understood thoroughly which in turn will conduce to the biological and methodological obstacles posed by host (to the production of foreign protein) would be easily overcome (Ferrer-Miralles et al., 2009). However, it would be very logical and ideal to survey different host organisms in parallel and then choosing the most appropriate one (Nel et al., 2009).

The diverse population of (micro)-organisms and their different physiologies lead to a broad spectrum of choices for production of recombinant proteins (Corchero et al., 2012). Microorganisms (eukaryotic and prokaryotic) together with cultured cells of higher organisms (mammalian, insect, or plant) are mostly used as expression hosts for production of heterologous as well as homologous proteins (Porro et al., 2005). The production platform also may be transgenic animals (Demain and Vaishnav, 2009). Recent advances in genomics, recently- emerged “omics” techniques and integrative approaches (system biology) have led to the development of novel microbial cell factories (Martinez et al. 2012). Figure 2.1 represents the share of mammalian and non-mammalian hosts in production of approved biopharmaceuticals in selected time intervals. Furthermore, the individual share of the recruited host organisms in production of biopharmaceuticals has been depicted based on two separate time intervals: 2010-2014 and in total from 1982 to 2014 (Figure 2.2).

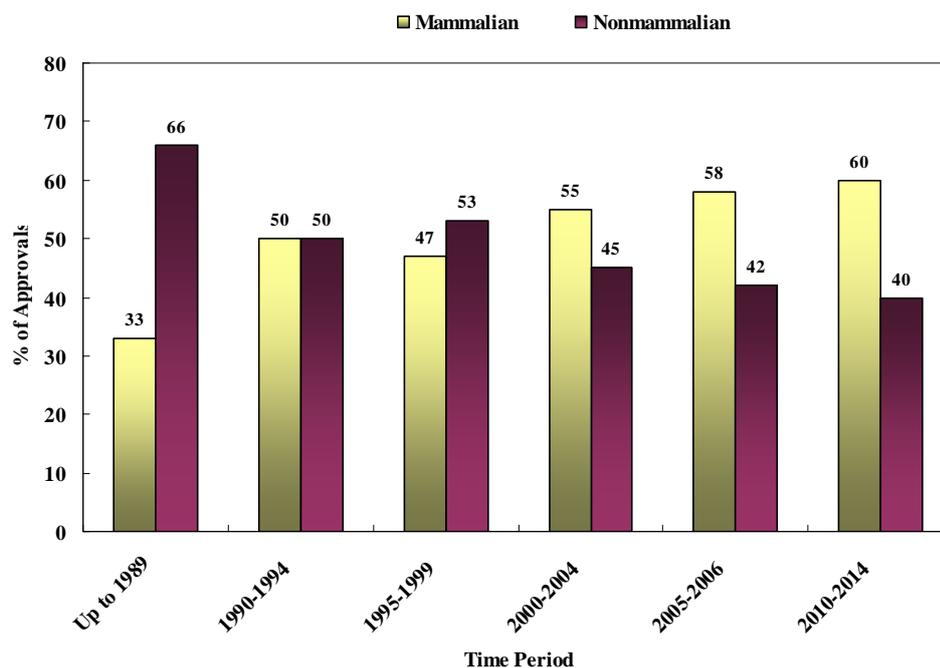


Figure 2.1. Relative application of mammalian versus non-mammalian-based expression systems in the production of approved biopharmaceuticals (Walsh, 2014).

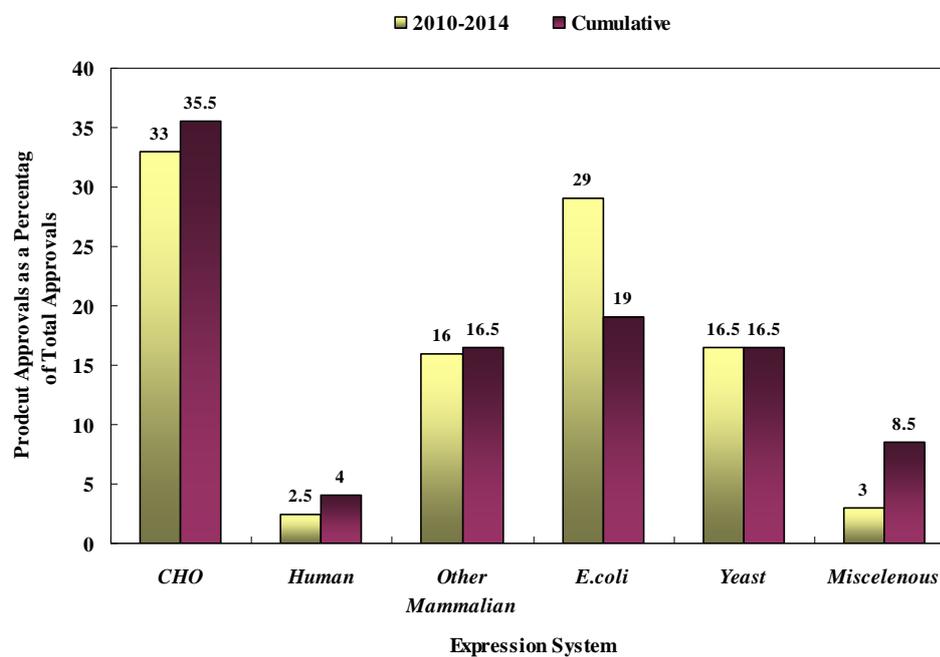


Figure 2.2. Biopharmaceutical products approvals, cumulative (1982-2014) and for the period of the study (2010 – July 2014) (Walsh, 2014).

2.1.1. Prokaryotes

In this part, especially bacterial cells are regarded as the major member. Gram- negative *E.coli*, Gram-positive *Bacillus* specific species, and other selected bacteria such as *Ralstonia eutropha* and *Staphylococcus carnosus* (Demain and Vaishnav, 2009) can be mentioned in this category.

E. coli is the most widely used microorganism among prokaryotes (Demain and Vaishnav, 2009; Schmidt, 2004) and the first choice for primary cloning and small-scale production in research purposes (Ferrer-Miralles et al., 2009). It is the workhorse of the biopharmaceutical industry by considering this point that the production of 17 out of 58 approved pharmaceutical products (from 2006 to June 2010) is carried out by its application (Corchero et al., 2012). *E. coli* B and K12 strains are the main platforms for expressing non-glycosylated peptides and proteins for industrial and academic aims (Graumann and Premstaller, 2006). Its genetics is the best understood one than any other microorganism and, thus, genetic manipulation can be easily conducted (Demain and Vaishnav, 2009; Graumann and Premstaller, 2006). Because of fast and easy growth on inexpensive medium (Demain and Vaishnav, 2009; Corchero et al., 2012) high cell densities can be achieved. Furthermore, accumulation of the expressed foreign proteins up to 80% of the dry weight makes them tolerant of a wide range of environmental conditions (Demain and Vaishnav, 2009). High cell density fermentations can be achieved by recruitment of strong inducible promoters system such as T7 promoter which are commonly used (Rezaei and Zarkesh-Esfahani, 2012). The level of production has been reported to be under 10 g/L in different cases even lower than 1 g/L and dry cell content can vary between 20 to 175 g/L. Although the production can be high but because of production of inclusion bodies (IBs) and low yields of renaturation, the recovery yield could be very low (Demain and Vaishnav, 2009). Lack of the required machinery for post-translational modifications (PTMs) which are the prevalent feature of most eukaryotic proteins, has made this system the cheapest, easiest, and quickest way for expression of

structurally simple proteins (Corchero et al., 2012). Expressing very large proteins, S-S rich proteins and proteins requiring PTMs will be problematic in this host (Demain and Vaishnav, 2009; Daly and Hearn, 2005). It should be emphasized that PTMs such as S-S bond formation, phosphorylation, and proteolytic cleavage may be a vital step to obtain a functional protein and a PTM-lacking version of protein may be inactive, unstable, or insoluble (Ferrer-Miralles et al., 2009). Acetate formation in high cell densities and its toxicity, and formation of IBs are the major drawbacks of working with *E. coli* (Demain and Vaishnav, 2009; Corchero et al., 2012, Ferrer-Miralles et al., 2009). The main reason of formation of IBs is the limited solubility and shortage of suitable chaperone and foldases system in the *E. coli* (Graumann and Premstaller, 2006; Schmidt, 2004). The presence as IBs or being soluble in cytoplasm depends mainly on protein intrinsic characteristics of sequence, promoter strength, and fermentation parameters such as temperature (Graumann and Premstaller, 2006). Inability in efficient secretion of the protein as the result of lack of general secretory system is the other important shortcoming of *E. coli* (Corchero et al., 2012). Directing to the periplasmic or extracellular medium can fade away the N-terminal methionylation problem and, furthermore, secretion to the periplasmic space can give correctly-folded proteins which can be enhanced by co-expression of chaperones. However, low level of chaperones and foldases, membrane structure, and high periplasmic protease leads to low product yield in *E. coli* secretion system (Schmidt, 2004). Proteolytic cleavage can also be another problem which can be conquered by mutations for deletion of several proteases (Graumann and Premstaller, 2006). Finally, it is worth to mention that both inducible and constitutive production systems have been utilized in *E. coli* because of possible toxicity of eukaryotic proteins in bacteria. (Demain and Vaishnav, 2009; Graumann and Premstaller, 2006). Another important issue in the case of *E. coli* as a prokaryote host is the different codon usage bias from human which is related to the specific tRNA pools (Demain and Vaishnav, 2009).

Different species of *Bacillus* lack outer membrane which is present in Gram-negative bacteria and contain the lipopolysaccharides (LPS) (Demain and Vaishnav, 2009) and, in addition, own inherently a powerful secretion capacity (Schmidt, 2004). In this system, effective secretion of recombinant proteins to medium can be achieved without production of harmful endotoxins and exotoxins; so, they are very efficient and cost-friendly regarding the product recovery procedure. *Bacillus* protein yields can reach 3 g/L but the major problem in this host is the production of many proteases which can cleave the protein (subtilisin and natural protease are responsible for 96-98% activity). However, strains of *Bacillus subtilis* have been developed which lack eight extracellular proteases (Demain and Vaishnav, 2009). *Bacillus brevis* can be considered because of its lower protease activity and also production of proteinase inhibitor. *Bacillus megaterium* and *Bacillus licheniformis* are also industrially important species (Demain and Vaishnav, 2009) where *B. megaterium* lacks membrane proteases (Schmidt, 2004).

Streptomyces (ex., *Str. lividans* and *Str. tendae*) are considered to be potentially amenable for introducing new secretion systems (Schmidt, 2004). But, the yields should be in cost-effective ranges in order to be employed in industry. Lactic acid bacteria considered as GRAS (Generally Recognized As Safe) can also be promising cell factories (Corchero et al., 2012). Cold-adapted (psychrophilic) bacteria (such as *Pseudomonas haloplanktis*) as a newly-emerging host (cold cell factories) should be paid more attention. They show improved protein folding, since the hydrophobic effect which underpins the IBs formation has relation with entropic driving forces that are reduced at low temperatures (Corchero et al., 2012).

2.1.2. Eukaryotes

Eukaryotic (micro) organisms have endomembrane structures that facilitate the compartmentalization of different groups of biochemical reactions and, thus, make pre-requisites such as PTMs feasible and, therefore, the foreign

product more genuine and functional. The members in this category are fungi (single cell and filamentous), mammalian cells, insect cells, and transgenic animals and plants. Because of the focus of the current research on r-protein production by *P. pastoris* as a genus of yeast group, all non-yeast eukaryote hosts are briefly discussed in one group and yeast group is discussed separately.

2.1.2.1. Non-yeast expression platforms

2.1.2.1.1. Mammalian cell lines

The industrial production of proteinaceous drugs is based on animal cell lines because of performed PTMs and product quality (Kunert et al., 2009). Simultaneous with early stages of biopharmaceutical efforts, onset of utilization of mammalian cell lines in rDNA technology was in 1980s as a result of failure in production of erythropoietin and tissue plasminogen activator in *E. coli* both of which required glycosylation (Demain and Vaishnav, 2009). As many therapeutic proteins are glycosylated or structurally complex (such as erythropoietin) with diverse modifications which are often essential for biological activity, their production requires the most appropriate system which can render mammalian type PTMs and this need has been the driving force for using mammalian cells (Gellissen et al., 1992; Corchero et al., 2012; Tanapongpipat et al., 2012). The consistent complete spectrum of modifications will lead to less structural discrepancy and, thus, less immunogenic properties (Schmidt, 2004) and, further, crossing out the necessity of renaturation because of correct folding (Demain and Vaishnav, 2009). Nowadays, nearly 60-70% of recombinant therapeutic proteins are expressed in mammalian cells (Rezaei et al., 2013).

Chinese hamster ovary (CHO) cells are the pioneers in this category and can be considered as a standard in protein production due to the rapid growth, process versatility, feasibility of culture in suspension or attached mode (Schmidt, 2004; Corchero et al., 2012), high production, and high

stability of foreign gene expression (Rezaei et al., 2013). The other cell types can be mentioned are: (i) baby hamster kidney cells (BHK), mentioning that most of the therapeutic proteins have been expressed in transgenic hamster cell lines (Ferrer-Miralles et al., 2009), (ii) various mouse myelomas such as NSO murine myeloma cells, used in production of monoclonal antibodies (iii) green monkey kidney cells (iv) human cell line such as human embryonic kidney (HEK) cells and human-retina-derived (PERC6) cells (Butler, 2005; Corchero et al., 2012). This kind of hosts needs serum for proliferation which increases the risk of contamination. Furthermore, since composition of serum is variable and undefined, it results in inconsistent growth and productivity (Butler, 2005). Passage of probable prions in bovine serum albumin and human serum albumin to the host can be overcome by usage of protein-free, and chemically defined media which enhances the biosafety of end product (Ferrer-Miralles et al., 2009). Major disadvantage in these hosts is the low level, a few kilograms per year of production (Houdebine, 2009) which is lower than productivity of prokaryotic hosts (Rezaei et al., 2013) furthermore, instable productivity due to the instability of the selected clones. The extended cultivation, prolonged gene expression, is possible by integrating the gene of interest in to the host genome (stable cell line) (Corchero et al., 2012). The other shortcomings of these platforms can be summarized as being shear sensitive which necessitates design of impellers for reduction of shear forces (Butler, 2005) arose from agitation and aeration (Keane et al., 2003), unsatisfactory secretion, high cost of production because of expensive materials specifically used in media (Corchero et al., 2012; Tanapongpipat et al., 2012), and prone to viral contamination (Demain and Vaishnav, 2009; Tanapongpipat et al., 2012). By improvement in medium design, the production has increased from 50 mg/L (in 1986) to 4.7 g/L (in 2004). In a new system with human cell lines protein titer of 26 g/L for monoclonal antibody was reported (Demain and Vaishnav, 2009).

2.1.2.1.2. Transgenic organisms

Due to the (potential) high production capacity in comparison with the mammalian cell culture-based expression platforms (Markaki et al., 2007), transgenic animals and plants would be regarded as suitable candidates for r-protein production.

2.1.2.1.2.1. Transgenic animals

This expression system is based upon using live animals. Generation of transgenic animals is hitherto relatively laborious and costly but is no longer an obstacle for recombinant protein production (Houdebine, 2009). Transgenic animals system is based upon using live animals and is employed to obtain r-protein in blood, urine, milk, egg white, seminal plasma, and silk worm cocoons. Wide variety of proteins have been produced by transgenic animals including, monoclonal antibodies, blood factors, vaccines, hormones, growth factors, cytokines, enzymes, milk proteins, collagen, fibrinogen, and other products (Houdebine, 2009) and in most cases the protein activity is the same as the native one (Demain and Vaishnav, 2009). Ruminants are potentially the most appropriate system to produce high amounts of protein but their reproduction is slow and they do not glycosylate the proteins and as well as pigs are sensitive to prion diseases (rabbits are insensitive to prion diseases) (Houdebine, 2009). Production in milk is more cost effective than production in mammalian cell cultures (Demain and Vaishnav, 2009) and r-proteins are stable in milk (Houdebine, 2009). At this time, production in milk is the most mature system; a level of 1 mg/mL or even lower is considered acceptable economically. At higher concentrations, the glycosylation is not performed fully perhaps due to saturation of mammary cellular machinery. The transfer of glycosylation enzymes to the transgenic host can be a solution for this imperfection. In addition, preference of milk to blood can be due to this fact that active r-proteins in blood may harm the animal and milk avoids this problem (Houdebine, 2009). One of the disadvantages of this platform is the time interval required for to

assess production level, 3.5 months in mice, 15 months in pigs, 28 months in sheep, and 32 months in cows, (Markaki et al., 2007; Demain and Vaishnav, 2009); difficult technical procedure faces another obstacle (Sanchez et al, 2004). The other imperfection may be the separation of human proteins from their animal counterparts (Houdebine, 2009). By taking safety concerns into consideration, transmission of infectious diseases (viral and prion infections) or adverse allergenic effects and immunogenic responses all are of prime importance in these hosts (Markaki et al., 2007; Ferrer-Miralles et al., 2009; Houdebine, 2009). In addition, some of the heterologous proteins may be active and deleterious for transgenic animals (Houdebine, 2009). Until 2009, ATryn was the only approved biopharmaceutical expressed in transgenic animals (milk of transgenic goat). However, if in coming years, pharmacovigilance (PV, or PhV also known as drug safety) after treatment of the patients did not revealed any adverse effect, it is expected that we encounter an increase in amount of transgenic animal-produced heterologous proteins (Ferrer-Miralles et al., 2009). The production of drugs in transgenic animals, by stopping the production of a lung drug in sheep for Bayer AG, has been stalled by the demise of PPL therapeutics (Demain and Vaishnav, 2009).

2.1.2.1.2.2. Transgenic plants

Plants can be considered as viable and competitive (non-microbial) alternatives for production of complex human r-proteins (Jez et al., 2013; Liu et al., 2013) as they are the least expensive biomass (Cunha et al., 2011). The first r-protein expressed in transgenic plants was hGH produced in transgenic tobacco (Ferrer-Miralles et al., 2009). Nowadays, a few enzymes used in researches or diagnosis are produced by transgenic plants at industrial scale (Houdebine, 2009). Transgenic plant cell systems can provide some advantages such as reduced risk of contamination by viruses, oncogens, and pathogens of mammalian cells or bacterial toxins, easy purification due to relatively simpler and cheaper downstream processing, lower cost of

required media for cultivation, ability to assemble multi-subunit proteins (such as antibodies) as efficiently as animal cells, ability to express functional proteins with the aid of eukaryotic PTMs, and retaining the native structure of the protein (Leite et al., 2000; Kim et al., 2008; Demain and Vaishnav, 2009; Ferrer-Miralles et al., 2009; Houdebine, 2009; Liu et al., 2013). It is possible to store the expressed r-protein in leaves, seeds or both regarding the type of promoter used (Houdebine, 2009). Disadvantages of the plant system are i) possible contamination by pesticides, herbicides, and toxic metabolites of plants (Demain and Vaishnav, 2009), ii) low expression level with plant systems which can be regarded as an obstacle in widespread use (Kim et al., 2008; Cunha et al., 2011), iii) stability of the proteins inside the plant which may be alleviated by targeting the protein to the specific organelles within the plant cell (Liu et al., 2013; Cunha et al., 2011), iv) as the main disadvantage, plant-specific PTMs that lead to expression of immunogenic products and, thus, adverse immune responses (Ferrer-Miralles et al., 2009; Yusibov and Mamedov, 2010) which can be regarded as an important bottleneck in production of humanized proteins, v) Post-transcriptional silencing of introduced transgenes (Markaki et al., 2007), vi) gene flow from transgenic plant to the wild type (Yusibov and Mamedov, 2010). Commercial production of proteinaceous drugs in plants was halted by closing down the PPL Therapeutics and exit of Monsanto Corporation from this area (Demain and Vaishnav, 2009).

2.1.2.1.2.3. Insect cell lines

Because of the need to milder culture conditions than mammalian cell lines such as lower temperatures and no CO₂ addition to the gas phase, insect cells have been increasingly used for the production of wide variety of proteins in recent years (Moraes et al., 2012). *Drosophila Schneider* cells along with the *baculovirus*-infected cells are the common host insect cells (Fernandez and Vega, 2013). The most widely used vector in this system is *baculovirus* along with the most usual host the fall armyworm (*Spodoptera frugiperda*,

Sf9), i.e., *baculovirus*-Sf9 insect cells, in suspension culture (Demain and Vaishnav, 2009; Houdebine, 2009). *Trichoplusia ni* (*Cabbage Looper*) cell lines also are used to produce heterologous proteins upon recombinant *baculovirus* infection (Moraes et al., 2012). Highly restricted host range of the *baculovirus* makes the *baculovirus*-insect cell system very safe (Yamaji et al., 2012). Reported amount of produced protein in this kind of host has increased from 600 mg/L in 1988 to 11 g/L in 2007 (Demain and Vaishnav, 2009). The advantages of the *baculovirus*-assisted insect cell expression can be: i) proper eukaryotic PTMs which is more complicated than fungi, ii) proper protein folding as a result of possessing the best machinery for folding, iii) high expression level due to presence of strong polyhedron protein promoter, iv) easy scale up because insect cells growth is not anchorage dependent, v) vertebrates and plants cannot be attacked by expression vectors, vi) no limitation on protein size, vii) simultaneous expression of multiple genes (Demain and Vaishnav, 2009), and viii) provision of safety advantages for the production of biological agents (including vaccines) because they do not aid the proliferation of mammalian viruses or mycoplasmas (Yamaji et al., 2012). In contrast, as some disadvantages, incorrect glycosylation (Houdebine, 2009) because of very simple and non-sialated N-linked glycosylation (Ferrer-Miralles et al., 2009), insufficient secretion of protein, presence of intracellular protein aggregates, high price of cultivation medium, low expression level because of slow growth (Demain and Vaishnav, 2009). In addition, since the viral infection is lethal for insect cells, in each batch of the cultivation fresh cells are required (Ferrer-Miralles et al., 2009). The *baculovirus*-insect cell system is employed for the industrial-scale production of a human papilloma virus vaccine, Cervarix, which has been approved for use against cervical cancer (Moraes et al., 2012; Yamaji et al., 2012). r-protein expression systems using of *D. melanogaster* cells are being increasingly employed lately. To date, about 100 cell lines derived from *D. melanogaster* have been obtained.

Amongst, 12 are easily cultivated; however, the only cell lines used for heterologous gene expression are Schneider's 2 and 3 (Moraes et al., 2012).

2.1.2.1.3. Filamentous fungi (molds)

As filamentous fungi are more organized than yeasts, their complex PTM machinery can resemble that of mammals. This category, such as *Aspergillus niger*, are very attractive choices as they have a high secretion capacity and also can perform glycosylation (Demain and Vaishnav, 2009). *A. awamori*, *A. oryzae*, *A. sydowii*, *Acremonium chrysogenum*, *Trichoderma species* (such as *T. reesei*), *Fusarium species*, *Penicillium Chrysogenum*, *Chrysosporium lucknowense* are the further examples which have been used for production of recombinant homologous (e.g., *A. niger* glucoamylase in *A. awamori*) or heterologous (e.g., human lactoferrin by *A. awamori*) proteins. Homologous proteins can be produced in the range of 30-40 g/L. However, heterologous proteins have the highest yield of 4-5 g/L (Schmidt, 2004). As a disadvantage, in these hosts, low copy number transformation can be expected as a result of foreign DNA degradation and low transformation rate (Schmidt, 2004). In contrast, since this kind of host can utilize cellulose, the fermentation cost can be diminished strikingly. Cultivation of these microorganisms lead to a viscous medium and one of the attempts to optimize the process can be on lowering the viscosity; use of strong homologous promoters, using protease-deficient strains, since the proteins are cleaved severely by endogenous fungal proteases (Corchero et al., 2012), screening for best producing strains by random mutagenesis, and fusion with an efficiently secreting protein can be other methods to improve the expression.

2.1.2.2. Yeast expression platforms

Table 2.1 has summarized features of available platforms for r-protein production.

Table 2.1. A comparison of pharmaceutical expression systems. Data updated in 2005 (Web 12).

Host	Advantages	Disadvantages	Applications	Cost / gram
Bacteria	Established regulatory track; well-understood genetics; cheap and easy to grow	Proteins not usually secreted; contain endotoxins; no posttranslational modifications	Insulin (<i>E. coli</i> ; Eli Lilly); growth hormone (Genentech); growth factor; interferon	-
Yeasts	Recognized as "safe;" long history of use; fast; inexpensive; posttranslational modifications	Overglycosylation can ruin bioactivity; safety; potency; clearance; contains immunogens/antigens	Beer fermentation; recombinant vaccines; hepatitis B viral vaccine; human insulin	\$50-100
Insect cells	Posttranslational modifications; properly folded proteins; fairly high expression levels	Minimal regulatory track; slow growth; expensive media; baculovirus infection (extra step); mammalian virus can infect cells	Relatively new medium; Novavax produces virus-like particles	-
Mammalian cells	Usually fold proteins properly; correct posttranslational modifications; good regulatory track record; only choice for largest proteins	Expensive media; slow growth; may contain allergens/contaminants; complicated purification	Tissue plasminogen activator; factor VIII (glycoprotein); monoclonal antibodies (Herceptin)	\$500-5,000
Transgenic animals	Complex protein processing; very high expression levels; easy scale up; low-cost production	Little regulatory experience; potential for viral contamination; long time scales; isolation/GMPs on the farm	Lipase (sheep, rabbits; PPL Therapeutics); growth hormone (goats; Genzyme); factor VIII (cattle)	\$20-50
Transgenic plants	Shorter development cycles; easy seed storage/scaling; good expression levels; no plant viruses known to infect humans	Potential for new contaminants (soil fungi, bacteria, pesticides); posttranslational modifications; contains possible allergens	Cholera vaccine (tobacco; Chlorogen, Inc.); gastric lipase (corn; Meristem); hepatitis B (potatoes; Boyce Thompson)	\$10-20

The innovative findings in the field of metabolic engineering and amelioration of process procedures have a supreme influence on the ability of yeasts to express r-proteins.

The production of foreign protein is carried out in yeast when the product is produced as aggregates (not soluble form) or requires a specific PTM that cannot be achieved artificially (Ferrer-Miralles et al., 2009). In general, yeasts lead to higher yields and better stability in comparison with mammalian cells (Corchero et al., 2012). Among the eukaryotic platforms, yeasts mingle the superiorities of single-cell organisms, i.e., fast growth and simple genetic manipulation, with protein processing capabilities of eukaryotic organisms, i.e., PTMs, subunit assembly, and protein folding (Porro et al., 2005; Çelik and Çalık, 2011). Furthermore, yeasts conform to safety aspects since they do not convey any pathogens, viral inclusions, or pyrogens because of absence of endotoxins and oncogenic and viral DNA (Porro et al., 2005; Böer et al., 2007). Absent and aberrant (N-linked) glycosylations, are the main problem plaguing the proper expression of recombinant glycoproteins in yeasts (Zhang et al., 2012); the yeast Golgi apparatus does not perform the trimming process of N-glycans such as mammals Golgi which leads to long mannose chains referred to as hyperglycosylation or hypermannosylation of core oligosaccharides. Furthermore, through the ER and Golgi, the added glycans may be located in the places other than the ones in the native protein and their composition may be substantially different in terms of mannose residues (Graumann and Premstaller, 2006). High-mannose side chain leads to affect the serum half-life and immunogenicity of end product (Ferrer-Miralles et al., 2009). Efforts to develop strains with humanized sugar content as what has been carried out for *P. pastoris* and *H. polymorpha* will pave the way for production of recombinant glycoproteins in yeasts. However, to date all the commercialized therapeutics produced in *S. cerevisiae* have been non-glycosylated (Corchero et al., 2012). Another problem is the difference between codon usage bias of the original gene donor and the expression host

which has led to poor expressions. However, artificially synthesized genes together with available codon usage tables have provided us with custom designed and host codon-optimized genes (Nel et al., 2009).

There are two general classifications of yeasts employed as expression hosts. In one of them yeasts are divided into: conventional and non-conventional; in the other one they are regarded as Crabtree positive (can produce ethanol under aerobic conditions) and Crabtree negative. *S.cerevisiae* is the only conventional host and together with *Zygosaccharomyces rouxii* and *Zygosaccharomyces bailii* are the Crabtree positive choices. Based on the recent insights yeasts can be divided into two other categories: methylotrophic and non-methylotrophic yeasts (Table 2.2). Non-methylotrophic yeasts are more familiar to specialists of different disciplines and also broad range of carbon and energy sources can be consumed in industrial processes by this group. However, large-scale production leads to some limitations and confines the scale-up of these hosts processes; therefore, most of the processes for r-protein production by these yeasts are in their early stages (Porro et al., 2005).

Table 2.2. Non-methylotrophic and methylotrophic yeast species recruited for r-protein production (Porro et al., 2005).

Nonmethylotrophic	Methylotrophic
<i>S. cerevisiae</i>	<i>H. polymorpha</i>
<i>K. lactis</i>	<i>P. pastoris</i>
<i>Y. lipolytica</i>	<i>P. methanolica</i>
<i>Z. rouxii</i>	<i>C. boidinii</i>
<i>Z. bailii</i>	
<i>S. occidentalis</i>	
<i>P. stipitis</i>	

Generally, yeasts are used by the help of *E.coli* since most of available vectors for yeast platforms have been designed as *E.coli* shuttle vectors (Graumann and Premstaller, 2006) to facilitate amplification of the plasmid. We proceed with introducing non-methylotrophic and methylotrophic yeasts separately.

2.1.2.2.1. *Saccharomyces cerevisiae*

S. cerevisiae has been a workhorse for both biochemists and industrial biotechnologists due to the broad knowledge about its physiology and genetic which conduces to being the best-characterized eukaryotic system (Nel et al., 2009); *S. cerevisiae* was the first eukaryotic organism sequenced (Corchero et al., 2012) and the initial yeast used for r-protein production (Böer et al., 2007). It has been in use as an expression platform since 1981 and is the most widely used yeast for the production of therapeutic proteins (Porro et al., 2005; Corchero et al., 2012). This yeast has been recognized as GRAS organism by FDA which has led to its popular application in food industry in the early 1990s (Porro et al., 2005; Böer et al., 2007). It has been reported recently that this yeast can be useful as a probiotic in humans for oral delivery of therapeutic proteins (Çelik and Çalık, 2011). Various products have been expressed in this yeast; insulin (commercially available since 1987), hepatitis B vaccines, anticoagulant hirudin, and *Aspergillus*-derived glucose oxidase, as a technical enzyme, are a few examples (Böer et al., 2007).

In addition to a promoter of a highly expressed gene of *S. cerevisiae*, for extracellular production of the r-protein (Çelik and Çalık, 2011), a secretion signal sequence (SP) should be recruited. Most widely used signal sequence is *S. cerevisiae* α -mating factor prepro sequence (α -MF) which is functional in all yeast platforms employed; but other choices such as pre-pro sequence of one derived from *S. occidentalis* *GAM1* gene can be mentioned. Secretion can also be facilitated by the co-expression of a secretory pathway gene.

Reduced biomass yield as a result of aerobic alcohol fermentation necessitates semi-batch fermentation to reach high cell densities which, in turn, leads to technical fermentation requirements and sophisticated equipments for this yeast processes (Romanos et al., 1992; Nel et al., 2009), hyperglycosylation of the protein which leads to activity reduction and decrease of receptor binding (Böer et al., 2007; Nel et al., 2009; Demain and Vaishnav, 2009), termination of sugar chains with α -1,3-mannose bond which acts as an allergen (Böer et al., 2007), retention of the product in periplasmic space which leads to partial degradation of the product (Porro et al., 2005), the limited capability of *S. cerevisiae* to secrete proteins (Sasagawa et al., 2011), strain instabilities because of common usage of episomal vectors which lead to batch inconsistencies (Böer et al., 2007; Nel et al., 2009), and the O-linked oligosaccharides just being comprised of mannose whereas the higher eukaryotic proteins have sialylated O-linked chains (Demain and Vaishnav, 2009) are all among the reasons that why *S. cerevisiae* is not the most suitable choice for large-scale production of r-proteins. However, by some modifications via invented strategies these limitations can be overcome (Nel et al., 2009). Nevertheless, the hyperglycosylation is still the major obstacle; the proportion of available therapeutic glycosylated proteins to non-glycosylated ones is 60 to 40 and, thus, this yeast (along with *E. coli*) can be used for expression of the 40% non-glycosylated (Nel et al., 2009).

Based on the described drawbacks of *S. cerevisiae* and regarding the great biodiversity among yeasts that 800 different yeast species with varying metabolic characteristics has been identified (Graumann and Permstaller, 2006), efforts have been focused on development of new host systems so-called "non-conventional" yeasts (Porro et al., 2005). Other yeasts as substitutes for *S. cerevisiae* have been used as sources of biopharmaceuticals due to the promoter strength, secretion efficiency, and ease of growth to high cell densities (Romanos et al., 1992): methylotrophic yeasts especially *P. pastoris*, *H. polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces*

lactis, *Candida maltosa*, two dimorphic yeasts *Yarrowia lipolytica* and *Arxula adenivorans* (Gellissen et al., 1992; Porro et al., 2005; Graumann and Premstaller, 2006). Clearly, no single yeast cell can provide us with all the required qualifications for foreign protein production (Çelik and Çalık, 2011). However, the main advantages of the alternative yeasts (in comparison to *S. cerevisiae*) are strong promoters, increased gene copy number, proper folding and secretory pathway (Corchero et al., 2012) which can offer a broader range of choices upon decision.

2.1.2.2.2. *Schizosaccharomyces pombe*

S. pombe aside from *S. cerevisiae* is the most thoroughly studied yeast species (Romanos et al., 1992). This host compared to *S. cerevisiae* has glycosylation and other PTMs more similar to mammalian cells (Graumann and Premstaller, 2006) and has very advanced genetics. The quality control system for protein folding in this yeast is closer to mammalian cells than *S. cerevisiae* (Çelik and Çalık, 2011); but as a result of less developed fermentation technology and relatively few inducible promoters (Çelik and Çalık, 2011), utilization of this yeast has been confined to the isolation and studying homologous mammalian genes rather than expression of foreign proteins. Promoters from *S. cerevisiae* genes generally exhibit less functionality and lead to inefficient transcription initiation in this host. Both constitutive and regulated expression has been reported for this host system (Romanos et al., 1992).

2.1.2.2.3. *Zygosaccharomyces*

The knowledge about *Zygosaccharomyces* is very poor but its six classified species are evolutionary very close to *S. cerevisiae* and are not so far from *K. lactis*. Some of the species have exceptional resistance to several stresses that confers them the suitability for industrial usage; *Z. rouxii* is salt tolerant (osmophilic) and *Z. bailii* can bear high sugar concentrations (osmotolerant), acidic conditions, and relatively high growth temperatures. Ethanol

production in these yeasts under aerobic conditions is lower compared to *S. cerevisiae*. Finally, there is no need to intricate fermentation strategies to obtain high cell densities (Porro et al., 2005).

2.1.2.2.4. *Kluyveromyces lactis*

K. lactis has been one of the main organisms grown in industry in bioreactors to produce bovine chymosin, known as rennin, on a large scale since 1950s. Its genome has been sequenced in 2004 and also has the GRAS status (Porro et al., 2005). Such as methylotrophs, the ability of metabolism adaptation in order to efficiently utilize a special carbon source is observable in *K. lactis* (Gellisen et al., 1992). This yeast as an industrial microbe has high secretion ability (Romanos et al., 1992) and can use a wide range of carbon sources such as galactose, raffinose, and maltose (Böer et al., 2007) but important discriminative characteristic is its ability to grow on lactose as the only carbon source where during growth in the lactose-containing medium, the lactose utilization pathway enzymes are strongly induced (Böer et al., 2007; Gellisen et al., 1992). The most widely used promoters of *K. lactis* are the *KIADH4* and inducible *LAC4* promoters (Nel et al., 2009) where the best recognized one is that of the *LAC4* gene (Romanos et al., 1992). It should be considered that several *S. cerevisiae* promoters are also active in *K. lactis* and have been used for production of proteins such as mouse α -amylase, D-amino acid oxidase, etc. which promoters of *PHO5*, *GALI*, *PGK*, and etc. can be some examples (Böer et al., 2007). Both genomic integration and episomal state have been reported for the related vectors. *K. lactis* has the advantage of possessing highly stable episomal vectors (Romanos et al., 1992).

2.1.2.2.5. *Arxula adeninivorans*

A. adeninivorans (*Blastobotrys adeninivorans*), is a non-pathogenic and also a dimorphic yeast species (Böer et al., 2007; Nel et al., 2009). The *Arxula* genus has two species: *A. terrestre* and *A. adeninivorans*. *A. terrestre* in

contrast to *A. adenivorans* does not have any fermentative ability (Nel et al., 2009). All *A. adenivorans* strains share unusual biochemical activities being able to assimilate a range of amines, adenine and several other purine compounds as sole energy and carbon source (Böer et al., 2007; Nel et al., 2009). Interesting features of this host are being osmo-tolerant and thermo-tolerant and temperature-dependent dimorph (budding and mycelial forms). Osmo-tolerance has been emanated in *AHSB4* gene and its promoter (Çelik and Çalık, 2011), is dependent on the presence and absence of NaCl, can be an attractive element for gene expression (Nel et al., 2009). Regarding the biotechnological importance of the morphology, O-glycosylation is performed in budding cells only; however, N-glycosylation occurs in both cell types (Nel et al., 2009). The utilized vectors in this host can be either episomal or integrative. The production of *E. coli*-derived β -galactosidase, green fluorescent protein (GFP), human serum albumin (HAS), and invertase enzyme with the aid of recombinant strains of *A. adenivorans* has been reported (Böer et al., 2007).

2.1.2.2.6. *Yarrowia lipolytica*

Y. lipolytica (*Candida lipolytica*) is dimorphic and is unicellular in minimal media containing glucose or n-hexadecane whereas forms mycelia in minimal media including olive oil or casein. A mixture of both forms is available in complex medium (Romanos et al., 1992). *Y. lipolytica* in contrast to *A. adenivorans*, in order to shift its morphology, needs additional factor than temperature such as change in pH or addition of serum to the medium (Nel et al., 2009). This yeast metabolizes only a few sugars (mainly glucose but not sucrose), alcohols, acetate, fatty acids, oils and unusual carbon sources, such as hydrocarbons. *Y. lipolytica* is a strictly aerobic microorganism (Böer et al., 2007) and with its intense secretion capability, specifically secretion of high molecular weight proteins, can be considered as a suitable host for secretion mode in the role of an industrial yeast such as *K. lactis* (Romanos et al., 1992; Nel et al., 2009). The high-

level secretion ability along with the capability to grow to high cell densities at industrial scale is a motivation for usage as host in heterologous protein production (Romanos et al., 1992). Furthermore, *Y. lipolytica* has the superiority of absence of hyperglycosylation over *S.cerevisiae* (Gellissen et al., 1992). *Y. lipolytica* strains do not ferment sugars and are considered as non-pathogenic species (Nel et al., 2009). Transformation of *Y. lipolytica* by both replicative (episomal) and integrative plasmids is feasible (Nel et al., 2009). The first tool for expression in this yeast was achieved by utilization of promoter of highly expressed *XPR2* gene which codes for secreted alkaline extracellular protease (AEP). One of the modifications of the industrial strain was to delete the gene of AEP because of the probability of degradation of the secreted r-protein by this enzyme (Nel et al., 2009). Additionally, two strong constitutive promoters including *TEF* and *RPS7* along with inducible ones have been obtained from *Y. lipolytica* (Çelik and Çalık, 2011). Promoters of isocitrate lyase (*ICLI*), 3-oxo-acyl-CoA thiolase (*POT1*), and acyl-CoA oxidase (*POX2*), *POX1* and *POX5* also codes for this enzyme, are the strongest inducible promoters of this yeast (Böer et al., 2007). Successful usage of *pICLI* and *pPOX2* for r-protein production has been reported (Böer et al., 2007).

2.1.2.2.7. *Schwanniomyces occidentalis*

Sw. occidentalis can degrade starch completely by the aid of two extracellular enzymes: α -amylase and glucoamylase; thus, *Sw. occidentalis* is widely known for its amylolytic system (Alvaro-Benito et al., 2013). Therefore, one of its advantages is the ability to grow on cheap media (Nel et al., 2009). Furthermore, this yeast can secrete proteins greater than 140 kDa efficiently. *Sw. occidentalis* also lacks hyperglycosylation (Gellissen et al., 1992; Nel et al., 2009) and does not secrete substantial amount of endogenous proteases (Nel et al., 2009).

2.1.2.2.8. *Pichia stipitis*

P. stipitis as a Crabtree negative yeast is suitable for r-protein production because it can effectively convert a carbon source to biomass (Nel et al., 2009). The complete sequencing of its genome was announced in 2007. *P. stipitis* is capable of both aerobic and oxygen-limited fermentation, and has the highest known natural ability of any yeast to directly ferment xylose and converting it to ethanol which can be considered as an economically valuable characteristic. Common promoters used for the expression in this system can be mentioned as: the promoter of native *XYLI* gene which codes for xylose reductase, the promoter of native *TKL* gene that codes for transketolase, *S. cerevisiae* *PDC1*, the promoter of *S. cerevisiae* *PGK1* gene which codes for phosphoglycerate kinase, the promoter of *S. cerevisiae* *ADHI* gene, and the promoter of *Schwanniomyces occidentalis* *GAMI* which codes for a glucoamylase. *XYLI* promoter is inducible in the presence of xylose. It should be emphasized that *P. stipitis* has a low level of glycosylation (Nel et al., 2009).

2.1.2.2.9. Methylotrophic yeasts

Just a few yeast species have methanol assimilating capabilities and, therefore, can use methanol as the sole carbon and energy source; these yeasts called methylotrophs belong to the genera *Candida*, *Torulopsis*, *Pichia*, and *Hansenula* (Gellissen et al., 1992; Hollenberg and Gellissen, 1997; Demain and Vaishnav, 2009). These yeasts can be good candidates for studying structural analysis by using ¹³C-labeled methanol as carbon source (Hollenberg and Gellissen, 1997). Methylotrophic yeasts are very attractive platforms for the industrial production of r-proteins (Demain and Vaishnav, 2009). Within this group, *P. pastoris* (syn. *Komagataella pastoris*, *Komagataella phaffii*) and *H. polymorpha* (also known as *Pichia angusta*), as two ascomycetous yeasts, have been widely used as expression platform in production of r-proteins. They can grow to high cell densities in cheap, chemically-defined medium. Methanol metabolism pathway is common in

both yeasts but significant differences in their genetics are observable (de Bruin et al., 2005; Porro et al., 2005). In addition, *P.methanolica*, and *C. boidinii* have been described as expression factories (Porro et al., 2005).

During growth in methanol-containing medium, key enzymes of the pathway are expressed in high amounts and peroxisomes proliferate; special abundance belongs to alcohol oxidase (AOX), formate dehydrogenase (FMD), and dihydroxyacetone synthase (DHAS) (Böer et al., 2007) but the most notable increase belongs to alcohol oxidase which is virtually absent in cells grown on glucose (Nel et al., 2009). AOX and DHAS proteins may constitute more than 60% of total cellular protein which shows the strength of related promoters (Nel et al., 2009); thus, promoter elements from the methanol utilization pathway can be suitable and robust candidates for control of heterologous gene expression. Nevertheless, other promoters can also be utilized for efficient expression (Böer et al., 2007). Totally, the attractiveness of methylotrophic yeasts in production of r-proteins substantially arises from their ability to reach high cell densities, as high as 130 g/L, without sophisticated fermentation processes as well as their strongly and strictly regulated promoters in methanol utilization pathway (Porro et al., 2005; Demain and Vaishnav, 2009).

2.1.2.2.9.1. *Pichia methanolica*

P. methanolica as a species of *Pichia* possesses two AOX genes designated as *AUG1* and *AUG2*, alcohol utilizing genes, induced in the presence of methanol (Böer et al., 2007). As yeast, it shares the advantages of molecular and genetic manipulations with *Saccharomyces*, and it has the added advantage of 10- to 100-fold higher r-protein expression levels.

2.1.2.2.9.2. *Candida boidinii*

C. boidinii possesses a single AOX gene represented as *AOD1* which is induced in the presence of methanol (Böer et al., 2007). Peroxisomes massively proliferate in the methylotrophic yeast *C. boidinii* when cultured

on methanol as the only carbon and energy source. It has been reported that peroxisomes are also induced by oleic acid and D-alanine as carbon sources, and that these peroxisomes contain increased amounts of the enzymes of fatty acid β -oxidation or D-amino acid oxidase, respectively (Kern et al., 2007).

2.1.2.2.9.3. *Hansenula polymorpha*

H. polymorpha is a facultative methylotrophic and can be cultivated using methanol, glycerol, or glucose as carbon sources. It has similar properties to *P. pastoris* (Romanos et al., 1992). In contrast to *P. pastoris*, *H. polymorpha* expresses just one methanol oxidase (*MOX*), as alcohol oxidase (*AOX*) gene and, thus, just has two phenotypes regarding methanol utilization (Porro et al., 2006). Three strains which possess different features and independent origins have been identified: CBS4732, DL-1, and NCYC495 (Böer et al., 2007). Products such as hirudin, hepatitis B vaccine, and interferon α -2b have been expressed successfully in this host and Hepatitis B vaccines, IFN α -2a, or insulin have reached the market. *H. polymorpha* can be recruited as a model organism for conducting researches on peroxisome biogenesis and function and nitrate assimilation, which is absent in other methylotrophs. Both inducible and constitutive promoters are available for this host (Çelik and Çalık, 2011). The promoter of peroxisome-located enzyme, methanol oxidase (*MOX*), is used to express foreign proteins as first generation. *MOX* gene is derepressed in absence of glucose and, therefore, makes a methanol-free process possible (Böer et al., 2007). As a novel promoter for this expression system, formate dehydrogenase (*FMD*) promoter has been isolated and used (Porro et al., 2005). The *MOX* promoter, *DHAS* promoter, and the *FMD* promoter are totally repressed in the presence of excess glucose and are highly induced by methanol; derepression of the mentioned promoters is possible by applying the glucose- or glycerol-limiting conditions (Nel et al., 2009). Furthermore, the nitrate assimilation pathway genes promoters can also be used in this yeast (Böer et al., 2007; Nel et al.,

2009). Constitutive promoters also can be mentioned such as the promoter of the plasma membrane H⁺-ATPase (*PMA1* promoter) and *GAP* gene promoter conditions (Nel et al., 2009).

The main difference with *P. pastoris* is that, the *MOX* gene is derepressed in absence of glucose (or glucose limitation) when using substrates such as glycerol, or sorbitol and therefore the tight regulation of promoter is lost during biomass increase phase (Romanos et al., 1992). However, its growth rate on simple media is higher in comparison with *P. pastoris* and *C. boidinii* (Nel et al., 2009). This yeast is more tolerant of higher temperatures (up to 49°C) and, therefore, the risk of contamination in large scale process may be decreased by elevated temperatures (Nel et al., 2009) and, further, makes it suitable for the expression of thermostable enzymes and proteins (Çelik and Çalık, 2011). In *H. polymorpha* it is possible to target the heterologous membrane proteins to localize in peroxisomal membranes by the targeting signal and then use of the membrane proteins in different researches; so, *H. polymorpha* can be regarded as a suitable platform for production of membrane proteins (Nel et al., 2009). A peroxisomal fate can also be helpful in the case of expressed proteins that may be deleterious to the host cell (Hollenberg and Gellissen 1997).

2.1.2.2.9.4. *Pichia pastoris* (*Komagataella phaffii*)

The methylotrophic yeast *P. pastoris* which has been reclassified as new genus of *Komagataella* is a Crabtree-negative (Mattanovich et al., 2009) with four chromosomes (De Schutter et al., 2009). This yeast is homothallic which remains haploid unless forced to mate (Nel et al., 2009). The *Pichia* genome has been sequenced for three different strains including GS115 which is auxotrophic for histidine, DSMZ 70382 and CBS 7435 with the respective genome size of 9.43 Mbp, 9.4 MbP, and 9.35 Mbp (De Schutter et al., 2009; Mattanovich et al., 2009; Küberl et al., 2011). All *P. pastoris* expression strains are derivatives of NRRL-Y 11430 strain (Northern Regional Research Laboratories, Peoria, IL) (Cereghino and Cregg, 2000)

which has been deposited as *P. pastoris* CBS 7435 (Küberl et al., 2011). During the 1970s, this yeast was considered as a promising source of single cell protein (SCP). Due to the oil crisis, such a process failed economically and, thus, the scenario changed and this yeast was considered as a potential host for r-protein production by the efforts and investigations conducted by Phillips Petroleum Company (Nel et al., 2009). Cell biologists initially became interested in *P. pastoris* because it provides a convenient system for studying peroxisome biogenesis (Sears et al., 1998). Ability to grow on a sole carbon source along with the presence of highly inducible methanol utilization pathway enzymes especially alcohol oxidase enzymes, *AOX1* and *AOX2*, which are strongly induced upon growth on methanol (Potvin et al., 2012; Staley et al., 2012), and, subsequently, their promoters makes it one of the preferences in rDNA technology.

The major advantage of *P. pastoris* over *E. coli* is the ability to produce disulfide bonds, glycosylation which inhibits the formation of misfolded proteins in the form of IBs, and extracellular production of protein (Demain and Vaishnav, 2009). In addition it conducts proteolytic processing present in other eukaryotic organisms which altogether leads to obtaining correctly folded and, thus, functional proteins. In comparison with mammalian and insect cells, this yeast needs less expensive media which makes the process more economical. Secretion pathway of *P. pastoris* resembles to eukaryotes more than *S. cerevisiae* and other non-conventional yeasts do (Corchero et al., 2012). Additionally, the notable difference between *P. pastoris* and *S. cerevisiae* is the (frequent) absence of hyperglycosylation in the former; N-glycan chain length in *P. pastoris* is shorter than that found in *S. cerevisiae*. Furthermore, *P. pastoris* expresses less antigenic proteins (Cereghino and Cregg, 2000). Moreover, *P. pastoris* secretes low levels of endogenous proteins to the extracellular medium which can lead to a less problematic purification procedure and therefore a less complex and cheaper downstream processing unit (Li et al., 2007). Cumulatively, three key reasons that make *P. pastoris* suitable for foreign

protein expression are (i) being easily manipulated at the molecular genetics, (ii) high-level protein expression due to the feasibility of high cell density fermentation, (iii) performing many "higher eukaryotic" protein modifications (Cereghino et al., 2002). Economical and well-defined mineral media containing only glycerol or methanol, biotin, salts, and trace elements (Chauhan et al., 1999; Cereghino et al., 2002) has made large-scale production of r-proteins by *P. pastoris* more convenient. Due to the preference of the respiratory pathway, contrary to *S. cerevisiae*, higher biomass production in *P. pastoris* is readily achieved which results in higher amount of recombinant product (Cregg, 1999). Fermentative pathway leads to the production of toxic fermentative product (i.e., ethanol) that is deleterious to the cells. *P. pastoris* has GRAS status and, thus, its utilization in food-related applications is allowable. From 2000 to 2010, over 700 proteins from bacteria to human have been produced in *P. pastoris* (Li et al., 2010). Some of the proteins that have been expressed in *P. pastoris* are in clinical trials to be used as human therapeutics (Nel et al., 2009). In 2009, a synthetic peptide expressed by *P. pastoris* was approved by FDA for the use in treatment of hereditary angioedema (Corchero et al., 2012).

The main problem in this yeast such as the other yeast platforms is the inability of the host to imitate the correct PTMs on the artificial product during passage from protein processing machinery (Nel et al., 2009). In addition, *P. pastoris* is unable to produce proteins that need chaperonins for correct folding. Remaining deficiencies in a glimpse can be summarized as high protease expression levels, high sensitivity to methanol levels, nutrient-deficiency when grown on defined media, and health and safety concerns associated with methanol (Potvin et al., 2012). To perform genuine glycosylation, *P. pastoris* has been genetically modified to result in human-type glycosylation (Demain and Vaishnav, 2009). Positive and negative features of *P. pastoris* as r-protein expression host have been summarized in Table 2.3.

Despite advances in protein expression field, to match a protein with a suitable expressing host and also increasing the expression yield is still a matter of "trial and error" and time and labour consuming approach (Boettner et al., 2007). In current study the selected host for r-protein expression is *P. pastoris* and, therefore, the remaining parts will be allocated specifically to this yeast.

Table 2.3. Advantages and disadvantages of *P. pastoris* expression system (Cregg, 1999).

Advantages	Disadvantages
Culturing Rapid growth High-cell density Clean medium Easy scale up to large volume Molecular genetics Availability of classical genetic methods Molecular methods similar to <i>S. cerevisiae</i> Stable integrative vectors Promoters Strong, tightly-regulated, and easily-controlled P_{AOXI} Strong constitutive P_{GAP} Expression Eukaryotic environment assists folding of r-proteins High expression level Secretion Proper PTMs Disulfide bond formation Secretion SP processing Folding Glycosylation High level fermentation product (g/L) Few endogenous proteins in medium	For high-level production of r-protein, bioreactor is often required Limited range of vectors Improper PTMs Native signals are not always processed Protein sticking in secretory pathway due to the misfolding Lower eukaryotic-type glycosylation Degradation of r-protein in medium by proteases

2.1.2.2.9.4.1. *Pichia pastoris* vectors

After choosing a host (*P. pastoris* in present study), a vector with suitable promoter (for transcription regulation) and a selectable marker should be selected for transformation of the host organism. A list of common vectors used in *P. pastoris* systems has been provided in Table 2.4.

Table 2.4. Common *P.pastoris* expression vectors (Cregg, 1999).

Name	Marker	Features
<u>Intracellular expression</u>		
pHIL-D2	HIS4	<i>AOX1</i> gene replacement happens
pAO815	HIS4	Able to generate multicopy vector
pPIC3K	HIS4 & kan ^r	Presence of MCS, G418 selection for multicopy strains
pPICZ	ble ^r	Presence of MCS, Zeocin selection for multicopy strains, potential for fusion of foreign protein to His ₆ and myc epitope tag
pHWO10	HIS4	GAP promoter
pGAPZ	ble ^r	GAP promoter, Presence of MCS, Zeocin selection for multicopy strains, potential for fusion of foreign protein to His ₆ and myc epitope tag
<u>Extracellular expression</u>		
pHIL-S1	HIS4	<i>AOX1</i> promoter, PHO1 secretion signal, several cleavage sites for insertion of foreign genes
pPIC9K	HIS4 & kan ^r	<i>AOX1</i> promoter, α -MF secretion signal, several cleavage sites for insertion of foreign genes, G418 selection for multicopy strains
pPICZ α	ble ^r	<i>AOX1</i> promoter, α -MF secretion signal, Presence of MCS, Zeocin selection for multicopy strains, potential for fusion of foreign protein to His ₆ and myc epitope tag
pGAPZ α	ble ^r	<i>GAP</i> promoter, α -MF secretion signal, Presence of MCS, Zeocin selection for multicopy strains, potential for fusion of foreign protein to His ₆ and myc epitope tag

The expression of any foreign gene in a host (here, *P. pastoris*) comprises (i) insertion of the gene in expression vector, (ii) transformation of the host cell and stable maintenance in host, (iii) selection of the desired gene-possessing strains and verifying their ability in production to select the best producing strain (Macauley-Patrick et al., 2005).

Presence of an expression cassette is one of the major features which is common in all *P. pastoris* expression vectors; it includes a promoter sequence, most often the *AOXI* gene promoter (P_{AOXI}), in the case of extracellular production a secretion signal sequence which is fused in-frame with the gene of interest, a transcriptional termination (TT) sequence derived from *AOXI* gene that directs efficient 3' processing and polyadenylation of the mRNAs, and a site for insertion of gene of interest called multiple cloning site (MCS). The MCS, containing multiple restriction sites in order to facilitate digestion by suitable restriction enzymes, is located between the promoter and terminator sequence for insertion of the desired gene. The commercial *P. pastoris* expression systems are supplied by Invitrogen Corporation; the vectors comprise either of inducible P_{AOXI} , or the constitutive *GAP* promoter (P_{GAP}) and the *HIS4* gene (or the *Sh ble* gene) for selection. *Sh ble* (or *Zeo*) gene confers resistance to the drug Zeocin™ and substantially decreases the size of plasmids. In the case of application of *HIS4*-based vectors, the *his4* auxotroph strains should be employed; however, when Zeocin™ -resistance-based vectors are used the prototrophic strain should be recruited (Cregg, 1999).

2.1.2.2.9.4.2. *Pichia pastoris* transfection

P. pastoris can be transfected by electroporation, or a spheroplast generation method, or whole cell methods such as those including lithium chloride (or lithium acetate) and polyethylene glycol (Nel et al., 2009). Since lithium acetate method does not work for *Pichia* transfection, lithium chloride method is selected as the method of choice (Invitrogen, 2010). At this time, the lithium chloride and electroporation methods are the common

transfection methods. High-frequency DNA transformation is one of the important features of *P. pastoris* (Nel et al., 2009).

In general, both of episomal and integrative vectors are available for the hosts; i.e., after transformation, the newly-introduced (recombinant) DNA can either integrate itself into the host genome (chromosomal DNA) or remain separately and replicate autonomously as a circular plasmid referred as episomal (Nel et al., 2009). Integrative vectors could provide more benefits as the problem of plasmid loss will be minimized in other words, the genetic instability due to segregational loss will be eliminated (Wu et al., 2003a) and, consequently, uniform expression will be achieved (Sears et al., 1998); in order to have a robust industrial process, the strain should be genetically stable without any need to selective pressure (Porro et al., 2005). Similar to *S. cerevisiae* there is a tendency for homologous recombination between genomic and introduced DNA in *P. pastoris* which leads to either single cross over or gene replacement events. This phenomenon is stimulated by plasmid digestion within a sequence shared by the host genome; free DNA end facilitates homologous recombination (Cregg, 1999; Nel et al., 2009). In order to have more stable transfectants, the expression is conducted preferentially by chromosomal integration with the help of the vectors designed to be integrative (Sreekrishna et al., 1997). However, utilization of a few episomal vectors has been reported in this host (Lee et al., 2005; Hong et al., 2006). All *P. pastoris* expression vectors have been designed as *E. coli* / *P. pastoris* shuttle vectors (Cereghino and Cregg, 2000). Gene insertion and gene replacement events are depicted in Figure 2.3. In *P. pastoris* gene replacement happens at lower frequencies compared to *S. cerevisiae* (Nel et al., 2009).

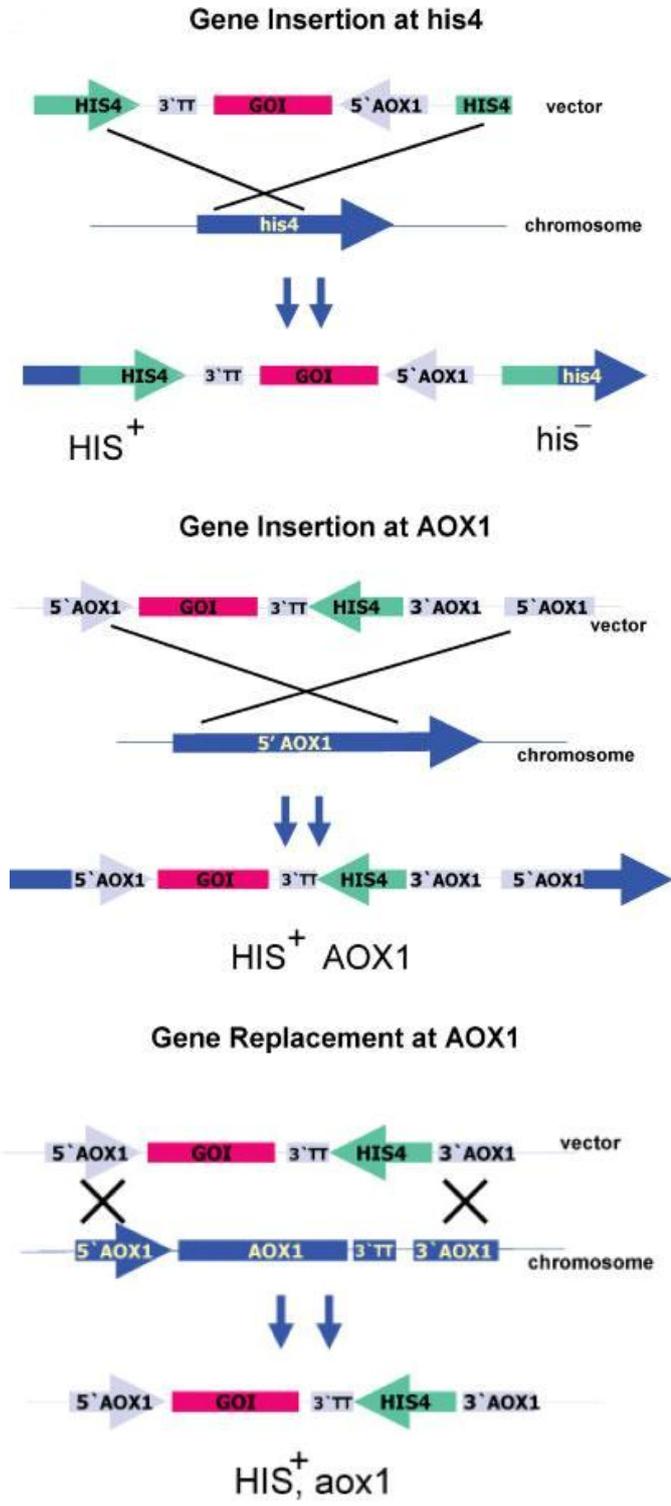


Figure 2.3. Linearized vector and its integration to the *P. pastoris* chromosome (Daly and Hearn, 2005).

Regarding the ability of *P. pastoris* growth on methanol the strain phenotype is different; Mut⁺ is the wild type which possesses both *AOX1* and *AOX2* genes, Mut^s has undergone *AOX1* gene deletion, and Mut⁻ which both *AOX* genes have been removed. The last strain is unable to grow on methanol. It should be reminded that foreign genes under P_{AOX1} control are still induced by methanol (Cregg, 1999). In the case of intracellular expression, Mut^s cells are preferred to be employed because of increased specific yield of r-protein. If secretion situation is intended, either one of Mut⁺ or Mut^s strains can be used (Chiruvolu et al., 1997).

2.2. Notable yield-affecting parameters

Considering the fermentation, parameters that can have effect on r-protein expression yield can be considered in two different levels: genetic level (or cell level) and process level (or cultivation level); virtually they can be supposed as micro-level and macro-level parameters (Macauley-Patrick et al., 2005). The former is related with the gene of interest, the expression host, the vector and the latter takes into account all factors that can affect the fermentation in shake flask (or bioreactor) after transformation completed.

2.2.1. Genetic parameters

Parameters in this category are genetic-related features and elements before initiation of cultivation for production; factors such as strain of the host cell, codon usage bias of the host in comparison with the gene of interest, A+T and G+C content of the gene of interest, inserted gene copy number, promoter, secretion signal peptide, and epitope tag in the vector can be included in this class.

2.2.1.1. Promoters

One of the essential elements required for any expression system is a strong promoter (Cregg, 1999). The site on the DNA to which the RNA polymerase molecule binds prior to initiation of transcription is called the promoter. In

other words, promoter is the mediator of transcription which provides specific binding site for transcription machinery and transcription factors (Vogl and Glieder, 2013); the schematic representation is represented in Figure 2.4. In addition to providing a binding site for the polymerase, the promoter contains the information that determines which of the two DNA strands is transcribed and the site at which transcription begins (Karp, 2008). The major distinction between transcription in prokaryotes and eukaryotes is the requirement in eukaryotes for a large variety of *accessory proteins*, or *transcription factors* (TFs) (Karp, 2008); cellular RNA polymerases are not capable of recognizing promoters on their own but require the help of these additional proteins.

Choice of the promoter for heterologous gene expression highly affects yield but the toxicity of the product must be considered in this selection (Çelik and Çalık, 2011). The promoter can be native to the host as well as from another microorganism. In the case of yeasts, generally, promoters of the yeast species are recruited (homologous promoters) since non-yeast promoters (heterologous promoters) often do not conduce to good efficiencies (Porro et al., 2005).

Promoters can either be inducible (regulated) or constitutive based on the mechanism of action. The former one is dormant and the gene under control is not expressed until the inducer compound would be present but the latter is permanently on and expression continues. Most foreign proteins are at least somewhat deleterious to the cell when expressed at high levels, thus, the ability to maintain the culture in a repressed or "expression off" mode is preferable (Cregg, 1999). Classically used promoters (benchmark promoters) in *P. pastoris* expression system are alcohol oxidase 1 gene promoter (P_{AOX1}) and glyceraldehyde-3-phosphate dehydrogenase gene promoter (P_{GAP}) (Vogl and Glieder, 2013). Most of the promoters in *P. pastoris* expression system such as P_{AOX1} , P_{AOX2} , P_{FLD1} , P_{PEX8} , P_{YPT1} , P_{DHAS} , P_{ICLI} , P_{TEF} and P_{GAP} are controlled by the carbon sources for foreign gene expression and some of them might also be regulated by nitrogen source, such as P_{FLD1} (Zhang et al.,

2009). It is useful to have a selection of different promoters suitable for recombinant expression of heterologous or homologous genes, varying from strong promoter activity to weak or reduced promoter activity for host engineering purposes (Stadlmayr et al., 2010). The promoters have been hitherto used in *P. pastoris* have been presented in Table 2.5.

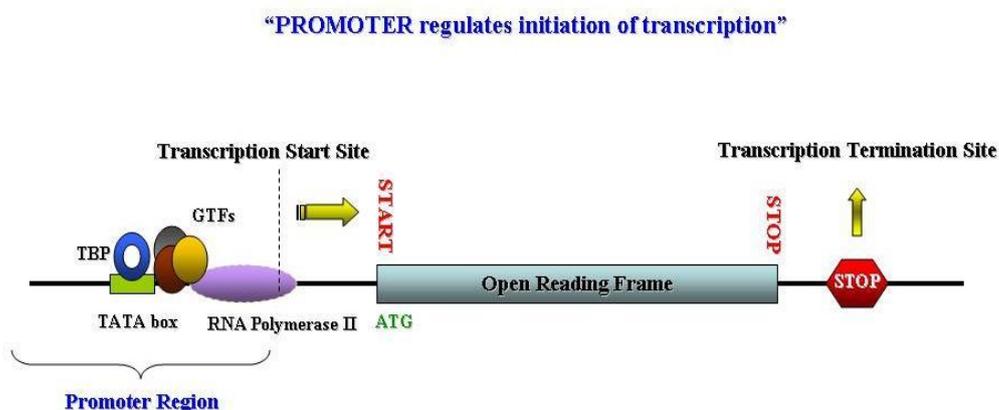


Figure 2.4. Simple representation and explanation of the promoter region of a gene with the general structure of the gene under control. There are untranslated regions (UTRs) before and after the coding sequence (ORF) which possess special responsibilities. GTFs: general transcription factors and TBP: TATA-binding protein.

2.2.1.1.1. Inducible promoters (Gene switches)

High level expression of a foreign protein can impose a significant metabolic burden on the host which leads to the reduced growth rate and gene expression. Furthermore, the recombinant product may be toxic for the host cell, thus, the use of tightly regulated promoters is advantageous due to the having a more controllable expression as the result of separated growth and induction phases.

Table 2.5. Constitutive and inducible promoters used in *P. pastoris* for expression of r-proteins. Expression level is relative to two classic P_{AOXI} and P_{GAP} (Çalık et al., 2015).

Gene Name	Gene Product	Regulation	Expression Level
Constitutive			
<i>GAP</i>	Glyceraldehyde 3-phosphate dehydrogenase	Expression on glucose, to a lesser extent on glycerol and methanol	Strong promoter
<i>ENO1</i>	Enolase	-	~ 20-70% of P_{GAP}
<i>GPM1</i>	Phosphoglycerate mutase	-	~ 15-40% of P_{GAP}
<i>HSP82</i>	Cytoplasmic chaperone (Hsp 90 family)	-	~ 10-40% of P_{GAP}
<i>KAR2</i>	ER resident chaperone (also termed BiP)	-	~ 10-70% of P_{GAP}
<i>PET9</i>	ADP/ATP carrier of the inner mito-chondrial membrane	-	~ 10-1700% of P_{GAP}
<i>PGK1</i>	Phosphoglycerate kinase	-	~ 10% of P_{GAP}
<i>SSA4</i>	Heat shock protein	-	~ 10-25% of P_{GAP}
<i>TEF1</i>	Translation elongation factor 1 α	Strong growth association	Strong (similar to P_{GAP})
<i>TPH1</i>	Triosephosphate isomerase	-	~ 10 -80% of P_{GAP}
<i>YPT1</i>	GTPase involved in secretion	-	Weak
<i>GCW14</i>	Potential glycosyl phosphatidyl inositol (GPI)- anchored protein	Expression on glucose, glycerol, and methanol	5-10 fold higher than P_{GAP}
<i>SDH</i>	Sorbitol dehydrogenase	-	~ Similar to P_{GAP}

Table 2.5. Constitutive and inducible promoters used in *P. pastoris* for expression of r-proteins. Expression level is relative to two classic P_{AOX1} and P_{GAP} (**Continued**).

Gene Name	Gene Product	Regulation	Expression Level
Inducible			
<i>AOX1</i>	Alcohol oxidase 1	Induction by methanol	Strong (~30% of total protein)
<i>AOX2</i>	Alcohol oxidase 2	Induction by methanol	~ 5-10% of P_{AOX1}
<i>AOD</i>	Alternative oxidase	Expression on glucose but not on methanol or upon glucose depletion	~ 40% of pGAP
<i>DAS</i>	Dihydroxyacetone synthase	Induction by methanol	Strong (similar to P_{AOX1})
<i>FLD1</i>	Formaldehyde dehydrogenase	Induction by methanol and methylamine	Strong (similar to P_{AOX1})
<i>ICL1</i>	Isocitrate lyase	Repression by glucose, derepression by glucose depletion and induction by ethanol	Not compared to classic promoters
<i>PEX8</i>	Peroxisomal matrix protein	Induction by methanol or oleate	Weak
<i>PHO89</i> or <i>NSP</i>	(putative) Sodium-coupled phosphate symporter	Induction by phosphate limitation	Strong (similar to P_{GAP})
<i>THI11</i>	Protein involved in thiamine biosynthesis	Complete repression by Thiamine	~ 70% of P_{GAP} on medium lacking thiamine; 63% P_{GAP}
<i>ADH1</i>	Alcohol dehydrogenase	Repression by glucose and methanol, induction by glycerol and ethanol	-

Table 2.5. Constitutive and inducible promoters used in *P. pastoris* for expression of r-proteins. Expression level is relative to two classic P_{AOXI} and P_{GAP} (**Continued**).

Gene Name	Gene Product	Regulation	Expression Level
<i>ADH3</i>	Protein involved in Ethanol utilization	Induced by ethanol	Strong (similar to P_{AOXI} and P_{GAP}) (Karaoglan et al., 2016b)
<i>GUT1</i>	Glycerol Kinase	Repression by methanol, induction by glucose, glycerol and ethanol	-
<i>G1</i>	High-affinity glucose transporter	Glycerol repression, induction by glucose limitation	Stronger than P_{GAP} (~ 230% P_{GAP})
<i>G6</i>	Putative aldehyde dehydrogenase	Glycerol repression, induction by glucose limitation	Weaker than P_{GAP} (~ 40% P_{GAP})
<i>THR1</i>	Homoserine kinase	Repression by addition of L-threonine, L-valine, L-leucine and L-isoleucine	13% P_{GAP} In non-repressing conditions
<i>MET3</i>	ATP sulfurylase	Repression by addition of L-methionine	13% P_{GAP} In non-repressing conditions
<i>SER1</i>	3-phosphoserine aminotransferase	Repression by addition of L-serine	1% P_{GAP} In non-repressing conditions
<i>PIS1</i>	Phosphatidylinositol synthase	Repression by addition of zinc	40% P_{GAP} In non-repressing conditions

2.2.1.1.1. Alcohol oxidase 1 gene promoter

P. pastoris possesses a highly inducible methanol utilization pathway (MUT) which has been depicted in Figure 2.5.

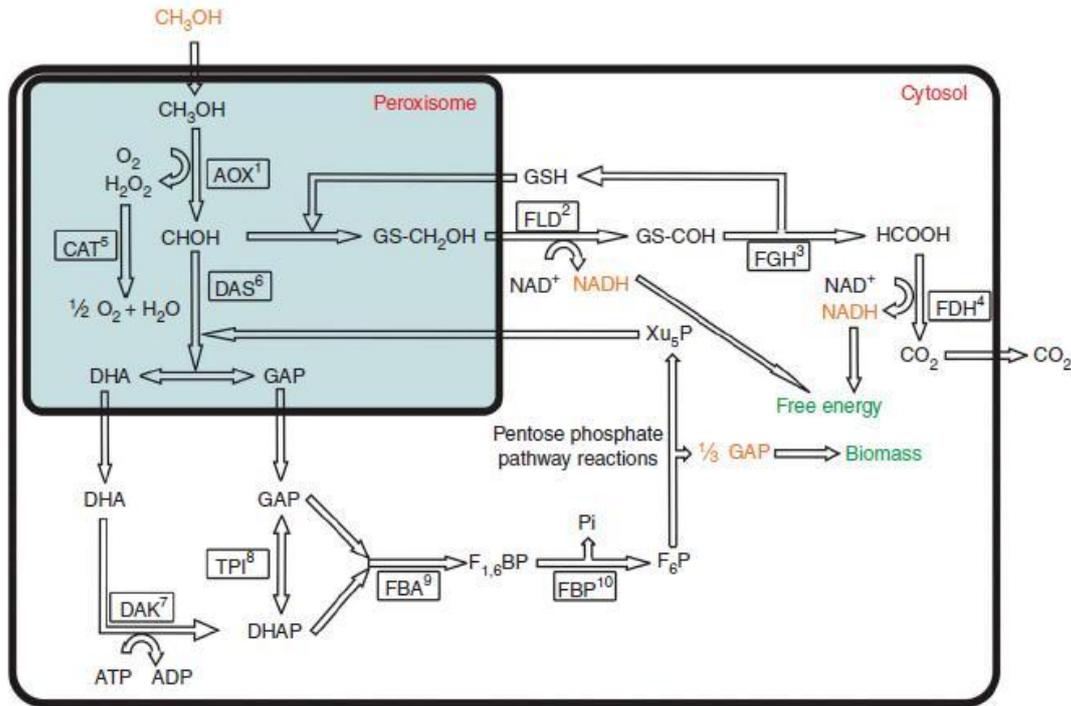


Figure 2.5. The methanol utilization pathway in *P. pastoris* (De Schutter et al., 2009).

1.AOX: alcohol oxidase, 2.FLD: formaldehyde dehydrogenase, 3. FGH: S-formylglutathione hydrolase, 4. FDH: formate dehydrogenase, 5. CAT: catalase, 6. DAS (~DHAS): dihydroxyacetone synthase, 7. DAK: dihydroxyacetone kinase, 8. TPI: triosephosphate isomerase, 9. FBA: fructose-1,6- biphosphate aldolase, 10. FBP: fructose-1,6- biphosphatase

The base for the heterologous protein production in *P. pastoris* is the availability of methanol metabolism enzymes when grown on methanol, i.e., they are induced by methanol. Most promoters used for *P. pastoris* up to date have been derived from genes that code for enzymes involved in the methanol metabolism (Stadlmayr et al., 2010). However, P_{AOX1} has been the most widely used because of its strength and tight regulation (Sreekrishna et

al., 1997; Nel et al., 2009). The highest expression level in *P. pastoris* have been reported by the use of P_{AOX1} which were 22 g/L and 15 g/L intracellularly and extracellularly respectively (Vogl and Glieder, 2013). Alcohol oxidase enzyme is coded by two genes *AOX1* and *AOX2*. According to the MUT, cells that grow on methanol have high oxygen requirement which urges to add pure oxygen to the medium which increases the cost of the process (Bollok et al., 2009). As the affinity of AOX1 enzyme for oxygen is low, the *P. pastoris* cell by overexpression of this enzyme omits this shortage and, thus, the enzyme can reach 30% of total cellular protein when grown on methanol as the only carbon source (Daly and Hearn, 2005). Cells that grow on repressing carbon compounds (such as glucose, glycerol, and ethanol) show complete absence of AOX activity. As soon as the repressing carbon source in the medium is going to be depleted or becomes very low in the medium, the promoter reaches the derepression state (Vogl and Glieder, 2013). P_{AOX1} is induced only by methanol as the only (and cheaper) carbon source but supplemental carbon sources may improve the production of r-protein (Zhang et al., 2009). Regulation and induction of heterologous gene expression by methanol have been shown to be simple, easy to scale-up and cost-effective for large-scale fermentations (Wu et al., 2003a). Mut^s phenotype has the benefit that leads to consumption of less amount of methanol for expression; growth can not be supported just by methanol and addition of an alternative carbon source, such as sorbitol, mannitol, alanine, and trehalose all of which have not any repressing effect on P_{AOX1} during expression phase, is inevitable (Daly and Hearn, 2005).

Although the P_{AOX1} has been successfully used to recombinant protein production, it should be reminded that (i) high methanol concentrations will conduce to accumulation of formaldehyde and hydrogen peroxide inside the cells as toxic substances and, so, methanol addition should be finely tuned, (ii) methanol is a potential fire hazard and storage and delivery of large amounts causes safety problems, (iii) methanol is not suitable for production of food stuffs (iv) relatively long fermentation times required along with

sophisticated feeding strategies for industrial applications, (v) strong induction on methanol may exceed the secretory capacity of the host cells (Menendez et al., 2003; Ahn et al., 2007; Nel et al., 2009; Stadlmayr et al., 2010), and (vi) in small scale productions such as shake flasks, the volatility of inducer methanol may cause problem also the limited oxygen supply in shake flasks can be another problem to provide sufficient oxygen for methanol metabolism (Resina et al., 2005); advantages and disadvantages of P_{AOX1} has been summarized in Table 2.6. Consequently, there is a strong demand for regulated promoters other than the methanol inducible, i.e. methanol-independent ones, where no toxic inducer would be required. It should be emphasized that, some inducers may be expensive or could interfere with the purification of the end product (Porro et al., 2005).

Table 2.6. Advantages and disadvantages of P_{AOX1} -based expression system (Macauley-Patrick et al., 2005).

Advantages	Disadvantages
Tight regulation of r-protein expression by repression/derepression mechanism	Difficulties in methanol monitoring during process
Expression of high levels of r-protein even if they are cytotoxic	As a fire hazard, storing large amounts of methanol is undesirable
Transcription repression by the initial carbon source leads to acceptable cell growth before overexpression of r-protein	Due to the mainly petrochemical origin of the methanol, its utilization in food products and additives is not suitable
Easy induction of transcription by methanol addition	Requirement for two carbon sources

2.2.1.1.1.2. Alcohol oxidase 2 gene promoter

The alcohol oxidase 2 gene promoter (P_{AOX2}) can be considered as another inducible promoter option. $AOX2$ gene also produces alcohol oxidase but with a 10-20 times less AOX activity than $AOX1$ gene (Macauley-Patrick et al., 2005); as the result of transcriptional regulation $AOX1$ is much more

strongly transcribed than *AOX2* (Vogl and Glieder, 2013) therefore, the methanol induction system is usually adopted for *AOX1* promoter. Two genes (*AOX1* and *AOX2*) are 97% identical in amino acid sequences but no clear similarity in promoter part is observed (Vogl and Glieder, 2013). The pattern of action in P_{AOX2} is similar to P_{AOX1} where it also requires absence of repressing carbon source and presence of inducer. If the limiting step in the way to reach an active protein is its folding or processing in secretory machinery, then it will be wise to use a weak promoter which causes weaker expression (Vogl and Glieder, 2013). Because of lower expression level of *AOX2*, the strains with this phenotype (Mut^s) grow slowly on methanol. In some deletion and insertion studies carried out for revealing the pattern of regulation in *AOX* promoters, in addition to finding positive and negative regulatory sequences, it was understood that deletion of *AOX1* gene profoundly decreased the growth rate on methanol; however, the deletion of *AOX2* led to a wild type-like growth (Vogl and Glieder, 2013). Physicochemical environment has effect on this promoter and as reported, addition of oleic acid to medium in production of human serum albumin has increased the protein expression by improving transcriptional regulation (Macauley-Patrick et al., 2005).

2.2.1.1.1.3. Dehydroxyacetone synthase gene promoter

In addition to the P_{AOX1} , other methanol-induced dehydroxyacetone synthase promoter (P_{DHAS}) has also been studied to express r-proteins. The DHAS protein solely can constitute up to 20% of the total cell protein when grown on methanol (Daly and Hearn, 2005). After sequencing of the whole genome of *P. pastoris*, it was comprehended that there are two very similar *DHAS* genes (91% similar) but the used one in researches has been *DHAS2* gene. It has been noticed that both *DHAS1* and *DHAS2* are induced to same extent by methanol (Vogl and Glieder, 2013). In contrast to P_{AOX1} which is induced by glucose starvation, such a pattern was not observed for P_{DHAS} ; the P_{DHAS} is not activated by carbon source starvation. In the production of β -

galactosidase when the transformants were grown on methanol, P_{AOXI} expressed five-time greater product in comparison to P_{DHAS} (Daly and Hearn, 2005).

2.2.1.1.1.4. Formaldehyde dehydrogenase gene promoter

Some of the yeasts such as *P. pastoris* are able to use methylated amines as sole nitrogen source. *FLDI* (glutathione-dependent formaldehyde dehydrogenase) gene is inducible by both methanol and methylamine (Macauley-Patrick et al., 2005) and, thus, allows the investigator to choose carbon or nitrogen source regulation with the same expressing strain. Formaldehyde has role in both methanol and methylamine pathways. Together with methanol metabolism, Formaldehyde dehydrogenase (FLD) enzyme plays role in assimilation of some C₁-amines such as methylamine as nitrogen source (Resina et al., 2005; Cos et al., 2006) which has been simply represented in Figure 2.6. The FLD enzyme can form up to 20% of the total cell protein (Daly and Hearn, 2005) and it clearly shows the strength of its promoter. Therefore, P_{FLDI} offers the flexibility of inducing high levels of protein expression with methylamine, as an inexpensive nitrogen source (Li et al., 2007). P_{FLDI} can be an attractive alternative for methanol-independent expression of r-protein (Macauley-Patrick et al., 2005; Vogl and Glieder, 2013). It should be reminded that yeasts are not able to use methylamine as the sole carbon and nitrogen source and a carbon source should accompany it (Cos et al., 2006).

Recently a P_{FLDI} -based vector has been designed (Figure 2.7) by Life Technologies Corporation (Vogl and Glieder, 2013). As a typical research regarding P_{FLDI} , the production of *Rhizopus oryzae* lipase in *P. pastoris* has been conducted in some recent researches in different conditions (Resina et al., 2005; Resina et al., 2009). Specific growth rate has proved to be a key parameter for the productivity of secreted proteins in *P. pastoris* and, in particular, of P_{FLDI} -based systems (Resina et al., 2009).

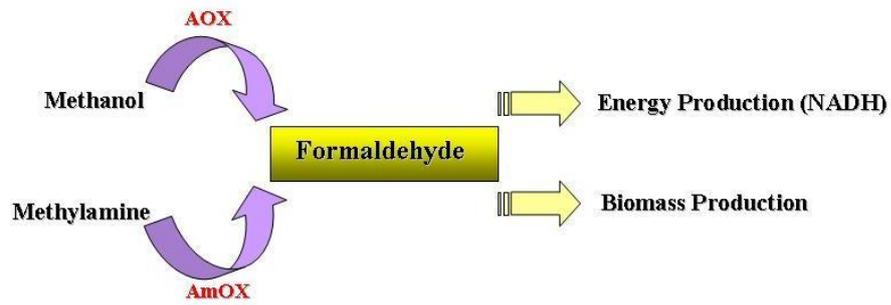


Figure 2.6. The role of intermediate formaldehyde (Macauley-Patrick et al., 2005). By the oxidation of methylamine in peroxisomes using methylamine oxidase (AmOX) formaldehyde is produced which either can be oxidized to CO₂ or changed to biomass (Cos et al., 2006).

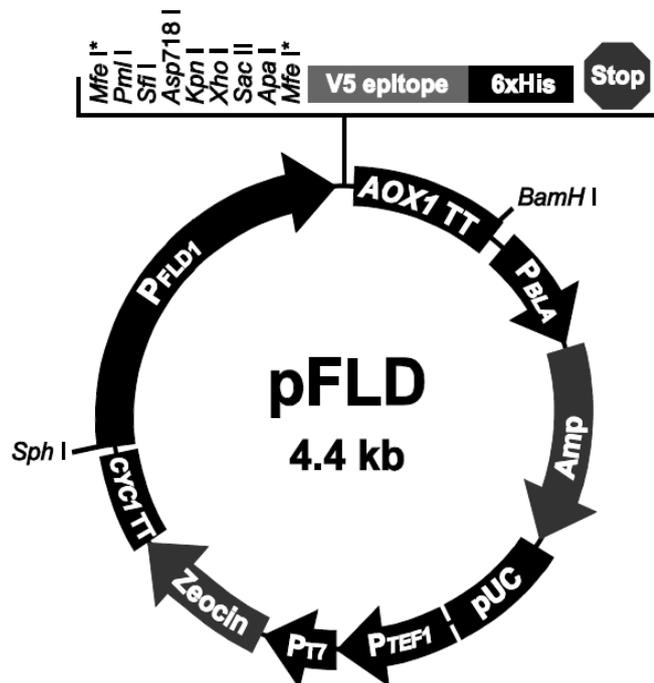


Figure 2.7. Map of the pFLD (Web 13).

2.2.1.1.1.5. Isocitrate lyase gene promoter

Isocitrate lyase is an enzyme in the glyoxylate cycle that catalyzes the cleavage of isocitrate to succinate and glyoxylate. Together with malate synthase, it bypasses the two decarboxylation steps of the tricarboxylic acid cycle (TCA cycle) and is used by bacteria, fungi, and plants. isocitrate lyase gene (*ICLI*) promoter (P_{ICLI}) is induced by ethanol (Stadlmayr et al., 2010), repressed by glucose, and is derepressed by glucose absence (Menendez et al., 2003). It can be regarded as a potential alternative for commonly used P_{AOXI} and P_{GAP} which needs more investigation. However, there are no comprehensive results about the direct comparison of P_{AOXI} and P_{ICLI} strengths and, therefore, in each case of production a preliminary screening may be troubleshooting. Although it is most probable that P_{AOXI} leads to better results but if avoiding methanol is inevitable, by the changing the gene copy number (gene dosage) the lower strength of other choices may be compensated. The *ICLI* gene from *P. pastoris* has been cloned and characterized and its promoter was used for the expression of dextranase from *Penicillium minioluteum* (Menendez et al., 2003).

2.2.1.1.1.6. *CUPI* gene promoter

In some cases it is not so desirable to alter the nitrogen or glucose source to regulate the expression of a protein and the use of another inducer is more pertinent (Koller et al., 2000). The benefit of copper-inducible system is that it does not interfere with the normal physiology of the affected organism. The copper-inducible *CUPI* promoter (P_{CUPi}) from *S. cerevisiae* has also been used in *P. pastoris* system that can act as an alternative and it was showed that the amount of expression depends on amount of copper in the medium (Koller et al., 2000). It is based on control elements that regulate the expression of copper detoxification genes in *S. cerevisiae* in response to elevated copper concentrations (Gatz and Lenk, 1998). Since, the addition of copper, as inducer, to the medium may cause health problems, thus, this promoter can be considered as the least important and compatible choice.

2.2.1.1.1.7. *PHO89* gene promoter

PHO89 gene codes for a putative sodium-coupled phosphate symporter. Phosphate-responsive promoters have been previously analyzed in *E. coli* and *S. cerevisiae*. The promoter of this gene is strongly dependent on phosphate and is regulated by phosphate available in growth medium (binding sites on promoter are available for transcription factor Pho4p). Using lipase as reporter it was revealed that at high phosphate concentrations no lipase activity was observed, but in reduced phosphate concentrations lipase activity similar to P_{TEFI} - and P_{GAP} - based systems was achieved. This expression system could be promising candidate for methanol-free systems but different reporter enzymes should be assessed to analyze the extent of expression (Vogl and Glieder, 2013).

2.2.1.1.1.8. alcohol dehydrogenase gene promoter

Two alcohol dehydrogenase (*ADH*) genes have been annotated from the genome of *P. pastoris*, *ADH* and *ADH3*, according to the available sequence homology to the *S. cerevisiae* *ADH* genes. *ADH3* gene is solely responsible for ethanol utilization in *P. pastoris* and *ADH* gene does not have any role in ethanol metabolism (Karaoglan et al., 2016a). *ADH3* gene promoter (P_{ADH3}) strength has been compared with P_{AOXI} and P_{GAP} in production of *A. niger* xylanase. The results revealed that P_{ADH3} can be a promising alternative for generally used P_{AOXI} and P_{GAP} (Karaoglan et al., 2016b).

2.2.1.1.1.9. Alternative oxidase gene promoter

Alternative oxidases (AODs) are present in the mitochondria of plants, fungi and many types of yeast. These enzymes transfer electrons from the ubiquinol pool directly to oxygen without contributing to the proton transfer across the mitochondrial membrane and, thus, act such as alternative shortcut to the standard respiratory pathway. This allows respiration even in the presence of complex III and IV inhibitors like antimycin A or cyanide (Kern et al., 2007). The findings suggest a major contribution of the AOD to *P.*

pastoris cell viability. By glucose depletion, the expression from P_{AOD} stops (Vogl and Glieder, 2013) and by considering that P_{AOD} is not induced by methanol, it is inferred the regulation is dependent upon carbon source. So, by the current knowledge it is postulated that P_{AOD} is a strong promoter that on glucose can be compared to P_{GAP} and encounters a repression or missing induction when glucose is exhausted or methanol is available (Vogl and Glieder, 2013).

2.2.1.1.2. Constitutive promoters

The type of the promoter is highly dependent on the nature of the product and, thus, its subsequent effect on the regular metabolic activity and proliferation of the host organism. In the case of expression of a host-friendly compound by the cell, the constitutive mode can be chosen where the production and biomass accumulation (cell growth) are synchronized.

2.2.1.1.2.1. Glyceraldehyde-3-phosphate dehydrogenase gene promoter

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH or GAP) is a key enzyme in the glycolytic and gluconeogenesis pathways (Figure 2.8). It is major protein in the organisms such as brown algae and cyanobacteria; in addition, presence in the amount of 2-5% of the poly(A) RNA in yeasts suggesting that the P_{GAP} is very strong and active (Jia et al., 2012). The nucleotide sequence of the *GAPDH* gene of the *P. pastoris* has been obtained with the aid of probing the genomic DNA with the *GAPDH* gene available from *S. cerevisiae* (Waterham et al., 1997); it has been indicated that *P. pastoris* genome contains only one *GAP* gene. The gene product has a molecular weight of 35.4 kDa (including 333 amino acids) and the P_{GAP} is an approximately 500-bp fragment (Waterham et al., 1997).

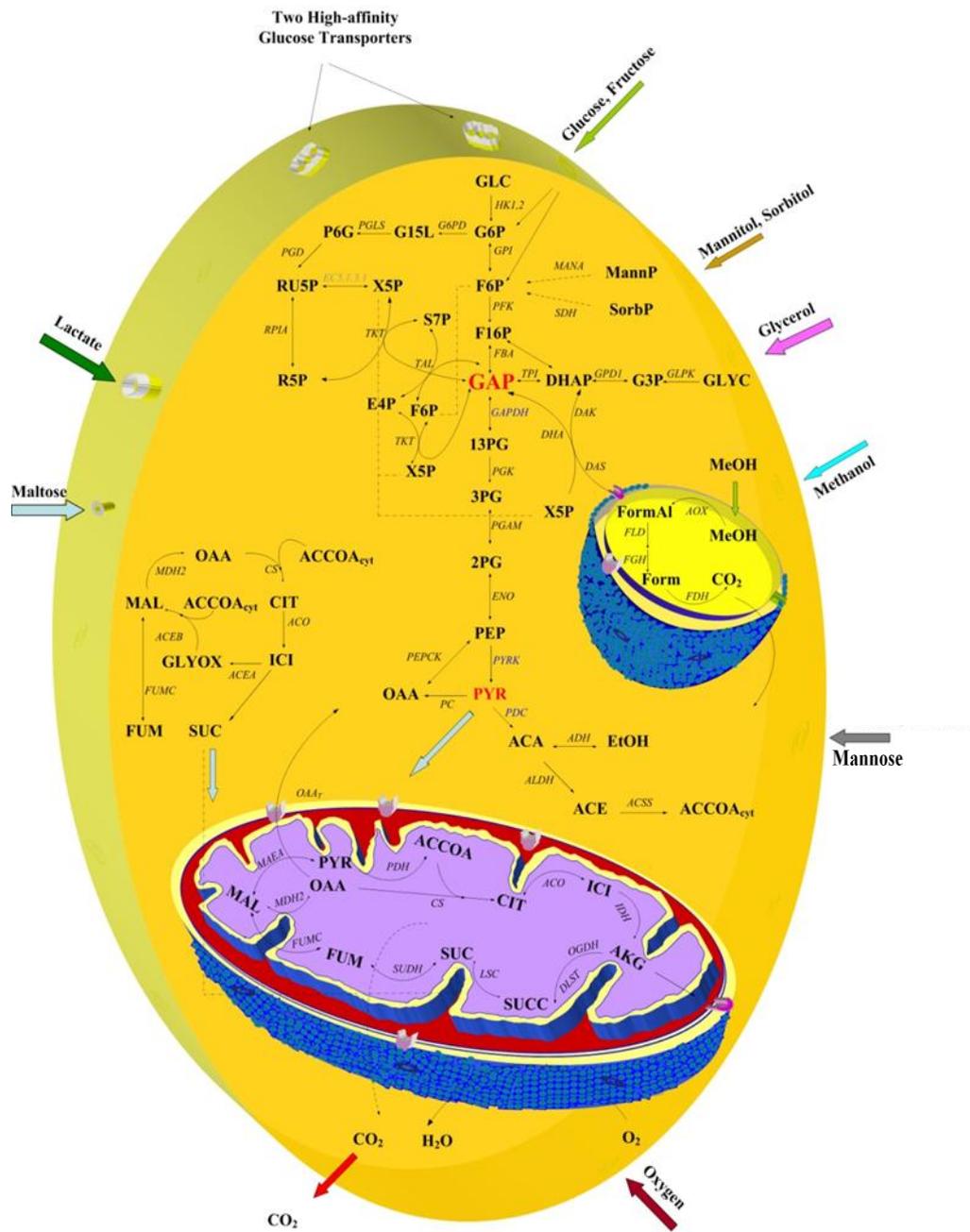


Figure 2.8. Central carbon metabolism of *P. pastoris* (Çalık et al., 2015).

P_{GAP} is one of the strongest constitutive promoters in yeasts and, consequently, *GAP* promoter-based system was developed as an alternative to *AOXI* promoter-based system (Wu et al., 2003a; Ahn et al., 2007; Khasa et al., 2007). P_{GAP} is not a suitable choice in the case of proteins that have toxic effects on the cell (cytotoxic effects) and is a convenient alternative in the case of non-toxic products (Wu et al., 2003a; Li et al., 2007; Driss et al., 2012). P_{GAP} -based production can cross out the imperfections related with inducible expression and, thus, can confer greater variability to the *P. pastoris*-based expression systems. Conflicting results have been reported by different authors in comparing efficiency of classic promoters, P_{AOXI} and P_{GAP} ; some data approve P_{AOXI} as better and some P_{GAP} (Zhang et al., 2009). Synchronization of product formation and biomass accumulation will never guarantee the uniformity of product yield ($Y_{P/X}$) in different r-proteins where some proteins may be "difficult to be expressed" (Pal et al., 2006). This system can be easily adapted for continuous production (Wu et al., 2003a). The P_{GAP} -containing vectors allow for continuous and large-scale production of the recombinant product avoiding the fed-batch fermentation by methanol inducible system which is complex and needs technically sophisticated carbon source shifts (Kottmeier et al., 2012). Thus, the microorganism growth and r-protein production become straightforward and the hazards and costs arose with the storage and delivery of large volumes of methanol are eliminated (Cos et al., 2006). In the case of production of recombinant human chitinase, the amount of intact protein increased perhaps because of continuous separation from the culture as a result of continuous operation mode (Goodrich et al., 2001). However, in some instances the negative effect of P_{GAP} on the specific growth rate of the microorganism showed the inefficiency of this promoter (Cos et al., 2006).

P_{GAP} -driven systems can be a solution for accumulation of formaldehyde and hydrogen peroxide in the host organism due to the methanol metabolism (Zhao et al., 2008) which in turn results in the less contaminating proteins as the direct consequence of the cell lysis in

comparison to P_{AOXI} -driven systems where methanol metabolism and high cell density cultures lead to lower cell viability (Wang et al., 2012).

Because of elimination of the induction phase, cultivation under controlled conditions is not so vital in the case of P_{GAP} and control requirements are minimized (Potvin et al., 2012). Furthermore, this system needs no washing to put aside non-methanolic carbon sources as in methanol induction phase (Cos et al., 2006).

GAP is actually a constitutively expressed enzyme, but, it has been shown that carbon source has effect on the strength of the expression (Waterham et al., 1997). Glucose, methanol, glycerol, and oleic acid have been reported as the sole carbon sources for systems use P_{GAP} (Zhang et al., 2009); where glucose resulted in the highest expression levels and glycerol and methanol as the subsequent substitutes for obtaining high expression levels (Potvin et al., 2012); generally, the used substrates are glucose or glycerol (Cos et al., 2006).

As described, usage of P_{GAP} instead of P_{AOXI} leads to replacing methanol with sources such as glucose. By using methanol, oxygen limitation could lead to decreased growth, as the growth on methanol is aerobic, regardless of methanol provided. However, when glucose is used as carbon source oxygen limitation leads to facultative anaerobic behavior, namely production of ethanol as by-product, by shifting to a different pathway; P_{GAP} gives the opportunity to examine the hypoxic conditions better (Baumann et al., 2008). The increased flux of glycolytic pathway in low oxygen availability conditions can be representative of increased expression of genes such as GAP gene (and hence, its promoter identity); thus P_{GAP} has the possible potential to be utilized in production of proteins in hypoxic conditions. The major advantage of this approach will be lower aeration rate (Baumann et al., 2008) and, thus, paving the way to a more economic process.

According to the homology present between yeast species, P_{GAP} of the *P. pastoris* has been recruited successfully in *S. cerevisiae* expression

system in order to intracellular expression of hepatitis B surface antigen (Vellanki et al., 2007). In addition, efficient gene expression by P_{GAP} from *P. pastoris* in a strain of *P. thermomethanolica*, has been reported recently (Tanapongpipat et al., 2012). As a different expression strategy, P_{AOXI} and P_{GAP} were utilized in combination in order to increase the expression level of P_{GAP} system; the purpose was to induce P_{AOXI} with methanol after exhaustion of glucose. As an example, the production of human granulocyte-macrophage colony-stimulating factor (hGM-CSF) increased when expressed under the combined P_{AOXI} and P_{GAP} (Wu et al., 2003a). In the case of presence of combined promoters, final cell concentration reached to lower amount than the condition of sole P_{GAP} but the secreted r-protein was higher in the former which suggests that instead of being used for growth, part of resources were diverted to express the r-protein when methanol was added to turn on P_{AOXI} (Wu et al., 2003a). A sequential expression strategy can also be developed to separately produce and recover two desired r-proteins (Wu et al., 2003b). A comprehensive and thorough review on P_{GAP} -based expression systems in *P. pastoris* from utilization of carbon source to different process conditions and bioreactor feeding strategies adopted hitherto is also available which can be used (Çalık et al., 2015).

2.2.1.1.2.2. *YPT1* gene promoter

YPT1 gene encodes a small GTPase involved in secretion (Sears et al., 1998); its promoter provides constitutive but low level (moderate constitutive) expression (Zhang et al., 2009). As an outcome of the probability of high-level expression resulting from P_{AOXI} (or P_{GAP} , P_{FLDI}) overwhelming the post-translational machinery of the host cell causing a high proportion of misfolded foreign protein, moderately expressing promoters (such as P_{YPT1} and P_{PEX8}) are considered as the best choices (Cereghino and Cregg, 2000). Maybe as a result of low expression levels P_{YPT1} and P_{PEX8} have not been widely used (Macauley-Patrick et al., 2005). Expression levels obtained from P_{YPT1} was reported to be lower than P_{GAP}

(Sears et al., 1998) which can be observed in Table 2.7. So, it can be a promising substitute for expressing genes that would be toxic when overexpressed; host cell tolerates the toxic effects of foreign gene expression in moderately-expressing promoters (Sung-Jae et al., 2007).

Table 2.7. Activities of the various promoters in *P. pastoris* cells grown on different carbon sources based on (intracellular) GUS (β -glucuronidase) activity (Sears et al., 1998).

Integrated plasmid	Carbon source	GUS activity
pIB2-GUS (P_{GAP})	Glucose	70.4
	Glycerol	48.8
	Methanol	11.3
	Mannitol	18.8
pIB3-GUS (P_{YPT1})	Glucose	0.63
	Glycerol	0.84
	Methanol	0.45
	Mannitol	1.67
pIB4-GUS (P_{AOX1})	Glucose	0.05
	Glycerol	0.34
	Methanol	587.6
	Mannitol	20.3

2.2.1.1.2.3. Translation elongation factor1- α gene promoter

Translation elongation factor1- α (TEF1) is one of the crucial components of the translation machinery of eukaryotes where it is responsible for the efficient transport of amino acyl transfer RNA complexes to ribosome to continue the elongation of the polypeptide chain (Vogl and Glieder, 2013). As one of the most abundant soluble proteins in eukaryotic cells, *TEF1- α* genes coding this protein were shown to have strong promoters (Ahn et al., 2007). Regarding its function, it has a growth-related expression behavior and level of the related mRNA is decreased in stationary phase (Stadlmayr et al., 2010; Vogl and Glieder, 2013).

A comparison was made between the strength of P_{TEFI} and P_{GAP} (by single-copy transformants) by expression of lipase in *P. pastoris* (Ahn et al., 2007). The results (Table 2.8) indicated that P_{TEFI} has the more constitutive characteristics when compared to P_{GAP} . Strength of P_{TEFI} in the carbon-limited semi-batch environment is higher than that of P_{GAP} and strong promoter activity produces r-proteins at levels similar to P_{GAP} . However, depending on the reporter protein, and cultivation time, the expression would be different and might be similar or lower than P_{GAP} (Vogl and Glieder, 2013).

Table 2.8. Comparison of the strength of P_{TEFI} and P_{GAP} in carbon-limited fed-batch culture (Ahn et al., 2007).

Construct	Carbon source	Cell density (A_{600})	Secreted lipase activity (U/mL)
P_{GAP} -CLLip	Glucose	393	100
	Glycerol	376	201
P_{TEFI} -CLLip	Glucose	397	226
	Glycerol	348	410

2.2.1.1.2.4. 3-phosphoglycerate kinase gene promoter

3-phosphoglycerate kinase (*PGKI*) gene, as a housekeeping gene, encodes PGK enzyme from glycolytic pathway; genes from glycolytic pathway are known to be highly expressed. *PGKI* gene falls under this category and therefore its promoter can be used in the construction of expression vectors (de Almeida et al., 2005; Stadlmayr et al., 2010) and according to its strength it represents a potential alternative to constitutive heterologous expression in *P. pastoris*. However, it is regarded as a rather weak promoter (Vogl and Glieder, 2013). It was shown that promoter is regulated moderately by the carbon source, since the mRNA levels were higher on glucose compared to

glycerol (Vogl and Glieder, 2013). The *PGKI* gene is represented by a single copy in *P.pastoris* genome (de Almeida et al., 2005).

In a conducted study, α -amylase production has been used as a probe to compare P_{PGKI} with P_{AOXI} . Figure 2.9 can be illustrative of the P_{PGKI} strength (de Almeida et al., 2005). A new expression vector (based on P_{PGKI}) has been developed for controlled constitutive expression in *P. pastoris* (Sung-Jae et al., 2007). P_{PGKI} (and also P_{TEFI}) in some cases can result in productions in the same order of P_{GAP} ; however the gene copy number should be included and also should be analyzed in each product case, since different authors have reported different results.

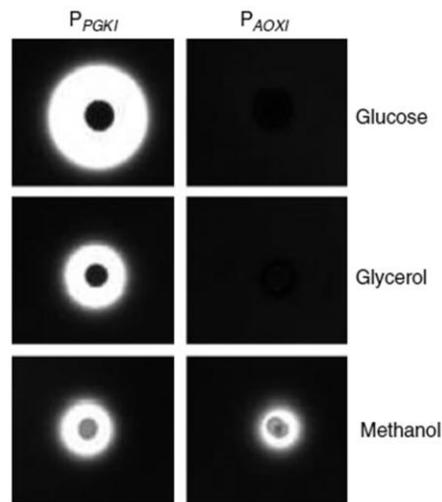


Figure 2.9. P_{PGKI} and P_{AOXI} activities in directing the expression of the *B. subtilis* α -amylase. The ability to produce starch hydrolysis halos by iodine vapor staining of starch-containing plates (de Almeida et al., 2005).

2.2.1.1.3. Novel promoters

In order to find novel promoters for r-protein expression in *P. pastoris* a research was performed and several promoters were selected based on both

microarray analysis and literature survey and were compared with reference promoters, i.e., P_{AOXI} and P_{GAP} , by production of reporter proteins (Stadlmayr et al, 2010). The thiamine responsive promoter (P_{THI11}) was shown to be a promising new alternative. The derepressible promoter can be used for the expression of target gene by the addition or absence of thiamine from the cultivation medium, independently from carbon and nitrogen sources, thereby enabling a very easy induction strategy. In addition, two types of novel inducible promoters have been introduced for r-protein production in *P. pastoris* which does not rely on the physical compounds added to the medium; oxygen-regulated promoters and heat-inducible promoters.

Promoters whom their activity can be controlled by oxygen availability appear to be interesting alternatives for heterologous protein production. Improving the production of heterologous protein by implementation of hypoxia can be so captivating from an applicative point of view, because of being cheap and easily-obtainable which leads to a straightforward induction (Camattari et al., 2007). Ethanol fermentation is illustrative of oxygen limitation in shake flask fermentations but hypoxia can be achieved by nitrogen sparging in bioreactors. In this category (alcohol dehydrogenase (*ADH2*) gene promoter (P_{ADH2}) of *P. stipitis* (Passoth and Hahn-hagerdal, 2000; Chien and Lee, 2005) and *KIPDC1* promoter (P_{KIPDC1}) of *K. lactis* (Camattari et al., 2007) both can be mentioned; the former has been checked in *P. pastoris* (Figure 2.10) and (Figure 2.11) but the latter has not been specifically used in *P. pastoris*. *KIPDC1* gene, another choice for oxygen limitation-induced expression, is the unique gene coding for pyruvate decarboxylase (PDC) in *K. lactis*. It can be induced by glucose and repressed by ethanol and autoregulation; i.e., the presence of KIPdc1p (*KIPDC1* gene product) is necessary for oxygen limitation-induced feature of P_{KIPDC} . Different reporter proteins were expressed by the aid of P_{KIPDC1} in *K. lactis*, *Zygosaccharomyces bailii*, and *S. cerevisiae*. The results showed the feasibility of usage of P_{KIPDC1} in other yeast species (Camattari et al., 2007).

The important point is that, the host must be proliferative in oxygen presence and then cease to growth in its absence and, thus, by oxygen limitation the gene starts to be expressed and the growth and production phases can be separated. This kind of induction is very easy and is just by interruption of oxygen flow to the system. Important consideration is keeping on aeration with low flow rate of oxygen (after sparging with nitrogen) in order to guarantee hypoxic conditions without any reduction in cell growth; strict anaerobic condition is not so favorable.

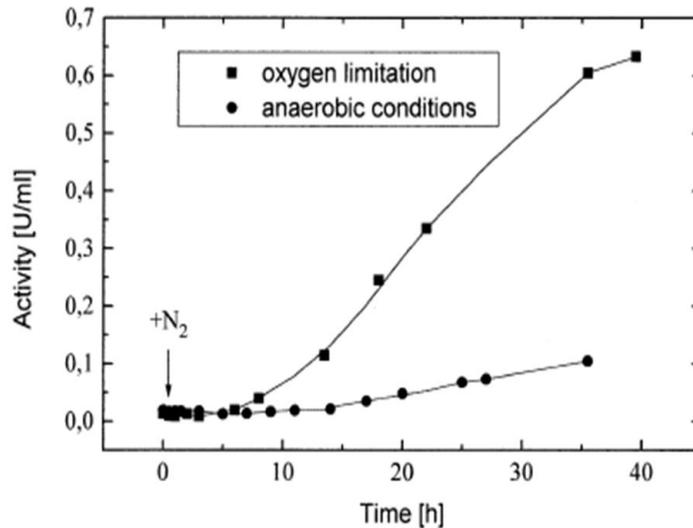


Figure 2.10. Induction of the endo-1,4- β -xylanase expression in *P. pastoris* host under the control of P_{ADH2} of *P. stipitis* (Passoth and Hahn-hagerdal, 2000). Enzyme activity is the base of judgment.

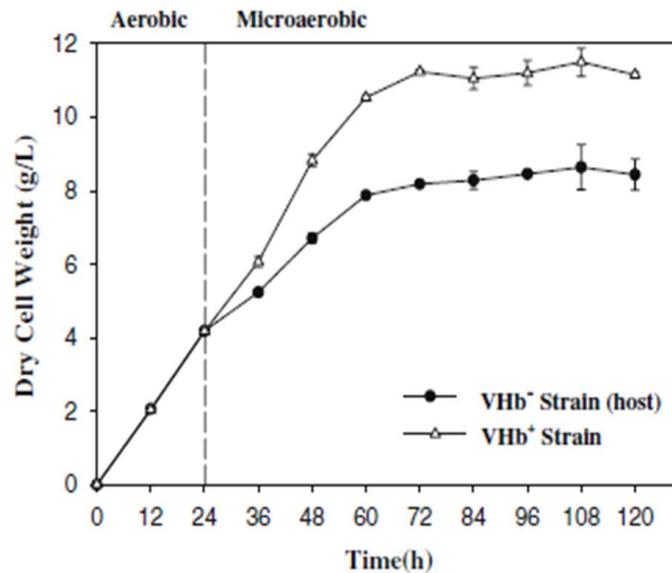


Figure 2.11. Expression of the bacterial hemoglobin (VHb) in *P. pastoris* under the control of P_{ADH2} of *P. stipitis*. Growth curve in microaerobic conditions for transformed and wild type strains. As VHb lets to survive and grow in O_2 -poor environment, thus, the amount of VHb-harboring cells increase in the microaerobic condition by the help of newly-introduced foreign gene (Chien and Lee, 2005).

The heat shock proteins (HSPs) are families of proteins that have role in protein folding and unfolding and their expression is elevated in increased temperatures or other stresses. They are also present in non-stressful conditions and just monitor the proteins of the cell and for example help old proteins to go to proteasomes which represents their housekeeping responsibilities as a part of cell stress response or "the heat-shock response". Heat-inducible promoters are the other new candidates, where the promoters of heat shock proteins are recruited. These promoters are induced just by shift in temperature. The heat shock promoters have proven to be among the strongest promoters ever tested, and temperature is the only requirement for their induced expression. Recently, a heat-inducible promoter drove the gene expression in *Tetrahymena thermophila* as expression host (Yu et al.,

2012). Usage of heat-inducible promoters introduces some difficulties such as the method of temperature shift and its duration. Furthermore, the heat may be benign to the foreign product; thus, again the nature of the product will be conducive. An interesting point is the P_{PGKI} which possesses a heat shock element that may lead to its usage as a component of this division.

In recent years, six novel regulated promoters have been characterized from *P. pastoris* cultures for glucose-based protein expression in order to provide methanol-free expression (Prielhofer et al., 2013). The strongest inducible promoter was found to be for the gene of high-affinity glucose transporter named *GTH1* gene. This promoter toolbox allows avoiding usage of inducer which in some cases can exert additional costs to the process.

Briefly, after selection of the product it will be easier to decide on the promoter. Product determines the type of the promoter as being inducible or constitutive; the product price in the market and its usage area defines the used materials in the medium for induction or regulation of the promoter. The product determines whether the combined use of promoters is beneficial or not. Selecting the best choice in each category needs screening of different promoters available since the contrary results have been reported in different cases about the strength of the promoters and, thus, the selection is case-dependent.

In addition to the available set of promoters for each r-protein expression system, artificially-designed promoters can be very beneficial. Promoter engineering can provide us with the opportunity of fine-tuning of the expression of heterologous genes in hosts and, thus, obtaining higher titers of desired r-protein (Vogl et al., 2014). Random mutagenesis as the common approach for manipulation of the selected yeast core promoter region, alteration of the upstream regulatory sequences, and combination of the cis-acting elements of a promoter with another natural promoters have been the prevalent methods in order to obtain engineered promoters in lower eukaryotes such as yeast (Qin et al., 2011a; Qin et al., 2011b; Vogl et al.,

2014). Afterwards, analyses of the function/sequence relationships can reveal the strength of each synthetic promoter. Recently, a synthetic core promoter has been designed for *P. pastoris* by aligning four core promoters and refining the sequence with insertion of common transcription factors binding sites (TFBS); this fully artificial sequence was used, subsequently, for engineering of P_{AOXI} in order to obtain a promoter library with different characteristics (Vogl et al., 2014).

2.2.1.2. Secretion signal peptides

As the second genetic factor, secretion signal peptides which guide the expressed proteins into the secretory pathway and facilitate their secretion to the extracellular medium are considered.

2.2.1.2.1. Importance of secretion

Any heterologous protein production places a substantial "metabolic burden" on the host organism influencing metabolism which ends with halted growth rate and, thus, affected gene expression. The effect can also be in the form of direct toxicity, hampering cell's normal working plan, in the case of intracellular expression (Boettner et al., 2002); contamination of target protein with endogenous compounds could also be another probable outcome. Intracellular location of the expressed protein also makes the cell lysis and cell debris removal steps of downstream purification inevitable which increases the capital investment of the whole process. Furthermore, entering the secretory pathway makes the final r-product further similar to its original eukaryote counterpart. Finally, the desired host, *P. pastoris*, secretes low levels of endogenous proteins and leads to an easily-handled separation (Cereghino and Cregg, 2000). As a consequence, researches aim finding the improved methods of directing the foreign protein to the outside of the *P. pastoris* cell, have occupied the spotlight recently.

The crucial requirement for extracellular expression of the r-protein is an efficient signal peptide, i.e. secretion signal peptide, which is

improved in the plasmid structure in the in-frame fusion form by the N-terminus of the desired gene. This expressed protein will be allowed to enter the endoplasmic reticulum (ER) to travel within until final destination. It worth mentioning that the choice of the secretory mode of expression is efficient and rational for proteins that are secreted in their native hosts, as the passage through the ER and Golgi can highly alter the structure and, consequently, the function of the proteins that are not secreted normally (Cereghino and Cregg, 2000). However, it has been reported that proteins which are not secreted naturally, have been secreted successfully; it can be problematic if there are possible glycosylation sites (Romanos et al., 1992). Additionally, the choice of an efficient SP for a product is rather arbitrary and a matter of trial and error which depends on the protein itself and the host (Sreekrishna et al., 1997; Damasceno et al., 2012); therefore, access to a repertoire of SPs will provide a better opportunity to try many choices and choosing the best option based on the secretion efficiency and protein authenticity.

2.2.1.2.2. Signal peptides

Generally it is possible to specify the localization of the expressed proteins by interpreting their amino acid sequences (Nakai 2000). Signal peptides are specific amino acid sequences that determine whether a protein will pass through a membrane into a particular organelle, become a transmembrane protein (TM), or be secreted to the extracellular medium (Chou, 2002). Although signal peptides usually locate in the N-terminal of the proteins for most secreted proteins and transmembrane (TM) proteins, they can also be detected within or at the C-terminal of the protein (Chou, 2002), such as nuclear localization signal peptides and peroxisomal targeting signals which can be found in both N-terminal and C-terminal (Nakai, 2000). N-terminal signal peptides lead to the entrance of virtually all secretory proteins to the secretory pathway (Chou, 2002). In a few exceptions there are proteins transported by other means than conventional ER-Golgi assembly so a signal

peptide can not be identified (Huang et al., 2011). In eukaryotes, the climax of the protein expression procedure is the entrance of the translated (or being translated) protein into the secretory pathway (starting from ER) in order to being further processed and folded to the final biologically active conformation and, then, being secreted. The ER is the hub of the protein secretion and posttranslational processing system which also entails the collaboration of Golgi apparatus. Entering to ER means residing in ER or being dispatched by vesicular transport system to: (i) different organelles, (ii) cell membrane, or (iii) extracellular medium; proteins that are transported to the nucleus or remain in cytosol are produced by free ribosomes and do not enter ER (Karp, 2008).

Briefly, N-terminal signal peptides are 20- to 30-amino acid sequences which are responsible for directing the secretory protein toward ER and mediating its passage through ER membrane. These signal peptides are referred to as secretion signal peptides (SPs) throughout current manuscript. SPs are of two types: “pre” type and “pre-pro” type; in contrast to the “pre” sequence cleaved while passing through the ER membrane, “pro” sequence is proteolytically cleaved from the intermediate (pro-protein) in the membrane of the trans Golgi or in the secretory vesicles (Chaudhuri et al. 1992). It has been found that in some cases “pro” sequence in SP can act as a molecular chaperone which aids protein folding (Chaudhuri et al., 1992; Ide et al., 2007). Also it is known to take part in the transportation of immature protein (Ide et al., 2007). As “pro” sequence remains after cleavage of “pre” sequence, it may undergo PTMs such as glycosylation as the case in α -MF, and quality control which both can affect protein secretion efficiency (Oka et al., 1999). The SPs should not be regarded as removable elements that their function terminates after cleavage, but instead some roles have been discovered for them after cleavage (Hegde and Bernstein, 2006).

2.2.1.2.3. Features of the secretion signal peptides

Although there is a common basic structure of SPs between bacteria and eukaryotes, there has been discovered no concrete consensus sequence among the pre-sequence of SPs, whether pre-type or prepro-type. However, a three-domain structure can be recognized: N-region, H-region, and C-region (Martoglio and Dobberstein, 1998; Nakai, 2000) where each can be assigned a specific feature (Figure 2.12). These three regions are conserved among the SPs derived from wide variety of organisms (Nagarajan, 1993). The N-region locates at the N-terminal of the SP and is a basic segment often with positively-charged polar residues. The H-region, hydrophobic domain, generally has a length of 7-15 amino acids. It has not been interrupted by charged residues with distinguished preference for leucines or alanines (Stroud and Walter, 1999; Hegde and Bernstein, 2006). This domain is the most essential region required for targeting and insertion into the ER membrane (Martoglio and Dobberstein, 1998). H-region spans the ER membrane and facilitates the translocation (Nagarajan, 1993). Providing that the hydrophobic core region retained, any change in amino acid sequence of SP can be tolerated (Stroud and Walter, 1999). The end of the H-region can be predicted by the presence of residues that favor the β -turn. The C-region in the C-terminal of the SP (binds to the N-terminal of the mature protein) is neutral but slightly polar and serves as a recognition site for cleavage of SP (Stroud and Walter 1999; Kjarulff and Jensen, 2005). The SP directs the molecule to the ER and then is cleaved meanwhile the protein molecule is translocated into the ER or shortly after completion of the translocation (Kjarulff and Jensen, 2005); there can be detected a "weak consensus pattern" to determine this cleavage site (Nakai, 2000). Potential SPs and their corresponding cleavage sites can be predicted by in-silico analyses with programs such as: SignalP, Phobius, and WolfPsort. Consequently, prediction of the SPs and their corresponding cleavage sites is considered as one of the crucial steps in directing r-proteins toward extracellular medium. Novel SPs are extracted from different secretome results by the help of in-

silico analyses and, subsequently, can be recruited in combination with different r-proteins in experiments in order to elucidate their applicability and efficiency to give credit to predictor softwares.

2.2.1.2.4. Mechanisms of secretion in eukaryotes

Upon emergence of the SP from the ribosome in cytoplasm during translation, the process of the targeting of the secretory protein is initiated. The basic mechanism for secretion of proteins across cell membrane appears to be universal containing many similar features between bacteria and eukaryotes (Nagarajan, 1993). Two common pathways are available for targeting in eukaryotes (Zimmermann et al., 2011): co-translational signal recognition particle (SRP)-dependent and post-translational SRP-independent. In the former, the SP is recognized by a signal recognition particle (SRP) while protein is being translated on the ribosome. In the latter, the SP is recognized by a protein complex on the ER membrane. However, since these two parallel pathways entail collaboration of Sec protein (as translocon) by converging at Sec61p translocon both of them are considered Sec-dependent. It should be reminded that some proteins require both routes (Figure 2.13).

In co-translational translocation, ribosome recognizes N-terminal amino acid of nascent polypeptide and recruits the SRP. SRP is a RNA-multiprotein complex (Buske et al., 2009) thought to prevent folding of newly (being) translated polypeptide since the folding of emerging polypeptide chain can pose an obstacle for translocation through ER membrane (Hegde and Bernstein, 2006). Upon binding SRP to SP, ribosome slows down the translation (Stroud and Walter, 1999; Buske et al., 2009). By interaction with SRP-receptor on ER membrane, docking protein (DP), the complex of ribosome, nascent polypeptide, and SRP is brought toward ER and specifically to the translocon complex, Sec61p, which provides a pore through the ER membrane; this happens via interaction between "M-site" on ribosome and Sec61p translocon complex (Stroud and Walter, 1999). Then,

SRP dissociates and DP slides away on the membrane. As the nascent polypeptide places at the entrance of the pore, translation elongation causes the protein to enter the ER.

In contrast, in post-translational translocation, the Sec62p-Sec63p complex in ER membrane performs the recognition of the SP. Since folding correctly or misfolding will prevent the recognition of SP, prevention of protein folding in the cytosol by the aid of cytosolic chaperones will be of supreme importance in this pathway (Martoglio and Dobberstein, 1998; Damasceno et al., 2012). Furthermore, it appears that pro-sequence could conduce to delay in folding and aid translocation. There is evidence that some assistant proteins in yeast play role in translocation probably by relaxation of any form of tertiary structure (Chaudhuri et al. 1992).

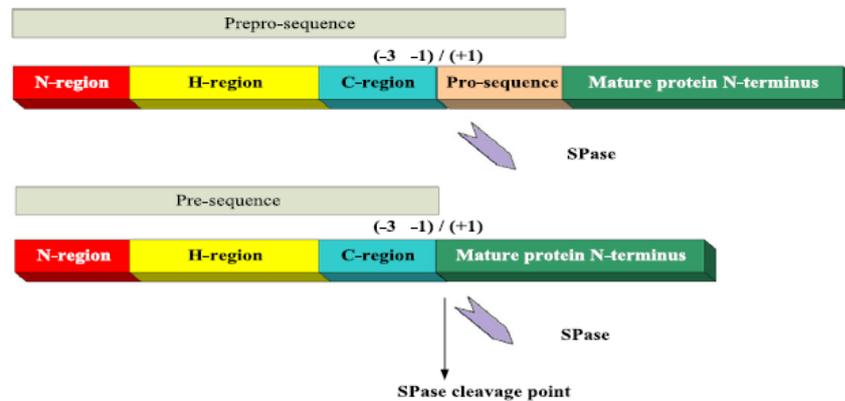


Figure 2.12. Tripartite structure of a secretion SP. Central hydrophobic H-region (yellow) and hydrophilic N- (red) and C-terminal (blue) flanking regions. SPs can be either “pre” or “pre-pro” type. ‘SPase’ is the signal peptidase type I resides in ER membrane eukaryotes (Massahi and Çalık, 2015).

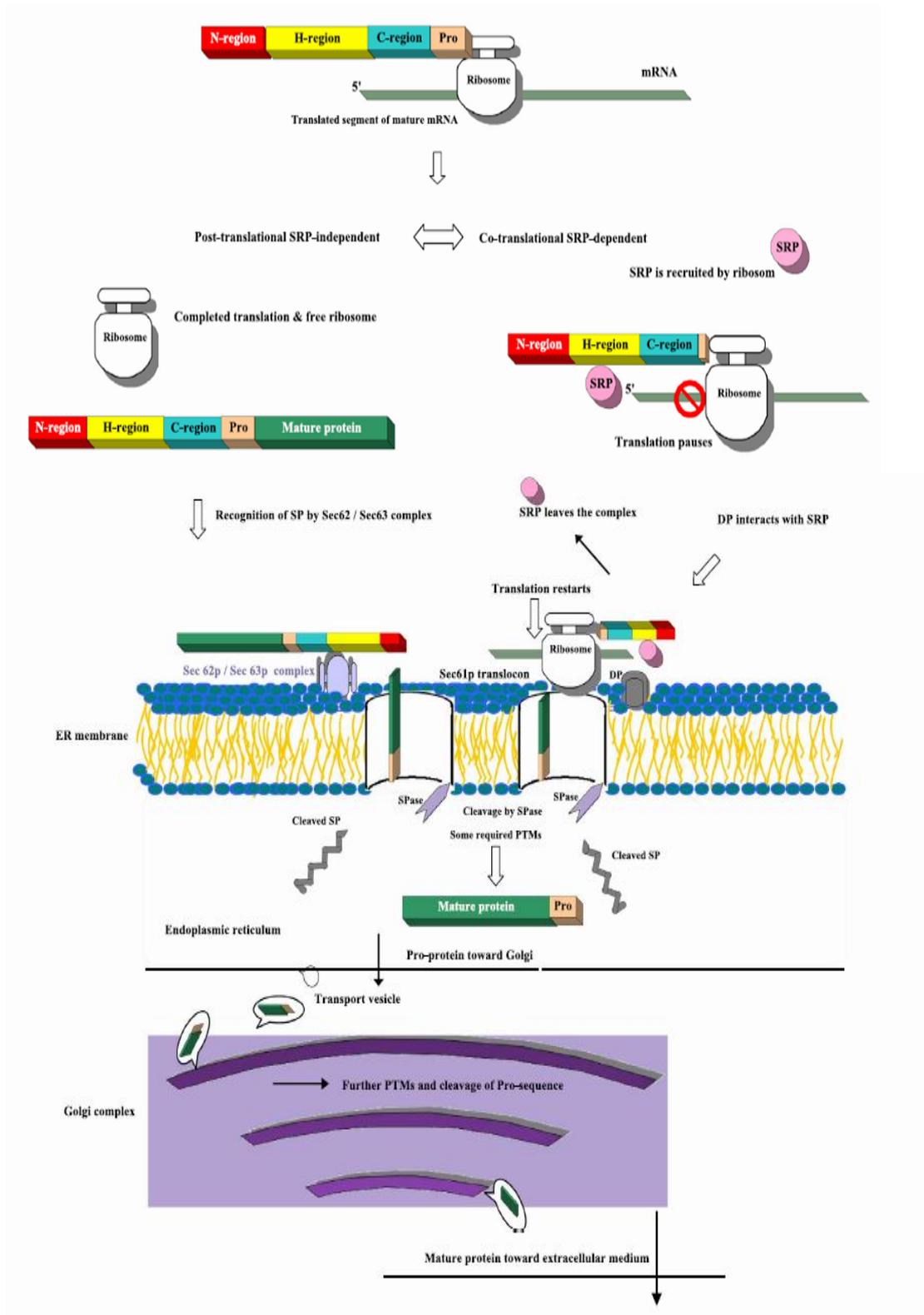


Figure 2.13. Two proposed mechanisms for protein secretion in eukaryotes (Massahi and Çalık, 2015).

The selection between abovementioned two translocation pathways is dictated by the features of the SP (Hegde and Bernstein, 2006) and mainly depends on the hydrophobicity of the hydrophobic core, H-region, of the SP (Ng et al., 1996); this factor plays more important role than the length of the H-region in recognition of the SP by SRP (Hatsuzawa et al., 1997). SP discrimination by SRP is available in microorganisms such as yeasts and bacteria and SRP binds to the SP primarily by its flexible hydrophobic region (located in SRP 54 kDa subunit) which has the ability to interact with a wide range of hydrophobic SPs. Yeast (and bacterial) SRPs bind preferentially to highly hydrophobic SPs (Hegde and Bernstein, 2006). It seems that this is the H-region of SPs which discriminates between SRP-dependent and SRP-independent pathways; SPs with highly hydrophobic H-region direct proteins toward the SRP-dependent pathway (Martoglio and Dobberstein 1998). Altering the hydrophobicity of H-region can change the pathway of the translocation (Hegde and Bernstein, 2006). However, hydrophobicity is not the only requirement for SRP-dependent pathway (Stroud and Walter, 1999). Furthermore, improvement of the secretion efficiency by increasing the hydrophobicity of SP has been reported (Xue et al., 1997). Nevertheless, it has been mentioned that N-terminus or main body of the protein can also be important (Nagarajan, 1993, Andrews et al., 1988; Ng et al., 1996; Matoba and Ogrydziak, 1998) especially presence of proline residues in positions near to the cleavage site may be decisive since they affect the positioning of the SP in the space and, in turn, influence the interaction with ER translocation machinery, specifically SRP. It is plausible that the net charge of the N-region can also play role in selection of the pathway, as it affects translocation efficiency while passing the ER membrane (Nakai, 2000). The efficiency of SP cleavage may be influenced by the N-terminal amino acids of the mature protein; substitution of especial residues that favor β -turn can affect the processing of SP (Nagarajan, 1993).

During the passage of polypeptide through ER membrane in both pathways, according to the composition of the hydrophobic H-region and

availability of a cleavage motif in the SP, the SP becomes a signal anchor in the case of TMs or is cleaved by the action of a signal peptidase enzyme (SPase). SPase is an integral membrane enzyme associates with the translocon (Stroud and Walter, 1999; Buske et al., 2009). SPases are of two types: class I and class II. SPases I that include ER-membrane SPase, cleave the usual SPs and SPases II cleave the SPs belong to lipoproteins (Nakai 2000). It worth mentioning that cleavage of SP is not an imperative prerequisite for protein secretion (Monod et al., 1989). By assuming cleavage between -1/+1 amino acid residues (Figure 2.11a), (-1) rule of cleavage can be inferred from the statistical analysis; the detachment of SP depends mainly on the amino acids at (-3) and (-1) positions. Presence of a short-side chain amino acid at (-1) position, presence of small or hydrophobic amino acid at (-3) position, and absence of a charged or aromatic amino acid at (-3) position will make the cleavage possible. In addition, the cleavage site should be in a 4- to 10- amino acid distance from the end of H-region. Generally Ala-X-Ala sequence can be recognized before cleavage point (Monod et al., 1989). However, it is difficult to forecast a cleavage event just by presence of a simple (-3, -1) pattern and it is not sufficient solely (Monod et al., 1989; Nakai, 2000). Correct processing of the SP is of great importance in protein activity (Kato et al., 2001) and can be tested by verifying the authenticity of the N-terminal of the protein expressed.

It should be considered that it is the structure of SPase that leads to the specificity for (-3) and (-1) residues (Stroud and Walter, 1999). The cleavage site must be sterically accessible for SPase (Monod et al., 1989) and any change in SP sequence which brings about steric hindrance would impair cleavage. Reaching cleavage site and recognizing it by SPase is dependent on the length of the SP and specifically H-region (Monod et al., 1989). Increase in H-region length can decrease the probability of cleavage (Martoglio and Dobberstein, 1998).

Due to the presence of a number of stages in the secretion procedure, in each step a problem can be faced. The yeast proteins which play role in

folding and disulphide bond formation differ from their counterparts in higher eukaryotes and this may affect folding of foreign proteins. There are also reports of proteins being retained in the Golgi, again possibly due to improper folding. Retention in the cell wall has also been a problem, especially with larger proteins, although factors other than molecular mass are known to be important (Romanos et al., 1992).

As the efficiently targeted, translocated, or cleaved SPs can conduce to higher amounts of the secreted proteins, by engineering the SPs it will be possible to control the amount of the secreted r-protein. In addition, SPs can be tailored based on specific requirements in the production of desired substrate attached with in terms of timing of cleavage, time span of association with the translocon which in turn leads to folding in a specific manner or specific modification required (Hegde and Bernstein, 2006). The effect of point mutations in last amino acid residue on the processing and secretion of yeast *PHO5* gene protein has been shown (Monod et al., 1989); deletion of a 4-amino acid region (-7 to -4 relative to the cleavage site) at the end of the SP also led to intracellular accumulation of the protein. SP optimization can be another option to enhance the yield of the r-protein production as was applied in *E. coli* (Klatt and Konthur, 2012).

2.2.1.2.5. Secretion signal peptides in *Pichia*-based systems

In secretory expression by *S. cerevisiae*, two approaches have been used: utilizing endogenous SPs and utilizing exogenous, non-*Saccharomyces*-specific SPs. It seems that homologous (with yeast origin) SPs obtained from naturally secreted proteins will produce better outcome than heterologous (with non-yeast origin) SPs (Chung et al., 1996; Zuyong et al., 2012) by regarding the specific features of the secretion pathway in each organism. However, by considering that *S. cerevisiae* has a low specificity requirement in secretory pathway and, thus, flexibly recognizes unspecific SPs (similarly may *P. pastoris*), many proteins have been expressed

successfully in *P. pastoris* by their native SPs (Daly and Hearn, 2005; Korona et al., 2006).

Overall, in *Pichia*-based systems, various SPs have been recruited until now, and they can be classified into two groups: exogenous SPs, endogenous SPs. Exogenous group comprises products own SPs and SPs from other yeast species, fungi, higher eukaryotes while endogenous SPs belong to the *P. pastoris* own secretome proteins. Products own SPs are case-dependent and can be exemplified by: *Bacillus licheniformis* α -amylase (Paifer et al., 1994), *Phaseolus vulgaris* agglutinin E-form (Raemaekers et al., 1999), *Bacillus Stearotherophilus* D-alanine carboxypeptidase, *Aspergillus awamori* glucoamylase, *S. cerevisiae* invertase, barley α -amylase, honey bee odorant-binding protein (ASP2), mouse major urinary protein complex (MUP), bile salt-stimulated lipase, gastric cathepsin E, human serum albumin (Cereghino and Cregg 2000), human midkine (Murasugi et al., 2001), *Aspergillus niger* exoinulinase (Moriyama et al., 2003), *A. niger* xylanase (Korona et al. 2006), thaumatin (Ide et al., 2007), interferon- α 2b (Ghosalkar et al., 2008), and *Rhizopus oryzae* α -amylase (Li et al., 2011).

S. cerevisiae α -MF has been the most widely used exogenous SP in *P. pastoris* and in some cases even has led to better secretion than protein's own SP (Macauley-Patrick et al., 2005; Ide et al., 2007). Mutated α -MF lacking Glu-Ala repeats, with only Kex2 cleavage site, enhanced the secretion level with valid N-terminus. The Glu-Ala repeats help to improve the protein processing by facilitating Kex2 cleavage. It has been shown that mutations in α -MF can alter the efficiency of the secretion where the deletion of amino acids 57-70 led to increase of protein secretion (Lin-Cereghino et al., 2013). The surrounding amino acids in α -MF can influence the cleavage of the "pro" sequence by Kex2 and Ste13 enzymes which cleave "pro" sequence and Glu-Ala residues, respectively. Furthermore, the tertiary structure of the protein can make the cleavage site inaccessible for respective protease enzymes (Cereghino and Cregg, 2000). The effectiveness of the α -

MF cleavage is reported to be very case-dependent and inefficient protein processing has been reported in both *Saccharomyces* and *Pichia* (Raemaekers et al., 1999). Therefore, α -MF may not be suitable for all proteins and if the authenticity of N-terminus of the protein is inevitable, other SP choices should be tested empirically.

The other common SPs are *P. pastoris* acid phosphatase (PHO1) SP and *S. cerevisiae* invertase (SUC2) SP. The available commercial vectors for expression in *P. pastoris* possess α -MF and PHO1 secretion signal peptides (Damasceno et al., 2012).

Even though available set of SPs for *Pichia* platform have conducted successful secretion of the heterologous proteins, searching for more efficient and more host-compatible SPs can lead to more efficient r-protein secretion. The endogenous SPs of the host microorganism, obtained from discovery of its secretome, are regarded as the first-line candidates recognized by the secretion machinery of the corresponding host (Mori et al., 2015). In consequence, recruitment of the endogenous SPs of *P. pastoris* for r-protein production seems to be a promising alternative for classic and previously utilized SPs. This necessitates having a secretome data for *P. pastoris*; secretome of the *P. pastoris* DSMZ 70382 strain has been predicted and validated using the microorganism grown in glucose-based medium (Mattanovich et al., 2009). Also a detailed profile of the host-secreted proteins has been provided for methanol-induced *P. pastoris* X-33 strain (Huang et al., 2011). By utilization of available secretome data and in-silico analyses for prediction of SPs in order to find their amino acid sequence, novel SPs can be found and recruited for r-protein production in *Pichia*-based systems. Until now, five endogenous SPs of *P. pastoris* have been utilized successfully in *Pichia* system for r-protein production and can be regarded as positive control in related in-silico analyses (Murasugi and Tohma-Aiba, 2001; Yoshimasu et al., 2002; Khasa et al., 2011; Liang et al., 2013).

2.2.1.3. Codon optimization

In heterologous protein expression, along with genetic factors such as signal peptide, promoter strength, AT-rich regions, and GC content, codon optimization has been considered as one of the useful strategies for efficiency improvement.

A raised problem in r-protein expression is the biased usage of codons, or biased codon usage (Lee et al., 2010); this synonymous codon usage bias can be a translational barrier which can decrease expression, although, high mRNA levels has been detected (Sinclair and Choy, 2002). Whole living systems have been dissected hitherto possess a general non-random preference toward a group of the 61 available sense codons, i.e., non-random usage of alternative synonymous codons, throughout a specific genome and this tendency is in correlation with the tRNA pool present in the cell. Based on a proposed model, so-called “major codon preference”, the major deciding factor for imposing such kind of preference is translational efficiency (Sinclair and Choy, 2002). During translation elongation, the aminoacyl-tRNA complex is brought to the ribosome based on the corresponding codon on mRNA; the limiting parameter for translation proceeding is the rate at which the aminoacyl-tRNA is made available on ribosome, so, the waiting time is proportional to the available aminoacyl-tRNA in the cell. Consequently, the protein translation rate can be increased by using the codons that match most common tRNAs in the host. Thus, the difference between the codon usage in host and gene of interest can greatly influence the expression level (Jia et al., 2012) and optimization of the codons of the gene of interest according to the codon bias of the host can greatly enhance expression efficiency. The full optimization of coding regions of gene of interest toward the bias of the host organism has lead to an increase about 10- to 50-fold for foreign protein production in several different host cells for different proteins (Sinclair and Choy, 2002).

The codon usage bias of *P. pastoris* from Kazusa database is represented in Table 2.9. The natural tendency of *P. pastoris* is toward the

A/T-ended codons instead of G/C-ended codons in mammalian genes (Hu et al., 2006), so removing rare codons in desired genes and replacing by frequently occurring ones (Outchkourov et al., 2002) will lead to lower GC content and, thus, higher AT content which renders the secondary structure of mRNA less stable; subsequently affects the initiation of translation (Hu et al., 2006); the alteration of the GC content can be considered as a by-product of the changing of the codons (Sinclair and Choy, 2002). In contrast, it should be reminded that high content of A+T leads to premature transcription termination; so, in order to overcome this problem, the redesign strategy would be to limit the A+T content in a specific range. To date, numerous efforts have been made in order to improve expression of r-proteins in *P. pastoris* with codon optimization which some examples can be mentioned: *Chondrus crispus* hexose oxidase (Wolff et al., 2001), hyperthermostable *Thermotoga maritima* xylanase (Jia et al., 2012), barley β -D-glucan exohydrolase HvExoI (Luang et al., 2010), human Zbtb7A (Wang et al., 2008), *Penicillium notatum* glucose oxidase (Gao et al., 2012).

In order to quantify the degree of bias in codon usage in each gene, many efforts have been made and different approaches and measures have been proposed. Codon adaptation index (CAI) is a simple and effective measure of synonymous codon usage bias. The index uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and, then, a score is calculated for a gene from the frequency of use of all codons in that gene (Sharp and Li, 1987). Constructing “relative synonymous codon usage” (RSCU) table from a set of highly expressed genes of the host organism as a reference table will be the first step to determine CAI; RSCU is calculated for each codon regarding specific amino acid. CAI for a gene is calculated as the geometric mean of RSCU values (from constructed reference table) corresponding to each of the codons used in the gene of interest divided by the maximum possible CAI for a gene of the same amino acid composition (Sharp and Li, 1987). This index will be useful in prediction of the probable success of expression of a foreign gene

in a host. Highly adapted genes with codon choice identical those preferred by the organism in question will yield CAI value of 1. While genes with lower overall magnitudes of bias, or bias towards a different set of preferred codons will yield values approaching zero. Therefore, in r-protein expression it will be beneficial to make CAI close to unity such as expression of *Thermus thermophilus* glucose isomerase (Ata et al., 2014).

2.2.1.4. Gene copy number (Gene dosage)

In order to improve r-protein expression, standard steps are codon optimization and increasing gene copy number (Inan et al., 2006). The heterologous protein yield is directly proportional to mRNA transcript which is further dependent on the inserted gene copy number on the vector plasmid. Multicopy integrants have the potential to express substantially higher foreign protein (Athmaram et al., 2012). Furthermore, in the case of probable mutation of one gene, its undesired product proportion in total protein fades by contribution of genuine genes action (Daly and Hearn, 2005). It should be emphasized that multicopy strains of host can exist naturally at a level of a few percent (1-10%) among transformants with single-copy-number vector (Cregg, 1999; Daly and Hearn, 2005).

Higher copy number will lead to overexpression of the protein which will not only exert more metabolic burden on the host but also result in problems such as saturation of secretory pathway that can lead to misfolded proteins. Therefore, there will be an optimum instead maximum point for gene dosage which is product- and expression construct-dependent (Hohenblum et al., 2004) where the increase in copy number will not further increase yield as a result of firstly, overwhelming the host secretory machinery capacity to process and fold the protein in authentic format and secondly, lack of required precursors and also energy (Macauley-Patrick et al., 2005); it is less probable to be because of the latter, as the expression level is low to medium in yeast expression system (Zhu et al., 2011).

Table 2.9. *Pichia pastoris* codon usage bias (<http://www.kazusa.or.jp>)

Triplet	Frequency (per thousand)	number (out of 81301)	Triplet	Frequency (per thousand)	number (out of 81301)	Triplet	Frequency (per thousand)	number (out of 81301)	Triplet	Frequency (per thousand)	number (out of 81301)
UUU	24.1	(1963)	UCU	24.4	(1983)	UAU	16.0	(1300)	UGU	7.7	(626)
UUC	20.6	(1675)	UCC	16.5	(1344)	UAC	18.1	(1473)	UGC	4.4	(356)
UUA	15.6	(1265)	UCA	15.2	(1234)	UAA	0.8	(69)	UGA	0.3	(27)
UUG	31.5	(2562)	UCG	7.4	(598)	UAG	0.5	(40)	UGG	10.3	(834)
CUU	15.9	(1289)	CCU	15.8	(1282)	CAU	11.8	(960)	CGU	6.9	(564)
CUC	7.6	(620)	CCC	6.8	(553)	CAC	9.1	(737)	CGC	2.2	(175)
CUA	10.7	(873)	CCA	18.9	(1540)	CAA	25.4	(2069)	CGA	4.2	(340)
CUG	14.9	(1215)	CCG	3.9	(320)	CAG	16.3	(1323)	CGG	1.9	(158)
AUU	31.1	(2532)	ACU	22.4	(1820)	AUU	25.1	(2038)	AGU	12.5	(1020)
AUC	19.4	(1580)	ACC	14.5	(1175)	AAC	26.7	(2168)	AGC	7.6	(621)
AUA	11.1	(906)	ACA	13.8	(1118)	AAA	29.9	(2433)	AGA	20.1	(1634)
AUG	18.7	(1517)	ACG	6.0	(491)	AAG	33.8	(2748)	AGG	6.6	(539)
GUU	26.9	(2188)	GCU	28.9	(2351)	GAU	35.7	(2899)	GGU	25.5	(2075)
GUC	14.9	(1210)	GCC	16.6	(1348)	GAC	25.9	(2103)	GGC	8.1	(655)
GUA	9.9	(804)	GCA	15.1	(1228)	GAA	37.4	(3043)	GGA	19.1	(1550)
GUG	12.3	(998)	GCG	3.9	(314)	GAG	29.0	(2360)	GGG	5.8	(468)

As increased gene dosage leads to increased flux of foreign protein to ER, chaperones such as PDI, BiP within the ER are depleted which, subsequently, leads to accumulation of improperly folded peptide in ER referred as protein folding stress or conformational stress (Zhu et al., 2011).

Totally, by high-copy transformations physiological effects are brought about. Although, generally the increase of gene copy number has led to increase in protein expression (Cos et al., 2005); but, there has also been conflicting results in assaying the effect of increasing copy number on yield and, thus, the result would remain unpredictable (Sreekrishna et al., 1997). It worth mentioning that gene copy number as an optimization strategy should be chosen by considering the identity (the strength) of the promoter (Macauley-Patrick et al., 2005); imperfection of weak and moderate promoters can be compensated by increasing the gene copy number.

2.2.2. Process parameters

All microorganisms have a set of biochemical reactions (anabolic and catabolic) which obey the general rules of reactions thermodynamics and, based on the chemical kinetic rules, have dependency on temperature, pH, and substrate concentration originating from the effect of temperature and pH on the activity of the corresponding enzymes. Unicellular organisms require specific internal conditions for optimal growth and function; however abrupt variations in the interacting environment can perturb the internal milieu, disrupting normal processes (Gasch, 2003). Cell, itself, controls all intracellular reaction network and reacts to the environmental changes via its signal transduction pathway and, therefore, adopts a suitable new internal state. In consequence, controlling the critical external parameters which influence cells biochemical reaction network is so crucial and determination of optimum condition or strategy developed for each factor increases the stability of the production. Thus, production of r-protein in controlled environment of a bioreactor compared to shake bioreactor brings higher yields. However, high-yield expression of heterologous protein is usually a

matter of "trial and error" (Boettner et al., 2007). As the normal functions of a cell directs the production of metabolites in required levels for cell, in order to reach economically acceptable levels of a product it may be necessary to increase the desired product production by a metabolic stress aiming abuse the cellular system.

One mechanism which is common in responding to the environmental perturbations is to initiate a common gene expression program that generally protects the cell during stressful times and conduces to the protection of internal system from fluctuations exerted by environment. This program is referred to as the environmental stress response (ESR) (Gasch, 2003) which has been observed as a genome-wide transcriptional change in *S. cerevisiae* (Mattanovich et al., 2004). It is of supreme importance to analyze the effects brought about by process variables (Mattanovich et al., 2004), ex., pH, temperature, osmolarity, and oxygenation on the physiology of the recombinant host by the combined study of the proteome, transcriptome and metabolome (Dragosits et al., 2009).

2.2.2.1. Medium composition (Nutrient consideration)

P. pastoris, like other yeasts, requires sources of carbon and nitrogen for its growth. Uptake of the suitable carbon and nitrogen source in order to continue the routine biological processes is a vital part of the organisms living program. The most prevalent carbon sources are glucose and glycerol and nitrogen sources are peptone, yeast extract, and yeast nitrogen base (Li et al., 2007). Glycerol is the preferred substrate for the *P. pastoris* cultivation (Ghosalkar et al., 2008). By affecting cell growth and viability, medium composition influences r-protein expression (Li et al., 2007) and, therefore, providing the correct chemical environment should be taken into account in designing a medium. Furthermore, specific needs or features of a process or some beneficial characteristics of ingredients may impose utilization of specific compounds. Based on the promoter type the composition of the

medium is variable; in the case of P_{AOXI} , a three-stage semi-batch strategy is very common (Inan and Meagher, 2001a, 2001b; Ghosalkar et al., 2008); usage of a repressing and non-fermentable carbon source (generally glycerol) in a defined medium in growth phase (glycerol batch), utilization of glycerol in transition phase (glycerol semi-batch) for further increase of the biomass and to prepare (derepress) the cells for induction (Cereghino et al., 2002), and addition of methanol in methanol semi-batch which acclimates the cells to methanol and starts expression (Cregg, 1999). In the case of r-protein production under P_{GAP} control the biomass production is carried out with glycerol as carbon source in batch phase and glucose is added in the fed-batch phase in order to produce r-protein. Although, there is also production in batch phase in presence of glycerol.

The most widely used medium for high cell density cultivation of *P. pastoris* is the BSM (basal salt medium) proposed by Invitrogen Corporation; the standard but not the optimum one for production of every protein (Ghosalkar et al., 2008). This medium is used in batch phase of *Pichia*-based systems for increase of biomass. BMGY (buffered minimal glycerol complex medium) is used generally for extracellular expression which simplifies pH control, decrease protease activity, and increase biomass amount. One of the imperfections of BSM media is the usage of ammonia as both pH controller and nitrogen source which may lead to starvation (Macauley-Patrick et al., 2005); thus, an optimized (chemically-defined) medium was provided (for maximum biomass production) with ammonium sulfate as nitrogen source in order to conquer nitrogen starvation (Ghosalkar et al., 2008). The constituents of above mentioned media used in *P. pastoris* expression systems (glycerol, methanol, biotin, salts, and trace elements) all are economical and well-defined and ideal for large-scale production.

As mainly complex medium is used for growth and induction (utilization of yeast extracts and peptones), batch-to-batch variations would be expected and, thus, elimination of any complex component will be desirable in order to standardize the production process (Macauley-Patrick et

al., 2005). In contrast, usage of defined media will increase the possibility of nutrient starvation which would result in "autophagic cell degradation" and lysis which will lead to release of vacuolar proteases (Potvin et al., 2012); furthermore, complex or (amino acid-) enriched media may provide competing substrates for proteases and, thus, inhibit proteases (Potvin et al., 2012).

Carbon sources such as glucose (as fermentative one) should be avoided in favour of non-fermentative ones (such as glycerol) to prohibit ethanol production which can repress the P_{AOXI} (even at levels around 10-50 mg/L) (Macauley-Patrick et al., 2005), and may harm the cells.

During methanol metabolism reactive oxygen species (ROS) are produced which can harm the cells (reduce viability) and increase the release of proteases; adding ascorbic acid to culture (in induction phase) as an antioxidant can substantially decrease proteolytic degradation (Potvin et al., 2012). In the case of utilization of alcohol oxidase defective strain (Mut^-), the cells can not grow on methanol as the sole carbon source and, thus, require an auxiliary carbon source for concomitant growth and expression (in expression phase); it is evident that the carbon source should not repress the promoter of expression. Glycerol which has widespread use as carbon source and added together with methanol during the expression may cause repression of P_{AOXI} . Mixed-carbon source feeding strategies are employed to improve the growth of the Mut^s strains (Xie et al., 2005). The candidates could be alanine, sorbitol, mannitol, trehalose (Inan and Meagher, 2001a), and lactic acid (Xie et al., 2005). Mixed-feed semi-batch may result in higher overall productivities but the maximum amount of protein can not be reached if a repressing substrate would be employed (Inan and Meagher, 2001b).

It has been observed that presence of yeast extract, casamino acids, and EDTA would enhance the protein accumulation by *P. pastoris* (Li et al., 2007).

In the cases which are intended to reduce the extent of glycosylation of an expressed protein, the host cells are cultured in the presence of

unicamycin, sugar analogue, which is a competitive inhibitor of UDP-GlcNAc dolichol P-GlcNAc transferase (Daly and Hearn, 2005).

From the viewpoint of a biotechnological process which aims high cell and product yield, there is a strict need for high concentrations of nutrients, i.e., salts and carbon sources which result in high osmolarity. The influence of osmolarity on the physiology of *P. pastoris* was analyzed in non-expressing, wild type, and an expressing strain (Dragosits et al., 2010). The results stated that the response is different to *S. cerevisiae* and an unfolded protein response (UPR)-like response was observed; there is a strong similarity between the mechanism of ESR and r-protein related stress (Dragosits et al., 2010).

2.2.2.2. pH

pH influences the activity of the enzymes in biochemical reactions and, thus, cell growth, protein formation, and protein stability (Li et al., 2007). *P. pastoris* is able to proliferate in a relatively broad range of pH (3.0-7.0) (Macauley-Patrick et al., 2005); this broadness makes it possible to conduct the fermentation in a pH which is not suitable for problem proteases (Macauley-Patrick et al., 2005) as the pH can affect protease activity in addition of stability of the expressed proteins (Potvin et al., 2012). *P. pastoris* serine and aspartic proteases are activated in low pH values which are an explanation for pH-dependency of proteolytic degradation (Potvin et al., 2012). Protein stability usually is increased by glycosylation and disulfide bond formation and, thus, affects the pH-mediated destabilization (Potvin et al., 2012). In the case of different heterologous proteins, different pH values have been reported to be optimum (Macauley-Patrick et al., 2005). The optimum pH specially depends on the stability of the protein expressed and is determined by performing a series of fermentations at different pH values (Li et al., 2007). Based on the performed studies, the optimum pH level for most of the foreign proteins ranges from 5.5 to 8 in order to decrease protease activity while preserving the protein stability (Potvin et al.,

2012). pH value of 5 was shown to be the optimum for recombinant hGH (rhGH) production (Çalık et al., 2010). According to what has been reported about products such as insulin-like growth factor (IGF-1), human midkine, growth-blocking peptide (GBP) from the armyworm *Pseudaletia separate* larvae, etc., it appears that during *P. pastoris* fermentation pH should be kept lower such as 3.0 in order not to harm the r-protein expressed (Li et al., 2007).

2.2.2.3. Temperature

Generally, *P. pastoris* fermentations are carried out in 30°C. Incubation temperatures of 30, 27, 25, and 23°C have also been reported in *P. pastoris* to decrease the proteolysis (Li et al., 2007); it can be propagated at temperatures as low as 15°C. Lower temperature leads to lower proteolytic degradation. In higher temperatures the intramolecular disulfide bonds are more prone to be formed which may result in protein aggregation, as well as the hydrophobic buried surfaces may be exposed and, thus, stimulate hydrophobic interactions which can be a prelude of protein aggregation; the aggregated and misfolded proteins are more susceptible to be degraded (Li et al., 2007). Lower fermentation temperatures can affect yield of the r-protein production (Macauley-Patrick et al., 2005). In lower temperature, the rate of protein expression is low and there is plenty of time for r-protein to be folded properly (Li et al., 2007). In lower temperatures cells are more viable and, therefore, the lysis-associated protease release is reduced (Potvin et al., 2012). Increase of the yield in production of herring antifreeze proteins by reducing the temperature from 30°C to 23°C and also increased activity of laccase by reducing the temperature from 30°C to 20°C can be illustrative examples (Macauley-Patrick et al., 2005). Low-temperature cultivation is beneficial in the case of foreign proteins that are aggregation prone and/or unstable, but the fermentation time would be increased compared to 30°C (Li et al., 2007).

Overall, cultivation temperature influences the stability and functionality of r-protein (Potvin et al., 2012); in other words, it seems that higher cell viability and lower proteolytic activity are the main reasons of increased productivity in lower temperatures (Dragosits et al., 2009). However, when the effect of temperature perturbation (20°C, 25°C, 30°C) on the proteome of an expressing and a non-expressing (as a control) *P. pastoris* strain was scrutinized led to this finding that there are other physiological parameters affect the productivity (Dragosits et al., 2009). Growth in lowered temperature may lead to higher stability of the proteins and, thus, the need for refolding and degradation machinery is decreased which as a positive side effect leads to higher secretion capacity. As a metabolic burden on host cell, r-protein production leads to higher need for NADH and ATP because of influence on folding and secretion system, in reduced temperature the stress which is exerted due to protein degradation is decreased and lower energy is wasted in this way and more energy is utilized for biomass production (Dragosits et al., 2009).

2.2.2.4. Dissolved oxygen

2.2.2.4.1. Importance of oxygen provision

Since, the design of a bioprocess has to meet both economic and technical requirements (Baumann et al., 2008), it is important to survey the effect of the key environmental factors on the physiology of the host cell while expressing r-protein (Baumann et al., 2010) which, in turn, can provide us with the opportunity of process optimization regarding product titer and bioactivity. Oxygen, as one of the factors in the center of attention, is an indispensable parameter especially in high cell density aerobic fermentations and the process performance can be influenced by its scarcity (Garcia-Ochoa and Gomez, 2009). As the simultaneous presence of the reactants in the reaction site is known the basic requirement of any (bio) chemical reaction (Nielsen et al., 2003), oxygen should also reach the desired location at

desired amount in order not to limit the reaction. In the bioreactors, often oxygen provision and transfer are the main limiting factors (Liang and Yuan, 2007) and, thus, continuous oxygen supply should be provided regarding to the low solubility of the oxygen in aqueous solutions (Table 2.10) (Nielsen et al., 2003). In addition, aeration aids either in system agitation or sweeping away the undesired volatile by-products along with carbon dioxide (Garcia-Ochoa et al., 2010).

Table 2.10. Oxygen solubility in the pure water at an oxygen pressure of 1 atm (Nielsen et al., 2003).

Temperature (°C)	Solubility of O ₂ (mmoles/L)
0	2.18
10	1.70
15	1.54
20	1.38
25	1.26
30	1.16
40	1.09

By considering the simplest equation for growth, Monod equation, and oxygen as the single limiting substrate we can describe the growth by following equation:

$$\mu = \frac{\mu_{max} \cdot C_{O_2}}{K_{O_2} + C_{O_2}} \quad (2.1)$$

where, μ is the specific growth rate and K_{O_2} is a concentration of oxygen where the specific growth rate becomes half of the maximum specific growth rate. The concentration of the oxygen below which the growth is dependent on oxygen is referred as “critical oxygen concentration” ($C_{Cr.o_2}$) and, thus, if

$C_{O_2} < C_{Cr.O_2}$ oxygen will be limiting, gene expression is negatively affected by oxygen limitation (Cereghino and Cregg, 2000) and cell metabolism will not function at its fastest rate (Doran, 1995). However, in the situations that $C_{O_2} > C_{Cr.O_2}$, the growth will be independent of oxygen concentration and oxygen will no longer be limiting.

In addition to affecting both of cell growth (biomass production) and protein expression and/or secretion, oxygen can influence the redox reactions within the cell which contribute to the protein folding and its presence can be of prime importance in order to have a biologically functional final product (Baumann et al., 2010).

2.2.2.4.2. Mathematical modeling of oxygen transfer

a. Diffusion in a stagnant layer

Because of insufficient mixing or in other words “malfunctioning hydrodynamics” of an air-liquid-intracellular/catalytic system, if there is a stagnant layer in the bioreactor, based on the Fick’s first law of diffusion, the flux of component A in a binary solution of A and B in one direction whenever the molecular diffusion is the only mechanism governs the mass transfer can be expressed as:

$$J_A = -D_{AB} \frac{dC_A}{dy} \quad (2.2)$$

The rate of transfer of component A will, therefore, be:

$$N_A = J_A \times a \quad (2.3)$$

Where, “ a ” is the mass transfer area perpendicular to the mass transfer direction. If the equation (2.2) became rearranged:

$$J_A \cdot dy = -D_{AB} \cdot dC_A \quad (2.4)$$

If we integrate the differential form of the equation (2.4):

$$\int_0^L J_A \cdot dy = \int_{C_{A0}}^{C_{AL}} -D_{AB} \cdot dC_A \quad (2.5)$$

Where, L is the thickness of the stagnant fluid film. By assuming steady state conditions and also constant diffusivity coefficient (which is temperature- and pressure-dependent), the equation can be written as:

$$J_A \int_0^L dy = -D_{AB} \int_{C_{A0}}^{C_{AL}} dC_A \quad (2.6)$$

By integrating:

$$J_A \cdot L = -D_{AB} \cdot (C_{AL} - C_{A0}) \quad ; \quad C_{A0} > C_{AL} \quad (2.7)$$

Finally:

$$J_A \cdot L = D_{AB} \cdot \Delta C_A \quad (2.8)$$

And consequently:

$$J_A = \frac{D_{AB}}{L} \Delta C_A \quad (2.9)$$

On the other hand, for mass transfer across a stagnant layer of a fluid the following correlation can be written:

“Transfer rate \propto (Transfer area) \times (Driving force)”

Which the correlation can be changed into an equality using a coefficient called “mass-transfer coefficient”. Therefore:

$$N_A = ka\Delta C_A \quad (2.10)$$

And the corresponding flux will be:

$$J_A = k\Delta C_A \quad (2.11)$$

By comparison of the equation (2.9) and (2.11) it can be realized that:

$$k = \frac{D_{AB}}{L} \quad (2.12)$$

Equation (2.12) is the simplest correlation between mass-transfer coefficient and the physical properties of the mass transfer system. If the presence of a resistance in front of the mass transfer is supposed, equation (2.11) can be written as:

$$J_A = \frac{\Delta C_A}{R_m} \quad (2.13)$$

Where, R_m represents the mass-transfer resistance which concentration difference (ΔC_A) acts to overcome it. Thus:

$$R_m = \frac{1}{k} = \frac{L}{D_{AB}} \quad (2.14)$$

b. Oxygen transfer in air-liquid-intracellular/catalytic bioreactor

As a special kind of mass transfer phenomenon, oxygen transfer from dispersed bubbles in fermentation medium into the suspended cells encompasses several steps each of them exerts a separate resistance to the

mass transfer and, thus, the mass transfer driving force should act to overcome these resistances:

$$OTR = \frac{\textit{Driving force}}{\textit{Overall mass transfer resistance}}$$

The available steps of the oxygen transfer phenomenon in a bioreactor can be mentioned as: 1) diffusion of oxygen molecules inside the bubble to the stagnant gas film inside bubble, 2) diffusion through the relatively stationary gas film inside bubble until gas-liquid interface, 3) traversing the gas-liquid interface, 4) diffusion through the relatively stationary liquid layer around the gas bubble, 5) transfer in the bulk well-mixed liquid toward the cells, 6) diffusion through the relatively immobile liquid layer around the cell, 7) moving along the liquid-cell interface in order to enter the cell interior (for individually-suspended cells), 8) transport inside the cell until the reaction location (Figure 2.14). If there were any cell aggregates, after reaching to the cell aggregates, the oxygen molecule should diffuse through the cells to reach the stagnant layer around the each cell (Garcia-Ochoa et al., 2010). Because of the low solubility of the oxygen in aqueous solutions, the resistance of the stagnant liquid layer surrounding the bubble is assumed to be the governing term of the overall mass transfer and this part changes to the rate-limiting step of the overall mass transfer process (Doran, 1995; Garcia-Ochoa and Gomez, 2009); therefore:

$$OTR \cong \frac{\textit{Driving force}}{\textit{Resistance of the liquid film around the bubble}}$$

Regarding the equation (2.14), decreasing the thickness of the fluid film and/or increasing the diffusivity in this layer will decrease the resistance and, in turn, will increase mass transfer coefficient.

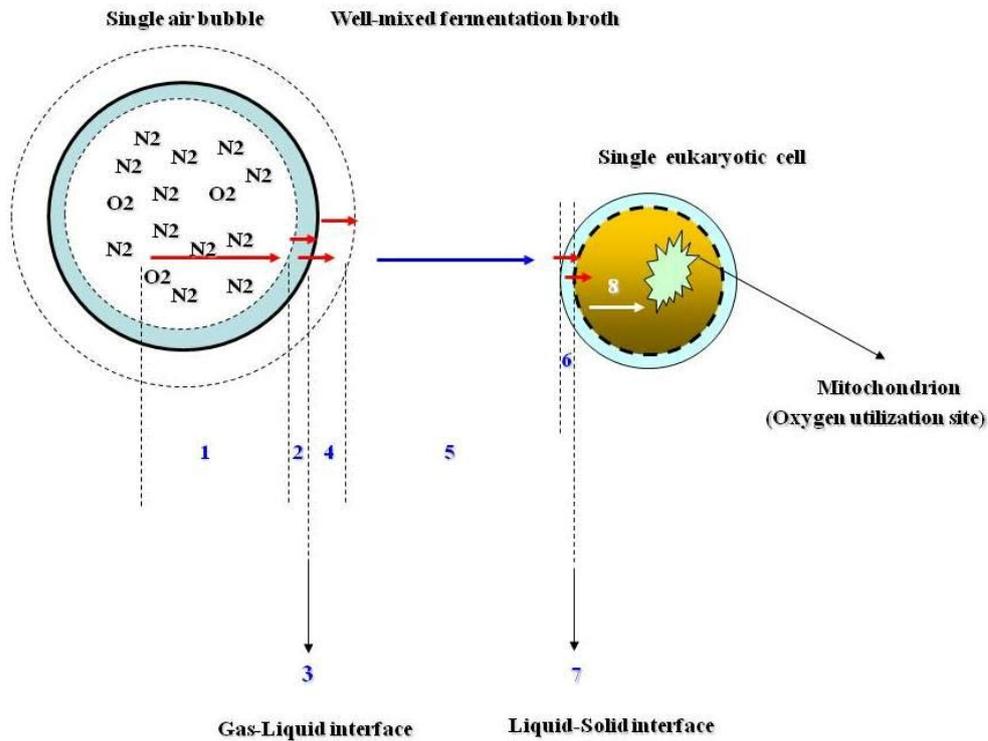


Figure 2.14. Available steps in oxygen transfer from a gas bubble to a single cell.

The mass transfer coefficient can be predicted by equations either empirical or theoretical. The empirical models generally are described in dimensionless style; the mass transfer coefficient, k_l , can be expressed in terms of dimensionless Sherwood number ($Sh = \frac{k_l L}{D_{O_2}}$) which is a function of Schmidt number ($Sc = \frac{\mu}{D_{O_2} \rho_l}$) and Reynolds number ($Re = \frac{\rho_l u d}{\mu}$):

$$Sh = f(Sc, Re) \quad (2.15)$$

Where, " μ " is the viscosity of the fermentation broth and represents the rheology of the fermentation medium.

On the other hand, the renowned theoretical models can be mentioned as: Whitman's two-film theory, Higbie's penetration theory, Danckwerts' surface renewal theory, and boundary-layer theory (Nielsen et al., 2003; Garcia-Ochoa and Gomez, 2009) where the generally admitted theory for gas-liquid transfer is Higbie's penetration theory (Garcia-Ochoa and Gomez, 2009). However, gas-liquid mass transfer is normally described by two-film theory (Nielsen et al., 2003). According to the two-film theory whenever two fluid phases come into contact boundary layers generated in both sides of the interface (phases boundary) in its close proximity which the turbulence of two phases ceases and a very thin film of relatively stationary fluid free of any kind of convective mass transfer is developed. The mass transfer through these stagnant layers is just governed by molecular diffusion (Doran, 1995). By applying two-film model (Figure 2.15) in oxygen transfer in bioreactors, we can write the oxygen mass transfer in both sides of the phase boundary.

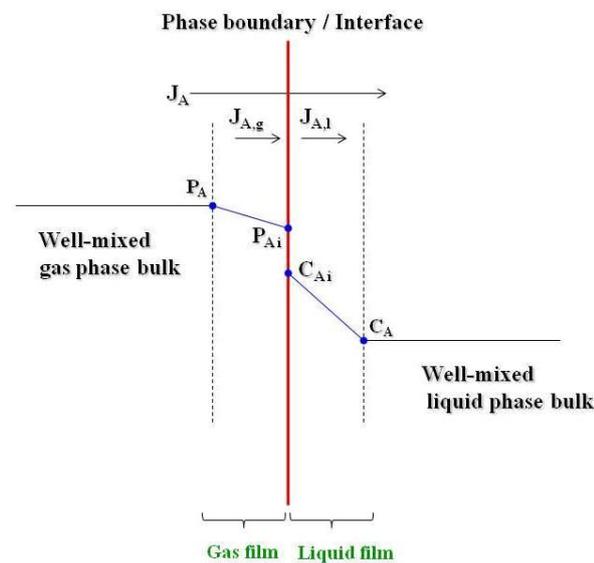


Figure 2.15. Two-film theory as the simplest model for explanation of the mass transfer from gas phase to liquid phase.

Based on the two-film theory, the mass flux of component A through the stagnant gas film can be written as:

$$J_{A,g} = k_g(P_A - P_{Ai}) \quad (2.16)$$

In addition, the mass flux of component A through the stationary liquid layer can be represented as follows:

$$J_{A,l} = k_l(C_{Ai} - C_A) \quad (2.17)$$

However, since the measurement of the interfacial concentrations, P_{Ai} and C_{Ai} , is not easily achievable, it is necessary to eliminate them in the obtained equations. Therefore, a correlation should be found between two interfacial amounts. For the components that are scarcely soluble in aqueous solutions, i.e. dilute solutions, Henry's law is applicable:

$$P_{Ai} = H_A \cdot C_{Ai} \quad (2.18)$$

Where, H_A is the Henry's constant for component A. In steady state conditions we can write $J_{A,l} = J_{A,g} = J_A$, that J_A is the overall flux from gas phase to the liquid phase regarding component A. Starting from equation (2.17):

$$J_{A,l} = J_A = k_l C_{Ai} - k_l C_A \quad (2.19)$$

$$C_{Ai} = \frac{J_A + k_l C_A}{k_l} \quad (2.20)$$

Insertion of the equation (2.20) in equation (2.18) gives:

$$P_{Ai} = H_A \frac{J_A + k_l C_A}{k_l} \quad (2.21)$$

By introducing C_A^* as the saturation concentration (maximum solubility) of the component A in the aqueous solution in bulk gas pressure of P_A , it can be written that:

$$P_A = C_A^* H_A \quad (2.22)$$

On the other hand,

$$J_{A,g} = J_A = k_g (P - P_{Ai}) \quad (2.23)$$

By replacing P and P_{Ai} with corresponding liquid concentrations:

$$J_A = k_g \left(C_A^* H_A - H_A \frac{J_A + k_l C_A}{k_l} \right) \quad (2.24)$$

By rearrangement of the equation (2.24), following equation is obtained:

$$J_A \left(1 + \frac{k_g H_A}{k_l} \right) = k_g H_A (C_A^* - C_A) \quad (2.25)$$

Dividing both sides to $k_g H_A$ leads to:

$$J_A \left(\frac{1}{k_g H_A} + \frac{1}{k_l} \right) = (C_A^* - C_A) \quad (2.26)$$

If we specify the overall flux from gas phase to liquid phase as a multiplication of driving force in liquid phase and liquid-side overall mass transfer coefficient it would be written as:

$$J_A = K_l (C_A^* - C_A) \quad (2.27)$$

Where, K_l is the liquid-side overall mass transfer coefficient and $(C_A^* - C_A)$ is the driving force for mass transfer. Therefore, by comparing equation (2.26) and (2.27) it can be deduced that:

$$\frac{1}{K_l} = \frac{1}{k_g H_A} + \frac{1}{k_l} \quad (2.28)$$

In the case of gases that are scarcely soluble in the aqueous medium which coincides with high values of corresponding H_A , ex., oxygen and carbon dioxide, the liquid side resistance overcomes the gas-side resistance, namely, k_g is much larger than k_l and, therefore, equation (2.28) can be simplified into:

$$\frac{1}{K_l} \cong \frac{1}{k_l} \quad (2.29)$$

In other words, in the case of low-solubility gases such as oxygen the overall liquid-side mass transfer coefficient, K_l , is approximated with mass transfer coefficient in liquid film, k_l . It is the parameter K_l , measured in reality but k_l is used in quantification of the mass transfer (Nielsen et al., 2003).

$$J_A = K_l (C_A^* - C_A) \quad (2.30)$$

By defining “ a ” as the gas-liquid interfacial area per unit liquid volume of the bioreactor, the mass transfer rate of the component A per unit liquid volume of the bioreactor will be calculated with following equation:

$$N_A = K_l a (C_A^* - C_A) \quad (2.31)$$

Rewriting the formula for oxygen:

$$OTR = N_A = K_l a (C_{O_2}^* - C_{O_2}) \quad (2.32)$$

Since individual measurement of the K_l and a is difficult, their product, " $K_l a$ ", which is called volumetric mass transfer coefficient is generally used to determine the volumetric mass transfer rate. Volumetric mass transfer coefficient can be determined by experimental methods or alternatively by empirical correlations which can be divided into two sets regarding the type of utilized bioreactor, i.e., stirred tank or pneumatically-agitated. In stirred tank bioreactors:

$$K_l a = f(\text{stirrer type, system geometry, } V_s, \left(\frac{P}{V}\right), \mu_a) \quad (2.33)$$

Where, V_s is the superficial gas velocity, (P/V) is the power input per volume of the fermentation broth, and μ_a is the apparent viscosity of the broth.

In bubble column and airlift bioreactors the correlation will be as:

$$K_l a = f(V_s, \mu_a) \quad (2.34)$$

Expressing the obtained equations in the form of dimensionless numbers is also prevalent.

2.2.2.4.3. Oxygen uptake rate and relation with OTR

The concentration of the oxygen in the fermentation medium (C_{O_2}) containing suspending respiring cells is dependent on its transfer from gas to liquid, OTR, and its consumption by the available microorganisms, oxygen uptake rate (OUR). By a constructing mass balance equation for oxygen the following equation is obtained (Bandyopadhyay and Humphrey, 1967):

$$\frac{dC_{O_2}}{dt} = K_l a (C_{O_2}^* - \bar{C}_{O_2}) - q_{O_2} \cdot C_x = OTR - OUR \quad (2.35)$$

OUR (representing oxygen demand of the microorganism) is regarded as one of the physiological characteristics of the culture growth and virtually reflects the viability of the fermenting cells (Garcia-Ochoa et al., 2010). By considering specific oxygen uptake rate (q_{O_2}) which specifies the rate of oxygen consumption per cell (Doran, 1995), OUR is expressed as the following product:

$$OUR = q_{O_2} \cdot C_x \quad (2.36)$$

Where, C_x is the cell concentration in the fermentation broth. Therefore, q_{O_2} is calculated from experimentally-determined OUR. The uptaken oxygen is utilized for oxidation of the substrate and generation of ATP which is, subsequently, utilized for production, biomass synthesis and cell maintenance (Garcia-Ochoa et al., 2010). As it is obvious, OUR will be dependent upon the cell amount and the specific oxygen uptake rate. It experiences an increase during exponential growth due to the high rate of substrate utilization (Garcia-Ochoa et al., 2010). Variation in C_x will lead to change in OUR. Typically, in batch fermentation, there is a maximum for q_{O_2} at the start of the logarithmic growth phase (Doran, 1995). However, q_{O_2} is supposed to be constant during growth (Garcia-Ochoa et al., 2010). q_{O_2} is the intrinsic need of a microorganism for oxygen and, in consequence, will be related to cell species (Doran, 1995; Garcia-Ochoa et al., 2010). In addition, the available nutrients as carbon source will also affect it. Dissolved oxygen concentration will also influence the q_{O_2} in the cases that $C_{O_2} < C_{Cr.O_2}$ and, consequently, the q_{O_2} will be zero order with respect to the dissolved oxygen concentration at $C_{O_2} > C_{Cr.O_2}$ (Doran, 1995).

OUR (and also OTR) can be determined via experimental techniques where the dynamic method seems to be the most practical technique because of its simplicity and the ability to be reproduced (Garcia-Ochoa et al., 2010). In this method, the inlet air is interrupted for a while ($OTR = 0$) and the change in dissolved oxygen concentration is recorded by the oxygen probe (Figure 2.16). OUR is obtained by the decrease in dissolved oxygen concentration as the slope of the dissolved oxygen versus time plot according to the equation (2.37) (Garcia-Ochoa et al., 2010):

$$\left(\frac{dC_{O_2}}{dt}\right)_{OTR=0} = OUR_d \quad (2.37)$$

OUR_d , is the measured OUR during dynamic method:

$$OUR_d = -q_{O_2} \cdot C_x \quad (2.38)$$

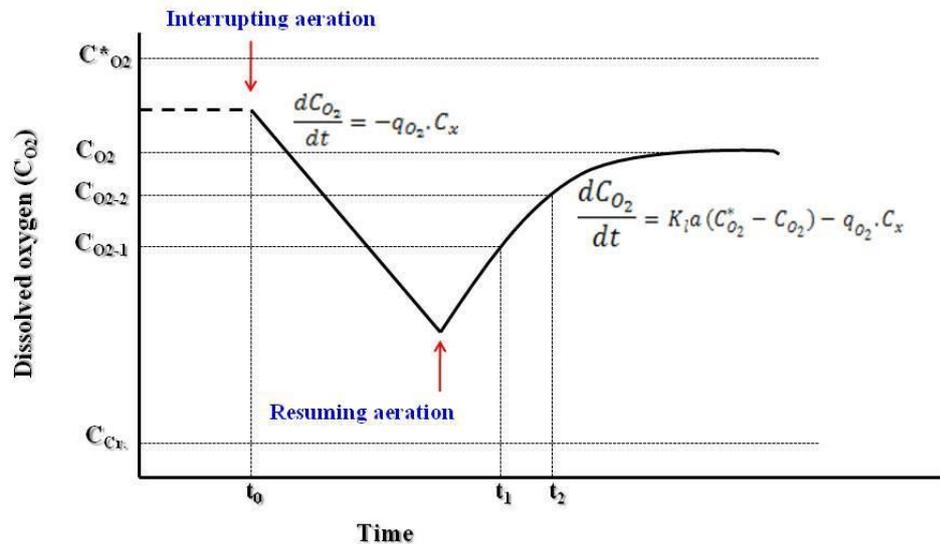


Figure 2.16. The schematization of dynamic method for determination of OUR and OTR in presence of microorganisms.

Then, the culture is re-oxygenated by constant flow of air/oxygen; the change in the oxygen concentration will again be the result of compromise between the OTR and OUR, by recording response of the system to the stimuli and obtaining the slope of the curve at a certain point (C_{O_2}) the amount of (dC_{O_2}/dt) would be determined which knowing OUR (from air interruption part) and $C_{O_2}^*$ will lead to calculation of $K_l a$ from equation (2.35).

In the other approach in this method for calculation of $K_l a$, by considering the final steady state concentration of dissolved oxygen (C_{O_2}) as a result of balance between OTR and OUR:

$$\frac{dC_{O_2}}{dt} = 0 = K_l a (C_{O_2}^* - \bar{C}_{O_2}) - q_{O_2} \cdot C_x \quad (2.39)$$

Therefore,

$$q_{O_2} \cdot C_x = K_l a (C_{O_2}^* - \bar{C}_{O_2}) \quad (2.40)$$

By replacing in the equation (2.35) and rearranging:

$$\frac{dC_{O_2}}{dt} = K_l a (\bar{C}_{O_2} - C_{O_2}) \quad (2.41)$$

Which can be solved by two available boundary conditions (Figure 2.16):

$$C_{O_2} = C_{O_2-1} \quad @ t = t_1$$

$$C_{O_2} = C_{O_2-2} \quad @ t = t_2$$

The final result will be:

$$K_l a = \frac{\ln\left(\frac{\bar{C}_{O_2} - C_{O_2-1}}{\bar{C}_{O_2} - C_{O_2-2}}\right)}{t_2 - t_1} \quad (2.42)$$

In both cases it is vital to maintain $C_{O_2} > C_{Cr.O_2}$ in order to have a constant q_{O_2} independent of dissolved oxygen concentration (Doran, 1995) since, the amount of q_{O_2} increases with increase of dissolved oxygen concentration up to $C_{Cr.O_2}$.

By knowing the OUR and OTR, the maximum cell concentration that can survive in the bioreactor can be determined by reminding that $K_l a$ represents the oxygen transfer capacity of the bioreactor (Doran, 1995) and its precise measurement is of paramount importance in design and operation of bioreactors (Garcia-Ochoa and Gomez, 2009). According to the equation (2.32), maximum oxygen transfer rate will be when the driving force ($C_{O_2}^* - C_{O_2}$) has its maximum, namely $C_{O_2} = 0$; therefore, the maximum cell concentration that can be handled by the bioreactor system in defined process conditions will be:

$$C_{x,max} = \frac{K_l a \cdot C_{O_2}^*}{q_{O_2}} \quad (2.43)$$

If this maximum cell concentration would not be able to fulfill the process requirements, the $K_l a$ should be improved.

Another important parameter that should be taken into consideration is the lowest $K_l a$ amount in order to have $C_{O_2} > C_{Cr.O_2}$, i.e., $(K_l a)_{Cr.}$:

$$(K_l a)_{Cr.} = \frac{q_{O_2} \cdot C_x}{(C_{O_2}^* - C_{Cr.O_2})} \quad (2.44)$$

2.2.2.4.4. Factors affecting oxygen transfer rate

The range of the parameters that influence the OTR is wide and encompasses the process operation conditions, bioreactor geometry, physicochemical properties of the gas phase and liquid phase, and presence of microorganisms (reflected in OUR) (Garcia-Ochoa and Gomez, 2009). Each of the three components of the OTR, namely K_l , a , and $(C_{O_2}^* - C_{O_2})$, can be the target for changing the oxygen transfer rate. However, increase of the interfacial area will have a more pronounced effect in OTR improvement (Doran, 1995). “ a ” is the interfacial area available for mass transfer which is in fact the area of the available bubbles dispersed in the fermentation broth. Thus, any action that affects the bubble sizes will conduce to change in interfacial area. Small bubbles have a lower rise velocity and, therefore, more time to transfer oxygen to the liquid phase (Doran, 1995). Furthermore, bubble size has also relation with “ K_l ”; a rigid immobile surface will be the result of the bubble diameters lower than 2-3 mm which decreases the K_l (Doran, 1995).

2.2.2.4.4.1. Operational conditions

In stirred tank bioreactors, stirrer speed, type and number of the stirrers, and gas flow rate are the major elements that play role in oxygen transfer rate adjustment (Garcia-Ochoa and Gomez, 2009). Increasing the stirrer speed leads to more energy dissipation into the liquid medium and exertion of higher shear force on the bubbles and, thus, smaller bubbles with higher “ a ”. On the other hand, increasing the stirrer speed decreases the thickness of the stagnant liquid film around the gas bubbles (Doran, 1995) and, therefore, decreases the liquid film resistance for mass transfer, i.e., increase of the “ K_l ”. However, it should be emphasized that bubbles smaller than 1mm should be avoided in processes since they reach to a immediate equilibrium with the liquid phase and their more stay in the liquid becomes futile (Doran, 1995). Gas flow have a direct impact on the surface area and its increase leads to higher gas holdup in bioreactor and, consequently, higher mass transfer area.

Temperature of the fermentation directly affects the saturation concentration where solubility decreases in fermentation medium by increase of the temperature (Nielsen et al., 2003) and, in consequence, the driving force for mass transfer is decreased. On the other hand, the temperature affects the diffusivity of the oxygen through aqueous medium; upon increase in temperature the diffusivity is increased and, thus, an increase in K_l is observed. However, the net effect of temperature on OTR depends on the range of the temperature of interest; above 40°C the net effect of temperature increase will be decrease of the oxygen transfer rate (Doran, 1995).

Total pressure of the inlet air or, in other word, the partial pressure of the oxygen in inlet air is also important by considering their effect on saturation concentration ($C_{O_2}^*$). Any increase in total pressure or oxygen partial pressure will lead to increase of solubility and, therefore, increase of driving force for mass transfer (Doran, 1995).

2.2.2.4.4.2. Presence of the cells and solutes in the medium

Any solutes available in the fermentation medium can alter oxygen solubility ($C_{O_2}^*$) and, therefore, its transfer rate from gas to liquid (Nielsen et al., 2003; Liang and Yuan, 2007). On the other hand, bubble coalescence, which is not desirable due to increase of bubble diameter, is suppressed by availability of the salts in the liquid phase which will be in favour of oxygen transfer rate; " $K_l a$ " is increased substantially by increasing ionic strength of the broth (Doran, 1995; Garcia-Ochoa and Gomez, 2009). Furthermore, concentration of the solutes (by causing viscosity variation in liquid phase) and ionic strength will influence the diffusivity (Bailey and Ollis, 1986). Increased liquid viscosity leads to decreased " $K_l a$ " (Garcia-Ochoa and Gomez, 2009) partially due to the increase of the stagnant liquid film thickness around the bubbles.

Ideal Newtonian rheological behavior can be deviated as the result of the presence of microorganisms in the broth. Although microorganisms are mostly small (1-20 μm) but their accumulation in large amounts in broth

leads to aberrations from ideal situation. Furthermore, excretion of polymeric compound also changes the medium rheological characteristics. Approximately-spherical cells in low concentrations lead to Newtonian rheology where the viscosity changes with cell concentration. Yeast and bacterial cultures exhibit this behavior (Blanch and Clark, 1997). Any probable increase in medium viscosity leads to higher thickness of the stationary layer around the air bubbles (lower K_l) bubble coalescence decreases and interfacial area “ a ” increases; overall, $K_l a$ decreases and OTR drops. In the absence of significant changes in rheological properties of the broth, the changes in “ K_l ” can be attributed to other factors. Presence of the cells in the medium not only affects the physical properties of the medium but also the cells along with their proteins and other (macro) molecules are gathered in the surface of the bubbles and interfere with related phenomena to bubbles like break-up and coalescence. Their presence in the interfacial area decreases the available area for mass transfer (Doran, 1995). In addition, microbial aggregates lead to decrease in diffusivity compared to pure water (Bailey and Ollis, 1986).

2.2.2.4.4.3. Surfactants

During the various phases of the fermentation, secretion of the numerous proteins, polysaccharides, and fatty acids into the medium which bind spontaneously to the mass transfer interface area leads to the decrease in the surface tension. Consequently, the average bubble diameter in the bioreactor decreases and the interfacial area “ a ” increases (Bailey and Ollis, 1986). Contrary to the area, the amount of the “ K_l ” is decreased as the direct consequence of the adsorption of the biomolecules to the surface. This biomolecular film leads to increase of the mass transfer resistance either by creation of an immobile surface or by introducing a new resistance by acting like a membrane. All-embracing, the product “ $K_l a$ ” increases by addition of surfactants (Bailey and Ollis, 1986).

2.2.2.4.4. Antifoaming substances

Accumulation of the cell-specific substances exported to the extracellular medium increase the foaming tendency of the fermentation media (Bailey and Ollis, 1986). By gradual increase of these compounds the non-coalescing character of the medium cannot be further preserved and bubbles tend to merge and become larger which is not favorable from mass transfer point of view. Antifoaming substances as the compounds that lower the surface tension, provide the fermentation system with lower bubble size (like surfactants) and, thus, higher interfacial area “ a ” for mass transfer. In contrast, their presence as a macromolecular film in the transfer area results in rigid interface and decreased liquid movement near interface and, therefore, decreased “ K_l ” where this decrease generally, conquers to the increase brought about by “ a ” and, in total, “ $K_l a$ ” is decreased dramatically (Doran, 1995; Garcia-Ochoa and Gomez, 2009).

2.2.2.4.4.5. Oxygen uptake rate

Oxygen uptake rate during the process can influence the mass transfer coefficient (Çalık et al., 1998). Enhancement of the OTR as the result of the consumption of the oxygen by the cells has been reported (Garcia-Ochoa and Gomez, 2009). As being proposed and confirmed by a mathematical model, the cells are being collected in the gas-liquid interface and enhance the oxygen transfer rate. In order to consider the improvement caused in OTR by presence of the cells in a system, biological enhancement factor (E) is defined (Çalık et al., 1998):

$$E = \frac{\text{Oxygen transfer flux in presence of consuming cells}}{\text{Oxygen transfer flux in absence of consuming cells}}$$

The hydrodynamic conditions and driving force for mass transfer should be similar in both conditions.

2.2.2.4.5. Oxygen-limited situations in bioprocesses

Severe oxygen limitation is one of the environmental stress factors so, dissolved oxygen level should be strictly controlled which is feasible only in controlled environment of a bioreactor (Cereghino and Cregg, 2000). Since oxygen in the role of the final electron acceptor in the TCA cycle plays a vital role in provision of ATP as the energy currency of the cell, shifting from complete aerobic conditions which leads to fully aerobic metabolism of glucose (via TCA cycle), toward oxygen-limited and low-oxygen availability conditions affects the core metabolism leading to ethanol production which, in turn, faces the cell with energy deprivation and makes it to adopt physiological mechanisms to compensate energy shortage (Baumann et al., 2010). As r-protein production (expression and secretion) is a multi-step cellular function which needs energy in the form of ATP, any perturbation in ATP/energy level because of change in oxygen provision can greatly influence the expression and/or secretion processes (Baumann et al., 2008; Baumann et al., 2010). As the dissolved oxygen level decreases during industrial fermentation, various efforts are carried out to maintain it at high levels (Camattari et al., 2007). Although air supplementation with pure oxygen is not so challenging in small scales but can be cost-prohibitive at large scale productions (Ferreira et al., 2012); the need for pure oxygen leads to higher costs and may conduce to problems in scale up where heat exchange and oxygen transfer capacities are limited. Therefore, implementation of a system with lower oxygen demand will be easier (Camattari et al., 2007). In consequence, low-oxygen availability conditions have been adapted for such reasons. Reaching such conditions does not require elaborate induction procedure and is just by closing the inlet air valve (Camattari et al., 2007). A two-stage mechanism (small amount of air inlet after oxygen starvation) for induction of low-oxygen availability can help cells to refresh (Camattari et al., 2007).

Reduced oxygen supply will lead to adaptation of a respiro-fermentative metabolism rather respiratory (Baumann et al., 2010). As an

advantage, partially fermentative pathway leads to lower biomass yield and, thus, less biomass is accumulated during process and consequently the cell removal is facilitated (Baumann et al., 2008). There will be cellular adaptations to low oxygen availability; cells have to shift growth to biomass reorganization in order to cope with the strongly reduced availability of ATP, and readjust their metabolic fluxes from cellular respiration to fermentation. The extent of such adaptation is important in the case of facultative anaerobe yeasts with different capacities in glucose fermentation (Crabtree effect) (Baumann et al., 2010). As previously mentioned, *P. pastoris* prefers a respiratory pathway rather than fermentative, thus, the oxygen demand is high during high cell density cultivation. Since *P. pastoris* is 'Crabtree-negative' (fermentation will only be induced when oxygen is severely limited (Baumann et al., 2008)) shows more sensitivity to the oxygen concentration in medium than 'Crabtree-positive' *S. cerevisiae*.

Sterols play role in the membrane fluidity and as molecular oxygen is required in the pathway of sterol synthesis (biosynthesis route), thus, oxygen scarcity will perturbate the sterol production and membrane characteristics. Both ergosterol pathway enzymes and sphingolipids synthesis genes transcription are affected by oxygen limitation. Any change in normal composition of the membrane sterol-sphingolipid will cause not only defects in membrane physical properties but also profound changes in related events such as cell signaling or secretion of proteins to the cell outside (Baumann et al., 2010).

2.3. Further considerations

Some important issues were not included among the major parameters considered hitherto, but, should be explained as they are inalienable parts of the r-protein production in any eukaryotic host including *P. pastoris*. Glycosylation of the protein (related with protein biological activity), ER stress (related with unfolded proteins in ER), and proteolysis (related with product stability) are the subjects which will be explained.

2.3.1. Glycosylation

Glycosylation is regarded as the most common and most intricate PTM which is in association with 40% of the approved pharmaceutical products (Macauley-Patrick et al., 2005; Corchero et al., 2012). Approximately 0.5-1% of the expressed proteins in eukaryotes are glycoproteins (Daly and Hearn, 2005) however, almost all of the secreted eukaryotic polypeptides are glycosylated (Demain and Vaishnav, 2009). Since most of the pharmaceuticals, over 70% of therapeutic proteins whether in clinical or preclinical stage (Chung et al., 2010), are classified as glycoproteins, and also the profound effect that the location and extent of the chains may have on the action of the protein, the recombinant host should possess a proper glycosylation pattern to obtain a biologically active product (Macauley-Patrick et al., 2005) and, thus, their production in mammalian cells will not be so surprising (Papini et al., 2010). A dissimilar pattern of glycosylation among the (micro) organisms is somewhat clear and glycosylation is species-, tissue-, and cell-type- specific (Demain and Vaishnav, 2009). It is worthy to put forward that there are some cases that carbohydrate moiety has no role in biological activity of the protein and, thus, glycosylation is not of prime importance (Demain and Vaishnav, 2009).

The parameters that glycosylation can alter and/or affect can be: serum half-life (stability), solubility, immunogenicity, (bio) activity, thermal stability, and receptor binding (Demain and Vaishnav, 2009). It has been suggested that protein glycosylation in general facilitates its secretion (Eylar, 1966). On the other hand, there are parameters which can influence glycosylation. Any delay in protein delivery from ER can bring about extended glycosylation, since the glycosylation is the inalienable part of the secretion pathway (Daly and Hearn, 2005). Composition of the medium that cells were grown can govern the pattern of the glycosylation of a protein (Demain and Vaishnav, 2009); it has been reported that O₂-limitation affects the N-glycan composition (Corchero et al., 2010). Furthermore, as there is a competition between glycosylation and folding in ER, the spatial

arrangement of the sequence (in 3D structure) is important (Daly and Hearn, 2005).

Adding specific sugar moieties to the protein, known as glycosylation, is a vital PTM carried out in “O” and “N” forms. In O-glycosylation, the residue is added to the hydroxyl group of serine or threonine and in N-glycosylation, the sugar is added in the specific recognition sequence (Asn-X-Ser/Thr) to (R-group of) asparagines (Asn) residue (Cregg 1996; Cereghino and Cregg, 2000; Demain and Vaishnav, 2009); regarding to the decreased K_m of glycosyltransferase enzyme in reaction with “Asn-X-Ser” the glycosylation is more readily achieved if the sequence would be “Asn-X-Thr” (Daly and Hearn, 2005). O-glycosylation takes place mainly in Golgi apparatus (Daly and Hearn, 2005). In higher eukaryotes, O-glycosylation encompasses addition of variety of sugars such as N-acetylgalactosamine, galactose, and sialic acid, in contrast, yeast O-glycosylation (Figure 2.17) is carried out by addition of one to five mannose oligosaccharides linked by α -1,2 linkage (Cregg 1996; Cereghino and Cregg, 2000; Macauley-Patrick et al., 2005; Daly and Hearn, 2005). It seems that there is no consensus primary amino acid sequence in order to forecast O-glycosylation residue (Cereghino and Cregg, 2000; Daly and Hearn, 2005); however, in the case of occurrence of unusual number of serine/threonine, presence of proline residues in proximity of serine/threonine, or occurrence of positively/negatively charged amino acids in the vicinity of serine/threonine the likelihood of O-glycosylation increases or it is promoted (Daly and Hearn, 2005). It is also probable that the O-glycosylation happens in a place other than native location, or a protein without any O-glycosylation would be glycosylated (Cregg, 1999).

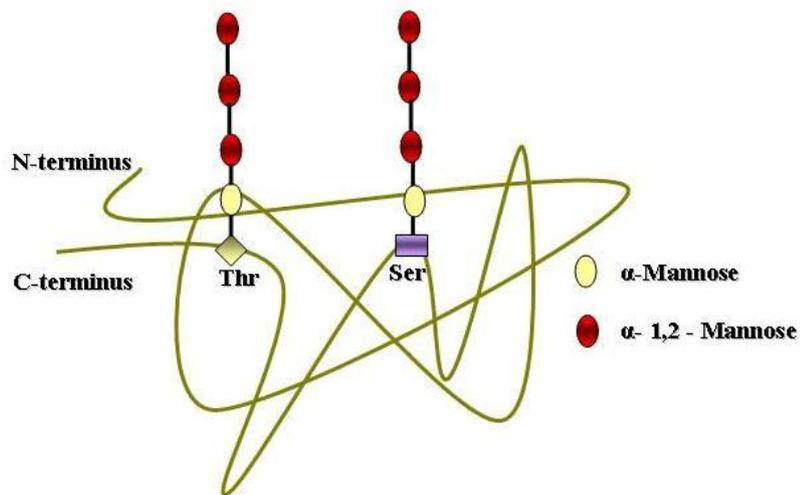


Figure 2.17. Common pattern of O-glycosylation in yeasts.

The main difference is the pattern of N-glycosylation in yeasts (and *P. pastoris*) compared to higher eukaryotic cells. The Golgi apparatus of yeasts does not carry out the trimming procedure of N-glycans like the mammalian Golgi and thus long mannose chains are added to the core glycan (oligosaccharide) (Corchero et al., 2010). N-glycosylation starts from ER and continues to the Golgi apparatus afterwards. First modification step is very similar and highly conserved in eukaryotes; after addition of similar oligosaccharide core unit ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (Cregg, 1999)) in both lower and higher eukaryotes it is trimmed to $\text{Man}_8\text{GlcNAc}_2$ (Figure 2.18) by removal of the three glucose residues with the aid of glucosidase II and I and removal of α -1, 2-linked mannose residue by the action of α -1, 2-mannosidase (Daly and Hearn, 2005). Further processing is performed in Golgi apparatus and the N-glycosylation pattern, N-glycan processing, diverges markedly from now on. In yeast and other fungi, high-mannose type chains are formed by the action of mannosyl-transferases and mannosylphosphate transferases, however, in mammalian cells, removing mannose residues from the glycans by α -1, 2-mannosidase located in Golgi

is the initial step which continues by addition of galactose, fucose, N-acetylglucosamine, and sialic acid to the N-glycan. Sialylation is the final step in human glycosylation (Bollok et al., 2009). In *S. cerevisiae* Golgi apparatus, contrary to mammalian Golgi, the addition of mannose outer chains continues without trimming reactions (Cregg 1996; Cereghino and Cregg, 2000; Daly and Hearn, 2005) which leads to obtaining long mannose outer chains with a broad range of size between 50-150 mannose residues (hyperglycosylation) (Cereghino and Cregg, 2000). The most notable difference between *P. pastoris* and *S. cerevisiae* is the (frequent) absence of hyperglycosylation in *P. pastoris* (Cereghino and Cregg, 2000; Boettner et al., 2002) and glycoproteins expressed in *P. pastoris* have an average mannose chain of 8-14 (Li et al., 2007). High-mannose N-linked chains (high-mannose type glycosylation) added during passage through yeast secretory pathway are drawbacks in production of pharmaceutical proteins in yeast (Cereghino and Cregg, 2000).

Additionally, presence of α -1, 3-linked mannosyl terminal linkages in *S. cerevisiae* leads to immunogenic reactions in patients. However, due to the lack of the α 1, 3-linked mannosyl transferase (Demain and Vaishnav, 2009), *P. pastoris* can not add terminal α -1, 3 mannose (potent antigen) to the oligosaccharide chain and, in consequence, these linkages are absent in *P. pastoris* (Cregg 1996; Cereghino and Cregg, 2000). Therefore, *P. pastoris* contains a non-allergenic α -1, 2 linkage (Böer et al., 2007). It should be noted that glycosylation is not a strict prerequisite for functionality of a protein but the issue of immunogenicity should be regarded (Banga et al., 2006). In addition to being immunogenic, therapeutic glycoprotein with different glycosylation pattern may experience complicated downstream purification or rapid clearance from body (Laukens et al., 2015).

Cumulatively, because of low shear resistance and low productivity of mammalian cells, yeast, especially *P. pastoris*, could be used as a promising alternative, by two strategies (Cereghino et al., 2002): removing the glycosylation sites of protein by altering its primary amino acid sequence

without interfering bioactivity or humanizing the glycosylation pattern through reengineering the pathway (Macauley-Patrick et al., 2005) to produce homogeneous and human-like glycans. Engineering of the glycosylation pathway in *P. pastoris* has been reported previously (Bobrowicz et al., 2004). Generally, the first target of engineering is devastating the hypermannosylation steps (De Pourcq et al., 2010) and the general strategy includes deletion of an endogenous glycosyltransferase gene and step-by-step introduction of heterologous glycosylation enzymes in order to confer glycosidase or glycosyltransferase activity into ER or Golgi (Jacobs et al., 2009). Engineering of the O-glycosylation pathway is not as advanced as N-glycosylation modification (Corchero et al., 2012).

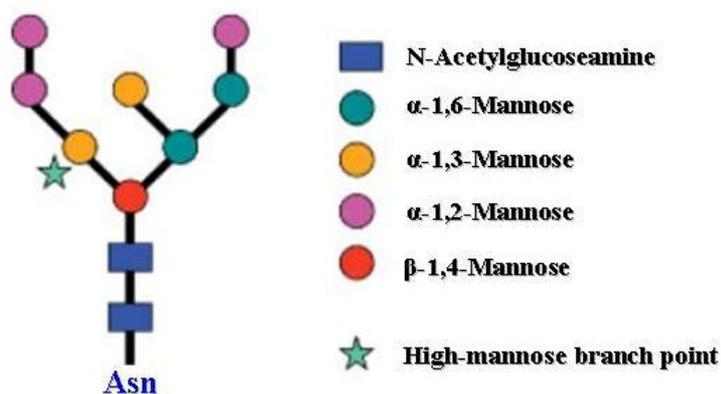


Figure 2.18. Structure of the core Asn-linked oligosaccharide (Daly and Hearn, 2005). The "star" shows the place where further oligosaccharides are added.

2.3.2. Endoplasmic reticulum under stress

Not all proteins fold simultaneously as they are synthesized in the cytosol and there should be an additional assistance for their proper folding. In eukaryotes, most secreted proteins and transmembrane proteins are directed toward ER. In addition, depending on the cell responsibilities there may be a huge amount of specific secreted protein molecules dispatched toward ER.

Furthermore, in the r-protein production processes the client protein is also expected in ER. Therefore, the flux through the ER may become formidable and eukaryotes ER should be equipped with special features to overcome the heavy load of the entering nascent polypeptides. ER possesses a complicate surveillance system, ER quality control (ERQC), whose proper functioning will guarantee folding and modification of secretory and membrane proteins and, further, elimination of finally-misfolded proteins (Araki and Nagata, 2011).

In yeasts the folding-assistant proteins (chaperones) reside in ER (Bollok et al., 2009) and the protein folding program proceeds in this compartment along with some PTMs such as N-glycosylation and disulfide bond formation. Protein folding is conducted via action of a group of chaperones and modifying enzymes (Ruggiano et al., 2014) as a network (Brodsky and Skach, 2011); these molecular chaperones either facilitate correct folding or provide a suitable microenvironment where folding can occur which results in minimized aggregation, facilitated native structure formation, and oligometric assembly (Brodsky and Skach, 2011) . ER-resident chaperones system includes i) protein disulfide isomerases (PDIs); ii) Hsp40 (J-proteins), Hsp70 (BiP/Grp78/Kar2), and Hsp90 chaperones, and iii) lectin-based chaperones such as calnexin (CNX) and calreticulin (CRT) (Gasser et al., 2008; Brodsky and Skach, 2011; Shen et al., 2012).

Upon entering ER, translocation, both endogenous and r-proteins, undergo specific pathway to be folded and modified correctly. However, an essential prerequisite for folding of a newly-synthesized protein is being soluble (Ruggiano et al., 2014). Kar2 in yeasts (BiP in mammals), aids the translocation of polypeptide chain into ER (Gasser et al., 2008), plays role in stabilization of immature protein (misfolded or newly-synthesized) through preferential binding to hydrophobic amino acid patches and preventing misfolding until ending the process of folding (Bollok et al., 2009; Idiris et al., 2010; Shen et al., 2012; Ruggiano et al., 2014). This enzyme mediates (in a cyclic process) the formation of disulfide bonds by a cycle of PDI and Ero

action which may cause formation of reactive oxygen species; Ero provides oxidizing equivalents for PDI (Brodsky and Skach, 2011). PDI is a multi-function oxidoreductase protein (yeast contains 5 PDIs) that takes part in disulfide bond formation, isomerization of primarily formed non-native disulfide bonds (Brodsky and Skach, 2011), and proper protein folding in oxidative environment of the ER. It also reduces disulfide bonds in misfolded proteins (Ruggiano et al., 2014). One of the crucial and preliminary modifications performed on nascent polypeptide is addition of a core oligosaccharide unit (mentioned in N-glycosylation) which facilitates interaction with PDI and provides a binding site for CNX and CRT (Brodsky and Skach, 2011). Glycan chain attachment is followed by successive cleavage of two glucose molecules from the core unit which renders the protein molecule a high-affinity substrate for CNX and CRT (Brodsky and Skach, 2011). CNX and CRT share a high structural homology although, CNX is a membrane protein and CRT is a soluble protein (Brodsky and Skach, 2011). Nascent glycoproteins bound by CNX are retained in the ER and are mediated for correct folding and processing of the N-glycans (Gasser et al., 2008; Brodsky and Skach, 2011). Removal of the final glucose residue generates a fragment that does not bind CNX/CRT. Folded substrates leave the ER but incompletely- or mis-folded N-glycosylated polypeptide substrates are sent back to be complexed with BiP (Kar2) chaperone or enter a second cycle of CNX/CRT, after reglucosylation by UGT (UDP-Glc glycoprotein glycosyltransferase) (Gasser et al., 2008; Brodsky and Skach, 2011). Enzymes responsible for reglucosylation and deglucosylation compete with mannosidases which catalyze the reaction slowly and reside in ER/Golgi; since mannose diminishes the reglucosylation by UGT, thus, a substrate can be generated for different set of lectin-like chaperones, EDEM and Yos9 for yeasts, in order to pass a second level of ERQC (Gasser et al., 2008; Brodsky and Skach, 2011). Providing that the conformation of the protein is certified by ERQC system resembling a molecular clock, a glucose-dependent surveillance mechanism which acts by trimming and

reattachment of glucose and confers the protein a limited time and multiple opportunities of refolding, only correctly folded, modified and assembled proteins are allowed to travel further along the secretory pathway (Gasser et al., 2008). Then, the protein would be transported to Golgi for further modification via vesicular transport (Gasser et al., 2008; Bollok et al., 2009) else, it will be get stuck in the ER. Prolonged BiP binding (and also CNX binding (Gasser et al., 2008)) is an indicator of mis-folding. The unusually high and non-physiological rates of r-protein production and the occurrence of significant amounts of mis-folded protein species, still an unsolved problem (Damasceno et al., 2012), drive the cells to a global conformational stress condition (Gasser et al., 2008). Providing that the accumulation of the misfolded proteins is not permanent or stress is brief, misfolded proteins will be handled with unfolded protein response (UPR) or will be destroyed by endoplasmic reticulum-aided degradation (ERAD) (Brodsky and Skach, 2011). The entrance of pre-mature proteins to ER, their quality control and subsequent degradation or secretion has been schematized in Figure 2.19.

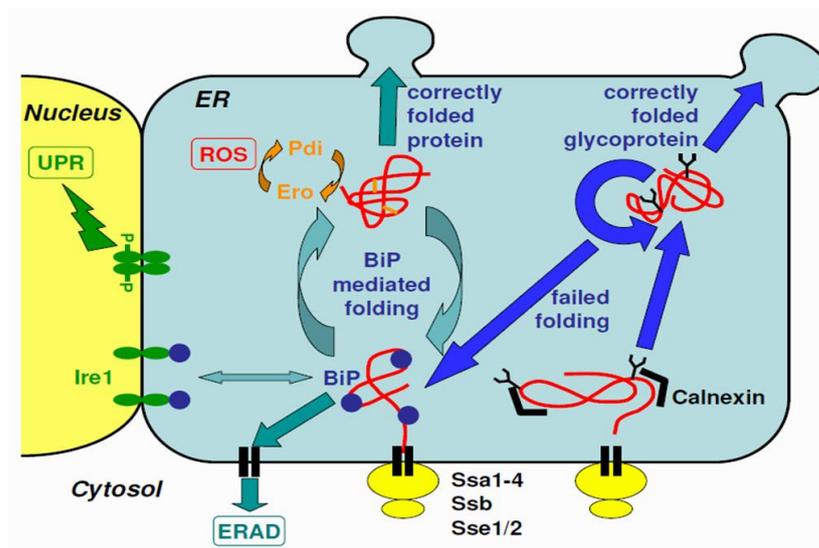


Figure 2.19. Representation of the protein folding, quality control, degradation, and secretion in yeasts (Gasser et al., 2008).

Disturbance of folding occurred by any reason, such as non-physiological expression rates of r-protein, (Gasser et al., 2008) inside ER, mutations, sub-stoichiometric amounts of binding partner or stoichiometric errors (Ruggiano et al., 2014; Brodsky and Skach, 2011), or shortage of a chaperone (Ruggiano et al., 2014) compelling the cells to commence UPR and also ERAD. Misfolded proteins are potentially harmful species and need to be removed.

UPR is a “cytoprotective” signal transduction cascade (Kimata et al., 2006) which entails series of responses for preserving homeostasis and cell survival (Mori, 2009) through alleviation of ER stress and restoring its function of protein folding (Hetz, 2012). Briefly, ER stressful condition, accumulation of unfolded/misfolded proteins, is sensed by UPR sensor (s). The received/sensed signal is transmitted to the nucleus (via cytosol) in order to adjust target genes expression in nucleus and, in consequence, provoking some compensating reactions finalized with buffering of ER stress (Figure 2.20) (Gasser et al., 2008; Damasceno et al., 2012; Hetz, 2012). All the target genes possess a segment called UPR element which signal transduction targets it in order to up-regulate intended genes transcription. Induced genes which exceed 300 genes comprise groups related to protein folding, ER-resident chaperone, vesicular transport, lipid biosynthesis, and degradation (ERAD) in order to prevent probable detrimental effects of mis-folded proteins and, subsequently, to recover the "cellular folding homeostasis" (Kimata et al., 2006; Gasser et al., 2008). The presence of Ire1 (inositol-requiring protein 1 α) as yeasts UPR sensor was proved for the first time in *S. cerevisiae* after initial characterization of ER stress in this budding yeast (Hetz, 2012); Ire1 is a single-pass transmembrane protein resides in ER membrane (Brodsky and Skach, 2011). Afterwards, a functional transcription factor, Hac1, is translated by the activity of Ire1 (Kimata et al., 2006; Hetz et al., 2012) which binds to the UPR elements of targeted genes in the nucleus (Kimata et al., 2006).

In sustained ER stress and, thus, extended UPR, accumulation of misfolded proteins especially those with disulfide bonds leads to increase of intracellular ROS as a consequence of repetitive oxidative folding attempts by foldases such as PDI; consequently, oxidative stress is brought about which obligates cellular protection. Therefore, the enzymes that relieve oxidative stress are also among the transcriptional targets of yeast Ire1-Hac1 pathway (Kimata et al., 2006; Gasser et al., 2008; Zhu et al., 2011; Brodsky and Skach, 2011).

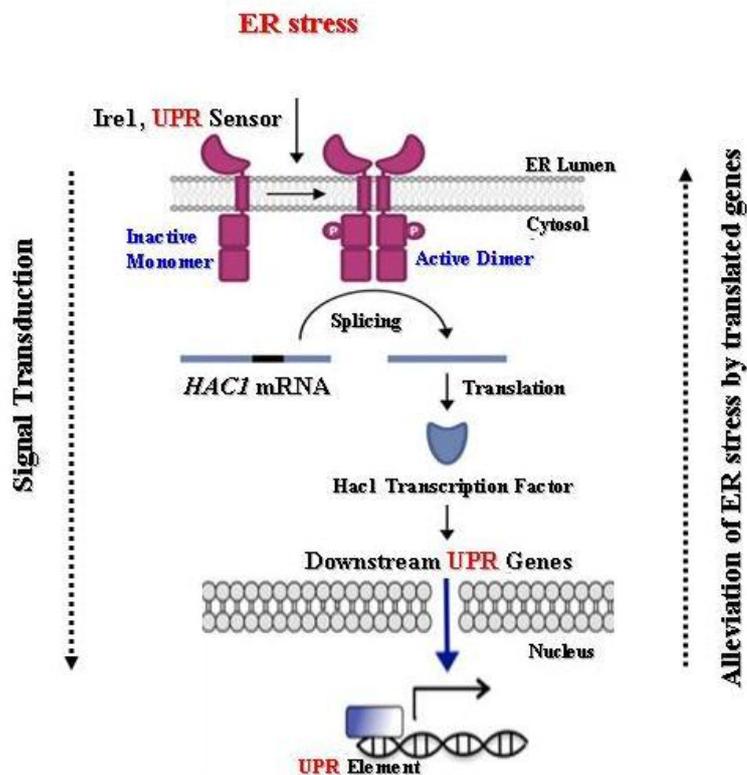


Figure 2.20. UPR as an intracellular signaling pathway from ER to nucleus. Ire1 is the unique UPR sensor found in yeasts. Mammals possess three UPR sensors as a result of evolution and, thus, we expect more elaborate adaptive responses facing ER stress (Mori et al., 2009).

If the increased handling capacity of the ER cannot afford the unfolded/misfolded proteins, they are gradually subjected to ERAD in order to get rid of their detrimental effects to the cell normal functioning (Ruggiano et al., 2014). ERAD is reverse translocation of mis-folded protein (could not pass ERQC (Brodsky and Skach, 2011)) through a “protein conducting channel” (Ruggiano et al., 2014) to the cytosol and tagging it by ubiquitin in order to proteasomal degradation by 26S proteasome in the cytosol (Gasser et al., 2008). So, ERAD encompasses four steps (Figure 2.21): i) recognition of substrate; ii) retranslocation to the cytosol; iii) ubiquitination in cytosolic side of ER, and iv) degradation by proteasome in cytosol (Brodsky and Skach, 2011). It should be emphasized that ERAD has different categories each specific for a different type of misfolded proteins. However, the principal and the sequence of events which lead to protein clearance are similar (Ruggiano et al., 2014). The whole steps are accomplished by orchestration with an ER membrane-embedded protein complex, E3 ligase complex (Ruggiano et al., 2014).

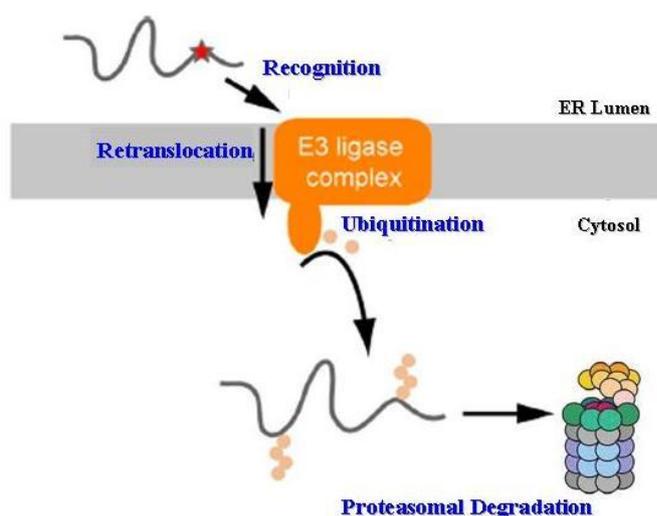


Figure 2.21. Simple representation of ERAD including its four well-defined steps. Reproduced from (Ruggiano et al., 2014)

The ER-associated responses have been reported to be activated while over-expression of r-proteins. These stress reactions thus prevent improvement of r-protein production via increasing gene dosage and prohibit the optimization by introducing a major bottleneck. It seems acceptable that co-expression of different chaperones as folding helpers with the target gene have a substantial effect on yield. However, it has been realized that influence of co-expressed BiP is protein or host specific. Regarding PDI, its sole over-expression has improved the secretion of some but not all proteins (Idiris et al., 2010; Huo et al., 2007).

2.3.3. Proteolysis (Product stability)

The vacuole of the yeast has been recognized as one of the “degradative organelles” dissected hitherto. Along with numerous physiological responsibilities, proteolysis, although in a nonspecific and uncontrolled manner, has been proved to be one of the functions of the vacuole and the level of various proteases reside in the vacuole, and thus proteolysis, changes with nutritional conditions of the yeast. Proteolytic degradation has been considered as an ever-present problem when yeasts are utilized as expression platforms for r-protein production (Van Den Hazel et al., 1996). Three main groups of the proteases have been identified in *P. pastoris*: i) cytosolic proteases or proteasomes; ii) secretory pathway proteases, and iii) vacuolar proteases. Excluding proteasomes, the other two groups have been summarized in Table 2.11 (Zhang et al., 2007).

Despite reaching high cell densities in bioreactor cultures of *P. pastoris*, proteolytic degradation of the heterologous protein would be a probable outcome of lysed cells and, thus, releasing their vacuolar proteases (excess production of the vacuolar proteases can lead to their secretion to cytosol and subsequently the extracellular medium (Sinha et al., 2005)). Furthermore, proteolytic degradation can be caused by extracellular or cell-bound proteases. Proteolysis will conduce either to truncated protein and, thus, decreased yield or problems in downstream processing because of the

release of intracellular constituents or intermediates of the protein degradation (Zhang et al., 2007). Lysed cells may be the inevitable consequence of methanol (a toxic material to cells) as a common inducer and carbon source in *Pichia*-based expression systems utilizing P_{AOX} -driven gene transcription; methanol and/or its subsequent metabolites such as formaldehyde, and hydrogen peroxide resulting from its catabolism in peroxisome are menace to available cells and are the major reasons of cell lysis (Mattanovich et al., 2009). Obviously, to overcome the proteolysis, especially in this case, using strategy other than methanol induction is the right choice (e.g., using glucose grown cells and P_{GAP}) for dampening the oxidative stress. In order to ameliorate the process by decreasing proteolysis, different strategies can be put forward; they can be classified into cell-level, cultivation-level (Macauley-Patrick et al., 2005) and, protein-level (Zhang et al., 2007). The cell-level strategies comprise recruitment of protease-deficient strains of *P. pastoris* in which genes of vacuolar protease have been knocked out which can lead to lower proteolysis up to a point (Macauley-Patrick et al., 2005; Potvin et al., 2012); vacuolar proteases seem to be the major cause of protein degradation (Li et al., 2007). Protease-deficient mutants have a disruption in the proteinase A gene (*PEP4*) and/or proteinase B gene (*PRB1*). SMD1163, SMD1165, and SMD1168 are the developed mutant strains which possess lower growth rate and viability. SMD1163 is a double-mutant which lacks both PrA and PrB proteases (Zhang et al., 2007). Using these mutant strains may put the production at stake where the protease itself is necessary for proper processing of the protein (Potvin et al., 2012). Therefore, these strains should be used only when other strategies failed to respond (Macauley-Patrick et al., 2005)

The cultivation-level strategies are related with medium composition, pH, temperature, and protease inhibitors (Macauley-Patrick et al., 2005; Potvin et al., 2012). Nutrient starvation should strictly be avoided during cultivation, as it brings about cell lysis. It has been unraveled that starvation, change of carbon sources, temperature and pH changes, or toxic chemicals

can impose stress on yeast cells (Sinha et al., 2005). r-protein expression is affected by medium composition (Zhang et al., 2007). Complex or enriched media, contrary to defined media, have the capacity of preventing nutrient limitation along with protease activity (Potvin et al., 2012). By adding amino acid-rich substances such as peptone or casamino acids to the medium they act as competing or excess substrate for proteases. Addition of skimmed milk to the induction medium has been reported to increase the yield of human bile salt stimulated lipase (Cereghino et al., 2002; Macauley-Patrick et al., 2005). Keeping the conductivity of the medium at a low constant level by salt feeding has also been reported to be a useful method in decreasing cell death and, therefore, protease concentration. However, salt can inversely influence protein titer (Zhang et al., 2007). Utilization of a carbon source with a less harmful nature and by-products can be another point to be considered. Since the growth of *P. pastoris* is not influenced remarkably by pH range (3.0-7.0), changing medium pH to a level (other than optimal for protease) in order to diminish protease activity could be an effective solution providing that the stability of the expressed protein would not be affected. Lower medium pH will result in lower protease activity (Zhang et al., 2007). Lowering the temperature can also increase the yield (Macauley-Patrick et al., 2005; Potvin et al., 2012) by decreasing protease activity. The improvement can also be ascribed to the poor stability of recombinant protein at higher temperatures, more release of protease, or folding problems at higher temperatures (Macauley-Patrick et al., 2005) because of hydrophobic interactions and consequent aggregation (Li et al., 2007). As a method to minimize proteolysis, expressions at temperatures of 27°C, 25°C, and 23°C have been carried out but with prolonged period of fermentation (Li et al., 2007). Addition of protease inhibitors to the medium can also be a solution for attenuating proteolysis. However, in an industrial scale, cost of utilization of such additives can be prohibitive (Zhang et al., 2007). Some protease inhibitors can be mentioned such as phenyl methyl sulfonyl fluoride (PMSF) as serine proteases inhibitor, EDTA, pepstatin (as aspartic proteases

inhibitor) (Sinha et al., 2005). Finally, providing that not competing for cell resource reservoirs, the co-expression of protease inhibitors in tandem with the desired gene can have a profound effect on the proteolytic activity and, therefore, increasing the yield.

Proteolysis is protein-dependent since the cleavage site (specific amino acid sequence) maybe absent in some products or it has been concealed because of unique three-dimensional structure of the protein product. Considering this fact, one method for alleviating the proteolytic degradation can be, as a protein-level strategy (Zhang et al., 2007), modifying the amino acid sequence of the expressed protein, without changing its bioactivity, to change the pattern recognized by the relevant protease(s) (Potvin et al., 2012).

Table 2.11. Non-cytosolic proteases of yeasts (extracted from (Zhang et al., 2007)).

Protease (abbreviation)		Gene	Protease features
Vacuolar			
Endoproteinase	Proteinase A (PrA)	<i>PEP4</i>	<ul style="list-style-type: none"> • Aspartyl proteinase • 405-aa precursor • 42-kDa glycoprotein • Possessing “pre-pro” SP • Three PTMs upon entering ER • 2 N-linked glycosylation • Necessary for activation of vacuolar proteases such as PrB and CpY
	Proteinase B (PrB)	<i>PRB1</i>	<ul style="list-style-type: none"> • Subtilisin family of serine protease • 69-kDa precursor • 31-33 kDa mature enzyme • Possessing “pre-pro” SP • Autocatalytic cleavage of “pro” sequence • Final cleavage takes place in vacuole to give mature PrB • One N-linked and one O-linked glycosylation

Table 2.11. Non-cytosolic proteases of yeasts (**Continued**).

Protease (abbreviation)	Gene	Protease features
Carboxypeptidase	Carboxypeptidase Y (CpY)	<i>PRCI</i> <ul style="list-style-type: none"> • Serine protease • 61-kDa glycoprotein • Possessing “pre-pro” SP • Four N-linked glycosylation • “pro” sequence and Carbohydrate side chains are required for proper folding • Final cleavage takes place in vacuole to give mature CpY
	Carboxypeptidase S (CpS)	<i>CPSI</i> <ul style="list-style-type: none"> • Metal ion-dependent carboxypeptidase • 576-aa membrane-associated precursor • 73- or 77-kDa mature enzyme depending on glycosylation degree • Proteolytic cleavage in vacuole to release soluble mature enzyme • N-linked glycosylation happens
Aminopeptidase	Aminopeptidase I (ApI)	<i>LAP4</i> <ul style="list-style-type: none"> • Metalloexopeptidase • 640-kDa glycoprotein with 12 subunits • Without a normal SP to be translocated into ER
	Aminopeptidase Co (ApCo)	- <ul style="list-style-type: none"> • 100-kDa metalloexopeptidase • Needs Co²⁺ for proper function
	Aminopeptidase Y (ApY)	- <ul style="list-style-type: none"> • 537-aa precursor • 70-kDa glycoprotein • Possessing “pre-pro” SP • Several N-linked glycosylation • “Pro” sequence has an inhibitory effect on enzyme function • Maybe identical to ApCo

Table 2.11. Non-cytosolic proteases of yeasts (**Continued**).

Protease (abbreviation)	Gene	Protease features
Depepdyl aminopeptidase B (DPAP-B)	<i>DAP2</i>	<ul style="list-style-type: none"> • Membrane-bound vacuolar protease • Luminal catalytic domain • Without proteolytic processing until the vacuole • Mostly six or seven N-linked glycosylation occurs, rarely five or eight
Secretory pathway		
Signal peptidase	<i>SEC11</i>	<ul style="list-style-type: none"> • Integral membrane protein • At least four subunits • Catalytic subunit is 18 kDa • Glycosylation of one subunit • <i>SEC11</i> gene codes for catalytic subunit
Kex2 endoprotease	<i>KEX2</i>	<ul style="list-style-type: none"> • Subtilisin family of serine protease • Integral membrane protein • Requires Ca²⁺ for activity • Glycoprotein with both N-linked and O-linked glycosylation • Thought to be at late Golgi
Kex1 carboxypeptidase	<i>KEX1</i>	<ul style="list-style-type: none"> • Integral membrane protein • Glycoprotein with just N-linked glycosylation • Locates at late Golgi • Specific for basic residues
Depepdyl aminopeptidase A (DPAP-A)	<i>STE13</i>	<ul style="list-style-type: none"> • Integral membrane protein • Glycoprotein anchored in the Golgi • Marked similarity to DPAP-B
Yeast aspartyl protease III (Yap3)	<i>YAP3</i>	<ul style="list-style-type: none"> • Upon overproduction and absence of Kex2 shows activity of cleavage • Topologically similar to Kex2 • Has a transmembrane anchor • An extra 45 amino acid segment compared to other aspartyl proteases

2.4. Bioprocess kinetics in fed-batch operation

In order to have a deeper insight to any kind of bioprocess, the major basic parameters such as biomass generation rate, specific growth rate, substrate consumption rate, and production rate, should all be clearly defined to develop a simple mathematical representation of the system behavior (Çalık et al., 2016; Öztürk et al., 2016). By applying mass balance for the bioreactor system according to the equation (2.45), equations describing the major parameters can be obtained.

$$\text{Input} - \text{Output} - \text{Consumption or Generation} = \text{Accumulation} \quad (2.45)$$

In any biological process conducted inside a bioreactor the volume change in the bioreactor can be depicted as (Yamané and Shimizu, 1984):

$$\frac{dV}{dt} = Q(t) + f_{Acid} + f_{Base} + f_{Antifoam} - f_{Vap.} - F_{out} \quad (2.46)$$

Where, $Q(t)$ is the volumetric flowrate of the inlet stream to the bioreactor. In the fed-batch operations, $F_{out} = 0$ as the output stream is not present and also for the volume change caused by addition of acid, base, and antifoam and loss because of vaporization we can assume:

$$f_{Acid} + f_{Base} + f_{Antifoam} - f_{Vap.} \ll Q(t) \quad (2.47)$$

Therefore:

$$\frac{dV}{dt} \approx Q(t) \quad (2.48)$$

Equation (2.48) attributes the bioreactor volume change in fed-batch operation just to inlet stream flowrate. Since the input and output stream is not available in batch mode, in batch operation:

$$\frac{dV}{dt} = 0 \quad (2.49)$$

2.4.1. Cell growth correlations

In batch-wise cultivation, after inoculation of the defined amount of the cell into the fermentation medium, the number of the cells increases by utilization of the provided substrate (mainly carbon source and nitrogen source). Since in the batch operation the input and output streams are both absent, by assuming that there is neither leakage from the system nor consumption of the cells, the mass balance related with the cell around the bioreactor system (by neglecting any kind of cell death and loss) can be written as follows:

$$r_x V = \frac{d(C_x V)}{dt} \quad (2.50)$$

In which, r_x is the volumetric rate of the biomass (cell) generation, V is the volume of the fermentation broth, C_x is the viable cell concentration inside the well-mixed bioreactor, and t is the fermentation time. Considering the definition of the volumetric reaction rate we can deduce the following equation:

$$r_x = \frac{1}{V} \frac{d(C_x V)}{dt} \quad (2.51)$$

Where, the “ $C_x V$ ” is the cell mass in the reaction volume in bioreactor. The reaction volume (V) can be either constant or variable. In addition, based on the definition of specific growth rate, μ :

$$\mu = \frac{1}{(C_x V)} \cdot \frac{d(C_x V)}{dt} \quad (2.52)$$

And therefore:

$$\frac{d(C_x V)}{dt} = \mu(C_x V) \quad (2.53)$$

By combining equation (2.51) and (2.53) it can be written that in the exponential growth phase:

$$r_x = \mu C_x \quad (2.54)$$

Where, the first order kinetic constant is the specific growth rate of the cell, μ . By insertion of the equation (2.54) in equation (2.50) the result will be:

$$\mu C_x V = \frac{d(C_x V)}{dt} \quad (2.55)$$

By applying the derivation into the term inside the brackets in right-hand side:

$$\mu C_x V = C_x \frac{dV}{dt} + V \frac{dC_x}{dt} \quad (2.56)$$

In the case of batch operation, based on the equation (2.49):

$$\mu C_x V = V \frac{dC_x}{dt} \quad (2.57)$$

By dividing both sides to V and rearranging:

$$\mu dt = \frac{dC_x}{C_x} \quad (2.58)$$

Thereafter, by integration between time interval $t=0$ until $t=t$:

$$\int_0^t \mu dt = \int_{C_{x0}}^{C_x} \frac{dC_x}{C_x} \quad (2.59)$$

During logarithmic growth, the specific growth rate is approximately constant (and maximum), thus:

$$\mu t = \ln C_x - \ln C_{x0} \quad \text{or} \quad \ln C_x = \ln C_{x0} + \mu t \quad (2.60)$$

This equation can be rewritten in exponential form:

$$C_x = C_{x0} e^{\mu t} \quad (2.61)$$

On the other hand, in fed-batch mode of the fermentation, although there is no output stream from the system, there is input to the bioreactor (contains sterile substrate) which is the reason of the change in the bioreactor volume, namely:

$$Q(t) = \frac{dV}{dt} \quad (2.62)$$

By insertion of equation (2.62) into the equation (2.56), the result will be as follows:

$$\mu C_x V = C_x Q(t) + V \frac{dC_x}{dt} \quad (2.63)$$

Rearranging the equation (2.63) will lead to the following equation:

$$\frac{dC_x}{dt} = \frac{C_x}{V} (\mu V - Q(t)) \quad (2.64)$$

Multiplication of the term $\left(\frac{C_x}{V}\right)$ by the terms inside the brackets:

$$\frac{dC_x}{dt} = C_x \left(\mu - \frac{Q(t)}{V} \right) \quad (2.65)$$

By solving equation (2.65) for μ , the specific growth rate of the cell in fed-batch fermentation can be represented in the following form:

$$\mu = \frac{1}{C_x} \frac{dC_x}{dt} + \frac{Q(t)}{V} \quad (2.66)$$

By assuming $Q(t)=0$, the batch mode of fermentation, the equation will be converted into:

$$\mu = \frac{1}{C_x} \frac{dC_x}{dt} \quad (2.67)$$

2.4.2. Correlations for substrate utilization

The desired substrate is utilized by the available microorganisms in the bioreactor in order to generate energy for cell. Produced energy is devoted to growth, product formation, and maintenance purposes. Substrate consumption rate can be expressed by the following equation:

$$(-r_s) = q_s C_x \quad (2.68)$$

The minus sign is just for manifestation of the decreasing nature of the rate. The term q_s in this equation represents the specific substrate utilization rate by the cells. Writing mass balance for the substrate in the bioreactor gives the following equation:

$$Q(t)C_{S_0} - (-r_s)V - mVC_x = \frac{d(C_S V)}{dt} \quad (2.69)$$

Where, $Q(t)$ is the volumetric flowrate of the input stream to the bioreactor, mVC_x is the term related to the cell maintenance (m is known as the maintenance coefficient), C_{S_0} concentration of the substrate in the inlet feed stream, C_s is the instantaneous concentration of the substrate in the bioreactor, and V is the working volume of the bioreactor. On the other hand:

$$\frac{d(C_S V)}{dt} = C_s \frac{dV}{dt} + V \frac{dC_s}{dt} \quad (2.70)$$

By ignoring the term peculiar to the maintenance, replacing r_s according to the equation (2.68) and insertion of the equation (2.70) into equation (2.69) the following equation is obtained:

$$Q(t)C_{S_0} - q_s C_x V = C_s \frac{dV}{dt} + V \frac{dC_s}{dt} \quad (2.71)$$

By replacing $\frac{dV}{dt}$ from equation (2.62) and rearranging the equation in order to obtain q_s :

$$q_s = \frac{1}{C_x} \left[\frac{Q(t)}{V} (C_{S_0} - C_s) - \frac{dC_s}{dt} \right] \quad (2.72)$$

If the flowrate of the (growth-limiting) substrate is increased proportional to the exponential growth rate, it will be possible to maintain a high rate of specific growth rate for a longer period of time (Yamané and Shimizu, 1984). Obtaining a correlation for such a flowrate as a function of time will be the next step.

By introducing a new parameter, $Y_{x/s}$, as the yield coefficient of the cell on the substrate we can relate the cell generation rate and substrate consumption rate according to the following equation:

$$Y_{x/s} = \frac{d(C_x V)/dt}{-d(C_s V)/dt} = \frac{r_x V}{(-r_s) V} = \frac{r_x}{(-r_s)} \quad (2.73)$$

In consequence it can be inferred that:

$$(-r_s) = \frac{r_x}{Y_{x/s}} \quad (2.74)$$

By neglecting maintenance and replacing $(-r_s)$ from equation (2.74) in equation (2.69), the obtained equation will be as:

$$Q(t)C_{S_0} - \frac{r_x}{Y_{x/s}} V = \frac{d(C_s V)}{dt} \quad (2.75)$$

By using equation (2.54) and equation (2.70) we get to the following equation:

$$Q(t)C_{S_0} - \frac{\mu C_x}{Y_{x/s}} V = C_s \frac{dV}{dt} + V \frac{dC_s}{dt} \quad (2.76)$$

Based on the equation (2.59) and solving the equation for $\frac{dC_s}{dt}$:

$$\frac{dC_s}{dt} = \frac{Q(t)}{V} (C_{s0} - C_s) - \frac{\mu C_x}{Y_{x/s}} \quad (2.77)$$

In order to have a constant specific growth rate, the concentration of the substrate inside the bioreactor should be constant, i.e., $\frac{dC_s}{dt} = 0$; therefore, equation (2.77) will be changed into the following equation:

$$\frac{Q(t)}{V} (C_{s0} - C_s) = \frac{\mu C_x}{Y_{x/s}} \quad (2.78)$$

In addition, immediate utilization of the entered substrate can lead to the assumption of $C_{s0} \gg C_s$ i.e., a very concentrated substrate inlet stream, the equation will be as follows:

$$\frac{Q(t)}{V} C_{s0} = \frac{\mu C_x}{Y_{x/s}} \quad (2.79)$$

Solving the equation for $Q(t)$ will lead to:

$$Q(t) = \frac{\mu V C_x}{C_{s0} Y_{x/s}} \quad (2.80)$$

In addition based on the definition of the specific growth rate we can write:

$$\mu(C_x V)_t = \frac{d(C_x V)}{dt} \quad (2.81)$$

$(C_x V)_t$ is the representative of the mass of the cell present in the bioreactor in specific time of “t” designated by “ X_t ”. Solving the equation for “X” and representing the initial cell amount in bioreactor as X_0 will give the following equation:

$$X_t = X_0 e^{\mu t} \quad (2.82)$$

In other words the necessary and sufficient conditions for exponential growth in fed-batch mode (where volume change exists) will be as:

$$(C_x V)_t = (C_x V)_{t=0} \cdot e^{\mu t} \quad (2.83)$$

So, by regarding the initial ($t=0$) volume of the bioreactor when semi-batch operation begins (V_0) and the initial concentration of the cells in the bioreactor (X_0) upon fed-batch start, the equation can be written as follows:

$$(C_x V)_t = C_{x_0} \cdot V_0 \cdot e^{\mu t} \quad (2.84)$$

By insertion of equation (2.84) into equation (2.80):

$$Q(t) = \frac{\mu C_{x_0} \cdot V_0 \cdot e^{\mu t}}{C_{s_0} Y_{x/s}} \quad (2.85)$$

The calculated flowrate for inlet stream including substrate can be utilized in order to keep an exponential growth with a pre-defined specific growth rate regarding known amounts of all the parameters in the equation (2.85).

2.4.3. Correlations for product formation

Regarding to the product, the balance equation just includes the accumulation and product formation rate as follows:

$$r_p V = \frac{d(C_p V)}{dt} \quad (2.86)$$

Where, C_p is the product concentration in well-mixed bioreactor in specific time and r_p is the rate of the product formation which can be expressed in the form of a first-order kinetic by introduction of specific product formation rate (q_p):

$$r_p = q_p C_x \quad (2.87)$$

By insertion of the equation (2.87) in equation (2.86) and taking derivatives:

$$q_p C_x V = C_p \frac{dV}{dt} + V \frac{dC_p}{dt} \quad (2.88)$$

Replacing the term ($\frac{dV}{dt}$) from equation (2.62) and solving for q_p :

$$q_p = \frac{1}{C_x} \left[\frac{C_p}{V} Q(t) + \frac{dC_p}{dt} \right] \quad (2.89)$$

2.4.4. Yield coefficients

In order to make the bioprocess easy to understand some parameters are defined which yield coefficients are among them. Yield coefficients clarify the nutritional requirements and organism-specific production characteristics (Doran, 1995) which are typical for each cell-substrate-product system. They can be calculated either as instantaneous parameters (which can vary during cultivation) or as overall parameters. Upon reporting the yields related with a bioprocess, the time or the time interval which the yields belong to should be specified (Doran, 1995).

$Y_{x/s}$, as cell yield on substrate in equation (2.73), is described as the mass of the cell generated per mass of the substrate utilized. $Y_{P/s}$ is defined as the mass of the produced product per mass of the substrate consumed for its production. $Y_{P/x}$ is defined as the mass of the product formed per mass of the cell generated at the same time interval. The instantaneous parameters can be expressed as follows in the infinitesimal time interval “ dt ”:

$$Y_{P/s} = \frac{dP}{dS} = \frac{d(C_p V)/dt}{-d(C_s V)/dt} = \frac{r_p V}{(-r_s) V} = \frac{r_p}{(-r_s)} \quad (2.90)$$

$$Y_{P/x} = \frac{dP}{dX} = \frac{d(C_p V)/dt}{d(C_x V)/dt} = \frac{r_p V}{r_x V} = \frac{r_p}{r_x} \quad (2.91)$$

Where, X , S , and P refer to the mass of the cell, mass of the substrate, and mass of the product, respectively. However, in addition to the instantaneous amounts, we can define overall yield coefficients corresponding to each of the abovementioned yield coefficients between defined time intervals (for example for whole process time, Δt) as constant average amounts:

$$\bar{Y}_{x/s} = \frac{\Delta X}{-\Delta S} \quad (2.92)$$

$$\bar{Y}_{P/s} = \frac{\Delta P}{-\Delta S} \quad (2.93)$$

$$\bar{Y}_{P/x} = \frac{\Delta P}{\Delta X} \quad (2.94)$$

These amounts can be replaced by concentration of each parameter in the bioreactor if the constant volume assumption persists.

2.5. Mathematical modelling of the metabolic network

Cellular metabolism is regarded as the whole network of coupled (bio) chemical conversions (reactions) inside the cell which the taken up materials and energy from the environment undergo (Steuer and Junker, 2009; Maarleveld et al., 2013) with the ultimate aim of maintaining life (Baart and Martens, 2011).

Mathematical models enable interpretation and prediction of any natural phenomenon and experimental result. They are based on a set of relationships between parameters of the system under investigation. Mathematical modeling is a kind of simplification. According to the aim of the study and presence of measurements, different models can be developed for the same phenomenon (Gombert and Nielsen, 2000). The role of the various components of a given system can be unraveled by usage of mathematical modeling. By recruitment of these models, the available interactions between different components of a system can be analyzed quantitatively (Nielsen and Olsson, 2002).

Mathematical modeling of a biological system as a mathematical representation of its behaviour is considered a promising tool in order to gain an in-depth insight related with cellular metabolism and deciphering its interaction with the environmental and process conditions. However, it is actually neither possible nor rational to expect that a single model could describe a complex network of the metabolic reactions (Steuer and Junker, 2009).

To optimize r-protein production, further insights into the cellular metabolism during heterologous gene expression are required; especially the effect of bioprocess parameters such as temperature, oxygen transfer, substrate choice, growth rate, and (foreign) gene expression on metabolism and its interplay with protein synthesis is of interest (Heyland et al., 2010).

By mathematical modelling of an expression system the pivotal cellular events that confine production and secretion and are, therefore, attractive targets for strain engineering in order to reach a host with enhanced desirable traits, can be pinpointed (Love et al., 2012). In consequence, striking enhancements can be achieved in the systems productivity by developing mathematical models by which the behavior of the system is accurately described (Çelik et al., 2009). Modelling will lead to prediction of the cell behaviour under different environmental and/or genetic conditions. In addition, by mathematical description of the fermentation processes (virtually cellular metabolism) and its subsequent comparison with experimentally-obtained data the performance of the model can be evaluated by assessing the consistency of the experimental data (Stephanopoulos et al., 1998; Nielsen and Olsson, 2002); the accuracy of the predictions will be greatly dependent on the underlying assumptions (hypotheses) for the developed model (Maarleveld et al., 2013).

In the field of cellular metabolism, mathematical models can be divided into stoichiometric and kinetic models. The former is based on the steady state (time invariant) properties of the metabolic networks, and the latter, in contrast to the former, deals with combined known stoichiometry and kinetics of the metabolic pathways and describes dynamic features of the network. Both stoichiometric and kinetic models have been used to investigate the cell metabolism (Gombert and Nielsen, 2000).

2.5.1. Kinetic (dynamic) modelling

A mathematical model for reaction rate, the kinetic expressions, relates the reaction rates with the variable parameters of the system. The model which is obtained by combining stoichiometric equations and kinetic expressions is called a kinetic model (Stephanopoulos et al., 1998). Kinetic models in the area of metabolic modeling can be classified as (simple) unstructured models and (advanced) structured models. In unstructured models the cell is regarded as an entity (black box) which exchanges material with the

extracellular environment; what happens inside is not considered and all the cellular metabolic reactions have been lumped into one that is the overall biomass growth (Stephanopoulos et al., 1998; Gombert and Nielsen, 2000). It is obvious that the ability of the method is limited in prediction of the cell responses in different cultivation conditions as the intracellular metabolism is neglected. The structured models approach incorporates more details of the intracellular metabolism of the cell and gradual inclusion of more structures can lead to even non-linear models (Gombert and Nielsen, 2000). These models are more realistic and mathematically more evolved (Stephanopoulos et al., 1998).

For growth of the recombinant *P. pastoris* on methanol an unstructured model, has been developed; the model describes the relationship between specific growth rate and methanol concentration and also relationships between specific methanol and ammonium consumption rates with specific growth rate (Zhang et al., 2000). A kinetic model in order to explain the effects of limiting factors on cell density in fed-batch cultures of *P. pastoris* has also been proposed (Jahic et al., 2002); the model developed could predict cell growth and oxygen consumption in processes with and without oxygen-enriched air. An unstructured model based on mass balance in fed-batch cultures of recombinant *P. pastoris* which relates total cell and induction time in both batch and fed-batch phases has also been introduced (Pais et al., 2003). The relationship between the specific growth rate and r-protein production in chemostat cultures of *P. pastoris* has also been investigated in single carbon source cultures (glycerol or methanol) which has led to development of another unstructured model (Jungo et al., 2006). The first structured kinetic model for growth of *P. pastoris* and recombinant human erythropoietin production in fed-batch fermentation has been proposed by (Çelik et al., 2009). It was shown that the key bottleneck in r-protein secretion is the capacity of the secretory machinery to transport folded protein out of the ER. Subsequently, a simple mechanistic model

describing the flux of folded protein through the ER in order to further supporting the experimental observations was proposed (Love et al., 2012).

2.5.2. Stoichiometric modelling and metabolic flux analysis

By considering the microorganism as the microbioreactor in the fermentation system, the state of the fermentation process depends specifically on the state of the cell (Çalık and Özdamar, 2002). The cells and organisms exist in a dynamic steady state but not at equilibrium with their surroundings (Nelson and Cox, 2005). Upon introduction of external stimuli and perturbations to a living system, the persisting steady state is disturbed and intracellular fluxes in metabolic pathways are transiently changed. Afterwards, the regulatory mechanisms, peculiar to each pathway, are triggered and, consequently, the system is brought to a new steady state to retrieve homeostasis (Stephanopoulos et al., 1998; Nelson and Cox, 2005). Metabolic fluxes are one of the major parameters by which the degree of engagement of each of the intracellular pathways in the whole cell function can be understood and, thus, the cellular physiology can be determined (Stephanopoulos et al., 1998); therefore, bioreaction network analysis, or shortly metabolic flux analysis (MFA) is required for the quantitative analysis of the cell metabolism which is controlled by the cell physiology (Çalık and Özdamar, 2002).

On the other hand, in order to reach a thorough understanding of a cell physiological behaviour and production capability via mathematical modelling, the expected pre-requisite is the availability of metabolic network which represents its cellular metabolism and the comprehensive interaction between the various pathways inside the cell (Förster et al., 2003); MFA is helpful to analyse the metabolic characteristics of the annotated genomes in order to analyse, interpret, and predict the relationship between the genotype and phenotype (Çalık and Özdamar, 2002). Thanks to the innovations in the genome sequencing techniques, the release of sequenced genomes have been accelerated and, in consequence, the available knowledge in gene-level and

protein-level have obtained the potential of being consolidated. As a consequence, genome-scale metabolic models/reconstructions which can describe the cell overall function as a cumulative result of the interaction of its individual components and, thus, present a holistic approach towards cellular metabolism of the organism has been developed (Förster et al., 2003; Balagurunathan et al., 2012). The output of such reconstruction procedure will be a set of biochemical reactions which can be incorporated into the construction of a stoichiometric model of cellular metabolism recruiting metabolite balance (Förster et al., 2003). Simply, the genome scale models (GSMs) play the role of a bridge connecting genotype and phenotype of an organism (Balagurunathan et al., 2012). In contrast to a group of enzymes or isolated pathways which are not capable of efficiently representing and describing the whole system because of the limited range of action and coverage, the “genome-wide scale” of the GSMs has enabled them to assess the whole response and behaviour of the cell confronting any kind of perturbation (Balagurunathan et al., 2012). The first GSM reconstructed belongs to Gram negative bacterium *Haemophilus Influenza* (Edward and Palsson, 1999). Regarding the whole-cell network covered by GSMs, metabolic engineering can be guided with them for adopting optimal strategies in order to reach desirable properties in the selected organism; the potential targets for deliberate manipulation of the genome (metabolic engineering) which lead to strain improvement can be rationally identified with the aid of characterization of cellular metabolism as the outcome of the GSMs (Oberhardt et al., 2009; Sohn et al., 2010; Chung et al., 2010). The optimization of the predictive capabilities of the GSMs which is reflected in their close reproduction of experimental data can be achieved by augmentation of available data about the organism (Caspeta et al., 2012) and, thus, casting light on the dark parts of the cellular metabolism.

In the case of *P. pastoris* three GSMs have been reconstructed after its genome became sequenced (Sohn et al., 2010; Chung et al., 2010; Caspeta

et al., 2012) and by their subsequent utilization the in-depth analysis of the *Pichia*-based systems has been more feasible.

Stoichiometric models of the cellular metabolism are derivatives of the reconstructed metabolic network of the organism and the stoichiometries of the reactions are the base for model development. Furthermore, in these models the “internal dynamic behaviour” of the cell is overlooked (Llaneras and Picó, 2008). By assuming a metabolic network consisting of “*m*” metabolites and “*n*” reactions, a stoichiometric matrix can be generated which its rows are the metabolites and its columns are the corresponding stoichiometric coefficients of each metabolite in each metabolic reaction and depends on the biochemistry of the microorganism:

$$\mathbf{A} = [\dots]_{m \times n} \quad (2.95)$$

This matrix interprets the available biological knowledge about the cellular metabolism into a mathematical term. Afterwards, a mass balance for intracellular metabolites can be written as follows (Çalık et al., 1999; Çalık and Özdamar, 2002):

$$\mathbf{A} \times \mathbf{r}(t) = \mathbf{C}(t) \quad (2.96)$$

In which, “ $\mathbf{r}(t)$ ” is the vector composed of the fluxes of the intended “*n*” metabolic reactions with “*n*” rows and “1” column and “ $\mathbf{C}(t)$ ” is the metabolites accumulation vector possessing “*m*” rows” and 1 column. The elements of $\mathbf{C}(t)$ are divided into two sub-vectors:

$$\mathbf{C}(t) = \mathbf{C}_1(t) + \mathbf{C}_2(t) \quad (2.97)$$

Where, $\mathbf{C}_1(t)$ and $\mathbf{C}_2(t)$ correspond to the extracellular and intracellular metabolite accumulation vectors, respectively. Equation (2.96) represents dynamic mass balance and is an indicator of the time-course

evolution of the metabolites and refers to the role of the cellular metabolism stoichiometry (\mathbf{A}) and the fluxes of the metabolic reactions for determination of the concentration profiles of the metabolites. The very fast dynamics of the intracellular reactions makes the transient state of the cell too short to be sensed; therefore, the pseudo steady state assumption comes into play, i.e., $\mathbf{C}_2(t) = 0$ in equation (2.97) and, thus:

$$\mathbf{A} \times \mathbf{r}(t) = \mathbf{C}_1(t) \quad (2.98)$$

A vectorial equation including “ n ” unknown (fluxes) and “ m ” equations where most often $n > m$; thus, there are “ $n-m$ ” degrees of freedom for equation and in this state system is underdetermined. Infinite set of answers can be found to fit in the equation (2.98).

By imposing some constraints on the system, the system can possess unique and particular results (fluxes); these results can be either related to the current situation of the system (phenotype determination) or be a prediction of the system behaviour as a predictive model.

If $m > n$, the solution of the model gives the exact solution. The solution of equation (2.98) can be determined by a constrained least-squares approach with the objective of minimizing the sum of squares of residuals from the stoichiometric mass balance. On the other hand, if $m < n$, metabolic flux distributions can be obtained by minimizing or maximizing the objective function, whereupon the best metabolic pathway utilization that would fulfill the stated objective is obtained. In this case, the mathematical formulation for the objective function Z is:

$$Z = \sum \alpha_i r_i \quad (2.99)$$

Where Z is a linear combination of the fluxes r_i , and α_i is the coefficient of the component- i in the stoichiometric equation of the corresponding reaction; the metabolic flux distributions are obtained either by minimizing or

maximizing the Z in the model. In least-squares method, the matrix is solved and a unique solution is obtained (Çalık and Özdamar, 2002).

As an important tool of the biochemical reaction engineering (Çalık and Özdamar, 2002) for the application of MFA, after selection of “major intracellular reactions”, one can construct mass balance equations for every intracellular metabolite, and obtain the general equation in the form of Eq-2.96, where a set of measured (known) extracellular fluxes are, then, used as inputs in Eq. 2.97 whereupon in the model in the form of Eq. 2.98 (constraints). The compounds represented by the column vector $C_1(t)$ are the exchanged compounds between the cell and its environment such as substrate uptake rate and metabolites, e.g., amino acids and organic acids excretion and extracellular proteins secretion rates. The outcome will be an estimation of the reaction rates (fluxes) of the biochemical reactions in the reaction network (Stephanopoulos et al., 1998; Çalık and Özdamar, 2002) through the different branches of the pathways, namely metabolic flux map (Gombert and Nielsen, 2000). Metabolic flux map obtained can be a comprehensive representative of “contribution of various pathways” in the overall processes comprised in substrate utilization toward product formation. However, such metabolic flux maps are of utmost importance in the case of flux differences obtained under different environmental or genetic conditions leading to full assessment of the impact of genetic and environmental perturbations. The important pathways and reactions will be determined by the aid of such comparisons (Stephanopoulos et al., 1998).

Metabolic engineering basically aims to drive the process in a path which conduces to as much useful product as possible from the substrate by redirecting the fluxes towards the desired end-product.

Regarding the MFA approach in *P. pastoris*, various researches have been carried out by concerning different objectives. A stoichiometric model for a recombinant *P. pastoris* by describing balances of some key metabolites, ATP, and NADH during glycerol and methanol metabolism pathways have been described; model could predict the cell growth and the

protein production with acceptable accuracy (Ren et al., 2003). Using MFA the central carbon metabolism and amino acid biosynthesis of *P. pastoris* were analyzed by considering only glycerol as substrate; it was reported that, amino acids synthesis and regulation of central carbon metabolism in *P. pastoris* are similar with those in *S. cerevisiae* (Sola et al., 2004). In another study using chemostat cultures and inclusion of methanol utilization pathway the MFA calculations were performed in recombinant *P. pastoris* expressing *Rhizopus oryzae* lipase (Sola et al., 2007). By MFA, the physiological adaptation of the wild type and recombinant *P. pastoris* X-33 expressing an antibody fragment to oxygen limitation conditions (including central carbon metabolism pathway) in glucose-limited chemostat culture was investigated (Baumann et al., 2010). Furthermore, the impact of the foreign gene expression on the metabolism of recombinant *P. pastoris* in batch and fed-batch cultures utilizing glucose as sole carbon and energy source was inspected by analyzing intracellular flux distribution quantified by MFA (Heyland et al., 2010). Besides, MFA has been used to investigate the effects of the different feeding rates and dual carbon sources on the intracellular reaction rate (flux) distribution in the synthesis of recombinant human erythropoietin by *P. pastoris* in fed-batch cultivation (Çelik et al., 2010). A constraint-based MFA using a developed GSM for wild type *P. pastoris* X-33 has been carried out and validated by two chemostat cultures with glucose minimal medium and mixed glycerol/methanol medium. The model was used in silico to predict the best candidate as carbon source for r-protein production in *P. pastoris* (Chung et al., 2010). Based on the knowledge that heterologous protein expression causes metabolic burden on the host organism, the effect of different levels of protein production on *P. pastoris* metabolism in batch cultures was compared; *P. pastoris* compensates for the additional resources required for r-protein production by reducing by-product formation and increasing energy generation via the TCA cycle (Heyland et al., 2011). The effect of methanol feeding rate on the bioreaction network of the host *P. pastoris* has been the parameter analyzed by the aid of MFA to

improve rhGH production (Çalık et al., 2011). MFA approach was also used to obtain metabolic flux distribution in central carbon metabolism of recombinant *P. pastoris* X-33 growing on mixed feed of glucose and methanol. The MFA revealed redistribution of carbon fluxes in the central carbon metabolism as a result of protein production and metabolic burden (Jorda et al., 2012). The r-protein production by *P. pastoris* was estimated using a constraint-based model by inclusion of metabolic network reactions stoichiometric description as well as additional restrictions such as reaction irreversibilities, thermodynamics and regulations (Tortajada et al., 2012); These types of models do not assume pseudo-steady state and describe accumulation of important metabolites and, furthermore, are not purely stoichiometric approach such as traditional MFA. In this study, protein production was related to the cell growth and ATP production rate obtained from MFA using main biochemical pathways of *P. pastoris*. In another research, by instationary ¹³C-labeled MFA approach which is different from NMR-derived stationary state labeling, the metabolic fluxes through the major pathways under mixed feed (co-assimilation) of glucose and methanol in chemostat cultures of *P. pastoris* X-33 harboring a mock plasmid were accurately investigated (Jorda et al., 2013). The intracellular flux distribution during continuous cultures of recombinant *P. pastoris* in mixed feeding medium of methanol/sorbitol (co-feeding process) with linear change of methanol was quantified by MFA recruiting a simplified metabolic network. The study demonstrated an optimum mixture for reducing specific oxygen uptake in one hand and increasing induction on the other hand (Niu et al., 2013). Very recently, the metabolic flux analysis was utilized in order to verify the predicted fluxes by in-silico model after incorporation of the r-protein formation reaction to one of the available GSMs (Sohn et al., 2010) for *P. pastoris* (Nocon et al., 2014).

2.6. Recombinant hGH production

Since hGH is a non-glycosylated protein, a prokaryotic platform is the first preference (Mukhija et al., 1995). The annual production of this hormone by conventional bacterial culture has been reported to be 4.4 kg (Salamone et al., 2006). Regarding to the utilized host system, the production in extracellular medium, as IBs (insoluble protein aggregates), and in periplasmic space (using a SP) would be possible. Although, being as IBs lead to straight-forward isolation and also protection from proteolytic attack (Kim et al., 2013), the IBs need to be solubilized and renatured before purified from the endogenous proteins of host (Mukhija et al., 1995); the overall yield of bioactive expressed protein after solubilization and refolding of IBs is generally low and also the aforementioned procedures are very expensive and time-consuming and need to be optimized to increase the overall yield (recovery) (Kim et al., 2013). However, in bacteria the synthesis of protein is initialized by formylmethionine which in many cases is not processed efficiently and, thus, in the case of usage for treatment can lead to antibody formation in patient (Shin et al., 1998). Among the bacterial hosts, the use of *E. coli* as a platform for expression of rhGH has been documented with higher triumph rates (Soorapaneni et al., 2010). In 1984, a conducted research was reported to express rhGH in *Pseudomonas aeruginosa* and secrete to the periplasm with an authentic N-terminus (Gray et al., 1984). In general, researches which have utilized *E. coli* as host organism hitherto have focused on wide range of parameters in order to enhance expression levels and analyze the effect of different parameters on rhGH production; most important selected parameters scrutinized in some selected researches can be mentioned as: different types of bioreactors either shake bioreactors or pilot-scale bioreactors, different fermentation modes, different fermentation parameters like culture medium, dissolved oxygen concentration, different promoters for driving expression, different leader sequences for facilitating secretion of rhGH to the extracellular medium, utilization of the fusion partners especially for aiding purification,

improvement of the solubilization of IBs and protein recovery, investigating purification steps by special emphasize on chromatographic methods (Mukhija et al., 1995; Uchida et al., 1997; Shin et al., 1998 ; Zhang et al., 1998; Patra et al., 2000; Ribela et al., 2000; Castan et al., 2002; Tabande et al., 2004; Young and Park, 2007; Shang et al., 2009; Soorapaneni et al., 2010; Rezaei and Zarkesh-Esfahani, 2012; Kim et al., 2013). *Bacillus subtilis*, as another eukaryote organism, has also been paid attention in order to express rhGH although not as much as *E. coli* (Honjo et al., 1986; Nakayama et al., 1988; Özdamar et al., 2009).

Although, the most reports on heterologous production of hGH are related with *E. coli* as host, but, the presence of formylmethionine in the N-terminus and its poor removal in many cases (Kadono-Okuda et al., 1995) could be a good motivation to express the hormone in another host (specifically eukaryotic) to overcome this obstacle. Higher organisms based on their developed and more advanced machinery for PTMs may be regarded as suitable candidates for production of complex proteins and, thus, have provided us with an efficient and economic method in order to produce biopharmaceuticals (Dyck et al., 1999). Transgenes including animal, plant, or insect along with non-transgenic animals each have been utilized as an expression host in rhGH production. Transgenic animals can led to large-quantity and high-quality end-products especially in secreted form in their body fluids (Zhu et al., 2003); secretion in urine, seminal fluid, blood, and milk has been reported widely (Kerr et al., 1998; Dyck et al., 1999; Zhu et al., 2003; Lipinski et al., 2003; Salamone et al., 2006; Lipinski et al., 2012). However, by taking into consideration the cumbersome procedure of production in milk and long time to express and also health risk (Hen et al., 2009; Sanchez et al., 2004), transient transfection of mammary epithelial cells, and developing non-transgenic expression host (Hens et al., 2000; Sanchez et al., 2004; Han et al., 2009), has been adopted as an alternative method. Plants, as the least expensive biomass (Cunha et al., 2011), have also been used as bioreactors for production of rhGH (Leite et al., 2000;

Russell et al, 2005; Kim et al., 2008; Cunha et al., 2011; Xu and Kieliszewski, 2011) because of providing a native structure for foreign protein. The observed level of production in plant systems has been lower than microbial systems such as *E. coli* and *P. pastoris* (Russell et al., 2005) and this low productivity as a result of truncated proteins accumulated has been the main drawback of these systems (Kim et al., 2008). Restriction of the production to a specific organ can be one of the most appropriate strategies in order to conquer protein degradation and maximizing protein accumulation (Cunha et al., 2011). Likewise, insect cells have been recruited in rhGH production mainly by considering their high-level PTMs and fast growth (Kadono-Okuda, 1995; Markaki et al., 2007).

About 60-70% of all recombinant pharmaceutical proteins are produced in mammalian cells (Rezaei et al., 2013). Rendering the protein more similar to native equivalent by PTMs and having a final product with high quality are the major factors play role in selection of animal cells as host organisms (Kunert et al., 2009). In comparison with prokaryotic cells the productivity of mammalian cells is low and there should be strategies to optimize these systems as much as possible by manipulating parameters such culture medium composition, cultivation temperature, engineering of bioreactor (Rezaei et al., 2013). The most widely used mammalian cell line, CHO cell line, has also been recruited for expression of rhGH in selected researches (Keane et al., 2002; Catzel et al., 2003; Kunert et al., 2009; Rezaei et al., 2013).

Yeasts, as dual-character microorganisms, have the ability of bacterial fast growth and higher-eukaryote-like PTMs. There have been reports related with production of rhGH mainly in yeast *S. cerevisiae* (Hiramatsu et al., 1990; Hahm and Chung, 2001) and *P. pastoris*. The selected researches related with rhGH production by *P. pastoris* along with their important features and outcomes have been summarized in Table 2.12.

Table 2.12. Summary of the selected researches conducted related with rhGH production using *P. pastoris*.

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> • Expression under P_{AOX1} • Secretion by α-MF • Fermentation in shake tubes and 2-L bioreactor • DO > 20% 	<p>11 mg/L (shake tube)</p> <p>49 mg/L (bioreactor)</p>	<ul style="list-style-type: none"> • Shake tubes culture was O₂-limited • C_x=73 gDCW/L • Lower percentage of hGH in bioreactor than in small scale mainly due to the higher cellular disruption in bioreactor • hGH was genuine regarding to size, amino acid sequence, immunological properties, and bioactivity • Biological activity was checked by adipogenic activity; capacity to induce the differentiation in one-day confluent cultures of 3T3-F442A/C4 cells into adipocytes • Insertion of hGH gene in <i>Pichia</i> genome by single-cross over at <i>HIS4</i> or <i>AOX1</i> 	<p>(Ecamilla-Trevino et al., 2000)</p>
<ul style="list-style-type: none"> • Investigating the effect of "EA" spacers in removal of α-MF in the secretory expression of hGH • Synthetic hGH gene according to codon preference of <i>E. coli</i>, <i>S. cerevisiae</i>, and <i>P. pastoris</i> • Expression under P_{AOX1} • Secretion by α-MF • Three constructs were made: one "EA" repeat, two "EA" repeats, without spacer 	<p>190 mg/L (optimized condition)</p>	<ul style="list-style-type: none"> • According to N-terminal sequencing, removal of spacer was not efficient in presence of spacer and just the construct without spacer led to complete cleavage of SP and genuine N-terminus • The optimum amount of methanol for induction was obtained to be 3% (v/v) • Addition of extra glutamic acid and alanine is not required for facilitating cleavage by Kex2 protease 	<p>(Eurwilachit et al., 2002)</p>

Table 2.12. Summary of the selected researches conducted related with rhGH production using *P. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> • Expression of <i>GH-V</i>, placental human growth hormone where cDNA obtained by RT-PCR • Expression in falcon • Expression under P_{AOX1} • Secretion by α-MF (without spacers) 	<p>~ 5 mg/L (15% of total protein in supernatant)</p>	<ul style="list-style-type: none"> • Biological activity was tested by Nb2 proliferation assay • According to Invitrogen, the double spacer "EAEA" is not required for cleavage by Kex2 when the amino acid after "DKR" (last ones in α-MF) is "F" like the case in <i>GH-V</i> protein 	<p>(Palma-Nicolas et al., 2005)</p>
<ul style="list-style-type: none"> • Developing a system for expression and purification • Fusion protein was expressed: His-tag with Factor Xa cleavage sequence with hGH • Shake flask experiments • Expression under P_{AOX1} • Secretion by α-MF • Ultra-filtration was used • Affinity purification was recruited • Digestion by Factor Xa protease • Analysis by "MOLDI ToF MS" 	<p>115 mg/L (original product, with 33% purity)</p> <p>65 mg/L (after affinity purification and Factor Xa digestion, with 88% purity)</p>	<ul style="list-style-type: none"> • Cx=5.74 g/L after t=24 h • Proper processing of SP regarding authentic N-terminus • Using dot blot for election of best expressing transfectant • Western-blot showed two hGH bands in supernatant • After digestion of purified protein with Factor Xa, the obtained result from MOLDI ToF was reasonable 	<p>(Çalik et al., 2008)</p>
<ul style="list-style-type: none"> • High-yielding process by methanol induction using surfactants during fermentation • Non-codon-optimized hGH • Developing multi-copy strain of <i>P. pastoris</i> • Addition of Tween-20 and Tween-80 during induction 	<p>500 mg/L (enhancement by Tween-20 addition)</p>	<ul style="list-style-type: none"> • Enhanced production in presence of Tween-20 in bioreactor • The obtained hGH band was checked in terms of immuno-reactivity, N-terminal sequence and length 	<p>(Apte-Dashpande et al., 2009)</p>

Table 2.12. Summary of the selected researches conducted related with rhGH production using *P. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> Studying the influence of carbon source on rhGH production in two methanol utilization phenotypes of <i>P. pastoris</i> expression Methanol in defined (solely) and complex media (BMMY without glycerol) and methanol/glycerol mixed defined medium was used Shake flask experiments were conducted Expression under P_{AOX1} Secretion by α-MF 	<p>Sole methanol in defined medium: 3.2 mg/L (Mut⁺; C_{Meth}=3%) Not detectable (Mut⁻)</p> <p>Methanol in complex medium (C_{Meth}=2%): 5.2 mg/L (Mut⁺) 1.60 mg/L (Mut⁻)</p> <p>Glycerol/methanol mixed defined medium (C_{Gly}=3.0 g/L): 60 mg/L (Mut⁺; C_{Meth}=4%) 1.10 mg/L (Mut⁻; C_{Meth}=1%)</p>	<ul style="list-style-type: none"> The results encourage utilization of Mut⁺ strain for rhGH production Methanol in complex medium was inhibitory at C_{Meth} ≥ 2% for both phenotypes μ_{max}=1.4 h⁻¹ for Mut⁺ sole methanol C_{Meth} ≤ 2% in defined medium In complex medium with methanol both phenotypes have higher cell concentration, compared to defined medium, due to the rich medium available Expression of rhGH with BMMY (no glycerol) and Mut⁺ is more desirable due to the slow utilization of methanol by this strain 	<p>(Orman et al., 2009)</p>
<ul style="list-style-type: none"> Investigation of pH influence on rhGH production Expression with Mut⁺ strain Expression levels of <i>AOX1</i> and <i>hGH</i> and some protease genes were quantified by qPCR Performing preliminary shaker experiments related with pH range Conducting bioreactor experiments Expression under P_{AOX1} Secretion by α-MF Calculating oxygen transfer characteristics 	<p>2.70 mg/L (pH=5.0)</p>	<ul style="list-style-type: none"> Keeping pH constant at optimum value (for hGH) of 5.0 inferred to be vital Protease secretion increased with pH increase Higher protein syntheses led to higher amino acid demand and, thus, lower cell growth at pH=5.0 Max K_{La} value obtained as 0.14 s⁻¹ at pH=5.0 Highest rhGH expression level was obtained at t=15 h There was a time shift between synthesis and transportation of rhGH The highest OUR was obtained at pH=5 due to the high specific utilization rate of the substrate methanol 	<p>(Çalik et al., 2010a)</p>

Table 2.12. Summary of the selected researches conducted related with rhGH production using *F. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> Investigating methanol feed rate on rhGH production in presence of sorbitol at different three μ values Expression with Mut^r strain Four-stage fermentation was accomplished. Glycerol batch, glycerol fed-batch, methanol pre-induction (methanol was added at once), and production phase (batch-wise addition of sorbitol and simultaneous exponential addition of methanol) Sorbitol batch and methanol fed-batch operation mode Accumulation of the sorbitol does not affect the r-protein production in expression under P_{AOX1} Preliminary shake flask experiments were conducted for finding highest non-inhibitory amount of sorbitol 	<p>270 mg/L ($\mu=0.03 \text{ h}^{-1}$)</p>	<ul style="list-style-type: none"> Maximum $C_x=48 \text{ g/L}$ at $\mu=0.04 \text{ h}^{-1}$ Amount of rhGH increased with cultivation time Higher amount of rhGH at $\mu=0.03 \text{ h}^{-1}$ was the result of lower extracellular protease availability and higher AOX1 activity OUR increased with increasing μ K_{La} had tendency to increase with μ Max K_{La} value obtained as 0.15 s^{-1} at $\mu=0.04 \text{ h}^{-1}$ Deleting sorbitol from the medium led to decrease in rhGH amount in $\mu=0.03 \text{ h}^{-1}$ Apparent viscosity of the fermentation medium affected K_{La} The secretion of proteins and metabolites led to limiting the contact surface between gas bubbles and cell surface, thus, oxygen transfer is decreased Mass transfer resistances dominated the biochemical reaction resistances. High O_2 demand cannot be satisfied by OTR_{max} due to the high methanol feed rate 	(Çalik et al., 2010b)

Table 2.1.2. Summary of the selected researches conducted related with rhGH production using *P. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> • Investigating influence of methanol feeding rate on intracellular reaction network of <i>P. pastoris</i> • Determining potential strategies to improve rhGH production by <i>P. pastoris</i> • Three different μ was checked • Using intracellular reaction network constructed by (Çalik et al., 2010) by inclusion of rhGH synthesis reaction: 102 metabolites and 141 reactions • Assumption: minimum in vivo rhGH accumulation rate • Obtaining metabolic flux distribution from the stoichiometric model • Obtaining metabolic flux distribution at principal nodes (Pyr and G3P) which flux partitioning in the node affects r-protein yield 	<p>270 mg/L ($\mu=0.03 \text{ h}^{-1}$) (Synthesis fluxes have been presented in different methanol feeding rates)</p>	<ul style="list-style-type: none"> • $\mu=0.03 \text{ h}^{-1}$ conditions seems to be favourable based on effective operation of bio-reaction network • Investigated nodes (G3P and Pyr) were obtained to be flexible • Well-defined perturbations were achieved by different methanol feeding rates (obtained by three different μ) • Based on the central metabolism, biomass synthesis, and rhGH synthesis fluxes, the strategy should be starting induction in a manner to have $\mu=0.03 \text{ h}^{-1}$ and from the declining part of the exponential growth phase (μ decrease) keeping with $\mu=0.02 \text{ h}^{-1}$ 	<p>(Çalik et al., 2011)</p>

Table 2.1.2. Summary of the selected researches conducted related with rhGH production using *P. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> • Investigating the effect of different feeding strategies of co-substrate sorbitol in production phase • Expression with Mut^r strain • Experimental groups were designed: just methanol fed-batch feeding, methanol fed-batch feeding with sorbitol pulse feeding, and fed-batch feeding of methanol and sorbitol together (0<t<15) then, methanol semi-batch 	<p>With the best strategy, the maximum amount: 640 mg/L at t=42 h</p>	<ul style="list-style-type: none"> • Feeding strategy of sorbitol, as cp-carbon source, is as important as methanol feeding rate strategy in r-protein production • Sorbitol feeding led to higher cell concentrations • C_{max}=105 g/L at t=42 h • μ was pre-determined amount kept with methanol feeding • In pulse sorbitol feeding, AOX activities increased until depletion of sorbitol and then decreased • Constant sorbitol period led to higher cell than pulse feeding • For all strategies, leucine, serine, glutamic acid and glutamine were not found in the medium • When cells grow on sole methanol and at high feeding rates, there was no lactic acid in the medium and, so, O₂ was utilized mainly in dissimilation pathway 	<p>(Çalik et al., 2013)</p>

Table 2.1.2. Summary of the selected researches conducted related with rhGH production using *P. pastoris* (Continued).

Features of Study	rhGH Amount	Outcome(s)	Reference
<ul style="list-style-type: none"> • Investigating novel methanol and co-substrate feeding strategies • Expression with Mut^r strain • Feeding strategies by inclusion of both mannitol and sorbitol as co-substrates and also by keeping methanol concentration in the bioreactor in constant amount of 5 g/L with continuous feeding of methanol 	<p>Methanol-stat strategy, the maximum amount:</p> <p>Without sorbitol, 1.3 g/L at t=5.4 h</p> <p>With sorbitol, 1.2 g/L at t=42 h</p>	<ul style="list-style-type: none"> • Mannitol and sorbitol both enhance cell growth in <i>Pichia</i> fermentations • Mannitol as co-substrate instead sorbitol in exponential feeding led to higher cell formation but was not as favourable as sorbitol regarding rhGH production • $C_{\text{max}}=1.57$ g/L at t=42 h with mannitol • For rhGH production under P_{Aox}. The methanol-stat feeding strategy with co-substrate sorbitol obtained as the most favourable fed-batch strategy • $C_{\text{max}}=1.34$ g/L at t=60 h with methanol-stat strategy in presence of sorbitol • $C_{\text{max}}=82$ g/L at t=5.4 h with methanol-stat strategy without sorbitol 	<p>(Güneş et al., 2016)</p>

Notes:

Abbreviations related with Figure 2.8

GLC: Glucose, G6P: Glucose-6-phosphate, F6P: Fructose-6-phosphate, F16P: Fructose-1,6-biphosphate, GAP: Glyceraldehyde-3-phosphate, DHAP: Dihydroxyacetone phosphate, G3P: Glycerol-3-phosphate, GLYC: Glycerol, 13PG: 1,3-bisphosphoglycerate, 3PG: 3-phosphoglycerate, 2PG: 2-phosphoglycerate, PEP: Phosphoenolpyruvate, PYR: Pyruvate, G15L: Glucono-1,5-lactone-6-phosphate, P6G: 6-phosphogluconate, RU5P: Ribulose-5-phosphate, R5P: Ribose-5-phosphate, X5P: Xylulose-5-phosphate, S7P: Sedoheptulose-7-phosphate, E4P: Erythrose-4-phosphate, ACA: Acetaldehyde, ACE: Acetate, EtOH: Ethanol, ACCOA: Acetyl-CoA, MeOH: Methanol, FormAl: Formaldehyde, Form: Formate, CIT: Citrate, ICI: Isocitrate, AKG: Ketoglutarate, SUCC: Succinyl-CoA, SUC: Succinate, FUM: Fumarate, MAL: Malate, OAA: Oxaloacetate, MannP: Mannitol Phosphate, SorbP: Sorbitol Phosphate, GLYOX: Glyoxylate, cyt: Cytoplasmic, OAA_i: Oxaloacetate transporter.

CHAPTER 3

MATERIALS AND METHODS

3.1. Substances and Stuffs

3.1.1. Enzymes, kits, and markers

All the restriction endonuclease enzymes, DNA ladders, and kits utilized in plasmid isolation, gel extraction, and PCR purification were procured from ThermoScientific. The total RNA isolation kit, cDNA synthesis kit, and SYBR Green were purchased from Roche Diagnostics. The pre-stained protein ladder used in SDS-PAGE analyses was also from ThermoScientific.

3.1.2. Chemicals

As the major constituents of the growth media, dextrose from DIFCA; Bacto™agar and Bacto™peptone; yeast extract, LB, and LB agar from Sigma; glycerol from Sigma-Aldrich were utilized.

Among the other compounds, sodium chloride and EDTA was from Riedel-de Haën. Sodium thiosulfate pentahydrate was from Fluka. All the remaining chemicals were purchased from Sigma and Merck. Bradford reagent utilized in total protein assays was also a product of Sigma.

3.1.3. Antibiotics

Zeocin™ was from InvivoGen and was stored at -20°C. Ampicillin (as sodium salt powder) was obtained from Sigma-Aldrich and was kept in +4°C. Chloramphenicol (as powder) was product of Sigma and was placed at room temperature.

3.1.4. Immunoassay compounds

Anti- growth hormone (GH-1) antibody (ab9821) as primary antibody in Western blot and dot-blot analyses, goat polyclonal secondary antibody to mouse IgG with HRP (ab6789) as secondary antibody antibody in Western blot and dot-blot analyses, and 3, 3'-Diaminobenzidine (DAB) all were purchased from abcam. Antibodies were stored at -20°C and DAB was kept at +4°C.

3.1.5. Buffers and stock solutions

All the required buffers and stock solutions were prepared with ultra-pure distilled water (dH₂O) with electrical resistivity of 18.3 MΩ.cm. They were autoclaved at 121°C for 20 minutes or alternatively, whenever needed, were filter-sterilized with 0.2 µm filter (Minisart, Sartorius stedim biotech, Goettingen, Germany). They were stored at room temperature, +4°C, or -20°C based on the instructions available in Appendix B. Phosphate buffered saline (PBS) 10X (pH ~7.4) was purchased as ready solution from Sigma and was stored at room temperature.

3.2. Strains and plasmids

Escherichia coli DH5α strain was utilized for cloning and amplification of the constructed plasmids. As expressing strain *Pichia pastoris* X-33 (Invitrogen, Carlsbad, Ca, USA) was recruited. pGAPZαA (Invitrogen) and pPICZαA::hGH (Çalık et al., 2008) plasmids were used as parent plasmids for development of the desired plasmid constructs which can be observed in Figure 3.1 and Figure 3.2, respectively.

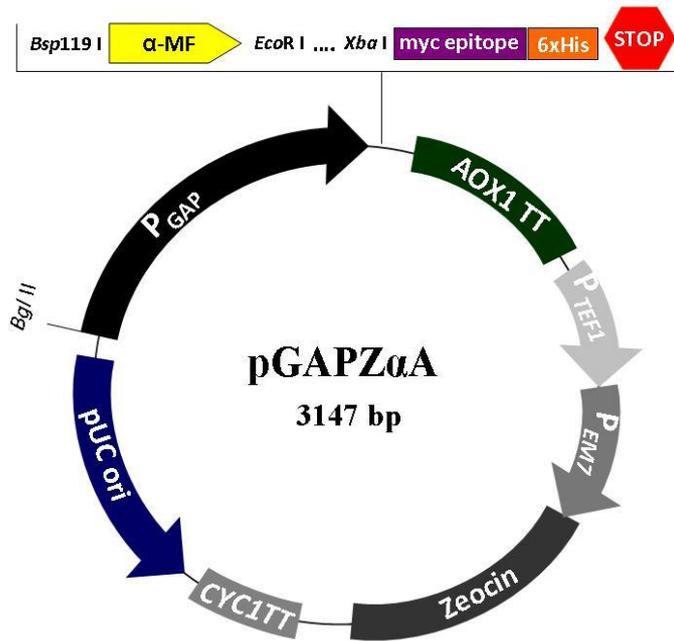


Figure 3.1. Schematic representation of pGAPZ α A.

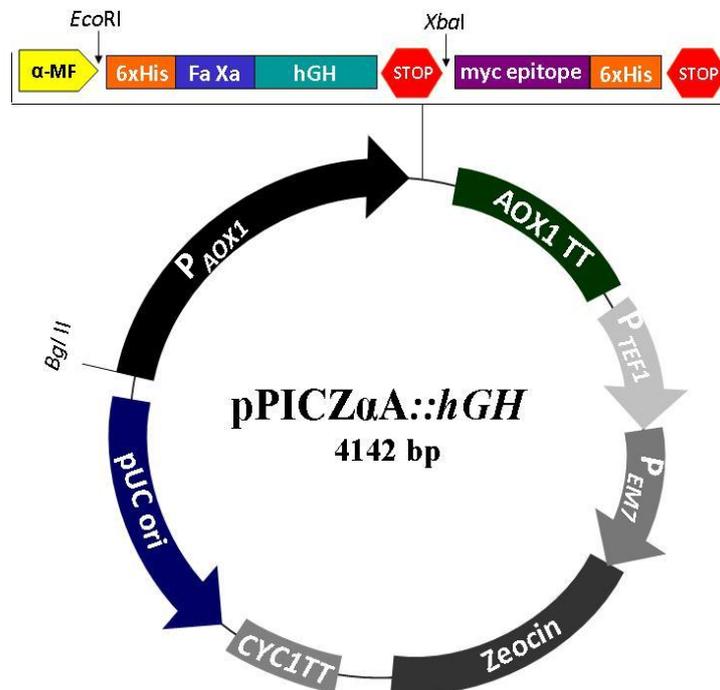


Figure 3.2. Schematic representation of pPICZ α A::hGH.

All the microbial strains were stored as long-lasting stocks in -80°C with Microbank™ which includes treated beads and a specific cryo-preserved solution (Web 14); however, corresponding temporary cultures were prepared directly from stocks in agar plates and were stored at +4°C for 15-30 days depending on the presence or absence of antibiotic in agar plate. The successive sub-culturing from plate to plate increases the risk of contamination and loss of rDNA. All the plasmids were stored at -20°C.

3.3. Media

The utilized media can be divided into two categories: media used for growth of the cells and media used in shake flask and laboratory-scale bioreactor experiments for rhGH production.

3.3.1. Growth media

E. coli strains were grown in low salt Luria broth (LSLB) whether in liquid or solid form (Table 3.1); Zeocin™ (25 µg/mL) and ampicillin (100 µg/mL) were added to this medium in order to grow true transformants whenever needed. *P. pastoris* strains growth was conducted in yeast extract peptone dextrose (YPD) medium (Table 3.2) which can be in either liquid or solid form; in order to make a selective medium, Zeocin™ was used in concentration of 100 (µg/mL).

3.3.2. rhGH production media

rhGH production experiments were conducted in both shake flask air-filtered bioreactor and laboratory-scale bioreactor. Each of them requires specific medium for production. However, they both share a pre-production step with so-called pre-cultivation medium which mainly increases the cell amount.

Table 3.1. LSLB agar medium composition for *E. coli* growth.

Compound	Concentration (g/L)
Tryptone	10
Yeast extract	5
NaCl	5
Agar	15

Table 3.2. YPD agar medium composition for *P. pastoris* growth.

Component	Concentration (g/L)
Yeast extract	10
Peptone	20
D-glucose (dextrose)	20
Agar	20

3.3.2.1. Pre-cultivation medium

In both shake flask and laboratory-scale bioreactor experiments the pre-cultivation medium for recombinant *P. pastoris* (r-*P. pastoris*) strains was buffered minimal glycerol-complex medium (BMGY) which was utilized prior to production (Table 3.3).

3.3.2.2. Production medium

After primary cell generation, the mediums in shake flask and laboratory-scale bioreactors were different.

Table 3.3. BMGY medium for pre-cultivation in shake flask and laboratory-scale bioreactor experiments.

Component	Amount (per 1 L)
Yeast extract	10 g
Peptone	20 g
Glycerol	10 mL
YNB (w/o amino acids)	3.4
Ammonium sulfate	10g
Biotin	0.0004 g
1M Potassium phosphate buffer, pH=6.0	100 mL

3.3.2.2.1. Shake flask air-filtered bioreactor

Defined medium for production in shake flask bioreactor experiments was taken from (Jungo et al., 2006) just with replacing $(\text{NH}_4)_2\text{SO}_4$ for NH_4Cl . Ammonium sulfate amount was calculated according to the glucose amount and regarding $C/N= 4.57$ (Jungo et al., 2006). The composition of the production medium for shake flask bioreactor experiments can be found in Table 3.4.

However, in production phase, addition of trace salts to the medium in order to fulfil the trivial needs of the microorganism is inevitable and, therefore, *Pichia* trace minerals (PTM1) solution (Table 3.5) was added to the production medium.

Table 3.4. Production medium for shake flask bioreactor experiments; the amount of ammonium sulfate was calculated based on pre-defined (C/N=4.57) ratio.

Component	Amount (per 1 L)
Glucose, C ₆ H ₁₂ O ₆	20 g
Calcium sulfate dihydrate	1.17 g
Magnesium sulfate heptahydrate	14.9 g
Ammonium sulfate, (NH ₄) ₂ SO ₄	9.63 g
1M Potassium phosphate buffer, pH=6.0	100 mL
PTM1	4.35 mL
Chloramphenicol (from 34 mg/mL stock)	1 mL

Table 3.5. PTM1 solution including trace salts for *P.pastoris* cultivation. (*Pichia* fermentation process guidelines, Invitrogen).

Component	Representation	Amount (per 1 L)
Cupric sulfate pentahydrate	CuSO ₄ .5H ₂ O	6 g
Sodium iodide	NaI	0.08 g
Manganese sulfate monohydrate	MnSO ₄ .H ₂ O	3 g
Sodium molybdate dihydrate	Na ₂ MoO ₄ .2H ₂ O	0.2 g
Boric acid	H ₃ BO ₃	0.02 g
Cobalt chloride	CoCl ₂	0.5 g
Zinc chloride	ZnCl ₂	20 g
Ferrous sulfate heptahydrate	FeSO ₄ .7H ₂ O	65 g
Biotin	-	0.2 g
Sulfuric acid	H ₂ SO ₄	5 mL

3.3.2.2.2. Laboratory-scale bioreactor

In laboratory-scale bioreactor experiments, after pre-cultivation using BMGY in shake flasks, for further increase of the cell amount, the batch phase of the process in bioreactor was accomplished by growth of the cells in basal salt medium (BSM) which again includes glycerol as carbon source (Table 3.6). However, the concentration of the glycerol in two sets of bioreactor runs was different; it was 40 g/L and 25 g/L for bioreactor experiments of SPs' strains (1st set) and promoters' strains (2nd set), respectively. After glycerol batch phase, glucose fed-batch phase of the production was started by exponential feeding of the glucose solution (Table 3.7).

Table 3.6. BSM for laboratory-scale bioreactor experiments; glycerol concentration for bioreactors related with SPs and promoters was decided to be 40 g/L and 25 g/L, respectively.

Component	Amount (per 1L)
Phosphoric acid 85%	26.7 mL
Calcium sulfate dehydrate	1.17 g
Potassium sulfate	18.2 gr
Magnesium sulfate heptahydrate	14.9 gr
Potassium hydroxide	4.13 gr
Glycerol	Variable
PTM1	4.35 mL
Antifoam 10%	1 mL
Chloramphenicol (from 34 mg/mL stock)	1 mL

Table 3.7. Ingredients of the glucose solution in fed-batch phase of the laboratory-scale bioreactor experiments.

Component	Amount (per 1 L)
Glucose	500g
PTM1	12 mL
Chloramphenicol	1 mL

3.4. Finding the nucleotide sequences of the desired biological elements

Major focus of the current research was on the SPs and promoters as two fundamental biological elements of the r-protein expression and, therefore, obtaining their corresponding nucleotide sequences was imperative in order to include them in structure of the new plasmid constructs.

3.4.1. Secretion signal peptides

The amino acid sequences of the endogenous proteins found in extracellular medium of the *P. pastoris* (Huang et al., 2011) along with their corresponding nucleotide sequences were down-loaded from NCBI (<http://www.ncbi.nlm.nih.gov>) and, then, were analyzed for determination of their secretion SPs. In addition to these proteins, the amino acid sequences of the exogenous proteins that their SPs were used for r-protein expression in *P. pastoris* and their related nucleotide sequences were obtained from UNIPROT (<http://www.uniprot.org>), were also analyzed. The summary of the procedure can be found in Figure 3.3.

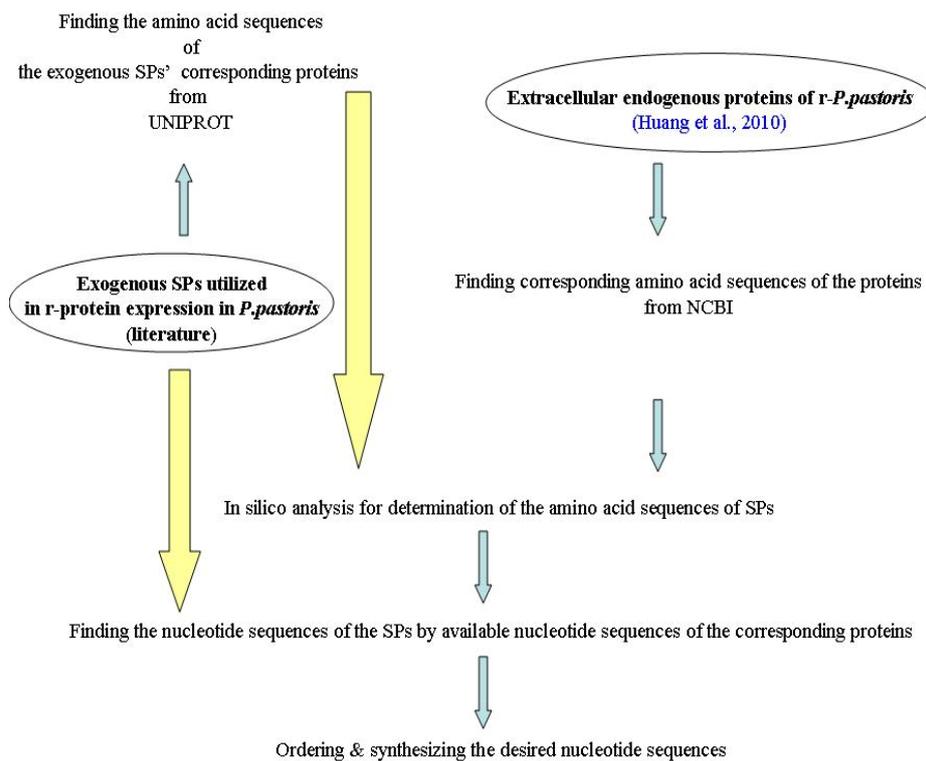


Figure 3.3. The sequence of events which were followed in order to obtain the nucleotide sequence of the endogenous and exogenous secretion SPs.

3.4.2. Promoters

According to the available data related with the effect of the different levels of oxygen provision on gene expression in *r-P. pastoris* (Baumann et al., 2010), the genes that had been upregulated more (in comparison to *GAP* gene) in lower oxygen availability were identified. The accession numbers of the selected genes and their corresponding proteins (enzymes) were identified. After finding the corresponding genes in *P. pastoris* GS115 (virtually X-33) by NCBI BLAST tool, the related promoters were determined and their nucleotide sequences were extracted. General procedure for identifying promoter regions has been summarized in Figure 3.4:

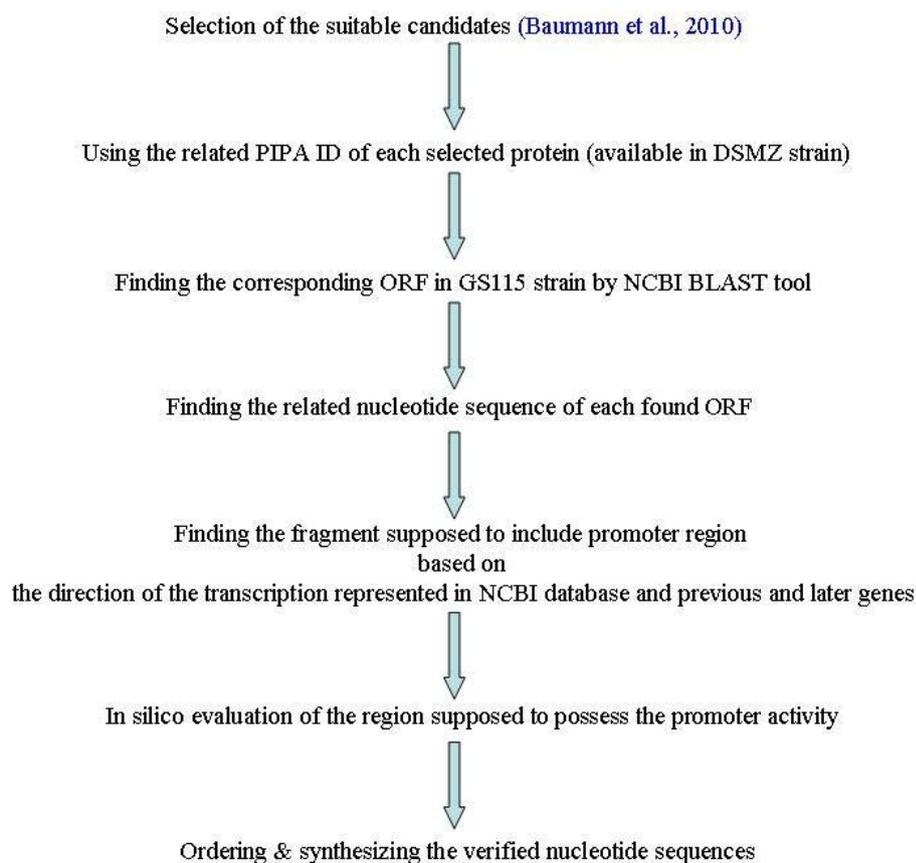


Figure 3.4. The sequence of events which were followed in order to obtain the nucleotide sequence of the desired promoters.

As a starting point, it was supposed that the sequence between the selected gene and its preceding gene is a rough approximation of the promoter region. However, it is obvious that this roughly found regions may contain nucleotides that codes for 5' and 3' un-translated regions (UTR) in mRNA. Furthermore, transcription activation factors, enhancers and inhibitors may reside in upstream or downstream of the gene which enters the other gene. The selected fragment between the genes is roughly referred to as promoter region and the promoter activity is attributed to this '*putative promoter*' region.

3.5. In-silico analyses

The SPs were determined by SignalP software v.4.1 available at (<http://www.cbs.dtu.dk>), WolfPsort (<http://wolfsort.org/>), and Phobius (<http://phobius.binf.ku.dk/>) programs. Furthermore, the D-scores of the SPs were determined by SignalP and the H-regions of the SPs were determined by Phobius program. ProP1.0 (<http://www.cbs.dtu.dk/services/ProP/>) was used in order to determine the probable “pro” sequence of the SPs.

In analyzing the results of the SPs experiments, GRand Average of hydropathY (GRAVY) and aliphatic index were computed using ExPASy ProtParam (<http://web.expasy.org/protparam/>). Isoelectric point (pI) and net charge were calculated by iep program (<http://www.bioinformatics.nl/cgi-bin/emboss/iep>) as a part of the EMBOSS bioinformatics package (Choo and Ranganathan, 2008). Mean charge was obtained by dividing the net charge by the polypeptide length.

Evaluation of the nucleotide sequences as ‘*putative promoter*’ regions were conducted by the software available at (www.fruitfly.org) as a part of Berkeley Drosophila genome project.

3.6. Genetic engineering techniques

The final product of these steps was transfected *P. pastoris* harboring newly-developed plasmid constructs.

3.6.1. Synthesizing of the determined nucleotide sequences

After provision of the required restriction sites in the nucleotide sequences of the selected endogenous secretion SPs and promoters of *P. pastoris*, they were ordered. The complete sequence was synthesized by GenScript inside unique plasmid, pUC57 (Appendix C). The synthesized plasmid was received in lyophilized form. In order to clone the synthesized plasmid, it was transformed into the *E. coli* DH5 α cell.

3.6.2. *E. coli* transformation

Newly developed plasmids were used in order to transform *E. coli* DH5 α cells and, thus, they could be amplified. The transformation procedure based on CaCl₂ method (Sambrook and Russell, 2001) with minor modification in recovery step is as follows:

1. Inoculation of the *E. coli* DH5 α strain from available glycerol stock at -80°C into LSLB agar medium
2. Overnight (16-20 hours) incubation at 37°C
3. Picking a single colony (2-3 mm in diameter) from the plate
4. Transferring the picked colony to 50 mL LB broth in 250 mL Erlenmeyer flasks
5. Incubation at 37°C with vigorous agitation at 175-200 rpm and monitoring the growth of the culture (measuring the OD₆₀₀ of the culture every 15-20 minutes)
6. Harvesting the culture when the OD₆₀₀ reaches 0.35-0.40
7. Transferring the cells to sterile, ice-cold 50-mL polypropylene tube
8. Cooling the culture to 0°C by keeping the tube on ice for 10 minutes
9. Recovering the cells by centrifugation at 3000g for 10 minutes at 4°C
10. Discarding the supernatant putting the tube in an inverted position on a paper towel for 1 minute to let the remained medium to be drained away
11. Re-suspending the cell pellets in 30 mL of ice-cold filter-sterilized (0.2 μ m) MgCl₂-CaCl₂ solution (80mM MgCl₂, 20mM CaCl₂) by gentle vortexing or swirling
12. Recovering the cells by centrifugation at 3000g for 10 minutes at +4°C.
13. Discarding the supernatant and putting the tube in an inverted position on a paper towel for 1 minutes to let the last traces of the media to be removed

14. Re-suspending the cell pellets in 2 mL of ice-cold filter sterilized (0.2 μm) 0.1 M CaCl_2 for each 50 mL of original culture (with gentle mixing or swirling)

At this point, the cells are chemically-competent and can be used directly or stored in aliquots with 12% glycerol at -80°C when liquid nitrogen is available.

15. Transferring 200 μL of the competent cells into a sterile, ice-chilled polypropylene tube by the aid of ice-chilled micropipette tips
16. Adding desired plasmid / ligation mixture (the DNA) by taking into consideration that the volume of the ligation reaction mixture should not exceed 10% of the competent cells volume
17. Mixing gently and storing on ice for 30 minutes
18. Transferring the tube to a rack placed in a 42°C pre-heated water bath and keeping the tube in the water bath for exactly 90 seconds without shaking. This heat shock step is very crucial; it is very important to raise the temperature to the exactly right point and at the correct rate
19. Rapid transferring of the tube to an ice bath and allowing the cells to be chilled for 1-2 minutes
20. Adding 800 μL LB medium to the tube and incubation for 60-80 minutes at 37°C in shaker incubator with agitation of 150 rpm (with loosely-capped tubes) to allow recovery of the cells and expressing the antibiotic resistance marker encoded by the foreign plasmid
21. Transferring appropriate volume, 100, 150 and 200 μL (up to 200 μL per plate) of the treated cells onto LSLB agar medium containing desired antibiotic at desired concentration
22. Gentle spreading the cells over the surface of the agar plate by a sterilized loop or bent glass rod and waiting until absorption of all the liquid
23. Incubation of the plates at 37°C ; colonies will appear in 16-24 hours

After appearing the colonies, single-colony cultures of the selected colonies were prepared in LSLB agar plates under selective pressure of the suitable

antibiotic; after overnight incubation at 37°C, false positive colonies were removed by failing to grow and, thus, true transformants were discriminated.

3.6.3. Plasmid DNA isolation from transformed *E. coli* cells

After transformation, a few (true) transformed colonies should be selected and the plasmid isolation should be performed using each of them to verify their precise transformation. In order to obtain high amounts of the desired plasmids, the transformed *E. coli* cells should be cultivated, harvested, and then disrupted. Plasmid isolation was conducted by two methods in current research; alkaline lysis and using plasmid isolation kits. Plasmid isolation by kit was conducted based on manufacturer's instruction. Alkaline lysis method was accomplished as follows:

Preparation of the cells:

1. Inoculation of 10 mL of the LB broth medium (in 50-mL sterile falcon) containing appropriate antibiotic (here, ampicillin or Zeocin™) with a single colony of the transformed *E. coli* cells
2. Overnight (~16 hours) incubation at 37°C with vigorous shaking at 200-225 rpm with loosely-capped tube(s) to ensure desirable aeration.
3. Centrifugation at 3000g for 10 minutes at +4°C.
4. Removing the supernatant by gentle aspiration and positioning the tube(s) in a manner that the remaining medium drains on a clean towel paper in order to leave the cells as dry as possible (cell wall components in the medium prevent the action of many of restriction endonuclease enzymes and, thus, the plasmid DNA will be resistant to cleavage).

Cell lysis:

5. Re-suspension of the pellet in 200 µL ice-cold alkaline lysis I solution (Appendix B) by vigorous mixing until making sure that the cells would be completely dispersed

6. Addition of 400 μL alkaline lysis II solution (Appendix B) which has been freshly- prepared to the bacterial solution and gentle mixing with rapid inversion for five times and storing the tube(s) on ice. Do not vortex in this step
7. Addition of 300 μL ice-cold alkaline lysis III solution (Appendix B) and gentle mixing by inversion of the tube(s) for several times and storing the tube(s) on ice for 3-5 minutes
8. Centrifugation of tube(s) at 14000g for 5 minutes at $+4^{\circ}\text{C}$ and careful transferring of 600 μL of the supernatant to a fresh (microfuge) tube in order not to disturb the cell debris during transfer
9. Adding equal volume of phenol/chloroform (at $+4^{\circ}\text{C}$) and mixing two phases by vortexing and, then, centrifugation of the emulsion at 14000g for a 2-minute period at $+4^{\circ}\text{C}$ and transferring the supernatant to a fresh tube

Recovery of the plasmid DNA:

10. Addition of 600 μL iso-propanol at room temperature (RT) and mixing the solution by vortexing and keeping the tube standed for 2 minutes at room temperature in order to precipitate nucleic acids from the supernatant
11. Collecting the precipitated nucleic acids by centrifugation at 14000g for 5 minutes at room temperature
12. Removal of the supernatant by gentle aspiration and positioning the tube in a manner that the remaining medium drains on a clean towel paper and, furthermore, removing of any fluid droplet present in the wall of the tube
13. Addition of 1 mL ethanol 70% (to wash the pellet) and recovering the DNA by centrifugation for 2 minutes at 14000g and at RT
14. Pouring off the supernatant by gentle aspiration (take care in this part, since the precipitate sometimes does not adhere to the tube wall tightly)

15. Removing remaining ethanol droplets from the sides of the tube and keeping the tube at room temperature for 10-15 minutes to let the ethanol be evaporated without DNA becoming dehydrated
16. Dissolving the final pellet (nucleic acids) in 30-100 μL TE buffer (pH=8.0) (Appendix B) containing (DNase-free) RNase at the final concentration of 20 $\mu\text{g}/\text{mL}$ and mixing by pipetting of the solution for a few seconds and storing the isolated plasmids at -20°C .
17. Determining the concentration of the isolated plasmid; it should be emphasized that $\text{OD}_{260}=1$ using a cuvette with light path of 10 mm refers to a DNA concentration of approximately 50 $\text{ng}/\mu\text{L}$.

3.6.4. Confirmation of the DNA fragments by agarose gel electrophoresis

The control of the isolated plasmids, isolated genomic DNA, and digested DNA fragments, regarding their length, was accomplished with agarose gel electrophoresis (AGE). AGE was conducted with Mini-sub[®] Cell GT Cell system (Bio-Rad, CA, USA). The general procedure of the AGE can be described as follows:

1. Cleaning the required apparatus and equipments and rinsing with reverse osmosis (RO) or ultrapure (UP) water
2. Assembling the gel casting tray and checking its horizontality
3. Preparation of 5X TBE (or TAE) (Appendix B) buffer and diluting to 1X buffer just before usage
4. Dissolving required amount (appropriate for separation of desired fragments) of agarose in 1X buffer to reach the desired resolution
5. Heating the mixture and boiling for a short while until being clear. Check that the volume of the solution has not been decreased by evaporation during boiling; replenish with dH_2O if necessary.
6. Addition of EtBr after cooling of the gel in order to have a final concentration of ca.0.4 $\mu\text{g}/\text{mL}$
7. Casting the gel in the plastic trays after reaching the molten gel $< 60^{\circ}\text{C}$
8. Putting the combs into the molten gel in order to form the wells

9. Allowing the gel to set completely for 30-45 minutes and then removing the comb carefully by pouring small amount of buffer on the gel; gelling temperature is 35-42°C
10. Mixing the DNA samples (5-10 μL) with 0.2 volume of the DNA gel loading dye 6X (Thermoscientific). DNA samples can be diluted if necessary. A typical sample well can hold about 40 μL sample
11. Loading the mixtures to the submerged gel in buffer
12. Loading DNA ladder (Thermoscientific) as a length standard (3-5 μL); using two different DNA markers will be beneficial
13. Closing the lid of the gel tank and attaching the electrical leads.
14. Performing electrophoresis at 90V for 40-100 minutes depending on the gel concentration and desired DNA length
15. Running the gel until the band of interest has migrated 40% - 60% down the length of the gel. Band broadening resulted from dispersion and diffusion results in a decrease in resolution in the lower third of the gel.
16. Visualizing the gel by UV light

3.6.5. Amplification of the target DNA

The amplification of the desired DNA segments was conducted with polymerase chain reaction (PCR). In order to amplify desired DNA fragments specific primers were designed. For this purpose, OligoAnalyzer3.1 (<http://eu.idtdna.com/calc/analyzer>) and NCBI primer design tool (www.ncbi.nlm.nih.gov) were utilized. In primer design, T_m values, ΔG values, GC contents, and possibility of self dimer and hetero dimer were also considered. The primers were synthesized by Sentegen (Ankara, Turkey) and Sentromer (Istanbul, Turkey) and were received in lyophilized form. They were prepared as 100 μM stocks with autoclaved filter-sterilized dH_2O according to the manufacturer's instruction; they were diluted to desired concentrations whenever needed. The complete list of the primers is available in Appendix D.

PCR buffers used in this research were without Mg^{2+} , and $MgSO_4$ was added separately in order to optimize Mg^{2+} ion concentration for improving efficiency. Utilized primers in PCRs of current study were 5 μ M concentration else mentioned. dNTPs were with concentration of 2 mM (Table 3.8). dH_2O utilized throughout the genetic analyses was autoclaved and filter-sterilized (0.2 μ m).

Table 3.8. Typical amounts of PCR constituents.

Component	Amount
10X amplification buffer with KCl	5 μ L
2 mM solution of dNTPs	5 μ L
5 μ M forward primer	Typically 2 μ L (can change)
5 μ M reverse primer	Typically 2 μ L (can change)
25 mM $MgCl_2$ / $MgSO_4$	2-4 μ L (can be optimized)
<i>Taq</i> / <i>pfu</i> DNA polymerase	1-2 U
dH_2O	Variable
Template DNA (Genomic DNA or plasmid DNA)	Variable
Total volume	50 μL

The PCR experiments were conducted in thermal block cycler (Techne[®], Flexigene and TC-3000X). The programs were changed according to the specific requirements of the reactions and the annealing temperatures and elongation times were situation-dependent. However, the general program can be seen in Table 3.9.

Table 3.9. Typical thermal profile for PCR. Variable amounts are decided based on the melting temperature of the designed primers, length of the fragment to be amplified, and type of the utilized DNA polymerase.

Number of Cycles	Temperature	Time
1 Cycle, Primary denaturation	94°C	3 min.
1 cycle	94°C	1 min.
	variable	1 min.
	72°C	variable
29 cycles	94°C	1 min.
	variable	1 min.
	72°C	variable
1 Cycle, Final extension	72°C	10 min.

3.6.6. Purification of the PCR products

The DNA fragments amplified via PCR were purified using PCR purification kit (Thermoscientific) in order to get rid of interfering agents which can halt subsequent experiments. Briefly:

1. Addition of 1:1 volume of the binding buffer to the final PCR product in PCR tube or a clean eppendorf tube and mixing the content thoroughly
2. Transferring (up to 800 μ L) the yellowish solution to the available column supplied in kit.
3. Centrifugation for 1 minute at 14000g at RT and discarding the flow-through
4. Addition of 700 μ L wash buffer to the column
5. Centrifugation for 1 minute at 14000g at RT and removing the flow-through
6. Additional centrifugation for 3 minutes to remove the residual ethanol

7. Addition of 20-50 μL elution buffer to the column
8. Centrifuged for 1 minute at 14000g at RT to elute the amplified DNA fragment. The eluted DNA was stored at -20°C .

3.6.7. Digestion of the plasmid DNA

In order to construct desired plasmids, first it is needed that some available plasmids be digested by restriction endonuclease enzymes. In addition to single-digestion performed for control purposes, cutting out a DNA fragment from an available plasmid was conducted by double-digestion with two suitable restriction enzymes. Each restriction digestion requires restriction enzyme, suitable buffer for optimum enzyme activity, sample DNA, and dH_2O to complete the reaction volume (Table 3.10). In double-digestion, the buffer should be suitable for both enzymes activity which can be accessed from the ThermoScientific website. Each restriction enzyme identifies a specific nucleotide sequence and cleaves the phosphodiester bond in that location in phosphate backbone. The free ends of the cleaved DNA strands may either be sticky or blunt where the sticky ends are preferred. Afterwards, these free (sticky) ends are used to insert the cut fragment within a new DNA backbone. The conditions for single- and double-digestions have been represented in Appendix E.

Table 3.10. Typical reaction mixture for single-digestion with *EcoRI* restriction enzyme.

Component	Amount
Nuclease-free water	Variable
<i>EcoRI</i> unique Buffer 10X	2 μL
DNA	Variable
<i>EcoRI</i>	0.5 - 2 μL
Total volume	20μL

EcoRI and *XbaI* restriction enzymes pair, was used for double-digestion of the parent plasmids at first step. After complete mixing of the restriction digestion reaction components in an eppendorf tube (Table 3.11), a brief span was conducted to collect the tube content in the bottom of the tube and, then, the mixture was incubated at 37°C in water bath for a suitable time period according to the amount of the plasmid DNA and restriction enzyme. The time should be enough to digest the whole DNA sample without any overdigestion. Inactivation of the enzymes was conducted by heating at 65°C for 20 minutes. .

Table 3.11. Typical reaction mixture for double-digestion with *EcoRI* and *XbaI* restriction enzymes.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable
10X Tango buffer	4 µL
DNA (parent plasmids)	Variable (up to 1 µg)
<i>EcoRI</i>	1 µL
<i>XbaI</i>	2 µL
Total volume	20µL

Bsp119I and *EcoRI* restriction enzymes pair was utilized to facilitate insertion of the selected endogenous signal peptides in the place of α -MF in order to develop a new set of plasmid constructs (Table 3.12). By ending the incubation time, inactivation of the enzymes in this case was conducted by heating at 80°C for 20 minutes.

Table 3.12. Typical reaction mixture for double-digestion with *EcoRI* and *Bsp119I* restriction enzymes.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable
10X Tango buffer	4 μL
DNA	Variable (up to 1 μg)
<i>EcoRI</i>	1 μL
<i>Bsp119I</i>	1 μL
Total volume	20μL

EcoRI and *NsiI* restriction enzymes pair was recruited to replace P_{GAP} by endogenous promoters to establish a distinctive set of novel plasmid constructs (Table 3.13). Upon completion of the incubation, inactivation of the enzymes was conducted by heating at 65°C for 20 minutes.

Table 3.13. Typical reaction mixture for double-digestion with *EcoRI* and *NsiI* restriction enzymes.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable
10X Tango buffer	4 μL
DNA	Variable (up to 1 μg)
<i>EcoRI</i>	1 μL
<i>NsiI</i>	2 μL
Total volume	20μL

The restriction digestion results, in all cases, were subjected to AGE in order to be verified. After verification, if the separated fragments were intended to be used in the next round of experiments, i.e., ligation, they were

cut from the gel and were subjected to the purification by gel extraction kit (Thermoscientific). Then, the double-digestion results were ready for ligation reaction.

3.6.8. Extraction of the DNA fragments from the agarose gel

After running the DNA samples on agarose gel and separation of the desired DNA fragments, the intended segments could be extracted from the gel, whenever needed, by recruiting gel extraction kit (Thermoscientific) according to the supplier instructions; in brief, the procedure can be summarized as following steps:

1. Cutting out DNA fragment from the gel in slice less than 400mg
2. Putting the slice in a clean pre-weighed 2-mL eppendorf tube
3. Addition of same volume of the binding buffer to the tube
4. Incubating the tube at 60°C for a 10-minute period by mixing with inversion a few times until complete dissolution of the gel
5. Transferring the clear yellowish solution into the supplied column
6. Centrifugation for 1 minute at 14000g at RT
7. Discarding the filtrate
8. Addition of 700 µL wash buffer to the column
9. Centrifugation for 1 minute at 14000g at RT
10. Centrifugation of the empty column for 1-3 minute(s) at 14000g at RT in order to remove the remained wash buffer
11. Transferring the column into a new clean 1.5-mL eppendorf tube
12. Addition of 20-50 µL elution buffer to the column
13. Centrifugation for 1 minute at 14000g at RT
14. Storage of the DNA fragment in the eluted liquid at -20°C in order to be used for subsequent ligation reactions.

3.6.9. Ligation of the DNA fragments

DNA fragments with sticky ends developed by single- or double-digestion can be joined together by the action of the T4 DNA ligase enzyme and, thus,

new DNA molecules are formed. Ligation of the desired DNA fragments in total volume of 20 μL was achieved by overnight incubation (~ 14 hours) of the mixture at 16°C and, then, inactivation by heating at 70°C for 10 minutes. The ligation results were directly used to transform *E. coli* cells. The constituents of a typical ligation reaction can be observed in Table 3.14.

Table 3.14. Typical ligation reaction mixture; both insert and backbone should be double-digested with similar restriction enzymes.

Component	Amount
10X ligation buffer	2 μL
Gene of interest (Insert)	Variable (regarding molar ratio)
Plasmid DNA (Backbone)	Up to 100 ng
T4 DNA ligase	1 μL
Autoclaved and filter-sterilized dH ₂ O	Variable
Total volume	20 μL

The molar ratio between the “*insert*” and “*backbone*”, represented by equation (3.1), is the most critical parameter in the ligation reaction and can be changed in order to reach successful ligation results. It is usually decided to be between 1 and 6. The amount of the gene of interest in Table 3.14 can vary to give different molar ratios:

$$\text{Molar ratio} = \left(\frac{\text{Amount (ng)}}{\text{Size (bp)}} \right)_{\text{Insert}} \times \left(\frac{\text{Size (bp)}}{\text{Amount (ng)}} \right)_{\text{Backbone}} \quad (3.1)$$

3.6.10. Sequencing of the isolated plasmids from *E. coli* cells

After transformation of the *E. coli* cells with developed plasmid constructs and subsequent plasmid isolation, verification of the identity of the plasmids

was achieved via AGE, restriction digestion, and PCR. Afterwards, the isolated plasmids were subjected to DNA sequence analysis in automatic DNA sequencer ABI Prism[®] 310 Genetic Analyzer (Applied Biosystems, California, USA) in METU central laboratory using the primers can be found in Table 3.15. The sequence authenticity was checked thoroughly with NCBI nucleotide BLAST tool (www.ncbi.nlm.nih.gov). Authentic plasmids were recruited in transfection of *Pichia pastoris* cells.

Table 3.15. The recruited primers for sequencing of the isolated plasmids from *E. coli* cells for their verification prior to *Pichia* transfection.

Primer Name	Primer Nucleotide Sequence
<i>GAP</i> forward	5'- GTCCCTATTTCAATCAATTGAA-3'
<i>AOX</i> reverse	5'- GCAAATGGCATTCTGACATCC-3'
<i>PDC-F</i>	5'...TCTATGCATGAGATCGGGACAAGCA...3'
<i>PDC-R</i>	5'...TGCGAATTCAGCTTCAGCCTCTCTTT...3'
<i>PYRK-F</i>	5'...TAAATGCATGAGATCTTCAGTGTGCGG...3'
<i>PYRK-R</i>	5'...TGAGAATTCAGCTTCAGCCTCTCTTTTCT...3'
<i>THI3-F</i>	5'...GAGATGCATGAGATCGTCTTTGTAAATAGT...3'
<i>THI3-R</i>	5'... TGAGAATTCAGCTTCAGCCTCTCTTTTC...3'
New <i>GAP</i> fwd	5'... GACGCATGTCATGAGATTATTGG ...3'
<i>PYRK</i> End-fwd	5'...GACCGTTCATGTACAGTAAATTG...3'
<i>PDC</i> End-fwd	5'...ACCTCAAATATCTAGCAACATCTT...3'
<i>THI3</i> End-fwd	5'...GGAATGTTGGTAGCAATTGGTAT...3'

3.6.11. Transfection of the *Pichia pastoris* X-33 cells

After developing new plasmids by genetic engineering techniques, subsequent cloning in *E. coli* cells, plasmid isolation, and sequencing, the plasmids were used to transfect *P. pastoris* X-33 cells. Enough plasmid DNA should be generated before transfection; 5-10 µg of each plasmid per each transfection is required. Furthermore, transfection with lithium acetate does not work with *P. pastoris* and the lithium chloride (LiCl) method is used (EasySelect™ *Pichia* Expression Kit). The general procedure of transfection includes plasmid preparation, cell preparation, and directing the plasmid into the cell. The procedure can be described as follows:

1. Restriction digestion of the desired plasmid with *NsiI* restriction endonuclease enzyme by incubation at 37°C for 5 hours and, then, inactivation at 65°C for 20 minutes in order to linearize the plasmid (Table 3.16).

It has been mentioned that homologous recombination can also occur with non-linearized plasmid, although, at a lower frequency. The linearized plasmid was purified with PCR purification kit (Thermoscientific) where at the elution step autoclaved and filter-sterilized dH₂O was utilized to elute the plasmid DNA. The linearization and purification was performed just one day before transfection.

Table 3.16. Typical reaction mixture for single-digestion with *NsiI* restriction enzyme.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable (complete to 20 µL)
R buffer 10X	2 µL
Circular plasmid DNA	Variable (up to 1 µg)
<i>NsiI</i>	0.5 - 2 µL
Total volume	20 µL

2. Inoculation of *P.pastoris* X-33 strain from -80°C glycerol stock into YPD agar plate and incubation for approximately 60 hours at 30°C
3. Picking up a single colony from the plate and growing the culture in a 250-mL baffled Erlenmeyer flask containing 50 mL YPD medium at 30°C and 200-225 rpm until reaching OD₆₀₀ of 0.8 -1.0 (~ 10⁸ cells / mL).
4. Harvesting the cells by centrifugation at 4000g for 5 minutes at +4°C
5. Washing the pellet with 25 mL sterile dH₂O and, then, centrifugation at 1500g for 10 minutes at RT
6. Pouring off the water (supernatant) and re-suspending the cells in 1 mL of 100 mM LiCl solution
7. Transferring the cell suspension to a 1.5-mL sterile eppendorf tube
8. Pelleting the suspended cells by centrifugation at 16000g for a 15-second period and removing LiCl with a pipette
9. Re-suspending the cells in 400 µL of 100 mM LiCl
10. Transferring 50 µL of the cell suspension into a sterile 1.5-mL eppendorf tube (for each transformation) and immediate utilization without storing on ice or freezing at -20°C
11. Boiling 25 µL sample of single-stranded DNA (salmon sperm DNA) for 5 minutes, quick chilling on ice and keeping on the ice until use
12. Centrifugation of the cell-LiCl suspension (from part 10) at 16000g speed for 15 seconds and removing the LiCl with a pipette
13. Adding the following reagents in exactly represented order to the pelleted cells in eppendorf tube:
 - 240 µL 50% (w/v) PEG
 - 36 µL 1M LiCl
 - 25 µL Salmon sperm 2 mg/mL
 - Linearized plasmid DNA (5-10 µg) in 50 µL sterile water
14. Vortexing tube(s) until thorough mixing of the cell pellets (~ 1 minute)
15. Incubation of the tube(s) at 30°C for 30 minutes without shaking

16. Exerting heat shock on the cells by incubation at 42°C in water bath for a period of 30-40 minutes
17. Centrifugation of the tube(s) at 4000g for 15 seconds and removal of the transfection solution with a pipette
18. Re-suspending the collected cells in 1 mL of the YPD medium (free of antibiotic) and incubation at 30°C and 150 rpm for 4 hours
19. Transferring 50-200 µL of the cell suspension onto the YPD agar plate(s) with Zeocin™ concentration of 100 µg/mL
20. Incubation of the plates for 2-3 days at 30°C

The putative transfectants were chosen and analyzed further in order to identify the real transfectants for subsequent shake flask and bioreactor experiments.

3.6.12. Genomic DNA isolation from *Pichia pastoris* cells

After transfection and selection of the putative transfectants, in order to verify the genomic integration of the desired plasmid constructs, the genomic DNA of the putative transfectants should be isolated. For this purpose, single-colony cultures of the selected putative transfectants were prepared by inoculation of the appeared colonies into the new YPD agar plates under Zeocin™ pressure (100 µg/mL). The plates were incubated for 20-24 hours at 30°C. Then:

1. Inoculation from single-colony cultures into 10 mL YPD medium in a 50-mL sterile falcon and cultivation at 30°C and 200 rpm until reaching $6 < OD_{600} < 10$
2. Harvesting the cells by centrifugation at 4000g for 10 minutes at RT
3. Removing the supernatant and re-suspending the cells in 0.5 mL distilled water and transferring to a 1.5-mL sterile eppendorf tube.
4. Centrifugation at 16000g for 2 minutes in order to collect the cells
5. Removing the supernatant and vortexing the cells in remaining supernatant

6. Addition of 200 μ L yeast lysis solution (Appendix B) to the tube(s) and mixing the tube(s) by inversion
7. Addition of 200 μ L phenol/chloroform/isoamyl alcohol (25:24:1) and 0.3 g acid-washed glass beads
8. Wrapping the tube(s) with parafilm and vortexing vigorously for about 4 minutes
9. Addition of 200 μ L TE buffer (pH=8.0) and centrifugation of tube(s) at 4000g for 2 minutes at RT
10. Transferring the upper aqueous layer to a new 2-mL eppendorf tube
11. Adding 1mL of 100% Ethanol (EtOH) and mixing by inversion
12. Centrifugation at 16000g for 2 minutes and pouring away the supernatant
13. Re-suspending the pellets in 400 μ L TE buffer with 5 μ L of 10 mg/mL solution of RNase A (a pancreatic ribonuclease cleaves single-stranded RNA)
14. Incubation of the solution at 37°C for 10 minutes
15. Addition of 14 μ L ammonium acetate (or sodium acetate) 3M and 1 mL of 100% EtOH and mixing the contents by inversion
16. Precipitation of the DNA by centrifugation at 16000g for a 2-minute period at RT and discarding the supernatant
17. Drying the DNA pellet in air and re-suspending in 50 μ L sterile distilled water (or TE buffer)

The isolated genomes were stored at -20°C.

3.6.13. Verification of the genomic integration in isolated genomes

Isolated genomes of the putative transfectants were subjected to PCR in order to verify insertion of the desired construct in the genome. In the case of plasmids harboring selected endogenous SPs, verification PCR was conducted by *GAP* forward and *AOX* reverse primers (Table 3.17). However, in the case of plasmids harboring selected promoters, verification PCR was accomplished by promoter-specific forward primers (Table 3.17) and *AOX*

reverse primer. After confirmation of the genomic integration, genomes concentration (A_{260}) and quality ($A_{260/280}$) were checked with NanoDrop[®] 2000 (Thermoscientific).

Table 3.17. The primers used in PCR for verification of the genomic integration.

Primer Name	Primer Nucleotide Sequence
<i>AOX</i> reverse	5'- GCAAATGGCATTCTGACATCC-3'
<u>Promoter-specific forward primers</u>	
<i>GAP</i> forward	5'- GTCCCTATTTCAATCAATTGAA-3'
<i>PDC-F</i>	5'...TCTATGCATGAGATCGGGACAAGCA...3'
<i>PYRK-F</i>	5'...TAAATGCATGAGATCTTCAGTGTGCGG...3'
<i>THI3-F</i>	5'...GAGATGCATGAGATCGTCTTTGTAAATAGT...3'

3.6.14. Determination of the hGH copy number in *Pichia* transfectants

In order to have a reliable comparison among different developed r-*P.pastoris* strains, based on the expressed rhGH, the copy number of the hGH gene in *Pichia* genome in all recombinant strains should be same and preferentially one. The copy number determination was achieved using real-time quantitative polymerase chain reaction (qPCR). However, regarding to the multiplicity of the selected *Pichia* transfectants, a pre-screening step was implemented prior to qPCR to get rid of probable multi-copy strains

3.6.14.1. Pre-screening of the *Pichia* transfectants

Pre-screening of the selected *P. pastoris* transfectants was conducted according to (Abad et al., 2010). The typical structure of single-digested desired plasmid and the resultant single or double insertions (as

representative of multiple insertions) in *Pichia* genome has been depicted in Figure 3.5. For pre-screening, PCR experiments was conducted by the isolated genomic DNAs of the selected transfectants as template along with two designed primers, i.e., *hGH*-R and *pUC* ori-F (Table 3.18). The location of the *hGH*-R and *pUC* ori-F primers can be observed in Appendix F.

Table 3.18. Utilized forward and reverse primers in pre-screening PCR.

Primer Name	Length (bp)	Nucleotide Sequence (5'→3')
<i>pUC</i> ori -F	35	GGATCTCAAGAAGATCCTTTGATCTTTTCTACG GG
<i>hGH</i> -R	32	AATGTCTCGACCTTGTCCATGTCCTTCCTGAA

Based on the Figure 3.5, while having multi-copy insertion in genome, does not matter in one side or both sides of the desired locus, there will be a band in pre-screening PCR results of the isolated genomes. However, if there is no band we can not strictly say something about single-insertion, since sometimes insertion happens in place other than expected locus. Therefore, the result of this pre-screening was just elimination of some of the multi-copy strains.

3.6.14.2. qPCR with genomes of the pre-screened *Pichia* transfectants

qPCR, specially absolute quantification, was the method of choice in current study for copy number determination. Furthermore, a reference gene was recruited and was measured parallel with target gene, i.e., *hGH* gene. *ARG4* gene (argininosuccinate lyase or argininosuccinase) was used as reference gene (Abad et al., 2010; Hartner et al., 2008) and was quantified parallel with the unknown *hGH* gene in genomic DNA samples. This gene acted as an endogenous control (housekeeping gene) because it is a single-copy gene

in *P. pastoris*. Using reference gene has the advantage of omitting the need for exact quantification and loading of the starting template. Consequently, the quantification was regardless to initial DNA amount. This compensates for any difference in the amount of the initial samples. Actually, *ARG4* represents the copies of the genome available in the sample and the *hGH* assayed copy is normalized by dividing, virtually, over the genome number. By inclusion of this gene in qPCR assays, the normalized *hGH* copy number can be calculated as follows:

$$\text{Copy number}_{hGH} = \frac{\text{Assayed copy quantity}_{hGH}}{\text{Assayed copy quantity}_{ARG4}} \quad (3.2)$$

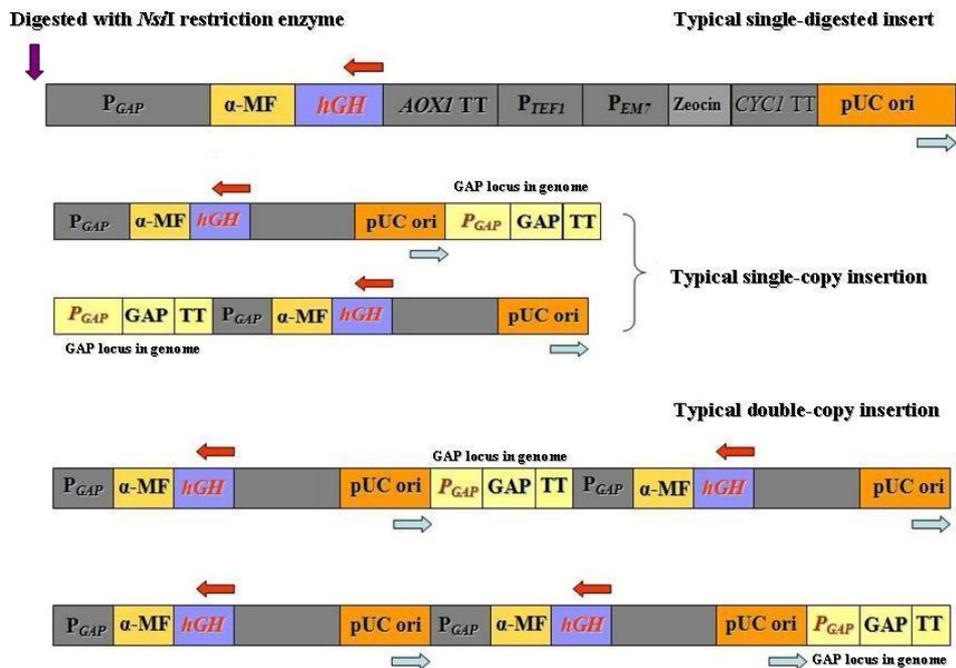


Figure 3.5. Schematic representation of the typical single- or double-copy insertion in *Pichia* genome and location of the *hGH* reverse (red arrow) and *pUC* ori forward (blue) primers. All constructs linear form is available in Appendix G.

3.6.14.2.1. Provision of the standard curve

Genomic DNAs of the selected transformants were the unknown samples in current research. Absolute quantification makes it necessary to have a standard curve (using a standard DNA solution) for copy number measurement in genomic DNA samples. However, since there are two genes that should be measured, *hGH* and *ARG4*, two standard curves should be prepared by two standard samples. The utilized primers are the same in both unknown samples (genomic DNA) and standard samples, thus, for making the conditions in standard samples closer to the conditions in genomic DNA samples in both cases, namely *hGH* and *ARG4*, standard DNA samples were prepared for both genes with the outer primer set by regarding the concept of the nested PCR (Wilhelm et al., 2003). In nested PCR two sets of primers called outer and inner primers are used; the outer primers prepare a template (standard DNA) that inner primers can amplify a part of it during qPCR experiment.

3.6.14.2.1.1. Preparation of the standard DNA sample for *ARG4* gene

There is just one copy of *ARG4* gene in *P. pastoris* genome; its nucleotide sequence can be found in Appendix H. *ARG4* gene chromosomal location is PAS_chr1-1_0389. For *ARG4* standard DNA preparation, a 330-bp fragment from this gene was selected. The outer primers for preparation of standard DNA for *ARG4* genes can be found in Table 3.19. *ARG4* outer primers location has been presented in Appendix H.

For preparation of the standard DNA, genomic DNA of *P. pastoris* was used as template. First a preliminary PCR with different combinations of primers, MgCl₂, template was performed in order to achieve the optimum conditions with absence of non-specific fragments in prepared *ARG4* standard. The PCR results, then, were purified using PCR purification kit and the concentration of the sample was measured using NanoDrop[®] 2000 (Thermoscientific).

Table 3.19. Outer primers for preparation of the standard DNA for *ARG4* gene.

Primer Name	Length (bp)	Nucleotide Sequence (5'→3')
<i>ARG4</i> -Std-F	22	CTTGAACATTGATGCCGAACGA
<i>ARG4</i> -Std-R	23	GACTCTAGCTTTTCATTTCAGTGC

3.6.14.2.1.2. Preparation of the standard DNA sample for *hGH* gene

pGAPZ α A::*hGH* as base plasmid (developed in current research) was used as template for preparation of the standard sample for *hGH* gene. *GAP* forward and *AOX* reverse primers (Invitrogen) were used as outer primer pair (Table 3.20) which resulted in a 1090-bp segment (Appendix I).

Table 3.20. Outer primers for preparation of the standard DNA for *hGH* gene.

Primer Name	Length (bp)	Nucleotide Sequence (5'→3')
<i>GAP</i> forward	22	GTCCCTATTTCAATCAATTGAA
<i>AOX</i> reverse	21	GCAAATGGCATTCTGACATCC

First a preliminary PCR was conducted with different combinations of primer, MgCl₂, and template amount in order to obtain the optimum conditions to eliminate non-specific fragments in desired standard. The PCR results, then, were purified using PCR purification kit and the concentration of the samples was measured using NanoDrop[®] 2000 (Thermoscientific).

3.6.14.2.2. Inner primers for qPCR

In order to render the qPCR experiments there should be a primer pair for amplifying a suitable selected amplicon. These primers are able to bind to a

segment inside the prepared standard DNA (with outer primers) and amplify it. The used inner primers in current research can be observed in Table 3.21. The inner primer pair for *ARG4* was the ones used in Abad et al., study (2010) which leads to a 84-bp fragment. Regarding *hGH* gene, based on previous research (Bayraktar, 2009), there was an available pair for qPCR analysis which results in a 238-bp fragment. Therefore, this pair was also ordered and used which can be seen in Appendix H.

Table 3.21. Inner primers for conducting qPCR experiments.

Primer Name	Length (bp)	Nucleotide Sequence (5'→3')
<u>For <i>hGH</i> standard</u>		
<i>Eda-qPCR-F</i>	20	GCCTTTGACACCTACCAGGA
<i>Eda-qPCR-R</i>	20	ACACCAGGCTGTTGGCGAAG
<u>For <i>ARG4</i> standard</u>		
<i>ARG-F</i>	21	TCCATTGACTCCCGTTTTGAG
<i>ARG-R</i>	19	TCCTCCGGTGGCAGTTCTT

3.6.14.2.2.1. Preliminary PCR with inner primers designed for *hGH*

The inner primer pair along with prepared *hGH* standard solution were used in qPCR experiments. However, prior to experiments, again a preliminary PCR was conducted with *hGH* standard and *P. pastoris* genomic DNA as template and inner primers in determined qPCR conditions in order to find out the optimum primer concentrations for qPCR. Based on the results, primer concentration of 200 nM was used in qPCR experiments.

3.6.14.2.2.2. Preliminary PCR with inner primers designed for *ARG4*

The same concept as *hGH* gene holds true for *ARG4* gene and, therefore, a preliminary PCR was conducted with *ARG4* standard and *P. pastoris*

genomic DNA in order to verify the optimum primer concentrations which can be used in related qPCR experiments. The PCR conditions were similar to the conditions used in the case of *hGH* gene. Primer concentration of 200 nM was again used in qPCR experiments as the optimum amount.

3.6.14.2.3. Rendering qPCR experiments

After finding the optimum concentration for inner primers, the qPCR mixture should be prepared. The fluorescent dye mixture (Faststart Universal SYBR Green Master) contains the optimum amount of buffer, MgCl₂ and dNTPs and just dH₂O, primers, and template should be added separately. Based on the aim of the measurements there should be six sets of tubes:

- Tubes with genomic DNA and *Eda-qPCR-F* and *Eda-qPCR-R* inner primers
- Tubes with genomic DNA and *ARG4-F* and *ARG4-R* inner primers
- Tubes with serial dilutions of the *hGH* standard sample and *Eda-qPCR-F* and *Eda-qPCR-R* inner primers
- Tubes with serial dilutions of the *ARG4* standard sample and *ARG4-F* and *ARG4-R* inner primers
- No-template tubes for control of any contamination, using dH₂O instead template, and *Eda-qPCR-F* and *Eda-qPCR-R* inner primers
- No-template tubes for control of any contamination, using dH₂O instead template, and *ARG4-F* and *ARG4-R* inner primers

According to the Table 3.22, a master mix was prepared by the components except template regarding the amount of the tubes needed in experiment; after mixing by pipetting, and a brief spin, 18 µL of the master mix was transferred into each qPCR tube and, afterwards, 2 µL template (standard DNA or genomic DNA) was added separately to the tubes. For non-template tubes, dH₂O was added. Genomic DNAs were used in dilutions between 1/200 and 1/300 to have the amount of 2 ng to 8 ng in the final mixture. In the case of the standard DNA samples, 1/100 dilutions of both standard samples were prepared and stored at -20°C. Subsequently, daily dilutions of

these stocks were prepared $1/10^3 \rightarrow 1/10^7$ in each qPCR experiment. qPCR experiments were conducted with QIAGEN Corbett Rotor-Gene 6000 series with 36 wells with the thermal profile available in Table 3.23.

Table 3.22. qPCR mixture used in copy number determination experiments. The amounts are for one tube and for master mix preparation they should be multiplied by (No. tubes + 1). 2 μ L of genomic DNA or standard DNA is added to the 18 μ L of the prepared mix.

Component	Amount (μ L)
dH ₂ O	6.4
Faststart Universal SYBR Green Master 2X (ROX)	10
Forward primer (5 μ M)	0.8
Reverse primer (5 μ M)	0.8

The copy number of the prepared standard samples and their corresponding dilutions were calculated with the following equation (Lee et al., 2006):

$$DNA \text{ (copy)} = \frac{6.02 \times 10^{23} (\text{copy/mol}) \times DNA \text{ amount (g)}}{DNA \text{ length (bp)} \times 660 \text{ (g/mol/bp)}} \quad (3.3)$$

Where, 660 g/mol is the mean molar mass of a base pair (Wilhelm et al., 2003) and 6.022×10^{23} is the Avogadro's number. Therefore, by knowing the copy number of the standards, the assayed copy number of the *hGH* gene and *ARG4* gene was measured.

Table 3.23. qPCR experiments thermal profile. After reaching to fluorescent level 100% before 45 cycles, the remaining steps could be skipped.

Denaturation	Amplification	Melting	Cooling
1 Cycle	45 Cycles	1 Cycle	1 Cycle
Hold: 95°C for 10 min	Denaturation: 95°C for 10 sec. Annealing: 55°C for 5 sec. Elongation: 72°C for 10 sec.	Melting: 50 °C to 99°C continuous with slope of 1°C/s.	Keeping: 40°C for 30 sec.

3.7. rhGH production with selected r-*P.pastoris* strains

After selection of the *P. pastoris* strains with similar (single) copy numbers, rhGH production was started. Production was conducted in both shake flask air-filtered bioreactor and laboratory-scale bioreactor.

3.7.1. Shake flask air-filtered bioreactor experiments

In the case of secretion SPs, the selected single-copy strains of r-*P.pastoris* were inoculated from glycerol stocks into YPD agar plates with Zeocin™ concentration of 100 µg/mL and after ca. 50 hours incubation at 30°C, inoculation was made into 25 mL BMGY medium with chloramphenicol and PTM1 in 150-mL glass flasks. Cultivation started at 30°C and 200 rpm. When the cells reached to OD₆₀₀ ~ 18 they were harvested at 2000g for 10 minutes at room temperature (RT) and, subsequently, were transferred into 50 mL production medium in 250-mL baffled Erlenmeyer flasks in a manner to have OD₆₀₀ of ~2. Production was conducted at 30°C and 200 rpm and lasted for 40-48 hours. Samples which were taken at t=24 h were used in dot-blot and SDS-PAGE analyses. Shake flask air-filtered bioreactor experiments were conducted in triplicate. The schematic of the procedure can be found in Figure 3.6.A.

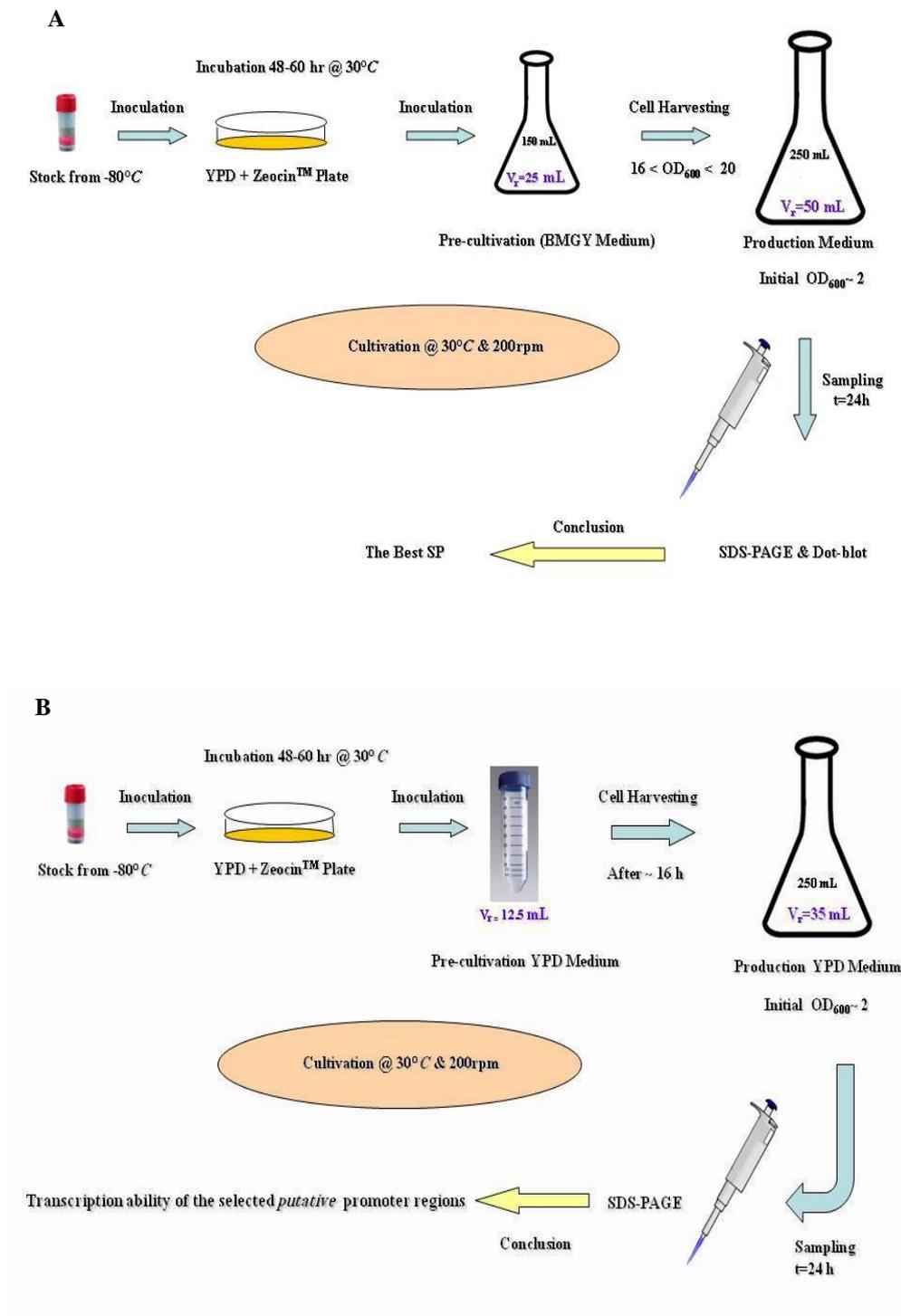


Figure 3.6. Schematic representation of the steps in shake flask bioreactor experiments. A: experiments related with SPs. B: preliminary experiment related with selected promoters.

On the other hand, as preliminary experiment related with promoter constructs prior to the bioreactor experiments, the selected single-copy strains of r- *P. pastoris* were inoculated from glycerol stocks into YPD agar plates containing Zeocin™ with concentration of 100 µg/mL; after incubation at 30°C for approximately 50 hours, inoculation was made into 50-mL sterile falcons containing 12.5 mL YPD broth with chloramphenicol and incubation was conducted at 30°C and 200 rpm as pre-cultivation step. After 16 hours the cells were harvested at 2000g for 10 minutes at RT. Harvested cells were transferred into 250-mL Erlenmeyer flasks with 35 mL YPD in a manner to have OD₆₀₀ of approximately 2 (Figure 3.6.B). Cultivation was conducted for 24 hours at 30°C and 200 rpm; samples were taken at t=24 h in order to be subjected to SDS-PAGE analysis.

3.7.2. Laboratory-scale bioreactor experiments

During conducted research, two sets of bioreactor experiments were carried out; one for confirmation of the shake flask experiments related with SPs and the other related with promoter constructs in order to choose the best oxygen limitation-induced promoter. Bioreactor experiments were conducted in a 3 L laboratory-scale bioreactor (Braun CT2-2) with working volume of 0.8 – 2.2 L. Bioreactor specifications are available in Table 3.24.

Table 3.24. Specifications of the laboratory-scale bioreactor used in current research.

Feature	
Tank diameter	T=12 cm
Impeller diameter	D=T/2.4
Impeller length	H=T/12
Off-bottom clearance	C=T/6
Average height of working volume	H=0.667T

3.7.2.1. Bioreactor experiments with r-*P. pastoris* strains of SPs

Desired r-*P. pastoris* strain was inoculated from glycerol stock into YPD agar plate with ZeocinTM concentration of 100µg/mL and was incubated for 48-60 hours at 30°C. Afterwards, inoculation was made from the plate into the BMGY medium for pre-cultivation. Reaching OD₆₀₀ to the range of 6-8 the cells were harvested by centrifugation at 2000g for 10 minutes at RT. The harvested cells were transferred with sterile ultra-pure water into the bioreactor containing BSM medium with pH of 5, adjusted with ammonium hydroxide 26%, to have initial OD₆₀₀ of ~2. The batch phase of the fermentation was started at 900 rpm and 30°C and pH 5. Dissolved oxygen was measured on-line with VisiFermTM DO 120 optical sensor (HAMILTON[®], Allston, MA, USA) and its content was kept constant at 20% air saturation. After reaching to $90 < OD_{600} < 100$, the fed-batch phase was started. The fed-batch phase of the fermentation was accomplished by exponential feeding of 500 g/L glucose solution containing chloramphenicol and PTM1 with a peristaltic pump controlled by a bioreactor system, to maintain the specific growth rate as 0.07 h^{-1} . At three-hour time intervals samples were taken from the bioreactor and were analyzed. Throughout the experiments temperature was controlled at $30.0 \pm 0.1 \text{ }^{\circ}\text{C}$ with PI controller of the bioreactor, pH was controlled at 5.0 ± 0.1 by 26% NH₄OH and dissolved oxygen concentration was controlled by air supply. 10% antifoam was added manually whenever needed in order to prevent foaming. The schematic illustration of the experiment has been presented in Figure 3.7.

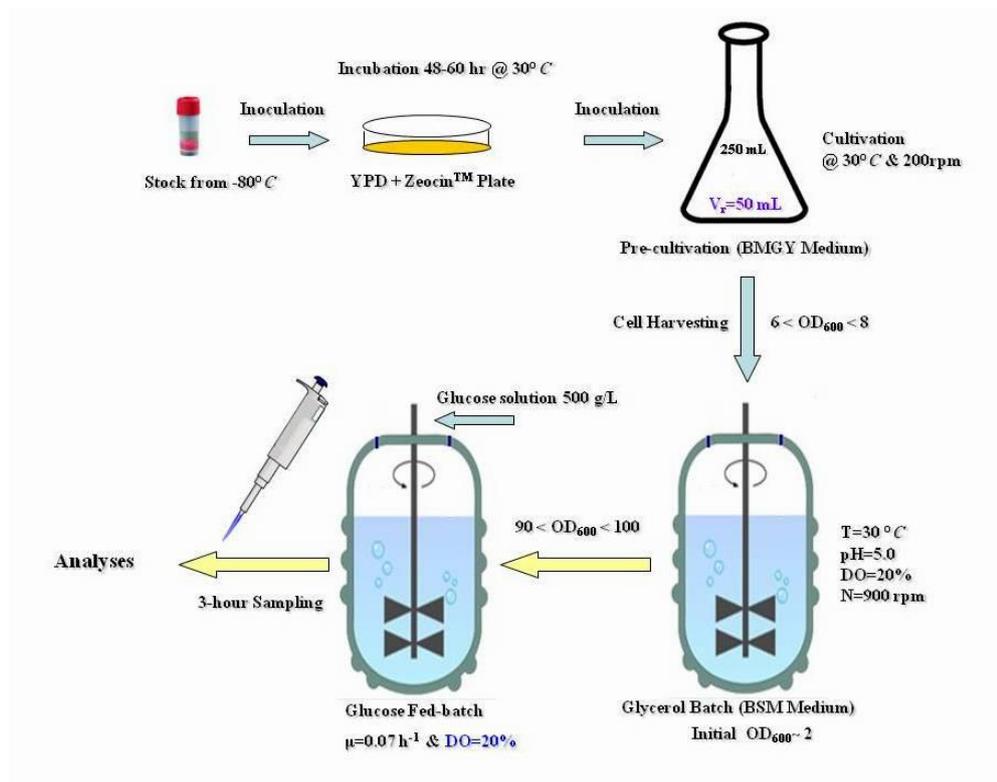


Figure 3.7. Schematic representation of the steps in bioreactor experiments related with SPs.

3.7.2.2. Bioreactor experiments with r-*P. pastoris* strains of promoters

Desired r-*P.pastoris* strain was inoculated from glycerol stock into YPD agar plate with ZeocinTM concentration of 100 μ g/mL and was incubated for 48-60 hours at 30°C. Afterwards, inoculation was made from the plate into 50 mL BMGY medium in 250-mL Erlenmeyer flask at 30°C and 200rpm for pre-cultivation. Reaching OD₆₀₀ to the range of 16-20 (~ 18) cell cycle synchronization (explained in section 3.7.2.2.1) was started by harvesting the cells and transferring to 50 mL YEP medium in 250-mL Erlenmeyer flasks and incubation at 30°C and 200rpm. After completion of the synchronization procedure, the harvested cells were transferred with sterile ultra-pure water into the bioreactor containing BSM medium with pH of 5, adjusted with ammonium hydroxide 26%, to have initial OD₆₀₀ of ~2. The batch phase of the fermentation was started at 900 rpm and 30°C and pH 5. Dissolved

oxygen content was kept constant at 20% air saturation. After reaching to $45 < OD_{600} < 50$, the fed-batch phase was started and DO set point was decreased to 4% in order to exert oxygen-limitation conditions. The fed-batch phase of the fermentation was accomplished by exponential feeding of 500 g/L glucose solution containing chloramphenicol and PTM1 with a peristaltic pump controlled by a bioreactor system, to maintain the specific growth rate as 0.15 h^{-1} . At three-hour time intervals samples were taken from the bioreactor and were analyzed. Throughout the experiments temperature was controlled at $30.0 \pm 0.1 \text{ }^\circ\text{C}$ with PI controller of the bioreactor, pH was controlled at 5.0 ± 0.1 by 26% NH_4OH and dissolved oxygen concentration was controlled by air. 10% antifoam was added manually whenever needed in order to prevent foaming. The summarized procedure can be seen in Figure 3.8.

3.7.2.2.1. Cell cycle synchronization

Briefly, the cells were grown to a specific density and, then, were subjected to starvation for a defined period. Afterwards, the cells were transferred to a fresh nutrient medium. Consequently, cells were supposed to be arrested in G1 phase (Banflavi, 2011). The protocol is as follows:

1. Growing *P. pastoris* cells in BMGY (pre-cultivation) medium until saturation (exhaustion of nutrients), approximately an OD_{600} of 16-20 ($OD_{600} \sim 18$). Liquid log phase cultures are typically used for most of synchronization protocols (Banflavi, 2011). Note that doubling time for wild-type *P. pastoris* is in the range of 2-3 hours.
2. Centrifugation at 1800-2000g for 1-2 minutes
3. Discarding the supernatant and repeating step 2
4. Re-suspending of the pellets in sterile YEP medium and incubation at 30°C for 6 hours
5. Centrifugation at 1800-2000g for 1-2 minutes
6. Discarding the supernatant and repeating step 5

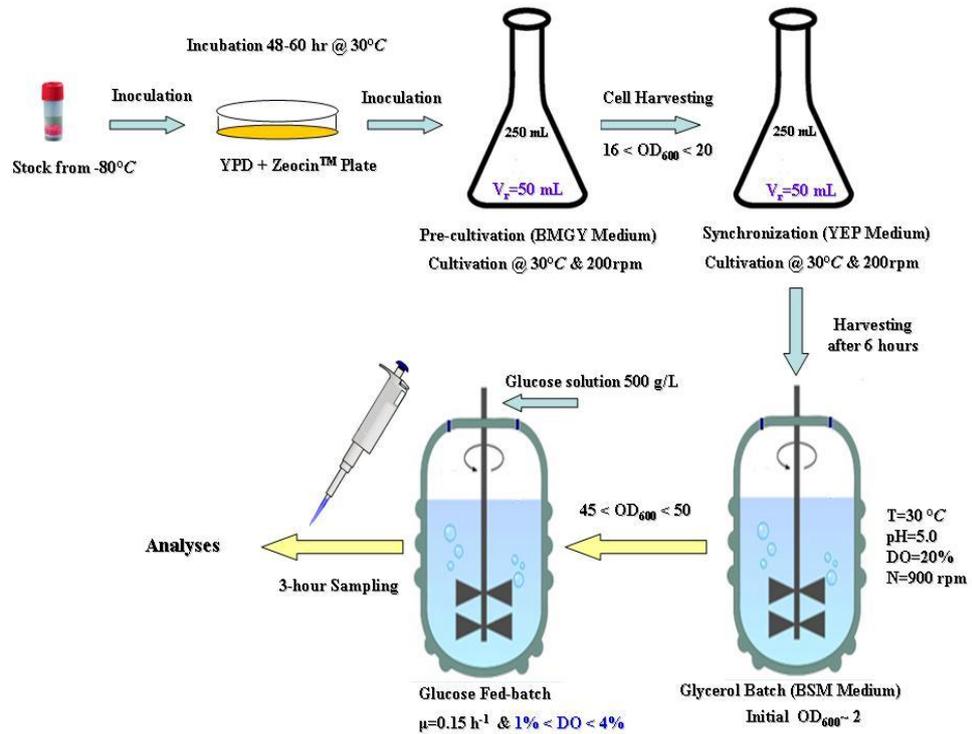


Figure 3.8. Schematic representation of the steps in bioreactor experiments related with promoters.

3.7.2.2.2. Total RNA isolation

Since the strength of the promoter is closely related with the mRNA level, in addition to the extracellular rhGH, the amount of mRNA was also measured to judge between selected promoters. mRNA should be extracted as quickly as possible after obtaining samples. For better results, either fresh samples or samples that have been quickly frozen in liquid nitrogen and stored at -70°C should be utilized.

High pure RNA isolation kit for small-scale (mini) preparations of RNA (Roche Diagnostics) was utilized. Isolation of total RNA from yeast has been suited for 1×10^8 cells. At the start point, use a dilution of cells which gives OD_{600} of 0.1 -0.15/mL (absorbance of 0.1 corresponds to

approximately 2×10^6 cells). It is better to harvest the yeast cells during mid- or late-log phase ($OD_{600} < 2.0$). After preparation of the working solutions based on manufacturer's instructions, the protocol can be followed:

1. Collecting the cells by centrifugation at 2000g for 5 minutes and removing the supernatant
2. Re-suspending the cells in 200 μ L PBS
3. Transferring the suspension to a sterile 1.5-mL eppendorf tube
4. Addition of 10 μ L Lyticase (Sigma #L4025) (1 mg/mL)
5. Incubation for 15 minutes at 30°C
6. Addition of 400 μ L Lysis/-Binding buffer (green cap)
7. Vortexing for 15 seconds
8. Inserting one filter tube in one collection tube and subsequent transferring of whole sample into the upper reservoir of the filter tube (max. 700 μ L)
9. Centrifugation of the tube assembly at 8000g for 15 seconds
10. Discarding the flow through liquid and re-assembling the filter tube and collection tube
11. Mixing 90 μ L DNase incubation buffer (white cap) with 10 μ L DNase I in a sterile reaction tube
12. Pipeting the solution on the glass filter fleece in the filter tube
13. Incubation at 15°C to 25°C for 15 minutes
14. Addition of 500 μ L wash buffer I (black cap) to the upper reservoir of the filter tube assembly
15. Centrifugation at 8000g for 15 seconds
16. Discarding flow through
17. Addition of 500 μ L wash buffer II (blue cap) to the upper reservoir of the filter tube assembly
18. Centrifugation at 8000g for 15 seconds
19. Discarding flow through
20. Addition of 200 μ L wash buffer II (blue cap) to the upper reservoir of the filter tube assembly

21. Centrifugation at 14000g for 2 minutes in order to remove residual wash buffer. Extra centrifugation time can be applied
22. Discarding collection tube and inserting the filter tube into a clean sterile 1.5-mL eppendorf tube
23. Addition of 50-100 μ L elution buffer to the upper reservoir of the filter tube
24. Centrifugation at 8000g for 1 minute
25. The eppendorf tube now contains the eluted RNA which can be directly utilized in subsequent experiments or can be stored at -80°C for later analyses.

Isolated total RNA can be used directly in first-strand cDNA synthesis. It is recommended to use 1-10 μ L of isolate in RT reaction. Performing an RT-minus control (without addition of reverse transcriptase to the cDNA synthesis reaction) shows the presence of residual genomic DNA.

3.7.2.2.3. Complementary DNA synthesis

The isolated RNA should be reverse transcribed to complementary DNA (cDNA) in order to be recruited in qPCR experiments. The isolated total RNA was used to prepare cDNA by utilizing Transcriptor First Strand cDNA Synthesis kit (Roche Diagnostics). Briefly, the procedure can be summarized as follows:

1. Thawing all the frozen reagents and brief centrifugation and keeping on ice during reaction set up
2. Preparing a sterile thin-walled PCR tube
3. Addition of components in the following order (Table 3.25) in order to prepare template-primer mixture (for one reaction):

Table 3.25. Order and amount of the components to prepare template-primer mixture. Suitable template concentration may range from 10 ng to 5 µg total RNA and from 1 to 100 ng mRNA.

Component	Volume	Final Concentration
Total RNA or poly[A] ⁺ mRNA		1 µg total RNA or 10 ng poly[A ⁺] RNA
Primer oligo(dT) ₁₈ 50 µM	1 µL	2.5 µM
PCR grade water	Variable	To make total volume 13 µL
Total volume	13 µL	

4. Denaturation of template-primer mixture by heating the tube for 10 minutes at 65°C and immediate cooling on ice for ensuring denaturation of RNA secondary structure (optional).
5. Addition of the components of the RT mix to the tube in the following order (Table 3.26):

Table 3.26. Order and amounts of RT mix component.

Component	Volume	Final Concentration
Transcriptor Reverse Transcriptase Reaction Buffer, 5X	4 µL	1X (8 mM MgCl ₂)
Protector RNase Inhibitor 40 U/µL	0.5 µL	20 U
dNTP mix 10 mM each	2 µL	1 mM each
Transcriptor Reverse Transcriptase 20 U/µL	0.5 µL	10 U
Total volume	20 µL	

6. Mixing the tube content without vortexing
7. Brief spinning
8. Putting the tube in a thermal block cycler with a heated lid (to minimize evaporation)
9. Incubating the RT reaction at 55°C for 30 minutes

10. Inactivating the Transcriptor reverse transcriptase by heating at 85°C for 5 minutes
11. Stopping the reaction by placing the tube on ice
12. Storing the tube at -15 to -25°C for long periods

The resultant cDNA can be used 1-5 µL as template in PCR without any purification.

3.8. Analysis

After production of rhGH, samples of the shake flask and bioreactor experiments were subjected to various analyses.

3.8.1. Measurement of the cell concentration

Dry cell concentration (g/L) was calculated by measuring OD₆₀₀ of the samples with Spectroquant[®] Pharo 300 UV-Vis spectrophotometer (Merck KGaA, Darmstadt, Germany) by applying the following equation:

$$C_x = 0.24 \times OD_{600} \times Dilution \ factor \quad (3.4)$$

3.8.2. SDS-PAGE

The samples taken from the fermentation broth were centrifuged at 2000g for 10 minutes at +4°C. Then, the supernatant was separated and filtered with 0.2 µm filter (Minisart, Sartorius stedim biotech, Goettingen, Germany) into a new sterile tube. The filtrate was used in SDS-PAGE analysis.

The primary used 3X SDS loading buffer was without reducing agent; 20 µL of the filtrate was mixed with 10 µL of 3X SDS loading buffer and 15 µL of the mixture was loaded into the prepared SDS-PAGE gel after thorough mixing.

In order to improve the results, 4X loading buffer was utilized (Appendix B); 13 µL of the filtrate was mixed with 5 µL of 4X SDS loading buffer and, then, 2 µL Dithiothreitol (DTT) 1M was added. Afterwards, the contents were mixed by pipetting and were put in water bath with

temperature of 95°C. Then, 15 µL of the mixture was loaded into the prepared SDS-PAGE gel. In parallel, hGH standard prepared like samples with concentration of 50 mg/L or 2.5 mg/L and was loaded for verification and quantification of rhGH. A protein marker (Thermoscientific) was also utilized to track the progression of the SDS-PAGE. Polyacrylamide gels were prepared by Bio-Rad's TGX™ and TGX stain-free™ FastCast™ acrylamide kits. Electrophoresis was conducted at 40 mA for approximately 45 minutes. The gels were silver-stained with the available protocol (Table 3.27). The taken images of the gels were visualized by imaging system (UVP, Upland, Canada) and the amount of the secreted rhGH was quantified.

Table 3.27. Sequential procedure of the silver nitrate staining. Composition of the solutions are available in Appendix B.

Name of the step	Used solution	Time
Fixing	Fixer solution	> 1 hour
Washing	50% EtOH	3×20 min.
Rinsing	dH ₂ O	2×20 sec.
Pretreatment	Pretreatment solution	1 min
Rinsing	dH ₂ O	3×20 sec (exactly)
Impregnation	Silver nitrate solution	20 min
Rinsing	dH ₂ O	2×20 sec (exactly)
Developing	Developing solution	~ 5 min
Washing	dH ₂ O	Addition to the developing solution to slow down the reaction
Stop	Stop solution	> 10 min

3.8.3. Dot-blot

The detailed procedure of the dot-blot is as follows:

1. Membrane preparation; cutting the membrane and marking one corner for easy locating of samples

2. Pre-wetting the membrane with methanol for 15 seconds
3. Immersing in dH₂O for 2 minutes and placing on a smooth clean surface
4. Dropping 5-10 μ L of the sample and repeating it for 3 rounds total amount of 15-30 μ L; letting the sample to be air-dried
5. Washing the membrane 2-3 times with TBS-T (Appendix B); Tween improves hybridization
6. Immersing the membrane in TBS-T milk (Appendix B) solution for 1 hour at RT with shaking in order to block unbound surface of the membrane
7. Washing the membrane with TBS-T three times with shaking at RT (15 minutes, 5 minutes, 5 minutes) with fresh changes of large volumes of TBS-T. Unspecific binding of antibodies is prevented using Tween in TBS-T
8. Dilution of the primary antibody (1:5000 or 1:10000) to the concentration of 1 μ g/mL in TBS-T
9. Transferring the membrane into a low volume container and incubation in the diluted primary antibody for 1 hour with shaking
10. Taking the membrane to another box and washing as step 7
11. Diluting the secondary antibody to 1:5000 (1:10,000) in TBS-T (Dilution factor should be determined empirically for each Antibody as 1:1000 – 1:10000; more dilution will increase the linearity and sensitivity)
12. Transferring the membrane to the small container and incubation in the diluted secondary antibody for 1 hour at RT with shaking
13. Transferring the membrane to a container and washing three times with fresh changes of large volumes of TBS-T buffer each for 10 minutes by shaking at RT
14. Visualizing the dots by using DAB substrate

After completion of the procedure, the photo of the membrane was taken and was visualized by imaging system to measure the rhGH concentration (UVP, Upland, Canada).

3.8.4. Total protein measurement

In order to measure total protein, Bradford assay was conducted. This procedure is based on the formation of a complex which is made by the dye (Brilliant Blue G) and protein in solution. There will be a shift in the maximum absorption of dye from 465 nm to 595 nm as the effect of the protein-dye complex. The protein amount is proportional to the absorption. The linear concentration range is 0.1-1.4 mg/mL of protein (Bovine serum albumin (BSA) as standard). The sample during the assay can be blank, standard, or unknown sample. The blank will be buffer without protein. Color development begins immediately after addition of the Bradford reagent (Sigma) to the protein sample. The procedure was as follows:

1. Gentle mixing of the Bradford reagent and bringing it to room temperature
2. Preparation of the standard samples by Bovine serum albumin as standard protein. For this purpose, 2 mg/mL BSA solution was prepared in the same buffer of the unknown samples and subsequently diluted (with buffer), afterwards.
3. Addition of 3 mL Bradford reagent to 100 μ L of each protein sample (standard, unknown, blank) in glass test tubes
4. Gentle Vortexing of the tubes for complete mixing (total volume of 3.1 mL)
5. Incubation at room temperature for 5-45 minutes; the protein-dye complex will be stable for 60 minutes
6. Transferring the test tube contents to cuvettes
7. Reading the absorbance of samples at 595 nm and plotting standard curve using standard samples' absorbance

8. Determination of the protein concentration of the unknown samples by comparison with standard curve.

3.8.5. Measurement of the rhGH concentration

Although, rhGH production was decided to be extracellular in present research, in order to check the efficiency of SPs, in bioreactor experiments related with SPs strains, intracellular rhGH was also measured.

3.8.5.1. Measurement of the intracellular rhGH

Intracellular rhGH measurement was conducted just for bioreactor samples related with *r-P.pastoris* strains of the endogenous SPs. The recruited procedure is as explained:

1. pelleting the 3-hour samples of bioreactor at 2000g and 4°C for 10 minutes
2. Washing the pelleted cells twice with 50 mM Tris-Cl buffer (pH 7.5)
3. Performing centrifugation at 2000g and RT for 10 minutes after each run of washing
4. Suspending the cells in 1 mL of the lysis buffer (Karaoglan et al. 2014) with minor pH modification : 20 mM HEPES pH 7.5 , 0.2 mM EDTA, 0.5 mM DTT (Dithiothreitol), 0.5 mM PMSF (Phenylmethanesulfonyl fluoride, as protease inhibitor), 0.42 M NaCl, 1.5 mM MgCl₂, 10% glycerol
5. Disrupting the cells with 0.5 mm glass beads in MM200 mixer mill (Retsch[®], Haan, Germany) in 3 cycles including 2 minutes disruption and 3 minutes cooling in ice with oscillation frequency of 30s⁻¹
6. Centrifugation at 16100g and +4°C for 15 minutes
7. Dilution of the supernatant samples to 1 µg/µL of total protein with lysis buffer
8. Conducting dot-blot analysis to determine the intracellular rhGH concentration.

3.8.5.2. Measurement of the extracellular rhGH

To determine the concentration of the extracellular rhGH, combination of SDS-PAGE, silver nitrate staining and image visualization was used. SDS-PAGE was performed according to (Laemmli, 1970) with some modifications such as use of ready solutions for gel preparation (explained in section 3.8.2). Silver nitrate staining (Blum et al., 1987) was performed for protein bands detection. The gels images were used to quantify rhGH by densitometric method.

3.8.6. Western blotting

In this analytical method the proteins which were separated by SDS-PAGE are transferred from gel into the membrane (nitrocellulose or PVDF) and, then, are stained with immunostaining with specific primary antibody of desired protein (here, hGH). Detailed protocol can be explained as follows:

1. Preparation of the protein sample whether from cell lysate or supernatant of fermentation medium (after pelleting the cells)
2. Running SDS-PAGE
3. Transferring gel, while still attached to the glass plates, into the transfer buffer/blotting buffer (Appendix B) in a tray and keeping for 15-20 minutes at room temperature in order to remove salts and SDS
4. Detaching the gel from the glasses
5. Cutting off the stacking gel and nick one corner of the resolving gel for orientation
6. Cutting the (PVDF) membrane and two blotting papers in suitable size and cutting one corner of the membrane
7. Pre-wetting the membrane in 100% methanol for a few seconds (5-15 seconds) and, then, incubation in ultra-pure water for 2 minutes
8. Putting the membrane in transfer buffer (in tray) for at least 10 minutes in order to replace water and equilibration
9. Putting two blotting paper and two sponges in tray containing transfer buffer in order to be wetted.

10. Opening the cassette and placing the opened cassette in tray filled with transfer buffer in a manner that the black plate of cassette sinks in buffer

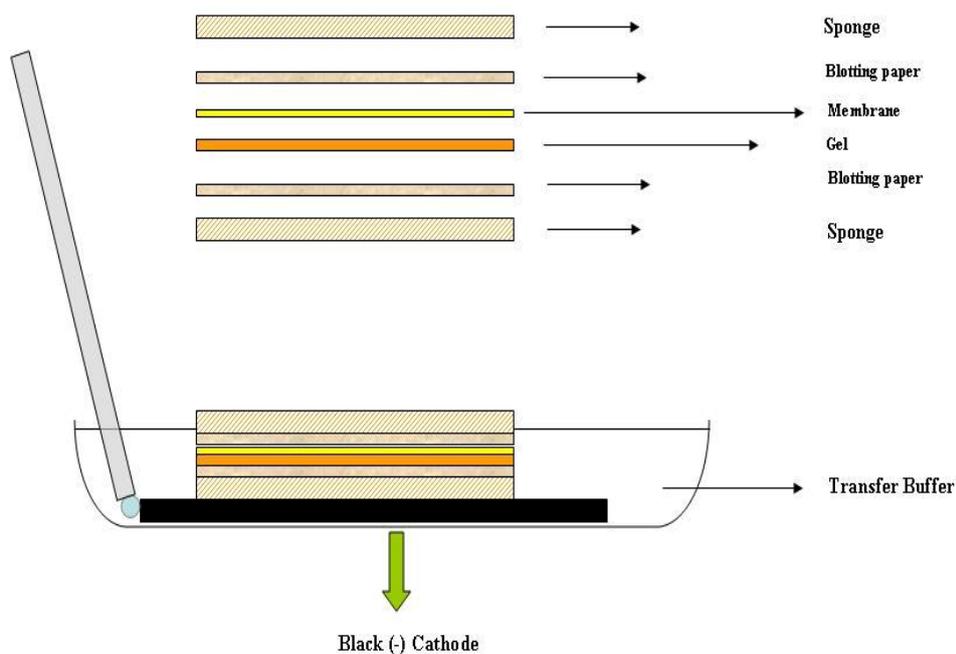


Figure 3.9. Order of the components of the transfer stack in Western blot analysis.

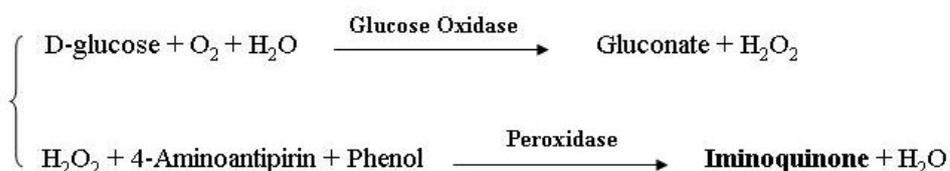
11. Assembling the transfer stack according to the Figure 3.9 in order to let the proteins to migrate from gel toward the membrane upon exerting an electric potential between cathode (-) and anode (+). Note that air bubbles should be removed from between gel and membrane by careful rolling of a glass pipette or glass test tube on each layer in stack
12. Closing the cassette and pressing lightly in order to be locked by its tabs
13. Putting the cassette in the transfer module in a manner that the hinges face up so that the black side of cassette faces the black cathode panel

14. Throwing a magnetic stirrer in transfer tank
15. Pouring buffer in transfer tank. It should be between min-max lines
16. Closing the lid of the transfer tank and turning on the stirrer and turning on the power supply and starting the blotting at cold room, at 50 V for 3 hours regardless of the number of the gels
17. Before starting immunostaining, let the blot air-dry to improve the protein binding.

After blotting the procedure will be like dot blot explained in section 3.8.3.

3.8.7. Measurement of the residual glucose

The glucose concentration in bioreactor samples was measured by Biyozim kits (Biasis, Turkey) based on the manufacturer's instructions. The analysis is based upon the following two reactions:



The red-color iminoquinone with maximum absorbance at 505 nm is the base of measurement. It should be emphasized that the samples should be homogenous, clear and colorless. Furthermore, the tubes, analysis solutions and dH₂O all should be in reaction temperature. The constituents and their amounts are available in Table 3.28.

Briefly, the procedure can be summarized as follows:

1. Preparation of the solutions based on manufacturer's instructions
2. Numbering the test tubes
3. Keeping tubes and solutions in reaction temperature
4. Addition of dH₂O to the tubes in order to have final volume of 2.5 mL

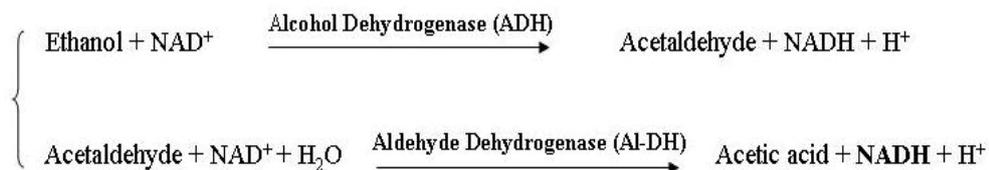
5. Addition of 400 μL buffer solution (containing KH_2PO_4 and phenol) to the tubes
6. Addition of 50 μL reactive solution (containing glucose oxidase, peroxidase and 4-aminoantipirin) to the tubes
7. Addition of 50 μL fermentation sample or standard to the corresponding tube
8. Vortexing gently for a few seconds
9. Incubation at 25°C for 40 minutes or at 37°C for 20 minutes (50 rpm)
10. Blanking spectrophotometer with blank sample
11. Reading absorbance of fermentation samples (and standards) at 505 nm

Table 3.28. The amount of the ingredients in glucose concentration measurement.

Ingredient	Blank tube	Sample tube
dH_2O	2.05 mL	2.00 mL
Analysis buffer	400 μL	400 μL
Analysis reactive solution	50 μL	50 μL
Fermentation sample	-	50 μL

3.8.8. Measurement of the produced ethanol

The ethanol concentration was measured by specific kits (Megazyme, Ireland) based on manufacturer's instructions. The analysis is based upon the following two reactions:



It is NADH which is measured at 340 nm absorbance and, thus, is related to ethanol with stoichiometric relation namely, 2 moles of NADH is produced for each mole of ethanol. It should be mentioned that this assay is linear over the range of 0.25 µg to 12 µg of ethanol in each assay. Cover the cuvettes during assay. The constituents and their amounts have been presented in Table 3.29. The measurement procedure can be described as follows:

1. Preparation of the solutions based on the instructions
2. Numbering the tubes specific for blank and fermentation samples
3. Addition of (defined amount) dH₂O (~25°C) to both blank and fermentation sample tubes
4. Addition of 100 µL fermentation sample to the corresponding tube
5. Addition of 200 µL solution 1 (buffer) to the tubes
6. Addition of 200 µL solution 2 (containing NAD⁺) to the tubes
7. Addition (after swirling) of 50 µL solution 3 (containing Al-DH) to the tubes
8. Mixing (by gentle inversion) and reading the absorbance (A_1) after approximately 2 minutes
9. Addition (after swirling) of 20 µL suspension 4 to the tubes
10. Mixing (by gentle inversion) and reading the absorbance (A_2) after approximately 5 minutes
11. It should be emphasized that reading can be against air (without a cuvette in light path) or against water. Note that with this amounts the sample volume will be 2.57 mL.

Ethanol concentration can be measured by the following equations:

$$\Delta A = A_2 - A_1 \quad (3.5)$$

$$\Delta A_{Ethanol} = \Delta A_{Sample} - \Delta A_{Blank} \quad (3.6)$$

$$C = 0.9397 \times \Delta A_{Ethanol} \quad (3.7)$$

where, “C” is the ethanol concentration (g/L). The amount of ethanol in 100 μL of fermentation sample should be between 0.25 μg to 12 μg so the samples should be diluted sufficiently. The amount of sample can be increased up to 2.00 mL by making sure that the sum of the sample and dH₂O volume is still 2.10 mL. Furthermore, in the calculation of ethanol concentration the new sample volume should be included.

Table 3.29. The amount of the ingredients in ethanol concentration measurement.

Ingredient	Blank tube	Sample tube
dH ₂ O	2.10 mL	2.00 mL
Fermentation sample	-	100 μL
Buffer solution	200 μL	200 μL
Solution 2 (NAD ⁺)	200 μL	200 μL
Solution 3 (Al-DH)	50 μL	50 μL
Mixing and reading the absorbance at 340 nm (A ₁)		
Suspension 4 (ADH)	20 μL	20 μL
Mixing and reading the absorbance at 340 nm (A ₂)		

3.8.9. Measurement of the organic acids concentrations

Concentrations of the organic acids were measured using reversed-phase high performance liquid chromatography utilizing Waters HPLC (Alliance[®] 2695, Milford, MA). This technique in combination with UV-Vis detection is a popular technique for organic acid analysis because it is easy to use and because organic acids are low in molecular weight, and have polar functionalities (Web 15).

The mechanism is based upon passage of the solution through a column filled with oktadecyl carbon chain-bonded silica (ODC) which has a hydrophobic nature. The more hydrophobic the solute the higher will be its interaction with the stationary phase and, in consequence, the higher will be

its retention time. The less hydrophobic solutes will appear earlier in the column exit. In order to elute more hydrophobic solutes from column, a non-polar organic solvent (acetonitrile) is fed to the column in order to perform gradient elution; in this case the eluting strength is varied gradually and, thus, it affects the interactions between solutes and stationary phase.

In current research, dH₂O 100% was used as needle wash, acetonitrile 20% was used as seal wash, acetonitrile 100% was used for column washing as organic solvent, and mobile phase was selected regarding (İleri and Çalık, 2006). The procedure briefly can be explained as follows:

1. Washing the sample vials of HPLC and rinsing with dH₂O
2. Preparation of the mobile phase, organic solvent, needle wash, and seal wash solutions
3. Filtration of the all prepared solutions, except acetonitrile 100%, with 0.45 µm filter
4. Degassing all the prepared solutions in ultrasonic cleaning bath (BRANSON[®] 2510E-MTH, Danbury, CA, USA) for 15-20 minutes
5. Pelleting 3-hour samples of bioreactor at 2000g and +4°C for 10 minutes
6. Filtration of the samples by 0.2 µm filter
7. Diluting the samples with mobile phase in order to obtain the results in acceptable ranges determined by the standard curves of each organic acid
8. Loading 200 µl of each sample into the HPLC vials
9. Running HPLC (Table 3.30).

In the case of each intended organic acid, the standard solution was prepared and loaded to the column in different dilutions in order to obtain calibration curve based on the areas of the peaks displayed by the system (Appendix J). Calibration curves relate the area of the peaks to the concentration of the organic acids. Afterwards, the peak areas obtained for the filtrate samples of the bioreactors are compared with standard curves in order to determine the concentration of the desired organic acid in unknown sample.

Table 3.30. Specifications and conditions of the HPLC analysis.

Column Type	Capital Optimal [®] ODS
Column dimension	4.6mm × 250mm
System	Reversed-phase Chromatography
Mobile phase	0.312% (w/v) NaH ₂ PO ₄ ; 0.062% (v/v) H ₃ PO ₄ 85%
Flow rate of mobile phase	0.8 mL min ⁻¹
Column temperature	30°C
Detector type	Waters 2487 Dual λ Absorbance Detector (UV/Vis)
Detector wavelength	210 nm
Detector temperature	30°C
Injection volume	5 μL
Analysis period	15 min
Delay time	5 min

3.8.10. Measurement of the protease activity

The proteolytic activity of the acidic proteases in the fermentation broth was measured by hydrolysis of casein. The procedure is as follows:

1. Preparation of 0.05 M sodium acetate buffer (Appendix B)
2. Taking required amount of the 0.05 M sodium acetate buffer in sterile cabin for samples dilutions and preparation of casein solution
3. Preparation of 0.5% (w/v) of casein solution in sodium acetate buffer, and 10% (w/v) trichloroacetic acid (TCA) solution in dH₂O
4. Centrifugation of the taken sample from bioreactor at 13500g for 10 min. in order to harvest the culture broth
5. Diluting the samples by mixing 300 μL sample with 700 μL buffer
6. Mixing 2 mL 0.5% (w/v) casein solution and 1 mL diluted sample.
Blank sample includes only 1 mL buffer

7. Incubation of the samples at 200 rpm at 30 °C for 20 min.
8. Putting samples on ice
9. Addition of 2 mL of 10% (w/v) TCA solution in order to terminate the reaction
10. Keeping on ice for 20 min.
11. Centrifugation at 10500 rpm for 10 min. at +4 °C
12. Keeping at RT for 5 min.
13. Reading the absorbance (of the supernatant) at 275 nm with quartz cuvettes.

One unit protease activity is defined as the activity that liberates 4 nmol of tyrosine per minute. The utilized equation in order to convert absorbance to protease activity (U/mL) is as follows (Çalık, 1998):

$$A_{protease} = Absorbance \times 15.625 \times Dilution\ factor \quad (3.8)$$

3.8.11. Metabolic flux analysis

In order to determine the fluxes of the intracellular reactions, metabolic flux analysis was conducted. The central metabolic reaction network of *P. pastoris* utilized previously (Çalık et al., 2011) was recruited with minor modifications; the reactions (Appendix U) totally comprise 102 metabolites and 146 reactions. GAMS 2.25 (General Algebraic Modeling System, GAMS Development Corp., Washington DC) was the optimization program used for solving the mass-flux balance equation (2.98). By minimizing the objective function (Z), optimum distributions of the fluxes were determined. The variables of the model were the fluxes of the metabolic reactions expressed in (mmol/gDCW.h). Furthermore, the flux towards biomass was represented with specific growth rate (μ).

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1. Development of the base plasmid

In order to obtain the primary plasmid construct that desired modifications can be performed on it, pGAPZ α A and pPICZ α A::*hGH* were used as two parent plasmids. *hGH* gene in pPICZ α A::*hGH* plasmid is the mature *hGH* gene, 573 nucleotides, with appendices in both ends in order to make the insertion of the gene into the desired plasmid feasible, facilitate the purification of the expressed protein, and, furthermore, provide an authentic N-terminus. The complete nucleotide sequence of the 618-bp *hGH* gene has been represented in Figure 4.1.

```
5' ...G...AATTC CACCATCACCATCACCAT ATTGAAGGGAGA
ttcccaactataccaactatctcgtctattcgataaacgctatgcttcgctgctcatcgtcttcatc
agctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaaacagaagtattc
attcctgcagaacccccagacctccctctgtttctcagagctctattccgacacctccaacagg
gaggaacacacaacagaaatccaacctagagctgctccgcctctccctgctgctcatccagtctg
ggctggagcccgtgcagttcctcaggagtgctctcgccaacagcctgggtgtacgggcctctga
cagcaacgtctatgaacctcctaaaggacctagaggaaggcatccaaacgctgatggggaggctg
gaagatggcagccccggactgggcagatcttcaagcagacctacagcaagtgcacacaaact
cacacaacgatgacgcactactcaagaactacgggctgctctactgcttcaggaaggacatgga
caaggtcgagacattcctgctcagctgctgagggcagctgtgggttc
TAG T...CTAGA...3'
```

Figure 4.1. Nucleotide sequence of the *hGH* gene utilized in current research. Small letters are mature *hGH* sequence. Black italic letters are restriction sites corresponding to *EcoRI* (at 5' end) and *XbaI* (at 3' end). Blue letters belong to 6-His tag. Red letters represent the factor Xa protease recognition sequence. Green letters are stop codon improvised at the end of the *hGH* gene.

EcoRI and *XbaI* restriction sites in 5' and 3' ends, respectively, provide suitable restriction digestion sites for subsequent steps. 6-His tag makes the final affinity purification possible and factor Xa protease recognition sequence leads to the cleavage of the N-terminus motif and an authentic N-terminus for expressed rhGH.

4.1.1. Isolation and validation of the parent plasmids

After inoculation of the recombinant *E. coli* (*r-E. coli*) stocks of the two parent plasmids from -80°C to suitable selective LSLB (with Zeocin) agar plates and subsequent inoculation of the selected colonies to selective LB broth, plasmid isolation (pGAPZαA and pPICZαA::hGH) was performed by alkaline lysis method. The isolated plasmids are observed in Figure 4.2. The location of two circular plasmids were checked and confirmed.

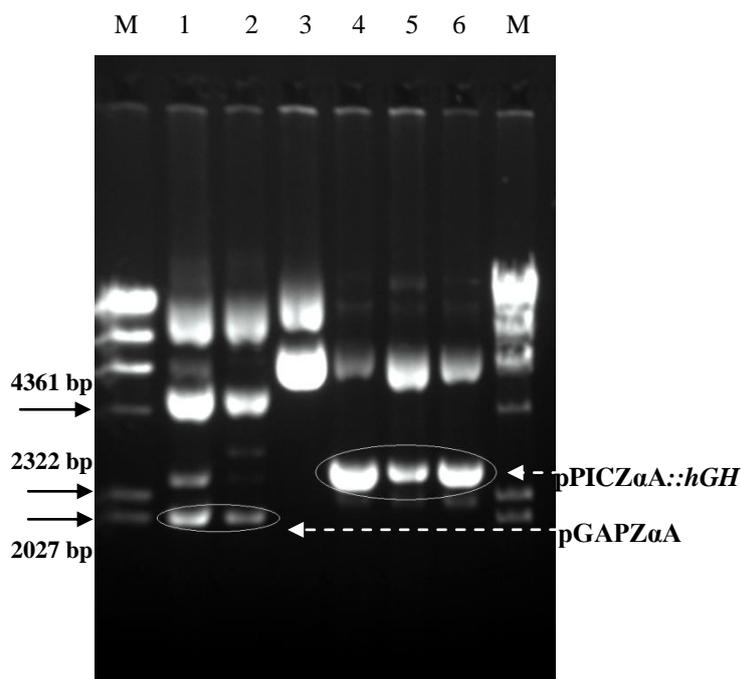


Figure 4.2. Agarose gel electrophoresis result of the isolated plasmids of two different colonies of pGAPZαA and four different colonies of pPICZαA::hGH . M: DNA marker λDNA/HindIII, 1: pGAPZαA colony 1, 2: pGAPZαA colony 2, 3: pPICZαA::hGH colony 1, 4: pPICZαA::hGH colony 2, 5: pPICZαA::hGH colony 3, 6: pPICZαA::hGH colony 4.

In order to further confirm the obtained plasmids, a single digestion with *EcoRI* enzyme was performed according to Table 3.10. The results can be clearly seen in the Figure 4.3. By combining the results shown in Figure 4.2 and Figure 4.3 the presence of the desired parent plasmids in isolated colonies were confirmed.

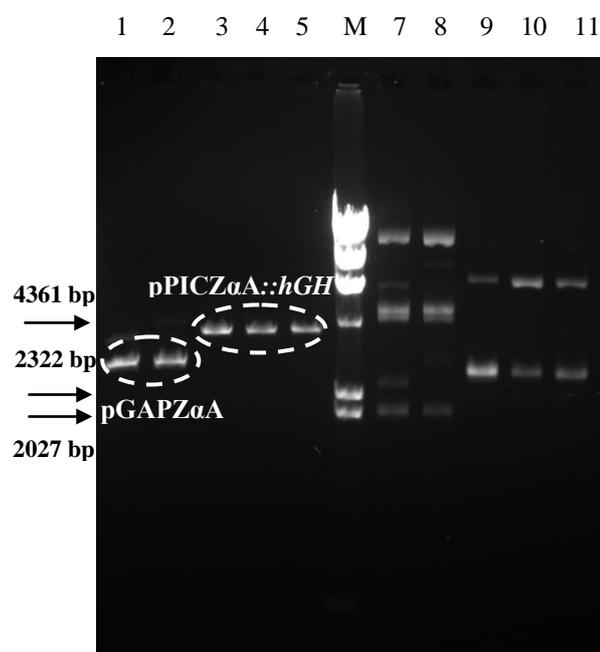


Figure 4.3. Further verification of pGAPZαA and pPICZαA::hGH plasmids after single digestion of putative corresponding isolated plasmids with *EcoRI* restriction enzyme. M: DNA marker λDNA/HindIII, 1, 2: single-digested putative pGAPZαA plasmids, 3, 4, 5: single-digested putative pPICZαA::hGH, 7, 8: circular pGAPZαA plasmids, 9, 10, 11: circular pPICZαA::hGH.

4.1.2. Double-digestion of the parent plasmids with *EcoRI* and *XbaI*

The next step was the double-digestion of the confirmed isolated plasmids of pGAPZ α A and pPICZ α A::*hGH*, parent plasmids, with *EcoRI* and *XbaI* restriction enzymes in order to release *hGH* gene from pPICZ α A::*hGH* and its subsequent insertion in to the pGAPZ α A. Double-digestion of the parent plasmids was performed by *EcoRI* and *XbaI* restriction enzymes as described in section 3.6.7 in Table 3.11. The result of the double-digestion can be found in Figure 4.4. The specified segments were cut and subjected to the gel extraction by gel extraction kit as described before. After gel elution, the confirmation of the purified segments was conducted by the aid of gel electrophoresis which can be seen in Figure 4.5.

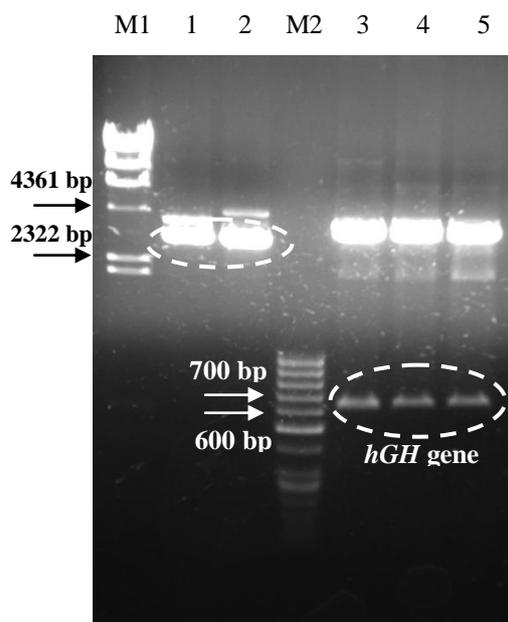


Figure 4.4. Double-digested (with *EcoRI* and *XbaI*) pGAPZ α A and pPICZ α A::*hGH* plasmids. M1: DNA marker λ DNA/HindIII, 1, 2: pGAPZ α A plasmid subjected to double-digestion, M2: GeneRuler™ 50bp DNA ladder, 3, 4, 5: pPICZ α A::*hGH* plasmid subjected to double-digestion.

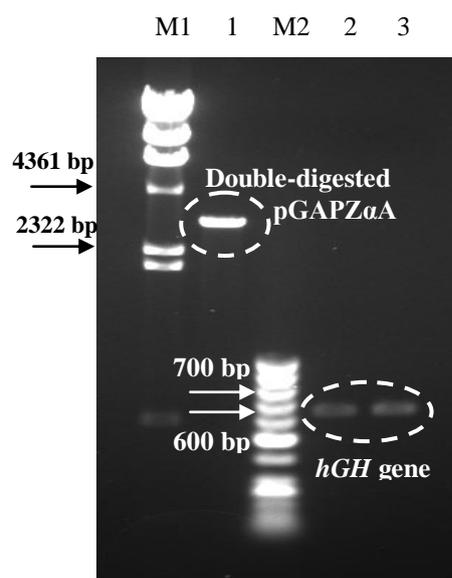


Figure 4.5. Control of the purified double-digested pGAPZ α A and pPICZ α A::*hGH*. M1: DNA marker λ DNA/HindIII, 1: purified double-digested pGAPZ α A, M2: GeneRuler™ 50bp DNA ladder, 2, 3: purified *hGH* gene as the result of the double-digested pPICZ α A::*hGH*.

4.1.3. Ligation of the double-digested fragments

After obtaining the desired double-digested segments with compatible ends, i.e., corresponding digested ends with same restriction enzymes, *Eco*RI and *Xba*I, the ligation of these segments was performed by T4 DNA ligase enzyme as described before in section 3.6.9 and Table 3.14. The molar ratio of the two isolated segments (equation 3.1), *hGH* gene as insert and double-digested pGAPZ α A as backbone, was set to be between 3:1 to 5:1. The ligation between two purified DNA fragments resulted in the base plasmid (BP) in current study (Figure 4.6). BP possesses 3696 nucleotides and, therefore, it should be located between pPICZ α A::*hGH* (4142 bp) and pGAPZ α A (3147 bp) in gel electrophoresis in circular form.

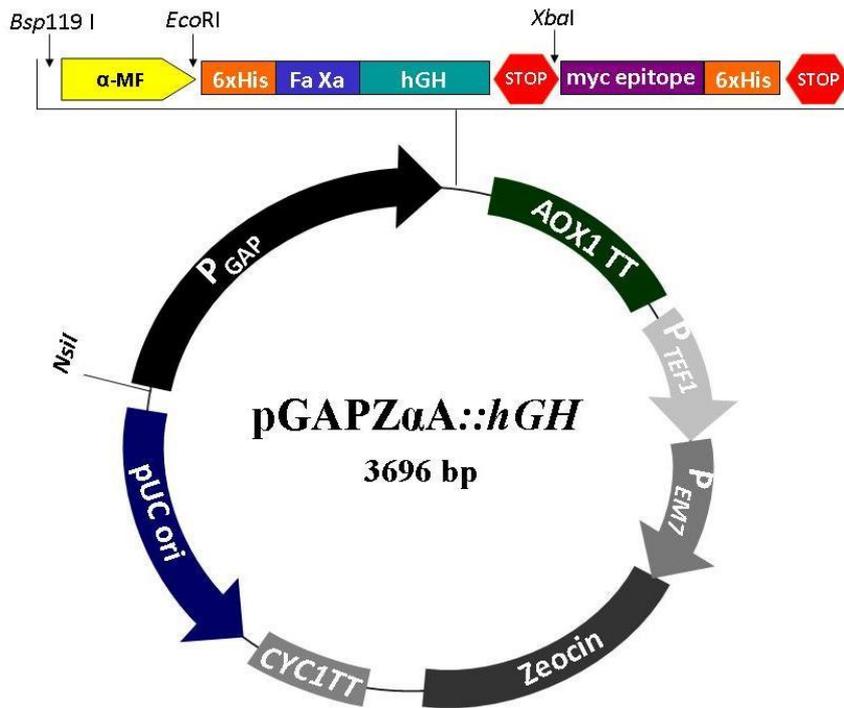


Figure 4.6. Schematic representation of the base plasmid (BP).

4.1.4. Transformation of the *E. coli* DH5 α cells with base plasmid

The ligation reaction result was used in order to transform *E. coli* DH5 α cells by CaCl₂ method as explained before (section 3.6.2). Appeared colonies after transformation, should be checked in order to get rid of false positive colonies. A few colonies were selected, among thousands, as putative transformants of BP and were sub-cultured in selective LSLB plates with Zeocin™ for further confirmation and analyses. Plasmid isolation (section 3.6.3) from the selected colonies, 6 colonies at first attempt, was performed and by agarose gel electrophoresis their length was verified (Figure 4.7).

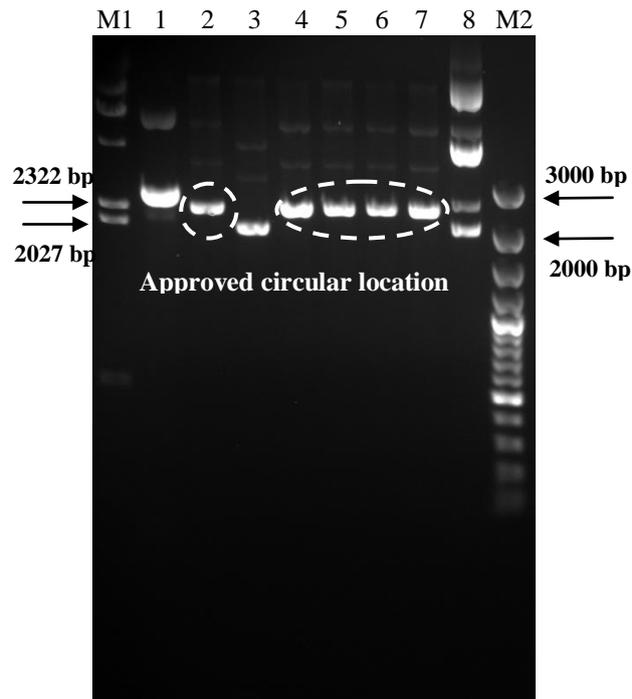


Figure 4.7. Isolated plasmids from 6 colonies related with putative transformants of BP. M1: DNA marker λ DNA/HindIII, 1: circular pPICZ α A::hGH plasmid, 2-7: 6 selected colonies 1, 2, 3, 4, 5, 6 respectively, 8: circular pGAPZ α A plasmid, M2: GeneRuler™ 100bp plus DNA ladder.

Five out of six selected colonies (except colony 2 as false positive result) would seem to be possible transformants. For further confirmation of precise ligation and transformation, these five putative transformant colonies (1, 3, 4, 5, 6) were double-digested with *Eco*RI and *Xba*I enzymes and then, the digestion result was run on the gel (Figure 4.8). Along with the selected colonies, double digested pPICZ α A::hGH was also used as an indicator of the release of hGH gene from the selected putative transformants. According to the gel electrophoresis result, all of the five colonies could be the desired transformant.

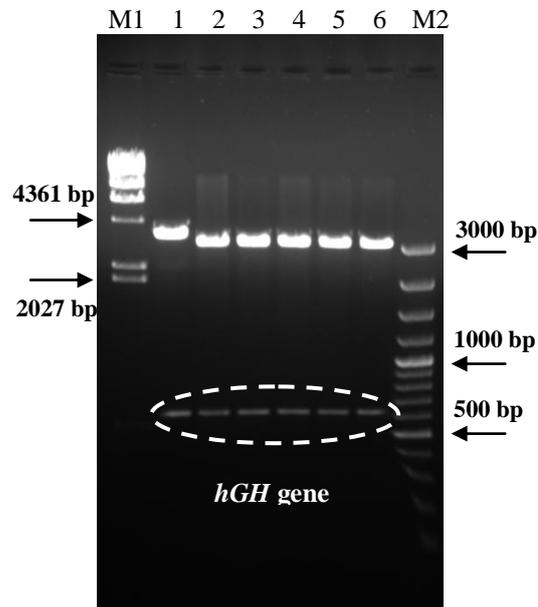


Figure 4.8. Double-digestion of five putative transformants' isolated plasmids by *EcoRI* and *XbaI*. M1: DNA marker λ DNA/HindIII, 1: double digested pPICZ α A::*hGH* with *EcoRI* and *XbaI*, 2-6: five putative transformants, M2: GeneRuler™ 100bp plus DNA ladder.

As final level of the confirmation, for two of the five approved colonies a PCR was conducted using *GAP* forward and *AOX* reverse primers (Appendix D). The PCR reaction mixture and PCR thermal profile can be found in Table 4.1 and Table 4.2, respectively. Based on the location of two primers which can be found in Appendix I, the region resides between two primers in original parent plasmid (pGAPZ α A) is approximately 540 bp. By insertion of the *hGH* gene and excluding the region between *EcoRI* and *XbaI* cleavage sites, the PCR result of the selected putative transformants (colony 3 and 6) should be approximately 1080 bp. Furthermore, this PCR can reveal the optimum conditions for sequencing the plasmids. The PCR results can be observed in Figure 3.9 which confirmed proper ligation and transformation. On the other hand, the gel result implies to the suitable conditions which can be used for sequencing. The isolated plasmids were sent to Middle East Technical University central laboratory for sequence analyses; the results can

be found in Appendix K. Both colonies were confirmed based on the obtained sequences and one of them was used for subsequent *Pichia* transfection.

Table 4.1. PCR mixture for verification of BP with *GAP* forward and *AOX* reverse primers.

Component	Amount
10X amplification buffer with KCl	5 μ L
2 mM solution of dNTPs	5 μ L
5 μ M <i>GAP</i> forward primer	2 μ L
5 μ M <i>AOX</i> reverse primer	2 μ L
25 mM MgCl ₂ / MgSO ₄	3 μ L / 4 μ L
<i>Taq</i> DNA polymerase	0.3 μ L
dH ₂ O	Complete to 50 μ L
Template DNA (Genomic DNA or plasmid DNA)	2 μ L
Total volume	50 μL

Table 4.2. PCR thermal profile for verification of the BP with *GAP* forward and *AOX* reverse primers.

Number of cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	50 °C	1 min.
29 cycles	72 °C	1 min.
	94 °C	1 min.
	55 °C	1 min.
Final extension	72 °C	1 min.
		10 min.

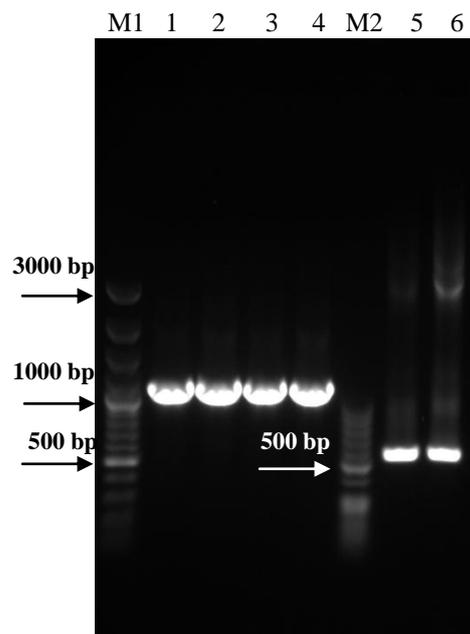


Figure 4.9. PCR results with isolated plasmids related with two putative transformant colonies of BP. M1: GeneRuler™ 100bp plus DNA ladder, 1-4: two different PCR results of two selected putative transformant colonies (1.5mM and 2 mM MgCl₂ concentration for each colony), M2: GeneRuler™ 50bp DNA ladder, 5,6: pGAPZ α A plasmid as positive control (1.5mM and 2 mM MgCl₂ concentration were tested).

4.2. In-silico determination of promising endogenous SPs of *P. pastoris*

Using available secretome data from previously conducted research (Huang et al., 2011), proteins found in extracellular medium of a *P. pastoris* culture, after finding their corresponding amino acid sequences, were analyzed by four computer programs including: SignalP4.1, WolfPsort, Phobius, and ProP1.0 to predict the SPs (both “pre” and “pro” sequence) along with the H-region of the SPs. SignalP program attributes a D-score to each SP which reflects its signal peptide-ness and Phobius specifies H-region of the SP. The determined endogenous SPs can be found in Table 4.3 as SP1-SP41. Furthermore, exogenous SPs utilized or have the potential to be used in *Pichia* system were also analyzed which have been represented as SP42-SP56 in Table 4.3. Deciding on the selected SPs was, primarily,

accomplished based on D-score reported by SignalP program. Then, SPs were regarded as potential candidates where all the predictions by programs led to a similar SP length in order to be confident about the cleavage point of the SP. Since *S. cerevisiae* α -MF is the widely used secretion SP in *P. pastoris* expression system, it was regarded as a scale for selection of the promising endogenous SPs of *P. pastoris*. It is of supreme importance to mention that as the SP locates in the N-terminus of the desired protein in expression cassette, its D-score should be reconsidered in combination with the intended protein.

As primary choices, 10 endogenous SPs were selected regarding their D-score, in their original protein, obtained higher than α -MF D-score and also consistency of the programs predictions (Table 4.4). However, the SPs were intended to be used in combination with the available *hGH* gene with its specific 12-amino acid N-terminus motif. Therefore, the D-scores were recalculated for combination of the selected SPs and the desired *hGH* gene. In addition D-scores for combination of SPs and mature hGH and hGH with two extra amino acids resulted from *EcoRI* restriction site were also recalculated in order to show the effect of N-terminus residues in D-score and SP cleavage point. As it is obvious, addition of 12 amino acids to the N-terminus had a profound effect on D-score. Regarding the calculated D-scores, SP23, SP24, SP26, SP34 were selected as the promising endogenous SPs. In addition, SP13, which was utilized successfully in *P. pastoris* (Liang et al., 2013), was also selected in order to check its capability in secretion of another protein, although its D-score is lower than α -MF. In total, five endogenous SPs were selected whom their secretion efficiency will be compared with α -MF. After selection of the endogenous secretion SPs, their nucleotide sequences should be determined. The amino acid sequences of the corresponding proteins and, subsequently, their related nucleotide sequences were obtained from NCBI (Appendix L). Consequently, the nucleotide sequences of the selected endogenous SPs as promising candidates were determined (Table 4.5).

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/localization of corresponding protein ³	Predicted SP (pre-sequence...prosequence) (# amino acids of pre-sequence)	Pro-sequence		SignalP D.score ^b	Wolf-Psort ^b	Phobius ^b	Reference
				No.	N-glyc.				
10	254566893	Endo-beta-1,3-glucanase, major protein of the cell wall, involved in cell wall maintenance (Huang et al., 2011)	MFNLKTLAAVAISIQVSA (20)	-	-	0.651	+	+	NYC
11	254570078	Protein of the SUN family (Sim1p, Uth1p, Nca3p, Sur4p) that may participate in DNA replication (Huang et al., 2011)	MKISALTACAVTLAQLAIA.....APAPKPEDCTIV QKRHQHKK (19)	21	-	0.694	+	+	NYC
12	254568684	Hypothetical protein (Huang et al., 2011)	MSYLKISALLSVLSVALA (18)	-	-	0.760	+	+	NYC
13	254567645	Cell wall protein with similarity to glucanases (Huang et al., 2011)	MLSTILNFTLLLEQASLQ (20)	-	-	0.925	+	+	(Liang et al., 2013)
14	254570357	Protein of unknown function, has similarity to Prip1 and Prip2 and to the plant PR-1 class of pathogen (Huang et al., 2011)	MKLSLNLLAIAAASAVVSA.....APVAPAEAAAH LHKK (20)	16	-	0.822	+	26 aa	NYC
15	254564915	Beta-1,3-glucanohydrolase, required for cell wall assembly (Huang et al., 2011)	MPKSLCMLIGSCLLSVLA (19)	-	-	0.918	+	+	NYC
16	254573228	O-glycosylated protein required for cell wall stability (Huang et al., 2011)	MKLAALSTIALTILPVALA (19)	-	-	0.861	+	+	NYC
17	254565329	Daughter cell-specific secreted protein with similarity to glucanases, endo-1,3-beta-glucanase (Huang et al., 2011)	MSFSSNVQQLFLLLVLLLVVSG (23)	-	-	0.893	+	+	(Liang et al., 2013)
18	254573224	Putative protein of unknown function (Huang et al., 2011)	MQLOYLAVLCAALLNVQS...KNVYDFSRFGDAK ISPDDIDLESREKRR (18)	28	-	0.699	+	+	NYC

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/localization of corresponding protein ^a	Predicted SP (pre-sequence, pro-sequence) (# amino acids of pre-sequence)	Pro-sequence		SignalP D scores ^b	Wolf- Psort ^b	Phobius ^b	Reference
				No.	N- glyc.				
19	254567547	Hypothetical protein (Huang et al., 2011)	<u>MKHSLLWNLEFP</u> SLG (19)	-	-	0.875	+	+	NYC
20	254573944	Hypothetical protein (Huang et al., 2011)	MS <u>ILLAVLLS</u> QNSALA (19)	-	-	0.910	+	+	NYC
21	254565023	Mucin family member (Huang et al., 2011)	MINLS <u>FLILVILL</u> SPALA...LPKNVLEEQAK DDLAKR (20)	18	-	0.868	+	+	NYC
22	254568260	Hypothetical protein (Huang et al., 2011)	MFSLAVG <u>ALLLTQ</u> AFC (16)	-	-	0.644	+	+	NYC
23	254565591	Protein disulfide isomerase, multifunctional protein resident in the endoplasmic reticulum lumen (Huang et al., 2011)	MK <u>LSALLL</u> FLAFA (16)	-	-	0.932	+	+	NYC
24	254572565	Hypothetical protein (Huang et al., 2011)	MKVSTTK <u>LAIVLLVRL</u> VCA (20)	-	-	0.897	+	+	NYC
25	254569896	Cell wall protein that contains a putative GPI- attachment site (Huang et al., 2011)	MQFGK <u>VAISALAVT</u> ALG (19)	-	-	0.761	+	26 aa	NYC
26	254572672	Hypothetical protein (unknown function) (Huang et al., 2011)	MW <u>SFSGLLIF</u> PLVLC (18)	-	-	0.883	+	+	NYC
27	254573438	Ferric reductase and cupric reductase (Huang et al., 2011)	MRNHLND <u>LVFLLLIV</u> AQA (21)	-	-	0.885	+	23 aa	NYC

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/localization of corresponding protein ^a	Predicted SP (pre-sequence...pro-sequence) (# amino acids of pre-sequence)	Pro-sequence		SignalP D.score ^b	Wolf- Psort ^b	Phobius ^b	Reference
				No.	N- glyc.				
28	254567898	Hypothetical protein (Huang et al., 2011)	<u>MFLKSL</u> <u>LFASIL</u> <u>ILCKA</u> (18)	-	-	0.863	+	+	NYC
29	254569230	Ferro-O ₂ -oxidoreductase (Huang et al., 2011)	<u>MFVFPYLLAVLVASTCYTA</u> (20)	-	-	0.735	+	+	NYC
30	254569662	Hypothetical protein (Huang et al., 2011)	<u>MVSLRSFTSSILAAGLTRAHG</u> (22)	-	-	0.540	20 aa	+	NYC
31	254567750	One of three repressible acid phosphatases, a glycoprotein that is transported to the cell surface (Huang et al., 2011)	<u>MFSPILSEIILALAILQSVFA</u> (22)	-	-	0.865	+	+	(Murasugi et al., 2001; Yoshimasa et al., 2002)
32	254570525	Hypothetical protein (Huang et al., 2011)	<u>MINHLVLTALSIALA</u> (16)	-	-	0.565	+	+	NYC
33	254567331	Phosphatidylycerol / phosphatidyinositol transfer protein (Huang et al., 2011)	<u>MLALVRSITLILLALTAASA</u> (19)	-	-	0.885	20 aa	+	NYC
34	254565679	Cell wall protein that functions in the transfer of chitin to beta(1-6) glucan (Huang et al., 2011)	<u>MRPYLSLILLASSVLA</u> (17)	-	-	0.932	+	+	NYC
35	254567499	Protein ROT1 (Huang et al., 2011)	<u>MVLQNFPLFAYLFFNQRAALA</u> (24)	-	-	0.546	45 aa	+	NYC
36	254570227	Hypothetical protein (Huang et al., 2011)	<u>MKFPVPLFLLOLFFFIATQC</u> (21)	-	-	0.852	+	+	NYC
37	254573778	Potative chitin transglycosidase, cell wall protein (Huang et al., 2011)	<u>MVSLTRLLITGIALALQVNA</u> (20)	-	-	0.708	+	+	NYC

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/localization of corresponding protein ^a	Predicted SP (pre-sequence...prosequence) (# amino acids of pre-sequence)	Pro-sequence		SignalP D.score ^b	Wolf-Port ^b	Phobius ^b	Reference
				No.	N-glyc.				
38	254572447	Vacuolar aspartyl protease (proteinase A) (Huang et al., 2011)	MFDGTTNIAIGLLSLGCAEA (24)	-	-	0.477	+	+	NYC
39	254572379	Putative integral membrane protein (Huang et al., 2011)	MVLVGLLRKLAPELVLLAGTVLILLVFLVLSGG	-	-	no SP ^b	31 aa	29 aa	NYC
40	254564917	Beta-1,3-glucanoyltransferase, required for cell wall assembly (Huang et al., 2011)	MLSLIALTLGLSCA (16)	-	-	0.818	+	+	NYC
41	254566331	Hypothetical protein (unknown function, has similarity to Prt1P and Prt3p) (Huang et al., 2011)	MRLLHSLSHLSYLTKANA (20)	-	-	0.822	+	+	NYC

42	<i>S. cerevisiae</i> (from source other than <i>P. pastoris</i>)	Alpha pheromone precursor, the active factor is excreted into the culture medium by haploid cells of the alpha mating type and acts on cells of the opposite mating type	MRPESIFTAVLEFASSALA...APVNTTIDEIAQIP AEAIVIGLDLEGFDAVLFPNSNTNGLLFINITI ASIAAKEGVSLDKR...EAEA (19)	66 ^j	3	0.885	+	+	(Batra et al., 2010)
43	<i>S. cerevisiae</i>	PHO5, repressible acid phosphatase (1 of 3) that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway-derived cell surface glycoprotein	MFKSVVYSILAAALANA (17) (Armas et al., 1983)	-	-	0.494	+	+	NYC
44	<i>Saccharomyces</i> spp.	SUC2 (invertase), hydrolysis of sucrose	MLLOAFLFLAGFAAKSA (19)	-	-	0.834	+	+	(Paifer et al., 1994; Kuwae et al., 2005)

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/localization of corresponding protein ^a	Predicted SP (pre-sequence...prosequence) (# amino acids of pre-sequences)	Pre-sequences		SignalP D score ^b	Wolf-Psort ^b	Phobius ^b	Reference
				No.	N-glyc.				
45	<i>Phaeoascus vulgaris</i>	PHA-E (lectin found in plants), protein agglutinates <i>Erythrocytes</i> ~ erythroagglutinating phytohemagglutinin	MASSNLLSLALFLVLLTHANS (21)	-	-	0.889	+	+	(Raemackers et al., 1999)
46	<i>Kluyveromyces fragilis</i>	Killer toxin , hypothetical protein	MNMFVFLLSFVQG.....LEHITRRGSLVKR (16)	13	-	0.843	+	+	(Kato et al., 2001)
47	<i>Pichia acacias</i>	Killer toxin , product of linear DNA plasmid pPacl.2, stimulates <i>Kluyveromyces fragilis</i> killer toxin	MLIVLLFLATLANS...LDCSGDVFFGYTRGDKTDVHKSQALTAVKVKKR (15)	33	-	0.882	+	+	(Crawford et al., 2003)
48	<i>S. cerevisiae</i>	Killer toxin , K28 preprototoxin (M28 virus)	MESVSSLNFSTFMVNYKSLVALLSVNLSKYARG.....MPTSERQQGLEER (36)	13 ^k	-	no SP	31 aa	no SP	(Eiden-Plach et al., 2004)
49	<i>S. carlsbergensis</i>	MEL1 (melibiase or α -galactosidase), pre- α galactosidase (melibiase) – probably extracellular	MEAFYFLTACISLKGVFG (18)	-	-	0.702	+	+	NYC
50	<i>S. cerevisiae</i>	BGL2 (endo-beta-1,3-galactanase), major protein of the cell wall, involved in cell wall maintenance	MRFTTLATATAALFTLASQVSA (23)	-	-	0.703	+	+	NYC
51	<i>Kluyveromyces fragilis</i>	INU (inulinase), hydrolyzing the beta-D-2,1-fructan fructoside of inulin - extracellular	MKEAYSLLELAGVSA.....SVNYKR (16)	7	-	0.685	+	+	NYC
52	<i>Trichoderma reesei</i>	Hydrophobin I , contributes to surface hydrophobicity. Secreted → cell wall	MKEFAIAALEFAAAVA.....QPLEDR (16)	7	-	0.805	+	+	(Kottmeier et al., 2011)

Table 4.3. List of the endogenous and exogenous SPs used, or have the potential to be used, in *P. pastoris* expression system (Continued).

SP No.	GI (<i>P. pastoris</i> ORF)	Function/Localization of corresponding protein ^a	Predicted SP (pre-sequence...pro-sequence) (# amino acids of pre-sequence)	Pro-sequence		SignalP D score ^b	Wolf-Psort ^b	Phobius ^b	Reference
				No.	N-glyc.				
53	<i>Trichoderma reesei</i>	Hydrophobin II , responsible for spore hydrophobicity and protection spore wall, secreted (cell wall)	MOFFAVALEATSALA (15)	-	-	0.854	+	+	(Kottmeier et al., 2011)
54	Human	Serum albumin , main function is the regulation of the colloidal osmotic pressure of blood. Major zinc transporter in plasma	MKVVVFISLLEESSAYS.....RGVFR (18)	6 P	-	0.848	+	+	(Xiong et al., 2008)
55	Chicken (<i>Gallus gallus</i>)	Lysozymes have primarily a bacteriolytic function; those in tissues and body fluids are associated with the monocyte-macrophage system and enhance the activity of immunogens.	MRSLLILVLCFPLAALG (18)	-	-	0.897	+	+	(Oka et al., 1999)
56	<i>Bos Taurus</i> (Bovine)	Beta casein	MKVLILACLVALALA (15)	-	-	0.900	+	+	(Zuyong et al., 2012)

Table 4.4. List of the potentially suitable endogenous SPs based on their D-scores in different combination. "AA" refers to amino acid.

SP	D-score in original Endogenous protein	Cleavage site in original Endogenous protein	D-score with mature hGH	Cleavage site + hGH	D-score with extra 2 AA + hGH	D-score with extra 12 AA + hGH	Comments
23	0.932	AA 16-17	0.940	AA 16-17	0.930	0.883	
34	0.932	AA 17-18	0.927	AA 17-18	0.914	0.870	
13	0.925	AA 20-21	0.891	AA 20-21	0.877	0.836	Checked (Liang et al., 2013) Cleavage point is changed by inclusion of 12 AA (Phobius)
15	0.918	AA 19-20	0.887	AA 19-20	0.861	0.793	
20	0.910	AA 19-20	0.878	AA 19-20	0.861	0.823	
9	0.904	AA 23-24	0.895	AA 23-24	0.881	0.851	
24	0.897	AA 20-21	0.910	AA 20-21	0.893	0.841	
17	0.893	AA 23-24	0.864	AA 23-24	0.847	0.811	Checked (Liang et al., 2013)
5	0.885	AA 25-26	0.877	AA 25-26	0.865	0.830	
26	0.883	AA18-19	0.909	AA 18-19	0.894	0.831	
α -MF	0.885	AA 19-20	0.885	AA 19-20	0.885	0.885	

Table 4.5. Amino acid sequence and nucleotide sequence of the selected endogenous SPs along with the chromosomal location of their corresponding gene.

SP	Chromosomal Location	Nucleotide Sequence	Amino acid Sequence (Length)
SP13	PAS_chr2-1_0052	ATGCTATCAACTA TCTTAAATATCTT TATCCTGTTGCTC TTCATACAGGCAT CCCTACAG	MLSTILNIFILLLFIQASLQ (20)
SP23	PAS_chr1-1_0160	ATGAAAATATTA AGTGCATTGCTTC TTCTTTTACGTT GCCTTTGCT	MKILSALLLFTLAF (16)
SP24	PAS_c131_0001	ATGAAAGTTTCTA CGACCAAATTTCT GGCTGTGTTCTTA TTAGTTAGACTCG TTTGGCT	MKVSTTKFLAVFLLVRLVCA (20)
SP26	PAS_chr4_0040	ATGTGGTCGCTGT TCATATCTGGACT ATTAATCTTCTAT CCTTGGTCCTTG GA	MWSLFISGLLIFYPLVLG (18)
SP34	PAS_chr1-1_0293	ATGAGACCAGTG CTTTCGTTATTAC TCTTGCTGGCTTC TTCGGTACTCGCT	MRPVLSLLLLLASSVLA (17)

4.3. Candidate promoters

Probable promoters that can be helpful in oxygen limitation conditions were selected and, then, their corresponding nucleotide sequences were determined.

4.3.1. Deciding on the promising oxygen limitation-induced promoters

Regarding problems arose in utilization of the methanol as inducer in *AOXI*-based *Pichia* expression systems, the methanol-free systems are preferred in recent years where P_{GAP} is considered as the pioneer in this area. Furthermore, increase of the transcription level of glycolytic genes, including *GAP*, under oxygen limitation conditions has been reported (Baumann et al., 2010) which casts light on the fact that P_{GAP} can be regarded as a promising choice for r-protein expression under oxygen limitation conditions. Considering the upregulation of *GAP* gene in oxygen limitation situations, in current research selection of the new promoters more active in oxygen limitation conditions were based on comparison to P_{GAP} . These promoters were utilized in subsequent r-protein production in oxygen limitation conditions.

Based on the realized positive correlation between transcription and expression level in core metabolic pathway genes of *P. pastoris* upon conforming to oxygen limitation conditions, by finding the proteins whom their expression level are increased in oxygen limitation conditions more than *GAP* protein, candidate promoters were determined. The list of the proteins with increased expression in oxygen limitation conditions has been represented in Table 4.6 where the third and fifth columns are representatives of relative increase of the proteins expression to *GAP* protein based on the conducted study (Baumann et al., 2010).

Table 4.6. List of the proteins with increased expression in shift from normal to oxygen limitation conditions in both wild type and recombinant *P. pastoris* X-33 strain. Two average ratio columns refer to the (Amount in oxygen limitation condition / Amount in normal oxygenation condition). “Av. ratio” refers to average ratio.

Protein	Recombinant Strain		Control Strain	
	Av. ratio	Av. ratio/ Av. ratio GAP	Av. ratio	Av. ratio/ Av. ratio GAP
Pyruvate decarboxylase	2.16	1.43	-	-
	1.99	1.32	1.25	0.61
	3.07	2.03	6.31	3.06
Pyrimidine precursor biosynthesis enzyme thi3	3.04	2.01	3.11	1.51
S-adenosylmethionine synthetase	1.2	0.79	2.88	1.4
Elongation factor 2	1.28	0.85	2.11	1.02
Myo-inositol-1-phosphate synthase	1.39	0.92	1.79	0.87
	2.07	1.37	2.27	1.10
Zinc binding oxidoreductase	1.87	1.24	2.47	1.2
Pyridoxine (pyridoxamine) phosphate oxidase	2.01	1.33	1.32	0.64
NADPH-dependent alpha-keto amide reductase	1.63	1.08	2.28	1.11
Pyruvate kinase	1.91	1.26	3.14	1.52
	2.19	1.45	2.98	1.45
Phosphoglucose isomerase	1.88	1.24	2.34	1.13
Enolase	1.69	1.12	2.86	1.39
	2.45	1.62	1.57	0.76
	2.4	1.59	2.09	1.01
GAP	1.51	1.0	2.06	1.0
Fructose-1,6-biphosphate aldolase	1.72	1.14	3.72	1.80
Phophoglycerate mutase	1.96	1.3	2.12	1.03
	3.38	2.24	1.98	0.96
Heat shock protein SSA4	1.36	0.9	3.06	1.48
Thioredoxin peroxidase	1.84	1.22	1.73	0.84
Hypothetical protein	1.35	0.9	2.2	1.07

At first round of the screening, the proteins that their relative average ratio to GAP in recombinant strain was smaller than unity were removed. Then, among the remaining candidates, myo-inositol-1-phosphate synthase was removed since the two isoforms of the protein showed non-uniform behavior ($0.92 < 1$ but $1.37 > 1$). Afterwards, the proteins that had lower average ratio over GAP protein in expressing strain were omitted except those that have isoforms, namely all of the isoforms were considered as a group as they are under the control of a single promoter. The remaining candidates have been presented in Table 4.7.

Table 4.7. List of the remained proteins with increased expression in shift from normal to oxygen limitation conditions in both wild type and recombinant *P. pastoris* X-33 strain after first round of the screening.

Protein	Recombinant Strain		Control Strain	
	<i>Av. ratio</i>	<i>Av. ratio/ Av. ratio GAP</i>	<i>Av. ratio</i>	<i>Av. ratio/ Av. ratio GAP</i>
Pyruvate decarboxylase	2.16	1.43	-	-
	1.99	1.32	1.25	0.61
	3.07	2.03	6.31	3.06
Pyrimidine precursor biosynthesis enzyme thi3	3.04	2.01	3.11	1.51
Pyridoxine (pyridoxamine) phosphate oxidase	2.01	1.33	1.32	0.64
Pyruvate kinase	1.91	1.26	3.14	1.52
	2.19	1.45	2.98	1.45
Enolase	1.69	1.12	2.86	1.39
	2.45	1.62	1.57	0.76
	2.4	1.59	2.09	1.01
Phosphoglycerate mutase	1.96	1.3	2.12	1.03
	3.38	2.24	1.98	0.96

Since a similar physiologic behaviour is expected in both wild type and recombinant strains in order to be able to attribute the changes in gene expression just to oxygen level oscillations, the control strain is also considered (actually it can be considered from start of the screening). At the second round of the screening, pyridoxine (pyridoxamine) phosphate oxidase

showed different trends in recombinant strain (expressing) and wild type and, thus, it was omitted. It can not be inferred whether the expression level change was due to an experimental error or just as the result of the expressed foreign protein. Counteracting effects may have masked the effect of the oxygen. Such inconsistencies between two strains, expressing and control, were also found regarding available isoforms of the pyruvate decarboxylase, enolase, and phosphoglycerate mutase. Therefore, they were removed and two promising candidates, i.e., pyrimidine precursor biosynthesis enzyme *thi3* and pyruvate kinase remained. Since pyruvate is an important branch point of the respiratory and fermentative pathways (Figure 4.10) and it may be fruitful to analyze the behavior of the promoters related with this critical conjunction, it was decided to include pyruvate decarboxylase as third promising candidate. The three selected proteins have been presented in Table 4.8.

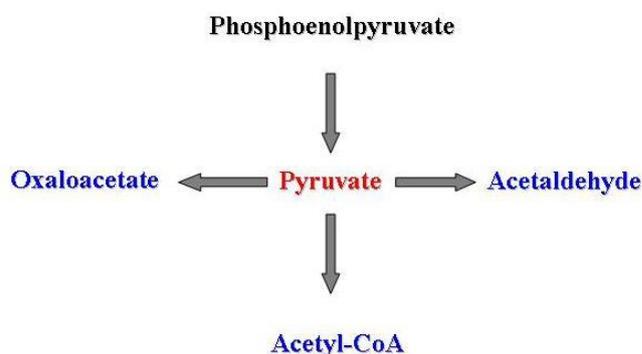


Figure 4.10. simple representation of the pyruvate node in yeasts.

However, there will be uncertainties about the strength of the pyruvate decarboxylase promoter by considering that *P. pastoris* pyruvate decarboxylase is less efficient at decarboxylation in comparison to all isozymes of *S. cerevisiae* (Agarwal et al., 2013).

Table 4.8. Final selected proteins which their corresponding promoters were regarded as the promising oxygen limitation-induced promoters can be utilized in *Pichia*-based expressions.

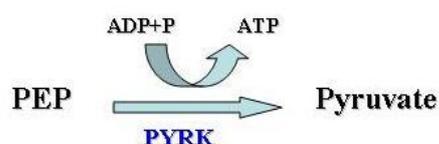
Protein name	Short Name	Protein ID
Pyruvate kinase	<i>PYRK</i>	PIPA00751
Pyruvate decarboxylase	<i>PDC</i>	PIPA01726
Pyrimidine precursor biosynthesis enzyme thi3	<i>THI3</i>	PIPA00420

4.3.2. Finding the nucleotide sequence of the selected promoters

The presented protein IDs in Table 4.8 belong to the *P. pastoris* DSMZ strain. However, the recruited strain in Baumann et al. study (2010) was *P. pastoris* X-33 strain. Therefore, determination of the selected promoters nucleotide sequence was based on the *P. pastoris* GS115 strain which is isogenic to *P. pastoris* X-33 strain with just one deleted gene. Genome sequence of the *P. pastoris* GS115 strain is publicly available at (www.ncbi.nlm.nih.gov).

4.3.2.1. Pyruvate kinase promoter (P_{PYRK})

PYRK enzyme catalyzes the final step in glycolysis. This process requires K^+ and Mg^{2+} or (Mn^{2+}):



PEP and F-1,6-biphosphate enhances enzyme activity allosterically. High concentrations of ATP, acetyl-CoA, and long-chain fatty acids (signs of the abundant energy supply) allosterically inhibit all isozymes of pyruvate kinase. Accumulation of the alanine which can be synthesized from pyruvate

in one step also, allosterically, inhibits pyruvate kinase, slowing the production of the pyruvate by glycolysis.

A BLAST analysis was performed between PIPA00751 (pyruvate kinase in DSMZ strain) and entire genome of the GS115 strain and 95% identical sequence was obtained between PIPA00751 and PAS_chr2-1_0769, which has been annotated as pyruvate kinase in GS115 strain. The location and direction of PAS_chr2-1_0769 in GS115 strain along with its preceding gene, i.e. PAS_chr2-1_0770, has been elucidated in Figure 4.11.

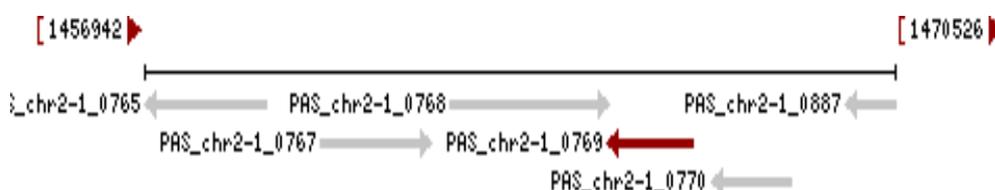


Figure 4.11. The chromosomal location of the pyruvate kinase gene and its neighboring genes. Red arrow shows the pyruvate kinase gene.

The preceding gene of PIPA00751 is PIPA00749 which is 92% similar to PAS_chr2-1_0770 based on the BLAST analysis in NCBI. Since the gene before PAS_chr2-1_0769 is PAS_chr2-1_0770, the ‘*putative promoter region*’ was supposed to be the segment between these two genes, the end of the PAS_chr2-1_0770 and the beginning of the PAS_chr2-1_0769. By utilization of the nucleotide sequences of these two genes the region between them was obtained in sequenced strand of chromosome 2; then, the complementary sequence was inferred and rearranged in 5’ to 3’ direction to obtain the following sequence as the promoter region of pyruvate kinase gene which precedes the “ATG” codon of the pyruvate kinase gene and has 345 bp length (Figure 4.12). The details of the sequences can be found in Appendix M.

“Pyruvate kinase *putative* promoter region”

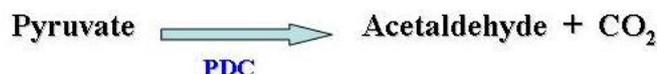
```
5'...TTCAGTGTGCGGGATACTGTATTCCGCTCGGGGTTCTAAAGAAATTGTTTAAACTAAACCAA  
TCGGATCAGAGGTTCCGTACGTTTTTCACATTC AAGGATGAGGGTTTTCCACGAGTGA ACTATTAC  
TCCGGTCTCCCACCATCATTTGCGGAATGAAACCTTTTGTGCTGAGATTGTATAGGGCGTGGGGAC  
GGACGCTTCTTAACCGTTC CCTAG AATGTCGTCCCCTGATCAAAATTTAATGGCATCCA ACTTTG  
CTGTAATAGGTATATATAACCTAGCAGGCGACCGTTCATGTACAGTAAATTGTTT TAGACTTTTTT  
TTAACTGAAATCAATCCA...3'
```

Figure 4.12. Nucleotide sequence of the putative promoter region peculiar to pyruvate kinase gene with 345 bp length in *P. pastoris* GS115 strain.

It should be expressed that the promoter region which was obtained from DSMZ strain as the sequence between PIPA00751 and PIPA00749 was 88% identical to the determined promoter region in GS115 strain.

4.3.2.2. Pyruvate decarboxylase promoter (P_{PDC})

PDC enzyme catalyzes the following reaction:



This enzyme activity depends on the cofactors thiamine pyrophosphate (TPP) and Mg^{2+} . The enzyme locates in the ethanol production pathway; in anaerobic conditions, it is part of the fermentation process that occurs in yeasts to produce ethanol by fermentation. The PDC activity in *P.pastoris* is induced by the presence of glycolytic substrates (ex., glycerol and glucose) and it has relatively higher activity on glucose (Agarwal et al., 2013).

A BLAST analysis was also performed between PIPA01726 (pyruvate decarboxylase in DSMZ strain) and entire genome of the GS115 strain and 95% identical sequence was obtained between PIPA01726 and PAS_chr3_0188, which has been annotated as pyruvate decarboxylase in

GS115 strain. The location and direction of PAS_chr3_0188 in GS115 strain along with its preceding gene, i.e. PAS_chr3_0189, has been represented in Figure 4.13. However, PIPA01726 has 813 nt. and PAS_chr3_0188 possesses 1683 nt.; therefore, a BLAST analysis was performed between PAS_chr3_0188 and the whole genome sequence of the DSMZ strain. Two separate segments each with high resemblance to half of the PAS_chr3_0188 were obtained; PIPA03164 and PIPA01726 where both of them reside at the end of the corresponding contigs, 00224 and 00063, respectively. The gene before PIPA03164 is PIPA09857 and the gene before PIPA01726 is PIPA01725 in DSMZ strain. There was no similarity between PIPA01725 and PAS_chr3_0189; in contrast, 93% identical sequence was obtained between PIPA09857 and PAS_chr3_0189.

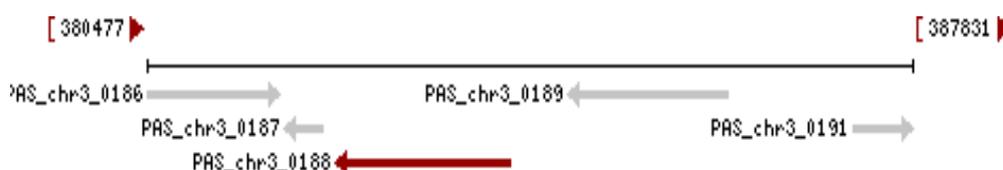


Figure 4.13. The chromosomal location of the pyruvate decarboxylase gene and its neighboring genes. Red arrow shows the pyruvate decarboxylase gene.

Since the gene before PAS_chr3_0188 is PAS_chr3_0189, the ‘*putative promoter region*’ was supposed to be the segment between these two genes, the end of the PAS_chr3_0189 and the beginning of the PAS_chr3_0188. By utilization of the nucleotide sequences of these two genes the region between them was obtained in sequenced strand of chromosome 3; then, the complementary sequence was inferred and rearranged in 5’ to 3’ direction to obtain the following sequence as promoter region of pyruvate decarboxylase gene which precedes the “ATG” codon of

the pyruvate decarboxylase gene (Figure 4.14) and has a length of 546 bp. The details of the sequences can be found in Appendix M.

“Pyruvate decarboxylase *putative* promoter region”

```
5' ...GGGACAAGCACACGATTACCCAATCACTTGATATGCACCAATTTGTTCCGTTGTTTATGC
CATATTTACCGAATTTTCTTCCCAGGTTTTTCCGAATGGACATCTGTAGTCCACTTTTTGGTT
ATCATAATCGTCCCACAAGTCGTGGATTTAACCAGAACCCTAGTAATTTTAAGTTCGCTATTAA
TCACTCAGAATGGTCTCACCTTGCTATTGGCCAAGTCTGGAGTCGCCAGCTACCACCTCAGAG
GCTACATAGACCTCCCAATGTCATCTCCTCAGTGGCTCTTCAATCTCGTGTCTTTTCCGTTA
AAACTCCGTTTCGTTTACCCTATACTGCCCTGGTTGTGCAGCTCTTACCACTTCGCGCCGCT
ACTATCCGTAGTGGTCGAGCCGCATCAATATCACGTTGAAATAGAATAACTCCCTACAAAAGC
CGCACGCAACCATCAAATCTATAAAGGAACCTCAAATATCTAGCAACATCTTTTCAATTTAC
TACAACATATTCGTTAATCATCAATCAATTAGCTAGTACACAACA...3'
```

Figure 4.14. Nucleotide sequence of the *putative* promoter region peculiar to pyruvate decarboxylase gene with 546 bp length in *P. pastoris* GS115 strain.

It should be emphasized that the nucleotide region between PIPA01726 and PIPA01725 which was obtained from DSMZ strain had no similarity to the *putative* promoter region represented in Figure 4.14. In contrast, the nucleotide region between PIPA03164 and PIPA09857 in DSMZ strain with 546 bp resulted in 92% similarity with the determined *putative* promoter region in GS115 strain.

4.3.2.3. Pyrimidine precursor biosynthesis enzyme thi3 promoter (P_{THI3})

Pyrimidine precursor biosynthesis enzyme thi3 (THI3~KID1) is probably involved in the biosynthesis of the pyrimidine moiety of the thiamin molecule (Figure 4.15). In thiamine, the thiazole and pyrimidine moieties are biosynthesized separately and then combined to form thiamine phosphate (ThMP) by the action of thiamine phosphate synthase. In most bacteria and in eukaryotes, ThMP is hydrolyzed to thiamine which then may be pyrophosphorylated to ThDP (TPP) by thiamine diphosphatase.

Recommended name for this enzyme in (www.uniprot.org) is thiamine metabolism regulatory protein THI3 (EC4.1.1.-).

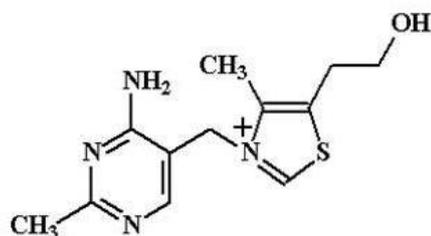


Figure 4.15. Schematic structure of the thiamine molecule.

In *S. cerevisiae* three positive regulators have been identified for thiamine biosynthesis pathway, *THI2*, *THI3*, and *PDC2*. Thi3p is a regulatory protein (in *S. cerevisiae*) that binds Pdc2p and Thi2p transcription factors and activates these thiamine biosynthesis transcription factors but release and deactivates them upon binding to TPP, the end-product of the biosynthesis (Web 16).

By finding the nucleotide sequence of the PIPA00420 and BLAST of the whole chromosome sequences of GS115 in NCBI with this sequence it was found that there is a significant alignment, 94% identical, in chromosome 4 between PAS_chr4_0065 sequence and PIPA00420. The attributed function for PAS_chr4_0065 was “protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (HMP)” which its nucleotide sequence can be found in Appendix M. According to the Figure 4.16 the genes that the promoter region should be searched between them were supposed to be PAS_chr4_0065 and PAS-chr4_0064 which were transcribed in opposite directions.

By utilization of the nucleotide sequences of these two genes (Appendix M) the region between them was obtained in sequenced strand of chromosome 4 and was presented as *putative* promoter region (Figure 4.17)

which precedes the “ATG” codon of the protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (HMP) gene and have length of 458 bp.



Figure 4.16. The chromosomal location of the *THI3* enzyme gene and its neighboring genes. Red arrow shows the gene of protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (HMP).

“Pyrimidine precursor biosynthesis enzyme *putative* promoter region”

```

5' ...GTCTTTGTAAGATTATAGTTT CAGAACTGGAATTGAGCTCAAAAACTGGAATCGAGC
GGATAATTTGAAGATTGATGCCCTTACTCATGAATTGATTGATAAGAGCTCCGTGATTCACCTCG
TCAATGATTACCCCTCTCCTACCCGATTTGGGACTTTTTCTTCAGTCTTGGGGACTTTTTTTC
ATATGACTTGACCTTGCTTTCCCAATAGGGAAGGACTCACCCATGGATGATTAAGTTTGGATT
ACTCGTTTAGGAAATAGTAGCCATGAATCAATTTGAATCATACCATCATGAAATAGGTTAGG
CTGTAAATGCCTCAAAAATGGCTCTTGAGGCTGGATTTTTGGGTATTGGAATGTTGGTAGCAA
TTGGTATAAAAAGGCCATTTGTATTTCACTTTTTTGTCTTCATACTTTACTCTTCTCAACTTT
GGAACTTCAATAAATCATC...3'

```

Figure 4.17. Nucleotide sequence of the *putative* promoter region peculiar to pyrimidine precursor biosynthesis enzyme gene with 458 bp length in *P. pastoris* GS115 strain.

Using BLAST in (www.pichiagenome.org) between the obtained promoter region from GS115 and whole genome sequence of DSMZ strain led to the sequence with 87% identical nucleotide sequence to the desired region. The result has been represented in the Figure 4.18. It is clear that the similar region locates between PIPA10077 and PIPA00420. BALST analysis between PIPA10077 and whole genome of the GS115 strain revealed 88% similarity with PAS_chr4_0064.

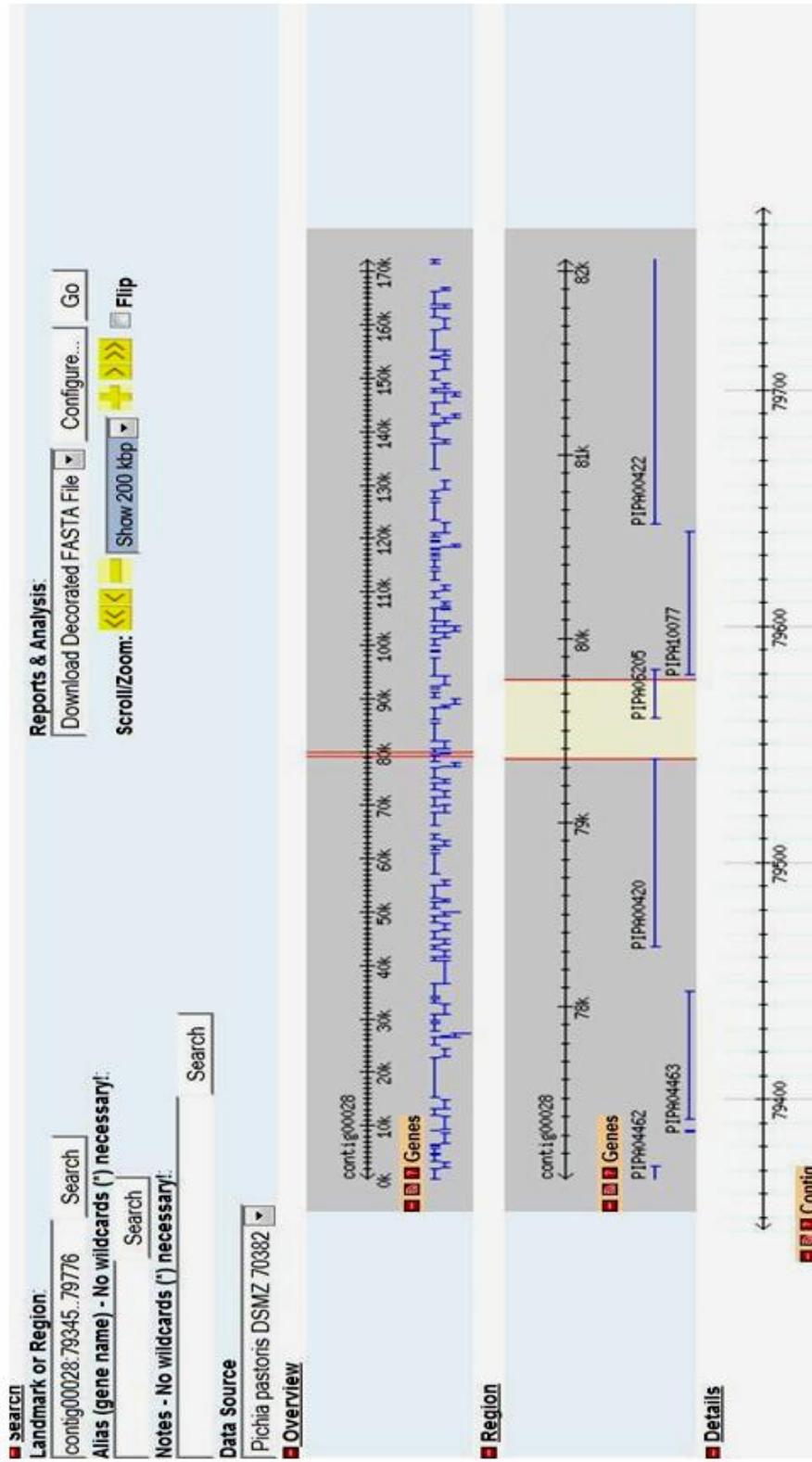


Figure 4.18. The BLAST analysis result between the determined *putative* promoter region in *P. pastoris* GS115 strain and the whole genome of the *P. pastoris* DSMZ strain.

4.3.3. In-silico appraisal of the selected putative promoter regions

In order to in-silico validation of the determined *putative* promoter regions, the sequences of P_{PYRK}, P_{PDC}, and P_{THI3} were analyzed by the program available at (http://www.fruitfly.org/seq_tools/promoter.html) which allocates a score to each sequence shows the “promoter-ness” of a region. The output of this program is a 50-nucleotide segment which pretends to have a promoter activity. In addition, the transcription start site is also identified in this fragment. Along with the selected promoter regions, two positive controls were also analyzed in order to verify the precision of the program prediction. Positive controls were P_{GAP} and sorbitol dehydrogenase promoter (P_{SDH}). The analyses results have been presented in Table 4.9. P_{GAP} sequence was purchased from (Invitrogen) and P_{SDH} was obtained from (Periyasamy et al., 2013); they can be found in Appendix M.

Table 4.9. The prediction results of the in-silico analyses conducted with *putative* promoter regions along with two positive controls.

Promoter	Start Nucleotide	End Nucleotide	Attributed Score
PYRK	263	313	0.99
PDC	420	470	0.91
	448	498	0.91
THI3	311	361	0.85
	370	420	0.84
GAP	91	141	0.81
	258	308	0.90
	322	372	0.98
	377	427	0.95
SDH	4	54	0.97
	58	108	0.82

According to the positive control results, the program predictions were supposed to be reliable and, therefore, all the selected regions have the potential of being a promoter region.

4.4. Provision of the restriction sites for endogenous promoters and SPs

By development of BP (section 4.1), choosing potentially more efficient endogenous SPs (section 4.2), and selection of the potential oxygen limitation-induced endogenous promoters (section 4.3), the way was paved for preparation of the new plasmids, based on BP, containing qualified promoters and SPs. Regarding Figure 4.6 it is obvious that in order to substitute α -MF in BP with new SPs, the original leader sequence, α -MF, should be removed by the action of *Bsp119I* and *EcoRI* restriction enzymes. In the case of the selected promoters, it was decided to remove the whole original promoter and leader sequence (P_{GAP} and α -MF together) and, then, insert new promoter in combination with α -MF (promoter construct) in order to preserve the original distance between transcription start site and translation start site; this makes expression genuine and similar to the real conditions in genome and the promoter binds exactly to the “ATG” codon of the α -MF (Appendix N). Therefore, the cleavage should be accomplished by *BglIII* and *EcoRI*. However, *BglIII* cuts the available *hGH* gene in BP and, in consequence, another restriction enzyme was found, *NsiI*. In the case of each SP and promoter, a pair of primers was also designed in order to be utilized whenever needed for control or amplification of the constructs.

4.4.1. Secretion signal peptides constructs and related primers

Some issues should be considered in developing SP constructs and their corresponding primers in order to make their amplification and, subsequent, insertion in BP possible. According to the available map of pGAPZ α A (Invitrogen) there is a 9-nucleotide distance between the end of the P_{GAP} and start of the α -MF which contains ribosome binding site (RBS). *Bsp119I* restriction enzyme has a restriction site in this region. Therefore, all the novel SPs (SP13, SP23, SP24, SP26, and SP34) constructs and their forward primers were provided with the restriction site of the *Bsp119I* in their 5' end. The restriction site of the *EcoRI* was also added to the 3' end of the constructs and reverse primers.

The recognition sequence of the *Bsp*119I enzyme is as follows:

5' ...TTCGAA... 3'

3' ...AAGCTT... 5'

After cutting:

5' ...TT CGAA...3'

3' ...AAGC TT...5'

*Eco*RI recognition sequence is:

5' ...GAATTC...3'

3' ...CTTAAG...5'

After cutting:

5' ...G AATTC...3'

3' ...CTTAA G...5'

It should be noticed that in original plasmid (BP) there is a "ACG" sequence after restriction site of *Bsp*119I and before "ATG" codon of α -MF. Therefore, this "ACG" was included in designed constructs and forward primers in order not to influence the efficiency of the P_{GAP} because of the change in distance between promoter and the translation start codon, ATG.

Since it was decided to amplify SPs with PCR and, then, digest them with restriction enzymes, for *Bsp*119I cleavage site, 1-5 bp was added as leader nucleotides, to the primers to increase the cleavage efficiency 50-100%. For *Eco*RI cleavage site, adding 1-5 bp leads to 50-100% increase in cleavage efficiency and it was taken into account in primers design. Three nucleotides were added for enhancing cleavage efficiency in all cases. At the end, melting temperature of the primers along with the homo dimer and hetero dimer production situations were also checked for primers. The designed constructs for SPs has been represented in Table 4.10. Based on the above mentioned explanations, after several times of trial and error to meet

the requirements of an optimum design, the primers were designed for amplification of the SPs in order to subsequent insertion in BP; they have been represented in Table 4.11.

Table 4.10. SP constructs after inclusion of the required nucleotides in both ends. The capital letters represent the restriction sites. The italic capital sequence is the three nucleotide devised in constructs in order to preserve the originality of the P_{GAP} .

SP Construct (5' → 3')
SP13 TTCGAAAC <i>CG</i> atgctatcaactatcttaaatactttatcctgttgctcttcatacaggcatccctacagGAATTC
SP23 TTCGAAAC <i>CG</i> atgaaaatattaagtgattgcttctcttttacgttggcctttgctGAATTC
SP24 TTCGAAAC <i>CG</i> atgaaagttctacgaccaaattctggctgtgtcttattagtagactcgttgcgctGAATTC
SP26 TTCGAAAC <i>CG</i> atgtggcgcgtgttcatactggactattaatctctatcctttggccttgaGAATTC
SP34 TTCGAAAC <i>CG</i> atgagaccagtgccttcgttattactcttgctggctcttcggactcgcctGAATTC

Table 4.11. The final form of the designed primers for SPs.

Primer Name	Corresponding Nucleotide Sequence (5' → 3')
SP13-F	<i>gcgttcgaaACG</i> atgctatcaactatc
SP13-R	agagaattcctgtagggatgcctgtat
SP23-F	<i>gcgttcgaaACG</i> atgaaaatattaagtg
SP23-R	acagaattcagcaaaggccaacgtaaa
SP24-F	<i>cagttcgaaACG</i> atgaaagtttctacg
SP24-R	atagaattcagcgcaaacgagtctaac
SP26-F	<i>atattcgaaACG</i> atgtggcgcgtgttc
SP-26-R	gaggaattctccaaggaccaaaggatagaag
SP34-F	<i>atattcgaaACG</i> atgagaccagtgc
SP34-R	aacgaattcagcgagtaccgaagaa

4.4.2. Promoters constructs and related primers

As mentioned, each promoter construct was regarded as a combination of the selected promoter and α -MF; some issues similar to SPs should also be considered in developing promoter constructs and their corresponding primers in order to make their insertion in BP possible. Considering the available map of BP, all the selected promoters (P_{PDC} , P_{PYRK} , P_{THI3}) constructs and their forward primers were provided with the restriction site of the *Nsi*I in their 5' end. The restriction site of the *Eco*RI was also added to the 3' end of the constructs and reverse primers. The recognition sequence of the *Nsi*I enzyme is as follows:

5'...ATGCAT...3'
3'...TACGTA...5'

After Cutting:

5' ...ATGCA T...3'
3' ...T ACGTA...5'

In addition, the restriction site of the *Nsi*I enzyme resides a few nucleotides before the start of the P_{GAP} ; therefore, in designing promoter constructs and corresponding forward primers these a few nucleotides were also included in order to mimic original BP. The extra leader nucleotides in order to increase the cleavage efficiency of the restriction enzymes were also taken into consideration in designing forward and reverse primers. The designed primers have been presented in Table 4.12.

Table 4.12. The final form of the designed primers for promoters. The underlined sequence is the preserved nucleotides to mimic the original BP. The bold sequences are the restriction sites of the *NsiI* and *EcoRI*

Primer Name	Corresponding Nucleotide Sequence (5' → 3')
<i>PDC-F</i>	TCTATGCAT <u>GAGATCGGGACA</u> AGCA
<i>PDC-R</i>	TGCGAATTCAGCTTCAGCCTCTCTTT
<i>PYRK-F</i>	TAAATGCAT <u>GAGATCTTCAGTGTGCGG</u>
<i>PYRK-R</i>	TGAGAATTCAGCTTCAGCCTCTCTTTTCT
<i>THI3-F</i>	GAGATGCAT <u>GAGATCGTCTTTG</u> TAAATAGT
<i>THI3-R</i>	TGAGAATTCAGCTTCAGCCTCTCTTTTC

4.5. Generation of r-*E. coli* strains with developed plasmid constructs

The plasmid constructs developed in current research are of two types: the ones with novel endogenous SPs in combination with P_{GAP} and, the others with α -MF in combination with selected oxygen limitation-induced promoter. After preparation of the BP, the promoters and SPs were prepared. Then, by performing required substitutions in BP the new plasmids were developed and, subsequently, were cloned in *E. coli* in order to proceed to *P. pastoris* transfection.

4.5.1. Synthesizing and preparation of the designed constructs

After pre-experimental in-silico determination of the desired endogenous promoters and secretion SPs of *P. pastoris* and subsequent designing of their corresponding constructs and primers (section 4.4), all the SPs and promoters constructs were synthesized in a unique plasmid, pUC57::*PRSP*, by GenScript (Appendix C). The synthesized sequence length was 2552 bp and, therefore, the length of the pUC57::*PRSP* became ~5200 bp. The plasmid was delivered in lyophilized form in order to be further utilized. Upon receiving the synthesized plasmid, it was cloned in *E. coli* DH5 α cells. Four colonies were selected among several putative transformants and plasmid

isolation was conducted. For further verification of the isolated plasmids, single-digestion of the isolated plasmids from four colonies was conducted by the *Bam*HI restriction enzyme according to the supplier information. The isolated circular and single-digested plasmids were subjected to the agarose gel electrophoresis and the result can be observed in Figure 4.19.

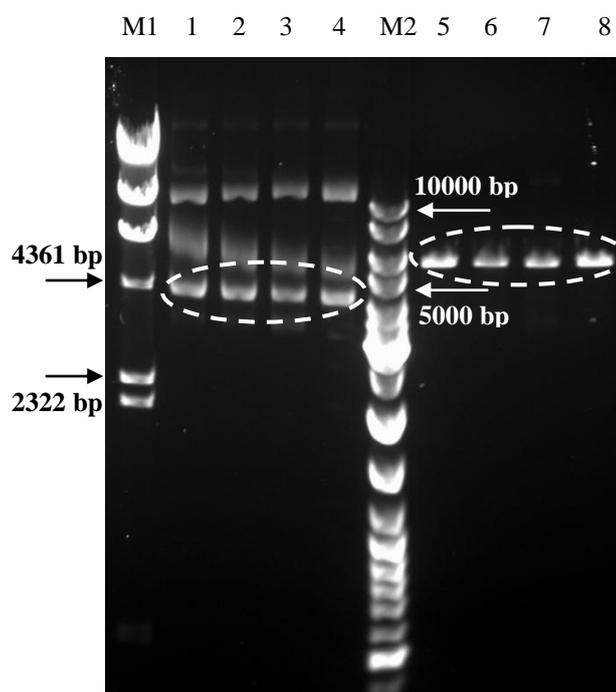


Figure 4.19. pUC57::PRSP plasmid verification. M1: DNA marker λ DNA/HindIII, 1, 2, 3, 4: Circular isolated pUC57::PRSP plasmid from four putative transformant colonies, M2: DNA ladder mix, 5, 6, 7, 8: pUC57::PRSP plasmids of four putative transformants single-digested by *Bam*HI restriction enzyme.

According to the obtained results, transformation has been successful and, therefore, Microbank™ stock from one of the four colonies was prepared for coming experiments.

4.5.1.1. Desired promoter constructs

As the desired constructs (SPs and promoters) were included in the pUC57::*PRSP* plasmid, it was used as template for preparation of the synthesized constructs. In order to get promoter constructs, the synthesized plasmid, pUC57::*PRSP*, was double-digested by the aid of suitable restriction enzymes, *NsiI* and *EcoRI*. Three promoter constructs were released simultaneously. The result of the double-digestion was run on the gel which can be found in Figure 4.20.

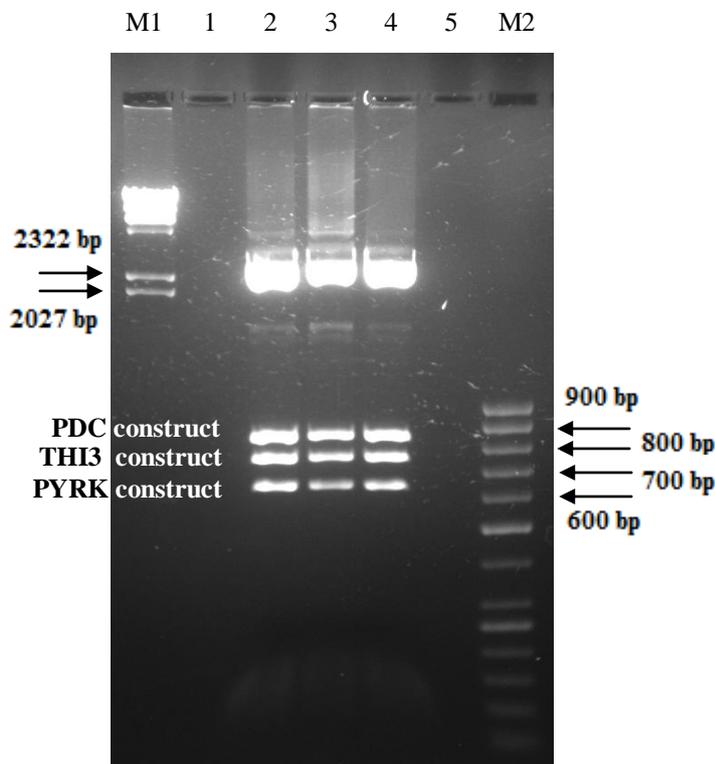


Figure 4.20. Double-digestion of the pUC57::*PRSP* plasmid with *NsiI* and *EcoRI* restriction enzymes. M1: DNA marker λ DNA/HindIII, 2, 3, 4: Double-digested pUC57::*PRSP* with *EcoRI* and *NsiI*, M2: GeneRuler™ 50bp DNA ladder. Expected sizes for P_{PDC} construct, P_{TH13} construct, and P_{PYRK} construct are: 831 bp, 743 bp, 630 bp, respectively.

The resultant bands were in the desired locations based on the expected sizes of the promoter constructs. After separation of the promoter constructs they were cut from the gel and, then, were extracted by gel extraction kit. The purified segments were again subjected to the gel electrophoresis which can be seen in Figure 4.21. The promoter constructs were ready to be used in next step.

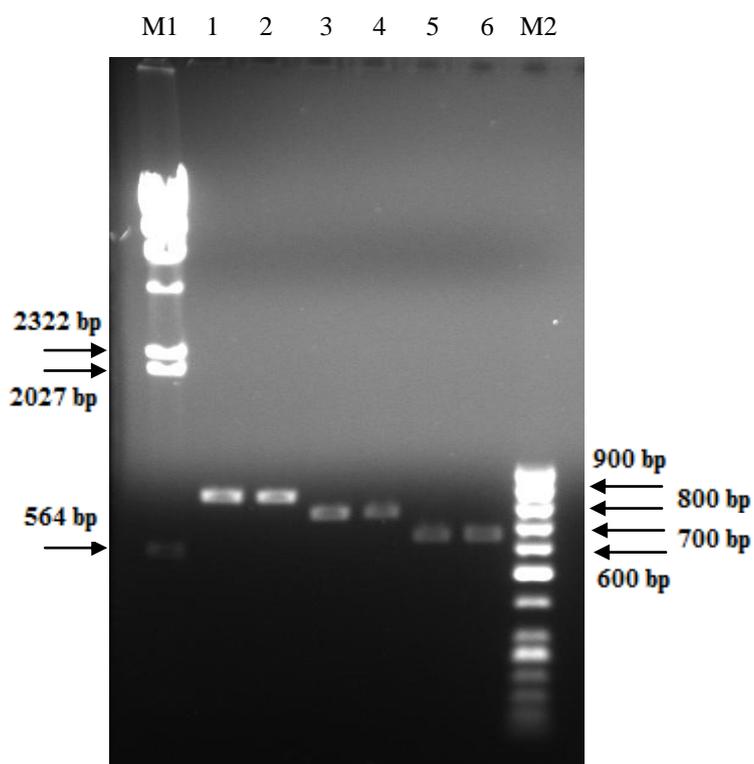


Figure 4.21. Verification of the eluted promoter constructs. M1: DNA marker λ DNA/HindIII, 1, 2: P_{PDC} construct, 3, 4: P_{THI3} construct, 5, 6: P_{PYRK} construct, M2: GeneRuler™ 50bp DNA ladder.

4.5.1.2. Desired signal peptide constructs

The selected endogenous signal peptides (SP13, SP23, SP24, SP26, and SP34) were improvised in pUC57::PRSP along with desirable restriction

sites in 5'- and 3'- ends. The desired SPs were amplified by PCR where pUC57::PRSP was recruited as template. The PCR conditions were tuned carefully for each individual SP by considering the possible similarities in order to prevent non-specific primer binding and obtaining undesirable results. As the results of the amplification were intended to be used in ligation, the PCR mixture was prepared by *Pfu* DNA polymerase instead of *Taq* DNA polymerase because of its lower error rate (Table 4.13). PCR thermal profile for amplification of SP13, SP23, SP24, SP26, and SP34 constructs can be found in Tables 4.14, 4.15, 4.16, and 4.17, respectively.

Table 4.13. PCR mixture for signal peptides amplification.

Component	Amount
10X <i>Pfu</i> buffer without MgSO ₄	5 µL
2 mM solution of dNTPs	5 µL
5 µM forward primer	3 µL or 5 µL
5 µM reverse primer	3 µL or 5 µL
25 mM MgSO ₄	2 µL or 3 µL or 4 µL
<i>pfu</i> DNA polymerase	1-2 U
dH ₂ O	23-29 µL (Variable)
Template DNA (pUC57::PRSP)	2 µL
Total volume	50 µL

For amplification of the SP13 construct, different primer amounts (3µL, 5µL) and salt concentrations (1 mM, 1.5 mM, and 2 mM) were checked. Template amount was kept constant 2µL. In the case of SP23, SP24, SP26, and SP34 constructs amplification, different salt concentrations (1 mM, 1.5 mM, and 2 mM) were checked. Template and primer amounts were kept constant 2 µL and 5 µL, respectively. The PCR results were, then,

subjected to the agarose gel electrophoresis for length verification which can be seen in Figures 4.22, 4.23, and 4.24.

Table 4.14. PCR thermal profile for SP13 amplification.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	55 °C	1 min.
	72 °C	15 sec.
29 cycles	94 °C	1 min.
	59 °C	1 min.
	72 °C	15 sec.
Final extension	72 °C	10 min.

Table 4.15. PCR thermal profile for SP23 and SP34 amplification.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	52.5 °C	1 min.
	72 °C	15 sec.
29 cycles	94 °C	1 min.
	56.5 °C	1 min.
	72 °C	15 sec.
Final extension	72 °C	10 min.

Table 4.16. PCR thermal profile for SP24 amplification.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	52 °C	1 min.
	72 °C	15 sec.
29 cycles	94 °C	1 min.
	56 °C	1 min.
	72 °C	15 sec.
Final extension	72 °C	10 min.

Table 4.17. PCR thermal profile for SP26 amplification.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	55 °C	1 min.
	72 °C	15 sec.
29 cycles	94 °C	1 min.
	59 °C	1 min.
	72 °C	15 sec.
Final extension	72 °C	10 min.

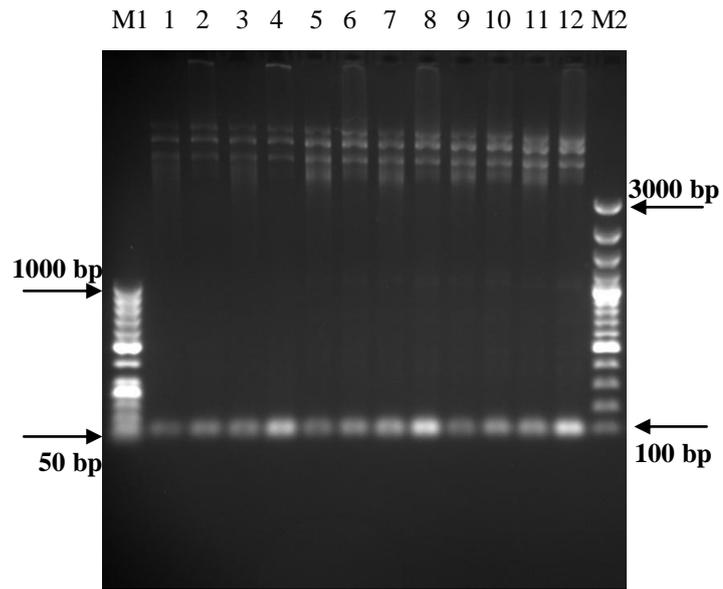


Figure 4.22. PCR results for amplification of SP13 (45 minutes run on 1% agarose gel). M1: GeneRuler™ 50bp DNA ladder, 1, 3, 5, 7, 9, 11: unpurified PCR results, 2, 4, 6, 8, 10, 12: purified PCR results, M2: GeneRuler™ 100bp plus DNA ladder.

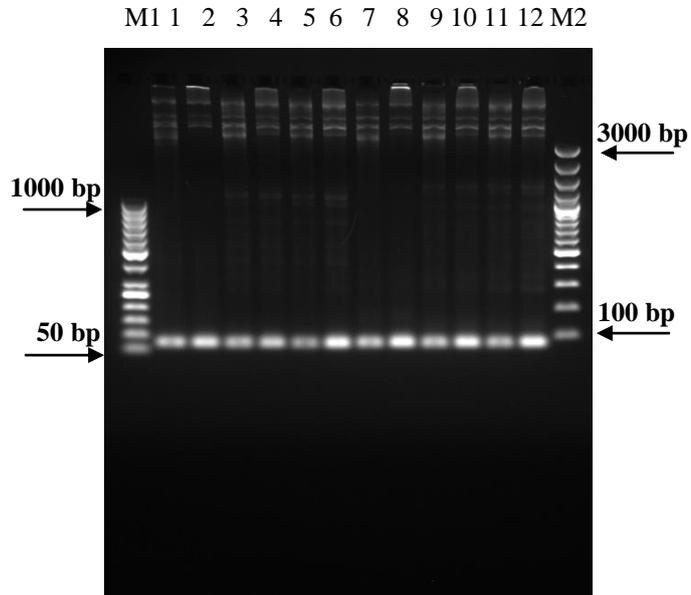


Figure 4.23. PCR results for amplification of SP23 and SP34 (45 minutes run on 1.5 % agarose gel). M1: GeneRuler™ 50bp DNA ladder, 1, 3, 5, 7, 9, 11: unpurified PCR results, 2, 4, 6, 8, 10, 12: purified PCR results, M2: GeneRuler™ 100bp plus DNA ladder.

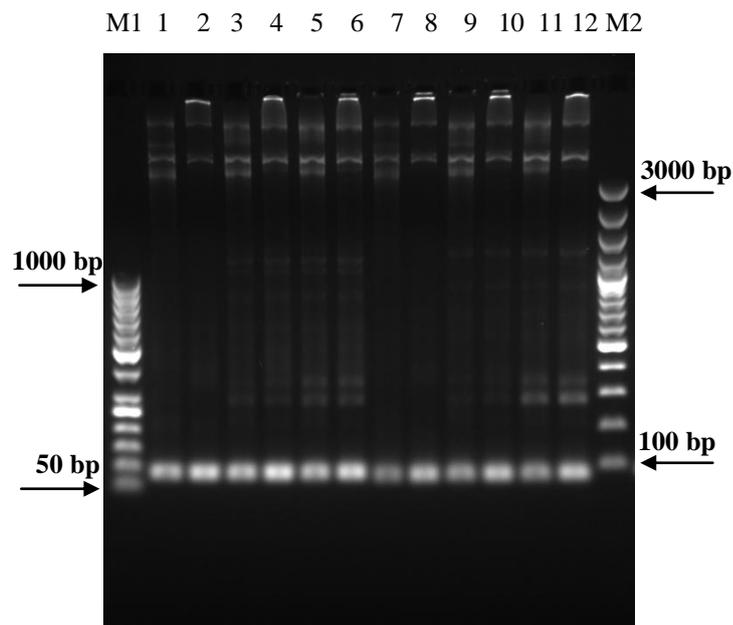


Figure 4.24. PCR results for amplification of SP24 and SP26 (45 minutes run on 1.5 % agarose gel). M1: GeneRuler™ 50bp DNA ladder, 1, 3, 5, 7, 9, 11: unpurified PCR results, 2, 4, 6, 8, 10, 12: purified PCR results, M2: GeneRuler™ 100bp plus DNA ladder.

4.5.2. Development of the new plasmids with selected promoters

In order to obtain plasmids that harbor desired promoters, substitution of the P_{GAP} in the BP with selected oxygen limitation-induced promoters (P_{PDC} , P_{THI3} , and P_{PYRK}) was conducted by double-digestion of the BP with *Nsi*I and *Eco*RI restriction enzymes and subsequent ligation of the promoter-free backbone with the prepared promoter constructs from section 4.5.1.1.

4.5.2.1. Preparation of the desired promoter-free backbone from BP

BP was isolated from the corresponding *r-E. coli* overnight culture by plasmid isolation kit. Afterwards, double-digestion of the isolated BP was performed with *Eco*RI and *Nsi*I restriction enzymes as described earlier. The result of the enzymatic digestion was run on the agarose gel (Figure 4.25). By cleavage, P_{GAP} along with α -MF was cleaved and the specified upper band remained as promoter-free backbone utilized for insertion of the new

promoters. Then, the procedure was followed by gel elution in order to obtain pure backbone fragment (Figure 4.26).

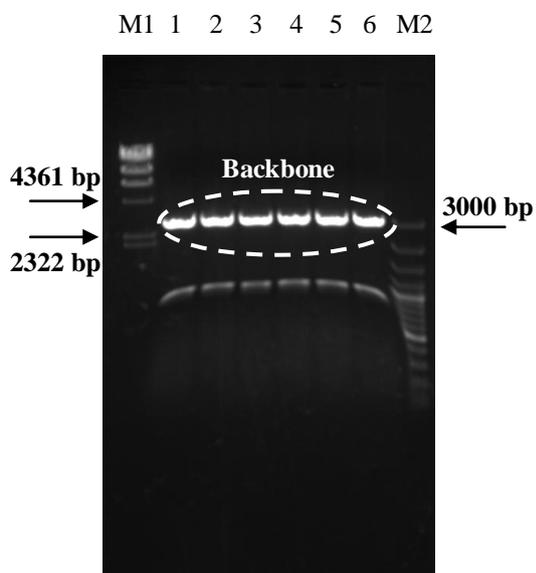


Figure 4.25. Double-digestion of the BP with *EcoRI* and *NsiI* restriction enzymes. The expected size of the generated backbone is ~ 3000 bp. M1: DNA marker λ DNA/HindIII, 1-6: double digested BP samples, M2: GeneRuler™ 100bp plus DNA ladder.

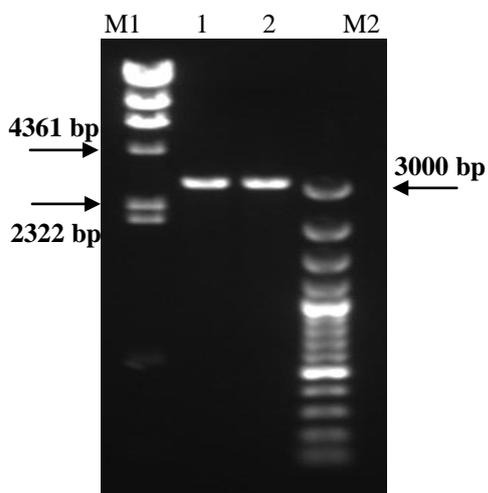


Figure 4.26. Developed promoter-free backbone after double-digestion of BP with *EcoRI* and *NsiI* restriction enzymes and subsequent gel extraction. M1: DNA marker λ DNA/HindIII, 1, 2: eluted backbone, M2: GeneRuler™ 100bp plus DNA ladder.

4.5.2.2. Preparation of the selected promoters

The selected promoter fragments were obtained from the pUC57::*PRSP* plasmid in the form of specific constructs (section 4.5.1.1). As *Nsi*I and *Eco*RI cleavage sites have been provided at the ends of the promoter constructs, they were released from pUC57::*PRSP* upon double-digestion with *Nsi*I and *Eco*RI in a ready-to-ligate form.

4.5.2.3. Ligation of the backbone and selected promoters constructs

The ligation reaction between insert (promoter constructs) and backbone, promoter-free result of the BP double-digestion with *Eco*RI and *Nsi*I, was accomplished by trying different molar ratios. After completion of the reaction, the mixture was directly used to transform the *E. coli* DH5 α cells and, based on the transformation results, the molar ratio between 4 and 5 seemed to be suitable to render ligation. The ligation results are the plasmids: pPDCZ α A::*hGH*, pPYRKZ α A::*hGH*, and pTHI3Z α A::*hGH*.

4.5.2.4. Transformation of the *E. coli* cells with developed plasmids

For all three promoters, transformation of the single-colony *E. coli* DH5 α cell was performed by ligation results from section 4.5.2.3. The expected sizes of the developed plasmids have been represented in Table 4.18.

Table 4.18. Expected size of the developed plasmids harboring desired promoters.

Plasmids Harboring Selected Promoters	Expected Size (bp)
pPDCZ α A:: <i>hGH</i>	3750
pTHI3Z α A:: <i>hGH</i>	3662
pPYRKZ α A:: <i>hGH</i>	3549

After transformation, for the entire three promoter constructs, putative transformants were observed on selective LSLB agar plates containing Zeocin™. In the case of each promoter, 12 colonies were selected at first step and were inoculated into new selective LSLB agar plates containing Zeocin™ in order to grow single-colony cultures. After 16-24 hours incubation at 37°C, the single-colony cultures were used to inoculate LB broth for plasmid isolation. Verification of the isolated plasmids of the putative transformants was performed in three rounds including: agarose gel electrophoresis of the isolated circular plasmids, double-digestion of the plasmids, and PCR. The gel electrophoresis of the selected 6 colonies of each promoter transformants was conducted and the results have been shown in Figure 4.27 and 4.28.

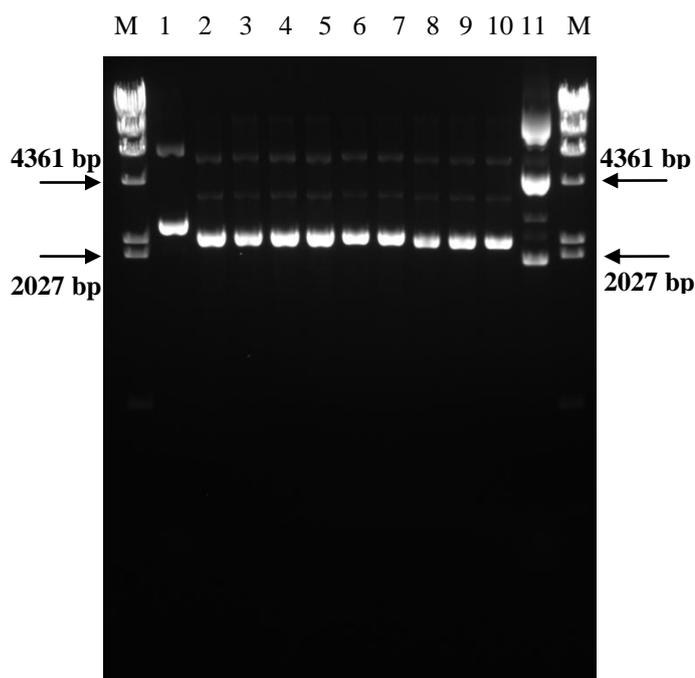


Figure 4.27. Verification of the isolated plasmids of putative transformants related with selected promoters. M: DNA marker λ DNA/HindIII, 1: Circular pPICZ α A::hGH (4142 bp), 2-7: Isolated circular plasmids of putative transformants of P_{PDC} construct (expected size 3750 bp), 8, 9, 10: Isolated circular plasmids of putative transformants of P_{TH13} construct (expected size 3662 bp), and 11: Circular pGAPZ α A (3147 bp).

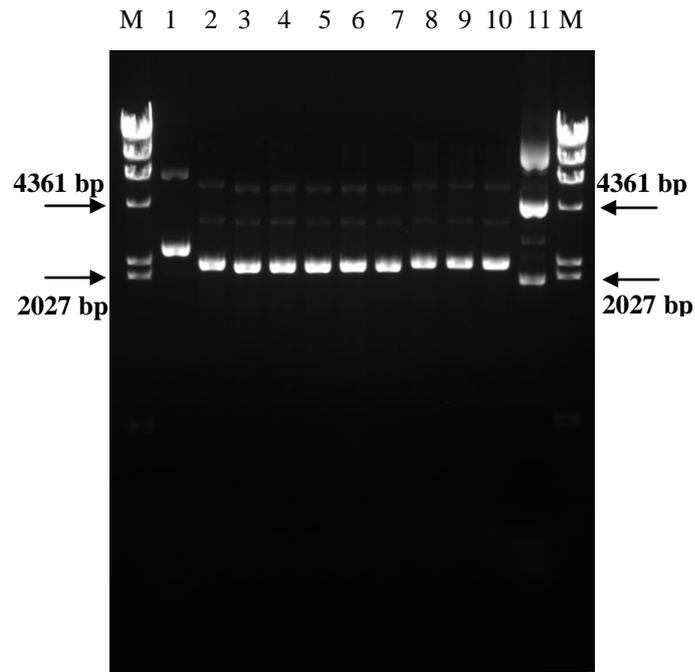


Figure 4.28. Verification of the isolated plasmids of putative transformants related with selected promoters. M: DNA marker λ DNA/HindIII, 1: Circular pPICZ α A::hGH (4142 bp), 2-7: Isolated circular plasmids of putative transformants of P_{PYRK} construct (expected size 3549 bp), 8, 9, 10: Isolated circular plasmids of putative transformants of P_{THI3} construct (expected size 3662 bp), and 11: Circular pGAPZ α A (3147 bp).

After verification of the circular plasmids of putative transformants, among the 6 colonies subjected to gel electrophoresis, 4 colonies from each group of the putative transformants were selected and double-digestion of their isolated circular plasmids with *EcoRI* and *XbaI* was conducted as the second round of the verification. It was expected that double-digestion results show released *hGH* gene in the agarose gel electrophoresis. All the digested plasmids exhibited a cleaved band (Figure 4.29).

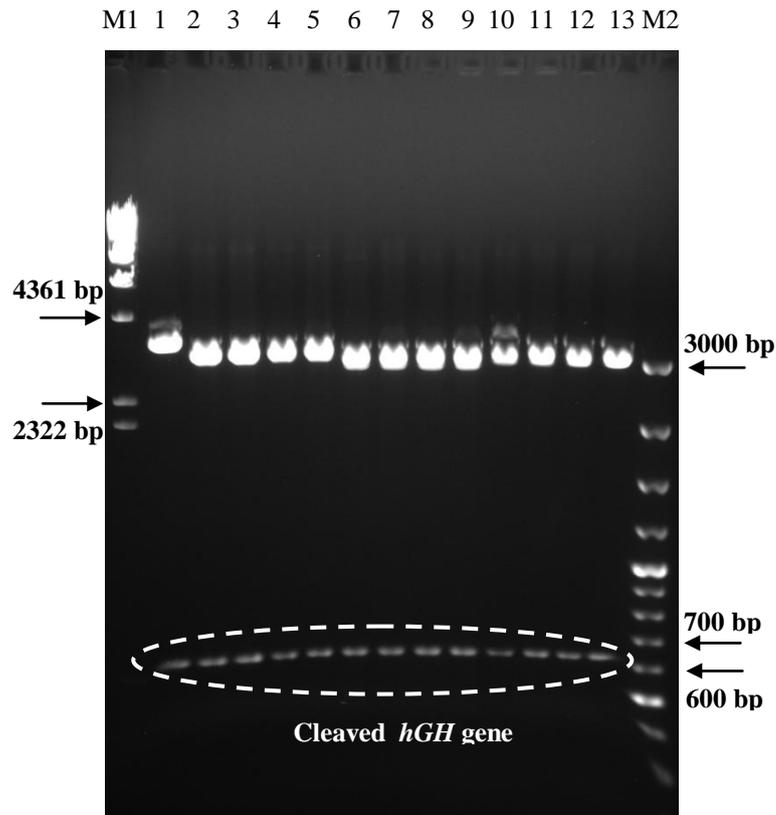


Figure 4.29. Double-digested (with *EcoRI* and *XbaI*) plasmids of putative transformants related with selected promoters (P_{PDC} , P_{PYRK} , and P_{THI3}). M1: DNA marker λ DNA/HindIII, 1: double-digested pPICZ α A::hGH utilized as positive control. 2-5: Digested plasmids of putative transformants related with P_{PDC} , 6-9: Digested plasmids of putative transformants related with P_{PYRK} , 10-13: Digested plasmids of putative transformants related with P_{THI3} , M2: GeneRuler™ 100bp plus DNA ladder.

After verification of the four colonies in each group, two colonies from each group (promoter) were selected and PCR was conducted recruiting the isolated circular plasmids as template. Two PCR runs were performed in the case of each plasmid isolated: 1st run with specific forward and reverse primers of promoters (Table 4.12), and 2nd run with the specific forward primers of the promoters and *AOX* reverse primer. The expected sizes of the PCR bands have been presented in Table 4.19. The 1st run PCR conditions have been presented in Table 4.20 and 4.21.

Table 4.19. Expected length of the bands in verification PCR of the putative transformants related with promoter constructs.

Intended Transformant	1st Run PCR (bp)	2nd Run PCR (bp)
P _{PDC} -containing transformants	831	1610
P _{THI3} -containing transformants	743	1522
P _{PYRK} -containing transformants	640	1409

Table 4.20. 1st run PCR thermal profile for verification of pPDCZαA::hGH transformants.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	48 °C	1 min.
29 cycles	72 °C	45 sec.
	94 °C	1 min.
29 cycles	51 °C	1 min.
	72 °C	45 sec.
Final extension	72 °C	10 min.

Table 4.21. 1st run PCR thermal profile for verification of pPYRKZαA::hGH and pTHI3ZαA::hGH transformants.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	51 °C	1 min.
29 cycles	72 °C	45 sec.
	94 °C	1 min.
29 cycles	53.5 °C	1 min.
	72 °C	45 sec.
Final extension	72 °C	10 min.

The results of the 1st run PCR can be found in Figure 4.30 and 4.31. After conducting 1st run of verification PCR, a length comparison was made between confirmed transformants plasmids by simultaneous loading of the samples of three promoters in one gel (Figure 4.32). All the obtained results referred to the proper insertion of the selected promoters and also proper length of the constructs.

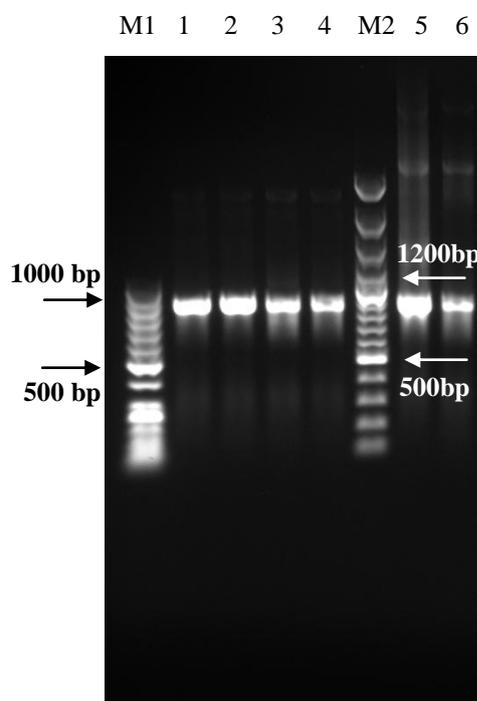


Figure 4.30. 1st run of the verification PCR specific to P_{PDC} -containing transformants. M1: GeneRuler™ 50bp DNA ladder, 1-4: Isolated plasmids of two selected colonies obtained with two different *Taq* buffers containing NH_4Cl or KCl , M2: GeneRuler™ 100bp plus DNA ladder, 5, 6: positive control using $\text{pUC57}::\text{PRSP}$ as template obtained with two different *Taq* buffers containing NH_4Cl or KCl .

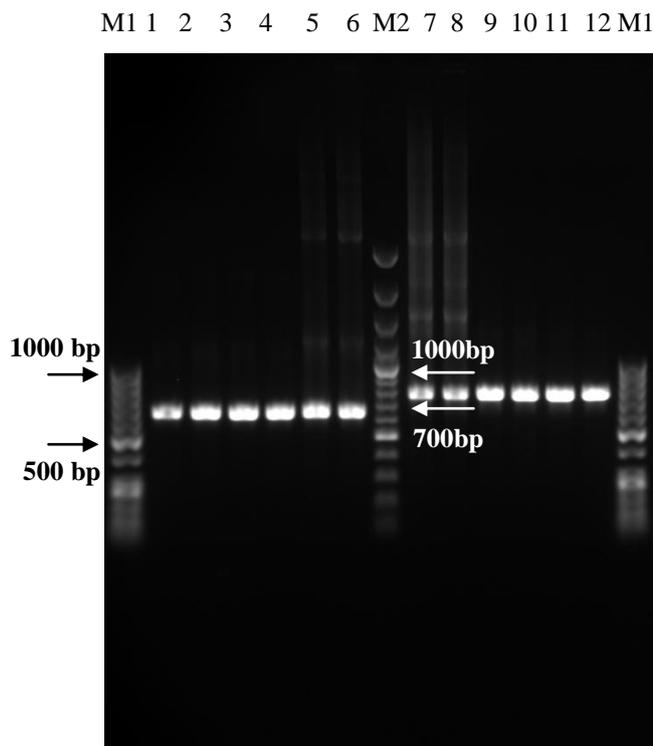


Figure 4.31. 1st run of the verification PCR specific to P_{PYRK} -containing and P_{THI3} -containing transformants. M1: GeneRuler™ 50bp DNA ladder, 1-4: Isolated plasmids of two selected colonies related with P_{PYRK} -containing transformants obtained with two different *Taq* buffers containing NH_4Cl or KCl , 5, 6: positive control using $pUC57::PRSP$ as template obtained with two different *Taq* buffers containing NH_4Cl or KCl with specific forward and reverse primers of P_{PYRK} , M2: GeneRuler™ 100bp plus DNA ladder, 7, 8: positive control using $pUC57::PRSP$ as template obtained with two different *Taq* buffers containing NH_4Cl or KCl with specific forward and reverse primers of P_{THI3} , 9-12: Isolated plasmids of two selected colonies related with P_{THI3} -containing transformants obtained with two different *Taq* buffers containing NH_4Cl or KCl .

The 2nd run of the verification PCR was accomplished in order to check the presence of *hGH* gene. 2nd run PCR conditions are available in Table 4.22. The results of the 2nd run of the verification PCR were subjected to gel electrophoresis which can be seen in Figure 4.33.

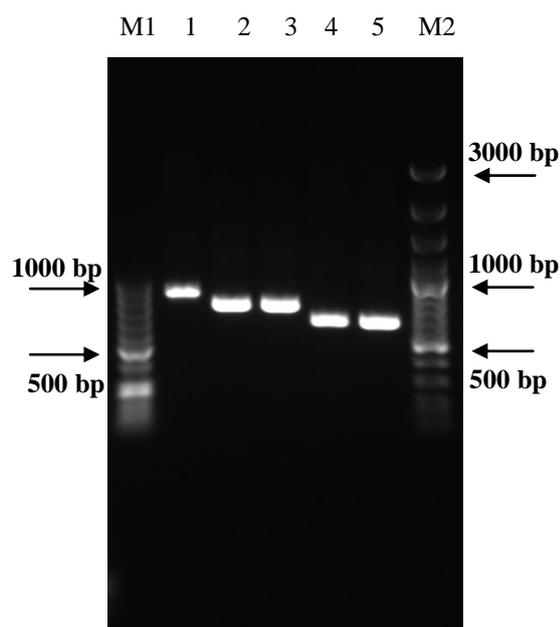


Figure 4.32. Comparison of the results of the 1st run of the verification PCR related with three promoter constructs. M1: GeneRuler™ 50bp DNA ladder, 1: P_{PDC} construct (831 bp), 2, 3: P_{THI3} construct (743 bp), 4, 5: P_{PYRK} construct (640 bp), M2: GeneRuler™ 100bp plus DNA ladder.

Table 4.22. 2nd run PCR thermal profile for verification of pPDCZαA::hGH, pPYRKZαA::hGH, and pTHI3ZαA::hGH plasmids as template utilizing specific forward primer of each promoter along with AOX reverse primer.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	49 °C	1 min.
29 cycles	72 °C	1 min.
	94 °C	1 min.
	52 °C	1 min.
Final extension	72 °C	1 min.
		10 min.

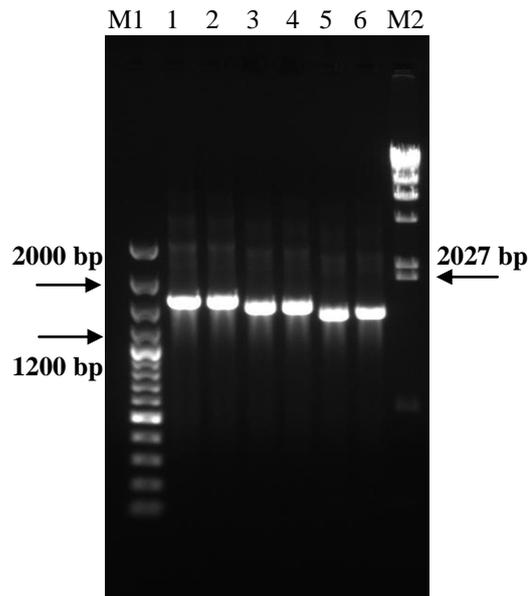


Figure 4.33. 2nd run PCR for promoter constructs verification. M1: GeneRuler™ 100bp plus DNA ladder, 1, 2: Putative P_{PDC} -harboring plasmid as template (expected size 1610 bp), 3, 4: Putative P_{THI3} -harboring plasmid as template (expected size 1522 bp), 5, 6: Putative P_{PYRK} -harboring plasmid as template (expected size 1409 bp), M2: DNA marker λ DNA/HindIII.

Based on the conducted triple controls, it was realized that the promoters have been inserted in the right place in BP. Therefore, the isolated plasmids of the selected two colonies for each of the promoter constructs were sequenced in METU central laboratory. All the isolated plasmids from selected putative transformant colonies possessed the desired nucleotide sequence based on the conducted BLAST analyses. Therefore, Microbank™ stocks were prepared from corresponding *r-E. coli* strains and were stored at -80°C for utilization in *Pichia* transfection. The results of the sequencing of the PDC-harboring plasmid (pPDCZ α A::*hGH*) can be found in Appendix K as an example.

4.5.3. Development of the new plasmids with selected endogenous SPs

In order to obtain plasmids that harbor selected endogenous SPs, substitution of α -MF in the BP with selected SPs (SP13, SP23, SP24, SP26, SP34) was conducted by double-digestion of the BP and prepared SPs constructs in section 4.5.1.2 with *Bsp*119I and *Eco*RI restriction enzymes and subsequent ligation of the digestion results.

4.5.3.1. Preparation of the desired SP-less backbone from BP

BP isolate obtained from corresponding r-*E. coli* strain was subjected to double-digestion with *Bsp*119I and *Eco*RI restriction enzymes. The result of enzymatic digestion was run on the agarose gel and, then, was subjected to the gel extraction to obtain purified SP-less backbone (Figure 4.34). Finally, a backbone was generated from BP which its SP (α -MF) had been removed and, therefore, it became ready for ligation with selected endogenous SP.

4.5.3.2. Preparation of the selected endogenous signal peptides

Since selected SPs constructs were amplified directly from pUC57::*PRSP*, the resultant fragments were intact and need to be digested in order to get compatible ends for ligation with prepared backbone from BP. Double-digestion of the amplified SPs was also conducted by *Bsp*119I and *Eco*RI restriction enzymes. After double-digestion, the digestion result was utilized directly in subsequent ligation reaction without any kind of purification.

4.5.3.3. Ligation of the SP-less backbone and selected SPs

After trying different molar ratios of insert (SP) and backbone (double-digested SP-less BP), molar ratio between 6 to 10 led to desirable ligation. The ligation reaction was accomplished between prepared backbone and double-digested endogenous SPs as explained earlier by T4 DNA ligase. The ligation mixture for development of the desired plasmids can be observed in Table 4.23.

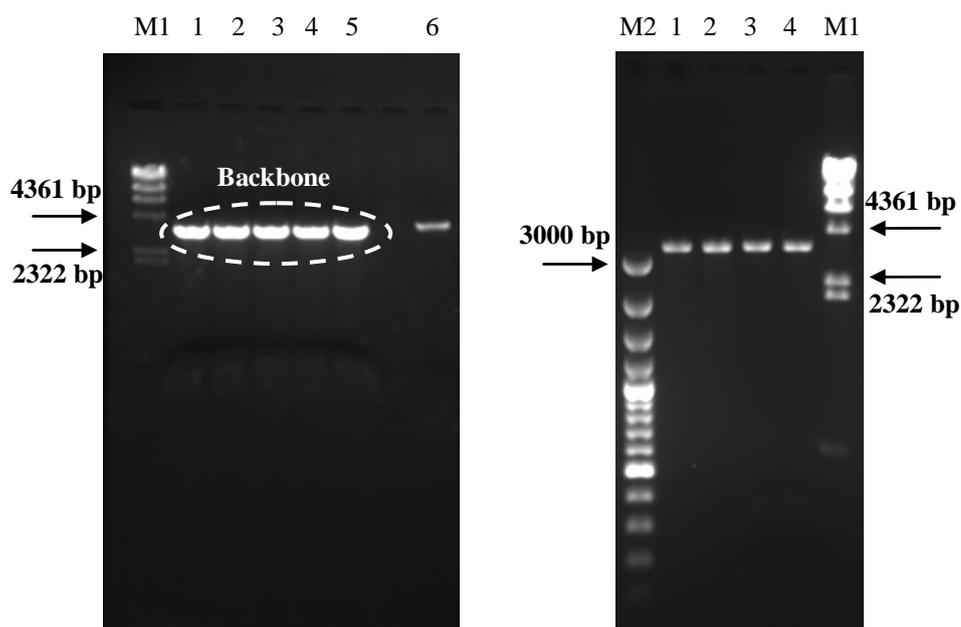


Figure 4.34. Double-digestion of the BP with *Bsp119I* and *EcoRI* restriction enzymes. The expected size of the generated backbone is ~ 3400 bp. Left-hand image - M1: DNA marker λ DNA/HindIII, 1-5: double digested BP samples, 6: single-digested (with *Bsp119I*) BP for validation of double-digestion. Right-hand image - M2: GeneRuler™ 100bp plus DNA ladder, 1-4: double-digested BP after gel extraction, M1: DNA marker λ DNA/HindIII.

Table 4.23. Ligation mixture for development of the plasmids harboring desired SPs.

Component	Amount
10X ligation buffer	2 μ L
Gene of interest (SP)	16-40 ng
Backbone plasmid (Double-digested BP)	200 ng
T4 DNA ligase	0.4 μ L (2U)
Autoclaved and filter-sterilized dH ₂ O	Variable
Total volume	20 μL

4.5.3.4. Transformation of the *E. coli* cells with developed plasmids

Transformation of the single colony *E. coli* DH5 α cells was performed using ligation results from previous section according to the CaCl₂ method described before. After transformation, for all five endogenous SPs there were putative transformants observed on selective LSLB agar plates containing ZeocinTM. In the case of each SP, 12 colonies were chosen and, subsequently, were inoculated into new selective LSLB agar plates containing ZeocinTM in order to prepare single-colony cultures. After 16-24 hours incubation at 37°C, six of the grown single-colony cultures (except SP13 that 12 single-colony cultures were used) were used for inoculation of the LB broth in order to prepare plasmid isolates of each colony by plasmid isolation kit for further analyses. Verification of the isolated plasmids of putative transformants was again achieved by gel electrophoresis of the circular isolated plasmids, double-digestion of the isolated plasmids with *Bsp119I* and *XbaI* restriction enzymes, and PCR; starting with higher number of colonies in agarose gel electrophoresis of circular plasmid isolates (for length control), toward double-digestion and PCR, the number of the selected colonies to be verified was decreased. The expected sizes of the developed plasmids can be found in Table 4.24. The length verification with gel electrophoresis can be observed in Figures 4.35, 4.36, and 4.37.

Table 4.24. Expected size of the developed plasmids harboring selected endogenous SPs.

Developed Plasmids Harboring Selected SPs	Expected Size (bp)
pGAPZ13A:: <i>hGH</i>	3489
pGAPZ23A:: <i>hGH</i>	3477
pGAPZ24A:: <i>hGH</i>	3489
pGAPZ26A:: <i>hGH</i>	3483
pGAPZ34A:: <i>hGH</i>	3480

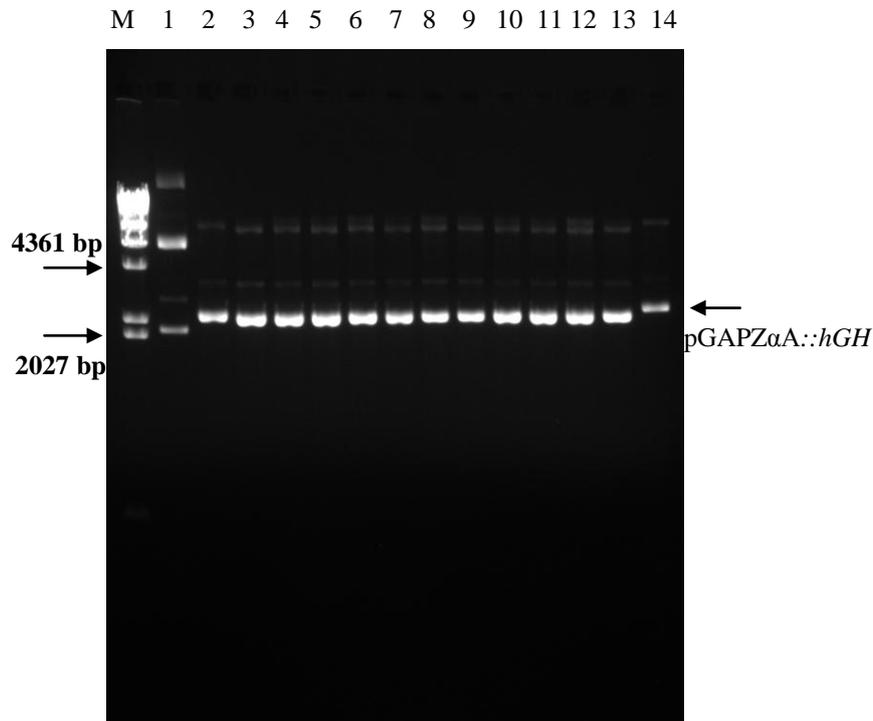


Figure 4.35 Isolated circular plasmids of twelve putative transformants of pGAPZ13A::hGH. M: DNA marker λ DNA/HindIII, 1: circular pGAPZ α A (3147 bp), 2-13: isolated plasmids of putative transformants, 14: circular pGAPZ α A::hGH (3696 bp).

The second step of the verification was conducted by double-digestion of the isolated plasmids with *Bsp*119I and *Xba*I in order to check cleavage and release of the SP and *hGH* gene together. The double-digestion mixture can be found in Table 4.25. The digestion results were subjected to the agarose gel electrophoresis and the results have been represented in Figure 4.38 and 4.39. In the case of pGAPZ13A::hGH, eight colonies were subjected to double-digestion but in other four cases, three colonies were used.

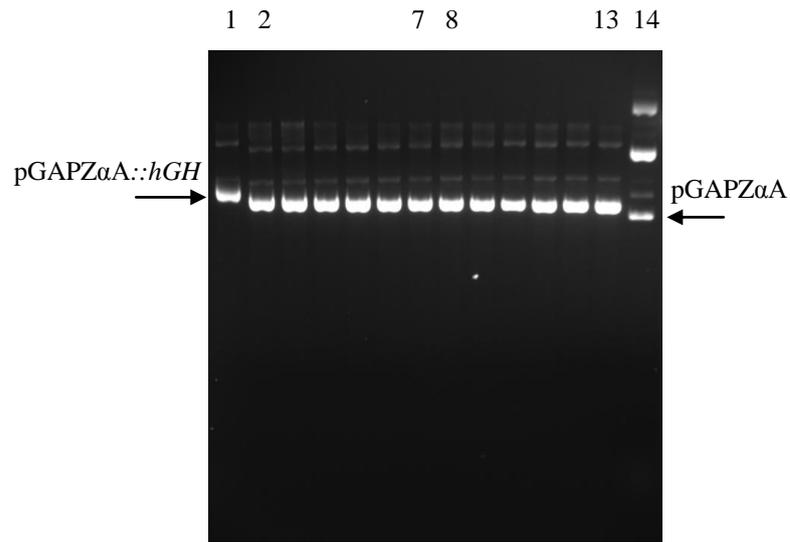


Figure 4.36. Isolated circular plasmids of six putative transformants of pGAPZ23A::hGH along with six putative transformants of pGAPZ24A::hGH, 1: circular pGAPZαA::hGH (3696 bp), 2-7: isolated plasmids of transformants of pGAPZ23A::hGH, 8-13: isolated plasmids of putative transformants of pGAPZ24A::hGH, 14: circular pGAPZαA (3147 bp).

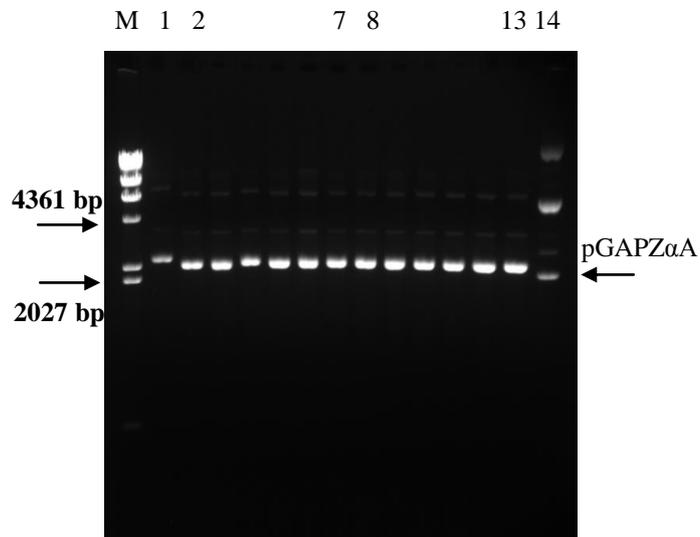


Figure 4.37. Isolated circular plasmids of six putative transformants of pGAPZ26A::hGH along with six putative transformants of pGAPZ34A::hGH, M: DNA marker λ DNA/HindIII, 1: circular pGAPZαA::hGH (3696 bp), 2-7: isolated plasmids of putative transformants of pGAPZ26A::hGH, 8-13: isolated plasmids of putative transformants of pGAPZ34A::hGH, 14: circular pGAPZαA (3147 bp).

Table 4.25. Double-digestion mixture for verification of the SP-harboring plasmids.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable
10X Tango buffer	2 μ L
Isolated plasmid	Variable (up to 1 μ g)
<i>Xba</i> I	1 μ L
<i>Bsp</i> 119I	1 μ L
Total volume	20μL

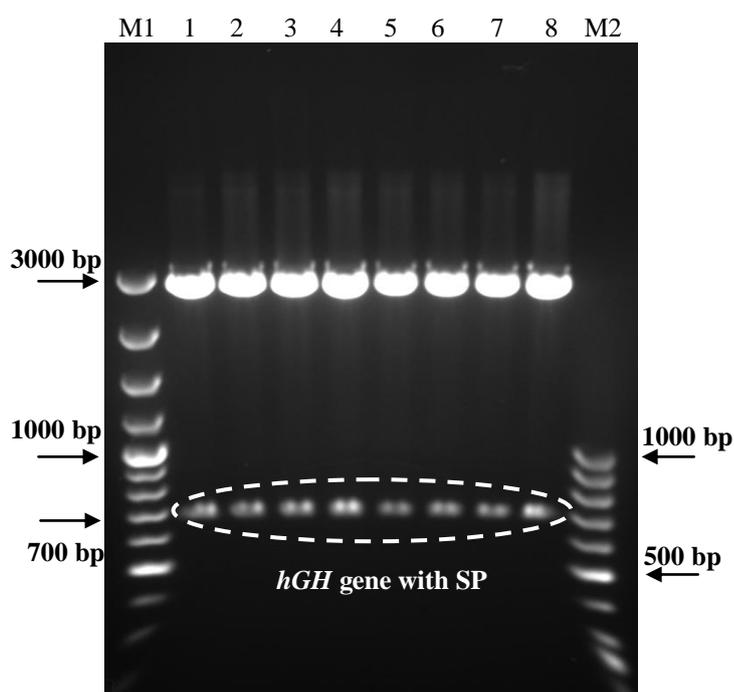


Figure 4.38. Double-digested (with *Bsp*119I & *Xba*I) plasmids of selected putative transformants (8 colonies) related to pGAPZ13::*hGH*. The small separated fragment includes SP and *hGH* gene (~ 700 bp). M1:GeneRuler™ 100bp plus DNA ladder, 1-8: isolated plasmids of 8 putative transformants, M2: GeneRuler™ 50bp DNA ladder.

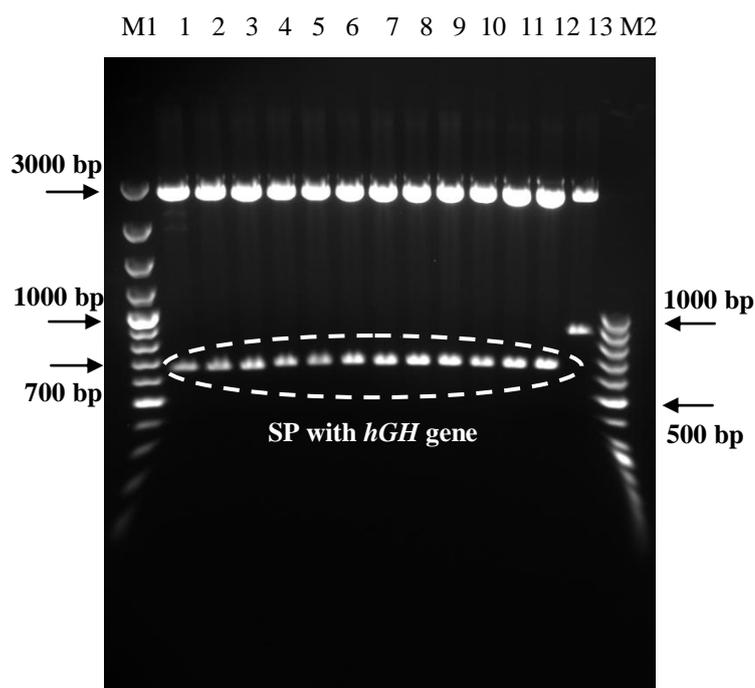


Figure 4.39. Double-digested (with *Bsp119I* & *XbaI*) plasmids of selected putative transformants related to pGAPZ23::*hGH*, pGAPZ24::*hGH*, pGAPZ26::*hGH*, and pGAPZ34::*hGH* (3 colonies for each plasmid). The small separated fragment includes SP and *hGH* gene. M1:GeneRuler™ 100bp plus DNA ladder, 1-3: isolated plasmids of 3 putative transformants of pGAPZ23::*hGH*, , 4-6: isolated plasmids of 3 putative transformants of pGAPZ24::*hGH*, , 7-9: isolated plasmids of 3 putative transformants of pGAPZ26::*hGH*, , 10-12: isolated plasmids of 3 putative transformants of pGAPZ34::*hGH*, 13: Double-digested BP (pGAPZ α A::*hGH*) as positive control (the released fragment ~ 900 bp), M2: GeneRuler™ 50bp DNA ladder.

The third round of the verification was PCR with *GAP* forward and *AOX* reverse primers using the isolated plasmids of putative transformants as template. The PCR mixture and the PCR thermal profile has been presented in Table 4.1 and 4.2, respectively. In the case of pGAPZ13::*hGH*, eight colonies were used in PCR (Figure 4.40) but in four other SPs, two colonies were used in PCR (Figure 4.41).

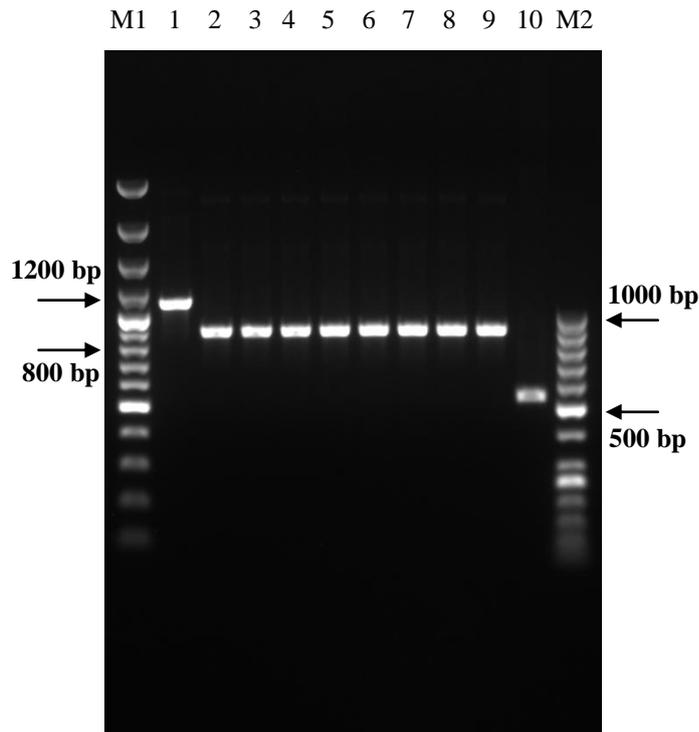


Figure 4.40. Verification PCR related to putative transformants of pGAPZ13A::*hGH*. The expected size of the PCR band is ~ 900 bp. M1: GeneRuler™ 100bp plus DNA ladder, 1: BP as positive control (expected size ~1080 bp), 2-9: PCR results of isolated plasmids of putative transformants, 10: pGAPZ α A as positive control (expected size ~550 bp), M2: GeneRuler™ 50bp DNA ladder.

After performing verification, in the case of each SP, two verified colonies were sent to METU central laboratory in order to be sequenced. Based on the available expected sequences of the developed plasmids, NCBI BLAST was recruited to check the results of the sequencing. All the isolated plasmids from selected putative transformant colonies possessed the desired nucleotide sequence and, therefore, Microbank™ stocks were prepared from the corresponding *r-E. coli* strains and were stored at -80°C for utilization in next steps. The sequencing results related with pGAPZ13A::*hGH* has been presented in Appendix K as example.

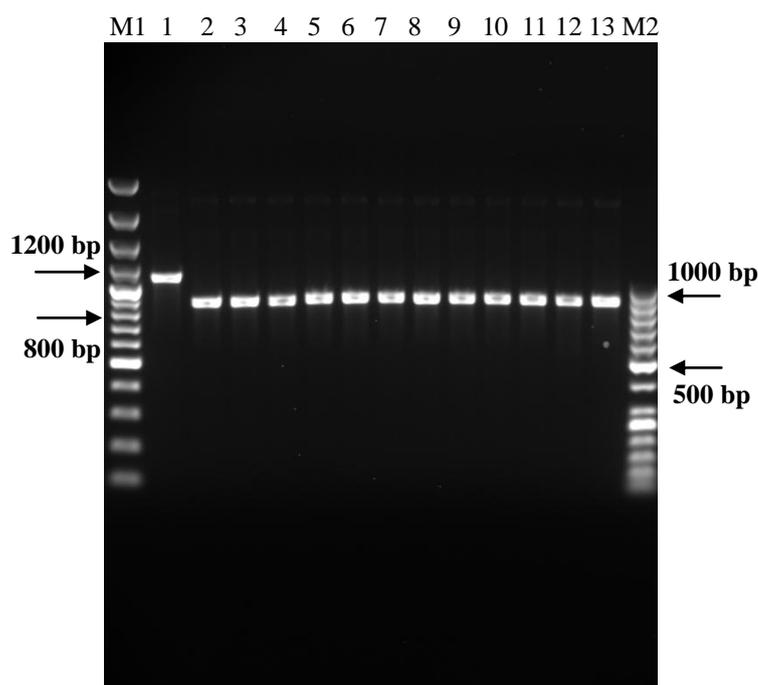


Figure 4.41. Verification PCR related to putative transformants of pGAPZ23::*hGH*, pGAPZ24::*hGH*, pGAPZ26::*hGH*, and pGAPZ34::*hGH* (3 colonies for each plasmid). The expected size of the PCR band is ~ 900 bp. M1: GeneRuler™ 100bp plus DNA ladder, 1: BP as positive control (expected size ~1080 bp), 2-4: PCR results related to pGAPZ23::*hGH*, 5-7: PCR results related to pGAPZ24::*hGH*, 8-10: PCR results related to pGAPZ26::*hGH*, 11-13: PCR results related to pGAPZ34::*hGH*, M2: GeneRuler™ 50bp DNA ladder.

4.6. Developing recombinant *P. pastoris* cells

After developing BP and eight new plasmids and subsequent preparation of the required r-*E. coli* strains possessing newly-developed plasmids harboring promoters and SPs of interest, the cloned plasmids were utilized in order to transfect *P. pastoris* X-33 cells via genomic integration. The developed plasmids were first linearized with *Nsi*I restriction enzyme and, then, were purified and used for transfection.

4.6.1. Single-digestion of the developed plasmids

Single-colony cultures of the developed r-*E. coli* strains were prepared and plasmid isolation was performed. The purified plasmids were single-digested with *NsiI* enzyme. The digestion mixture for plasmids harboring selected promoters and SPs can be found in Tables 4.26 and 4.27. After digestion, the digestion reaction result was purified with PCR purification kit where at the final elution step instead of the elution buffer, autoclaved and filter-sterilized ultra-pure water was utilized for eluting DNA fragments. After elution of the desired linearized plasmids, the elution results were loaded in the agarose gel in order to verify the linearization; the typical results of the gel electrophoresis can be observed in Appendix P.

Table 4.26. Digestion mixture for plasmids harboring desired endogenous SPs.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	18 µL
R buffer 10X	30 µL
DNA (plasmid)	250 µL (regarding plasmid amount)
<i>NsiI</i>	2 µL
Total volume	300 µL

Table 4.27. Digestion mixture for BP and plasmids harboring selected promoters.

Component	Amount
Autoclaved and filter-sterilized dH ₂ O	Variable
R buffer 10X	25 µL
DNA (plasmid)	180-200 µL (regarding plasmid amount)
<i>NsiI</i>	2 µL
Total volume	250µL

According to the gel electrophoresis results, digestion was not complete (in some cases very poor digestion); however, the experiments proceeded by considering the possibility of the transfection either in circular form of the transfecting plasmid. Therefore, the eluted and verified linearized plasmid samples were utilized in *Pichia* transfection.

4.6.2. Transfection of *P. pastoris* X-33 cells with linearized new plasmids

Linearized plasmids were used in transfection of the *P. pastoris* X-33 strains according to LiCl method. Transfection results were checked after approximately 60 hours. 18 colonies were selected for each developed plasmid. Single-colony cultures of these selected colonies were prepared and after 24 hours incubation in 30°C in YPD agar plate under Zeocin[™] pressure (100µg/mL) the grown colonies among them were used for genomic DNA isolation and further verification analyses.

Putative transfectants were used to verify the insertion of the desired plasmid in the genome of the *P. pastoris*. Therefore, six colonies out of 18 colonies in each case were selected and were subjected to genomic DNA isolation as described before. The results of the genomic DNA isolation of the putative transfectants peculiar to pGAPZ13A::*hGH* have been represented in Figure 4.42 which is a typical image and all the other genome isolates were similar to it.

Verification of the integration to the genome was achieved by PCR where the isolated genomes were used as template. In the case of plasmids harboring endogenous SPs, recruited primers were *GAP* forward and *AOX* reverse. However, in the case of plasmids harboring selected promoters, the utilized primers were specific corresponding forward primer of each promoter along with *AOX* reverse primer.

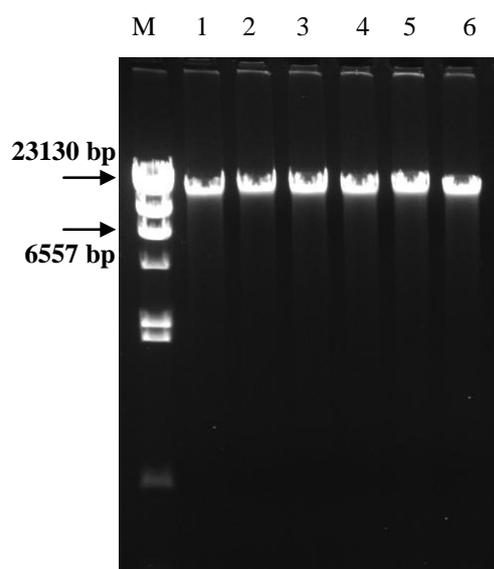


Figure 4.42. Genomic DNA isolates of 6 putative transfectants of pGAPZ13A::*hGH*. M: DNA marker λ DNA/HindIII, 1-6: Genomic DNA isolates of 6 selected colonies.

The composition of the components in the verification PCR of the plasmids related with endogenous SPs can be observed in Table 4.28.

Table 4.28. PCR mixture for verification of the genomic integration of the plasmids possessing selected endogenous SPs. PCR thermal profile is available in Table 4.2.

Component	Amount
10X amplification buffer with KCl	5 μ L
2 mM solution of dNTPs	5 μ L
5 μ M <i>GAP</i> -Forward primer	2 μ L
5 μ M <i>AOX</i> -Reverse primer	2 μ L
25 mM MgCl ₂ / MgSO ₄	3 μ L
<i>Taq</i> DNA polymerase (5 U/ μ L)	1-2 U
dH ₂ O	32
Template DNA (Genomic DNA or plasmid DNA)	1 μ L
Total volume	50 μL

The PCR results, then, were subjected to the agarose gel electrophoresis. For genome isolates of pGAPZ13A::*hGH* putative transformants the results of the verification PCR have been represented in Figure 4.43 as a typical image; since the length of the developed plasmids is very close to each other in the case of SP-harboring plasmids, the other verification PCRs led to similar gel images upon genomic integration; so, they were not presented. In the case of false positive colonies, the wrong colonies were replaced by new ones in order to have at least six confirmed colonies in each case.

By conducting similar verification PCR for other plasmids harboring selected SPs, approved colonies were determined and the genomes were stored for copy number determination experiments.

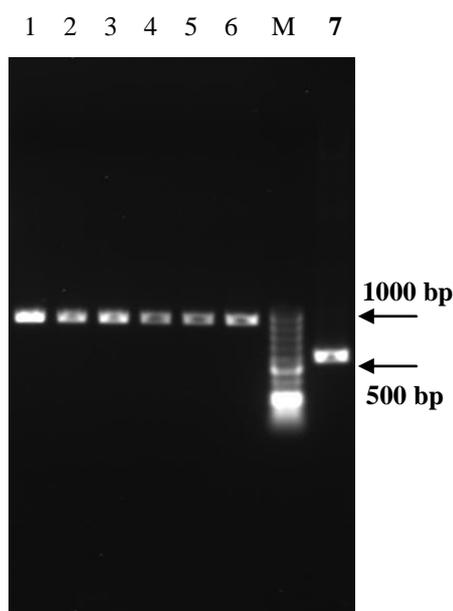


Figure 4.43. Results of the verification PCR to confirm genomic integration related with pGAPZ13::*hGH* transfectants. Expected size of the desired PCR band is ~900 bp. 1-6: PCR bands obtained from isolated genomes, M: GeneRuler™ 50bp DNA ladder, 7: PCR band obtained by pGAPZαA plasmid as template, as positive control.

Related with the plasmids harboring selected promoters, the verification PCR was a little different in terms of the utilized primers. The PCR mixture and the PCR conditions have been represented in Tables 4.29 and 4.30, respectively.

Table 4.29. PCR mixture for verification of the genomic integration of the plasmids possessing selected promoters.

Component	Amount
10X <i>Taq</i> buffer without MgCl ₂	5 µL
2 mM solution of dNTPs	5 µL
5 µM promoter-specific forward primer	2 µL
5 µM <i>AOX</i> reverse primer	2 µL
25 mM MgCl ₂	3 µL
<i>Taq</i> DNA polymerase	1-2 U
dH ₂ O	31 µL
Template DNA (genome)	1 µL (up to 1 µg)
Total volume	50 µL

Table 4.30. Thermal profile of the PCR for verifying the insertion of the plasmids harboring selected promoters in *Pichia* genome.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	49 °C	1 min.
29 cycles	72 °C	1 min.
	94 °C	1 min.
	52 °C	1 min.
Final extension	72 °C	1 min.
	72 °C	10 min.

The corresponding results of the verification PCR have been shown in Figures 4.44 and 4.45. As it is obvious from the results, there are false positive colonies which were replaced with new ones to complete the number of the approved transfectants to six colonies.

After verification of the genomic insertion of the developed plasmids (including BP), at least six colonies were available in the case of each newly-developed plasmid which were utilized in order to find single-copy *P. pastoris* transfectant for each plasmid.

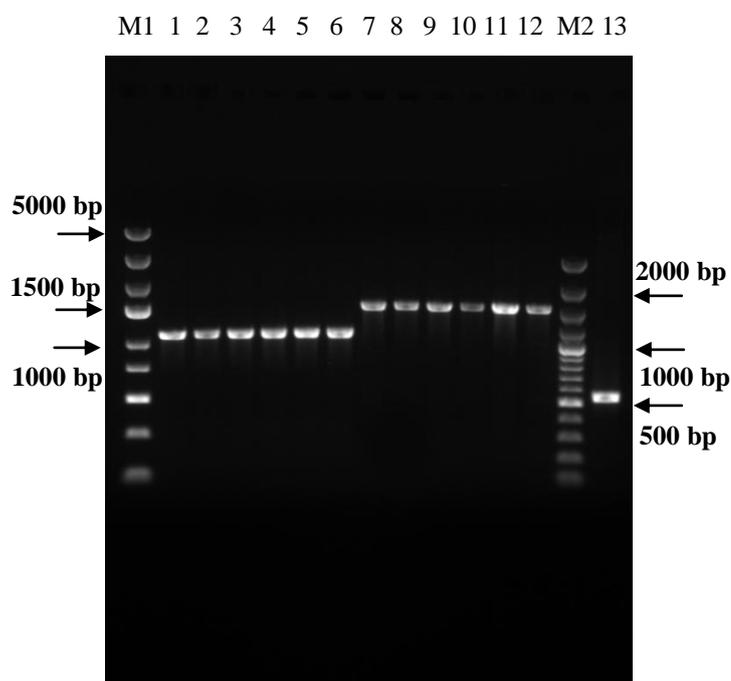


Figure 4.44. Results of the verification PCR to confirm genomic integration related with pGAPZ α A::hGH and pTHI3Z α A::hGH transfectants. M1: GeneRuler™ express DNA ladder, 1-6: PCR bands obtained from isolated genomes of putative transformants specific to pGAPZ α A::hGH with expected size of 1090 bp, 7-12: PCR bands obtained from isolated genomes of putative transformants specific to pTHI3Z α A::hGH with expected size of 1522 bp, M2: GeneRuler™ 100 bp plus DNA ladder, 13: PCR band obtained by pGAPZ α A plasmid as template, as positive control.

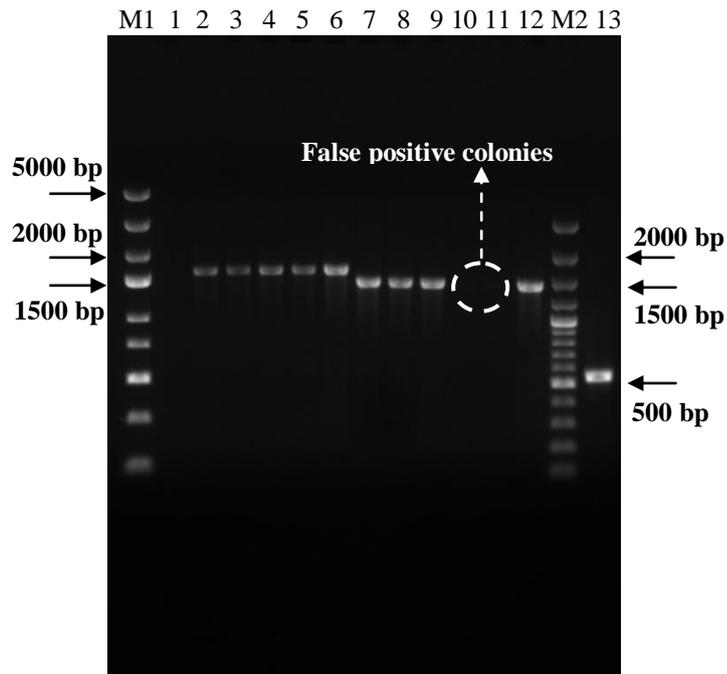


Figure 4.45. Results of the verification PCR to confirm genomic integration related with pPDCZ α A::hGH and pPYRKZ α A::hGH transfectants. M1: GeneRuler™ express DNA ladder, 1-6: PCR bands obtained from isolated genomes of putative transformants specific to pPDCZ α A::hGH with expected size of 1610 bp, 7-12: PCR bands obtained from isolated genomes of putative transformants specific to pPYRKZ α A::hGH with expected size of 1409 bp, M2: GeneRuler™ 100 bp plus DNA ladder, 13: PCR band obtained by pGAPZ α A plasmid as template, as positive control.

4.7. Pre-screening of the selected *P. pastoris* transfectants

After verification of the genomic integration in (at least) six transformant colonies of each of the developed plasmids, a pre-screening step was conducted according to section 3.6.14.1 in order to determine some of the probable multi-copy strains of the developed *r-P. pastoris*. The PCR mixture and the the PCR conditions are available in Table 4.31 and 4.32, respectively. The results of the pre-screening PCR was subjected to agarose gel electrophoresis which have been presented in Figures 4.46 to 4.48 and Figures 4.50 to 4.53. According to these figures, it was realized that, except a few colonies, pallid bands also appeared along with bright bands in gel

images. It was supposed that the bright bands simply refer to multi-copy integrations. However, since the single-digestion of the plasmids with *NsiI* enzyme prior to *Pichia* transfection was not complete, the pallid bands might be due to be the results of the circular plasmids that had remained after transfection and, subsequently, were isolated along with genomic DNA.

Table 4.31. PCR mixture for the pre-screening experiments.

Component	Amount
10X Taq buffer with KCl - MgCl ₂	5 µL
2 mM solution of dNTPs	5 µL
5 µM <i>pUC ori</i> -F primer	3 µL
5 µM <i>hGH-R</i> primer	3 µL
25 mM MgCl ₂	3 µL
<i>Taq</i> DNA polymerase (5 U/ µL)	1-2 U
dH ₂ O	29
Template DNA (Genomic DNA or plasmid DNA)	1.5 µL
Total volume	50 µL

Table 4.32. PCR thermal profile for the pre-screening experiments.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	54 °C	80 sec.
29 cycles	72 °C	4 min.
	94 °C	1 min.
	59 °C	80 sec.
Final extension	72 °C	4 min.
	72 °C	10 min.

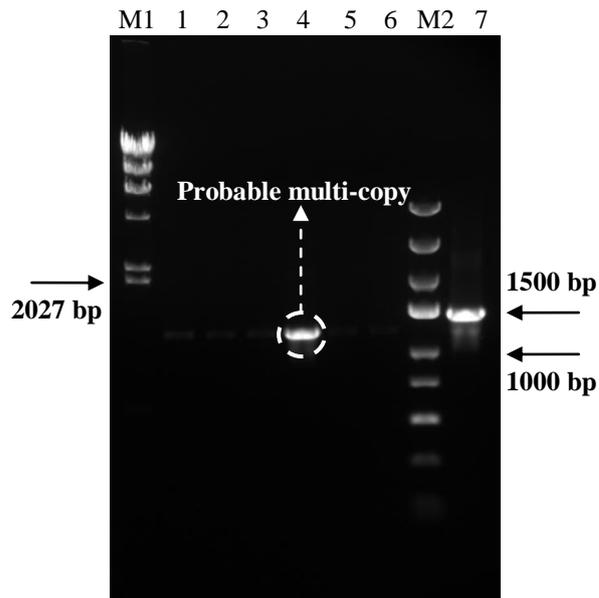


Figure 4.46. Pre-screening results of the verified transfectants of pGAPZ13A::*hGH*. M1: DNA marker λ DNA/HindIII, 1-6: corresponding results of the colonies number 2, 3, 4, 6, 17, 18, respectively, M2: GeneRuler™ express DNA ladder, 7: pGAPZ α A::*hGH* as template in pre-screening PCR, positive control with expected size of 1420 bp.

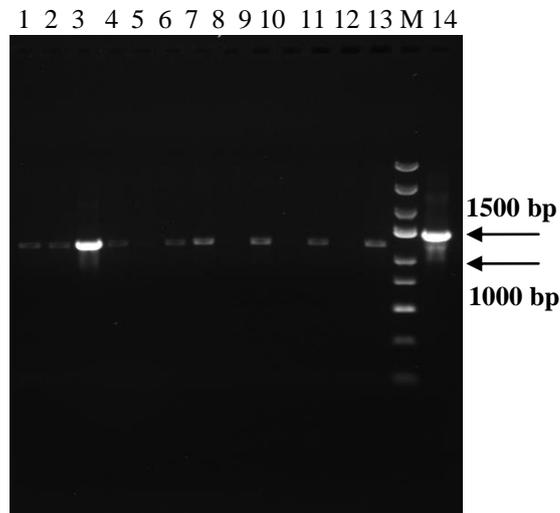


Figure 4.47. Pre-screening results of the verified transfectants of pGAPZ23A::*hGH* and pGAPZ24A::*hGH*. 1-6: corresponding results of the colonies number 22, 23, 24, 32, 33, 34, respectively (for pGAPZ23A::*hGH*), 7-13: corresponding results of the colonies number 40, 48, 49, 50, 51, 53, 54, respectively (for pGAPZ24A::*hGH*), M: GeneRuler™ express DNA ladder, 14: pGAPZ α A::*hGH* as template in pre-screening PCR, positive control with expected size of 1420 bp.

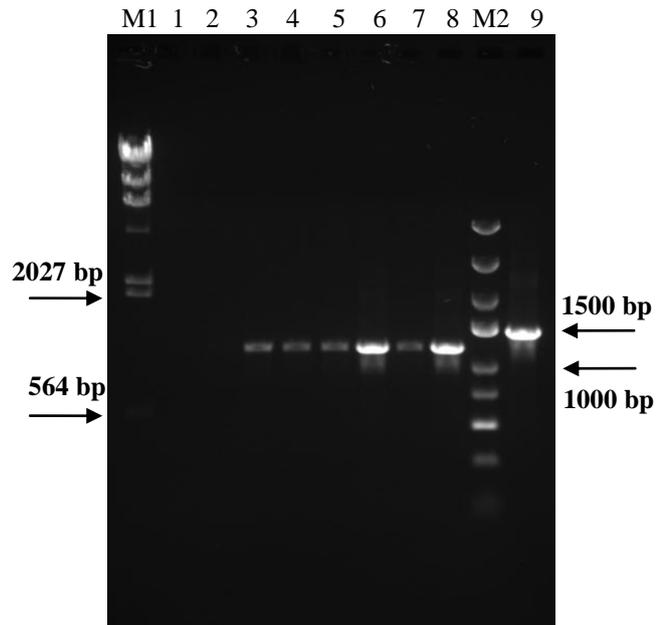


Figure 4.48. Pre-screening results of the verified transfectants of pGAPZ26A::hGH. M1: DNA marker λ DNA/HindIII, 1-8: corresponding results of the colonies number 61, 64, 65, 66, 68, 70, 71, 72, respectively, M2: GeneRuler™ express DNA ladder, 9: pGAPZ α A::hGH as template in pre-screening PCR, positive control with expected size of 1420 bp.

In order to have a sense about the presence of pallid bands, a single-digestion was made with *EcoRI* enzyme on pGAPZ34A::hGH and pGAPZ23::hGH. According to the digestion results it appears that there is no circular plasmid in the single-digestion reaction result (Figure 4.49); namely, the circular plasmid has been completely digested. Since it is not expected any band by the single-digested plasmid (free of circular plasmid), this mixture was used as negative control for pre-screening PCR experiments

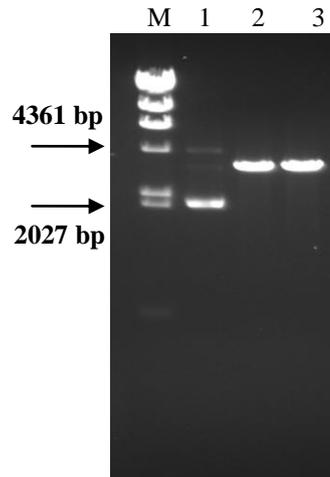


Figure 4.49. Single-digestion (by *EcoRI* restriction enzyme) results of pGAPZ23A::*hGH* (3477 bp) and pGAPZ34A::*hGH* (3480 bp). M: DNA marker λ DNA/HindIII, 1,2: circular and single-digested pGAPZ34A::*hGH*, respectively, 3: single-digested pGAPZ23A::*hGH*.

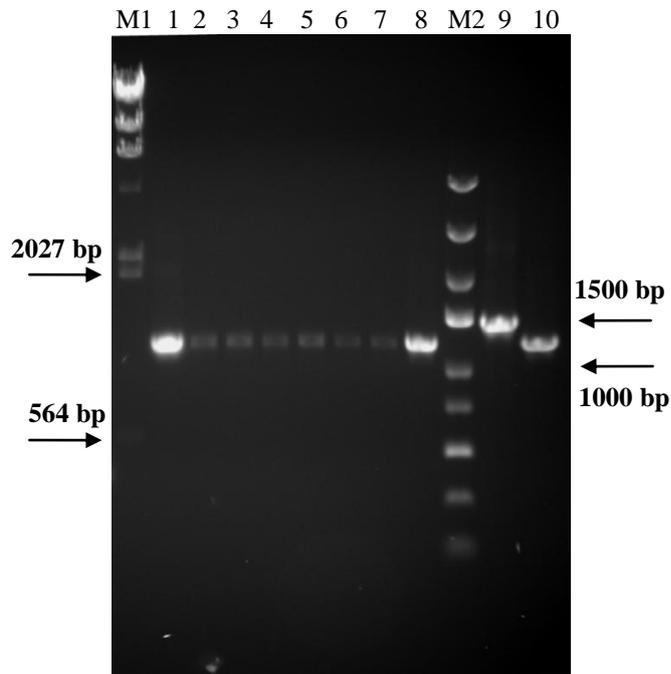


Figure 4.50. Pre-screening results of the verified transfectants of pGAPZ34A::*hGH*. M1: DNA marker λ DNA/HindIII, 1-8: corresponding results of the colonies number 74, 75, 76, 77, 78, 81, 83, 84, respectively, M2: GeneRuler™ express DNA ladder, 9: pGAPZ α A::*hGH*

as template in pre-screening PCR, positive control with expected size of 1420 bp, 10: single-digested pGAPZ34A::*hGH* as negative control.

Figure 4.50 implies that in spite of the absence of any detectable circular plasmid in gel image after single-digestion, PCR confirms presence of circular undigested plasmid even in very negligible quantity. Therefore, presence of pallid bands in the pre-screening results as the result of circular (undigested) plasmids, although they were not observed in genomic DNA isolate, is not so irrational.

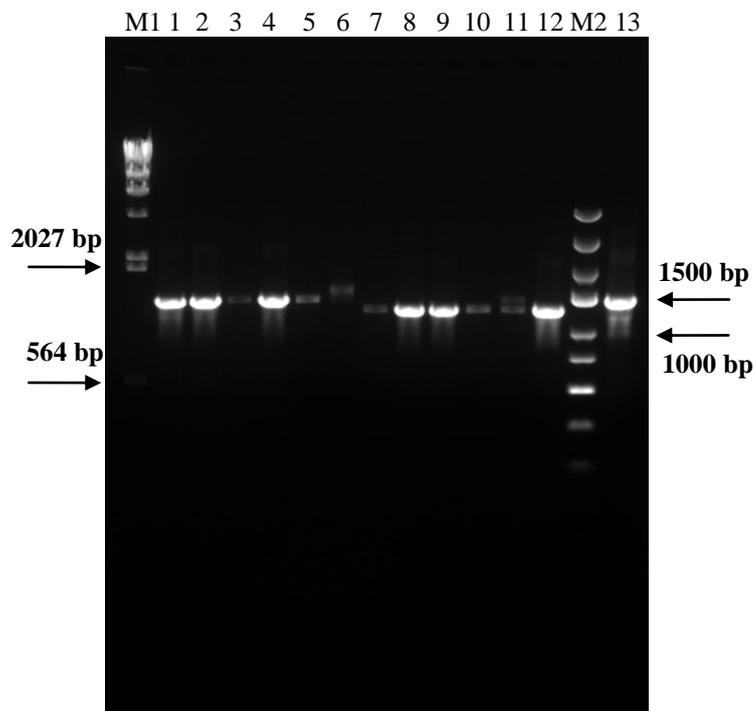


Figure 4.51. Pre-screening results of the verified transfectants of pGAPZ α A::*hGH* (BP) and pPYRKZ α A::*hGH*. M1: DNA marker λ DNA/HindIII, 1-6: corresponding results of the colonies number 107, 108, 109, 110, 111, 112, respectively (for pGAPZ α A::*hGH*), 7-12: corresponding results of the colonies number 144, 147, 149, 150, 151, 154, respectively (for pPYRKZ α A::*hGH*), M2: GeneRuler™ express DNA ladder, 13: pGAPZ α A::*hGH* as template in pre-screening PCR, positive control with expected size of 1420 bp.

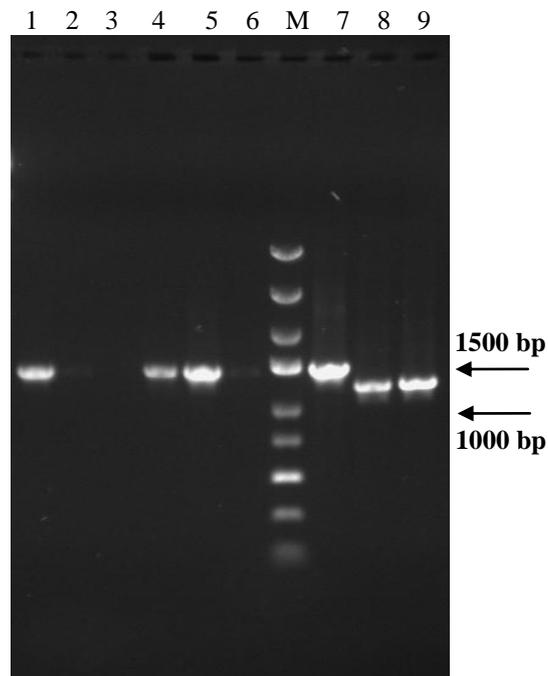


Figure 4.52. Pre-screening results of the verified transfectants of pTHI3Z α A::*hGH*. 1-6: corresponding results of the colonies number 159, 160, 163, 166, 171, 172, respectively, M: GeneRuler™ express DNA ladder, 7: pGAPZ α A::*hGH* as template in pre-screening PCR, positive control with expected size of 1420 bp, 8: single-digested pGAPZ23A::*hGH* as negative control, 9: single-digested pGAPZ34A::*hGH* as negative control.

In the case of pPDCZ α A::*hGH* verified transfectants, the precision of the pre-screening was checked more with utilization of another reverse primer in parallel with *hGH*-R; *Eda-qPCR*-R (Table 3.21) was used as another alternative reverse primer. The pallid bands were present again in the gel image (Figure 4.53) and, therefore, there was no problem related with primers used.

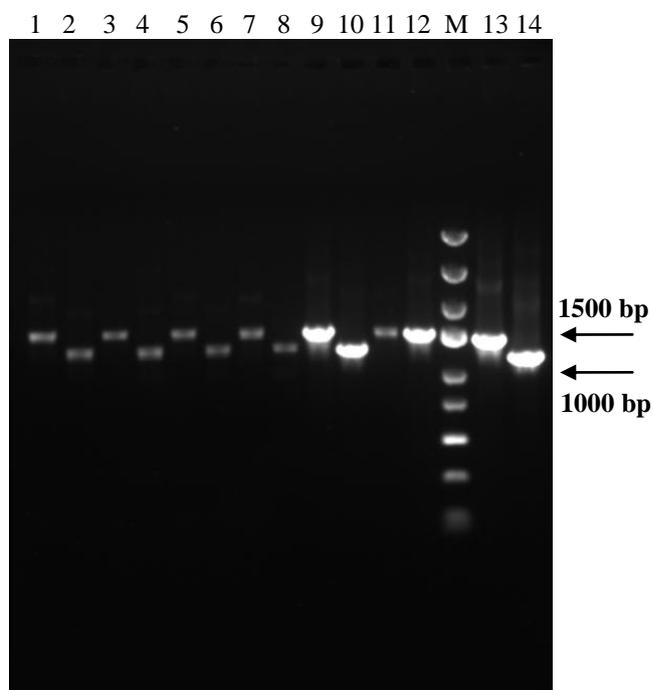


Figure 4.53. Pre-screening results of the verified transfectants of pPDCZ α A::*hGH*. In order to double-check the insertion, two different reverse primers including *hGH*-R (Table 3.18) or *Eda-qPCR*-R (Table 3.21) were used along with *pUC-Ori-F* as forward primer. 1, 3, 5, 7, 9, 11 and 12: corresponding results of the colonies number 120, 121, 122, 123, 124, 128, 130, respectively obtained with *hGH*-R, 2, 4, 6, 8, 10: corresponding results of the colonies number 120, 121, 122, 123, 124 obtained with *Eda-qPCR*-R, M: GeneRuler™ express DNA ladder, 13: pGAPZ α A::*hGH* as template in pre-screening PCR (with *hGH*-R), positive control with expected size of 1420 bp, 14: pGAPZ α A::*hGH* as template in pre-screening PCR (*Eda-qPCR*-R), positive control with expected size of 1220.

The obtained results from the conducted pre-screening PCRs have been summarized in Table 4.33. In the case of each developed plasmid, three probable single-copy transfectant strains were selected (Table 4.34) and genomic DNA isolation was conducted again in order to obtain isolates with better quality.

Table 4.33. Final pre-screening results of the verified *P. pastoris* transfectants.

Developed Plasmid	Single/Multi-copy	Attributed Colony Number
pPDCZ α A::hGH	Probable single-copy	120 – 121 – 122 – 123 - 128
	Multi-copy	124 - 130
pPYRKZ α A::hGH	Probable single-copy	144 – 150 - 151
	Multi-copy	143 – 147 – 148 – 149 - 154
pTHI3Z α A::hGH	Probable single-copy	160 – 163 - 172
	Multi-copy	159 - 166 - 171
pGAPZαA::hGH	Probable single-copy	109 – 111 - 112
	Multi-copy	107 – 108 - 110
pGAPZ13A::hGH	Probable single-copy	2 – 3 – 4 – 17 - 18
	Multi-copy	6
pGAPZ23A::hGH	Probable single-copy	22 – 23 – 32 – 33 - 34
	Multi-copy	24
pGAPZ24A::hGH	Probable single-copy	40 – 48 – 49 – 50 – 51 – 53 - 54
	Multi-copy	None!
pGAPZ26A::hGH	Probable single-copy	61 – 64 – 65 – 66 – 68 - 71
	Multi-copy	70 - 72
pGAPZ34A::hGH	Probable single-copy	75 – 76 – 77 – 78 – 81 - 83
	Multi-copy	74 - 84

The probability of the insertion of the plasmid in other site(s) than desired locus prevents us from strict decision-making about being single-copy in the cases that there were not any bands in PCR results gel images. The new isolated genomes were used in subsequent qPCR experiments for copy number determination.

Table 4.34. Selected three probable single-copy r-*P. pastoris* strains for each of the developed plasmids.

Developed Plasmid	Selected Colonies
pGAPZ13A:: <i>hGH</i>	4, 17, 18
pGAPZ23A:: <i>hGH</i>	32, 33, 34
pGAPZ24A:: <i>hGH</i>	48, 50, 53
pGAPZ26A:: <i>hGH</i>	61, 64, 65
pGAPZ34A:: <i>hGH</i>	75, 81, 83
pGAPZαA::<i>hGH</i>	109, 111, 112
pPDCZ α A:: <i>hGH</i>	120, 122, 123
pPYRKZ α A:: <i>hGH</i>	144, 150, 151
pTHI3Z α A:: <i>hGH</i>	160, 163, 172

4.8. *hGH* copy number determination in selected verified strains

The single-copy transfectants of developed *P. pastoris* strains were determined by conducting qPCR experiments as explained earlier in section 3.6.1.4.2. Regarding the requirements of the absolute quantification, standard samples for *ARG4* and *hGH* genes were prepared in order to have a standard curve for accomplishment of the qPCR experiment. The standard samples were prepared by considering the concept of the nested PCR.

4.8.1. Standard DNA sample preparation for *ARG4* gene

The outer primer pair was utilized for *ARG4* gene standard sample preparation was presented in Table 3.19. The optimum PCR mixture and PCR conditions have been illustrated in Tables 4.35 and 4.36, respectively.

Table 4.35. Optimum PCR mixture for preparation of the standard DNA for *ARG4* using *P. pastoris* genomic DNA as template.

Component	Amount
10X amplification buffer with KCl	2 μ L
2 mM solution of dNTPs	2 μ L
5 μ M <i>ARG4</i> -Std-F primer	1.2 μ L
5 μ M <i>ARG4</i> -Std-R primer	1.2 μ L
25 mM MgCl ₂ / MgSO ₄	0.8 μ L
<i>Taq</i> DNA polymerase (5 U/ μ L)	1-2 U
dH ₂ O	11 μ L
Template DNA (Genomic DNA or plasmid DNA)	2 μ L
Total volume	20 μL

Table 4.36. PCR thermal profile for preparation of the *ARG4* standard DNA.

Number of Cycles	Temperature	Time
1 cycle	94 °C	3 min.
1 cycle	94 °C	1 min.
	49 °C	1 min.
	72 °C	30 sec.
29 cycles	94 °C	1 min.
	51 °C	1 min.
	72 °C	15 sec.
Final extension	72 °C	30 min.

The result of the PCR has been represented in Figure 4.54. However, in order to have a pure sample free of any contaminating reagents, the resultant PCR fragments were purified with PCR purification kit. The purified fragment has been represented in Figure 4.55. After purification, the quality and the concentration of the prepared purified sample was checked; the average

concentration was measured as approximately 12.3 ng/ μ L which is equal to approximately 3.41×10^{10} copy/ μ L (equation 3.3).

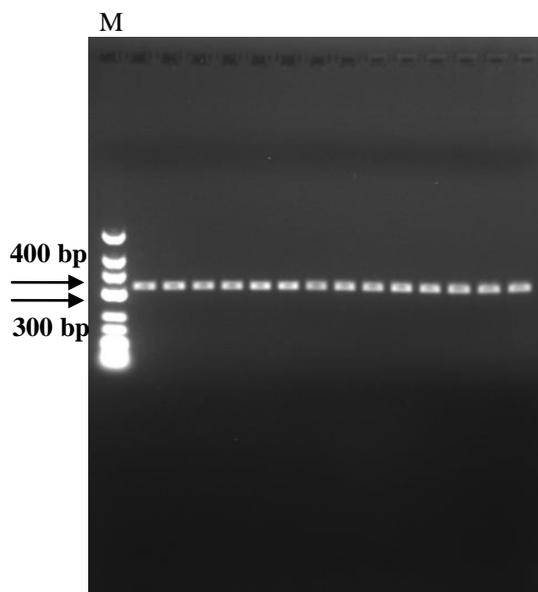


Figure 4.54. Preparation of the *ARG4* gene standard sample (expected length of 330 bp) by using *P. pastoris* genomic DNA as template. M: GeneRuler™ low-range DNA ladder, all the other lanes belong to the prepared standard sample.

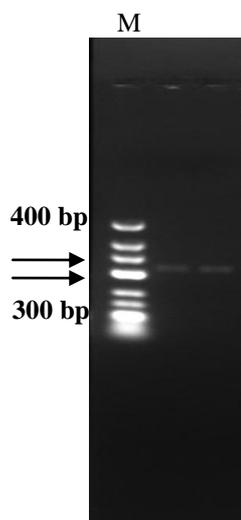


Figure 4.55 Purified sample of the *ARG4* gene standard for qPCR experiments. M: GeneRuler™ low-range DNA ladder, two other lanes are purified standard samples.

4.8.2. Standard DNA sample preparation for *hGH* gene

The outer primer pair utilized for *hGH* gene standard sample preparation was presented in Table 3.20. The optimum PCR mixture and the thermal profile of the conducted PCR can be found in Table 4.37 and 4.2, respectively.

Table 4.37. Optimum PCR mixture for preparation of the standard DNA for *hGH* gene using base plasmid.

Component	Amount
10X amplification buffer with KCl	5 μ L
2 mM solution of dNTPs	5 μ L
5 μ M <i>GAP</i> Forward primer	2 μ L
5 μ M <i>AOX</i> Reverse primer	2 μ L
25 mM MgCl ₂ /MgSO ₄	3 μ L
<i>Taq</i> DNA polymerase (5 U/ μ L)	1-2 U
dH ₂ O	32 μ L
Template DNA (base plasmid)	1 μ L
Total volume	50 μL

The result of the PCR, has been represented in (Figure 4.56). However, a purified sample was prepared by purification of the PCR result with PCR purification kit. The purification results can be observed in (Figure 4.57). After purification, the quality and the concentration of the prepared sample was checked; the average concentration was measured as approximately 106.5 ng/ μ L which is equal to approximately 8.91×10^{10} copy/ μ L (equation 3.3).

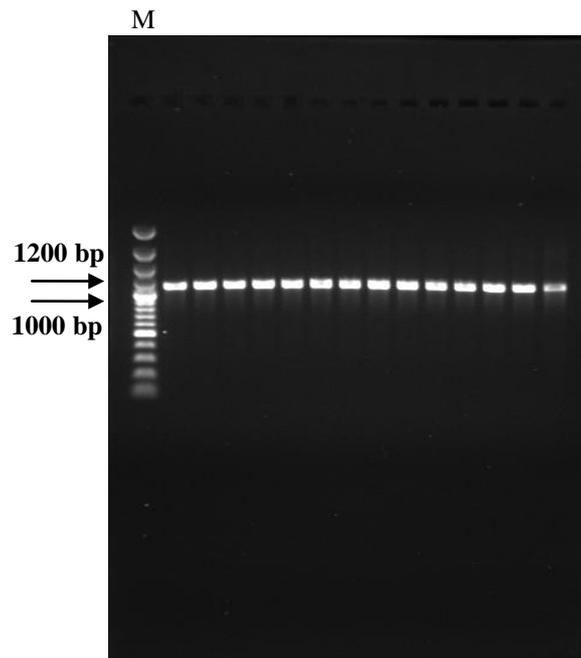


Figure 4.56. Preparation of the *hGH* gene standard sample (expected length of 1090 bp) by using BP as template. M: GeneRuler™ 100bp plus DNA ladder., all the other lanes belong to the prepared standard sample.

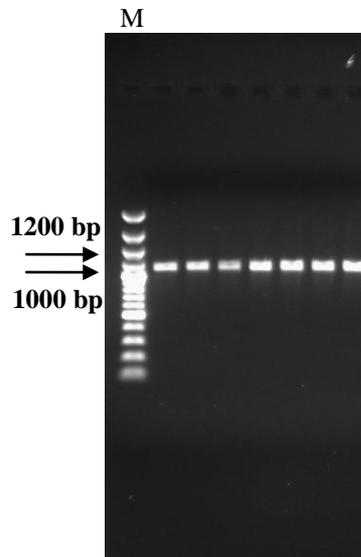


Figure 4.57. Purified sample of the *hGH* gene standard for qPCR experiments. M: GeneRuler™ 100bp plus DNA ladder, seven other lanes are purified samples.

After preparation of the standard samples for both *ARG4* and *hGH* genes and conducting preliminary PCRs with the mixtures represented in Tables 4.38 and 4.39, optimum concentration of the (inner) primers for qPCR experiments was obtained as 200 nM where any kind of non-specific bands were absent and, therefore, the fluorescence in qPCR will be just as the result of the desired DNA fragment. The recruited PCR conditions can be found in Table 4.40.

Table 4.38. Optimum amount of the ingredients for conducting qPCR using *hGH* standard (1/100 and 1/1000 dilution) and *r-P. pastoris* genomic DNA (1/100 and 1/1000 dilutions) as template. MgCl₂ was used with constant concentration of 2 mM. Since the other ingredients are available in the SYBR Green mixture and cannot be changed, the major parameter is the primer amount.

Component	Amount
10X amplification buffer with KCl	2 µL
2 mM solution of dNTPs	2 µL
5 µM <i>Eda-qPCR-F</i> primer	0.8 µL
5 µM <i>Eda-qPCR-R</i> primer	0.8 µL
25 mM MgCl ₂ / MgSO ₄	1.6 µL
<i>Taq</i> DNA polymerase (5 U/ µL)	1-2 U
dH ₂ O	10.8 µL
Template DNA	2 µL
Total volume	20 µL

Table 4.39. Optimum amount of the ingredients for conducting qPCR with *ARG4-F* and *ARG4-R* primers, and *ARG4* standard (1/1000, 1/10000, and 1/100000 dilutions), and *r-P. pastoris* genomic DNA (1/100 and 1/1000 dilution) as template. MgCl₂ was used with constant concentration of 2 mM. Since the other ingredients are available in the SYBR Green mixture and cannot be changed, the major parameter is the primer amount.

Component	Amount
10X amplification buffer with KCl	2 µL
2 mM solution of dNTPs	2 µL
5 µM <i>ARG4-F</i> primer	0.8 µL
5 µM <i>ARG4-R</i> primer	0.8 µL
25 mM MgCl ₂ / MgSO ₄	1.6 µL
<i>Taq</i> DNA polymerase (5 U/ µL)	1-2 U
dH ₂ O	10.8 µL
Template DNA	2 µL
Total volume	20 µL

Table 4.40. Preliminary PCR conditions for obtaining optimum concentration of the inner primers utilized in qPCR.

Number of Cycles	Temperature	Time
1 cycle	95 °C	10 min.
45 cycles	95 °C	10 sec.
	55 °C	5 sec.
	72 °C	15 sec.
Final extension	72 °C	10 sec.

4.8.3. qPCR experiments

The preparation of the samples was conducted as explained completely (section 3.6.14.2.3.). Each colony's isolated genome was analyzed with both *ARG4* inner primers and *hGH* inner primers and regarding obtained copy numbers for *ARG4* and *hGH*, the actual *hGH* copy number inserted into the

genome was determined (equation 3.2). By taking into consideration the obtained results, the qPCR experiments were conducted in three sets.

Regarding the tube capacity of the equipment for qPCR experiments, i.e. 36 wells, it was decided to start with four colonies from different groups (either those harboring SPs or those harboring selected promoters) each in duplicate (in total 16 tubes). Thereafter, 4 serial dilutions of each standard each in duplicate (in total 16 tubes); the copy number of the prepared serial dilutions of the standard samples is available in Table 4.41. Finally, two non-template control (NTC) tubes were also prepared. The configuration of 4 selected colonies can be found in Table 4.42. It was the first set of the qPCR experiments and performed in duplicate. The final results of the duplicated runs as 1st set of experiments can be seen in Table 4.44.

Table 4.41. Copy number of the *ARG4* and *hGH* genes in serial dilutions of the corresponding standard DNA samples. The grey highlighted dilutions were recruited in qPCR experiments.

<i>hGH</i> standard dilution	copy/ μL	<i>ARG4</i> standard dilution	copy/ μL
1/100	8.91×10^8	1/100	3.41×10^8
1/10 ³	8.91×10^7	1/10 ³	3.41×10^7
1/10 ⁴	8.91×10^6	1/10 ⁴	3.41×10^6
1/10 ⁵	8.91×10^5	1/10 ⁵	3.41×10^5
1/10 ⁶	8.91×10^4	1/10 ⁶	3.41×10^4
1/10 ⁷	8.91×10^3	1/10 ⁷	3.41×10^3

Table 4.42. Configuration of the samples in 1st set of the qPCR experiments. Standards start from 1/10⁴ dilution (correspond to hGH-1 and ARG-1) to 1/10⁷ dilution (correspond to hGH-4 and ARG-4). Each genomic DNA with both *hGH* and *ARG4* inner primers were assayed in duplicate. The corresponding developed plasmid of the colonies can be found in (Table 4.34).

Well	Sample	Well	Sample
1	hGH-1 Standard	21	Colony 64 with hGH
2	hGH-1 Standard	22	Colony 64 with hGH
3	hGH-2 Standard	23	Colony 64 with ARG
4	hGH-2 Standard	24	Colony 64 with ARG
5	hGH-3 Standard	25	Colony 53 with hGH
6	hGH-3 Standard	26	Colony 53 with hGH
7	hGH-4 Standard	27	Colony 53 with ARG
8	hGH-4 Standard	28	Colony 53 with ARG
9	ARG-1 Standard	29	Colony 33 with hGH
10	ARG-1 Standard	30	Colony 33 with hGH
11	ARG-2 Standard	31	Colony 33 with ARG
12	ARG-2 Standard	32	Colony 33 with ARG
13	ARG-3 Standard	33	dH ₂ O with hGH (NTC)
14	ARG-3 Standard	34	dH ₂ O with ARG (NTC)
15	ARG-4 Standard		
16	ARG-4 Standard		
17	Colony 163 with hGH		
18	Colony 163 with hGH		
19	Colony 163 with ARG		
20	Colony 163 with ARG		

According to the measured copy numbers lower than one, it was decided to change the number of the colonies in each experiment (Table 4.43). Therefore, just two colonies were included in each run both in quadruplicate for each set of the primers (in total 8 tubes for each colony). The colonies were again selected randomly among the developed plasmids.

These runs were regarded as 2nd set of the qPCR experiments which their results can be seen in (Table 4.44).

Table 4.43. Typical configuration of the prepared samples for 2nd set of the qPCR experiments. Standards start from 1/10⁴ dilution (correspond to hGH-1 and ARG-1) to 1/10⁷ dilution (correspond to hGH-4 and ARG-4). Each genomic DNA with both *hGH* and *ARG4* inner primers were assayed in quadruplicate. Number 17 to 24 belongs to one colony and numbers 25 to 32 belong to the other colony.

Well	Sample	Well	Sample
1	hGH-1 Standard	21	Colony with ARG
2	hGH-1 Standard	22	Colony with ARG
3	hGH-2 Standard	23	Colony with ARG
4	hGH-2 Standard	24	Colony with ARG
5	hGH-3 Standard	25	Colony with hGH
6	hGH-3 Standard	26	Colony with hGH
7	hGH-4 Standard	27	Colony with hGH
8	hGH-4 Standard	28	Colony with hGH
9	ARG-1 Standard	29	Colony with ARG
10	ARG-1 Standard	30	Colony with ARG
11	ARG-2 Standard	31	Colony with ARG
12	ARG-2 Standard	32	Colony with ARG
13	ARG-3 Standard	33	dH ₂ O with hGH (NTC)
14	ARG-3 Standard	34	dH ₂ O with ARG (NTC)
15	ARG-4 Standard		
16	ARG-4 Standard		
17	Colony with hGH		
18	Colony with hGH		
19	Colony with hGH		
20	Colony with hGH		

It was noticed that colony number 163 led to a copy number different than the others, another colony from corresponding plasmid group

(pTHI3ZαA::hGH) was selected, colony number 160. Second set of the qPCR runs was completed with analyzing this colony.

Table 4.44. Summarized results of 1st and 2nd sets of the qPCR experiments. The efficiencies are for each of the standard samples separately. The (+) and (-) signs in NTC column refer to the presence or absence of the fluorescence in the corresponding NTC tube, respectively. The copy numbers are the respective amounts obtained for the colonies. The grey part is the results of the 1st set of the qPCR experiments.

Colonies	Efficiency (%)	R ² -value	NTC	Copy Number
163, 64, 53, 33	ARG 92.1	0.9976	-	1.08, 0.55, 0.61, 0.61
	hGH 91.6	0.9993	-	
163, 64, 53, 33	ARG 88.7	0.9995	+	
	hGH 90.8	0.9998	-	
163, 64	ARG 93.4	0.9999	+	1.29, 0.64
	hGH 92.4	0.9999	-	
53, 33	ARG 89.2	0.998	+	0.55, 0.56
	hGH 86.4	0.9995	-	
144, 120	ARG 92.8	0.9999	-	0.56, 0.61
	hGH 96.3	0.9976	-	
109, 81	ARG 87.5	0.9977	-	0.47, 0.47
	hGH 95.6	0.9994	-	
160, 4	ARG 91.8	0.9983	-	0.59, 0.64
	hGH 88.6	0.9998	(dimmer)	

By finalizing two abovementioned qPCR sets, it was realized that the copy numbers are still lower than unity which makes the inference a little problematic as the copy numbers have been obtained in different runs. The reason for these amounts can be any kind of mistake especially in concentration measurement of the prepared standards. As the final solution,

it was decided to collect all the colonies belong to the same group (plasmids harboring SPs or plasmids harboring promoters) in one qPCR run and, therefore, any copy number in same range can be acceptable and rational; we were exploring same copy numbers and preferentially one in each group in order to have a robust comparison based on the rhGH production. This set was 3rd set of the qPCR experiments. Verified transfectant strains of *P. pastoris* were divided into two groups: group of the strains with integrated plasmids harboring available promoters (promoter group) and group of the strains with integrated plasmids harboring available SPs (SP group). The former includes the colonies possessing pGAPZ α A::hGH (colony 109), pPDCZ α A::hGH (colony 120), pPYRKZ α A::hGH (colony 144), and pTHI3Z α A::hGH (colony 160) and the latter consists of the colonies possessing pGAPZ α A::hGH (colony 109), pGAPZ13A::hGH (colony 4), pGAPZ23A::hGH (colony 33), pGAPZ24A::hGH (colony 53), pGAPZ26A::hGH (colony 64), and pGAPZ34A::hGH (colony 81). Colony number 109 is the interface of the two groups where in promoter group its “P_{GAP}” is considered and in SP group its “ α -MF signal peptide” is considered. The configuration of the samples in the 3rd set has been presented in Tables 4.45 and 4.46. The 3rd set of the qPCR experiments was conducted in triplicate.

The obtained *hGH* gene copy numbers in separate runs and the average copy number of the *hGH* gene in the intended strains can be observed in Table 4.47 and 4.48, respectively. The tabulated organized qPCR results related with the triplicate runs of both promoter and SP group (in total 6 runs) has been presented in Appendix Q. The original raw data of the typical qPCR experiments of the 3rd set, produced by qPCR program, can also be found in Appendix R.

Table 4.45. Samples configuration in 3rd set of the qPCR experiments related to promoter group. Standards start from 1/10⁴ dilution (correspond to hGH-1 and ARG-1) to 1/10⁷ dilution (correspond to hGH-4 and ARG-4). Each genomic DNA with both *hGH* and *ARG4* inner primers were assayed in duplicate. The run was performed in triplicate.

Well	Sample	Well	Sample
1	hGH-1 Standard	21	Colony 120 with hGH
2	hGH-1 Standard	22	Colony 120 with hGH
3	hGH-2 Standard	23	Colony 120 with ARG
4	hGH-2 Standard	24	Colony 120 with ARG
5	hGH-3 Standard	25	Colony 144 with hGH
6	hGH-3 Standard	26	Colony 144 with hGH
7	hGH-4 Standard	27	Colony 144 with ARG
8	hGH-4 Standard	28	Colony 144 with ARG
9	ARG-1 Standard	29	Colony 160 with hGH
10	ARG-1 Standard	30	Colony 160 with hGH
11	ARG-2 Standard	31	Colony 160 with ARG
12	ARG-2 Standard	32	Colony 160 with ARG
13	ARG-3 Standard	33	dH ₂ O with hGH (NTC)
14	ARG-3 Standard	34	dH ₂ O with ARG (NTC)
15	ARG-4 Standard		
16	ARG-4 Standard		
17	Colony 109 with hGH		
18	Colony 109 with hGH		
19	Colony 109 with ARG		
20	Colony 109 with ARG		

Table 4.46. Samples configuration in 3rd set of the qPCR experiments related to SP group. Standards start from 1/10⁴ dilution (correspond to hGH-1 and ARG-1) to 1/10⁷ dilution (correspond to hGH-4 and ARG-4). Just hGH-3 and ARG-3 were in duplicate. Each genomic DNA with both *hGH* and *ARG4* inner primers were assayed in duplicate. The run was performed in triplicate.

Well	Sample	Well	Sample
1	hGH-1 Standard	23	Colony 53 with hGH
2	hGH-2 Standard	24	Colony 53 with hGH
3	hGH-3 Standard	25	Colony 53 with ARG
4	hGH-3 Standard	26	Colony 53 with ARG
5	hGH-4 Standard	27	Colony 64 with hGH
6	ARG-1 Standard	28	Colony 64 with hGH
7	ARG-2 Standard	29	Colony 64 with ARG
8	ARG-3 Standard	30	Colony 64 with ARG
9	ARG-3 Standard	31	Colony 81 with hGH
10	ARG-4 Standard	32	Colony 81 with hGH
11	Colony 109 with hGH	33	Colony 81 with ARG
12	Colony 109 with hGH	34	Colony 81 with ARG
13	Colony 109 with ARG	35	dH ₂ O with hGH (NTC)
14	Colony 109 with ARG	36	dH ₂ O with ARG (NTC)
15	Colony 4 with hGH		
16	Colony 4 with hGH		
17	Colony 4 with ARG		
18	Colony 4 with ARG		
19	Colony 33 with hGH		
20	Colony 33 with hGH		
21	Colony 33 with ARG		
22	Colony 33 with ARG		

Table 4.47. Results of 3rd set of the qPCR experiments including promoter and SP groups separately. The copy numbers are the respective amounts obtained for the colonies. The (+) and (-) signs in NTC column refer to the presence or absence of the fluorescence in the corresponding NTC tube, respectively.

Colonies	Efficiency (%)	R ² -value	NTC	Copy Number
109, 120, 144, 160	ARG 96.8	0.9996	-	0.54, 0.59,
	hGH 91.7	0.9995	-	0.55, 0.58
	ARG 93.7	0.9999	-	0.59, 0.62,
	hGH 93.3	0.9998	-	0.62, 0.61
109, 4, 33, 53, 64, 81	ARG 89.2	0.9996	-	0.50, 0.54,
	hGH 90.4	0.9998	-	0.56, 0.51
	ARG 92.7	0.9999	-	0.57, 0.67, 0.67,
	hGH 90.9	0.9998	-	0.64, 0.74, 0.67
109, 4, 33, 53, 64, 81	ARG 92.2	0.9990	-	0.50, 0.52, 0.54,
	hGH 90.4	0.9997	-	0.52, 0.54, 0.52
	ARG 88.2	0.9997	-	0.62, 0.58, 0.70,
	hGH 86.9	0.9999	-	0.61, 0.76, 0.62

Table 4.48. The evaluated average copy number results deduced from triplicate experiments presented in Table 4.47. The tabulated organized data can be found in Appendix Q.

Colony		4	33	53	64	81	109	120	144	160
Evaluated Copy Number	SP group	0.59	0.64	0.58	0.67	0.60	0.56	-	-	-
	Promoter group	-	-	-	-	-	0.54	0.58	0.58	0.57

Although the evaluated copy numbers are lower than 1, but they are in same order and are very close to each other and it can be inferred that all the colonies in the promoter and SP groups possess the same copy number and with high probability one. After confirmation of the copy number similarity in selected r-*P. pastoris* strains, Microbank™ stocks were prepared from each of the verified strains of r- *P.pastoris* and, subsequently, were stored at -80°C in order to be used in fermentation experiments.

4.9. Experiments for selection of the most efficient SP

After choosing desired single-copy strains of *P. pastoris* corresponding to the selected SPs, determination of the most efficient SP was accomplished. At first step, preliminary shake flask bioreactor experiments were conducted which were, then, completed by subsequent laboratory-scale bioreactor experiments.

4.9.1. Shake flask air-filtered bioreactor experiments

The experiments were conducted as explained in section 3.7.1. The corresponding strains of the selected five endogenous SPs of *P.pastoris* (SP13, SP23, SP24, SP26, SP34) along with BP (possess α -MF as secretion SP) were recruited in these experiments. In all strains the r-protein production was driven by P_{GAP} . The shake flask bioreactor experiments were conducted in triplicate. The measurement of the cell concentration until t=48 h revealed that after approximately t=24 h the process enters its stationary phase. Furthermore, SDS-PAGE analysis of the t=36 h and t=48 h samples confirmed r-protein (r-hGH) degradation due to the cell lysis and subsequent action of intracellular proteases (Appendix S). In consequence, 24th hour samples were subjected to the desired measurements and making inferences. The cell concentration of the conducted shake flask bioreactor experiments can be observed in (Figure 4.58). The obtained results point of the similar growth trend of the six strains in triplicate runs. The cell-free supernatant samples of the t=24 h of the triplicate shake flask experiments were

subjected to the SDS-PAGE and dot-blot analyses in order to check the presence of the rhGH in the extracellular medium. The SDS-PAGE result of the 1st shake flask bioreactor experiment can be observed in (Figure 4.59) and the results related with the 2nd and 3rd experiments have been presented in (Appendix S). In addition, the results related with the dot-blot analysis can also be seen in (Figure 4.60). Converging point of the two sets of the results is the success of all selected endogenous SPs in secretion of the rhGH which is in agreement with their D-score > 0.8.

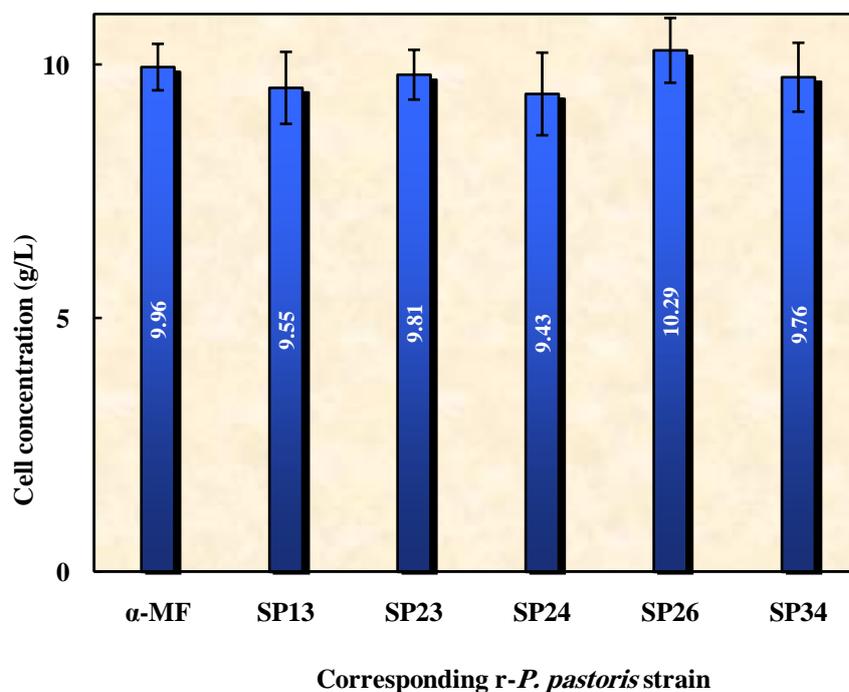


Figure 4.58. The average cell concentration of the cultivated single-copy *r-P. pastoris* strains related to SPs at t=24 h of production in triplicate shake flask bioreactor experiments.

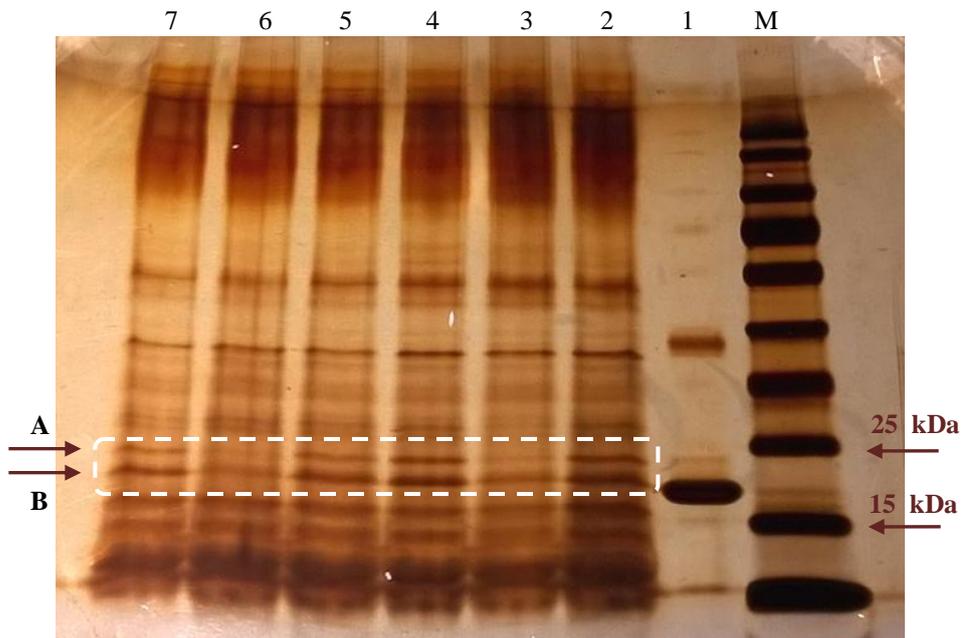


Figure 4.59. SDS-PAGE results of the 1st shaker experiment at t=24 h of production. M: protein marker. 1: hGH standard 0.05 g/L, cell-free supernatant of the r-*P. pastoris* strains of 2: pGAPZ α A::hGH , 3: pGAPZ13A::hGH, 4: pGAPZ23A::hGH , 5: pGAPZ24A::hGH, 6: pGAPZ26A::hGH, 7: pGAPZ34A::hGH.

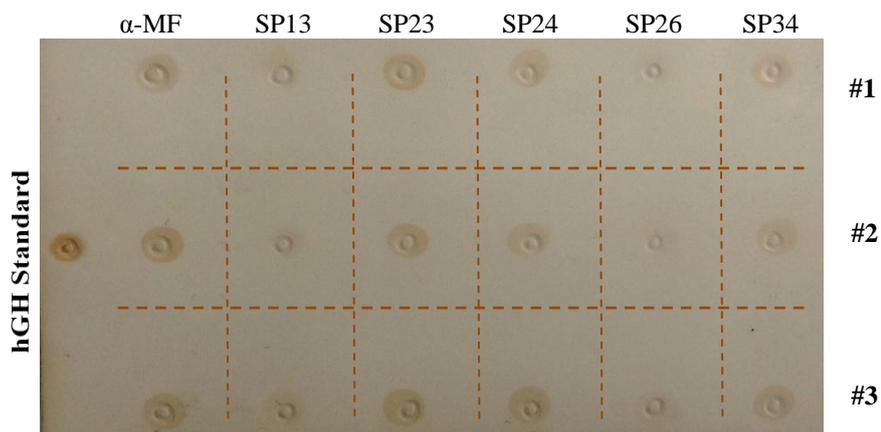


Figure 4.60. Dot-blot analysis for t=24 h samples of three shake flask bioreactor experiments. The number of the shake flask bioreactor experiment has been designated in right-hand side of the figure. hGH standard was used with concentration of 2.5 mg/L.

The important point in the (Figure 4.59) is the presence of two bands (A and B) around hGH standard; the upper band (A) is the expressed rhGH in original form, i.e., with fusion tag (including 6His-tag and factor Xa protease cleavage sequence), in the extracellular medium and the lower band (B) is the rhGH with removed his-tag as the result of the cleavage by available machinery in the cell (Çalık et al., 2008). The dot-blot results all confirm each other and triplicate runs all led to similar outcomes.

After using densitometric method and measuring the amount of the secreted rhGH determined by each of the two recruited analysis methods, the secreted rhGH at t=24 h was normalized by the amount of the cell present in the fermentation medium at t=24 h. Afterwards, the normalized rhGH quantity was used in order to express the effectiveness of the endogenous SPs in directing secretion of rhGH as the percentage of the α -MF secretion (Figure 4.61). Based on the measured results the most efficient endogenous SP was SP23 with the highest D-score of 0.883. However, its secretion efficiency was lower than α -MF with the D-score of 0.885. The ordered SPs regarding decreasing D-score can be written as: α -MF > SP23 > SP34 > SP24 > SP13 > SP26; on the other hand, the ordered SPs regarding decreasing secretion efficiency was obtained as: α -MF > SP23 > SP24 > SP34 > SP13 > SP26. Therefore, the higher D-score does not necessarily refers to a higher secretion efficiency (Liang et al., 2013).

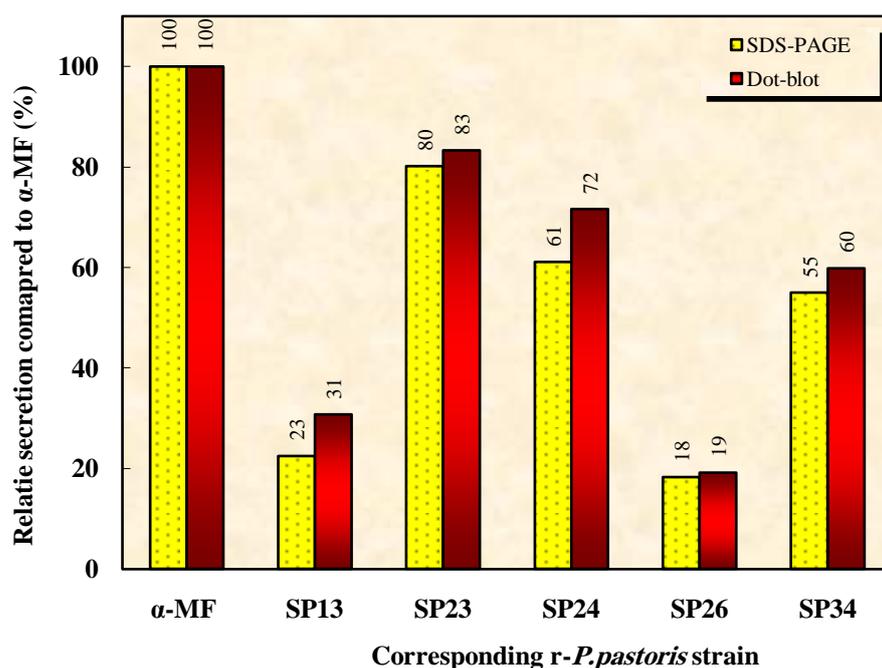


Figure 4.61. Relative amount of the normalized (by cell amount) secreted r-hGH by endogenous SPs as a percentage of the secretion directed by α -MF. The results obtained by SDS-PAGE and dot-blot analyses. The measurements belong to the t= 24 h samples.

4.9.2. Laboratory-scale bioreactor experiments

Following the obtained results in shake flask bioreactor experiments, in order to compare the secretion abilities of α -MF and SP23 in a more controlled environment, laboratory-scale bioreactor experiments were conducted as explained in section 3.7.2.1. The conducted experiments corresponding to α -MF and SP23 were named BR-1 and BR-2, respectively. After glycerol batch phase, the glucose fed-batch phase was started by exponential feeding of 500 g/L glucose solution according to the equation (2.85) in order to keep a constant specific growth rate of $\sim 0.07 \text{ h}^{-1}$. After taking samples in 3-hour intervals during glucose fed-batch phase of the fermentation, the samples were subjected to the cell concentration measurement, residual glucose measurement, intracellular rhGH measurement by dot-blot analysis, and extracellular rhGH measurement by SDS-PAGE analysis.

The maximum cell concentration was achieved at $t=15$ h and, then, started to decline in both strains. Its maximum amount was measured as 82 g/L and 75 g/L for corresponding strains of SP23 and α -MF, respectively (Figure 4.62). Approximately similar growth behavior of both strains and absence of any conspicuous difference was expected and can be attributed to the similar genetic structure of the two strains and, also, similar metabolic load exerted by a foreign protein (single-copy rhGH) to the normal physiologic functions of the cell. Meanwhile, concentration of the residual glucose in the medium was measured and until $t=15$ h there was no detectable glucose in the medium in both strains; however, glucose concentration started to increase at $t=15$ h owing to the decreased cell concentration (Figure 4.62).

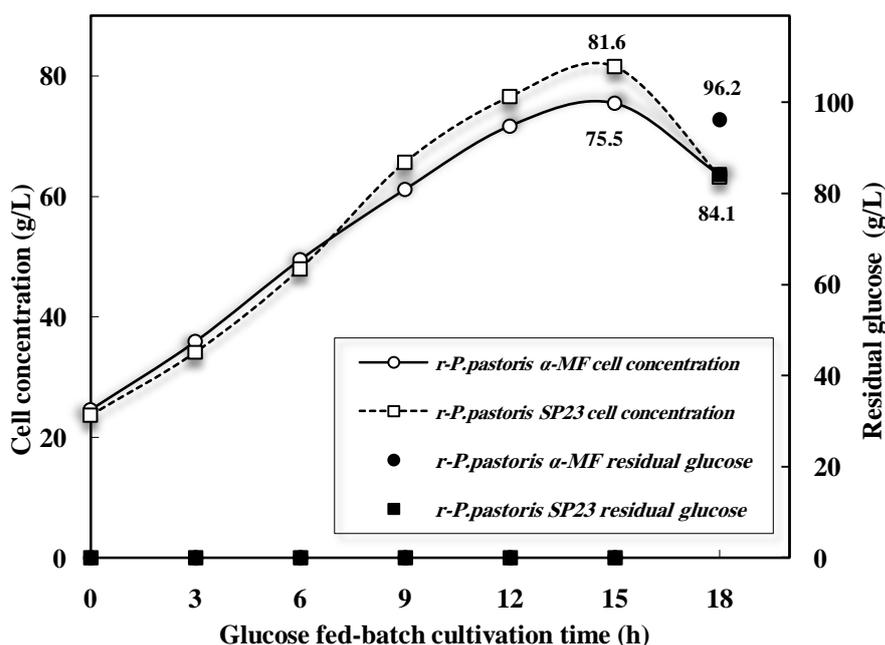


Figure 4.62. Cell concentration and residual glucose concentration during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of α -MF (BR-1) and SP23 (BR-2).

The SDS-PAGE results of the filtrate samples of two bioreactor experiments, BR-1 and BR-2, have been represented separately in (Figure 4.63) and (Figure 4.64), respectively. Furthermore, by visualizing the bands in each of the figures, it was decided to perform an extra SDS-PAGE analysis with the samples showing the highest amount of the secreted r-hGH in two bioreactor experiments (Figure 4.65) in one gel, in order to measure the rhGH simultaneously in a single gel to have a more precise densitometric results.

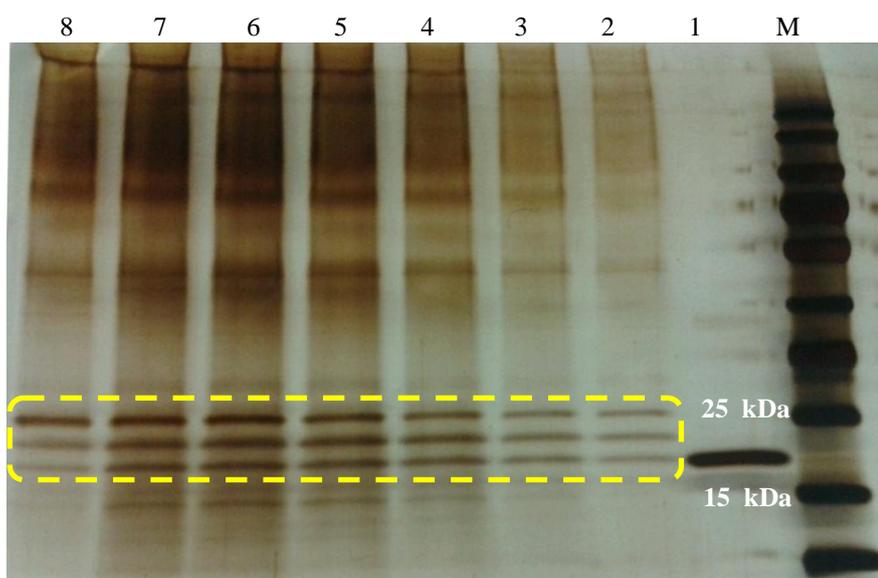


Figure 4.63 SDS-PAGE results related with BR-1 (*r-P. pastoris* pGAPZ α A::hGH strain). Samples of the glucose fed-batch phase with $\mu=0.07 \text{ h}^{-1}$. M: protein marker, 1: hGH standard 0.05 g/L, 2-8: represent the samples for $t=0, 3, 6, 9, 12, 15,$ and 18 h after fed-batch start, respectively.

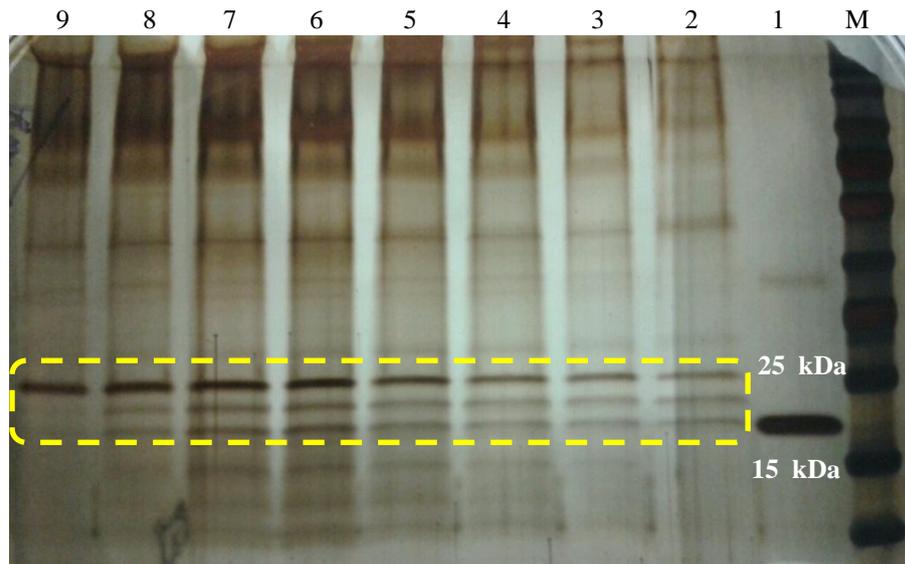


Figure 4.64. SDS-PAGE results related with BR-2 (*r-P. pastoris* pGAPZ23A::*hGH* strain). Samples of the glucose fed-batch phase with $\mu=0.07 \text{ h}^{-1}$. M: protein marker, 1: hGH standard 0.05 g/L, 2-9: represent the samples for $t= 0, 3, 6, 9, 12, 15, 18,$ and 20 h after fed-batch start, respectively.

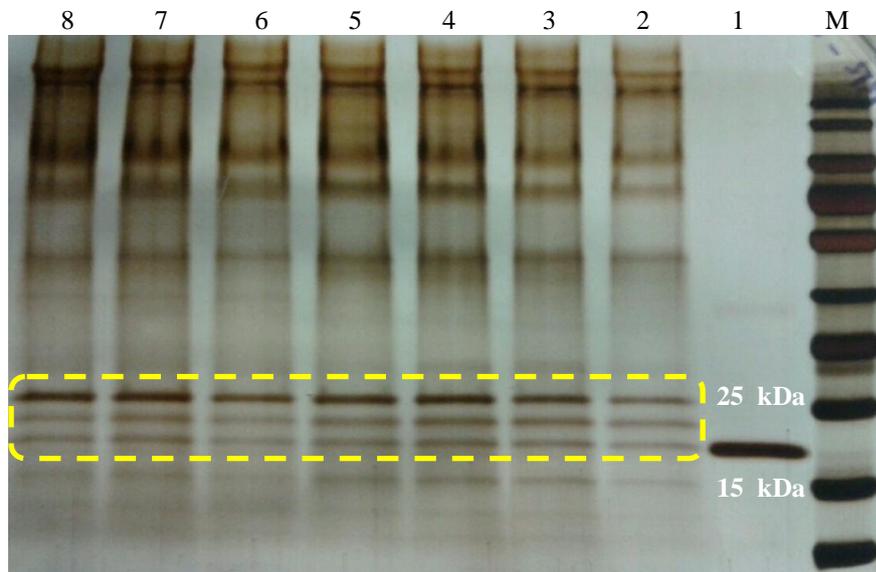


Figure 4.65. SDS-PAGE results related with *r-P. pastoris* pGAPZ α A::*hGH* strain (BR-1) and *r-P. pastoris* pGAPZ23A::*hGH* strain (BR-2) samples loaded in one gel. Samples were obtained from the glucose fed-batch phase with $\mu=0.07 \text{ h}^{-1}$. M: protein marker, 1: hGH standard 0.05 g/L, 2-5: BR-1 sample for $t= 6, 9, 12,$ and 15 h after fed-batch start, respectively. 6-8: BR-2 sample for $t= 9, 12,$ and 15 h after fed-batch start, respectively.

The SDS-PAGE results all refer to the presence of three bands around the expected location of the hGH while increasing during fed-batch phase. In order to cast light on the identity of these bands a Western blot analysis was also conducted using the selected samples of two bioreactor experiments. The result of the analysis has been presented in (Figure 4.66). The sensitivity and high specificity of the Western blot method was enough convincing to include the upper band (designated as ‘C’) in the calculation of the secreted rhGH in conducted bioreactor experiments.

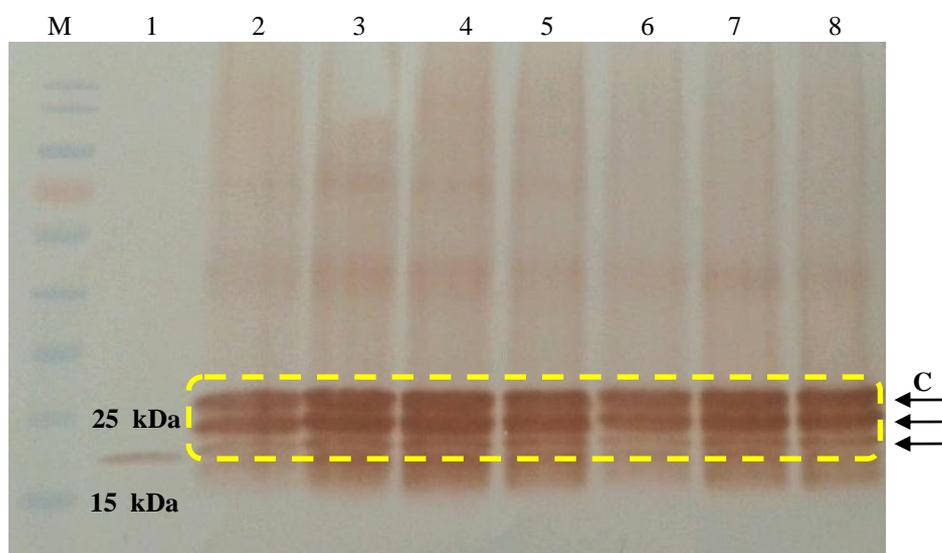


Figure 4.66. Western blot analysis result related with *r-P. pastoris* pGAPZ α A::hGH strain (BR-1) and *r-P.pastoris* pGAPZ23A::hGH strain (BR-2). Samples of the glucose fed-batch phase with $\mu=0.07$ h⁻¹. M: protein marker, 1: hGH standard 2.5 mg/L, 2-5: BR-1 sample for t= 6, 9, 12, and 15 h after fed-batch start, respectively. 6-8: BR-2 sample for t= 9, 12, and 15 h after fed-batch start, respectively. Samples were prepared as 1/3 dilution of the original supernatant sample.

By unraveling the problem related with the third band in SDS-PAGE results, the amount of the secreted rhGH was measured by utilization of the images of the conducted SDS-PAGE analyses. The measured secreted rhGH

(by densitometry) in BR-1 and BR-2 experiments can be found in (Figure 4.67). Both strains again showed a similar trend in secretion of rhGH; secreted rhGH by both strains increased from start of the fed-batch phase until $t=12$ h where reached its maximum in both strains 70 mg/L and 56 mg/L for BR-1 and BR-2, respectively. The secreted rhGH under conduction of SP23 was about 80% of the protein secreted by α -MF. However, in order to have a sound judgment, the amount of the cells present in the medium was also taken into consideration in order to prevent any effect which can be caused by the number of the expressing cells. In consequence, the normalized secreted rhGH by amount of the cell was calculated for two bioreactors which can be seen in (Figure 4.68).

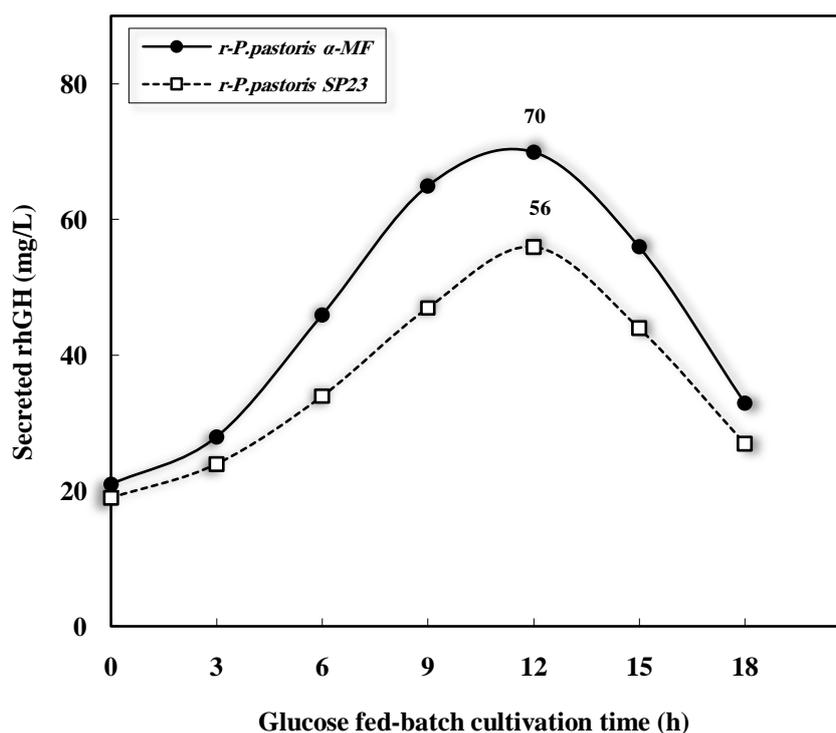


Figure 4.67. Concentration of the secreted rhGH during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of α -MF (BR-1) and SP23 (BR-2).

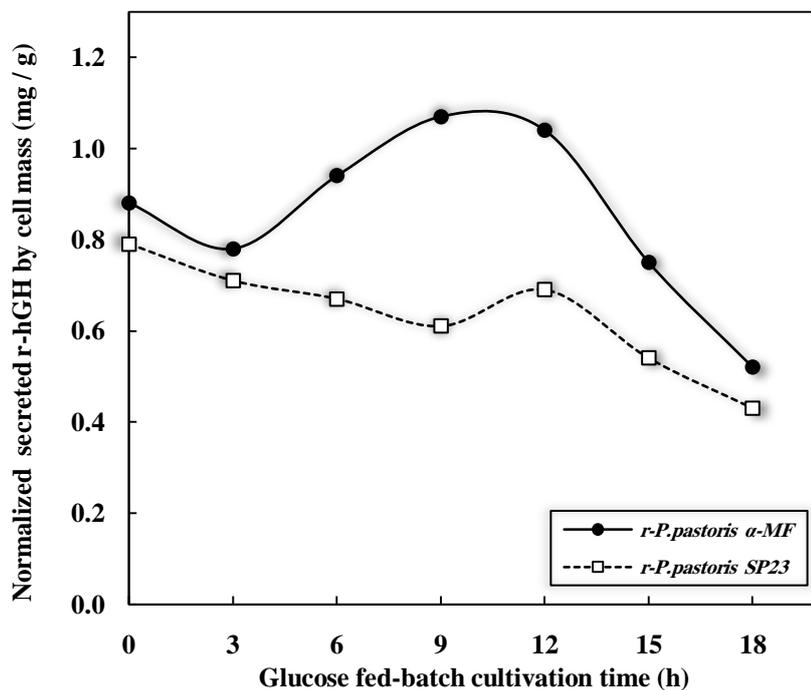


Figure 4.68. The normalized (by cell amount) secreted rhGH during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of α -MF (BR-1) and SP23 (BR-2).

The results again confirmed a more effective secretion by α -MF compared to SP23. The ratio of the normalized secreted rhGH in two conditions was measured to be 75% in average which was very close to the 80% obtained by secreted rhGH without normalization. In total, extracellular rhGH measurements revealed the higher secretion ability of the α -MF in comparison with SP23.

Parallel to higher extracellular amount of a protein as a scale for deciding about secretion efficiency of a SP, lower intracellular amount of that protein can also be considered as another indicator. Therefore, the intracellular rhGH concentration was also measured as an extra affirmative factor (Figure 4.69). Both profiles possessed a similar trend and after a decrease until $t=6$ h showed an increase toward the end of the process. The ratio of the intracellular rhGH in two SPs was very close and oscillated

between 0.8 and 1, the results confirmed the lower intracellular rhGH by utilization of the α -MF as secretion guide.

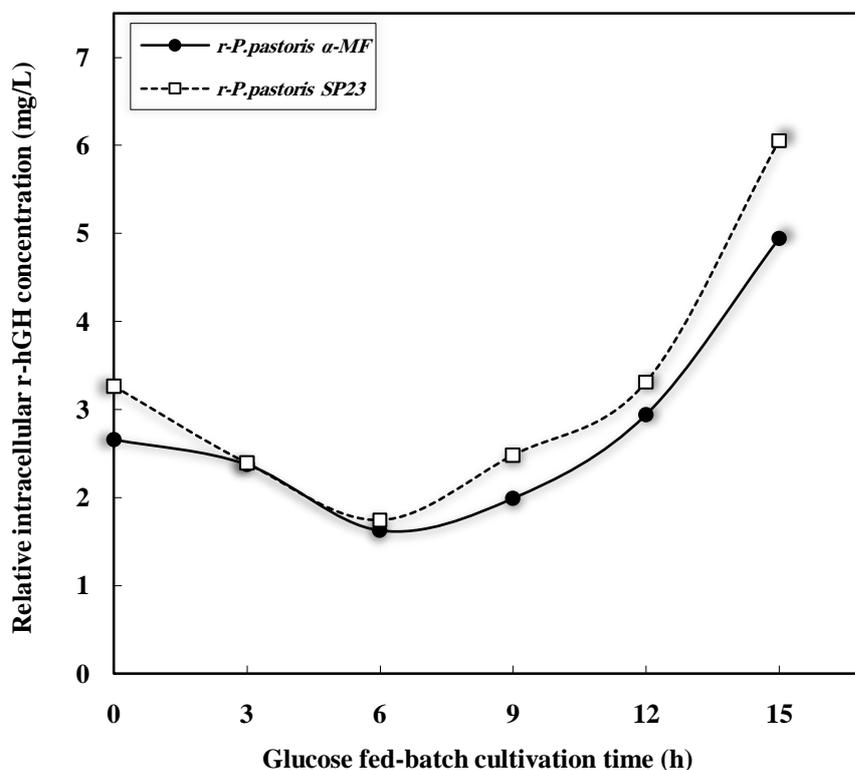


Figure 4.69. Relative intracellular rhGH concentration during glucose fed-batch phase of the fermentation with corresponding *r-P.pastoris* strains of α -MF (BR-1) and SP23 (BR-2).

By combining the results obtained from shake flask bioreactor experiments and laboratory-scale bioreactor experiments, SP23 secretion efficiency was comparable with α -MF. Roughly, SP23 secretion efficiency was 70-80% of α -MF secretion efficiency (Massahi and Çalık , 2016).

4.9.3. Correlating SP efficiency with its physicochemical properties

As proposed previously, the D-score does not necessarily correlate with the real amount of the secreted r-protein. Furthermore, it can be realized that

SP13, which seemed to be a more efficient choice compared to α -MF in secretion of the enhanced green fluorescent protein (Liang et al., 2013) was not very successful in secretion of rhGH. These different efficiencies of a single SP, here SP13, in secretion of different r-proteins support the idea of arbitrary selection of SP in the case of different r-proteins.

By conducting the experiments in the same process conditions under the control of the same promoter, the contribution of the parameters like transcription rate, translation rate in secreted r-protein amount is diminished; however, because of the direct correlation between the intrinsic characteristics of the SP and the rate of the targeting toward ER and passage through the ER membrane, the source of the differences in secretion efficiency was supposed to be the physicochemical properties of the SPs.

The targeting of the secretory proteins toward ER is initiated upon emergence of the SP from the ribosome in cytoplasm while translation is being carried out. At the beginning, SRP checks the availability of a SP in nascent polypeptide which is imperatively co-translational (Ng et al., 1996); afterwards, the selection of the pathway which mainly depends on the hydrophobicity of the hydrophobic core, H-region, of the SP (Ng et al., 1996) is performed. This factor plays more important role than the length of the H-region in recognition of the SP by SRP (Hatsuzawa et al., 1997). The governing parameter of the interactions occurring while translocation and maturation events take place is supposed to be the information in the H-region of the SP (Rutkowski et al., 2003). Nevertheless, the effect of the N-terminus or main body of the protein should not be overlooked (Andrews et al., 1988; Matoba and Ogrydziak, 1998; Ng et al., 1996) especially proline residues in close proximity of the cleavage point of the SP may be decisive because of their potential effect in spatial positioning of the SP and, in turn, influencing the interaction with ER translocation machinery, specifically SRP.

In order to scrutinize the parameters contribute in secretion by influencing the chemistry of the nascent polypeptide and, therefore, its

interaction with secretion pathway elements and subsequent secretion, the major physicochemical properties (Biro, 2006), GRAVY score, aliphatic index (representative of the size), and pI (charge of the peptide chain) were analyzed.

The approach of the SP (or cargo protein/SP complex) to the negatively-charged phosphate groups in the outer part of the ER membrane via electrostatic forces can be affected by the net charge of the SP (and polypeptide), as a result of its pI and microenvironment pH (Kajava et al., 2002). The hydrophobicity of the SP is not only important in the selection of the targeting pathway, but also it may influence the passage through the hydrophobic trans-membrane part of the translocon in ER membrane. In addition, the aliphatic index as an indicator of the bulkiness of the side chains of the amino acids is important by considering the steric hindrance that these side chains may exert while approaching different translocation machinery and, thus, bulky amino acids may impair the maturation of the protein (Kajava et al., 2002).

It seems that over the course of the evolution, microorganisms have tailored their SPs, becoming more efficient in translocation of the corresponding protein domain (Andrews et al., 1988). Therefore, preliminary in-silico analyses were performed on hGH (mature, immature and its original SP) and the 12-amino acid fusion partner known as N-terminus motif (Figure 4.70) which is presented in Table 4.49. Furthermore, In-silico calculations related with pI, GRAVY score, and aliphatic index of the utilized SPs in this work and SPs used in Liang et al. study (2013) have been presented in Tables 4.50 and 4.51, respectively.

Table 4.49. Calculated physicochemical properties of hGH (mature and immature), its SP, and the N-terminus motif.

Sequence (length, aa)	Mw (Da)	GRAVY Score	Aliphatic Index	pI	Net Charge @ pH = 7	Mean Charge @ pH = 7
Immature hGH (217)	24847	-0.269	87.65	5.084	- 4.44	- 0.020
Mature hGH (191)	22129	-0.411	83.72	5.065	- 4.41	- 0.023
hGH SP (26)	2736	0.777	116.54	6.232	- 0.05	- 0.0019
N-terminus Motif (12)	1572.6	- 1.983	32.50	7.181	0.42	0.035

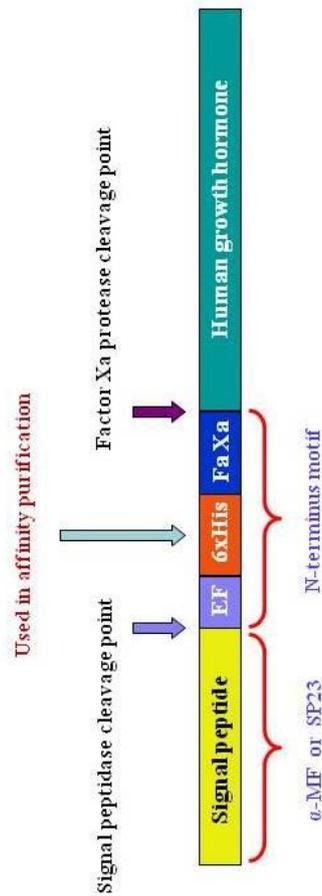


Figure 4.70. Schematization of the N-terminus motif in expressed rhGH.

Table 4.50. In-silico calculation of pI, GRAVY score, and aliphatic index related with utilized SPs.

SP	Isoelectric Point					Net Charge @ pH=7 & Mean Charge @ pH=7						
	SP with original protein	SP alone	SP + EFHHHHHIEGR	SP + EFHHHHHIEGR + hGH	SP with original protein	SP alone	SP + EFHHHHHIEGR	SP + EFHHHHHIEGR + hGH	SP + EFHHHHHIEGR	SP + EFHHHHHIEGR + hGH		
					Net	Mean	Net	Mean	Net	Mean		
SP13	4.197	6.090	7.181	6.229	-18.43	-0.040	-0.02	-0.001	0.42	0.013	-3.96	-0.018
SP23	6.201	9.700	7.978	6.417	-3.42	-0.009	0.98	0.061	1.42	0.051	-2.96	-0.014
SP24	5.920	10.793	9.678	6.786	-0.86	-0.006	2.94	0.147	3.39	0.106	-0.99	-0.004
SP26	4.525	6.093	7.180	6.229	-23.90	-0.106	-0.02	-0.001	0.42	0.014	-3.96	-0.018
SP34	3.722	10.549	7.978	6.417	-49.56	-0.111	0.98	0.058	1.42	0.049	-2.96	-0.013
Pre α -MF	-	10.549	-	-	-	-	0.98	0.052	-	-	-	-
Prepro α -MF	4.697	3.730	4.486	4.772	-7.04	-0.043	-9.01	-0.101	-8.57	-0.085	-12.95	-0.044

SP	GRAVY Score				H-region of SP
	SP with original protein	SP alone	SP + EFHHHHHIEGR	SP + EFHHHHHIEGR + hGH	
SP13	-0.019	1.865	0.422	-0.291	2.883
SP23	-0.366	2.175	0.393	-0.308	3.033
SP24	-0.217	1.555	0.228	-0.319	3.400
SP26	-0.627	1.861	0.323	-0.311	2.064
SP34	-0.423	1.882	0.283	-0.320	3.111
Pre α -MF	-	1.389	-	-	1.933
Prepro α -MF	-0.371	0.275	0.007	-0.266	1.933

SP	Aliphatic Index				H-region of SP
	SP with original protein	SP alone	SP + EFHHHHHIEGR	SP + EFHHHHHIEGR + hGH	
SP13	84.18	200.00	137.19	91.39	260.00
SP23	92.98	189.38	122.14	88.63	271.11
SP24	91.64	146.00	103.44	86.55	231.25
SP26	80.93	167.78	113.67	87.78	177.27
SP34	65.27	206.47	134.48	90.41	303.33
Pre α -MF	-	103.16	-	-	114.17
Prepro α -MF	77.09	97.75	90.00	85.89	114.17

Table 4.51. In-silico calculation of pI, GRAVY score, and aliphatic index related with SPs utilized in Liang et al. Study (2013).

SP	Isoelectric Point					Net Charge @ pH=7 & Mean Charge @ pH=7						
	SP with original protein	SP alone	SP + EF	SP + EF + CALB	SP with original protein		SP alone		SP + EF		SP + EF + CALB	
					Net	Mean	Net	Mean	Net	Mean	Net	Mean
SP4	4.301	5.919	3.849	5.203	-28.83	-0.070	-0.06	-0.004	-1.05	-0.058	-1.99	-0.006
SP13	4.197	6.090	3.849	5.203	-18.43	-0.040	-0.02	-0.001	-1.02	-0.046	-1.96	-0.006
SP17	4.393	6.099	3.849	5.203	-37.85	-0.037	-0.02	-0.0009	-1.02	-0.041	-1.96	-0.006
Pre α -MF	-	10.549	-	-	-	-	0.98	0.052	-	-	-	-
Prepro α -MF	4.697	3.730	3.693	4.337	-7.04	-0.043	-9.01	-0.101	-10.01	-0.110	-10.95	-0.027

SP	GRAVY Score				H-region of SP
	SP with original protein	SP alone	SP + EF	SP + EF + CALB	
SP4	-0.448	1.700	1.472	0.114	2.478
SP13	-0.019	1.865	1.664	0.142	2.883
SP17	-0.139	1.348	1.212	0.123	3.233
Pre α -MF	-	1.389	-	-	1.933
Prepro α -MF	-0.371	0.275	0.262	0.087	-

SP	Aliphatic Index				H-region of SP
	SP with original protein	SP alone	SP + EF	SP + EF + CALB	
SP4	79.61	158.75	141.11	87.13	227.78
SP13	84.18	200.00	181.82	90.41	260.00
SP17	81.08	156.52	144.00	88.45	292.22
Pre α -MF	-	103.16	-	-	114.17
Prepro α -MF	77.09	97.75	95.60	86.64	-

Isoelectric point: The least efficient SPs, SP13 and SP26, have $pI < pH = 7$ (common pH of the cytoplasm), while the efficient ones (including pre α -MF) have $pI > 7$ (Table 4.50) but without an order similar to their corresponding efficiency. $pI > 7$ leads to slightly positive net charge of the SP which can accelerate approach of the nascent protein to the ER membrane. The net charge of the SP/N-terminus motif/rhGH is negative in all six SPs but in the case of α -MF a higher negative net charge is observed which is not compatible with its higher efficiency as a result of electrostatic interactions. Therefore, SPs net charge seems to be more illustrative. In contrast to our findings, Liang et al. (2013) reported that the efficiency of the SPs ranked as $SP4 > SP17 > \alpha$ -MF $> SP13$ in secretion of *Candida Antarctica* lipase B (CALB) by $pI_{SP4} = 5.919 < 7$ (Table 4.51). In addition, the difference between pI of SP/N-terminus motif/rhGH combination and pI of SP/original protein combination points out that α -MF, SP23, and SP24 as the most efficient SPs have change percentages of 1.61%, 3.48%, and 14.68% respectively; the most efficient SP has the least change percentage, which is not valid for SP34, SP13, and SP26. In contrast, the change percentage for SP4 as the best SP in Liang et al. (2013) study is higher than SP17 and is not the least. In total, the presence of the N-terminus motif shifted the charges toward positive. Nevertheless, the net charge of the SP/N-terminus motif shows a remarkable difference between endogenous SPs and α -MF. The formers have a positive net charge and the later has a negative net charge. This is because in α -MF the “pre” part is not directly linked to the positively charged N-terminus motif. In order to mimic the conditions in endogenous SPs, twelve initial amino acids of the α -MF “pro” part was considered along with its “pre” part; the pI , net charge, and mean charge were obtained as 3.927, -2.02, and -0.065, respectively (data not shown in table). Therefore, such as prokaryotic SPs where the nature of the charged amino acids in the early sequence after SP is of prime importance (Li et al., 1988), the positive charge of the proximate region of the SP may be the reason of the lower efficiency of the endogenous SPs compared to the α -MF.

However, a rational relationship is not observable between SP efficiency and positive net charge among the endogenous SPs. In contrast, in Table 4.51 the net charge of the SP/EF residue is negative in all four cases and, hence, α -MF does not possess better secretion efficiency. It seems that the net negative charge of the SP and its nearby sequence in the mature protein improves the secretion; although, net charge bias in N-terminus of the mature proteins has been confirmed for Gram-negative bacteria (Kajava, 2000). At the end, it should be considered that pI of the SP/N-terminus motif/rhGH combinations in none of the cases resembles to the original evolutionary-gained pI of the immature hGH regarding Table 4.49.

GRAVY score: This parameter was regarded as the representative of hydrophobicity of a specific amino acid sequence. SP23 has the highest and α -MF “pre” part has the lowest GRAVY scores (Table 4.50). GRAVY scores of the H-regions of the SPs imply the possible divergence in the translocation pathway which is mainly governed by the hydrophobicity of the H-region of the SP; α -MF leads to the SRP-independent targeting because of GRAVY score < 2 (Ng et al., 1996) and endogenous SPs conduce to SRP-dependent targeting which means faster targeting in SRP-independent pathway. Excluding α -MF, the GRAVY scores of the SPs’ H-region are in the order of SP24 $>$ SP34 $>$ SP23 $>$ SP13 $>$ SP26 where except SP23, higher GRAVY score of the H-region coincides with the higher efficiency of the endogenous SP. The assumption of the “pathway as the main reason of the efficiency difference” cannot be valid in the Liang et al. (2013) study where the GRAVY score of 2.478 for SP4 refers to SRP-dependent pathway (Ng et al., 1996) and it still leads to a better result than α -MF. Finally, it is seen that GRAVY score of the α -MF/N-terminus motif/rhGH complex (- 0.266) is very close to GRAVY score of the immature hGH (- 0.269). However, there is no apparent similarity and relation between GRAVY scores of the recruited endogenous SPs and immature hGH considering Table 4.49.

Aliphatic index: The relative volume occupied by the side chains of the aliphatic amino acids of the peptide chain is represented by this value.

The side chains can have effect on the hydrophobic interactions. It is obvious in Table 4.50 that α -MF has relatively small aliphatic index compared to endogenous SPs. Less interference with the interaction between the SP and translocation machinery may be another reason for more efficient secretion by α -MF. Such an inference cannot be made for Linag et al. (2013) studies where α -MF still has the lowest aliphatic index (97.75) but SP4 with aliphatic index of 158.75 has the highest efficiency. Furthermore, there is no evident relationship between secretion efficiency and aliphatic index among the investigated endogenous SPs as SP26 with aliphatic index of 167.78 has lower efficiency compared to SP23 with aliphatic index of 189.38. In the end, the aliphatic indices of the SP/N-terminus motif/rhGH complexes are in the same range and very close to the immature hGH and α -MF has the closest aliphatic index to the original SP of hGH.

It should be emphasized that, the inconsistencies present between results of the current study and Liang et al. (2013) results can be because of the use of different fusion proteins and, consequently, different amino acid context around SP cleavage point. Some SPs seem to be very protein specific and the others have a general ability in secretion of different proteins; the amino acid sequence of the mature protein downstream of the cleavage point can affect SP efficiency.

Because of the “pre-pro” identity, it should be reminded that α -MF and the endogenous SPs may have basic differences in their mode of action which can prevent us to reach a reasonable inference. Although the privilege of the α -MF in current research can be attributed to above mentioned properties such as lower GRAVY score of the H-region, lower aliphatic index, or negative net charge around the SP cleavage point, related with the examined endogenous SPs, the results obtained from the in-silico analyses were not so conclusive regarding singular effect of the included physicochemical properties. Therefore, inclusion of more factors and even their possible interactions seems to be imperative where their combinatory effects may be more illustrative. Such insight can be achieved by

experimental design including the parameters that are supposed to play role in the proteins secretion and series of experiments with different r-proteins and SPs; the final result will be the r-protein secretion amount as a function of the selected parameters and their possible interactions. This function will be a start to reach the objective of the most efficient SP at least for each r-protein which may be designed artificially.

The preserved twelve-amino acid sequence (EFHHHHHHIEGR) in the N-terminus (referred as N-terminus motif) of rhGH contained a polyhistidine-tag for rapid purification and a target site for factor Xa protease. Thus, the *in vitro* cleavage will produce a mature form of the protein which has the native N- and C-termini; otherwise, there will be two extra amino acids (EF) resulting from the restriction site used in cloning, which is not plausible for the recombinant therapeutic protein. In consequence, this N-terminus motif may be the preferred tag not only for extracellular rhGH production but also for the production of other r-proteins. For further analysis of the effect of this N-terminus tag on SP identification and cleavage, the D-scores were calculated for SPs recruited in this study in combination with two extra selected proteins, *Aspergillus niger* xylanase (Xyl) and *Candida Antarctica* lipase B (CALB), both with and without N-terminus motif (Table 4.52). The percentage of the D-scores changes caused by addition of N-terminus motif in front of rhGH compared to rhGH without tag is greater than changes caused by insertion of a new protein in place of rhGH in presence of tag which implies the probable masking effect of the N-terminus motif on the effect of cargo protein N-terminus.

It has been previously illustrated that at least 22 amino acids from the N-terminus of the mature protein may affect the secretion efficiency in eukaryotes (Andrews et al., 1988). However, there is no evident that remote residues have any effect on SP or not. The new D-scores of the SPs with new proteins are also sufficiently above 0.8 and show a high likelihood of being a suitable SP.

Table 4.52. Calculation of the D-scores of utilized SPs in combination with two extra proteins (CALB and Xyl) in presence of the N-terminus motif in order to check the effect of this motif on D-score.

SP	D-score					
	With 12aa + hGH	With 12aa + CALB	With 12aa + Xyl	% Change with 12 aa compared to hGH	% Change when CALB replaces hGH after motif	% Change when Xyl replaces hGH after motif
SP13	0.836	0.829	0.837	-6.17	-0.84	0.12
SP23	0.883	0.884	0.888	-6.06	0.11	0.57
SP24	0.841	0.840	0.850	-7.58	-0.12	1.07
SP26	0.831	0.830	0.844	-8.58	-0.12	1.56
SP34	0.870	0.872	0.878	-6.15	0.23	0.69
<i>α</i> -MF	0.885	0.885	0.885	0	0	0

For acquiring more insight, the first amino acid of the three representative proteins (hGH, Xyl, CALB) were changed with other 19 available amino acids and the D-scores were re-calculated in presence of SP23 as the most promising endogenous SP addressed by this study (Table 4.53). The changes in D-score caused by replacement of the first amino acid (virtual mutation) of the proteins were not conspicuous and reflected the trivial effect of the first amino acid after N-terminus motif on D-score where still D-score > 0.8. Q (Glutamine), W (Tryptophan), Y (Tyrosine), N (Asparagine), P (Proline), V (Valine), C (Cysteine) did not decrease the D-score. Mutations which lead to start with these amino acids, after the motif, are less prone to decrease the D-score.

Table 4.53. D-scores of the SP23 with three proteins having different amino acids after N-terminus motif. White, grey, and blue colors refer to constant, increased, decreased D-scores compared to original protein, respectively. The numbers in parenthesis refer to the length of the intended protein. “Original” means: D-score calculated in presence of the amino acid that originally is the first amino acid of the N-terminus of the protein.

Amino acid	hGH (191 aa)	Xyl (207 aa)	CALB (317 aa)
V	0.885	0.888 (original)	0.889
P	0.885	0.888	0.889
H	0.882	0.884	0.885
D	0.885	0.887	0.888
S	0.883	0.886	0.887
E	0.885	0.887	0.888
R	0.883	0.885	0.886
A	0.884	0.887	0.887
K	0.884	0.886	0.887
L	0.881	0.883	0.884 (original)
T	0.885	0.887	0.888
G	0.885	0.887	0.888
N	0.886	0.888	0.889
F	0.883 (original)	0.885	0.886
Y	0.885	0.888	0.889
W	0.887	0.889	0.890
Q	0.886	0.888	0.889
I	0.883	0.886	0.887
M	0.884	0.886	0.887
C	0.888	0.890	0.892

Generalization of such explanations needs more analyses, inclusion of the amino acid context of the protein and simultaneous mutation of several amino acids until making sure that there will be no substantial effect on the SP cleavage efficiency at least in terms of D-score. These outcomes can be used for initiation of any other similar analysis with any available N-terminus motif.

In order to accept SP23 as a generalized endogenous SP which can compete with α -MF in r-protein production with *P. pastoris* in presence of available N-terminus tag, in-silico (virtual) mutations in the N-terminus of hGH (and any other candidate protein) will be rewarding. Subsequent measurements of D-scores will reveal the effect of the presence of the N-terminus amino acid motif. D-score > 0.8 refers to high probability of a successful secretion. If such an analysis would point out the insignificance of the protein N-terminus, the mature protein N-terminus was neglected and SP23, following α -MF, could be considered as a highly efficient candidate endogenous SP in the cases that present N-terminus motif is recruited for downstream affinity purification purposes.

On the other hand, the eighty-nine amino acid length of the α -MF compared to the SP23 length with 16 amino acids seems to be enough convincing to choose SP23 for secretion of r-proteins in starvation and situations that microorganism will experience lack of adequate nutrients.

4.10. Experiments for selection of the best oxygen limitation-induced promoter

In order to check the ability of the selected promoters in expression of rhGH, the corresponding single-copy strains of r-*P. pastoris* were utilized in both shake flask and bioreactor experiments.

4.10.1. Preliminary shake flask air-filtered bioreactor experiments

Before starting bioreactor experiments, in order to be confident about the proper function of the selected *putative* promoter regions, their capability in

driving transcription of *hGH* gene and, thus, subsequent expression was checked through preliminary shake flask experiments depicted in Figure 3.6.B. The obtained SDS-PAGE result can be observed in Figure 4.71.

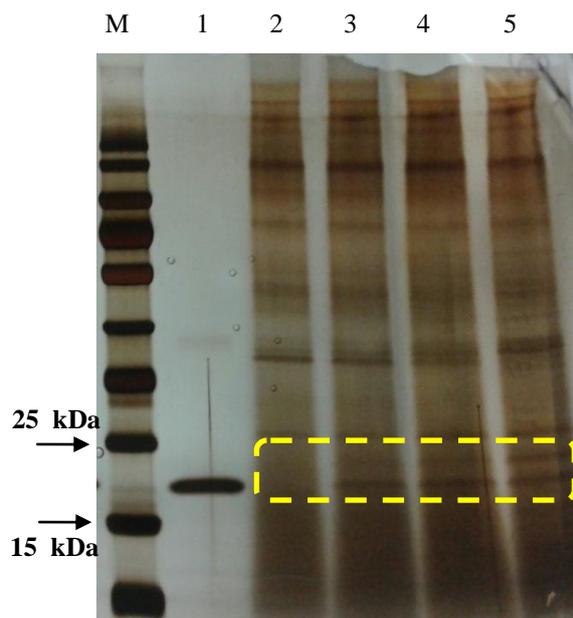


Figure 4.71. SDS-PAGE results related with preliminary shake flask bioreactor experiment for validation of promoter activity of the selected *putative* promoter regions (t=24 h). M: protein marker, 1: hGH standard 0.05 g/L, 2: pTHI3ZαA::hGH, 3: pPYRKZαA::hGH, 4: pPDCZαA::hGH, 5: pGAPZαA::hGH.

According to the SDS-PAGE result, it is obvious that *putative* promoter region related with “thiamine precursor biosynthesis enzyme THI3” did not express rhGH. However, it cannot be attributed strictly to the wrong promote region since this promoter is repressed by exogenous thiamine (Stadlmayr et al., 2010; Landes et al., 2016) and, therefore, the presence of thiamine in medium constituents (yeast extract) can lead to

promoter repression and switching off rhGH expression. As a result, P_{THI3} was discarded and other two promoters (P_{PDC} and P_{PYRK}) along with P_{GAP} were utilized in laboratory-scale bioreactor experiments.

4.10.2. Laboratory-scale bioreactor experiments

The experiments were conducted as explained in section 3.7.2.2 (Figure 3.8) using three selected single-copy strains of *r-P. pastoris* corresponding to promoter constructs ($pGAPZ\alpha A::hGH$, $pPDCZ\alpha A::hGH$, and $pPYRKZ\alpha A::hGH$). The runs were named as BR-3, BR-4 and BR-5 for the strains carrying $P_{PYRK}::hGH$, $P_{PDC}::hGH$, and $P_{GAP}::hGH$, respectively. After glycerol batch phase, glucose fed-batch phase was started by feeding 500 g/L glucose solution into the bioreactor. Dissolved oxygen was kept in desired range ($1\% < DO < 4\%$) until $t=9$ h; however, the system was not capable of controlling DO in desired level and it fell to zero at $9 \text{ h} < t < 15 \text{ h}$ despite of aeration with flowrate of 10 L/min (the maximum capacity of the system). The period $0 < t < 9 \text{ h}$ is regarded as low oxygen availability and $9 \text{ h} < t < 15 \text{ h}$ is regarded as severe oxygen limitation where the extent of oxygen limitation is increased towards the end of the period ($t=15 \text{ h}$).

Taken samples from the bioreactor were subjected to the required analyses in order to verify the efficiency of the selected promoters. The cell concentration profile during fed-batch phase of the three experiments was obtained which can be observed in Figure 4.72. All the strains exhibited similar trend and reached their maximum at $t=15 \text{ h}$ and then the cell concentration decreased.

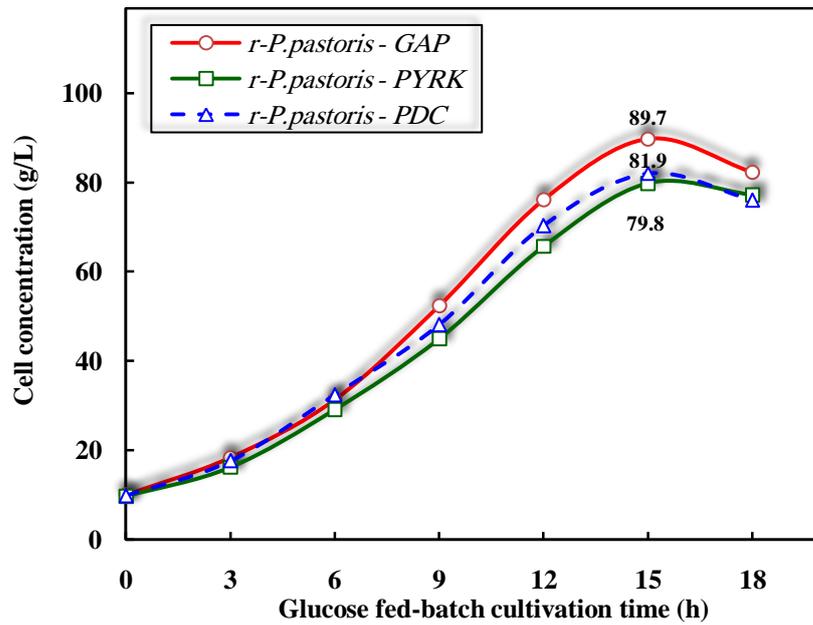


Figure 4.72. Cell concentration during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5).

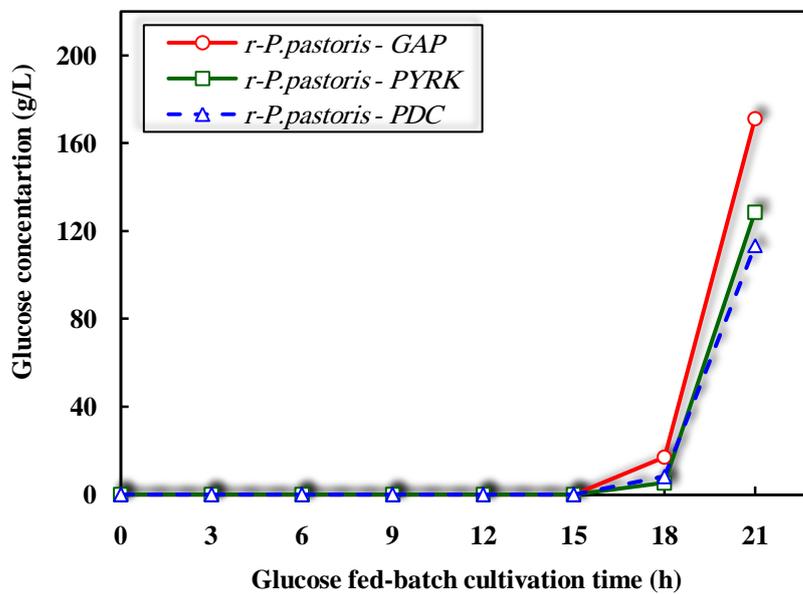


Figure 4.73. Residual glucose concentration during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5).

The overall yield of the cell on substrate ($\bar{Y}_{X/S}$) at $t=15$ h of the fed-batch phase was calculated to be 0.37, 0.37, and 0.43 g g^{-1} in BR-3, BR-4, and BR-5, respectively. The residual glucose concentration in the fermentation medium was also measured in order to verify the complete utilization of the inlet glucose by the cells. The glucose concentration profile in conducted experiments has been represented in Figure 4.73; the profiles pointed out no glucose accumulation until $t=15$ h and, then, upon decrease in cell concentration, glucose started to accumulate in the fermentation medium. In consequence, all the glucose fed to the bioreactor was consumed effectively by the cells until $t=15$ h.

In the scarcity of the oxygen, in order to recover the oxidation ability of the cell which has been diminished by slowing down the TCA cycle, yeast cells shift their metabolism towards fermentative pathway instead respiratory pathway to regenerate the NAD^+ required for proper and effective functioning of the glycolysis pathway. Ethanol is one of the end-products of this fermentative pathway. To verify the extent of shift in metabolism, the evolved ethanol during glucose fed-batch phase was also monitored. Concentration of the produced ethanol can be seen in Figure 4.74. In all three experiments, ethanol concentration showed a slight increase in $t=3$ h and, then, decreased until $t=6$ h, and started to another increase at $t=9$ h. However, the P_{GAP} -driven strain produced less ethanol compared to other two strains in contrast to its higher cell concentration. The maximum point of the ethanol production in three experiments was $t=18$ h with 28.0, 28.6, and 16.3 g/L for the strains express rhGH under control of P_{PYRK} , P_{PDC} , and P_{GAP} , respectively. As a kind of ‘Crabtree negative’ yeast, *P. pastoris* shifts the metabolism to ethanol production in low oxygen availability and the production of ethanol in these conditions refers to oxygen limitation conditions.

In order to examine the extracellular rhGH production in three runs, filtrate samples of the bioreactor experiments were subjected to SDS-PAGE analysis. The result of the SDS-PAGE is represented in Figure 4.75.

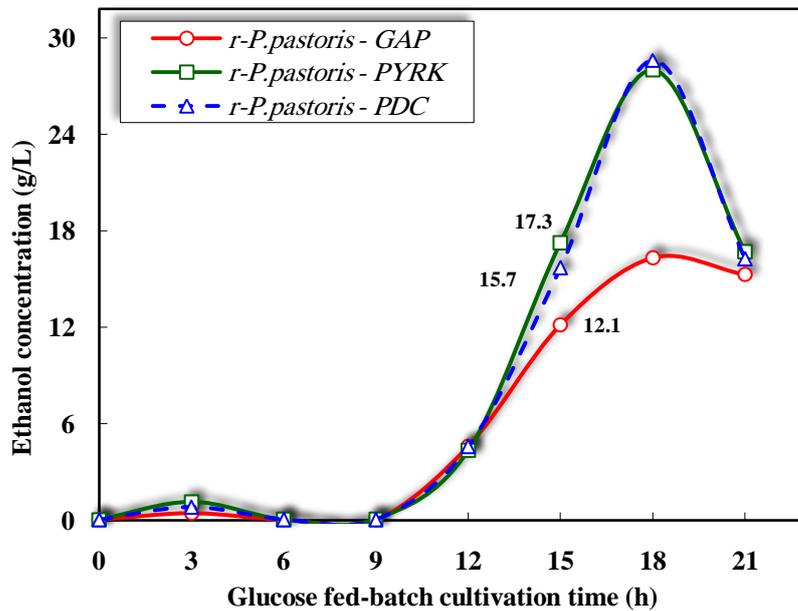


Figure 4.74. Produced ethanol concentration during glucose fed-batch phase of the fermentation in BR-3, BR-4, and BR-5.

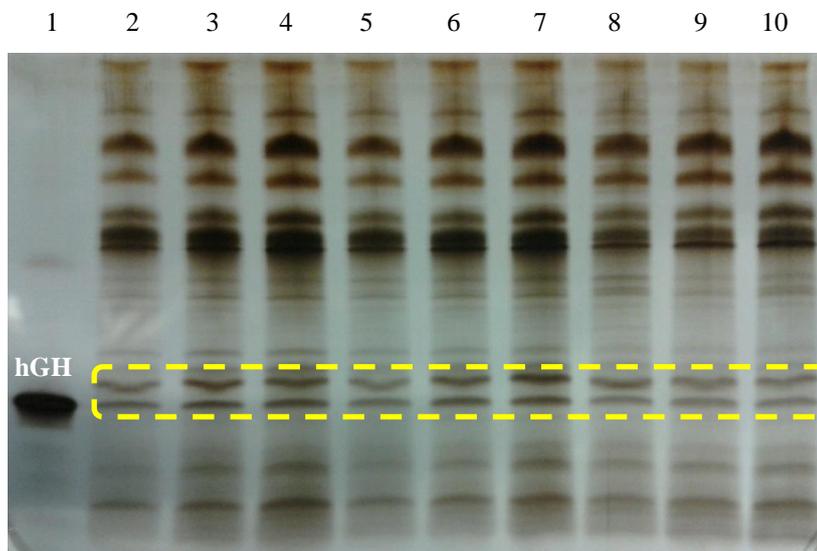


Figure 4.75. SDS-PAGE results related with BR-3, BR-4, and BR-5. Samples of glucose fed-batch phase with $\mu=0.15 \text{ h}^{-1}$. 1: hGH standard 0.05 g/L; 2, 3, 4: BR-3 sample for $t=9, 12,$ and 15 h after fed-batch start, respectively; 5, 6, 7: BR-4 sample for $t=9, 12,$ and 15 h after fed-batch start, respectively; 8, 9, 10: BR-5 sample for $t=9, 12,$ and 15 h after fed-batch start, respectively

The image refers to this point that there are two (expected) bands, i.e., mature hGH without any extra amino acid and the His-tagged hGH. Further verification of these two bands was accomplished by Western-blot analysis using $t=15$ h samples of three bioreactor experiments (Figure 4.76). The results confirmed that appeared two bands are hGH. The amount of the extracellular rhGH was measured using densitometric method and the results are available in Figure 4.77. In addition, productivity and specific production rate of rhGH (q_{rhGH}) in BR-3, BR-4, and BR-5 has been presented in Table 4.54. Furthermore the overall product yield on substrate (glucose) ($\bar{Y}_{P/S}$) was calculated as 0.87, 0.59, and 0.82 mg g^{-1} in the time of maximum product concentration for BR-3, BR-4, and BR-5, respectively; meanwhile, the overall yield of the cell on substrate ($\bar{Y}_{X/S}$) was calculated to be as 0.42, 0.37, and 0.55 for BR-3 ($t=12$ h), BR-4 ($t=15$ h), and BR-5 ($t=9$ h), respectively.

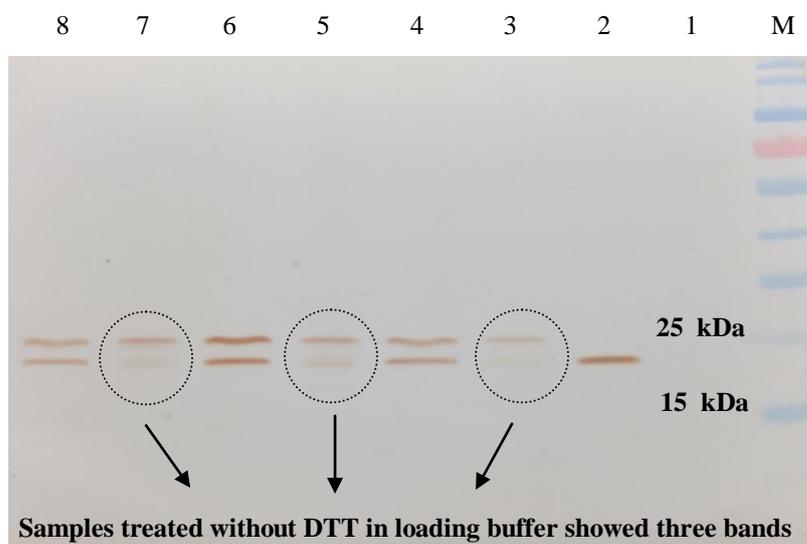


Figure 4.76. Western-blot results related with $t=15$ h samples of triple bioreactor experiments of *r-P. pastoris* strains corresponding to promoter constructs. M: protein marker, 2: hGH standard 2.5 mg/L, 3, 4: BR-3 sample without and with DTT, respectively; 5, 6: BR-4 sample without and with DTT, respectively; 7, 8: BR-5 sample without and with DTT, respectively. DTT: dithiothreitol.

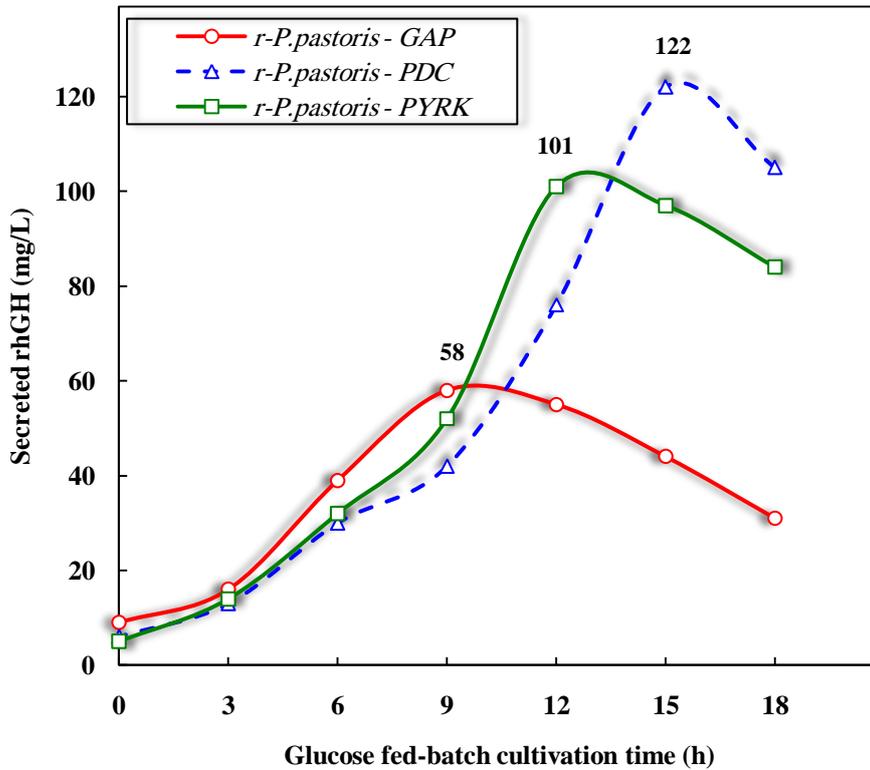


Figure 4.77. Concentration of the secreted rhGH during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5).

Table 4.54. The comparison of three bioreactor experiments related with promoter constructs based on productivity (mg/L.h) and specific production rate of rhGH (q_{rhGH}). The unit of specific productivity is mg/g.h.

Run		Glucose fed-batch cultivation time (h)						
		t=0	t=3	t=6	t=9	t=12	t=15	t=18
BR-3	Productivity	0.31	3.00	4.50	5.22	8.00	6.13	4.39
	q_{rhGH}	0.28	0.26	0.21	0.24	0.12	-0.04	-
BR-4	Productivity	0.37	2.33	4.00	4.00	5.83	7.73	5.50
	q_{rhGH}	0.17	0.21	0.15	0.14	0.19	0.073	-
BR-5	Productivity	0.44	3.00	4.83	5.67	4.00	2.47	1.33
	q_{rhGH}	0.28	0.36	0.22	0.06	-0.03	-0.05	-

The organic acids are carboxylic acids produced as intracellular metabolites inside the cell and are excreted to the extracellular medium regarding the metabolic state of the cell in order to either maintain the intracellular balance of the different pathways or adjust the intracellular pH (Nielsen et al., 2003). In addition to the abovementioned analysis, organic acids concentration in extracellular medium was also measured as explained in section 3.8.9. The results have been represented in Table 4.55.

The analysis results confirmed that due to the low oxygen availability conditions and, thus, incomplete TCA cycle, the cycle components are generally present in the extracellular medium in order to alleviate their excess intracellular amount where towards the end of the process they generally exhibit an increasing trend which emphasizes on the severe oxygen scarcity and imbalance between oxygen demand for respiratory chain and the supplied oxygen at the final hours of the process. The component of the TCA cycle found in comparatively larger amount was succinic acid. On the other hand, the components related with fermentative pathway were also present in the medium including: pyruvic acid, lactic acid, and acetic acid; formic acid was also detected in the medium which its relevant enzyme's gene (pyruvate formate lyase) has not yet been annotated in *P. pastoris*. These components have greater share compared to the TCA cycle elements which implies the preference of the fermentative pathway when oxygen deficiency occurs. In all three runs, towards the end of the cultivation, the total amount of the organic acids increased substantially which mainly included pyruvic acid, acetic acid and succinic acid.

Table 4.55. Concentration of the organic acids in the fermentation medium related with production of rhGH with *P. pastoris* using three different promoters. Units are (g/L).

	Cultivation time (h)	Oxalic acid	Formic acid	Pyruvic acid	Malic acid	Lactic acid	Acetic acid	Maleic acid	Citric acid	Fumaric acid	Succinic acid	Glutaric acid	Total
BR3	0	0.0800	0.0250	0.1425	0.0095	0.0645	0.1676						0.4891
	3	0.0009		0.0011		0.0496	0.0069			0.0039	0.8178		0.8802
	6	0.0686	0.0287	0.4073	0.0227	0.0593	0.2762		0.0451	0.0005	0.0529		0.9613
	9	0.1340	0.1139	0.1562	0.0302	0.1640	0.4076	0.0009	0.0558	0.0009	0.1707		1.2342
	12	0.0016		0.0560		0.5094				0.0043			0.5713
	15	0.4324		0.6822			1.5201				0.1931		2.8278
18	0.8881	0.9413	4.1873	0.0532	0.5401	2.7	0.0020	0.0879	0.0241	2.8132	0.0145		12.2517
BR4	0	0.0589	0.0072	0.0951	0.1090	0.0252	0.1868			0.0002	0.1216		0.6040
	3	0.0036		0.0264			0.0914			0.0013			0.1227
	6	0.1037	0.0753	0.4289	0.0351	0.0859	0.3364	0.0007	0.0278	0.0017	0.1346	0.0018	1.2319
	9	0.1631	0.0872	0.376	0.1080	0.2059	0.4656	0.0010	0.0590	0.0009	0.0951		1.5618
	12	0.2197	0.1057	0.3839	0.0436	0.1722	0.5893	0.0014	0.0975	0.0016	0.0858		1.7007
	15	0.3736	0.1231	0.9317	0.0816		1.0401			0.0012			2.5513
18	0.1102	0.1355	0.8878	0.0693		1.3991	0.0003	0.1136	0.0008	2.5913	0.0039		5.3118
BR5	0	0.0469	0.0080	0.0994	0.0097	0.0524	0.1404			0.0001	0.0904		0.4473
	3	0.0760	0.0180	0.3276	0.0764		0.0202	0.0003	0.0344	0.0003			0.5532
	6	0.0999	0.0370	0.4304	0.0320	0.0565	0.2925	0.0005	0.0218	0.0005	0.1392		1.1103
	9	0.1067	0.0478	0.0169	0.0517	0.0921	0.3764	0.0007	0.0475	0.0009	0.1258		0.8665
	12	0.2760		0.3911						0.0010	0.1950		0.8631
	15	0.6658	0.0718	1.2262	0.2204		1.0585			0.0019	0.2157		3.4603
18	0.7601	0.5600	4.2152	0.0890	0.8013	2.9314	0.0017	0.0923	0.0033	3.0234	0.0096		12.4873

4.10.2.1. mRNA level of hGH in fed-batch fermentations

The central dogma of molecular biology considers the information flow from DNA to protein through mRNA (Maier et al., 2009). On the other hand, the mRNA transcript level represents the strength of the promoter which is responsible for transcription of the gene of interest (i.e., hGH). Therefore, it is plausible to check the mRNA level of hGH in conducted bioreactor runs to extract any available correlation between secreted rhGH level (Figure 4.77) and its corresponding mRNA level. According to the cell concentration during fermentation (C_x) the samples of t=9, 12, and 15 h of the glucose fed-batch phases were analyzed for mRNA level. Total RNA isolation and cDNA synthesis was conducted as explained previously in sections 3.7.2.2.2 and 3.7.2.2.3, respectively. The qPCR experiments were accomplished by prepared cDNA samples as template instead genomic DNA (Table 3.22) and the samples were put in the instrument in the order represented in Table 4.56 by using *hGH* inner primers. The qPCR condition was similar to Table 3.23. The results of the triple runs can be seen in Table 4.57; the mRNA level change in selected samples has been depicted in Figure 4.78 where the average amounts of the copies in each time have been used. It is inferred from the mRNA level and secreted rhGH concentration at t=9 h that higher mRNA (transcript) level led to higher secreted rhGH (Table 4.58); hence at t=9 h the bioprocess just enters severe oxygen limitation. In consequence, in low oxygen availability and before severe oxygen limitation a positive correlation was obtained between mRNA level and secreted rhGH concentration. However, at t=12 h and t=15 h a clear-cut correlation between transcript level and protein level was not observable.

Considering the secretion of rhGH, in order to analyze the parameters that influence the protein level, ER capacity, folding and secretion processes should also be taken into consideration. The ER initiates UPR upon being exposed to any event that conduces to disrupt or overload its folding function (Baumann et al., 2010). Severe oxygen limitation can be regarded as an

environmental stress imposed on the cell which will, in turn, lead to expected cellular stress response and possible impairment of the protein secretion.

Table 4.56. Configuration of the samples for qPCR experiments related to hGH mRNA level measurement. Standards start from $1/10^5$ dilution (correspond to hGH-1) to $1/10^8$ dilution (correspond to hGH-4). Each total RNA sample along with *hGH* inner primers was assayed in duplicate. The run was performed in triplicate.

Well	Sample		Well	Sample	
1	hGH-1	Standard	21	5	(BR-3 t=15 h)
2	hGH-1	Standard	22	5	(BR-3 t=15 h)
3	hGH-2	Standard	23	10	(BR-4 t=15 h)
4	hGH-2	Standard	24	10	(BR-4 t=15 h)
5	hGH-3	Standard	25	15	(BR-5 t=15 h)
6	hGH-3	Standard	26	15	(BR-5 t=15 h)
7	hGH-4	Standard	27	dH ₂ O with hGH (NTC)	
8	hGH-4	Standard	28	dH ₂ O with hGH (NTC)	
9	3	(BR-3 t=9 h)			
10	3	(BR-3 t=9 h)			
11	8	(BR-4 t=9 h)			
12	8	(BR-4 t=9 h)			
13	13	(BR-5 t=9 h)			
14	13	(BR-5 t=9 h)			
15	4	(BR-3 t=12 h)			
16	4	(BR-3 t=12 h)			
17	9	(BR-4 t=12 h)			
18	9	(BR-4 t=12 h)			
19	14	(BR-5 t=12 h)			
20	14	(BR-5 t=12 h)			

Table 4.57. The mRNA level of hGH during rhGH expression under P_{FHE} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5) at selected times of glucose fed-batch phase of the fermentation. The results are for triplicate real time polymerase chain reaction (qPCR) runs where each sample was duplicated in each run. Empty cells are the discarded numbers in order to make $CV\% < 10$. $CV\%$ is the coefficient of variation percent which is defined as (standard deviation (SD) / mean) $\times 100$. Sample numbers are arbitrary attributed numbers to the specific selected samples of experiment; they are the numbers which have been used in related qPCR report ([Appendix R](#)).

$t=9$		$t=12$			$t=15$			Cultivation time (h)	
3 (BR-3)	8 (BR-4)	13 (BR-5)	4 (BR-3)	9 (BR-4)	14 (BR-5)	5 (BR-3)	10 (BR-4)	15 (BR-5)	Sample # (Run)
	38658	92272	73215		110345				
	38203	91402	69462	119591	114337				
53975	43497	114458	82923	141186	123778	61346	110028	169359	mRNA level (copy)
60629	42245	109067	83487	130182	119365	63763	111055	171446	
63440	46133	111449	85681	133207	120647	68109	116672	184452	
64733	45792	112156	85134	142559	125870	68793	111926	176994	
60694	41747	105134	79984	137345	119057	65503	112420	175563	Mean
4795.9	3341.0	10446.0	6876.3	12845.8	5822.5	3555.4	2938.7	6745.4	SD
7.9	8.0	9.9	8.6	9.4	4.9	5.4	2.6	3.8	CV%

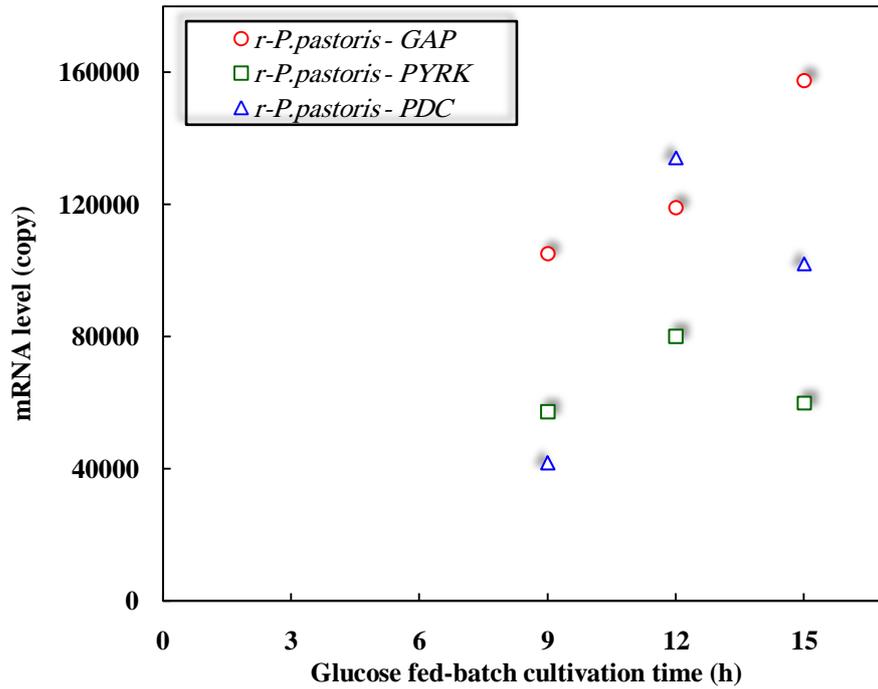


Figure 4.78. hGH mRNA level in selected samples of bioreactor runs. The mean of mRNA copy in Table 4.57 has been utilized.

Table 4.58. The summarized results of the conducted measurements related with extracellular rhGH level and its corresponding mRNA level. The promoter name represents the bioreactor run which rhGH expression was conducted utilizing corresponding promoter; P_{PYRK} represents BR-3, P_{PDC} represents BR-4, and P_{GAP} represents BR-5.

Cultivation Time (h)	t=9	t=12	t=15
rhGH production	$P_{GAP} > P_{PYRK} > P_{PDC}$	$P_{PYRK} > P_{PDC} > P_{GAP}$	$P_{PDC} > P_{PYRK} > P_{GAP}$
mRNA level	$P_{GAP} > P_{PYRK} > P_{PDC}$	$P_{PDC} > P_{GAP} > P_{PYRK}$	$P_{GAP} > P_{PDC} > P_{PYRK}$
Bioprocess DO state	Low oxygen availability	Severe oxygen limitation	Severe oxygen limitation

4.10.2.2. Proteases secretion in fed-batch fermentations

During the fermentations, the intracellular proteases were measured in the extracellular bioreactor fermentation media; where, the undesired proteases lead to the proteinaceous product degradation. Based on the optimal pH of the proteases activity, they can be classified into three groups as acidic-, neutral-, and basic- proteases. Since the experiments in the current research were accomplished at pH 5, only acidic protease activity was considered and measured. In the bioreactor experiments related with the promoters the activity of the acidic proteases in the bioreactor culture broth was measured as explained in section 3.8.10. The results have been presented in Figure 4.79.

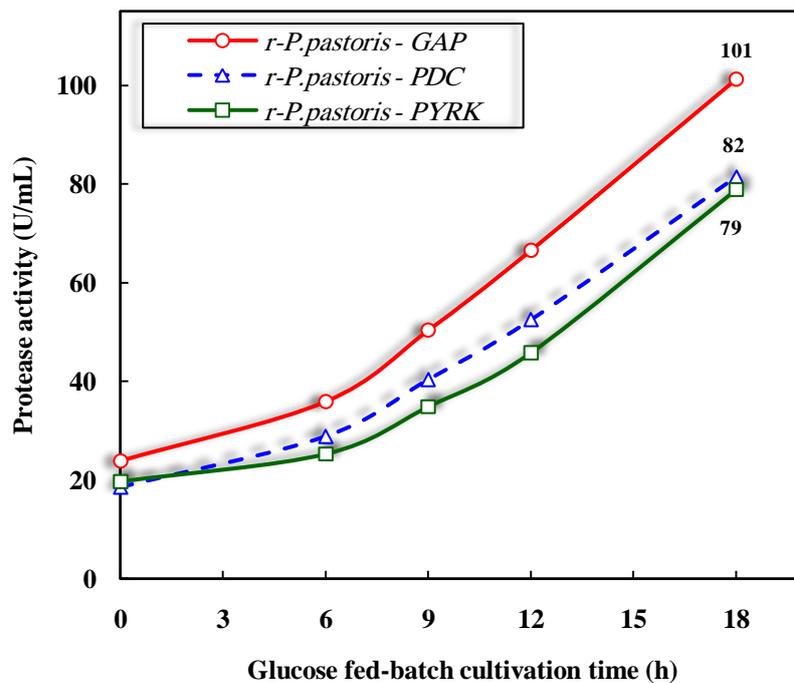


Figure 4.79. Protease activity during glucose fed-batch phase of the fermentation with corresponding *r-P. pastoris* strains of P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5).

It was understood that the trend was similar in three conducted experiments and the protease activity was increased towards the end of the process and

reached its maximum at $t = 18$ h. However, the maximum protease activity in the experiment related with P_{GAP} (BR-5) was approximately 20% higher than the experiments conducted with P_{PDC} (BR-4) and P_{PYRK} (BR-5).

4.10.2.3. Metabolic flux analysis

According to the available detailed biochemical reactions related with cell proliferation and rhGH production, distribution of the intracellular reaction rates (fluxes) reveals the behaviour of the rhGH-producing *P. pastoris* strains with different promoters. The comparison of the fluxes in three different experiments enables a more in-depth study of the metabolic burden that the host protein exerts on the producing strain and casts lights on the dark sides of the recombinant strain physiology when is subjected to oxygen limitation. The map of the postulated reaction network of *P. pastoris* has been represented in Figure 4.80 by emphasizing that some reactions were lumped into a single reaction by preserving the accuracy.

The accumulation rates of the organic acids, rhGH, and cell were used as inlets of the model in order to find the intracellular reaction rates. The period $0 < t \leq 3$ h was considered as period I where rhGH just started to increase and the data at $t = 3$ h virtually represents this period. In addition, the period $3 \text{ h} < t \leq t_i$ defines a period where cell synthesis enters declination and is regarded as period II; the behavior of the system at this period is approximated by data at t_i . In order to obtain the intracellular metabolic fluxes t_i was selected as $t = 12$ h, $t = 15$ h, and $t = 6$ h for BR-3, BR-4, and BR-5, respectively. The model-calculated fluxes have been presented in Table 4.59 in normalized form. The results revealed the pathways that were active or inactive, the pathways that were affected more by change in process conditions, and the energy state of the cell; generated ATP throughout the bioprocess in each period can be found in Table 4.60.

First, each run was analyzed individually and, then, a comparative analysis was conducted between triple runs (BR-3, BR-4, and BR-5) in order to dissect the details of the differences and similarities between the runs to

extract data related with the possible privilege of one of the selected promoters in expression of rhGH. At the end, the important nodes, Pyr and G3P, were considered in order to analyse the flux partitioning in them during changes occur in process conditions. All the fluxes, except total ATP generation, have been mentioned in normalized form with respect to glucose uptake rate and have been represented as percentage (%). Total ATP generation has been mentioned in the text normalized with respect to glucose and not in percentage form.

4.10.2.3.1. Metabolic flux analysis for rhGH expression under P_{PYRK}

The glucose uptake rate (R171) was 1.50 and 1.54 mmol/gDCW.h in period I and II, respectively. Glycolysis was active with increasing fluxes from period I to period II. In contrast, gluconeogenesis pathway was completely inactive (R6, R9, R11, and R13). However, as the process proceeded, the share of the pentose phosphate pathway (PPP) decreased which implies to decreased need for reducing equivalent, NADPH, which is the direct product of the oxidation phase of the PPP and used in anabolic reactions.

Considering anaplerotic reactions, because of the excess amount of the pyruvate which was confirmed by presence of pyruvic acid in extracellular medium (Table 4.55) the conversion of Mal to Pyr (R34) was inactive in both periods. Furthermore, by progression of the process and increased production of host protein (rhGH) the need for aromatic amino acids especially tyrosine and phenylalanine, which possess a high percentage in rhGH composition (Table 4.61), led to need for PEP supplied via conversion of OA (R36). The higher fluxes of R57-R62, originate from OA, in period I compared to period II revealed that, in period I, OA is required in higher amounts which should be supplied also from Mal (R46) and in order to cope with the high need; by decreasing the need for OA in period II, the amount of the Mal which was supplied from Fum (R44) was enough to fulfill the requirements and, thus, R 174 which was active in period I, became inactive.

Table 4.59. Normalized flux distributions (with respect to glucose uptake rate) of rhGH production by *P. pastoris* with different promoters in low oxygen availability. In BR-3, the glucose uptake rate was 1.50 and 1.54 at period I and II, respectively. In BR-4, the glucose uptake rate was 1.52 and 1.51 at period I and II, respectively. In BR-5, the glucose uptake rate was 1.62 and 1.19 at period I and II, respectively. The uptake rate unit is mmol/gDCW.h.

R #	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	29.34	0.00	0.00	0.00
7	3.67	48.51	0.00	0.00	0.00	0.00
8	53.60	70.65	35.33	45.80	47.16	36.47
9	0.00	0.00	0.00	0.00	0.00	0.00
10	130.80	149.16	101.71	113.83	116.54	89.33
11	0.00	0.00	0.00	0.00	0.00	0.00
12	117.20	131.56	90.99	110.99	88.09	69.08
13	0.00	0.00	0.00	0.00	0.00	0.00
14	111.20	129.81	85.07	108.21	82.47	61.43
15	9.00	12.53	0.00	0.00	0.00	0.00
16	91.53	47.92	124.61	97.75	95.62	93.95
17	49.93	22.21	64.67	45.80	47.16	36.47
18	0.00	0.00	0.00	0.00	0.00	0.00
19	26.47	14.48	33.82	23.63	24.94	20.08
20	0.00	0.00	0.00	0.00	0.00	0.00
21	26.47	14.48	33.82	23.63	24.94	20.08
22	0.00	0.00	0.00	0.00	0.00	0.00
23	23.47	7.73	30.86	22.17	22.16	16.30
24	0.00	0.00	0.00	0.00	0.00	0.00
25	26.47	14.48	33.82	23.63	24.94	20.08
26	0.00	0.00	0.00	0.00	0.00	0.00
27	25.40	13.70	1.64	55.99	21.11	17.39
28	0.00	0.00	1.64	1.52	21.11	32.02
29	25.40	13.70	0.00	54.47	0.00	0.00
30	0.00	0.00	0.00	0.00	21.11	32.02
31	7.80	0.00	13.09	16.28	18.77	35.55
32	0.00	66.04	0.00	0.00	0.00	0.00
33	0.00	0.00	0.00	0.00	0.00	0.00
34	0.00	0.00	0.00	0.00	0.00	0.00
35	32.20	24.03	0.00	20.05	0.00	21.34
36	0.00	11.88	0.00	0.00	0.00	0.00
37	11.60	8.64	4.14	0.00	13.95	28.99
38	11.60	8.64	4.14	0.00	13.95	28.99
39	11.60	8.64	4.14	0.00	13.95	28.99
40	0.00	0.00	0.00	2.05	0.00	32.69

Table 4.59. Normalized flux distributions (with respect to glucose uptake rate) of rhGH production by *P. pastoris* with different promoters in low oxygen availability (**Continued**).

R #	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
41	0.00	0.00	0.00	7.54	0.00	32.69
42	0.00	0.00	0.00	0.00	0.00	0.00
43	0.00	0.00	0.00	7.54	0.00	40.84
44	26.40	24.22	41.05	55.00	43.40	92.44
45	0.00	0.00	0.00	0.00	0.00	0.00
46	37.07	24.22	81.71	55.00	83.40	92.44
47	0.00	0.00	0.00	0.00	0.00	0.00
48	13.60	17.60	10.72	2.78	28.46	20.25
49	10.93	15.58	8.09	1.52	25.99	16.89
50	0.07	0.06	0.07	0.04	0.06	0.08
51	5.20	3.90	5.13	2.45	4.81	6.55
52	6.40	4.74	6.32	2.98	5.93	8.07
53	3.00	2.27	2.96	1.39	2.78	3.78
54	3.40	2.53	3.36	1.59	3.15	4.29
55	15.13	11.30	26.12	28.39	23.52	37.39
56	0.73	0.58	11.91	21.64	0.68	19.33
57	57.67	28.05	77.30	74.72	69.14	84.71
58	1.13	0.84	1.12	0.53	1.11	1.43
59	24.93	3.70	30.00	24.29	19.88	25.13
60	24.33	3.25	29.47	24.02	19.32	24.37
61	2.20	1.62	2.17	6.49	2.04	2.77
62	0.80	0.58	0.79	0.40	0.74	1.01
63	3.00	6.82	2.96	1.39	2.78	3.78
64	1.53	6.56	1.51	1.26	1.42	1.93
65	1.13	0.00	1.12	0.00	1.05	1.43
66	0.33	0.26	0.33	0.13	0.31	0.42
67	105.20	70.97	125.33	107.21	129.07	147.73
68	32.33	24.09	46.97	50.23	48.95	59.16
69	1.87	1.43	1.84	0.86	1.73	2.35
70	3.60	2.66	7.43	15.49	3.33	4.54
71	1.87	1.36	5.66	14.69	1.73	2.35
72	3.27	2.40	3.22	1.52	6.17	4.12
73	0.00	0.00	0.00	0.00	0.00	0.00
74	0.00	0.00	3.88	13.83	0.00	0.00
75	0.00	0.00	0.00	0.00	0.00	0.00
76	0.00	0.00	0.00	0.00	0.00	0.00
77	0.00	0.00	0.00	0.00	0.00	0.00
78	0.00	0.00	0.00	0.00	0.00	0.00
79	0.00	0.00	0.00	0.00	0.00	0.00
80	0.00	0.00	0.00	0.00	0.00	0.00

Table. 4.59. Normalized flux distributions (with respect to glucose uptake rate) of rhGH production by *P. pastoris* with different promoters in low oxygen availability (**Continued**).

R #	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
81	0.00	0.00	11.18	21.31	0.00	18.32
82	0.00	0.00	0.00	5.49	0.00	0.00
83	0.00	0.00	0.00	0.00	0.00	0.00
84	0.00	5.39	0.00	0.53	0.00	0.00
85	0.00	0.00	0.00	0.00	0.00	0.00
86	0.00	0.00	0.00	0.00	0.00	0.00
87	20.00	0.00	25.13	16.48	15.25	18.91
88	0.00	0.00	0.00	0.00	0.00	0.00
89	0.00	4.55	0.00	0.00	0.00	0.00
90	0.00	0.00	0.00	0.00	0.00	0.00
91	0.00	0.00	0.00	0.00	3.15	0.00
92	12.27	9.16	12.11	5.76	20.86	15.46
93	13.00	9.74	24.01	27.40	21.54	34.79
94	0.73	0.58	0.72	0.33	0.68	0.92
95	0.20	0.13	0.20	0.07	0.19	0.25
96	12.27	9.16	23.29	27.00	20.86	33.78
97	0.00	0.00	0.00	0.00	0.00	0.00
98	0.27	0.19	0.26	0.13	0.25	0.34
99	0.20	0.13	0.20	0.07	0.19	0.25
100	12.27	9.16	23.29	27.00	20.86	33.78
101	29.40	21.95	51.32	56.32	46.23	73.78
102	1.80	1.30	1.78	0.86	1.67	2.27
103	1.00	0.78	0.99	0.46	0.93	1.26
104	0.87	0.65	0.86	0.40	0.80	1.09
105	0.87	0.65	0.86	0.40	0.80	1.09
106	0.67	0.45	0.66	0.26	0.62	0.84
107	0.00	0.00	0.00	0.00	0.00	0.00
108	0.33	0.26	0.33	0.20	0.31	0.42
109	0.27	0.19	0.26	0.13	0.25	0.34
110	0.20	0.13	0.20	0.07	0.19	0.25
111	0.20	0.13	0.20	0.07	0.19	0.25
112	15.33	3.96	17.83	10.72	17.35	16.22
113	0.80	0.58	0.79	0.40	0.74	1.01
114	25.33	18.83	24.93	11.85	42.41	31.85
115	25.33	18.83	24.93	11.85	42.41	31.85
116	226.33	140.78	277.63	205.49	297.65	368.07
117	0.00	0.00	0.00	13.04	0.00	40.84
118	114.33	86.88	111.78	159.43	108.77	173.19
119	0.00	0.00	0.00	0.00	0.00	0.00
120	120.87	90.06	134.21	91.66	159.44	170.59

Table. 4.59. Normalized flux distributions (with respect to glucose uptake rate) of rhGH production by *P. pastoris* with different promoters in low oxygen availability (**Continued**).

R #	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
121	0.00	0.00	0.00	0.00	0.00	0.00
122	0.67	0.52	0.66	0.33	0.62	0.84
123	18.67	13.90	40.72	51.29	26.79	60.25
124	68.67	51.17	35.99	37.13	62.22	67.31
125	0.00	0.00	0.00	0.00	0.00	0.00
126	168.60	95.39	191.78	148.51	200.31	199.41
127	4.73	3.51	4.67	2.18	4.38	5.97
128	7.27	5.45	7.17	3.38	6.73	9.16
129	0.67	0.52	0.66	0.33	0.62	0.84
130	0.53	0.39	0.53	0.26	0.49	0.67
131	0.33	0.26	0.33	0.13	0.31	0.42
132	0.20	0.13	0.20	0.07	0.19	0.25
133	0.05	0.04	0.05	0.02	0.05	0.06
134	0.20	0.13	0.20	0.07	0.19	0.25
135	0.13	0.13	0.13	0.07	0.12	0.17
136	0.13	0.06	0.13	0.06	0.12	0.17
137	0.07	0.05	0.07	0.03	0.06	0.08
138	0.06	0.04	0.06	0.03	0.05	0.07
139	11.33	8.44	11.18	5.29	10.49	14.29
140	0.0013	0.0007	0.0006	0.0005	0.0010	0.0008
141	0.00	0.00	0.00	0.00	35.56	0.00
171	100.00	100.00	100.00	100.00	100.00	100.00
172	0.00	0.00	0.00	0.00	0.00	0.00
173	0.00	0.00	0.00	0.00	0.00	14.54
174	10.67	0.00	40.66	0.00	40.00	0.00
175	0.00	0.00	0.00	0.00	0.00	0.00
176	0.00	0.00	0.00	0.00	0.00	0.00

Table 4. 60. Variation in generated ATP throughout the bioprocess.

Amount of generated ATP (mmol/gDCW.h)						
R#	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
10	1.962	2.297	1.546	1.72	1.888	1.063
14	1.668	1.999	1.293	1.635	1.336	0.731
41	0	0	0	0.114	0	0.389
97	0	0	0	0	0	0
106	0.01	0.007	0.01	0.004	0.01	0.01
108	0.005	0.004	0.005	0.003	0.005	0.005
116	6.79	4.336	8.44	6.21	9.644	8.76
117	0	0	0	0.197	0	0.486
Total	10.435	8.643	11.294	9.883	12.883	11.444
Normalized*	6.957	5.612	7.430	6.540	7.952	9.616

Percent of generated ATP (%)						
R#	BR-3		BR-4		BR-5	
	Period I	Period II	Period I	Period II	Period I	Period II
10	18.80	26.58	13.69	17.40	14.65	9.29
14	15.98	23.13	11.45	16.54	10.37	6.39
41	0.00	0.00	0.00	1.15	0.00	3.40
97	0.00	0.00	0.00	0.00	0.00	0.00
106	0.10	0.08	0.09	0.04	0.08	0.09
108	0.05	0.05	0.04	0.03	0.04	0.04
116	65.07	50.17	74.73	62.84	74.86	76.55
117	0.00	0.00	0.00	1.99	0.00	4.25
Total	100	100	100	100	100	100

* Normalized with respect to glucose uptake rate.

Regarding branches from Pyr, R27 was always active with fluxes of 25.40 and 13.70 in period I and II, respectively. Produced Acet was completely devoted to ethanol production (R29) and neither acetic acid (R28) nor AcCoA (R30) was produced.

The TCA cycle was not completed in both of the considered periods and SucCoA was not synthesized from α KG. However, Fum to Mal (R44) and Mal to OA (R46) were both active to provide the cell with required amounts of OA as an important precursor for biosynthesis of amino acids.

The required Fum for resuming TCA cycle was provided by R92 and R100 regarding their high similar fluxes of 12.27 and 9.16 in period I and period II, respectively.

ATP generation normalized flux (with respect to glucose uptake rate) decreased from 6.957 in period I to 5.612 in period II which can be attributed mainly to decrease in oxidative phosphorylation; in total, 37.8% decrease in oxidative phosphorylation was observed.

The biomass synthesis flux (R139) was affected by decrease in ATP supply from period I to period II and showed a decreasing trend from 11.33 to 8.44 as expected (Buamann et al., 2010). rhGH secretion flux (R140) decreased from 0.0013 in period I to 0.0007 in period II. In contrast, R14 catalyzed by pyruvate kinase enzyme exhibited a 16.7% increase in flux from 111.20 in period I to 129.81 in period II. The fluxes variations of R14 and R140 from period I to period II, both of which are virtually the result of the transcription activity of P_{PYRK} , were not parallel presumably as a result of the other factors play role in enzyme-catalyzed reactions. The increased flux of pyruvate kinase-catalyzed reaction (R14) may be partly attributed to the decrease in the intracellular citrate level by decreasing the flux of R37 from 11.60 in period I to 8.64 in period II; it relieves the allosteric inhibition effect exerted by accumulated citrate on pyruvate kinase enzyme. Alternatively, the increased amount of the fructose 1,6-biphosphate, via increased flux of R8 from 53.60 in period I to 70.65 in period II, which is a strong activator of pyruvate kinase enzyme can be another explanation (Hackett et al., 2016).

4.10.2.3.2. Metabolic flux analysis for rhGH expression under P_{PDC}

The glucose uptake rate decreased very negligible from 1.52 mmol/gDCW.h in period I to 1.51 mmol/gDCW.h in period II. Regarding to the entrance of the majority of G6P to PPP, the flux of G6P to F6P (R7) was zero in both periods and the glycolysis pathway was interrupted at this point; glycolysis was resumed from F6P regarding ever-present flux through its supplying reactions in PPP. In gluconeogenesis pathway the only active reaction was

F6P to G6P (R6) in period I which was the result of the exceptional high flux through the PPP and as the result of the higher flux of the two reactions R23 and R25; owing to the increased flux in upstream, the flux of the R14 increased from 85.07 in period I to 108.21 in period II which implies to increased production of pyruvate. The flux of the formation of AcCoA_m from Pyr (R15) was zero in both periods. The low oxygen availability through current study made the TCA cycle work inefficiently and, thus, the metabolism directed towards synthesis of by-products of the fermentation through branched pathways from Pyr. Since R30 flux was zero in both periods, AcCoA was supplied through the increasing flux of R82 from 0 (period I) to 5.49 (period II) and led to the fueling of AcCoA_m and, therefore, the Pyr was channeled into the TCA which may be one of the reasons that *P. pastoris* is a “Crabtree-negative” yeast (Agarwal et al., 2013). Presence of the ethanol in the extracellular medium validated redirection of the metabolism towards the by-products in Pyr node.

Taking into account the branches from Pyr, the reactions responsible for production of lactate and formate were inactive (R32 and 172) in both periods. The flux of the transformation of Pyr to Acet (R27) which is catalyzed by pyruvate decarboxylase enzyme increased strikingly from 1.64 in period I to 55.99 in period II. Pyruvate decarboxylase enzyme is regulated allosterically and is activated by its substrate, Pyr (Agarwal et al., 2013); increased flux of R14 from 85.07 in period I to 108.21 in period II and accumulation of Pyr due to the lower consumption via TCA cycle as a result of oxygen limitation increased the flux of R27. Excess Pyr was redirected towards fermentative pathway in order to be metabolized.

Regarding anaplerotic reactions, because of the excess amount of the pyruvate confirmed by presence of pyruvic acid in the extracellular medium, the conversion of Mal to Pyr (R34) was inactive in both time points.

The conversion of Pyr to OA (R35) and Pyr to Mal (R174) were both expected to be active because of preparation of the substrate for PEP. In period I, Pyr to Mal (R174) ensured the fulfillment of the needs for OA

production and it was obvious from the fluxes of the R46 which was higher in period I due to the provision of Mal from Pyr; in period II, Pyr, without any intermediate, supplied directly the supply of OA (R35).

In TCA cycle, two different behaviours were observed. In period I, where the oxygen is more or less available and the process has not yet entered the severe oxygen limitation situation, the cycle proceeded until important intermediate, α KG, namely the active reactions were R37, 38, and 39. Afterwards, the active reactions in the metabolic circuit were R44 and 46 which end with the other important intermediate OA. In any circumstance they should serve as precursors of several amino acids. Fum required for restarting the cycle was again prepared via R92 and R100. The ineffective function of the TCA cycle limited the amount of the α KG where it was just recruited as precursor and was hardly found in extracellular medium according to the organic acids concentrations in extracellular medium (Table 4.55). On the other hand, interestingly, in period II the reactions end with α KG all were dormant (R37, 38, and 39) and the cycle did not generate precursor for glutamic family amino acids. Thus, α KG should be provided from other pathways; the increased flux of R56 from 11.91 in period I to 21.64 in period II can be a plausible explanation for generation of α KG. The lower rate of the transport of NADH from mitochondrion to cytoplasm (R126) was a rational representation of the less efficient function of TCA cycle in period II, i.e., severe oxygen limitation situation. The reactions R41, 43, 44, and 46 all worked in concert to prepare OA for cell vital reactions as a normal behavior.

ATP generation normalized flux (with respect to glucose uptake rate) experienced a decrease from 7.430 in period I to 6.540 in period II due to the 21.3% decrease in oxidative phosphorylation from 277.63 in period I to 218.53 in period II. The contribution of the oxidative phosphorylation to the cell energy content decreased about 10% from period I to period II by changing its amount from 5.553 to 4.240 (normalized but not as %). However, in period II the ATP production as the result of the electron

donation from FADH₂ (R117) also had a share in total ATP generation. In period II, FADH₂ was produced via R43 which is catalyzed by the succinate dehydrogenase as the only membrane-bound enzyme in the TCA cycle which resides in the inner membrane of the mitochondrion and participates in respiratory chain by aiding electron flow (Nelson and Cox, 2005).

The biomass synthesis flux (R139) diminished about 53% from 11.18 in period I to 5.29 in period II which was the expected outcome of the less available energy represented in the form of decreased ATP generation flux (R116) mentioned above. Furthermore, rhGH secretion flux (R140) decreased 17% from 0.0006 in period I to 0.0005 in period II. Contrary to rhGH synthesis flux, the flux of the reaction R27 catalyzed by pyruvate decarboxylase enzyme showed approximately 34-fold increase from period I to period II (from 1.64 to 55.99); although rhGH was expressed using the promoter of the pyruvate decarboxylase enzyme the fluxes variation were not parallel in R27 and R140 probably due to the other parameters influence enzyme-catalyzed reactions; as mentioned earlier, allosteric regulator(s), like Pyr itself, can enhance the activity of the pyruvate decarboxylase enzyme which is not the case in the rhGH synthesis reaction.

4.10.2.3.3. Metabolic flux analysis for rhGH expression under P_{GAP}

Glucose uptake rate decreased from 1.62 mmol/gDCW.h in period I to 1.19 mmol/gDCW.h in period II. The glycolysis pathway started from F6P to G3P (R8) as a consequence of the high inlet flux to PPP (R16) and subsequent emergence as F6P (R23 and 25). The gluconeogenesis pathway was inactive completely in this run. The zero flux of the transformation of Pyr to AcCoA_m (R15) devolved the fueling TCA cycle to cytoplasmic AcCoA (R31); this is supposed to be the direct result of the accumulation of Pyr and shifting the metabolism towards fermentation because of the low availability of oxygen.

In branches from Pyr, acetate was produced at both periods via R28; however, flux of the reaction increased in period II due to the extra flux came into the scene by utilization of the ethanol (R173). AcCoA from Ac

(R30) was actively generated in both of the periods with higher flux in period I; AcCoA is used for production of lipids required for cell especially membrane synthesis and the increase in fluxes of R128 to R138 from period I to period II confirmed this point.

The explanation about the anaplerotic reactions in this case is similar to BR-4 since they both followed a similar trend and R174 was active in period I and R35 was active in period II where both finally lead to OA for either biosynthesis of amino acids or transformation to PEP.

Analysis of the TCA cycle revealed that in both of the inspected periods, the reactions were active till α KG. In period I the cycle was completed where in period II the cycle was interrupted and R44 and R46 resumed it until production of intermediate OA. The required Fum for R44 originated R100 which its flux increased from 20.86 in period I to 33.78 in period II.

Energy currency of the cell, in the form of total ATP flux (with respect to glucose uptake rate), changed from 7.952 to 9.616 from period I to period II (21% increase) mainly due to the 37.4% increase of the oxidative phosphorylation flux from 297.65 in period I to 408.91 in period II. Both of these periods are in the low oxygen availability regime prior to severe oxygen limitation. However, the part of the ATP which was generated from substrate-level oxidations (ATP_{SP}) in R10 and R14 decreased 24.2% from 199.01 in period I to 150.76 in period II.

The biomass synthesis reaction (R139) experienced a 36.2% increase from 10.49 in period I to 14.29 in period II; the 37.4% increase in ATP generation flux through oxidative phosphorylation confers the microorganism the ability of more biomass synthesis while the host protein (rhGH) synthesis flux has a decreasing trend. rhGH secretion reaction flux decreased 20% from 0.0010 in period I to 0.0008 in period II which was in accordance with the decreased flux (23.4%) of the R10 from 116.54 in period I to 89.33 in period II. R10 entails the activity of the glyceraldehyde-3-phosphate dehydrogenase enzyme. Therefore, the reactions which have

been based on the transcriptional activity of the same promoter (P_{GAP}) worked in parallel.

4.10.2.3.4. Comparative analysis of the metabolic fluxes distributions

At the start of the central metabolism after glucose uptake and its subsequent phosphorylation (R171), the main discrepancy between three runs happened at redirection of the flux through PPP where R16 in BR-4 and BR-5 had the whole flux and R7 was inactive and the PPP reentered the flux to glycolysis via F6P. In all cases the PPP flux through R16 decreased towards the end of the process while the decrease in BR-5 was very slight. Therefore, entering severe oxygen limitation conditions, the case in BR-3 and BR-4, and decreased level of oxidative stress lead to lower flux through PPP which one of its responsibilities is to produce NADPH for conquering oxygen free radicals. In BR-4 and BR-5 prior to the maximum production time, the flux in PPP is higher compared to BR-3 which is indicator of the increase of the reductive power of the cell and better implementation of the anabolic (biosynthetic) reactions.

The glycolytic flux (R8, R10, R12, and R14) in BR-3 was higher than BR-4 and BR-5. In BR-3, the R8 fluxes were 53.60 and 70.65 in period I and II, respectively. Period I flux (53.60) was 1.51-fold and 1.14-fold higher than period I fluxes of BR-4 and BR-5, respectively. Further, period II flux (70.65) was 1.54-fold and 1.94-fold higher than period II fluxes of BR-4 and BR-5, respectively. In BR-3, the R10 fluxes were 130.80 and 149.16 in period I and II, respectively. Period I flux (130.80) was 1.29-fold and 1.12-fold higher than period I fluxes of BR-4 and BR-5, respectively. Further, period II flux (149.16) was 1.31-fold and 1.67-fold higher than period II fluxes of BR-4 and BR-5, respectively. In BR-3, the R12 fluxes were 117.20 and 131.56 in period I and II, respectively. Period I flux (117.20) was 1.29-fold and 1.33-fold higher than period I fluxes of BR-4 and BR-5, respectively. Further, period II flux (131.56) was 1.18-fold and 1.90-fold higher than period II fluxes of BR-4 and BR-5, respectively. In BR-3, the

R14 fluxes were 111.20 and 129.81 in period I and II, respectively. Period I flux (111.20) was 1.31-fold and 1.35-fold higher than period I fluxes of BR-4 and BR-5, respectively. Further, period II flux (129.81) was 1.20-fold and 2.11-fold higher than period II fluxes of BR-4 and BR-5, respectively. In the case of BR-3 and BR-4 the glycolytic flux increased during process from period I to period II while in BR-5 the trend was different.

Working at severe oxygen limitation situation in period II for BR-3 and BR-4 confronted the cell with decreased provision of ATP through oxidative phosphorylation (ATP_{OP}) due to the incomplete TCA cycle; the ratio of $\left(\frac{ATP_{OP}}{ATP_{SP}}\right)$ for BR-3 decreased from 1.89 in period I to 0.99 in period II corresponding to 1.91-fold decrease. While in the case of BR-4, $\left(\frac{ATP_{OP}}{ATP_{SP}}\right)$ decreased from 2.96 in period I to 1.84 in period II corresponding to 1.61-fold decrease. ATP shortage was compensated to some extent by increasing the flow through glycolytic pathway (R10 and 14) and, in consequence, increased ATP generation share with substrate-level phosphorylation. It was a consistent outcome with previously reported results (Baumann et al., 2010). The inactive reaction of transformation of Pyr to AcCoA was another different point between triple runs. Low oxygen availability imposed the cell to go to the branch from Pyr towards by-products in order to oxidize glucose and extract its energy for the cell proliferation and rhGH production and, thus, AcCoA was generated as one of the end-products. The excess produced AcCoA can have inhibitory effect on pyruvate dehydrogenase complex (Nelson and Cox, 2005) which undertakes the conversion of Pyr to AcCoA and, consequently, the required AcCoA for TCA cycle was supplied with R31 in the cases where the R15 flux was zero.

The maintenance energy just appeared in period I of BR-5. It indicates the ATP in this period was utilized in processes other than product formation such as preserving electrical potential across membranes, adjusting pH gradient, or intracellular futile cycles (Nielsen et al., 2003). The surveillance of the energy state of the cell at different periods in triple runs

unveiled that period I of BR-5, after subtracting the maintenance energy, was better compared to other first periods based on the normalized total ATP generated (Table 4.60). However, BR-5 did not face severe oxygen limitation even in period II and, thus, its ATP level increased. Regarding energy aspect, BR-4 possessed a more favorable situation compared to BR-3 either in period I (before severe oxygen limitation) or period II (during severe oxygen limitation).

Dissimilar trends observed between rhGH secretion flux under selected promoters and the flux of the reactions catalyzed by enzymes which are under the control of the same promoters that transcribe rhGH, can be analysed, in the simplest form, regarding the single-substrate enzyme-catalyzed reactions explained by Michaelis-Menten kinetics (Nelson and Cox, 2005) according to equation (4.1):

$$r_{er} = \frac{r_{er.max} C_s}{K_m + C_s} \quad (4.1)$$

Where, r_{er} is the enzymatic reaction rate, $r_{er.max}$ is the maximum rate of the enzymatic reaction, C_s is the substrate concentration, and K_m , Michaelis-Menten constant, is a constant specific to the enzyme-substrate couple. In addition, advanced analyses incorporate the allosteric regulators and products into the reaction rate still in a Michaelis-Menten-style reaction equation. Apparently, equation (4.1) relates the enzymatic reaction rate to the substrate concentration (C_s); however, the $r_{er.max}$ bears the total enzyme concentration used in the reaction. Therefore, ignoring products and allosteric effectors, any change in the reaction rate (flux) should not be just attributed to one of these parameters (C_s and/or total available enzyme) and their synergistic effect should be taken into consideration. The flux of the reactions which prepare the substrate for desired reactions should be considered; the increase or decrease of their rates will affect the substrate amount available for desired reactions. Therefore, any change of the flux in

reaction could be either the result of the substrate concentration and/or total available enzyme concentration and cannot be attributed solely to the enzyme amount which represents its corresponding promoter activity. In addition, the enzymes which catalyze reactions are intracellular whereas rhGH is secreted to the extracellular medium and follows a different fate. Consequently, it would not be rational to expect that rhGH secretion flux acts parallel to the reaction catalyzed by the enzyme of its promoter.

Based on the reaction engineering principles every biochemical reaction in the intracellular reaction network can occur through an enzyme-catalytic reaction mechanism with three consecutive steps: i) adsorption of the reactants (metabolites) onto the active sites of the biochemical catalyst (enzyme), ii) biochemical reaction in the active site of the enzyme, and iii) desorption of the reaction products from the surface of the enzyme. In consequence, in total, it is the expression of the related genes which encode the enzymes (enzyme activity) further their turn-over numbers, and flow (flux) of carbon molecules (or metabolite concentrations) that determine the magnitude of the metabolic fluxes of the intracellular reactions. However, Hackett et al. (2016) has recently reported in their work for a yeast, the limiting parameter for the magnitude of fluxes is primarily the metabolite concentration or carbon molecule flux rather than the enzyme concentration if the expression of the gene encoding the related enzyme occurs to an extent in the cell, which is indeed noteworthy. The allosteric regulators bind, irreversibly, to the regulatory (allosteric) enzymes and affect their activity through conformational change (Nelson and Cox, 2005); therefore, they are post-translational modulators of the enzyme activity which just influence the protein after being transcribed and translated. The regulatory effect on promoters which may affect the amount of the expressed protein cannot be explained or analysed by their action.

About branching in Pyr node, regarding the pathways led to by-products (R27, R32, and R172) the lowest flux belonged to period I of BR-4 and the highest flux belonged to period II of BR-3. The flux towards lactate

(R32) as one of the main by-products could just be seen in period I of BR-3. R27 was the only reaction which was active throughout investigated periods in all three runs and, thus, implied to the unsuccessful transfer of electrons via respiratory chain to final electron acceptor, i.e., oxygen, in TCA cycle because of low oxygen availability which led to imbalance between oxygen demand and supply; thus, NADH could not transfer its electron to the oxygen and NAD^+ required for proper functioning of the glycolysis pathway could not be regenerated. The required NAD^+ was obtained by alternative pathway from Pyr towards Acet. The reactions in this branch all include incorporation of cofactors such as ATP, NADH, or NADPH. Consequently, the cell based on the metabolic state and its energy needs regulates the reactions in a way for example to conquer ATP deprivation or to adjust redox balance represented by $[\text{NADH}/\text{NAD}^+]$ and $[\text{NADPH}/\text{NADP}^+]$; R30 was inactive in both BR-3 and BR-4 and was expected acetate in the broth of BR-4 regarding to its accumulation. This reaction consumes 2 moles of ATP per mole of acetate and regarding the ATP generation rate in BR-3 and BR-4 they were all lower than the ATP generation rate in BR-5. In order to save energy the cell shut down this reaction in BR-3 and BR-4. The lactate production was observed just in period II of BR-3 (R32) which can be attributed to the requirements of cell for NAD^+ produced from this reaction; regarding R126 which represents transport of the NADH from cytoplasm to mitochondrion, the lowest rate belonged to BR-3 in period II and, thus, the cells use NADH by some other means (R32 and R29) in order to prevent its accumulation in cytoplasm and subsequent distraction of redox balance.

Focusing on the TCA cycle revealed that in period I in all runs wherein the cell faced low oxygen availability (but not severe oxygen limitation), the trend was similar; the TCA cycle operated properly until production of αKG (R39) which is of prime importance in anabolic reactions for the amino acids synthesis. The cycle was resumed from Fum toward OA (R44 and R46). Here we could expect secretion of succinate to the extracellular medium. Since in period II the cell was still in low oxygen

availability condition the same behaviour as period I was expected but, in contrast, BR-5 acted differently and TCA cycle functioned effectively in this case. By commencement of the severe oxygen limitation in period II, the BR-3 behaviour did not change and followed similar trend of the period I. However, in period II of BR-4 the situation was completely oxygen-limited and the initial reactions of TCA cycle shut down. It was resumed from R40 in order to generate OA.

Regarding the fatty acid synthesis reactions, it is obvious that in BR-3 and BR-4 severe oxygen limitation led to decreased flux through the corresponding reactions; in contrast, in BR-5 where the process was at the verge of severe oxygen limitation (starts at $t=9$ h) the fluxes increased from period I to period II.

The highest synthetic flux in all runs belonged to Glu (R67), or the glutamic family of amino acids (R67 to R72), which was the common trend among all the periods and was consistent with the previous findings; it reflected the pivotal role of the Glu as the precursor of many other amino acids and also Glu and Gln as donors of the virtually all amino and amide groups in the cellular components (Çelik et al., 2010; Çalık et al., 2011). The understanding of the fluxes through the pathways of amino acids biosynthesis helps to ameliorate the candidate pathways, as potential targets of the metabolic engineering, end with amino acids that have more participation in the r-protein structure (Irani et al., 2016). The percentage of the amino acids in hGH structure by inclusion of the fusion partner (N-terminus motif) can be seen in Table 4.61. Regarding Table 4.61 leucine and serine are the most abundant amino acids in hGH structure and, thus, fluxes towards them seem to be very informative. They are produced from R54 and R48, respectively. R54 flux decreased upon emergence of severe oxygen limitation for BR-3 and BR-4. On the other hand, R48 flux increased in period II of BR-3 but decreased drastically in period II of BR-4.

Table 4.61. The amino acid composition of hGH expressed in current research by inclusion of its N-terminus fusion tag. Mature hGH possesses 191 amino acids and in combination with the N-terminus motif its length increases to 203 amino acids.

Amino acid	Abbr.	Number (#)	Percent (%)
Alanine	A	7	3.45
Arginine	R	12	5.91
Asparagine	N	9	4.43
Aspartic acid	D	11	5.42
Cysteine	C	4	1.97
Glutamic acid	E	16	7.88
Glutamine	Q	13	6.40
Glycine	G	9	4.43
Histidine	H	9	4.43
Isoleucine	I	9	4.43
Leucine	L	26	12.81
Lysine	K	9	4.43
Methionine	M	3	1.48
Phenylalanine	F	14	6.89
Proline	P	8	3.94
Serine	S	18	8.87
Threonine	T	10	4.93
Tryptophan	W	1	0.49
Tyrosine	Y	8	3.94
Valine	V	7	3.45

4.10.2.3.5. Metabolic flux distribution at important nodes

Identification of the possible rigid points in the biochemical reaction network and finding of the critical branch points and bottlenecks in the overall flux distribution can also be achieved by the use of MFA (Çalık and Özdamar, 2002). In addition to the analysis of the central metabolism fluxes which gives a general view of the cell function, the metabolic fluxes around important branching points convey valuable information about the cell metabolic state and the preference of different pathways by cell regarding available bioprocess conditions.

As two important branch points in central metabolism G3P and Pyr were considered. Flux 100 was assigned to R8 and R14 and other inward and outward fluxes were calculated accordingly. Consequently, all the incoming and outgoing fluxes were equal. The calculations were conducted in period I and II of the three bioreactor runs. The normalized flux results related to G3P and Pyr node have been presented in Figures 4.81 and 4.82, respectively.

In G3P node, the main point about three runs was the stable fluxes in the case of rhGH production under control of P_{GAP} (BR-5) in two considered periods; in other two runs, BR-3 and BR-4, the fluxes changed strikingly with a decreasing trend going from period I to period II. BR-3 and BR-4 experienced severe oxygen limitation in period II and the decreasing fluxes revealed the effect of the oxygen availability on the fluxes around this node. In all runs PPP had interplay with glycolysis; the inward and outward fluxes decreased from period I to II showing either the less genetic material required in period II or the less NADPH required in oxygen scarcity. Although, the share of the amino acid and lipid synthesis pathways were negligible in all cases, the decrease in lipid metabolism flux can be attributed to the molecular oxygen scarcity because of its role in these reactions. In an overall view, the major fluxes around node in BR-4 were higher compared to BR-3 and BR-5 which delivered higher fuel to the following reactions in the glycolysis pathway.

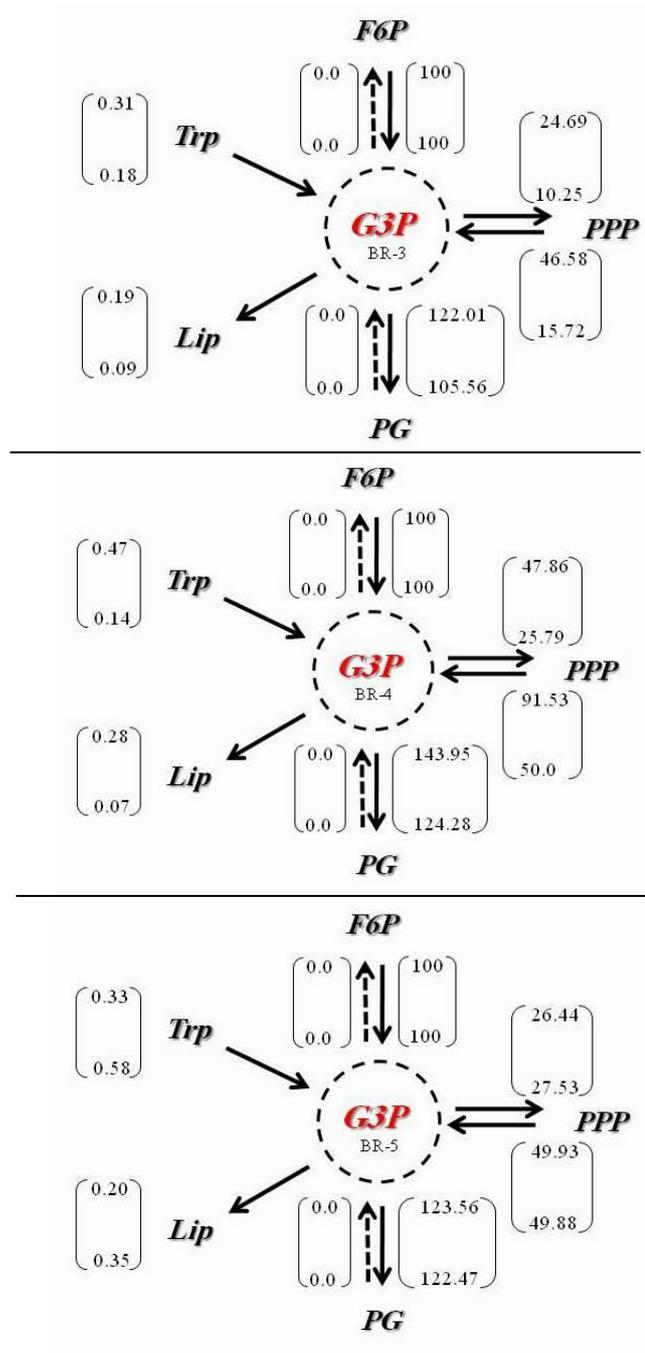


Figure 4.81. The normalized fluxes around G3P node in two periods (I, II) during rhGH expression under three different promoters P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5). The upper numbers belong to the period I and the lower ones belong to the period II.

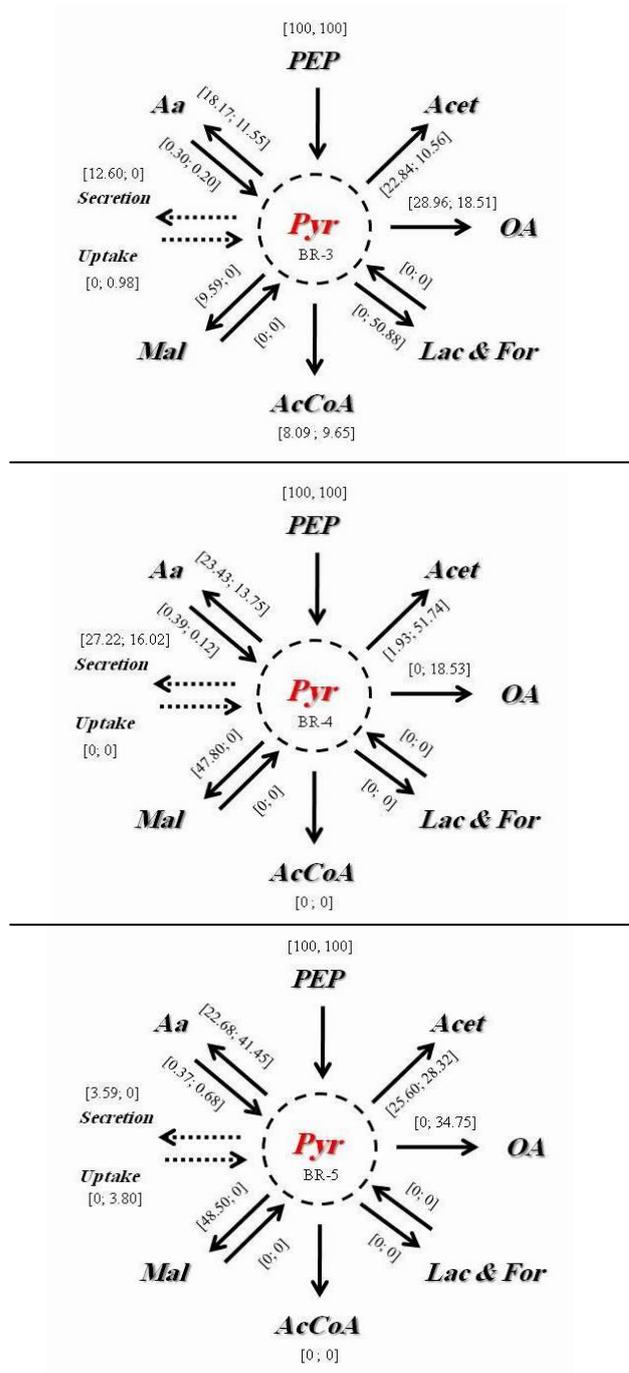


Figure 4.82. The normalized fluxes around *Pyr* node in two periods (I, II) during rhGH expression under three different promoters P_{PYRK} (BR-3), P_{PDC} (BR-4), and P_{GAP} (BR-5). The numbers in the left belong to the period I and the ones in right belong to the period II.

In Pyr node, the common feature of all runs in both periods was the presence of flux towards acetaldehyde and amino acid synthesis pathway. The inward flux from amino acids pathway was very infinitesimal in all cases. High percentage of the flux was sent towards amino acid synthesis in low oxygen availability before severe oxygen limitation. The flux of the reaction towards acetaldehyde indicates the effect of oxygen availability on deciding the route from pyruvate; even in BR-5 which both periods were in low oxygen availability, this reaction was active and implied to the incomplete TCA cycle due to the oxygen shortage. In period I of all runs, the flux of the anaplerotic reaction from Pyr to Mal was active. In period I, the by-product flux was the lowest in BR-4; it remained approximately constant in BR-5 whereas in period II of BR-3 more flux was redirected to the branched pathway of by-products. Therefore, while entering severe oxygen limitation, BR-3 had c.a. 20% more flux in fermentative pathways compared to BR-4.

In total, these two nodes possess high degree of flexibility according to the flux changes during conducted runs; regarding the cell requirements the fluxes through these nodes are subjected to adjustment.

4.10.2.4. Deciding on the most efficient promoter

The relative successfulness of rhGH expression using three different promoters (P_{PYRK} , P_{PDC} , and P_{GAP}) under low oxygen availability conditions can be argued in two ways: judgment based on secreted rhGH or deciding based on mRNA level. The former deals with the secreted rhGH to the extracellular medium by ignoring what happens in preceding steps in molecular level and the latter considers the molecular bases of the expression and secretion. In Table 4.58 the summarized results of the related experiments and calculations have been presented in order to show the privilege of each promoter in each case.

As mentioned before, at $t=9$ h where the bioprocess is about to finish low oxygen availability period and enter severe oxygen limitation period, the

rhGH production and mRNA level both led to the same outcome although q_{rhGH} is slightly different (Table 4.54). The net result was the better activity of P_{GAP} either considering secreted rhGH or transcript (mRNA) level.

By commencement of the severe oxygen limitation period ($9 \text{ h} < t$) in bioprocess the inconsistencies between secreted rhGH and its corresponding mRNA level started. Measurements of the mRNA and protein levels as the constituents of a biological system complete each other and describe the physiological state of the cell (Gygi et al., 1999; Greenbaum et al., 2003).

Although, considering intracellular expression, a general correlation has not yet been derived between mRNA level and protein level and their abundances do not follow a normal distribution, some facts and hints have been deduced hitherto. The major parameters that lead to a poor correlation between mRNA level and protein level are supposed to be post-transcriptional factors, post-translational factors, and noises and errors related to the experimental methods used for mRNA and protein measurement (Greenbaum et al., 2003; Maier et al., 2009). The proteins that are expressed in high amounts have large codon biases (Gygi et al., 1999) and codon bias is supposed to be a mechanism to enhance translation efficiency (Maier et al., 2009). More variable mRNA levels point out to high degree of correlation between mRNA level and protein level, while steady level of mRNA reveals that the control is exerted at protein level and, therefore, there is a weak correlation between mRNA level and protein level (Greenbaum et al., 2003). “Ribosomal occupancy” (association of mRNAs with ribosome) and CAI are also the factors that have been used to derive a correlation; transcripts with higher ribosomal occupancy exhibit higher degree of correlation with their corresponding protein abundance (Maier et al., 2009); furthermore, in genes with higher CAI higher correlation between mRNA level and protein level is observed although the correlation is not so strong. Individual half-life of the protein is presumed as the main post-translational parameter that affects correlation between abundances of mRNA and protein (Maier et al., 2009). In addition, it has been mentioned

that the major source of the inconsistency of mRNA-protein level is the fluctuating accuracy arose from mRNA measurement and, therefore, if the mRNA estimation accuracy increases the mRNA-protein correlation will more likely be high enough (Gry et al., 2009).

In intracellular space the rout form mRNA to protein involves series of dynamic processes entailing protein synthesis and degradation which should be thoroughly analyzed to comprehend that the change in protein abundance is the result of the change in protein synthesis rate, mRNA level, or protein degradation/turnover (Greenbaum et al., 2003).

However, related with secreted proteins, the scenario is a little different and more parameters come into play i.e., ER folding capacity, secretion and degradation by proteases. The intracellular factors are not so important in current research since the gene of interest is the same in all conditions and the reference organism is also the same; therefore, the factors related with secretion supposed to be more vital and deterministic

By utilization of the unique SP (α -MF) in all expressions, the SP effect may be excluded. The ER capacity may be overwhelmed by increased amount of entry via translocons; therefore, it is possible that high transcript number be represented by lower amount of the secreted protein because of insufficient capacity of the ER to deal with the high cargo proteins which have entered ER in order to be secreted. Furthermore, the folding process inside ER is disturbed during (severe) oxygen limitation conditions and, thus, accumulation of the un/misfolded proteins leads to initiation of UPR (Baumann et al., 2010) which ends with back flux of the un/misfolded proteins to the cytoplasm and degradation by proteasomes as explained before. Since all the strains under investigation are single-copy, the mRNA level is just attributed to the promoter activity and its elevated level is supposed to be the result of the increased promoter activity. At t=12 h the transcript level of rhGH under control of P_{PDC} and P_{GAP} are higher than P_{PYRK} (Table 4.58). However, presumably due to the ER occupation and possible induction of UPR the translated proteins have been stuck in ER or have been

degraded which has been led to less secreted rhGH compared to P_{PYRK} . Similarly, at $t=15$ h, as a result of the same phenomena, the higher transcript level related to P_{GAP} has led to lower level of secreted rhGH compared to other two promoters. It will be possible to enhance the capacity of the ER and, subsequent, secretion of the target protein by co-expression of the helper proteins (folding assistants) (Inan et al., 2006; Huo et al., 2007).

The intracellular proteases explained in section 2.3.3 are the other causative agents that may increase the inconsistency between the mRNA and corresponding secreted protein abundance. The cell death rate (equation 4.2) has a first order kinetic (Doran, 1995) with respect to the cell concentration (C_x) and by increasing the cell concentration during bioprocess the death rate will increase:

$$r_d = k_d C_x \quad (4.2)$$

where, r_d is the cell death rate, k_d is the death reaction rate constant which is a function of temperature, and C_x is the viable cell concentration. According to the Figure 4.79 and the observed increase of the proteases activity in all the triple experiments (BR-3, BR-4, and BR-5) as the result of the higher cell death towards the end of the bioprocess, the degradation of the rhGH can be plausible. Higher cell death will lead to the distribution of the more proteases in the fermentation medium which, in turn, degrades the secreted proteins in higher rate. So, the difference between mRNA level and rhGH level in severe oxygen limitation period of the bioprocess can be partly ascribed to the presence of proteases in the medium as a result of the higher number of dead cell because of environmental stress caused by oxygen scarcity.

Based on the above mentioned explanations the correlation between mRNA and rhGH level can be increased if: i) the capability of ER in handling secretory proteins would be increased by any means like co-expression of helper molecules such as protein disulfide isomerase (PDI)

and/or ii) the proteases in the fermentation medium would be inhibited by any means like addition of protease inhibitors to the medium.

The criteria for selection of the most efficient product can be related to the nature of the product, i.e., hGH, as a low volume/high value added pharmaceutical product.

In general, in any implemented fermentation process, the design parameters that are of an utmost degree of importance in decision-making can be mentioned as: yield of the product on substrate, productivity, and final titer of the product (Nielsen et al., 2003). In order to make a microbial process economically viable specially in the case of commodity chemicals, which are the products with large amount but low value such as industrial enzymes and secondary metabolites, final titer of the product, productivity, and yield should be improved as much as possible (Chubukov et al., 2016). Yield virtually represents the efficiency by which the raw initial material has been transformed into the final product and becomes more important when a large volume of the substrate is utilized. The productivity is a scale to point out the efficient utilization of the bioreactor by generating more product in a defined time period. Titer of the product is the concentration of the product in the fermentation medium at the end of the bioprocess time.

In contrast, the above-mentioned three parameters are not so vital regarding products with therapeutic importance, with low volume but high value; the quality of the product, virtually its purity and bioactivity, in this case is considered more (Nielsen et al., 2003); the explained design criteria are taken into account at the start of the design but afterwards the change in the process conditions will be too hard to perform due to the regulations. However, the titer of the product at the end of the bioprocess greatly influences the downstream purification. High final titer is especially of prime importance in the case of biopharmaceuticals where the purification cost is responsible for more than 90% of the final cost of the product (Nielsen et al., 2003).

The higher final titer will make utilization of the bioreactors with smaller volume justifiable and, thus, will lead to less capital investment and also reduced time required for cleaning and sterilization of the equipment as well. The prolonged bioprocess time seems not to be a promising strategy for increasing the final titer since the viable cells decrease dramatically based on the cell concentration profile (Figure 4.72). Alternatively, the increase in the product titer can either be achieved by process engineering like optimization of the bioprocess conditions, changing medium composition, and increasing biomass, or by host engineering with rational strain improvement by metabolic engineering tools in order to, for example, trigger the pathways ending with desired product (Nielsen et al., 2003; Liu et al., 2010; Chubukov et al., 2016). However, it should be reminded that increasing the biomass will lead to more sophisticated downstream processing and increased cost for separation of the biomass. As a solution, from process engineering point of view, working in environments with reduced oxygen content in the case of aerobic organisms like *P. pastoris* will cause decrease in biomass, as illustrated by biomass synthesis flux and decreased energy content of the cell; the reduced biomass can be compensated transcriptionally by utilization of the oxygen limitation-induced promoters (ex., P_{PYRK} and P_{PDC}) in rhGH expression, which exert their control in transcriptional level and are regulated partially by oxygen availability. It seems to be a rational strategy in order to improve the product titer which not only leads to decrease in biomass and, thus, facilitate final treatment/purification steps but also enhances the maximum achievable product amount.

In total, without any modification and improvement of the utilized r-*P. pastoris* strains in current research, considering the mRNA level in selection of the most efficient promoter will be vain since the higher mRNA, virtually the stronger promoter, will not necessarily lead to the higher secretion. Without focusing on DO, with adopted strategies and conditions in this research, in spite of the higher rhGH productivity and higher product yield on substrate obtained by P_{PYRK} , application of P_{PDC} for expression of

rhGH seems to be more reasonable by considering final product titer of 122 g/L at t=15 h which may be attributed to the higher ATP synthesis flux (Table 4.60) and lower loss of carbon source as by-products in Pyr node (Figure 4.82).

Notes:

Small letter superscripts used in Table 4.3

- ^a Functions of SP1 to SP41 have been obtained from Huang et al. research in 2011.
- ^b The (+) signs in WolfPsort and Phobius analyses refer to the same cleavage site prediction obtained by SignalP; different cleavage site prediction results (than SignalP) by WolfPsort and Phobius were indicated with the number of amino acids in each program individual column.
- ^c The underlined part of the SPs is the "H-region" predicted by Phobius program.
- ^d The bold sequences are the ones that have been predicted in this work.
- ^e NYC: has not yet been controlled in *P. pastoris* expression system.
- ^f In the case of presence of "pro" sequence, it has been separated from the "pre" part by "...."
- ^g aa, refers to amino acid.
- ^h no SP": no secretory signal peptide was predicted.
- ^j In the column of pro-sequence, the underlined numbers show that the length of the pro-sequence was obtained based on literature or data banks not ProP program.
- ^k obtained from (Riffer et al., 2002).
- ^p The pro-sequence is the prediction of ProP, however, in the UNIPROT it has been mentioned to be 4 amino acids.

CHAPTER 5

CONCLUSION

The present study was an endeavour to achieve an improved expression and secretion by *P. pastoris* host system through utilization of endogenous SPs and oxygen limitation-induced promoters.

The study was started by developing base plasmid (pGAPZ α A::hGH) from parent plasmids, pGAPZ α A and pPICZ α A::hGH. The hGH gene in the base plasmid had a fusion partner in its N-terminus comprised of 6His-tag and factor Xa cleavage site for facilitating subsequent purification and obtaining native N-terminus, respectively. P_{GAP} controlled the expression of rhGH in base plasmid and its subsequent secretion was directed by α -MF.

The in-silico analyses carried out by recruitment of secretome and proteome datasets related to *P. pastoris* were the next round of the study. Potential endogenous SPs which can be used in *Pichia*-based systems were determined by combination of the results of the SignalP, Phobius, WolfPsort, and ProP programs and based on the D-score (signal peptide-ness score) as output of SignalP. SP23, SP24, SP26, and SP34 were the SPs that their D-score were higher compared to α -MF as the most common SP in *Pichia* expression systems; SP13 was also added to this group in order to check the validity of previous inferences obtained from other studies. Besides, the potential oxygen limitation-induced promoters that presumably lead to better expression than the benchmark promoter P_{GAP} were selected as: P_{PDC}, P_{PYRK}, and P_{THI3}. The original SP and promoter in the base plasmid were substituted by the selected endogenous SPs and promoters, respectively. The resultant

plasmids were cloned in *E. coli* and were integrated to the genome of the *P. pastoris*, afterwards. For the sake of making a rational comparison based on expressed/secreted rhGH, between developed r-*P. pastoris* strains, the single-copy strains were selected by the aid of qPCR. Subsequently, they were used in fermentation experiments.

SP group strains including the strains harbouring *hGH* gene under control of P_{GAP} and secreted by selected endogenous SPs in addition to α -MF, were compared by shake flask bioreactor experiments regarding their ability in secretion of rhGH, at the first step.

- The successful secretion of the rhGH by all endogenous SPs was in agreement with their D-score > 0.8.
- The rank of the endogenous SPs based on their secretion efficiency was as SP23 > SP24 > SP34 > SP13 > SP26 which was not completely consistent with the D-score rank, as expressed by other similar studies, but the most and the least successful SPs were corresponding to the highest and lowest D-score.
- SP23 (D-score=0.883) had the most comparable efficiency to α -MF (D-score=0.885).

SP23 and α -MF as the most favoured candidates were tested further in controlled environment of the bioreactor:

- The shake flask experiments results were confirmed and SP23 secretion efficiency was obtained to be 70-80% of α -MF.
- The highest cell concentration was obtained as 82 g/L and 75 g/L for α -MF and SP23 strains, respectively at t=15 h.
- The highest amount of the secreted rhGH was 70 mg/L and 56 mg/L for α -MF and SP23 strains, respectively at t=12 h.
- By further in-silico analyses, the relative successfulness of the α -MF in secretion of the rhGH could be attributed to some extent to the chosen physicochemical properties such as pI, hydrophobicity, and aliphatic index.

- A straight-forward correlation between the examined endogenous SPs efficiency and their corresponding physicochemical properties could not be extracted.
- In order to be able to decide better about the efficiency of SPs in secretion of the proteins regarding their intrinsic properties, more factors like length of the SP, secondary structure of the SP, and interactions of selected properties and conducting further experiments with several proteins to comprehensively evaluate the parameters that influence SP efficiency seems to be imperative.
- The N-terminus of the protein affects the cleavage of the SP which is represented in D-score.
- The available N-terminus motif in N-terminal of the hGH protein showed masking effect which faded away the effect of the N-terminus of the mature protein on the D-score. It preserves the D-score > 0.8 related with SP23 which has been accepted as a threshold for expecting a successful secretion.
- When the preferred purification method is affinity chromatography, the current N-terminus motif, by inclusion of the effect of the protein amino acid sequence (since distal mutations has shown also profound effects on SP function), can be a promising fusion tag for secretion of other host proteins by SP23 in *Pichia*-based system.
- The strikingly shorter length of the SP23 compared to α -MF makes it a promising candidate for conditions that the *r-P. pastoris* strain experiences nutrient limitation like nitrogen starvation.

Promoter group strains including the strains harbouring *hGH* gene under control of P_{THI3} , P_{PYRK} , P_{PDC} , and P_{GAP} secreted by α -MF to the extracellular medium, were checked by a preliminary shake flask bioreactor experiment which lead to disqualification of P_{THI3} regarding absence of rhGH in extracellular medium. The effectiveness of the remained promoters in expression of rhGH was determined by laboratory-scale bioreactor experiments:

- The maximum cell concentration obtained in the experiments was 80 g/L, 82 g/L, and 90 g/L for strains express rhGH under control of P_{PYRK} , P_{PDC} , and P_{GAP} , respectively.
- The maximum secreted rhGH was measured to be 101 mg/L, 122 mg/L, and 58 mg/L for strains expressed rhGH under control of P_{PYRK} , P_{PDC} , and P_{GAP} , respectively.
- Ethanol was produced in high levels during severe oxygen limitation in all three experiments with the highest amount of 28 g/L in both strains with P_{PYRK} and P_{PDC} .
- The presence of the metabolites of branches from pyruvate node and also TCA cycle referred to the inefficient operation of the TCA cycle in low oxygen availability and severe oxygen limitation conditions and shift of the metabolism towards fermentation.

In order to have a better understanding of the cell behaviour in oxygen limitation state, metabolic flux analysis (MFA) and mRNA level measurement was also conducted:

- MFA results, in general, pointed out the decrease in biomass synthesis flux and diminished cell energy content in the form of ATP when cell confronted severe oxygen limitation.
- The increased flux through glycolysis in severe oxygen limitation confirmed the partial adaptation of the cell to low oxygen availability and, thus, low ATP content by compensation of the ATP obtained through oxidative phosphorylation (in TCA cycle) with ATP obtained via substrate-level phosphorylation.
- The flux changes in the reactions catalyzed by the enzymes which are under the control of the promoters that transcribe rhGH did not show parallel trend with rhGH secretion flux. It cannot be attributed to the sole substrate or enzyme level, both of them have share in flux changes, although, the effect of the substrate concentration is stronger.

- mRNA level during cultivations showed fluctuation. The rhGH transcript level by three promoters was in good agreement with the corresponding secreted rhGH until t=9 h. However, the mRNA-rhGH abundances were not correlated afterwards.
- According to the obtained results, based on the importance of the final product titer in production of biopharmaceuticals, with adopted strategy and conditions in current research in terms of product titer, P_{PDC} was selected as the most efficient promoter in expression of rhGH in collaboration with α -MF. On the other hand, in terms of the productivity and yield of the product on substrate, P_{PYRK} showed a better performance.

RECOMMENDATIONS

- The most efficient endogenous SP, SP23, can be attached to the “pro” part of the α -MF in order to analyse the effect of the “pro” part of α -MF in its higher successfulness in secretion.
- Artificial SP(s) can be designed considering the codon usage bias of *P. pastoris*; it can, then, be also used in different combinations with “pro” parts.
- Regarding the revealed strength of P_{PYRK} and P_{PDC} especially in severe oxygen limitation conditions, based on secreted rhGH, the upstream region can also be included as promoter region (up to 1000 bp) in order to inspect the probable missed transcription factor sites for improvement of their transcription efficiency.
- Combination of SP23 as SP and P_{PDC} as promoter and comparing the expression results with qualified complex of α -MF and P_{PDC} under nitrogen-limitation in oxygen limited conditions can be very informative in assessment of SP23 compared to α -MF.

REFERENCES

Abad, S., Kitz, K., Schreiner, U., Hoermann, A., Hartner, F., Glieder, A. (2010) "Real time PCR based determination of gene copy numbers in *Pichia pastoris*", *Biotechnology Journal* 5(4): 413- 420.

Agarwal, P., K., Uppada, V., Noronha, S., B. (2013) "Comparison of pyruvate decarboxylase from *Saccharomyces cerevisiae* and *Komagataella pastoris* (*Pichia pastoris*)", *Applied Microbiology and Biotechnology* 97: 9439-9449.

Ahn, J., Hong, J., Lee, H., Park, M., Lee, E., Kim, C., Choi, E., Jung, J., Lee, H. (2007) "Translation elongation factor 1- α gene from *Pichia pastoris*: molecular cloning, sequence, and use of its promoter", *Applied Microbiology and Biotechnology* 74: 601-608.

Alvaro-Benito, M., Fernandez-Lobato, M., Baronian, K., Kunze, G. (2013) "Assessment of *Schwanniomyces occidentalis* as a host for protein production using the wide-range Xplor[®]2 expression platform", *Applied Microbiology and Biotechnology* 97(10): 4443-4456.

Amersham Biosciences (1999) "Reversed phase chromatography: principles and methods", Piscataway, NJ, USA.

Andrews, D.W., Perara, E., Lesser, C., Lingappa, V. R. (1988) "Sequences beyond the cleavage site influence signal peptide function", *The Journal of Biological Chemistry* 263(30): 15791-15798.

Apte-Dashpande, A., Rewanwar, S., Kotwal, P., Raiker, V., A., Padmanabhan, S. (2009) "Efficient expression and secretion of recombinant human growth hormone in the methylotrophic yeast *Pichia pastoris*: potential applications for other proteins", *Biotechnology and Applied Biochemistry* 54(4): 197-205.

Araki, K., Nagata, K. (2011) "Protein folding and quality control in the ER", *Cold Spring Harbor Perspectives in Biology* 3(11): a007526.

Arima, K., Oshima, T., Kubota, I., Nakamura, N., Mizunaga, T., Toh-e, A. (1983) "The nucleotide sequence of the yeast *PHO5* gene: A putative precursor of repressible acid phosphatase contains a signal peptide", *Nucleic Acids Research* 11(6): 1657-1672.

Ascacio-Martinez J., A. and Barrera-Saldana, H., A. (2004) "Production and secretion of biologically active recombinant canine growth hormone by *Pichia pastoris*", *Gene* 340: 261-266.

Ata, Ö., Boy, E., Güneş, H., Çalık, P. (2015) "Codon optimization of *xylA* gene for recombinant glucose isomerase production in *Pichia pastoris* and fed-batch feeding strategies to fine-tune bioreactor performance", *Bioprocess and Biosystems Engineering*. 38(5): 889-903.

Athmaram, T., N., Saraswat, S., Singh, A., K., Rao, M., K., Gopalan, N., Suryanarayana, V., V., Rao, P., V. (2012) "Influence of copy number on the expression levels of pandemic influenza hemagglutinin recombinant protein in methylotrophic yeast *Pichia pastoris*", *Virus Genes* 45(3): 440-451.

Avery, S., V. (2006) "Microbial cell individuality and underlying sources of heterogeneity", *Nature Review Microbiology* 4(8): 577-587.

Baart, G., J., E., Martens, D., E. (2010) "Genome-scale metabolic models: Reconstruction and analysis", *Methods in Molecular Biology* 799: 107-126.

Bailey, J., E., Ollis, D., F. (1986) "Biochemical engineering fundamentals", 2nd Edn., McGraw-Hill, New York, USA.

Balagurunathan, B., Jonnalagadda, S., Tan, L., Srinivasan, R. (2012) "Reconstruction and analysis of a genome-scale metabolic model for *Scheffersomyces stipitis*", *Microbial Cell Factories* 11: 27.

Bandyopadhyay, B. and Humprey, A., E. (1967) "Dynamic measurement of the volumetric oxygen transfer coefficient in fermentation systems", *Biotechnology and Bioengineering* 9: 533-544.

Banflavi, G., (ed.) (2011) "Cell cycle synchronization: Synchronization of yeast, Chapter 12", *Methods in Molecular Biology* 761, pp 173-199.

Banga, A., K. (2006) "Therapeutic peptides and proteins- Formulation, processing, and delivery systems", 2nd Edn., CRC press.

Barsh, G., S., Seeburg, P., H., Glinas, R., E. (1983) "The human growth hormone gene family: structure and evolution of the chromosomal locus", *Nucleic Acids Research* 11(12): 39-39-3958.

Baumann, G., P. (2009) "Growth hormone isoforms", *Growth Hormone & IGF Research* 19: 333-340.

Baumann, K., Carnicer, M., Dragosits, M., Graf, A., B., Stadlmann, J., Jouhten, P., Maaheimo, H., Gasser, B., Albiol, J., Mattanovich, D., Ferrer, P. (2010) "A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions", *BMC Systems Biology* 4: 141.

Baumann, K., Maurer, M., Dragosits, M., Cos, O., Ferrer, P., Mattanovich, D. (2008) "Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins", *Biotechnology and Bioengineering* 100(1): 177-183.

Bayraktar, E., (2009) "effects of pH on Human growth hormone production by *Pichia pastoris* considering the expression levels of regulatory genes", MSc. Thesis, Middle East Technical University, Ankara.

Berkeley *Drosophila* genome project, accessed 5 April 2013, <www.fruitfly.org>.

BIO-RAD (2006) "Real-time PCR Application Guide", Bio-Rad Laboratories, Inc.

Biro, J.C. (2006) "Amino acid size, charge, hydrophathy indices and matrices for protein structure analysis", *Theoretical Biology and Medical Modeling* 3, 15.

Blanch, H., W., Clark, D., S. (1997) "Biochemical engineering", Marcel Dekker, Inc.

Bobrowicz, P., Davidson, R., C., Li, H., Potgieter, T.,I., Nett, J., H., Hamilton, S.,R., Stadheim, T., A., Miele, R., G., Bobrowicz, B., Mitchell, T., Rausch, S., Renfer, E., Wildt, S. (2004) "Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose", *Glycobiology* 14(9): 757-766.

Böer, E., Steinborn, G., Kunze, G., Gellissen, G. (2007) "Yeast expression platforms", *Applied Microbiology and Biotechnology* 77:513–523.

Boettner, M., Prinz, B., Holz, C., Stahl, U., Lang, C. (2002) "High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*", *Journal of Biotechnology* 99: 51-62.

Boettner, M., Steffens, C., von Mering, C., Bork, P., Stahl, U., Lang, C. (2007) "Sequence-based factors influencing the expression of heterologous genes in the yeast *Pichia pastoris* - a comparative view on 79 human genes", *Journal of Biotechnology* 130: 1-10.

Bollok, M., Resina, D., Valero, F., Ferrer, P. (2009) "Recent patents on the *Pichia pastoris* expression system: Expanding the toolbox for recombinant protein production", *Recent Patents on Biotechnology* 3: 192-201.

Brodsky, J., L., Skach, W., R. (2011) "Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems", *Current Opinion in Cell Biology* 23: 464-475.

Burke, D., Dawson, D., Stearns, T. (2000) "Methods in yeast genetics", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Buske, F., A., Maetschke, S., Boden, M. (2009) "It's about time: Signal recognition in staged models of protein translocation", *Pattern Recognition* 42: 567-574.

Butler, M. (2005) "Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals", *Applied Microbiology and Biotechnology* 68: 283-291.

Çalık, G., Kocabaş, P., Afşar, H., Çalık, P., Özdamar, T., H. (2016) "Parametric continuous feed stream design to fine-tune fed-batch bioreactor performance: recombinant human growth hormone production in *Bacillus subtilis*", Journal of Chemical Technology and Biotechnology 91: 2740-2750.

Çalık, P. (1998) "Bioprocess development for serine alkaline protease production", PhD. Thesis, Ankara University, Ankara.

Çalık, P. and Özdamar, T., H. (2002) "Bioreaction network flux analysis for industrial microorganisms: A review", Reviews in Chemical Engineering 18(6): 553-596.

Çalık, P., Ata, Ö., Güneş, H., Massahi, A., Boy, E., Keskin, A., Öztürk, S., Zerze, G., H., Özdamar, T., H. (2015) "Recombinant protein production in *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter: From carbon source metabolism to bioreactor operation parameters", Biochemical Engineering Journal 95: 20-36.

Çalık, P., Bayraktar, E., İnankur, B., Soyaslan, E., Ş., Şahin, M., Taşpınar, H., Açıık, E., Yılmaz, R., Özdamar, T., H. (2010a) "Influence of pH on recombinant human growth hormone production by *Pichia pastoris*", Journal of Chemical Technology and Biotechnology 85: 1628-1635.

Çalık, P., Bozkurt, B., Zerze, G., H., İnankur, B., Bayraktar, E., Boy, E., Orman, M., A., Açıık, E., Özdamar, T., H. (2013) "Effect of co-substrate sorbitol different feeding strategies on human growth hormone production by recombinant *Pichia pastoris*", Journal of Chemical Technology and Biotechnology, 88(9): 1631-1640.

Çalık, P., Çalık, G., Özdamar, T., H. (1998) "Oxygen transfer effects in serine alkaline protease fermentation by *Bacillus licheniformis*: use of citric acid as the carbon source", *Enzyme and Microbial Technology* 23: 451-461.

Çalık, P., Çalık, G., Takaç, S., Özdamar, T., H. (1999) "Metabolic flux analysis for serine alkaline protease fermentation by *Bacillus licheniformis* in a defined medium: Effect of the oxygen transfer rate", *Biotechnology and Bioengineering* 64(2): 151-167.

Çalık, P., İnankur, B., Soyaslan, E., Ş., Şahin, M., Taşpınar, H., Açık, E., Bayraktar, E. (2010b) "Fermentation and oxygen transfer characteristics in recombinant human growth hormone production by *Pichia pastoris* in sorbitol batch and methanol fed-batch operation", *Journal of Chemical Technology and Biotechnology* 85(2): 226-233.

Çalık, P., Orman, M., A., Çelik, E., Halloran, S., M., Çalık, G., Özdamar, T., H. (2008) "Expression system for synthesis and purification of recombinant human growth hormone in *Pichia pastoris* and structural analysis by MALDI-ToF mass spectrometry", *Biotechnology Progress* 24: 221-226.

Çalık, P., Şahin, M., Taşpınar, H., Soyaslan, E., Ş., İnankur, B. (2011) "Dynamic flux balance analysis for pharmaceutical protein production by *Pichia pastoris*: Human growth hormone", *Enzyme and Microbial Technology* 48: 209-216.

Camattari, A., Bianchi, M., M., Branduardi, P., Porro, D., Brambilla, L. (2007) "Induction by hypoxia of heterologous-protein production with the *KIPDCI* promoter in yeasts", *Applied and Environmental Microbiology* 73(3): 922-929.

Caspeta, L., Shoaie, S., Agren, R., Nookaew, I., Nielsen, J. (2012) "Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in-silico evaluation of their potentials", BMC Systems Biology 6: 24.

Castan, A., Nasman, A., Enfors, S. (2002) "Oxygen enriched air supply in *Escherichia coli* processes: Production of biomass and recombinant human growth hormone", Enzyme and Microbial Technology 30: 847-854.

Catzel, D., Lalevski, H., Marquis, C., P., Gray, P., P., Van Dyk, D., Mahler, S., M. (2003) "Purification of recombinant human growth hormone from CHO cell culture supernatant by Gradiflow preparative electrophoresis technology", Protein Expression and Purification 32: 126-134.

Çelik, E. and Çalık, P. (2011) "Production of recombinant proteins by yeast cells", Biotechnology Advances 30(5): 1108-1118.

Çelik, E., Çalık, P., Oliver, S., G. (2009) "A structured kinetic model for recombinant protein production by Mut⁺ strain of *Pichia pastoris*", Chemical Engineering Science 64: 5028-5035.

Çelik, E., Çalık, P., Oliver, S., G. (2010) "Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: effect of methanol feeding rate", Biotechnology and Bioengineering 105(2): 317-329.

Cereghino, G., P., L., Cereghino, J., L., Ilgen, C., Cregg, J., M. (2002) "Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*", Current Opinion in Biotechnology 13: 329-332.

Cereghino, J., L. and Cregg, J.,M. (2000) "Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*", FEMS Microbiology Reviews 24: 45-66.

Chaudhuri, B., Steube, K. and Stephan, C. (1992) "The pro-region of the yeast prepro- α -factor is essential for membrane translocation of human insulin-like growth factor 1 *in vivo*", European Journal of Biochemistry 206: 793-798.

Chauhan, A., K., Arora, D., Khanna, N. (1999) "A novel feeding strategy for enhanced protein production by fed-batch fermentation in recombinant *Pichia pastoris*", Process Biochemistry 34: 139-145.

Chien, L. and Lee, C. (2005) "Expression of bacterial hemoglobin in the yeast, *Pichia pastoris*, with a low O₂-induced promoter", Biotechnology Letters 27: 1491-1497.

Chiruvolu, V., Cregg, J., M., Meagher, M., M. (1997) "Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fed-batch fermentations", Enzyme and Microbial Technology 21: 277-283.

Choo, K., H., Ranganathan, S. 2008 "Flanking signal a mature peptide residues influence signal peptide cleavage", BMC Bioinformatics 9 (Suppl 12): S15.

Chou, K., C. (2002) "Review: prediction of protein signal sequences", Current Protein and Peptide Science 3: 615-622.

Chubukov, V., Mukhopadhyay, A., Petzold, C., J., Keasling, J., D., Martin, H., G. (2016) "Synthetic and systems biology for microbial production of commodity chemicals", *npj Systems Biology and Applications* 2: 16009.

Chung, B., H., Nam, S., W., Kim, B., M., Park, Y. (1996) "Highly efficient secretion of heterologous proteins from *Saccharomyces cerevisiae* using inulinase signal peptides", *Biotechnology and Bioengineering* 49: 473-479.

Chung, B., K., S., Selvarasu, S., Camattari, A., Ryu, J., Lee, Hy., Ahn, J., Lee, Ho., Lee, D. (2010) "Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement", *Microbial Cell Factories* 9:50.

Corchero, J., L., Gasser, B., Resina, D., Smith, W., Parrilli, E., Vázquez, F., Abasolo, I., Giuliani, M., Jäntti, J., Ferrer, P., Saloheimo, M., Mattanovich, D., Schwartz Jr., S., Tutino, M., L., Villaverde, A. (2012) "Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics", *Biotechnology Advances* 31 (2): 140-153.

Cos, O., Ramon, R., Montesinos, J., L., Valero, F. (2006) "Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review", *Microbial Cell Factories* 5:17.

Cos, O., Serrano, A., Montesinos, J., L., Ferrer, P., Cregg, J., M., Valero, F. (2005) "Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures", *Journal of Biotechnology* 116: 321-335.

Crawford, K., Zaror, I., Bishop, R., J., Innis, M., A. (2003) “*Pichia* secretory leader for protein expression”, *US Patent*.

Cregg, J., M. (1999) "Expression in the methylotrophic yeast *Pichia pastoris*", Gene expression systems – using nature for the art of expression, Ed. Fernandez, J., M. and Hoeffler, J., P., Academic Press, San Diego, California, USA. pp 157-191.

Cunha, N., B., Murad, A., M., Cipriano, T., M., Araujo, A., C., G., Aragao, F., J., L., Leite, A., Vianna, G., R., McPhee, T., R., Souza, G., H., M., F., Waters, M., J., Rech, E., L. (2011) "Expression of functional recombinant human growth hormone in transgenic soybean seeds", *Transgenic Research* 20: 811-826.

Daly, R. and Hearn, M., T., W. (2005) "Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production", *Journal of Molecular Recognition* 18: 119-138.

Damasceno, L., M., Huang, C., Batt, C., A. (2012) "Protein secretion in *Pichia pastoris* and advances in protein production", *Applied Microbiology and Biotechnology* 93: 31-39.

de Almeida, J., de Moraes, L., Goncalves Torres, F. (2005) "Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*", *Yeast* 22: 725-737.

de Bruin, E., C., Duitman, E., H., de Boer, A., L., Veenhuis, M., Bos, I., G., A., Hack, C., E. (2005) "Pharmaceutical proteins from methylotrophic yeasts", *Methods in Molecular Biology* 308: 65-76.

De Pourcq, K., De Schutter, K., Callewaert, N. (2010) "Engineering of glycosylation in yeast and other fungi: Current state and perspectives", *Applied Microbiology and Biotechnology* 87:1617–1631.

De Schutter, K., Lin, Y., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., Van de Peer, Y., Callewaert, N. (2009) "Genome sequence of the recombinant protein production host *Pichia pastoris*", *Nature Biotechnology* 27(6): 561-566.

Demain, A., L. and Vaishnav, P. (2009) "Production of recombinant proteins by microbes and higher organisms", *Biotechnology Advances* 27: 297–306.

Dieffenbach, C., W., Dveksler, G., S. (2003) "PCR primer: A laboratory manual" Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2nd Edn., New York, USA.

Dimitrov, D., S. (2012) "Therapeutic proteins", *Methods in Molecular Biology* 899: 1-26.

Doran, P., M. (1995) "Bioprocess engineering principles", Academic Press Inc., San Diego, CA, USA.

Dragosits, M., Stadlmann, J., Albiol, J., Baumann, K., Maurer, M., Gasser, B., Sauer, M., Altmann, F., Ferrer, P., Mattanovich, D. (2009) "The effect of temperature on the proteome of recombinant *Pichia pastoris*", *Journal of Proteome Research* 8: 1380-1392.

Dragosits, M., Stadlmann, J., Graf, A., Gasser, B., Maurer, M., Sauer, M., Kreil, D., P., Altmann, F., Mattanovich, D. (2010) "The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*", BMC Genomics 11: 207.

Driss, D., Bhiri, F, Ghorbel, R., Chaabouni, S., E. (2012) "Cloning and constitutive expression of His-tagged xylanase GH 11 from *Penicillium occitanis* Pol6 in *Pichia pastoris* X33: Purification and characterization", Protein Expression and Purification 83: 8-14.

Dyck, M., K., Gagne, D., Ouellet M., senechal, J., Belanger, E., Lacroix, D., Sirard, M., Pothier, F. (1999) "Seminal vesicle production and secretion of growth hormone into seminal fluid", Nature Biotechnology 17: 1087-1090.

Ecamilla-Trevino, L., L., Viader-Salvado, J., M., Barrera-Saldana, H., A., Guerrero-Olazaran, M. (2000) "Biosynthesis and secretion of recombinant human growth hormone in *Pichia pastoris*", Biotechnology Letters 22: 109-114.

Edward, J., S., Palsson, B., O. (1999) "Systems properties of the *Haemophilus influenza* Rd metabolic genotype", Journal of Biological Chemistry 274: 17410-17416.

Eiden-Plach A., Zagorc, T., Heintel, T., Carius, Y., Breinig, F., Schmitt, M., J. (2004) "Viral preprotoxin signal sequence allows efficient secretion of green fluorescence protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*", Applied and Environmental Microbiology 70(2): 961-966.

Erickson, B., Nelson, J., E., Winters, P. (2012) "Perspective on opportunities in industrial biotechnology in renewable chemicals", *Biotechnology Journal* 7: 1-10.

Eurwilaichitr, L., Roytrakul, S., Suprasongsin, C., Manitchotpisit, P., Panyim, S. (2002) "Glutamic acid and alanine spacer is not necessary for removal of MF α -1 signal sequence fused to the human growth hormone produced from *Pichia pastoris*", *World Journal of Microbiology and Biotechnology* 18: 493-498.

Eylar, E., H. (1966) "On the biological role of glycoproteins", *Journal of Theoretical Biology* 10(1): 89–113.

Fernandez, F., J., Vega, M., C. (2013) "Technologies to keep an eye on: Alternative hosts for protein production in structural biology", *Current Opinion in Structural Biology* 23: 365-373.

Ferreira, A., R., Ataide, F., von Stosch, M., Dias, J., M., L., Clemente, J., J., Cunha, A., E., Oliveira, R. (2012) "Application of adaptive DO-stat feeding control to *Pichia pastoris* X33 cultures expressing a single chain antibody fragment (ScFv)", *Bioprocess and Biosystems Engineering* 35(9): 1603-1614.

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J., L., Vázquez, E., Villaverde, A. (2009) "Microbial factories for recombinant pharmaceuticals", *Microbial Cell Factories* 8: 17.

Förster, J., Famili, I., Fu, P., Palsson, B., Ó., Nielsen, J. (2003) "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network", *Genome Research* 13: 244-253.

Gao, Z., Li, Z., Zhang, Y., Huang, H., Li, M., Zhou, L., Tang, Y., Yao, B., Zhang, W. (2012) "High-level expression of the *Penicillium notatum* glucose oxidase gene in *Pichia pastoris* using codon optimization", *Biotechnology Letters* 34: 507-514.

Garcia-Ochoa, F. and Gomez, E. (2009) "Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview", *Biotechnology Advances* 27: 153-176.

Garcia-Ochoa, F., Gomez, E., Santos, V., E., Merchuk, J., C. (2010) "Oxygen uptake rate in microbial processes: An Overview", *Biochemical Engineering Journal* 49: 289-307.

Gasch, A., P. (2003) "The environmental stress response: A common yeast response to diverse environmental stresses", *Topics in Current Genetics* 1: 11-70.

Gasser, B., Saloheimo, M., Rinas U., Dragosits, M., Rodriguez-Carmona, E., Baumann, C., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M., L., Mattanovich, D., Villaverde, A. (2008) "Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview", *Microbial Cell Factories* 7: 11.

Gatz, C. and Lenk, I. (1998) "Promoters that respond to chemical inducers", *Trends in Plant Science* 3(9): 352-358.

Gellissen, G., Melber, K., Janowicz, Z., A., Dahlems, U., M., Weydemann, U., Piontek, M., Strasser, A., W., M., Hollenberg, C., P. (1992) "Heterologous protein production in yeast", *Antonie van Leeuwenhoek* 62: 79-93.

GenScript, WoLFPSORT: Advanced protein subcellular localization prediction tool, accessed 13 May 2014, <<http://wolfpsort.org/>>.

Ghosalkar, A., Sahai, V., Srivastava, A. (2008) "Optimization of chemically defined medium for recombinant *Pichia pastoris* for biomass production", *Bioresource Technology* 99: 7906-7910.

Goeddel, D., V., Heyneker, H., L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D., G., Ross, M., J., Miozzari, R., C., Seeburg, P., H. (1979) "Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone", *Nature* 281: 544–548.

Gombert, A., K. and Nielsen, J. (2000) "Mathematical modeling of metabolism", *Current Opinion in Biotechnology* 11: 180-186.

Goodrich, J., C., Xu, M., Finnegan, R., Schilling B., M., Schiavi, S., Hoppe, H., Wan, N., C. (2001) "High-level expression and stabilization of recombinant human chitinase produced in a continuous constitutive *Pichia pastoris* expression system", *Biotechnology and Bioengineering* 74(6): 492-497.

Graumann, K. and Premstaller, A. (2006) "Manufacturing of recombinant therapeutic proteins in microbial systems", *Biotechnology Journal* 1: 164-186.

Gray, G., L., McKeown, K., A., Jones, A., J., S., Seeburg, P., H., Heyneker, H., L. (1984) "*Pseudomonas aeruginosa* secretes and correctly processes human growth hormone", *Nature Biotechnology* 2: 161-165.

Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M., Nilsson, P. (2009) "Correlation between RNA and protein expression profiles in 23 human cell lines", *BMC Genomics* 10: 365.

Güneş, H., Boy, E., Ata, Ö., Zerze, G., H., Çalık, P., Özdamar, T., H. (2016) "Methanol feeding strategy design enhances recombinant human growth hormone production by *Pichia pastoris*", *Journal of Chemical Technology and Biotechnology* 91: 664-671.

Gygi, S., P., Rochon, Y., Franza, B., R., Aebersold, R. (1999) "Correlation between protein and mRNA abundance in yeast", *Molecular and Cellular Biology* 19(3): 1720-1730.

Hackett, S., R., Zanutelli, V., R., T., Xu, W., Goya, J., Park, J., O., Perlman, D., H., Gibney, P., A., Botstein, D., Storey, J., D., Rabinowitz, J., D. (2016) "System-level analysis of mechanisms regulating yeast metabolic flux", *Science* 354(6311): 432.

Hahm, M., S. and Chung, B., H. (2001) "Secretory expression of human growth hormone in *Saccharomyces cerevisiae* using three different leader sequences", *Biotechnology and Bioprocess Engineering* 6: 306-309.

Han, Z., Wu, S., Li, Q., Li, J., Gao, D., Li, K., Liu, Z., W., Zhao, H. (2009) "Efficient human growth hormone gene expression in the milk of non-transgenic goats", *Folia Biologica* 55: 17-22.

Handwerger, S. and Freemark, M. (2000) "The roles of placental growth hormone and placental lactogen in the regulation of human fetal growth and development", *Journal of Pediatric Endocrinology Metabolism* 13 (4): 343-356.

Hartner, F., S., Ruth, C., Langenegger, D., Johnson, S., N., Hyka, P., Lin-Cereghino G., P., Lin-Cereghino, J., Kovar, K., Cregg, J.,M., Glieder, A. (2008) "Promoter library designed for fine-tuned gene expression in *Pichia pastoris*", Nucleic Acids Research 36(12): e76.

Hatsuzawa, K., Tagaya, M., Mizushima, S. (1997) "The hydrophobic region of signal peptides is a determinant for SRP recognition and protein translocation across the ER membrane" Journal of Biochemistry 121(2): 270-277.

Hegde, R., S., Bernstein, H., D. (2006) "The surprising complexity of signal sequences", Trends in Biochemical Sciences 31(10): 563-571.

Hens, J., R., Amstutz, M., D., Schanbacher, F., L., Mather, I., H. (2000) "Introduction of the human growth hormone gene into the guinea pig mammary gland by in vivo transfection promotes sustained expression of human growth hormone in the milk throughout lactation", Biochimica et Biophysica Acta 1523: 161-171.

Hepner, F., Cszasar, E., Roitinger, E., Lubec, G. (2005) "Mass spectrometrical analysis of recombinant human growth hormone (Genotropin[®]) reveals amino acid substitution in 2% of the expressed protein", Proteome Science 3(1).

Hetz, C., (2012) "The unfolded protein response: Controlling cell fate decisions under ER stress and beyond", Nature Reviews 13: 89-102.

Heyland, J., Fu, J., Blank, L., M., Schmid, A. (2010) "Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production", Biotechnology and Bioengineering, 107(2): 357-368.

Heyland, J., Fu, J., Blank, L., M., Schmid, A. (2011) "Carbon metabolism limits recombinant protein production in *Pichia pastoris*", *Biotechnology and Bioengineering* 108(8): 1942-1953.

Hiramatsu, R., Yamashita, T., Aikawa, J., Horinouchi, S., Beppu, T. (1990) "The prepro-peptide of *Mucor* rennin directs the secretion of human growth hormone by *Saccharomyces cerevisiae*", *Applied and Environmental Microbiology* 56(7): 2125-2132.

Hohenblum, H., Gasser, B., Maurer, M., Borth, N., Mattanovich, D. (2004) "Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*", *Biotechnology and Bioengineering* 85 (4): 367-375.

Hollenberg, C., P. and Gellissen, G. (1997) "Production of recombinant proteins by methylotrophic yeasts", *Current Opinion in Biotechnology* 8: 554-560.

Honjo, M., Akaoka, A., Nakayama, A., Furutani, Y. (1986) "Secretion of human growth hormone in *Bacillus subtilis* using prepropeptide coding region of *Bacillus amyloliquefaciens* neutral protease gene", *Journal of Biotechnology* 4(2): 63-71.

Houdebine, L. (2009) "Production of pharmaceutical proteins by transgenic animals", *Comparative Immunology, Microbiology and Infectious Diseases* 32: 107-121.

Hu, S., Li, L., Qiao, J., Guo, Y., Cheng, L., Liu, J. (2006) "Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*", *Protein Expression and Purification* 47: 249–257.

Hu, X., Chu, J., Zhang, Z., Zhang, S., Zhuang, Y., Wang, Y., Guo, M., Chen, H., Yuan, Z. (2008) "Use of different carbon sources in cultivation of recombinant *Pichia pastoris* for angiostatin production", *Journal of Biotechnology* 137: 44-49.

Huang, C., Damasceno, L., M., Anderson, K., A., Zhang, S., Old, L., J., Batt, C., A. (2011) "A proteomic analysis of the *Pichia pastoris* secretome in methanol-induced cultures", *Applied Microbiology and Biotechnology* 90: 235-247.

Huo, X., Liu, Y., Wang, X., Ouyang, P., Niu, Z., Shi, Y., Qiu, B. (2007) "Co-expression of human protein disulfide isomerase (hPDI) enhances secretion of bovine follicle-stimulating hormone (bFSH) in *Pichia pastoris*", *Protein Expression and Purification* 54: 234-239.

Hong, I., P., Anderson, S., Choi, S., G. (2006) "Evaluation of a new episomal vector based on the GAP promoter for structural genomics in *Pichia pastoris*", *Journal of Microbiology and Biotechnology* 16: 1362-1368.

Ide, N., Masuda, T., Kitabatake, N. (2007) "Effect of pre- and pro-sequence of thaumatin on the secretion by *Pichia pastoris*", *Biochemical and Biophysical Research Communication* 363: 708-714.

Idiris, A., Tohda, H., Kumagai, H., Takegawa, K. (2010) "Engineering of protein secretion in yeast: Strategies and impact on protein production", *Applied Microbiology and Biotechnology* 86: 403-417.

Inan, M. and Meagher, M., M. (2001a) "Non-repressing carbon sources for alcohol oxidase (*AOX1*) promoter of *Pichia pastoris*", *Journal of Bioscience and Bioengineering* 92(6): 585-589.

Inan, M. and Meagher, M., M. (2001b) "The effect of ethanol and acetate on protein expression in *Pichia pastoris*", Journal of Bioscience and Bioengineering 92 (4): 337-341.

Inan, M., Aryasomayajula, D., Sinha, J., Meagher, M., M. (2005) "Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase", Biotechnology and Bioengineering 93(4): 771-778.

Integrated DNA Technologies, OligoAnalyzer 3.1, accessed 10 July 2015, <<http://eu.idtdna.com/calc/analyzer>>.

Invitrogen Corporation, *Pichia* fermentation process guidelines, accessed 15 June 2016, <www.invitrogen.com>.

Invitrogen™, (2010) "EasySelect™ *Pichia* Expression Kit", Cat. No. K1740-01, by life technologies™, 18 June.

Invitrogen user manual (2010) "pGAPZA, B, C and pGAPZαA, B, C *Pichia* expression vectors for constitutive expression and purification of recombinant proteins", last update 28 June 2010.

Irani, Z., A., Kerkhoven, E., J., Shojaosadati, S., A., Nielsen, J. (2016) "Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins", Biotechnology and Bioengineering 113(5): 961-969.

İleri, N., Çalık, P. (2006) "Effects of pH strategy on endo- and exo-metabolome profiles and sodium potassium hydrogen ports of β -lactamase-producing *Bacillus licheniformis*", Biotechnology Progress 22: 411-419.

Jacobs, P., P., Geysens, S., Vervecken, W., Contreras, R., Callewaert, N. (2009) "Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology", *Nature Protocols* 4(1): 58-70.

Jahic, M., Rotticci-Mulder, J., C., Martinelle, M., Hult, K., (2002) "Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein", *Bioprocess and Biosystem Engineering* 24: 385-393.

Jez, J., Castilho, A., Grass, J., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., Steinkellner, H. (2013) "Expression of functionally active sialylated human erythropoietin in plants", *Biotechnology Journal* 8: 371-382.

Jia, H., Fan, G., Yan, Q., Liu, Y., Yan, Y., Jiang, Z. (2012) "High-level expression of a hyperthermostable *Thermotoga maritima* xylanase in *Pichia pastoris* by codon optimization", *Journal of Molecular Catalysis B: Enzymatic* 78: 72– 77.

Jia, Y., Li, S., Allen, G., Feng, S., Xue, L. (2012) "A novel glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) promoter for expressing transgenes in the halotolerant alga *Dunaliella salina*", *Current Microbiology* 64: 506-513.

Jorda, J., Jouhten, P., Camara, E., Maaheimo, H., Albiol, J., Ferrer, P. (2012) "Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose methanol mixtures", *Microbial Cell Factories* 11:57.

Jorda, J., Suarez, C., Carnicer, M., Pierick, A., Heijnen, J., J., van Gulik, W., Ferrer, P., Albiol, J., Wahl, A. (2013) "Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary ^{13}C flux analysis", *BMC Systems Biology* 7:17.

Jungo, C., Rerat, C., Marison, I., W., von Stockar, U. (2006) "Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of recombinant avidin in a *Pichia pastoris* Mut⁺ strain", *Enzyme and Microbial Technology*, 39(4): 936-944.

Kadono-Okuda, K., Yamamoto, M., Higashino, Y., Taniai, K., Kato, Y., Chowdhury, S., Xu, J., Choi, S., K., Sugiyama, M., Nakashima, K., et al. (1995) "*Baculovirus*-mediated production of the human growth hormone in larvae of the silkworm, *Bombyx mori*", *Biochemistry and Biophysics Research Communications* 213(2): 389-396.

Kajava, A.V., Zolov, S.N., Kalinin, A.E., Nesmeyanova, M.A. (2000) "The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of Gram-negative bacteria", *Journal of Bacteriology* 182(8), 2163-2169.

Karaoglan, M., Karaoglan, F., E., Inan, M. (2016a) "Functional analysis of alcohol dehydrogenase (ADH) genes in *Pichia pastoris*", *Biotechnology Letters* 38: 463-469.

Karaoglan, M., Karaoglan, F., E., Inan, M. (2016b) "Comparison of *ADH3* promoter with commonly used promoters for recombinant protein production in *Pichia pastoris*", *Protein Expression and Purification* 121: 112-117.

Karp, G. (2008) "Cell and molecular biology concepts and experiments", 5th Edn. John Wiley and Sons, Inc. Hoboken, NJ, USA.

Kato, S., Ishibashi, M., Tatsuda, D., Tokunaga, H., Tokunaga, M. (2001) "Efficient expression, purification, and characterization of mouse salivary α -amylase secreted from methylotrophic yeast, *Pichia pastoris*", *Yeast* 18: 643-655.

Kazusa DNA Research Institute, accessed 4 March 2013, <www.kazusa.or.jp>.

Keane, J., T., Ryan, D., Gray, P., P. (2003) "Effect of shear stress on expression of a recombinant protein by Chinese hamster ovary cells", *Biotechnology and Bioengineering* 81(2): 211-220.

Kern, A., Hartner, F., S., Freigassner, M., Spielhofer, J., Rumpf, C., Leitner, L., Fröhlich K., Glieder, A. (2007) "*Pichia pastoris* 'just in time' alternative respiration", *Microbiology* 153:1250-1260.

Kerr, D., E., Liang, F., Bondioli, K., R., Zhao, H., Kreibich, G., wall, R., J., Sun, T. (1998) "The bladder as a bioreactor: Urothelium production and secretion of growth hormone into urine", *Nature Biotechnology* 16: 75-79.

Khasa, Y., P., Conrad, S., Sengul, M., Plautz, S., Meagher, M., M. and Inan, M. (2011) "Isolation of *Pichia pastoris* PIR genes and their utilization for cell surface display and recombinant protein secretion", *Yeast* 28: 213-226.

Khasa, Y., P., Khushoo, A., Srivastava, L., Mukherjee, K., J. (2007) "Kinetic studies of constitutive human granulocyte-macrophage colony stimulating factor (hGM-CSF) expression in continuous culture of *Pichia pastoris*", *Biotechnology Letters* 29: 1903-1908.

Khatri, N., K., Hoffmann, F. (2005) "Impact of methanol concentration on secreted protein production in oxygen-limited cultures of recombinant *Pichia pastoris*", *Biotechnology and Bioengineering* 93(5): 871-879.

Kim, M., Park, H., S., Seo, K., H., Yang, H., Kim, S., Choi, J. (2013) "Complete solubilization and purification of recombinant human growth hormone produced in *Escherichia coli*", *PLOS ONE* (e56168) 8 (2): 1-8.

Kim, T., Baek, M., Lee, E., Kwon, T., Yang, M. (2008) "Expression of human growth hormone in transgenic rice cell suspension culture", *Plant Cell Reports* 27: 885-891.

Kimata, Y., Ishiwata-Kimata, Y., Yamada, S., Kohno, K. (2006) "Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins", *Genes to Cells* 11: 59-69.

Kjarulff, S. and Jensen, M., R. (2005) "Comparison of different signal peptides for secretion of heterologous proteins in fission yeast", *Biochemical and Biophysical Research Communication* 336: 974-982.

Klatt, S. and Konthur, Z. (2012) "Secretory signal peptide modification for optimized antibody-fragment expression-secretion in *Leishmania tarentolae*", *Microbial Cell Factories* 11: 97.

Klein, D. (2002) "Quantification using real-time PCR technology: Applications and limitations", Trends in Molecular Medicine 8(6): 257-260.

Kleine, B., Rossmannith, W., G. (2016) "Hormones and the endocrine system: Textbook of Endocrinology", Springer, Switzerland.

Koller, A., Valesco, J., Subramani, S. (2000) "The *CUPI* promoter of *Saccharomyces cerevisiae* is inducible by copper in *Pichia pastoris*", Yeast 16: 651-656.

Korona, B., Korona, D. and Bielecki, S. (2006) "Efficient expression and secretion of two co-produced xylanases from *Aspergillus niger* in *Pichia pastoris* directed by their native signal peptides and the *Saccharomyces cerevisiae* α -mating factor", Enzyme and Microbial Technology 39: 683-689.

Kottmeier, K., Günther, T., J., Weber, J., Kurtz, S., Ostermann, K., Rödel, G., Bley, T. (2012) "Constitutive expression of hydrophobin HFB1 from *Trichoderma reesei* in *Pichia pastoris* and its purification by foam separation during cultivation", Engineering in Life Sciences 12(2): 162-170.

Kottmeier, K., Ostermann, K., Bley, T., Rodel, G. (2011) "Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in *Pichia pastoris*", Applied Microbiology and Biotechnology 91:133-141.

KS Chung, B., Selvarasu, S., Camattari, A., Ryu, J., Lee, J., Ahn, J., Lee, H., Lee, D. (2010) "Genome-scale metabolic reconstruction and in-silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement", Microbial Cell Factories 9:50.

Küberl, A., Schneider, J., Thallinger, G., G., Anderl, L., 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", Journal of Biotechnology 154 : 312– 320.

Kunert, R., Steinfeldner, W., Altmann, F., Wallner, J., Katinger, H., Vorauer-Uhl, K. (2009) "CHO-recombinant human growth hormone as a protease sensitive reporter protein", Applied Microbiology and Biotechnology 84: 693-699.

Kuwae, S., Ohyama, M., Ohya, T., Ohi, H. and Kobayashi, K. (2005) "Production of recombinant human antithrombin by *Pichia pastoris*", Journal of Bioscience and Bioengineering 99(3): 264-271.

Lacroix, M., C., Guibourdenche, J., Frenedo, J., L., Muller, F., Evain-Brion, D. (2002) "Human placental growth hormone - a review", Placenta. Suppl A: S87-94.

Landes, N., Gasser, B., Vorauer-Uhl, K., Lhota, G., Mattanovich, D., Maurer, M. (2016) "The vitamin-sensitive promoter P_{THI} enables pre-defined autonomous induction of recombinant protein production in *Pichia pastoris*", Biotechnology and Bioengineering 113(12): 2633-2643.

Laukens, B., De Wachter, C., Callewaert, N. (2015) "Engineering the *Pichia pastoris* N-glycosylation pathway using the GlycoSwitch technology", Methods in Molecular Biology 1321: 103-122.

Leader, B., Baca, Q., J., Golan, D., E. (2008) "Protein therapeutics: a summary and pharmacological classification", Nature Reviews Drug Discovery 7: 21-39.

Lee, C., Kim, J., Shin, S., G., Hwang, S. (2006) "Absolute and relative qPCR quantification of plasmid copy number in *Escherichia coli*", Journal of Biotechnology 123: 273-280.

Lee, C., C., Williams, T., G., Wong, D., W., Robertson, G., H. (2005) "An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*", Plasmid 54: 80-85.

Lee, S., G., Koh, H., Y., Han, S., J., Park, H., Na, D., C., Kim, I., Lee, H., K., Yim, J., H. (2010) "Expression of recombinant endochitinase from the Antarctic bacterium, *Sanguibacter antarcticus* KOPRI 21702 in *Pichia pastoris* by codon optimization", Protein Expression and Purification 71: 108-114.

Leite, A., Kemper, E., L., da Silva, M., J., Luchessi, A., D., Siloto, R., M., P., Bonaccorsi, E., D., El-Dorry, H., F., Arruda, P. (2000) "Expression of correctly processed human growth hormone in seeds of transgenic tobacco plants", Molecular Breeding 6: 47-53.

Li, P., Anumanthan, A., Gao, X., Ilangovan, K., Suzara, V., V., Duzgunes, N., Renugopalakrishnan, V., (2007) "Expression of recombinant proteins in *Pichia pastoris*", Applied Biochemistry and Biotechnology 142: 105-124.

Li, P., Beckwith, J., Inouye, H. (1988) "Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in *Escherichia coli*", Proceedings of the National Academy of Sciences 85, 7685-7689.

Li, S., Sing, S., Wang, Z. (2011) "Improved expression of *Rhizopus oryzae* α -amylase in the methylotrophic yeast *Pichia pastoris*", Protein Expression and Purification 79: 142–148.

Li, Z., Leung, W., Yon, A., Nguyen, J., Perez, V., C., Vu, J., Giang, W., Luong, L., T., Phan, T., Salazar, K., A., Gomez, S., R., Au, C., Xiang, F., Thomas, D., W., Franz, A., H., Lin-Cereghino, J., Lin-Cereghino, G. (2010) "Secretion and proteolysis of heterologous proteins fused to *Escherichia coli* maltose binding protein in *Pichia pastoris*", Protein Expression and Purification 72: 113-124.

Liang, J. and Yuan, J. (2007) "Oxygen transfer model in recombinant *Pichia pastoris* and its application in biomass estimation", Biotechnology Letters 29: 27-35.

Liang, S., Li, C., Ye, Y., Lin, Y. (2013) "Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in *Pichia pastoris*", Biotechnology Letters 35: 97–105.

Lin-Cereghino, G., P., Stark, C., M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., Agari, K., Moua, P., Low, L., K., Tran, N., Huang, A., D., Nattestad, M., Oshiro, K., T., Chang, J., W., Chavan, A., Tsai, J., W., Lin-Cereghino, J. (2013) "The effect of α -mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*", Gene 519(2): 311-317.

Lipinski, D., Jura, J., Kalak, R., Plawski, A., Kala, M., Szalata, M., Jarmuz, M., Korcz, A., Slomska, K., Jura, J., Gronek, P., Smorag, Z., Pienkowski, M. (2003) "Transgenic rabbit producing human growth hormone in milk", Journal of Applied Genetics 44(2): 165-174.

Lipinski, D., Zeyland, J., Szalata, M., Plawski, A., Jarmuz, M., Jura, J., Korcz, A., Smorag, Z., Pienkowski, M., Slomski R. (2012) "Expression of human growth hormone in the milk of transgenic rabbits with transgene mapped to the telomere region of chromosome 7q", *Journal of Applied Genetics* 53: 435-442.

Liu, H., F., Ma, J., Winter, C., Bayer, R. (2010) "Recovery and purification process development for monoclonal antibody production", *mAbs* 2(5): 480-499.

Liu, X., Sun, L., Li, C., Yang, A., Zhang, J. (2013) "Enhanced expression of the human CD14 protein in tobacco using a 22-kDa alpha-zein signal peptide", *Plant Cell Tissue and Organ Culture* 112: 9-18.

Llaneras, F., Picó, J. (2008) "Stoichiometric modelling of cell metabolism", *Journal of Bioscience and Bioengineering* 105(1): 1-11.

Love, K., R., Politano, T., J., Panagiotou, V., Jiang, B., Stadheim, T., A., Love, J., C. (2012) "Systemic single-cell analysis of *Pichia pastoris* reveals secretory capacity limits productivity", *PLOS ONE* (e37915) 7 (6): 1-11.

Luang, S., Hrmova, M., Cairns, J., R., K. (2010) "High-level expression of barley β -D-glucan exohydrolase HvExoI from a codon-optimized cDNA in *Pichia pastoris*", *Protein Expression and Purification* 73: 90-98.

Maarleveld, T., R., Khandelwal, R., A., Olivier, B., G., Teusink, B., Bruggeman, F., J. (2013) "Basic concepts and principles of stoichiometric modeling of metabolic networks", *Biotechnology Journal* 8: 997-1008.

Macauley-Patrick, S., Fazenda, M.,L., McNeil, B., Harvey, L., M. (2005) "Heterologous protein production using the *Pichia pastoris* expression system", *Yeast* 22: 249-270.

Maier, T., Güell, M., Serrano, L. (2009) "Correlation of mRNA and protein in complex biological samples", *FEBS Letters* 583: 3966-3973.

Markaki, M., Drabek, D., Livadaras, I., Craig, R., K., Grosveld, F., Savakis, C. (2007) "Stable expression of human growth hormone over 50 generations in transgenic insect larvae", *Transgenic Research* 16: 99-107.

Martinez, J., L., Liu, L., Petranovic, D., Nielsen, J. (2012) "Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation", *Current Opinion in Biotechnology* 23(6): 965-971.

Martoglio, B. and Dobberstein, B. (1998) "Signal sequences: More than just greasy peptides", *Trends in Cell Biology* 8: 410-415.

Massahi, A. and Çalık, P. (2015) "In-silico determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production", *Journal of Theoretical Biology* 364: 179-188.

Massahi, A. and Çalık, P. (2016) "Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: From in-silico analysis to fermentation", *Journal of Theoretical Biology* 408: 22-33.

Matoba, S. and Ogrydziak, D.M. (1998) "Another factor besides hydrophobicity can affect signal peptide interaction with signal recognition particle", *Journal of Biological Chemistry* 273(30), 18841-18847.

Mattanovich, D., Gasser, B., Hohenblum, H., Sauer, M. (2004) "Stress in recombinant protein producing yeast", *Journal of Biotechnology* 113:121-135.

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*", *Microbial Cell Factories* 8:29.

Menendez, J., Valdes, I., Cabrera, N. (2003) "The *ICLI* gene of *Pichia pastoris*, transcriptional regulation and use of its promoter", *Yeast* 20: 1097-1108.

Molnar, I. and Horvath, C. (1976) "Reverse-phase chromatography of polar biological substances: Separation of catechol compounds by high-performance liquid chromatography", *Clinical Chemistry* 22(9): 1497-1502.

Monod, M., Haguenaer-Tsapis, R., Rauseo-Koenig, I., Hinnen, A. (1989) "Functional analysis of the signal-sequence processing site of yeast acid phosphatase", *European Journal of Biochemistry* 182: 213-221.

Moraes, A., M., Jorge, S., A., C., Astray, R., M., Suazo, C., A., T., Riquelme, C., A., C., Augusto, E., F., P., Tonso, A., Pamboukian, M., M., Piccoli, R., A., M., Barral, M., F., Pereira, C., A. (2012) "*Drosophila melanogaster* S2 cells for expression of heterologous genes: From gene cloning to bioprocess development", *Biotechnology Advances* 30: 630-628.

Mori, A., Hara, S., Sugahara, T., Kojima, T., Iwasaki, Y., Kawarasaki, Y., Sahara, T., Ohgiya, S., Nakano, H. (2015) "Signal peptide optimization tool for the secretion of recombinant protein from *Saccharomyces cerevisiae*" *Journal of Bioscience and Bioengineering* 120(5): 518-525.

Mori, K. (2009) "Signalling pathways in the unfolded protein response: development from yeast to mammals", *Journal of Biochemistry* 146(6):743-50.

Moriyama, S., Tanaka, H., Uwataki, M., Muguruma, M. And Ohta, K. (2003) "Molecular cloning and characterization of an exoinulinase gene from *Aspergillus niger* strain 12 and its expression in *Pichia pastoris*", *Journal of Bioscience and Bioengineering* 96(4): 324-331.

Mukhija, R., Rupa, P., Pillai, D., Garg, L., C. (1995) "High-level production and one-step purification of biologically active human growth hormone in *Escherichia coli*", *Gene* 165: 303-306.

Murasugi, A. and Tohma-Aiba, Y. (2001) "Comparison of three signals for secretory expression of recombinant human midkine in *Pichia pastoris*" *Bioscience, Biotechnology, and Biochemistry* 65 (10): 2291-2293.

Nagarajan, V. (1993) "Protein secretion , in *Bacillus subtilis* and other Gram-positive bacteria, biochemistry, physiology, and molecular genetics", Ed. Sonenshein, A., L., Hoch, J., A., and Losick, R. pp. 713-726. Washington DC: American Society for Microbiology.

Nakai, K. (2000) "Protein sorting signals and prediction of subcellular localization", *Advances in Protein Chemistry* 54: 277-344.

Nakayama, A., Ando, K., Kawamura, K., Mita, I., Fukazawa, K., Hori, M., Honjo, M., Furutani, Y. (1988) "Efficient secretion of the authentic mature human growth hormone by *Bacillus subtilis*", *Journal of Biotechnology* 8(2): 123-134.

National Center for Biotechnology Information, Sequence analysis, accessed 10 December 2016, <<https://www.ncbi.nlm.nih.gov/>>.

Nel, S., Labuschagne, M., Albertyn, J. (2009) "Advances in gene expression in non-conventional yeasts" T. Satyanarayana, G. Kunze (Eds.), *Yeast Biotechnology: Diversity and Applications*, Chapter 18 369-403.

Nelson, D., L., Cox, M., M. (2005) "Lehninger principles of biochemistry" 4th Edn., W. H. Freeman and Company, New York, NY.

Ng, D.T.W., Brown, J.D., Walter, P. (1996) "Signal sequences specify the targeting route to the endoplasmic reticulum membrane", *Journal of Cell Biology* 134(2), 269-278.

Nielsen, J. and Olsson, L. (2002) "An expanded role for microbial physiology in metabolic engineering and functional genomics: Moving towards systems biology", *FEMS Yeast Research* 2: 175-181.

Nielsen, J., Villadsen, J., Lidén, G. (2003) "Bioreaction engineering principles", 2nd Edn., Kluwer academics / Plenum publishers, New York, USA.

Niu, H., Jost, L., Pirlot, N., Sassi, H., Daukandt, M., Rodriguez, C., Fickers, P. (2013) "A quantitative study of methanol/sorbitol co-feeding process of a *Pichia pastoris* Mut⁺/pAOX1-lacZ strain", *Microbial Cell Factories* 12:33.

Nocon, J., Steiger, M., G., Pfeffer, M., Sohn, S., B., Kim, T., Y., Maurer, M., Rußmayr, H., Pflügl, S., Ask, M., Haberhauer-Troyer C., Ortmayr, K., Hann, S., Koellensperger, G., Gasser, B., Lee, S., Y., Mattanovich, D. (2014) "Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production", *Metabolic Engineering* 24: 129-138.

Oberhardt, M., A., Palsson, B., ø., Papin, J., A. (2009) "Application of genome-scale metabolic reconstructions", *Molecular Systems Biology* 5: 320.

Oka, C., Tanaka, M., Muraki, M., Harata, K., Suzuki, K. and Jigami, Y. (1999) "Human lysozyme secretion increased by alpha factor pro-sequence in *Pichia pastoris*", *Bioscience, Biotechnology, and Biochemistry* 63(11): 1977-1983.

Orman, M., A., Çalık, P., Özdamar, T., H. (2009) "The influence of carbon sources on recombinant human growth hormone production by *Pichia Pastoris* is dependent on phenotype: A comparison of Mut^s and Mut⁺ strains", *Biotechnology and Applied Biochemistry* 52: 245-255.

Outchkourov, N., S., Stiekema, W., J., Jongsma, M., A. (2002) "Optimization of the expression of equistatin in *Pichia pastoris*", *Protein Expression and Purification* 24: 18–24.

Özdamar, T., H., Şentürk, B., Yilmaz, Ö., D., Çalık, G., Çelik, E., Çalık, P. (2009) "Expression system for recombinant human growth hormone production from *Bacillus subtilis*", *Biotechnology Progress* 25(1): 75-84.

Öztürk, S., Çalık, P., Özdamar, T., H. (2016) "Fed-batch biomolecule production by *Bacillus subtilis*: A state of the art review", Trends in Biotechnology 34(4): 329-345.

Paifer, E., Margolles, E., Cremata, J., Montesino, R., Herrera, L., Delgado, J., M. (1994) "Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences", Yeast 10: 1415-1419.

Pais, J., M., Varas, L., Valdes, J., Cabello, C., Rodriguez, L., Mansur, M. (2003) "Modeling of mini-proinsulin production in *Pichia pastoris* using the AOX promoter", Biotechnology Letters 25: 251-255.

Pal, Y., Khushoo, A., Mukherjee, K., J. (2006) "Process optimization of constitutive human granulocyte-macrophage colony-stimulating factor (hGM-CSF) expression in *Pichia pastoris* fed-batch culture", Applied Microbiology and Biotechnology 69: 650-657.

Papini, M., Salazar, M., Nielsen, J. (2010) "Systems biology of industrial microorganisms", Advanced Biochemical Engineering and Biotechnology 120: 51-99.

Passoth, V. and Hahn-hagerdal, B. (2000) "Production of a heterologous endo-1,4- β -xylanase in the yeast *Pichia stipitis* with an O₂ -regulated promoter", Enzyme and Microbial Technology 26: 781-784.

Patra, A., K., Mukhopadhyay, R., Mukhija, R., Krishnan, A., Garg, L., C., Panda, A., K. (2000) "Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*", Protein Expression and Purification 18: 182-192.

Periyasamy, S., Govindappa, N., Sreenivas, S., Sastry, K. (2013) "Isolation, characterization and evaluation of the *Pichia pastoris* sorbitol dehydrogenase promoter for expression of heterologous proteins", Protein Expression and Purification 92: 128-133.

Porro, D., Sauer, M., Branduardi, P., Mattanovich, D. (2005) "Recombinant protein production in yeasts", Molecular Biotechnology 31: 245-259.

Potvin, G., Ahmad, A., Zhang, Z. (2012) "Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review", Biochemical Engineering Journal 64: 91– 105.

Prielhofer, R., Maurer, M., Klein, J., Wenger, J., Kiziak, C., Gasser, B., Mattanovich, D. (2013) "Induction without methanol: Novel regulated promoters enable high-level expression in *Pichia pastoris*", Microbial Cell Factories 12:5.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., Chu, J. (2011a) "GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*", Applied and Environmental Microbiology 77 (11): 3600-3608.

Qin, X., Qian, J., Zhuang, Y., Zhuang, S., Chu, J. (2011b) "Reliable high-throughput approach for screening of engineered constitutive promoters in yeast *Pichia pastoris*", Letters in Applied Microbiology 52: 634-641.

Raben, M., S. (1962) "Clinical use of human growth hormone", The New England Journal of Medicine 266(2): 82-86.

Raemaekers, R., J., M., de Muro, L., Gatehouse, J., A., Fordham-Skelton, A., P. (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", *European Journal of Biochemistry* 265: 394-403.

Ren, H., T., Yuan, J., Q., Bellgardt, K., H. (2003) "Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance", *Journal of Biotechnology* 106: 53-68.

Resina, D., Cos, O., Ferrer, P., Valero, F. (2005) "Developing high cell density fed-batch cultivation strategies for heterologous protein production in *Pichia pastoris* using the nitrogen source-regulated *FLD1* promoter", *Biotechnology and Bioengineering* 91(6): 760-767.

Resina, D., Maurer, M., Cos, O., Arnau, C., Carnicer, M., Marx, H., Gasser, B., Valero, F., Mattanovich, D., Ferrer, P. (2009) "Engineering of bottlenecks in *Rhizopus oryzae* lipase production in *Pichia pastoris* using the nitrogen source-regulated *FLD1* promoter", *New Biotechnology* 25(6): 396-403.

Rezaei, M., Zarkesh-Esfahani, S., H. (2012) "Optimization of production of recombinant human growth hormone in *Escherichia coli*", *Journal of Research in Medical Sciences* 17(7): 681-685.

Rezaei, M., Zarkesh-Esfahani, S., H., Gharagozloo, M. (2013) "The effect of different media composition and temperatures on the production of recombinant human growth hormone by CHO cells", *Research in Pharmaceutical Sciences* 8(3): 211-217.

Ribela, M., T., C., P., Camargo, I., M., C., Oliveira, J., E., Bartolini, P. (2000) "Single-step purification of recombinant human growth hormone (hGH) directly from bacterial osmotic shock fluids, for the purpose of I-hGH preparation", *Protein Expression and Purification* 18: 115-120.

Riffer, F., Eisfeld, K., Breinig, F., Schmitt, M., J. (2002) "Mutational analysis of K28 preprotoxin processing in the yeast *Saccharomyces cerevisiae*", *Microbiology* 148(5): 1317-1328.

Romanos, M., A., Scorer, C., A., Clare, J., J. (1992) "Foreign gene expression in yeast: A review", *Yeast* 8: 423-488.

Ruggiano, A., Foresti, O., Carvalho, P. (2014) "ER-associated degradation: Protein quality control and beyond", *Journal of Cell Biology* 204(6): 869-879.

Russell, D., A., Spatola, L., A., Dian, T., Paradkar, V., M., Dufield, D., R., Carroll, J., A., Schlittler, M., R. (2005) "Host limits to accurate human growth hormone production in multiple plant systems", *Biotechnology and Bioengineering* 89(7): 775-782.

Rutkowski, D., T., Ott, C., M., Polansky, J., R., Lingappa, V., R. (2003) "Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen", *Journal of Biological Chemistry* 278(32): 30365-30372.

Salamone, D., Baranao, L., Santos, C., Bussmann, L., Artuso, J., Werning, C., Prynck, A., Carbonetto, C., Dabsys, S., Munar, C., Salaberry, R., Berra, G., Berra, I., Fernandez, N., Papouchado, M., Foti, M., Judewicz, N., Mujica, I., Munoz, L., Alvarez, S., F., Gonzalez, E., Zimmermann, J., Criscuolo, M., Melo, C. (2006) "High level expression of bioactive recombinant human growth hormone in the milk of a cloned transgenic cow", *Journal of Biotechnology* 124: 469-472.

Sambrook, J., Russell, D.W. (2001) "Molecular Cloning: A Laboratory Manual", 3rd Edn., Cold Spring Harbor Library Press, Cold Spring Harbor, New York.

Sanchez, O., Toledo, J., R., Rodriguez, M., P., Castro, F., O. (2004) "Adenoviral vector mediates high expression levels of human growth hormone in the milk of mice and goats", *Journal of Biotechnology* 114: 89-97.

Sasagawa, T., Matsui, M., Kobayashi, Y., Otagiri, M., Moriya, S., Sakamoto, Y., Ito, Y., Lee, C., C., Kitamoto, K., Arioka, M. (2011) "High-throughput recombinant gene expression systems in *Pichia pastoris* using newly developed plasmid vectors", *Plasmid* 65: 65-69.

Schauer, R. (1972) "The mode of action of hormones", *Angewandte Chemie International Edition*, 11(1): 7-16.

Schmidt, F., R. (2004) "Recombinant expression systems in the pharmaceutical industry", *Applied Microbiology and Biotechnology* 65: 363-372.

Sears, I., B., O'Connor, J., Rossanese, O., W., Glick, B., S. (1998) "A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*", *Yeast* 14: 783-790.

Shang, L., Tian, P., Kim, N., Chang, H., N., Hahm, M., S. (2009) "Effects of oxygen supply modes on the production of human growth hormone in different scale bioreactors", *Chemical Engineering Technology* 32(4): 600-605.

Sharp, P., M., Li, W. (1987) "The codon adaptation index – a measure of directional synonymous codon usage bias, and its applications", *Nucleic Acids Research* 15(3): 1281-1295.

Shen, Q., Wu, M., Wang, H., Naranmandura, H., Chen, S. (2012) "The effect of gene copy number and co-expression of chaperone on production of albumin fusion proteins in *Pichia pastoris*", *Applied Microbiology and Biotechnology* 96: 763-772.

Shin, N., Kim, D., Shin, C., Hong, M., Lee, J., Shin, H. (1998) "High-level production of human growth hormone in *Escherichia coli* by a simple recombinant process", *Journal of Biotechnology* 62: 143-151.

Sinclair, G. and Choy, F., Y., M. (2002) "Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*", *Protein Expression and Purification* 26: 96–105.

Sinha, J., Plantz, B., A., Inan, M., Meagher, M., M. (2005) "Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant Ovine Interferon- τ ", *Biotechnology and Bioengineering* 89(1): 102-112.

Soetaert, W., Vandamme, E., J. (2010) "The scope and impact of industrial biotechnology, in industrial biotechnology: Sustainable growth and economic success (Eds. W. Soetaert and E. J. Vandamme), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Sohn, S., B., Graf, A., B., Kim, T., Y., Gasser, B., Maurer, M., Ferrer P., Mattanovich, D., Lee, S., Y. (2010) "Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in-silico analysis of heterologous protein production", *Biotechnology Journal* 5: 705-715.

Sola, A., Jouhten, P., Maaheimo, H., Sanchez-Ferrando, F., Szyperski, T., Ferrer, P. (2007) "Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates", *Microbiology* 153: 281-290.

Sola, A., Maaheimo, H., Ylönen, K., Ferrer, P., Szyperski, T. (2004) "Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*", *European Journal of Biochemistry* 271: 2462-2470.

Soorapaneni, S., Apte-Deshpande, A., Sabnis-Prasad, K., Kumar, J., Raiker, V., A., Kotwal, P., Padmanabhan, S. (2010) "Arabinose promoter based expression of biologically active recombinant human growth hormone in *E. coli*: strategies for overexpression and simple purification methods", *Journal of Microbial and Biochemical Technology* 2(2): 38-45.

Sreekrishna, K., Brankamp, R.,G., Kropp, K.,E., Blankenship, D.,T., Tsay, J., Smith, P.,L., Wierschke, J.,D., Subramaniam, A., Birkenberger, L.,A. (1997) "Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*", *Gene* 190: 55-62.

Stadlmayr, G., Mecklenbrauker, A., Rothmuller, M., Maurer, M., Sauer, M., Mattanovich, D., Gasser, B. (2010) "Identification and characterization of novel *Pichia pastoris* promoters for heterologous protein production", *Journal of Biotechnology* 150: 519-529.

Staley,C.,A., Huang, A., Nattestad, M., Oshiro, K.,T., Ray, L.,E., Mulye, T., Harry Li, Z., Le, T., Stephens, J.,J., Gomez, S., R., Moy, A., D., Nguyen, J., C., Franz, A., H., Lin-Cereghino, J., Lin-Cereghino, G. (2012) "Analysis of the 5' untranslated region (5' UTR) of the alcohol oxidase 1 (*AOX1*) gene in recombinant protein expression in *Pichia pastoris*", *Gene* 496: 118-127.

Stephanopoulos, G., N., Aristidou, A., A., Nielsen, J. (1998) "Metabolic engineering: Principles and methodologies" Academic Press, San Diego, CA.

Steuer, R., Junker, B., H. (2009) "Computational models of metabolism: Stability and regulation in metabolic networks", *Advances in Chemical Physics* 142: 105–251.

Stockholm Bioinformatics Center, Phobius: a combined transmembrane and signal peptide predictor, accessed 16 May 2014, <<http://phobius.binf.ku.dk/>>.

Stroud, R., M. and Walter, P. (1999) "Signal sequence recognition and protein targeting", *Current Opinion in Structural Biology* 9: 754-759.

Sung-Jae, L., Hong, I., Baek, S., Choi, S. (2007) "Deletion analysis of *Pichia PGK1* promoter and construction of an episomal vector for heterologous protein expression in *Pichia pastoris*", *Korean Journal of Microbiology and Biotechnology* 35 (3): 184-190.

Susanne Hiller-Sturmhöfel, S., Bartke, A. (1998) "The endocrine system: An overview", *Alcohol Health and Research World* 22(3): 153-164.

Swiss Institute of Bioinformatics, ExPASy ProtParam tool, accessed 12 July 2016, <<http://web.expasy.org/protparam/>>.

Tabandeh, F., Shojaosadati S., A., Zomorodipour, A., Khodabandeh, M., Sanati, M., H., Yakhchali, B. (2004) "Heat-induced production of human growth hormone by high cell density cultivation of recombinant *Escherichia coli*", *Biotechnology Letters* 26: 245-250.

Tanapongpipat, S., Promdonkoy, P., Watanabe, T., Tirasophon, W., Roongsawang, N., Chiba, Y., Eurwilaichitr, L. (2012) "Heterologous protein expression in *Pichia thermomethanolica* BCC16875, a thermotolerant methylotrophic yeast and characterization of N-linked glycosylation in secreted protein", FEMS Microbiology Letters 334: 127-134.

Tang, W., L., Zhao, H. (2009) "Industrial biotechnology: Tools and applications", Biotechnology Journal 4: 1725-1739.

Technical University of Denmark, Center for Biological Sequence Analysis, ProP 1.0, accessed 13 May 2014, <<http://www.cbs.dtu.dk/services/ProP/>>.

Technical University of Denmark, Center for Biological Sequence Analysis, SignalP 4.1, accessed 13 May 2014, <<http://www.cbs.dtu.dk/services/SignalP/>>

ThermoFischer Scientific, Restriction enzymes, accessed 14 December 2013, <www.thermoscientific.com/en/home.html>.

Tortajada, M., Llaneras, F., Ramon, D., Pico, J. (2012) "Estimation of recombinant protein production in *Pichia pastoris* based on a constraint-based model", Journal of Process Control 22: 1139-1151.

Tsakraklides, V., Brevnova, E., Stephanopoulos, G., Shaw, A., J. (2015) "Improved gene targeting through cell cycle synchronization", PLOS ONE 10(7): e0133434.

Uchida, H., Naito, N., Asada, N., Wada, M., Ikeda, M., Kobayashi, H., Asanagi, M., Mori, K., Fujita, Y., Konda, K., Kusuhara, N., Kamioka, T., Nakashima, K., Honjo, M. (1997) "Secretion of authentic 20-kDa human growth hormone (20K hGH) in *Escherichia coli* and properties of the purified product", *Journal of Biotechnology* 55: 101-112.

Van Den Hazel, H., B., Kielland-Brandt, M., C., Winther, J., R. (1996) "Review: Biosynthesis and function of yeast vacuolar proteases", *Yeast* 12: 1-16.

Vellanki, R., N., Narayana, K., Tatineni, R., Mangamoori, L., N. (2007) "Expression of hepatitis B surface antigen in *Saccharomyces cerevisiae* utilizing glyceraldehyde-3-phosphate dehydrogenase promoter of *Pichia pastoris*", *Biotechnology Letters* 29: 313-318.

Vogl, T. and Glieder, A. (2013) "Regulation of *Pichia pastoris* promoters and its consequences for protein production", *New Biotechnology* 30(4): 385-404.

Vogl, T., Hartner, F., S., Glieder, A., (2013) "New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*", *Current Opinion in Biotechnology*, 24: 1094-1101.

Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., Glieder, A. (2014) "Synthetic core promoters for *Pichia pastoris*", *ACS Synthetic Biology* 3: 188-191.

Walker, G., M. (1999) "Synchronization of yeast cell populations", *Methods in Cell Science* 21: 87-93.

Wallis, M. (1981) "The molecular evolution of pituitary growth hormone prolactin and placental lactogen: A protein family showing variable rates of evolution", *Journal of Molecular Evolution* 17: 10-18.

Walsh, G. (2014) "Biopharmaceutical benchmarks 2014", *Nature Biotechnology* 32(10): 992-1000.

Wang, H., Wang, Q., Zhang, F., Huang, Y., Ji, Y., Hou, Y. (2008) "Protein expression and purification of human Zbtb7A in *Pichia pastoris* via gene codon optimization and synthesis", *Protein Expression and Purification* 60: 97-102.

Wang, X., Sun, Y., Ke, F., Zhao, H., Liu, T., Xu, L., Liu, Y., Yan, Y. (2012) "Constitutive expression of *Yarrowia lipolytica* lipase LIP2 in *Pichia pastoris* using *GAP* as promoter", *Applied Biochemistry and Biotechnology* 166: 1355-1367.

Waterham, H., R., Digan, M., E., Koutz, P.,J., Lair, S., V., Cregg, J., M. (1997) "Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter", *Gene* 186: 37-44.

Web 1: Drugs.com, accessed 20 October 2016, < <http://www.drugs.com>>.

Web 2: BCC Research, accessed 11 September 2014, <<http://www.bccresearch.com/market-research>>.

Web 3: Global Business Intelligence, accessed 11 July 2016, <<http://www.marketresearch.com/GBI-Research-v3759/>>.

Web 4: The official site for the protocols of Dr. William D. Kelley, Victory over cancer, accessed 8 August 2016, <<http://www.drkelley.info/>>.

Web 5: The endocrine system, accessed 17 August 2016, <<http://thebiologyzone.wordpress.com>>.

Web 6: Ipsen Biopharmaceuticals, pituitary gland and pituitary hormones, access 19 August 2016, <<http://www.true-2-me.ca/>>.

Web 7: Leonid Andronov, Human growth hormone (Somatotropin), accesse 22 September 2016, <www.123rf.com>.

Web 8: The Universal Protein Resource, accessed 12 September 2016, <www.uniprot.org/uniprot/P01241/>.

Web 9: Tayyab, M., Process of recombinant DNA technology (Genetic engineering), 12 July 2016, <www.simplebiologyy.blogspot.com>.

Web 10: ACGT Inc., recombinant DNA technology, accessed 12 April 2012, <www.acgtinc.com>.

Web 11: Accessed 14 May 2013, <www.somatropin.net>.

Web 12: BioPharm International, accessed 19 October 2016 <<http://www.biopharminternational.com/>>.

Web 13: Invitrogen Co. , accessed 23 June 2013, <www.invitrogen.com>.

Web 14: Pro-Lab Diagnostics Inc., accessed 11 October 2016, <<http://www.pro-lab.com/>>.

Web 15: RESTEK Corporation, Pure Chromatography, access 7 November 2016, <<http://www.restek.com/>>.

Web 16: *Saccharomyces* genome database, accessed 18 November 2014, <www.yeastgenome.org>.

Wilhelm J., Pingoud A., Hahn, M. (2003) "Real-time PCR-based method for the estimation of genome sizes", *Nucleic Acids Research* 31(10): e56.

Wolff, A., M., Hansen, O., C., Poulsen, U., Madrid, S., Stougaard, P. (2001) "Optimization of the production of *Chondrus crispus* hexose oxidase in *Pichia pastoris*", *Protein Expression and Purification* 22: 189-199.

Wu, J., M., Chieng, L., L., Hsu, T., A., Lee, C., K. (2003b) "Sequential expression of recombinant proteins and their separate recovery from a *Pichia pastoris* cultivation", *Biochemical Engineering Journal* 16: 9-16.

Wu, J., M., Lin, J., C., Chieng, L., L., Lee, C., K., Hsu, T., A. (2003a) "Combined use of *GAP* and *AOX1* promoter to enhance the expression of human granulocyte-macrophage colony-stimulating factor in *Pichia pastoris*", *Enzyme and Microbial Technology* 33: 453-459.

Xie, J., Zhou, Q., Du, P., Gan, R., Ye, Q. (2005) "Use of different carbon sources in cultivation of recombinant *Pichia pastoris* for angiostatin production", *Enzyme and Microbial Technology* 36: 210-216.

Xiong, R., Chen, J., Chen, J. (2008) "Secreted expression of human lysozyme in the yeast *Pichia pastoris* under the direction of the signal peptide from human serum albumin", *Biotechnology and Applied Biochemistry* 51: 129-134.

Xu, J., Kieliszewski, M. (2011) "Enhanced accumulation of secreted human growth hormone by transgenic tobacco cells correlates with the introduction of an N-glycosylation site", *Journal of Biotechnology* 154: 54-59.

Xue, G., Johnson, J.S., Bransgrove, K.L., Gregg, K., Beard, C.E., Dalrymple, B.P., Gobius, K.S. and Aylward, J.H. (1997) "Improvement of expression and secretion of a fungal xylanase in the rumen bacterium *Butyrivibrio fibrisolvens* OB156 by manipulation of promoter and signal sequences", *Journal of Biotechnology* 54: 139–148.

Yamaji, H., Segawa, M., Nakamura, M., Katsuda, T., Kuwahara, M., Konishi, E. (2012) "Production of Japanese encephalitis virus-like particles using the *baculovirus*-insect cell system", *Journal of Bioscience and Bioengineering* 114(6): 657-662.

Yamanè, T., Shimizu, S. (1984) "Fed-batch techniques in microbial processes", *Advances in Biochemical Engineering/Biotechnology* 30: 147–194.

Yoshimasu, M.A., Ahn, J., Tanaka, T. and Yada, R.Y. (2002) "Soluble expression and purification of porcine pepsinogen from *Pichia pastoris*", *Protein Expression and Purification* 25: 229–236.

Young, K., T. and Park, T., H. (2007) "Expression of recombinant human growth hormone in a soluble form in *Escherichia coli* by slowing down the protein synthesis rate", *Journal of Microbiology and Biotechnology* 17(4): 579-585.

Yu, T., Barchetta, S., Pucciarelli, S., La Terza, A., Miceli, C. (2012) "A novel robust heat-inducible promoter for heterologous gene expression in *Tetrahymena thermophila*", *Protist* 163: 284-295.

Yusibov, V., M. and Mamedov, T., G. (2010) "Plants as an alternative system for expression of vaccine antigens", Proceedings of ANAS (Biological Sciences) 65(5-6): 195-200.

Zhan, X., Giorgianni, F., Desiderio, D., M. (2005) "Proteomics analysis of growth hormone isoforms in the human pituitary", Proteomics 5: 1228-1241.

Zhang, L., Shi, J., Jiang, D., Stupak, J., Ou, J., Qiu, Q., An, N., Li, J., Yang, D. (2012) "Expression and characterization of recombinant human alpha-antitrypsin in transgenic rice seed", Journal of Biotechnology 164: 300-308.

Zhang, W., Bevins, M., A., Plantz, B., A., Smith, L., A., Meagher, M., M. (2000) "Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A", Biotechnology and Bioengineering 70(1): 1-8.

Zhang, X., Sun, T., Liu, X., Gu, D., Huang, X. (1998) "Human growth hormone production by high cell density fermentation of recombinant *Escherichia coli*", Process Biochemistry 33(6): 683-686.

Zhang, Y., Liu, R., Wu., X. (2007) "The proteolytic systems and heterologous proteins degradation in the methylotrophic yeast *Pichia pastoris*", Annals of Microbiology 57(4): 553-560.

Zhao, W., Wang, J., Deng, R., Wang, X. (2008) "Scale-up fermentation of recombinant *Candida rugosa* lipase expressed in *Pichia pastoris* using the *GAP* promoter", Journal of Industrial Microbiology and Biotechnology 35: 189-195.

Zhu, T., Guo, M., Zhuang, Y., Chu, J., Zhang, S. (2011) "Understanding the effect of foreign gene dosage on the physiology of *Pichia pastoris* by transcriptional analysis of key genes", *Applied Microbiology and Biotechnology* 89: 1127-1135.

Zhu, X., Cheng, J., Huang, L., Gao, J., Zhang, Z., Pak, J., Wu, X. (2003) "Renal tubule-specific expression and urinary secretion of human growth hormone: a kidney-based transgenic bioreactor", *Transgenic Research* 12: 155-162.

Zimmermann, R., Eyrisch, S., Ahmad, M., Helms, V. (2011) "Protein translocation across the ER membrane", *Biochimica et Biophysica Acta*. 1808: 912-924.

APPENDICES

A. Details on immature hGH nucleotide sequence.

Human growth hormone (hGH) amino acid sequence was obtained from Uniprot database (<http://www.uniprot.org/uniprot/P01241>). The grey sequence represents the secretion signal peptide with D-score of 0.819 (calculated by SignalP program):

MATGSRTSLLLAFLGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAY
IPKEQKYSFLQNPQTSLCFSESIPTPSNREETQOKSNLELLRISLLLIQSWLEPVQFLRSV
FANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDALL
KNYGLLYCFRKDMDKVVETFLRIVQCRSVEGSCGF

Mature hGH nucleotide sequence used in current research:

TTC CCA ACT ATA CCA CTA TCT CGT CTA TTC GAT AAC GCT
ATG CTT CGT GCT CAT CGT CTT CAT CAG CTG GCC TTT GAC
ACC TAC CAG GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA
CAG AAG TAT TCA TTC CTG CAG AAC CCC CAG ACC TCC CTC
TGT TTC TCA GAG TCT ATT CCG ACA CCC TCC AAC AGG GAG
GAA ACA CAA CAG AAA TCC AAC CTA GAG CTG CTC CGC ATC
TCC CTG CTG CTC ATC CAG TCG TGG CTG GAG CCC GTG CAG
TTC CTC AGG AGT GTC TTC GCC AAC AGC CTG GTG TAC GGC
GCC TCT GAC AGC AAC GTC TAT GAC CTC CTA AAG GAC CTA
GAG GAA GGC ATC CAA ACG CTG ATG GGG AGG CTG GAA GAT
GGC AGC CCC CGG ACT GGG CAG ATC TTC AAG CAG ACC TAC
AGC AAG TTC GAC ACA AAC TCA CAC AAC GAT GAC GCA CTA
CTC AAG AAC TAC GGG CTG CTC TAC TGC TTC AGG AAG GAC
ATG GAC AAG GTC GAG ACA TTC CTG CGC ATC GTG CAG TGC
CGC TCT GTG GAG GGC AGC TGT GGC TTC

Following amino acid sequence was obtained by translating available nucleotide sequence of mature hGH (without secretion signal peptide) by EMBOSS Transeq tool (<http://www.ebi.ac.uk/>):

```
FPTIPLSRLFDNAMLRHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPT  
PSNREETQOKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEG  
IQTLMGRILEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIV  
QCRSVEGSCGF
```

In order to compare two sequences, The NCBI protein BLAST was performed and 100% identical result was obtained.

In order to further confirm the available sequence of the secretion signal peptide, the abovementioned nucleotide sequence was searched in NCBI and the following sequence was obtained for immature protein mRNA nucleotide sequence:

```
AGGATCCCAAGGCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCTCACCTAGCTGC  
AATGGCTACAGGCTCCCGGACGTCCCTGCTCCTGGCTTTTGGCC  
TGCTCTGCCTGCCCTGGCTTCAAGAGGGCAGTGCCTTCCCAACCAT  
TCCCTTATCCAGGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACCAGCTGGCC  
TTTGACACCTACCAGGAGTTTGAAGAAGCCTATATCCCAAAGGAACAGAAGTATTCATTCC  
TGCAGAACCCCCAGACCTCCCTCTGTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGA  
GGAAACACAACAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCG  
TGGCTGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCT  
CTGACAGCAACGTCTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAACGCTGATGAG  
GAGGCTGGAAGATGGCAGCCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTC  
GACACAAACTCACACAACGATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCA  
GGAAGGACATGGACAAGGTCGAGACATTCCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGG  
CAGCTGTGGCTTCTAGCTGCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCCTCCT  
GGCCCTGGAAGTTGCCACTCCAGTGCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATC  
A
```

The grey underlined parts were the start and the end which corresponds to our sequence. The two sequences (one we have and the one between two grey underlined parts) were a little different, 16 nucleotides; however, we just wanted the secretion signal peptide part which will be immediately before mature sequence namely, the bolded blue part with 78 bp length

which is equivalent to 26 amino acids. This part was translated into amino acid sequence by EMBOSS Transeq tool (<http://www.ebi.ac.uk/>) and the following sequence was obtained:

MATGSRTSLLLAFLGLLCLPWLQEGSA

This sequence is the same as the sequence obtained from Uniprot at the start of this appendix for complete human growth hormone amino acid sequence. Therefore, we can use the nucleotide sequence in front of our mature hGH nucleotide sequence to have complete nucleotide sequence of immature hGH; in the following nucleotide sequence of the immature hGH, the underlined blue part is the secretion signal peptide:

ATG GCT ACA GGC TCC CGG ACG TCC CTG CTC CTG GCT TTT
GGC CTG CTC TGC CTG CCC TGG CTT CAA GAG GGC AGT GCC
TTC CCA ACT ATA CCA CTA TCT CGT CTA TTC GAT AAC GCT
ATG CTT CGT GCT CAT CGT CTT CAT CAG CTG GCC TTT GAC
ACC TAC CAG GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA
CAG AAG TAT TCA TTC CTG CAG AAC CCC CAG ACC TCC CTC
TGT TTC TCA GAG TCT ATT CCG ACA CCC TCC AAC AGG GAG
GAA ACA CAA CAG AAA TCC AAC CTA GAG CTG CTC CGC ATC
TCC CTG CTG CTC ATC CAG TCG TGG CTG GAG CCC GTG CAG
TTC CTC AGG AGT GTC TTC GCC AAC AGC CTG GTG TAC GGC
GCC TCT GAC AGC AAC GTC TAT GAC CTC CTA AAG GAC CTA
GAG GAA GGC ATC CAA ACG CTG ATG GGG AGG CTG GAA GAT
GGC AGC CCC CGG ACT GGG CAG ATC TTC AAG CAG ACC TAC
AGC AAG TTC GAC ACA AAC TCA CAC AAC GAT GAC GCA CTA
CTC AAG AAC TAC GGG CTG CTC TAC TGC TTC AGG AAG GAC
ATG GAC AAG GTC GAG ACA TTC CTG CGC ATC GTG CAG TGC
CGC TCT GTG GAG GGC AGC TGT GGC TTC

B. Media, buffers, and stock solutions.

Medium / Solution	Recipe
LSLB	<p>Dissolve 1 g tryptone, 0.5 g yeast extract, and 0.5 g NaCl in 80 mL dH₂O, after pH adjustment to 7.5 with 1 N NaOH (0.2 N NaOH can be used), finalize the volume to 100 mL and autoclave.</p> <p>If making LSLB slants or plates, include 1.5 g agar in 100 mL.</p>
LSLB + ZeocinTM / ampicillin for <i>E. coli</i>	<p>Add 25 µL ZeocinTM 100 mg/mL to 100 mL of LSLB after being autoclaved and cooled to < 55 °C.</p> <p>Add 100 µL ampicillin 100 mg/mL to 100 mL LSLB after being autoclaved and cooled to < 55 °C.</p>
YPD	<p>Dissolve 1 g yeast extract, 2 g peptone, and 2 g dextrose (glucose) in 100 mL dH₂O and autoclave.</p> <p>If making YPD slants or plates, include 2 g agar in 100 mL.</p> <ul style="list-style-type: none"> • If possible autoclave glucose separately (or filter-sterilize separately).
YPD+ ZeocinTM for <i>P. pastoris</i>	<p>Add 100 µL ZeocinTM 100 mg/mL to 100 mL of YPD after being autoclave and cooled to < 55 °C.</p> <ul style="list-style-type: none"> • Liquid mediums without ZeocinTM can be stored at room temperature. • Mediums with ZeocinTM should be stored at +4 °C in the dark. • YPD slants or plates have shelf life of several months. Mediums with ZeocinTM have shelf life of one to two weeks.

BSM	<p>Dissolve all the ingredients mentioned in Table 3.6 in dH₂O and complete the volume to 1L. Sterilize by autoclaving and, then, add PTM1 and antifoam and chloramphenicol.</p> <ul style="list-style-type: none"> • BSM will be a clear solution with pH of 2.0
YEP for starvation	<p>Dissolve 1 g yeast extract and 2 g peptone in 100 mL dH₂O. Sterilize by autoclaving.</p>
KiP buffer 1 M (pH=6)	<p>Dissolve 0.92 g K₂HPO₄ and 4.73 g KH₂PO₄ in 35 mL dH₂O and adjust pH to 6 and, then, complete the volume to 40 mL. Sterilize by autoclaving.</p> <ul style="list-style-type: none"> • pH adjustment to 6.0±0.1 with phosphoric acid or KOH.
10X YNB + (NH₄)₂SO₄ solution	<p>Dissolve 1.36 g YNB (without amino acids and ammonium sulfate) and 4 g (NH₄)₂SO₄ in 40 mL dH₂O. Cover the container with aluminum foil. Autoclave separately.</p>
10X glycerol solution	<p>Mix 4.6 mL glycerol 87% and 35.4 mL dH₂O. Sterilize by Autoclaving separately.</p> <ul style="list-style-type: none"> • 4 mL Glycerol 100% is equivalent to 4.6 mL Glycerol 87%; based on the density of the 87% solution, 1.23 g/mL, 5.66 g glycerol 87% was mixed with 35.4 mL dH₂O.
Chloramphenicol (34 mg/mL)	<p>Dissolve 0.34 g chloramphenicol in 10 mL EtOH 100% and filter-sterilize. Store at -20°C in a foiled dark container.</p>

Biotin 500X	Dissolve 0.01 g biotin in 50 mL dH ₂ O and filter-sterilize. Store at +4°C in a foiled container.
BMGY (300 mL)	Dissolve 3 g yeast extract and 6 g peptone in ~210 mL dH ₂ O and, then, add 30 mL KiP buffer 1 M, 30 mL YNB + (NH ₄) ₂ SO ₄ 10X solution, 30 mL glycerol 10X solution, 0.6 mL biotin 500X, and 0.3 mL chloramphenicol 34 mg/mL. <ul style="list-style-type: none"> This recipe is the same as the one in the “<i>Pichia</i> expression kit for expression of recombinant proteins in <i>Pichia pastoris</i>”, User Guide 2014 for 1L BMGY (just edited for 300 mL)!
PTM1	Dissolve all the components mentioned in (Table 3.) in 80 mL dH ₂ O and after adjusting volume to 100 mL, filter-sterilize and store at +4°C in a foiled container. Discard the solution when it turned to yellowish green. (<i>Pichia</i> fermentation process guidelines - Invitrogen)
Salts solution (CaSO₄.2H₂O + MgSO₄.7H₂O)	Dissolve 0.35 g CaSO ₄ .2H ₂ O and 4.47 gr MgSO ₄ .7H ₂ O in 180 mL dH ₂ O. Sterilize by autoclaving. <ul style="list-style-type: none"> They are hardly dissolved and after autoclaving there may be a cloudy suspension in the bottom of the container which can be dissolved by shaking vigorously.
Production medium for shake flask bioreactor experiments (300 mL)	Mix 30 mL KiP buffer 1M (pH=6), 180 mL salts solution, 40 mL (NH ₄) ₂ SO ₄ solution (containing 2.88 g ammonium sulfate and sterilized by autoclaving), 50 mL glucose solution (containing 6 g glucose and filter-sterilized) and add 1.305 mL PTM1 solution 0.15 mL chloramphenicol 34 mg/mL.

80 mM MgCl₂ – 20mM CaCl₂	Dissolve 7.6 g MgCl ₂ and 2.2 g CaCl ₂ in 800 mL dH ₂ O and finalize the volume to 1000 mL. After filter-sterilization store at +4°C.
0.1 M CaCl₂	Dissolve 11 g CaCl ₂ in 800 mL dH ₂ O and finalize the volume to 1 L. Filter-sterilize and store at +4°C.
1 M Tris-Cl (various pHs)	<p>Dissolve 121 g Tris base in 800 mL of dH₂O. Adjust pH to the desired value by adding approximately the following:</p> <p>pH = 7.4 about 70 ml of concentrated HCl pH = 7.6 about 60 ml of concentrated HCl pH = 8.0 about 42 ml of concentrated HCl</p> <ul style="list-style-type: none"> • Make sure solution is at room temperature before making final pH adjustments (http://userpages.umbc.edu) <p>Finalize the volume to 1L. Sterilize by autoclaving</p> <ul style="list-style-type: none"> • If the 1 M solution of Tris has a yellow color, discard it and obtain Tris of better quality. The pH of Tris solutions is temperature-dependent and decreases ~0.03 pH units for each 1°C increase in temperature. For example, a 0.05 M solution has pH values of 9.5, 8.9, and 8.6 at 5°C, 25°C, and 37°C, respectively (http://cshprotocols.cshlp.org).
Alkaline lysis solution I	Prepare 100 mL. 50 mM glucose, 25 mM Tris-Cl (pH=8) ~ 2.5 mL Tris-Cl 1M; EDTA (pH=8) 10 mM. Sterilize by autoclaving. Store at +4°C.
Alkaline lysis solution II	Prepare 5 or 10 mL. Dissolve 0.1 g Sodium dodecyl sulfate in 10mL NaOH 0.2 N. (Should be freshly-prepared)

Alkaline lysis solution III	Mix 60 mL potassium acetate 5M, 11.5 mL glacial acetic acid, and 28.5 mL dH ₂ O. Store the solution at +4°C and transfer to an ice bucket just prior to use.
EDTA 0.5 M	Dissolve 37.22 g Na ₂ EDTA.2H ₂ O in 160 mL dH ₂ O. Stir vigorously using magnetic stirrer and add NaOH to adjust the pH to 8. Toward the end of the pH adjustment it is better to use NaOH solution instead solid NaOH. EDTA goes slowly to the solution as pH becomes closer to 8. Then, complete the volume to 200 mL. Autoclave and store at room temperature. Na ₂ EDTA salt is a chelator of divalent metal cations.
TE buffer (pH=8)	Add 1 mL Tris-Cl 1M and 0.2 mL EDTA 0.5 M to 90 mL dH ₂ O adjust the pH by NaOH or HCL and finalize the volume to 100 mL. Autoclave and store at room temperature.
5X TBE buffer	Dissolve 54 g Tris base and 27.5 g boric acid in 900 mL dH ₂ O. Add 20 mL 0.5 M EDTA. Finalize the volume to 1 L by dH ₂ O. The pH of the concentrated stock should be ~ 8.3. <ul style="list-style-type: none"> • 1X solution should be prepared just before usage.
1M LiCl	Dissolve 0.424 g LiCl in 8 mL dH ₂ O and complete the volume to 10 mL. Filter-sterilize and store at +4°C. Dilute by sterile dH ₂ O for preparation of 0.1 M LiCl.

<p>50% PEG</p>	<p>Dissolve 25 g PEG in 30 mL dH₂O by gentle heating and after dissolution finalize the volume to 50 mL. Filter-sterilize. Seal tightly (to prevent dH₂O loss). Can be stored at room temperature, +4°C or -20°C.</p> <ul style="list-style-type: none"> • PEG shields the yeast cells from the harsh effects of the high concentration of LiCl. • PEG is crucial ingredient. Prepare in small batches; do not use old PEG.
<p>Yeast lysis solution (for genomic DNA isolation)</p>	<p>Dissolve 2 g Triton X-100, 1 g SDS, 0.58 g NaCl in approximately 80 mL dH₂O. Then, add 1 mL Tris-Cl (pH=8), 0.2 mL EDTA 0.5 M. Complete the volume to 100 mL. Autoclave (or filter-sterilize) and store at room temperature.</p>
<p>Sodium acetate (3M)</p>	<p>Dissolve 246 g CH₃COONa in 700 mL dH₂O and complete the volume to 1 L. Filter-sterilize. Store at room temperature.</p>
<p>5X SDS-PAGE running buffer</p>	<p>Dissolve 15 g Tris-base, 72 g glycine, and 5 g SDS in 600 mL dH₂O and complete the volume to 1 L by dH₂O. Store at +4°C.</p> <ul style="list-style-type: none"> • 1X solution should be 25 mM Tris base; 192 mM glycine; 0.1% SDS.

4X SDS-PAGE loading buffer	<p>Mix 8 mL 1M Tris-HCl (pH=6.8) with 16 mL glycerol 100% and 16 mL 1% (w/v) bromophenol blue and dissolve 3.2 g SDS in it. Store the buffer without DTT at room temperature. Add DTT from a 1 M stock just before the buffer is used to have a DTT concentration of 100 mM.</p> <ul style="list-style-type: none"> • 200 mM β-mercaptoethanol can be used instead of DTT.
Fixer solution	<p>Mix 100 mL MeOH, 24 mL acetic acid, 100 μL 37% formaldehyde and complete the volume to 200 mL with dH₂O. Solution can be used during one month.</p> <ul style="list-style-type: none"> • Treating the gel with acid makes the macromolecules insoluble in the gel and prevents them diffusing out of the gel. Furthermore, interfering substances are removed in this step.
EtOH solution (50%)	<p>Mix 75 mL pure EtOH with 75 mL dH₂O. (Should be freshly-prepared)</p>
Pretreatment solution	<p>Dissolve 0.05 g Na₂S₂O₃·5H₂O in 250 mL dH₂O; take 2 mL for further use in developing solution. (Should be freshly-prepared)</p> <ul style="list-style-type: none"> • This step in SDS renders the proteins more reactive toward silver leading to faster reduction of silver ions.
Silver nitrate solution	<p>Dissolve 0.2 g AgNO₃ in 100 mL dH₂O and add 75μL 37% formaldehyde. (Should be freshly-prepared)</p> <ul style="list-style-type: none"> • Mild acidic conditions prevent silver ions from being reduced to metallic silver. Excess silver is removed from the gel surface.

Developing solution	<p>Dissolve 2.25 g potassium carbonate in 1mL dH₂O, add 2 mL from pretreatment solution and 75 µL 37% formaldehyde.</p> <p>(Should be freshly-prepared)</p> <ul style="list-style-type: none"> • Formaldehyde renders silver ion to metallic silver. Since high pH is required, therefore, potassium carbonate is included.
Stop solution	<p>Mix 50 mL MeOH, 12 mL acetic acid and complete the volume to 100 mL by dH₂O.</p> <p>(Should be freshly-prepared)</p> <ul style="list-style-type: none"> • Prevents further reduction of silver ions.
Transfer buffer (Blotting buffer)	<p>25 mM Tris; 190 mM Glycine; 20% Methanol. Adjust pH to 8.3 if necessary.</p> <ul style="list-style-type: none"> • 20% methanol is generally optimal for protein binding.
MES buffer	<p>20 mM 2-(N-morpholine)-ethanesulfonic acid, 0.1 M sodium chloride; pH 5.0.</p>
Mobile phase (HPLC)	<p>Dissolve 3.12 g NaH₂PO₄ in 900 mL dH₂O and adjust the pH by H₃PO₄ 85% (620 µL) to ~ 2.7; complete the volume to 1L. ; filter with 0.45 µm membrane and degas for 15 minutes.</p>
Seal wash (HPLC)	<p>Mix 200 mL acetonitrile with 800 mL dH₂O; filter with 0.45 µm membrane and degas for 15 minutes.</p>

10X TBS (pH~7.6)	<p>Dissolve 24 g Tris base and 87.66 g NaCl in 800 (to 900) mL dH₂O and adjust pH to 7.6 with HCl. Complete the volume to 1 L and store at room temperature or +4°C.</p> <ul style="list-style-type: none"> • The pH value should be a little alkaline, between 7.4 and 8.0 (www.abcam.com). • 1X TBS should be 150 mM NaCl and 20 mM Tris-base.
TBS-T	<p>Mix 100 mL TBS 10X with 900 mL dH₂O and add 1 mL Tween 20. Adjust the pH to 7.6. Prepare at the day of usage.</p>
TBS-T milk	<p>Dissolve 5% (w/v) non-fat milk powder in TBS-T. Prepare at the day of usage. Store at 2-8°C.</p>
Lyticase enzyme (≥ 200U/mL)	<p>Dissolve 2 mg powder (#L4025 from Sigma) in 0.2 mL potassium phosphate buffer 1M (pH=7.5) then add 0.040 mL NaCl 5M and 1 mL glycerol. Complete the volume to 2 mL by dH₂O.</p>
Stacker gel	<p>It is prepared based on the information in Bio-Rad's TGX™ and TGX stain-free™ FastCast™ acrylamide kits.</p>
Resolving gel	<p>It is prepared based on the information in Bio-Rad's TGX™ and TGX stain-free™ FastCast™ acrylamide kits.</p>
APS 10%	<p>Dissolve 0.1 g APS in 1 mL dH₂O in a clean 1.5-mL eppendorf tube. It is better to be prepared freshly.</p>

<p>Sodium acetate buffer (0.05 M)</p>	<p>Prepare 0.5 M solution of acetic acid by mixing 2.85 mL acetic acid with dH₂O and completing the volume to 100 mL by dH₂O. Prepare 0.5 M solution of sodium acetate by dissolving 4.103 g sodium acetate in 40 mL dH₂O and completing the volume to 100 mL by dH₂O. Mix approximately 16 mL of acetic acid 0.5 M by approximately 34 mL of sodium acetate 0.5 M to reach to pH of 5. Dilute 10 times (to 500 mL). Autoclave and store at +4 °C.</p>
<p>Casein solution 5% (w/v)</p>	<p>Dissolve 0.5 g casein in 100 mL of 0.05 M sodium acetate buffer. (Should be freshly prepared)</p> <ul style="list-style-type: none"> • Casein is hardly soluble in acidic buffer and the suspension should be homogenized by the help of a magnet in a bottle.
<p>Trichloroacetic acid solution 10% (w/v)</p>	<p>Dissolve 10 g TCA in 40 mL dH₂O and complete the volume to 100 mL. (Should be freshly prepared)</p>

D. Primers recruited for PCR and DNA sequencing.

Primer Name	Primer Nucleotide Sequence
GAP forward	5'- GTCCCTATTTCAATCAATTGAA-3'
AOX reverse	5'- GCAAATGGCATTCTGACATCC-3'
PDC-F	5'...TCTATGCATGAGATCGGGACAAGCA...3'
PDC-R	5'...TGCGAATTCAGCTTCAGCCTCTCTTT...3'
PYRK-F	5'...TAAATGCATGAGATCTTCAGTGTGCGG...3'
PYRK-R	5'...TGAGAATTCAGCTTCAGCCTCTCTTTTCT...3'
THI3-F	5'...GAGATGCATGAGATCGTCTTTGTAAATAGT...3'
THI3-R	5'... TGAGAATTCAGCTTCAGCCTCTCTTTTC...3'
New GAP fwd	5'... GACGCATGTCATGAGATTATTGG ...3'
PYRK End-fwd	5'...GACCGTTCATGTACAGTAAATTG...3'
PDC End-fwd	5'...ACCTCAAATATCTAGCAACATCTT...3'
THI3 End-fwd	5'...GGAATGTTGGTAGCAATTGGTAT...3'
SP13-F	5'....gcgttcgaaACGatgctatcaactatc....3'
SP13-R	5'....agagaattcctgtagggatgcctgtat.... 3'
SP23-F	5'.... gcgttcgaaACGatgaaaatattaagtgc...3'
SP23-R	5'....acagaattcagcaaaggccaacgtaaa....3'
SP24-F	5'.... cagttcgaaACGatgaaagtttctacg....3'
SP24-R	5'.... atagaattcagcgcaaacgagttaac.... 3'
SP26-F	5'... atattcgaaACGatgtggtcgtgttcat...3'
SP26-R	5'...gaggaattctccaaggaccaaaggatagaag...3'
SP34-F	5'.... atattcgaaACGatgagaccagtgc... 3'
SP34-R	5' aacgaattcagcgagtaccgaaga.... 3'

E. Reaction condition for single-digestion/double-digestion with utilized restriction enzymes in current research.

E.1. *EcoRI* restriction enzyme

The *EcoRI* restriction enzyme cuts best at 37°C in its own unique buffer. Thermal inactivation occurs at 65°C in 20 minutes.

Reaction conditions:

Recommended buffer for 100% activity	Optimal Temperature	Enzyme activity in ThermoScientific buffers, %						Tango buffer For double-digestion
		B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)		
						1X	2X	
<i>EcoRI</i> buffer (Unique)	37°C	0-20	NR	100	100	NR	100	2X

NR: Not recommended, because of high star activity or restriction enzyme activity is less than 20%.

E.2. *XbaI* restriction enzyme

The *XbaI* restriction enzyme cuts best at 37°C in Tango buffer. Thermal inactivation occurs at 65°C in 20 minutes.

Reaction conditions:

Recommended buffer for 100% activity	Optimal Temperature	Enzyme activity in ThermoScientific buffers, %						Tango buffer For double-digestion
		B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)		
						1X	2X	
Tango	37°C	50-100	50-100	20-50	0-20	100	50-100	1X or 2X

E.3. *Bsp119I* restriction enzyme

The *Bsp119I* (*BstBI*) restriction enzyme cuts best at 37°C in Tango buffer. Thermal inactivation occurs at 80°C in 20 minutes.

Reaction conditions:

Recommended buffer for 100% activity	Optimal Temperature	Enzyme activity in ThermoScientific buffers, %						Tango buffer For double-digestion
		B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)		
						1X	2X	
Tango	37°C	20-50	0-20	0-20	0-20	100	100	1X or 2X

E.4. *NsiI* (*Mph1103I*) restriction enzyme

The *Mph1103I* (*NsiI*) restriction enzyme cuts best at 37°C in R buffer. Thermal inactivation occurs at 65°C in 20 minutes.

Reaction conditions:

Recommended buffer for 100% activity	Optimal Temperature	Enzyme activity in ThermoScientific buffers, %						Tango buffer For double-digestion
		B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)		
						1X	2X	
R (Red)	37°C	0-20	50-100	20-50	100	50-100	50-100	1X or 2X

E.5. Double-digestion with *EcoRI* & *XbaI*

The prepared mixture for reaction:

2X Tango buffer, *EcoRI*, 2-fold excess of *XbaI*

Incubate at 37°C.

Reaction conditions for restriction enzymes in ThermoScientific five buffer system:

Restriction enzyme	Optimal temperature	Units for overnight incubation	Thermal inactivation	Enzyme activity in ThermoScientific buffers, %					
				B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)	
								1X	2X
<i>EcoRI</i>	37°C	0.2	65°C	NR	NR	100	100	NR	100
<i>XbaI</i>	37°C	0.1	65°C	50-100	50-100	20-50	NR	100	50-100

NR: Not recommended, because of high star activity or restriction enzyme activity is less than 20%.

E.6. Double-digestion with *EcoRI* & *Bsp119I*

The prepared mixture for reaction:

2X Tango buffer, *Bsp119I* (*BstBI*), *EcoRI*

Incubate at 37°C

Reaction conditions for restriction enzymes in ThermoScientific five buffer system:

Restriction enzyme	Optimal temperature	Units for overnight incubation	Thermal inactivation	Enzyme activity in ThermoScientific buffers, %					
				B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)	
								1X	2X
<i>Bsp119I</i>	37°C	0.1	80°C	20-50	NR	NR	NR	100	100
<i>EcoRI</i>	37°C	0.2	65°C	NR	NR	100	100	NR	100

NR: Not recommended, because of high star activity or restriction enzyme activity is less than 20%.

E.7. Double-digestion with *EcoRI* & *NsiI*

The prepared mixture for reaction:

2X Tango buffer, *EcoRI*, 2-fold excess of *Mph1103I* (*NsiI*)

Incubate at 37°C.

Reaction conditions for restriction enzymes in ThermoScientific five buffer system:

Restriction enzyme	Optimal temperature	Units for overnight incubation	Thermal inactivation	Enzyme activity in ThermoScientific buffers, %					
				B (blue) 1X	G (Green) 1X	O (Orange) 1X	R (Red) 1X	Tango (Yellow)	
								1X	2X
<i>EcoRI</i>	37°C	0.2	65°C	NR	NR	100	100	NR	100
<i>NsiI</i>	37°C	0.3	65°C	NR	50-100	20-50	100	50-100	50-100

NR: Not recommended, because of high star activity or restriction enzyme activity is less than 20%.

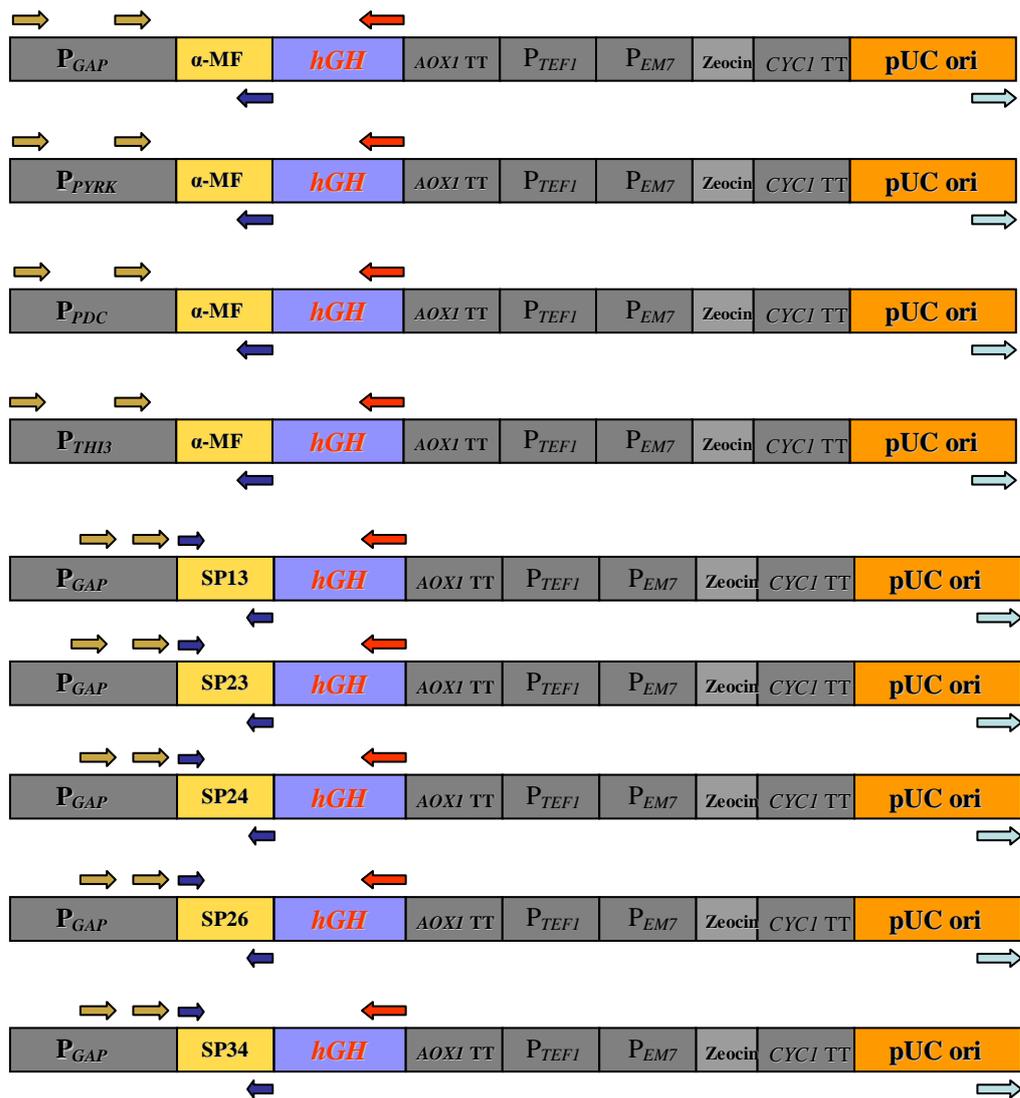
F. Location of the pUC-ori forward and hGH reverse primers used in pre-screening PCRs

The red highlights are location of the primers:

```
AATTCACCATCACCATCACCATATTGAAGGGAGA ttcccaactataccactatctcgtctattcgataacgctagt
cttcgctgctcatcgtcttcatcagctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaaca
gaagtattcattcctgcagaacccccagacctcctctgttttctcagagtctatcccgacacctccaacagggagg
aaacacaacagaaatccaacctagagctgctccgcatctccctgctgctcatccagtcgtggctggagcccgtgag
ttctcaggagtgcttctgccaacagcctgggtgtacggcgctctgacagcaacgtctatgacctcctaaggacct
agaggaaggcatccaaacgctgatggggaggctggaagatggcagccccggactgggcagatcttcaagcagacct
acagcaagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggctgctctactgc ttcaggaag
gacatggacaaggtcggagacat tctgctcagctgctgagtgccgctctgtggagggcagctgtggcttcTAGT. CTA
GAACAAAACTCATCTCAGAAGAGGATCTGAATAGCGCGTTCGACCATCATCATCATCATCATGATTGAGTTTGTAGCCT
TAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGAGGATGT
CAGAATGCCATTTGCCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTT
TTTTTGTCATTTTTGTCTTCTCCTGACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGTGTA
GGGGTTTGGGAAAATCATTCGAGTTTGTATTTTTTCTTGGTATTTCCCACTCCTCTCAGAGTACAGAAGATTAAGT
GAGACCTTCGTTTTGTGCGGATCCCCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTT
TCTCGGACTCCGCGCATCGCCGTACCCTTCAAAACACCCAAGCACAGCATACTAAATTTCCCTCTTCTCCTCT
AGGGTGTCTGTTAATACCGTACTAAAGTTTGGAAAAAGAAAAAGAGACCGCCTCGTTTTCTTTTCTTCGTCGAAA
AAGGCAATAAAAAATTTTTATCAGTTTTCTTTTTCTTGAATTTTTTTTTTTAGTTTTTCTCTTTCAGTGACCTCC
ATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAAGTTTCAATTTCTTCTGTTCTATTACAACTTTTT
TTACTTCTTGTTCATTAGAAAAGCAATAGCAATCTAATCTAAGGGCGGTGTTGACAATTAATCATCGGCATAGTA
TATCGGCATAGTATAATACGACAAGGTGAGGAATAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGTCCACCG
CGCGCAGCTCGCCGAGCGGTGAGTTCTGACCGACCGGCTCGGGTCTCCCGGACTTCGTGGAGGACGACTTC
GCCGGTGTGGTCCGGGACGACGTGACCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCTGGC
CTGGGTGTGGGTGCGCGCCTGGACGAGTGTACGCCGAGTGGTCCGAGGTCTGTCCACGAACCTCCGGGACGCT
CCGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGCGGGAGTTCCGCTGCGCGACCCGGCCGCACTGC
GTGCACCTTCGTGGCCGAGGAGCAGGACTGACACGTCCGACGGCCGCCCCAGGGTCCAGGCCTCGGAGATCCGTCCC
CCTTTTCTTTGTGATATCATGTAATTAGTTATGTACGCTTACATTCAGGCCCTCCCCCACATCCGCTCTAACCC
GAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTATAGTTATGTTAGTATTAAGAACGTT
ATTTATATTTCAAATTTTTCTTTTTTCTGTACAGACGCGGTGACGCATGTAACATTAATACTGAAAACCTTGCTG
AGAAGTTTTTGGGACGCTCGAAGCTTTAATTTGCAAGCTGGAGACCAACATGTGAGCAAAAAGGCCAGCAAAAAGCC
AGGAACCGTAAAAAGGCCGCTTGTGCGGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAGAAAATCGAGC
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGC
TCACGCTGTAGGTATCTCAGTTCCGTTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCAGAACCCCGCTTACGCC
CGACCGCTGCCCTTATCCGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACAGACTTATCGCCACTGGCAGCAC
CCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTTCTTGAAAGTGGTGGCCTAACTACGGC
TACACTAGAAGGACAGTATTGGTATCTGCGCTGTGTAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTG
ATCCGGCAAAACAAACCACCGCTGTAGCGGTGGTTTTTTTTGTTGCAAGCAGCAGATTACCGCCAGAAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTCTACGGGTCTGACGCTCAGTGGAAACGAAAACCTACGTTAAGGATTTTGGTGC
ATGCATGAGATC
```

G. The linearized constructs (with *NsiI* restriction enzyme).

The schematic representation of the single-digested constructs in order to imagine the pre-screening step. The arrows show the approximate location of the primers designed in this study:



H. Location of the designed primers for preparation of standard solutions and rendering qPCR experiments.

H.1. *hGH* gene specific primers

After preparation of the *hGH* gene standard solution by *GAP* forward and *AOX* reverse primers (Appendix I), the primer pair utilized for qPCR experiments was designed as follows (yellow highlights) in the nucleotide sequence of the available *hGH* gene:

1. The size of the region between the primers is 238 bp and was used in qPCR experiments

```
CACCATCACCATCACCATATTGAAGGGAGAattcccaactataaccaactatctcgtctattcgataacgctatgcttcg
tgctcatcgtcttcatcagctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaacagaagt
atcattctctgcagaacccccagacctccctctgtttctcagagtctattccgacacctccaacagggaggaaaca
caacagaaatccaacctagagctgctccgcactctccctgctgctcatccagtcgtggctggagcccgtgcagttcct
caggagtgtcttcgccaacagcctgggtgacggcgctctgacagcaacgtctatgacctcctaaggacctagagg
aaggcatccaaacgctgatgggaggctggaagatggcagccccggactgggcagatcttcaagcagacctacagc
aagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggctgctctactgcttcaggaaggacat
ggacaaggtcgagacattcctgcgcatcgtgcagtgccgctctgtggaggcagctgtggttcTAG
```

2. Extra primer pair designed in this study for further analysis; the size of the region between the primers is 202 bp. The reverse primer was utilized in pre-screening experiments along with *pUC-ori* forward primer (Table 3.18).

```
CACCATCACCATCACCATATTGAAGGGAGAattcccaactataaccaactatctcgtctattcgataacgctatgcttcg
tgctcatcgtcttcatcagctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaacagaagt
atcattctctgcagaacccccagacctccctctgtttctcagagtctattccgacacctccaacagggaggaaaca
caacagaaatccaacctagagctgctccgcactctccctgctgctcatccagtcgtggctggagcccgtgcagttcct
caggagtgtcttcgccaacagcctgggtgacggcgctctgacagcaacgtctatgacctcctaaggacctagagg
aaggcatccaaacgctgatgggaggctggaagatggcagccccggactgggcagatcttcaagcagacctacagc
aagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggctgctctactgcttcaggaaggacat
ggacaaggtcgagacattcctgcgcatcgtgcagtgccgctctgtggaggcagctgtggttcTAG
```

H.2. *ARG4* gene specific primers

Nucleotide sequence of the *ARG4* gene from (NCBI) and the location of the primers (green highlights) designed in present study as outer primers (*ARG4*-

Std-F and *ARG4*-Std-R) for preparation of *ARG4* standard in qPCR experiments. The size of the region between the primers is 330 bp.

>gi|254567652:c1280445-1279048 *Pichia pastoris* GS115
chromosome 1, complete sequence

```
ATGTCGAATCAAGAAGAAGGACTTAAACTGTGGGGTGGCAGGTTTACTGGGGCTACTGACCCCTTGATGGATTTGTA
TAACGCTTCCTTACCTTACGACAAGAAAATGTACAAGGTGGATTTAGAAGGAACAAAAGTTTACACTGAGGGCCTGG
AGAAAATTAATTTGCTAACTAAAGACGAACTAAGTGAGATTTCATCGTGGTCTCAAATTTGATTGAAGCAGAGTGGGCA
GAAGGGAAGTTTGTGAGAAGCCAGGGGATGAGGATATTCACACTGCTAATGAACGTCGCTTGGGTGAGTTGATTGG
TCGTGGAATCTCTGGTAAGGTTTCATACCGGAAGGCTAGAAAATGATCAAGTTGCCACTGATATGCGGTTGTATGTCA
GAGACAATCTAACTCAGTTGGCTGACTATCTGAAGCAGTTCATTCAAGTAATCATCAAGAGAGCTGAACAGGAAATA
GACGCTCTTGATGCCCGTTTATACTCACTTGGCAAAGAGCTCAACCAATCAGATGGTCTCACTGGTTGAGCATGTATGC
TACCTATTTCACTGAAGATTATGAGAGACTGAATCAAATCGTTAAAAGGTTGAACAAATCCCATTGGGAGCTGGAG
CTTTGGCTGGTCATCCTTATGGAATGATCGTGAATACATTGCTGAGAGATTAGGGTTTGATTCTGTATTGGTAAT
TCTTTGGCCGCTGTTTCAGACAGAGATTTTGTAGTCGAAACCATGTTCTGGTCTTCGTTGTTTATGAATCATATTTTC
TCGATTTCTCAGAAGATTTGATCATTACTCCACTGGAGAGTTTGGATTTATCAAGTTGGCAGATGCTTATTTCTACTG
GATCTTCTCTGATGCCTACAAAAAAAACCCAGACTCTTTGGAGTTATTGAGGGGTAATCTGGTAGATGTTTTGGG
GCCTTGGCTGGTTTCTCATGTCTATTAAGTCCATTCGGTCAACCTATAACAAAGATATGCAAGAGGATAAGGAGCC
TTTATTTGATACTCTAATCACTGTAGAGCACTCGATTTTGTAGCATCCGGTGTAGTTTCTACCTTGAACATTTGATC
CCGAACGATGAAGAATGCTCTAACTATGGATATGCTGGCTACAGATCTTGCCGACTATTTAGTTAGAAGGGGAGTT
CCATTGAGAGAACTCACCACATTTCTGGTGAATGTGTGAGACAAGCCGAGGAGTTGAACCTTTCTGGTATTGATCA
GTTGTCCCTCGAACAATGAAATCCATTGACTCCCGTTTTGAGGCTGATGTGGCTTCAACGTTTACTTTGAGCCCA
GTGTTGAAAAAAGAAGTCCACCAGGAGAACTTCTAAGACTGCTGTTTTAAAGCAATTGGATGCACCTGAATGAAAAA
CTAGAGCTTTGA
```

Nucleotide sequence of the *ARG4* gene from (NCBI) and the location of the primers used as inner primers in qPCR experiments (grey highlights); the size of the region between the primers is 84 bp.

>gi|254567652:c1280445-1279048 *Pichia pastoris* GS115
chromosome 1, complete sequence

```
ATGTCGAATCAAGAAGAAGGACTTAAACTGTGGGGTGGCAGGTTTACTGGGGCTACTGACCCCTTGATGGATTTGTA
TAACGCTTCCTTACCTTACGACAAGAAAATGTACAAGGTGGATTTAGAAGGAACAAAAGTTTACACTGAGGGCCTGG
AGAAAATTAATTTGCTAACTAAAGACGAACTAAGTGAGATTTCATCGTGGTCTCAAATTTGATTGAAGCAGAGTGGGCA
GAAGGGAAGTTTGTGAGAAGCCAGGGGATGAGGATATTCACACTGCTAATGAACGTCGCTTGGGTGAGTTGATTGG
TCGTGGAATCTCTGGTAAGGTTTCATACCGGAAGGCTAGAAAATGATCAAGTTGCCACTGATATGCGGTTGTATGTCA
GAGACAATCTAACTCAGTTGGCTGACTATCTGAAGCAGTTCATTCAAGTAATCATCAAGAGAGCTGAACAGGAAATA
GACGCTCTTGATGCCCGTTTATACTCACTTGGCAAAGAGCTCAACCAATCAGATGGTCTCACTGGTTGAGCATGTATGC
TACCTATTTCACTGAAGATTATGAGAGACTGAATCAAATCGTTAAAAGGTTGAACAAATCCCATTGGGAGCTGGAG
CTTTGGCTGGTCATCCTTATGGAATGATCGTGAATACATTGCTGAGAGATTAGGGTTTGATTCTGTATTGGTAAT
TCTTTGGCCGCTGTTTCAGACAGAGATTTTGTAGTCGAAACCATGTTCTGGTCTTCGTTGTTTATGAATCATATTTTC
TCGATTTCTCAGAAGATTTGATCATTACTCCACTGGAGAGTTTGGATTTATCAAGTTGGCAGATGCTTATTTCTACTG
GATCTTCTCTGATGCCTACAAAAAAAACCCAGACTCTTTGGAGTTATTGAGGGGTAATCTGGTAGATGTTTTGGG
GCCTTGGCTGGTTTCTCATGTCTATTAAGTCCATTCGGTCAACCTATAACAAAGATATGCAAGAGGATAAGGAGCC
TTTATTTGATACTCTAATCACTGTAGAGCACTCGATTTTGTAGCATCCGGTGTAGTTTCTACCTTGAACATTTGATC
CCGAACGATGAAGAATGCTCTAACTATGGATATGCTGGCTACAGATCTTGCCGACTATTTAGTTAGAAGGGGAGTT
CCATTGAGAGAACTCACCACATTTCTGGTGAATGTGTGAGACAAGCCGAGGAGTTGAACCTTTCTGGTATTGATCA
GTTGTCCCTCGAACAATGAAATCCATTGACTCCCGTTTTGAGGCTGATGTGGCTTCAACGTTTACTTTGAGCCCA
GTGTTGAAAAAAGAAGTCCACCAGGAGAACTTCTAAGACTGCTGTTTTAAAGCAATTGGATGCACCTGAATGAAAAA
CTAGAGCTTTGA
```

I. Nucleotide sequence of the pGAPZαA::hGH.

The nucleotide sequence of the developed base plasmid (BP) has been presented. Light green highlight is the P_{GAP}. Red highlights are GAP forward and AOX reverse primer sites. Blue highlights are the cleavage sites of Bsp119I and NsiI and “...” represent cleavage sites of EcoRI and XbaI restriction enzymes. Yellow highlighted sequence is the α-MF secretion signal peptide. His-tag and factor Xa cleavage sequence reside before grey highlighted hGH sequence as brown and dark green highlights, respectively:

```
AGATCTTTTTGTAGAAATGCTCTTGGTGTCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCA
ACGCCGAACGACCTGCTGGCAACGTAAAAATTCTCCGGGTAAAACTTAAATGTGGAGTAATGGAACAGAAACGTC
TCTTCCCTTCTCTCCTTCCACCGCCCGTTACCGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTACCGG
CCCCCTTGCAGCAATGCTCTTCCAGCATTACGTTGCCGGTAAAAACGGAGGTCGTGTACCCGACCTAGCAGCCAG
GGATGGAAGTCCCGCCGTCGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAACCACCAGA
ATCGAATATAAAAGGCCAACCCTTCCCAATTTGGTTTCTCTGACCCAAAGACTTTAAATTTAATTTATTTGT
CCCTATTCAATCAATTGAACTCACTATTTGGAAACGATGAGATTTCCTTCAATTTTACTGCTGTTTTATTTCGCAG
CATCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAATTCGGCTGAAGCTGTCAAT
CGGTTACTCAGATTAGAAAGGGGATTTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAAATACGGGTTATTG
TTTATAAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTG
...AATTCACCATCACCATCACCATATTGAAGGGAGAttcccaactataaccactatctcgtcttattcgataacgcta
tgcttcgtctcatcgtcttcatcagctggcctttgacacctaccaggagtttgaagaagcctatatcccagaagga
acagaagtattcattctcgcagaacccccagacctccctctgtttctcagagtctattccgacacctccaacagga
gaggaacaacacagaaatccaaacctagagctgctccgcaatctccctgctgctcatccagctcgtggctggagcccg
tgagttctcagagagtgtcttcgccaacagcctgggtgacggcgctctgacagcaacgtctatgacctcctaaa
ggacctagaggaagcattccaaacgctgatggggaggctggaagatggcagccccggactgggcagatcttcaag
cagacctacagcaagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggctgctctactgct
tcaggaagagacatggacaaggtcgagacattcctgcgcatcgtgcatgcccctctggtggagggcagctggtgctt
cTAGT...CTAGAACAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATATTAG
TTTCTAGCTTAGACATGACTGTTCCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGCTCTGTGATATTCTAAT
CAAGAGGATGTCAGAATGCCATTTGCTGAGAGATGCAGGCTTCATTTTGTACTTTTTTATTGTAACCTATATAGTATAG
GATTTTTTTTGTGATTTTGTCTTCTCTGACGAGCTTGTCTCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGTGGTAGGGT
TTGGGAAATCATTGAGTTGATGTTTTCTTGGTATTTCCACTCCTCTTCCAGTACAGAAGATTAAGTGAACCTTCGTTGTG
GCGGATCCCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTAC
CACTTCAAAACACCAAGCACAGCATAAAATTTCCCTCTTCTCTCTAGGGTGTGTTAATACCCGTAATAAGGTTTGGAA
AAAGAAAAAGAGACCGCTCGTTTCTTCTTCTCGTCAAAAAGGCAATAAAAAATTTTATCAGTTCCTTTTCTTGAATTTTT
TTTTTTAGTTTTTCTCTTTTCAAGTACCTCCATTGATATTTAAGTAAATAAACGGTCTTCAATTTCTCAAGTTTCAAGTTTCAATTT
TCTTGTCTATTACAATTTTTTACTTCTGTTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGGGCGGTGTTGACAATTAATC
ATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGAACTAAACCATGGCCAAGTTGACAGTGCCTGCTCAC
CGCGCGCAGCTGCGCGGAGCGGTGAGTTCGGACCGACCGCTCGGGTTCCTCCGGGACTTCGTGGAGGACGACTTCGCGCGGTG
GGTCCGGGACGACGTGACCTGTTTCATCAGCGCGTCCAGGACCGAGGTGGTGGCGGACAAACCCCTGGCTGGGTGTTGGTGGCGG
CTTGACGAGCTGTACGCGAGTGGTTCGGAGGTCGTGTCCAGAACTTCCGGGACCGCTCCGGGCCGGCCATGACCGGATCGGGCA
GCAGCGTGGGGCGGGAGTTCGCCCTGCGCGACCGCGCGGCAACTGCGTGCATCTCGTGGCGGAGGACGAGCTGACACGTCGG
ACGGCGGCCACGGTCCAGGCTCGGAGATCCGTCGCCCTTTTCTTGTGATATCATGTAATAGTTATGTCACGCTTACATT
CAGCGCTCCCCACATCCGCTTAACGAAAAGGAAGGAGTTAGACAACTGAAGTCTAGTCCATATTTATTTTTTATAGTTA
TGTTAGTATTAAGAAGCTTATTTATATTCAAAATTTTCTTTTTTCTGTACAGACCGGTGTACGATGTAACATATACTGAAAA
CCTTGCTTGAGAAGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCTGGAGACCAACATGTGAGCAAAAGGCCAGAAAAAGCCA
GGAACCGTAAAAAGCCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGA
GGTGGCAAAACCCGACGAGCTATAAAGATACAGGCGTTTCCCTTGGAAAGTCCCTCGTGGCTCTCCGTGTCGACCTGCCG
TTACCGGATACCTGTCGCTTCTCCCTCGGGAAGCGTGGCGCTTCTCAATGCTCAGCTGTAGGTATCTCAGTTCGGGTGAGG
TCGTTGCTCCAAAGCTGGGCTGTGTGCACGAACCCCGTTACGCCGACCGCTGCGCTTATCCGGTAACATPCGCTTGGTCCA
ACCCGGTAAGACAGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGAGGCGGTGCTACAGAT
TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTGATATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAA
GAGTTGGTAGCTTGTATCCGGCAACAAACCCGCTGGTAGCGGTGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCTTTGATCTTTTCTACGGGCTGACGCTCAGTGGAAACGAAACTCAGTTAAGGATTTTTGGTCATG
ATGAGATC
```

Appendix J. The standard curves used during experiments.

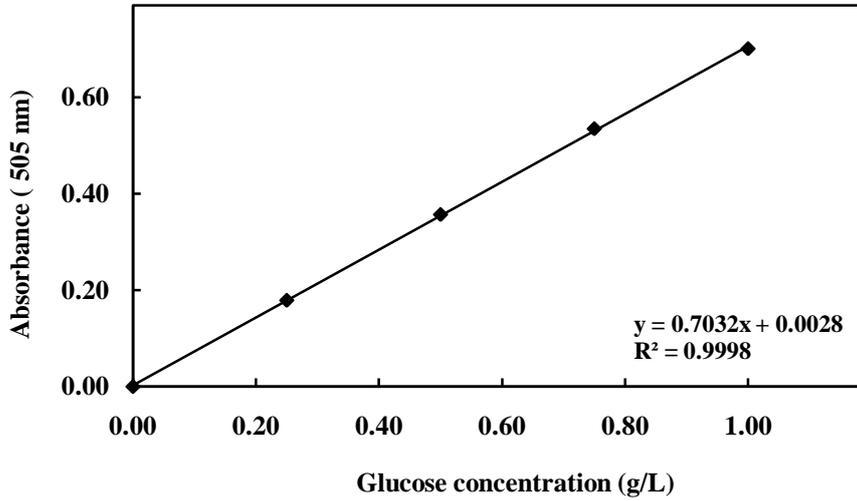


Figure J.1. Glucose standard curve used in residual glucose concentration measurement in laboratory-scale bioreactor experiments.

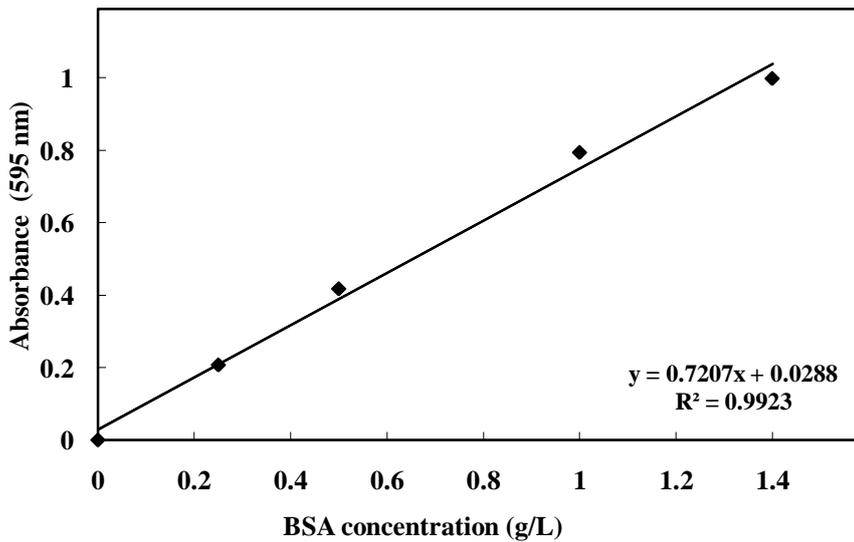


Figure J.2. Bovine serum albumin (BSA) standard curve used in total protein measurement.

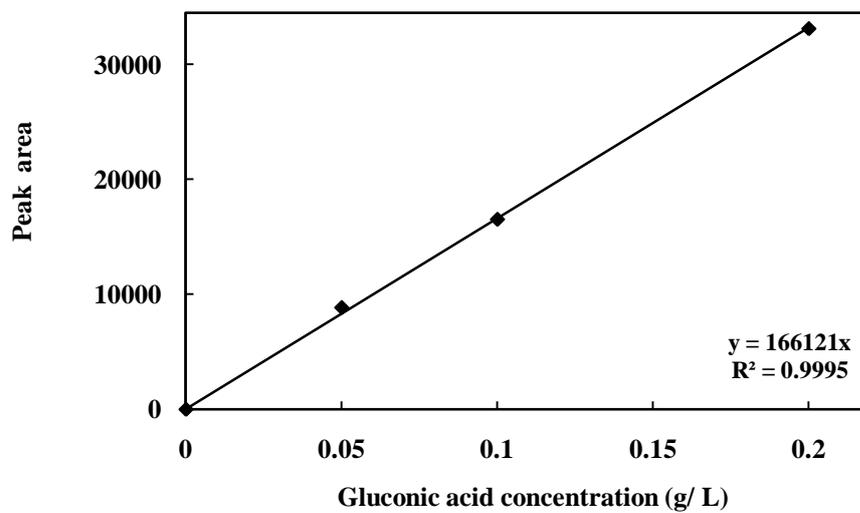


Figure J.3. Gluconic acid standard curve used in HPLC analysis.

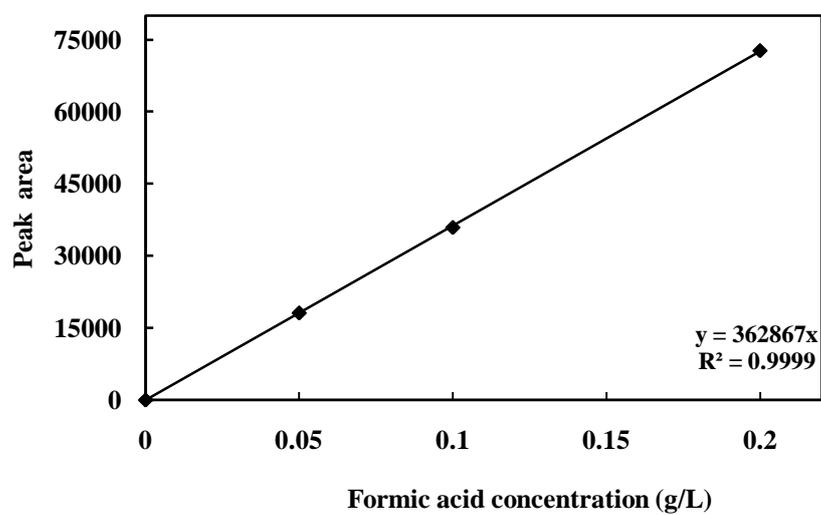


Figure J.4. Formic acid standard curve used in HPLC analysis.

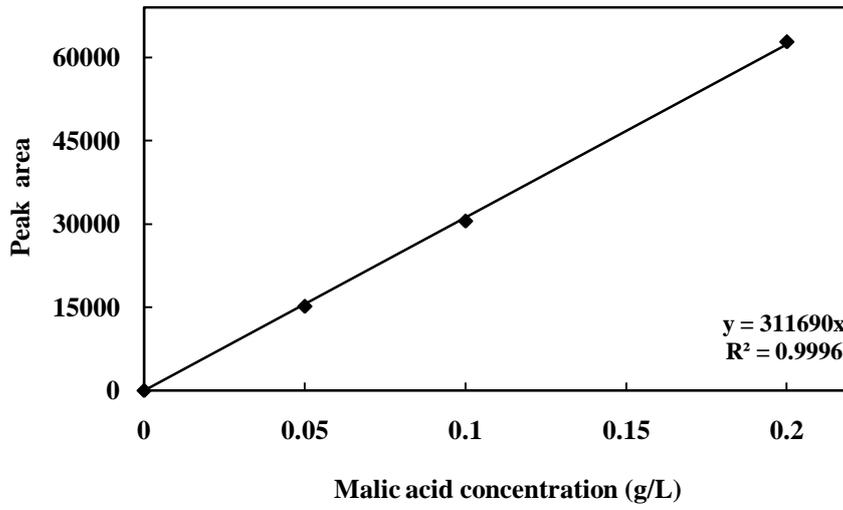


Figure J.5. Malic acid standard curve used in HPLC analysis.

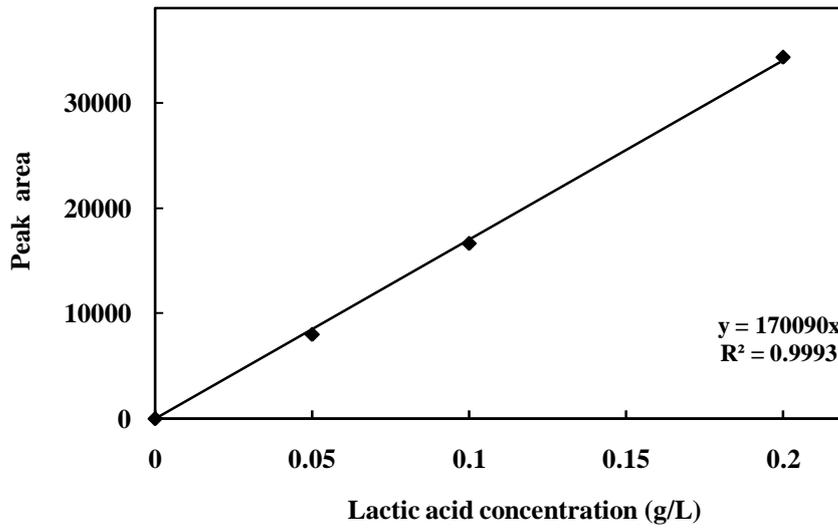


Figure J.6. Lactic acid standard curve used in HPLC analysis.

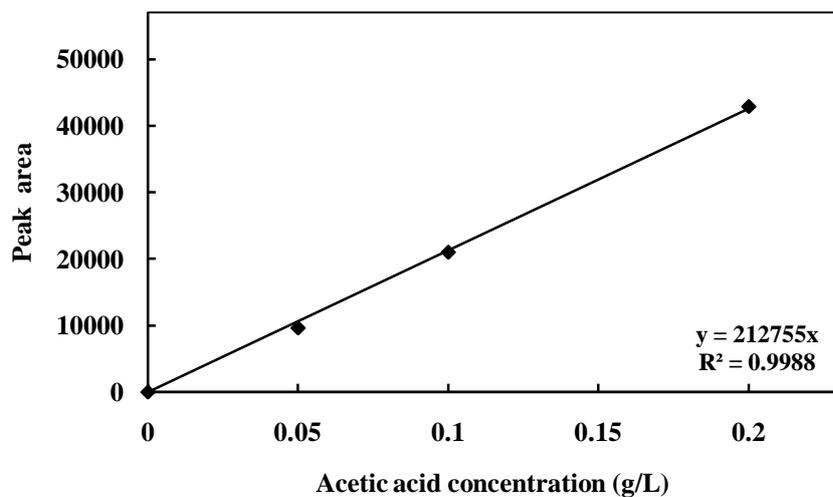


Figure J.7. Acetic acid standard curve used in HPLC analysis.

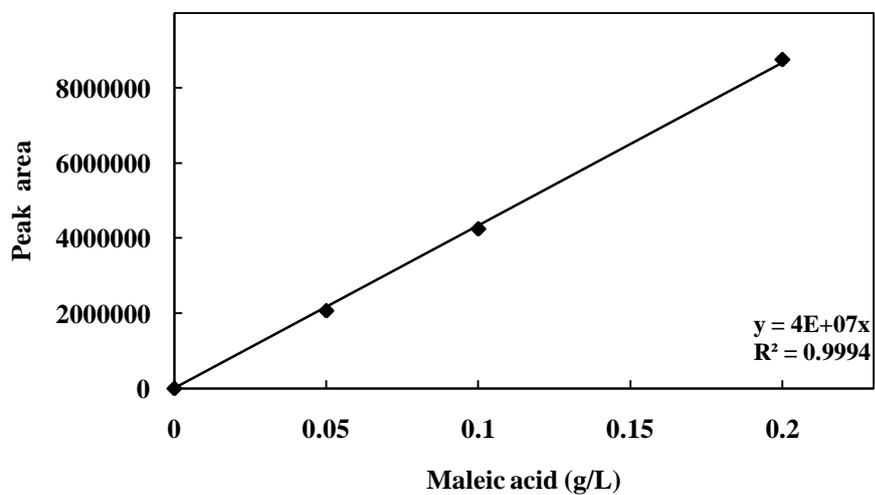


Figure J.8. Maleic acid standard curve used in HPLC analysis.

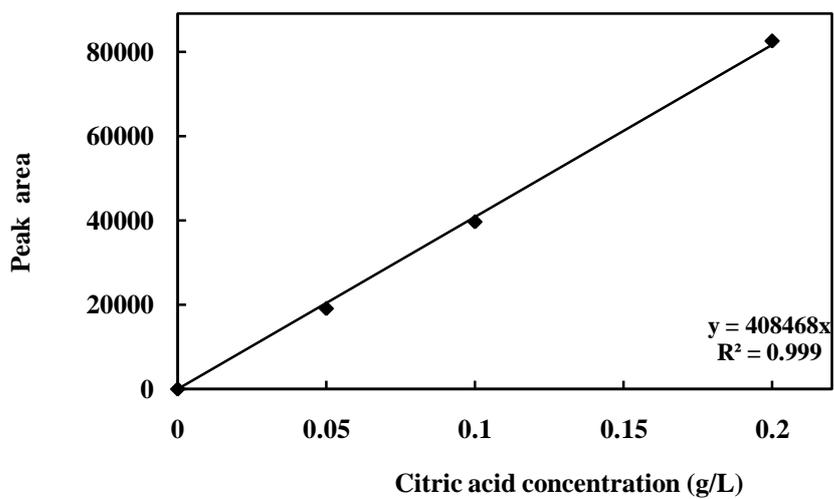


Figure J.9. Citric acid standard curve used in HPLC analysis.

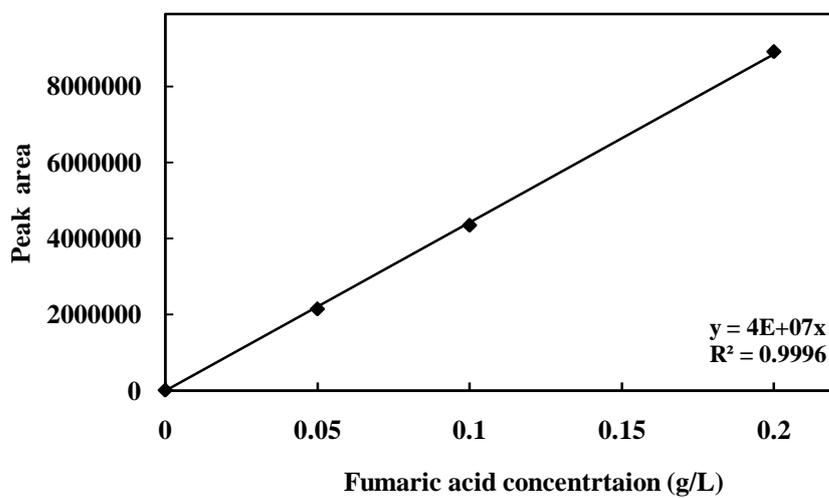


Figure J.10. Fumaric acid standard curve used in HPLC analysis.

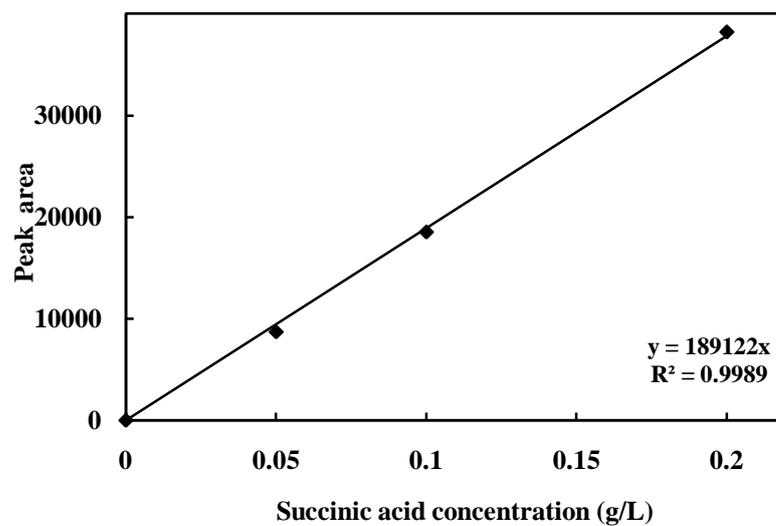


Figure J.11. Succinic acid standard curve used in HPLC analysis.

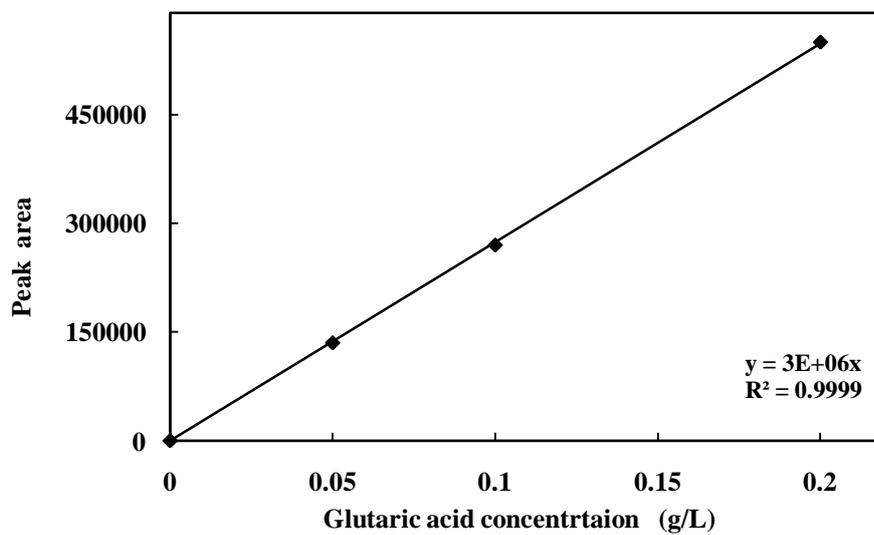


Figure J.12. Glutaric acid standard curve used in HPLC analysis.

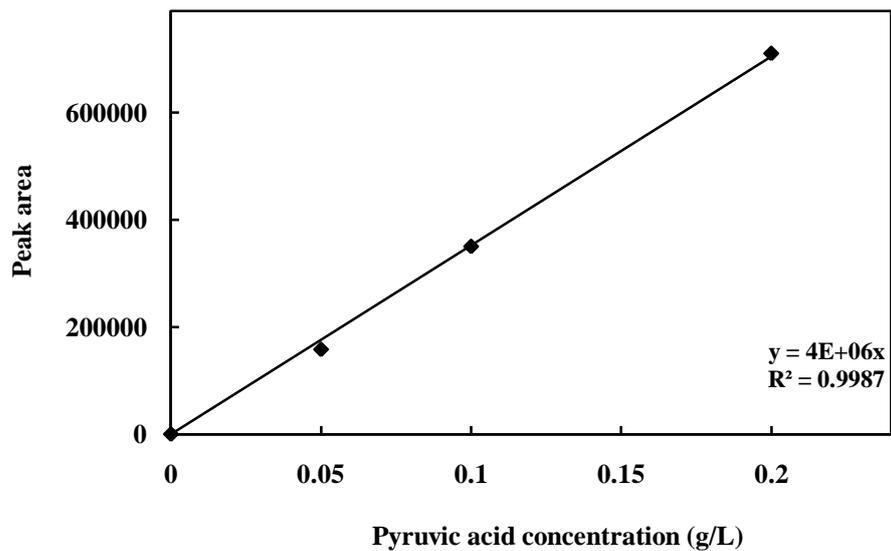


Figure J.13. Pyruvic acid standard curve used in HPLC analysis.

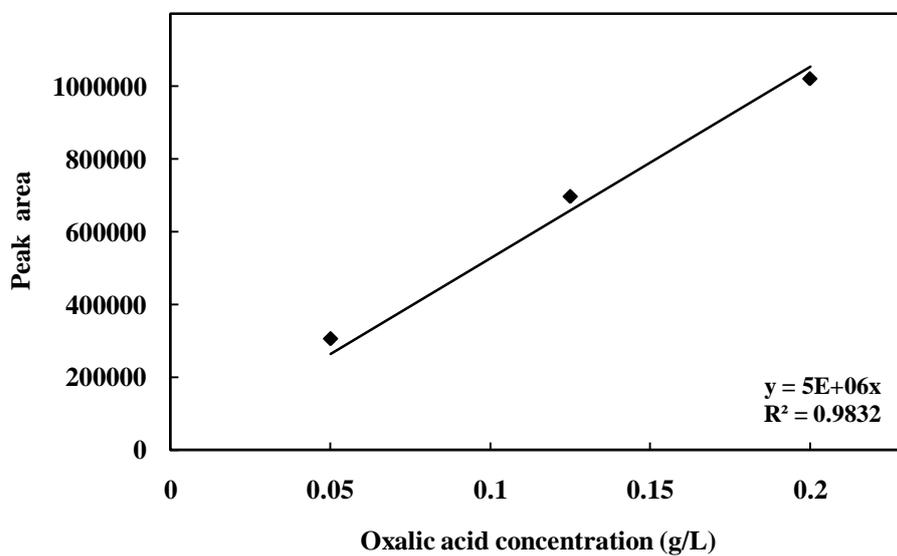


Figure J.14. Oxalic acid standard curve used in HPLC analysis.

K. Verification of the sequences of typical developed plasmids by BLAST analysis.

K.1. pGAPZaA:*hGH* (Base plasmid)

Forward sequencing with *GAP* forward primer; the location of the primer can be found in Appendix I. The term “query” has been attributed to the expected sequence.

```
Query 28  ATTCGAAACGATGAGATTCCTTCAATTTTACTGCTGTTTTATTTCGCAGCATCCTCCG 87
Sbjct 7   ATTCG-AACGATGA-ATTCCTTC-ATTTTT-CTGCTGTTTTATTTCGCAGCATCCTCCG 62

Query 88  CATTAGCTGCTCCAGTCAACTACAACAGAAGATGAAACGGCACAATTCGGGCTGAAG 147
Sbjct 63  CATTAGCTGCTCCAGTCAACTACAACAGAAGATGAAACGGCACAATTCGGGCTGAAG 122

Query 148  CTGTCATCGGTTACTCAGATTTAGAAGGGGATTCGATGTTGCTGTTTTGCCATTTCCA 207
Sbjct 123  CTGTCATCGGTTACTCAGATTTAGAAGGGGATTCGATGTTGCTGTTTTGCCATTTCCA 182

Query 208  ACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAG 267
Sbjct 183  ACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAG 242

Query 268  AAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCACCATCACCATCACC 327
Sbjct 243  AAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCACCATCACCATCACC 302

Query 328  ATATTGAAGGGAGATTCCTCACTATAACCACTATCTCGICTATTGATAACGCTATGCTTC 387
Sbjct 303  ATATTGAAGGGAGATTCCTCACTATAACCACTATCTCGICTATTGATAACGCTATGCTTC 362

Query 388  GTGCTCATCGTCTTCATCAGCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTATA 447
Sbjct 363  GTGCTCATCGTCTTCATCAGCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTATA 422

Query 448  TCCCAAAGGAACAGAAGTATTCATTCTGCAGAACCCCGAGACCTCCCTCTGTTTCTCAG 507
Sbjct 423  TCCCAAAGGAACAGAAGTATTCATTCTGCAGAACCCCGAGACCTCCCTCTGTTTCTCAG 482

Query 508  AGTCTATTCCGACACCCCTCCAACAGGGAGGAAACACAACAGAAATCCAACCTAGAGCTGC 567
Sbjct 483  AGTCTATTCCGACACCCCTCCAACAGGGAGGAAACACAACAGAAATCCAACCTAGAGCTGC 542

Query 568  TCCGCATCTCCCTGCTGCTCAICCACTCGTGGCTGGAGCCCGTGCAATTCCCTCAGGAGTG 627
Sbjct 543  TCCGCATCTCCCTGCTGCTCAICCACTCGTGGCTGGAGCCCGTGCAATTCCCTCAGGAGTG 602

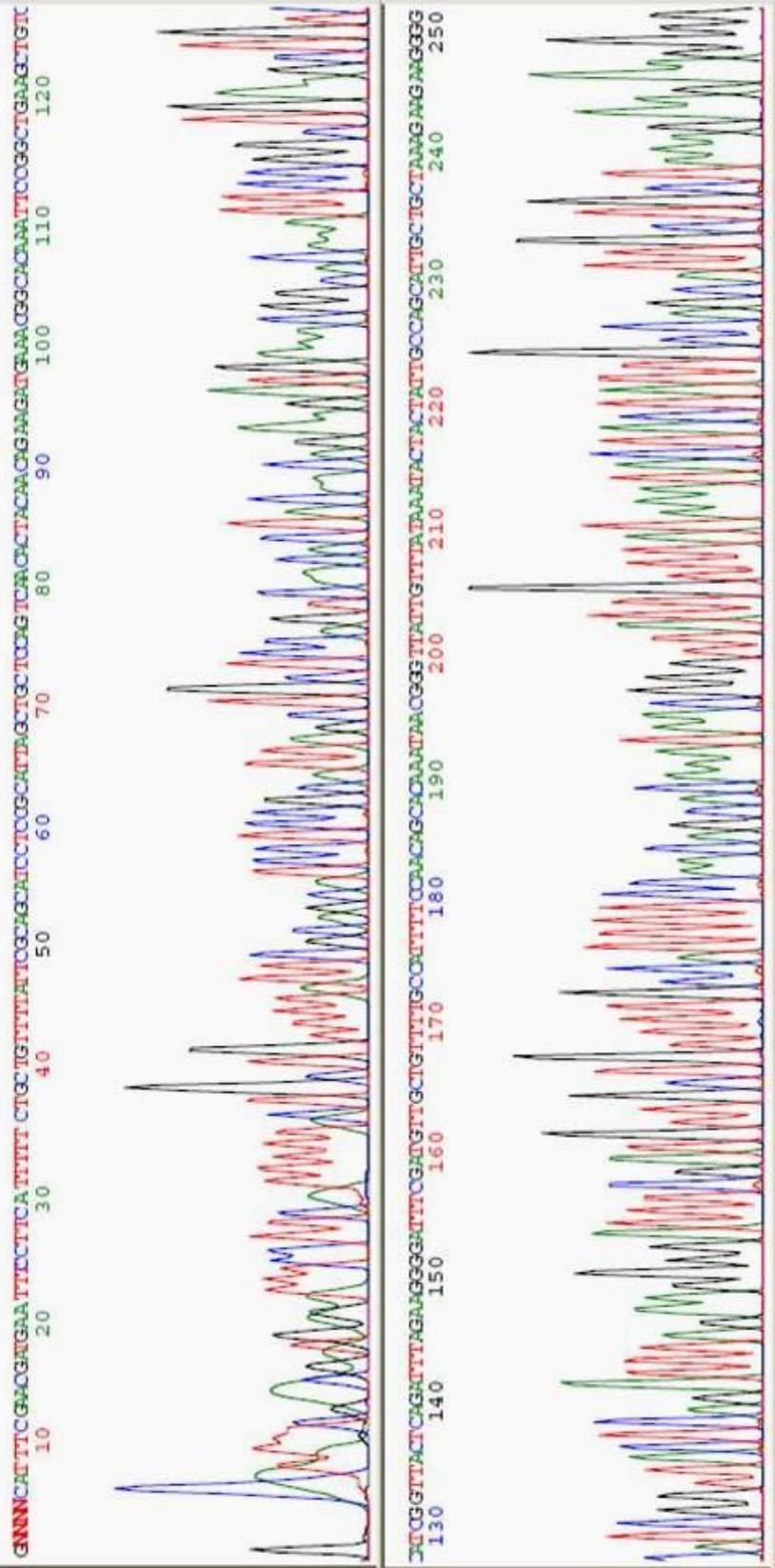
Query 628  TCTTCGCCAACAGCCTAGTGTACGGCGCCTCTGACAGCAACGTCTATGACCTCCTAAAGG 687
Sbjct 603  TCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGACCTCCTAAAGG 662

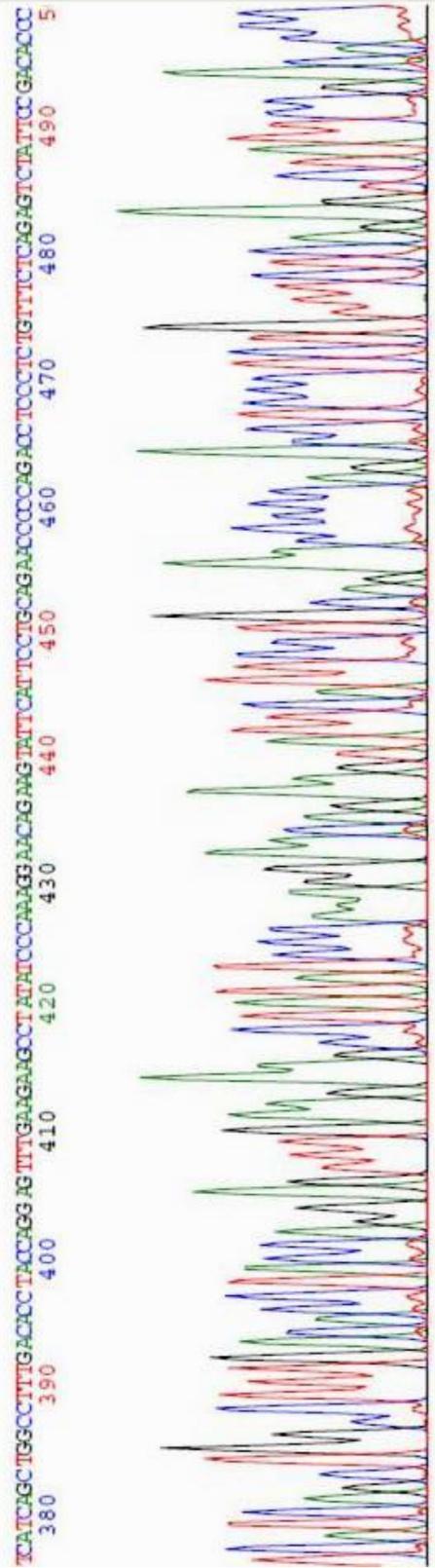
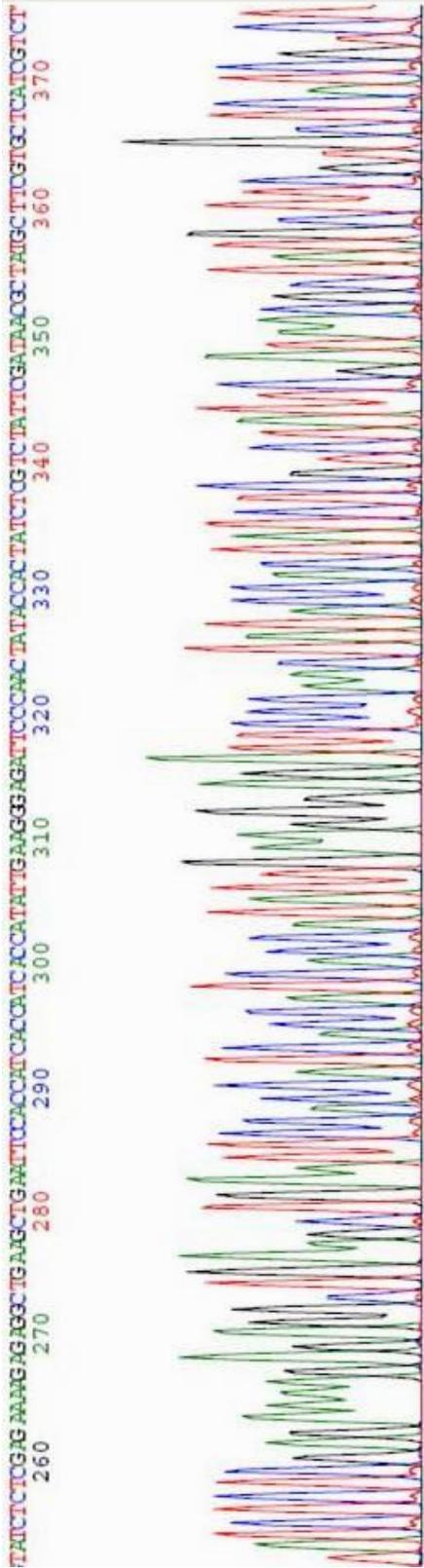
Query 688  ACCTAGAGGAAGGCATCCAACGCTGATGGGAGGCTGGAAGATGGCAGCCCGGACTG 747
Sbjct 663  ACCTAGAGGAAGGCATCCAACGCTGATGGGAGGCTGGAAGATGGCAGCCCGGACTG 722

Query 748  GGCAGATCTTCAAGCAGACCTACAGCAAGTTTCGACACAACTCACACAACG-ATGACGCA 806
Sbjct 723  GGCAGATCTNC-AGCAGACCTACAGCAAGTTTCGACACAACTCACACNNA-CGNATGACGCA 780

Query 807  CTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGGTCGAGA 864
Sbjct 781  CTACTCCAGAACTACGGGCTGCTCTACTGCTTC-GGNANGANATGGA-AAGGTCGAGA 836
```

NCBI BLAST: 820/838 (98% identical), gaps: 9/838 (1%). The gaps were checked with chromatograms obtained by sequencing.





K.2. pPDCZαA::hGH

Two specific forward primers were utilized separately specified in dark blue highlights (*PDC-F* and *PDC End-fwd* in Appendix D) along with specific reverse primer (*PDC-R*) and *AOX* reverse primer (red highlights) for sequencing.

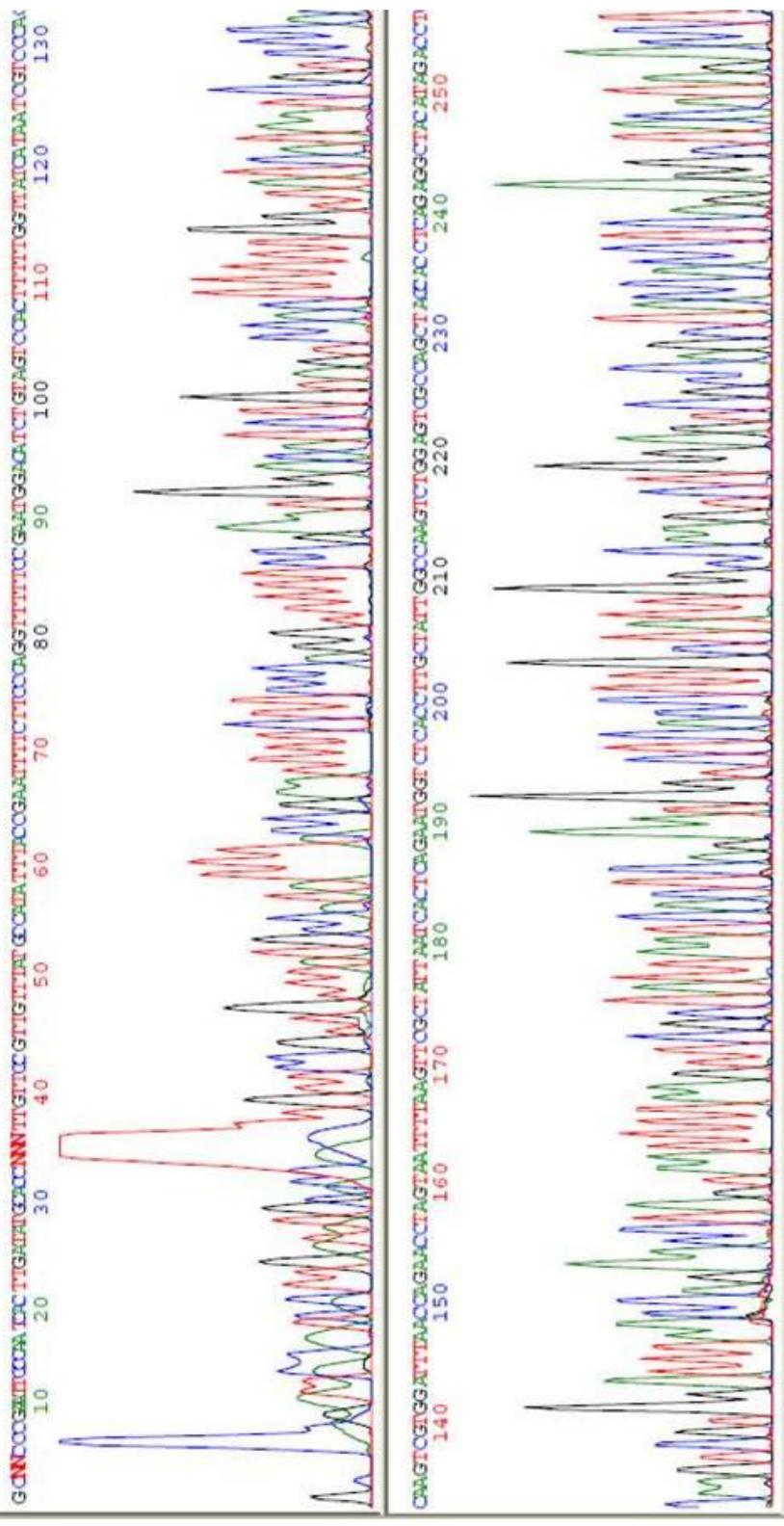
```
GGGACAAGCACAGATTACCCAATCACTTGATATGCACCAATTTGTTCCGTTGTTTATGCCATATTTACCGAATTT
TCTTCCCAGGTTTTTCCGAATGGACATCTGTAGTCCACTTTTTGGTTATCATAATCGTCCCACAAAGTCGTGGATTT
AACCAGAACCTAGTAATTTAAGTTCGCTATTAATCACTCAGAATGGTCTCACCTTGCATATGGCCAAGTCTGGAG
TCGCCAGCTACCACCTCAGAGGCTACATAGACCTCCCAATGTCACTCCTCAGTGGCTCTCAATCTCGTGTCTT
TTCCGTTAAAACCTCCGTTTCACCCCTATACTGCCCTGGTTGTGCAGCTCTTACCACCTTCGGCCCGTACTAT
CCGTAGTGGTCGAGCCGCATCAATATCAGTGTGAAATAGAATAACTCCCTACAAAAGCCGCACGCAACCATCAAA
CTATATAAGGAACCTCAATATCTAGCAACATCTTTTCAATTTACTACAACATATTCGTTAATCATCAATCAATTA
GCTAGTACACAACATGAGATTTCCCTCAATTTTTACTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGTGTCTCC
AGTCAACACTACAACAGAAGATGAAACGGCACAAATTCGGCTGAAGCTGTATCGGTTACTCAGATTTAGAAGGG
GATTTTCATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTATTGTTTATAAATACTACTATTGCCA
GCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGA AAAGAGAGGCTGAAGCTGAATTC CACCATCACCATCACC
TATTGAAGGGAGA tcccaactataccaactatctcgtctattcgataacgctatgcttcggtcctcctcctc
cagctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaacagaagtattcattcctgcaga
acccccagacctccctctgtttctcagagtctattccgacacctccaacaggaggaaacacacagaaatccaa
cctagagctgctccgcatctccctgctgctcatccagctgctggctggagcccctgcagttcctcaggagtgcttc
gcaaacagctggtgtacggcgctctgacagcaacgtctatgacctcctaaaggacctagaggaaggcatccaaa
cgtgatggggaggctggaagatggcagccccggactgggcagatcttcaagcagacctacagcaagttcgacac
aaactcacacaacgatgacgcactactcaagaactacgggctgctctactgcttcaggaaggacatggacaaggctc
gagacattcctgcgcatcgtgcagtgccgctctgtggagggcagctgtggcttc TAGTCTAGAACAAAACTCAT
CTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGATTTCTAGCCTTAGACATGACTGT
TCCTCAGTTCAAGTTGGGCACCTTACGAGAAGACCGGCTTGTGCTAGATTCTAATCAAGA GGATGTCAGAATGCCATT
TGC
```

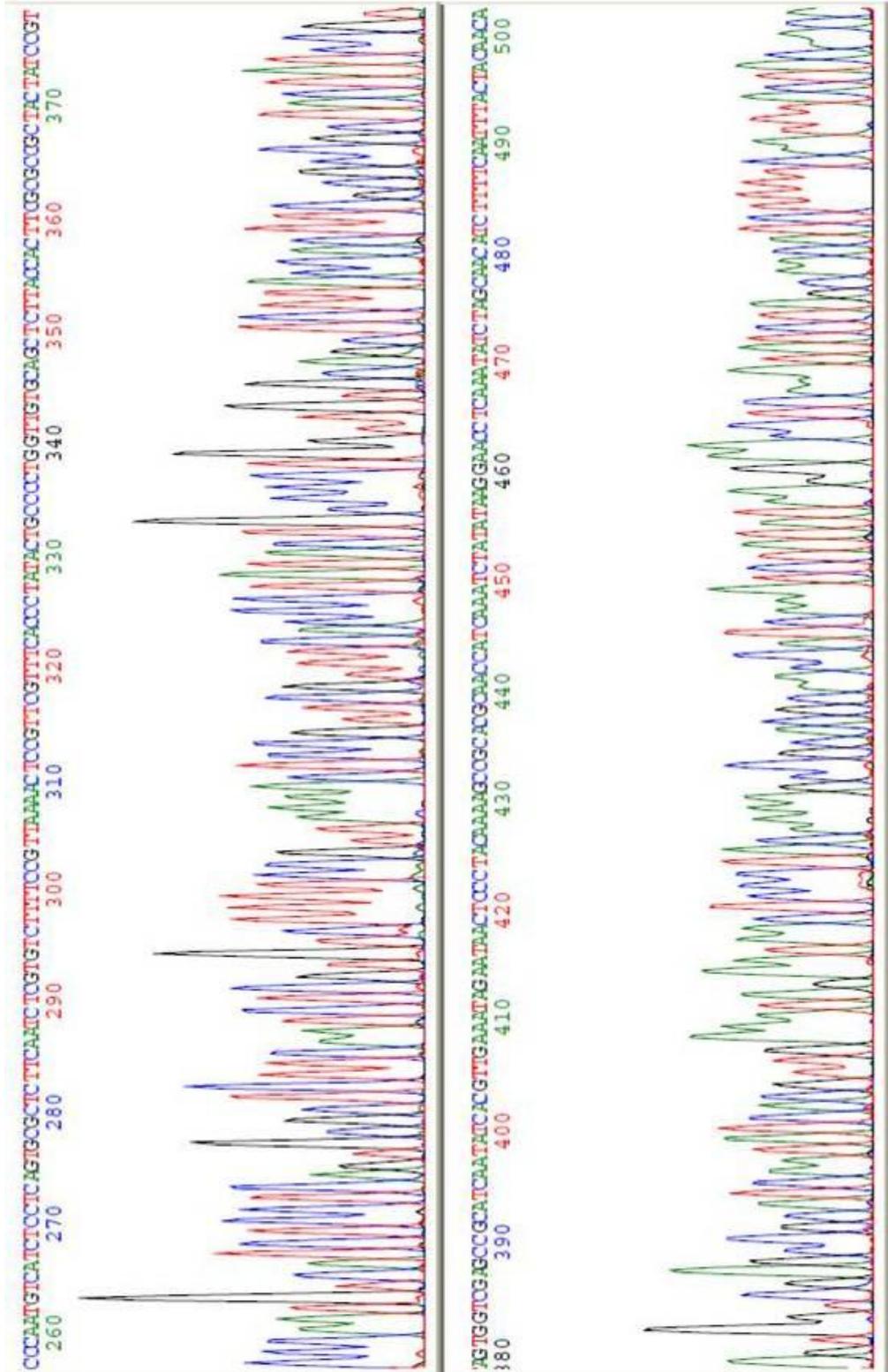
Forward sequencing with normal forward primer of PDC starts at the beginning of the promoter:

Range 1: 10 to 761 Graphics		▼ Next Match ▲ Previous Match		
Score	Expect	Identities	Gaps	Strand
1267 bits(686)	0.0	732/753(97%)	16/753(2%)	Plus/Plus
Query 27	ATTACCCAATCACTTGATATGCACCAATTGTTCCGTTGTTTATGCCATATTTACCGAAT	86		
Sbjct 10	ATT-CCCAATCACTTGATATGCACCNNNTTGTCCGTTGTTTATGCCATATTTACCGAAT	68		
Query 87	TTTCTTCCCAGTTTTTCCGAATGGACATCTGTAGTCCACTTTTTGGTTATCATAATCGT	146		
Sbjct 69	TTTCTTCCCAGTTTTTCCGAATGGACATCTGTAGTCCACTTTTTGGTTATCATAATCGT	128		
Query 147	CCCACAAGTCGTGGATTTAACCAGAACCTAGTAATTTAAGTTCGCTATTAATCACTCAG	206		
Sbjct 129	CCCACAAGTCGTGGATTTAACCAGAACCTAGTAATTTAAGTTCGCTATTAATCACTCAG	188		
Query 207	AAATGGTCTCACCTTGCTATTGGCCAAGTCTGGAGTCGCCAGCTACCACCTCAGAGGCTAC	266		
Sbjct 189	AAATGGTCTCACCTTGCTATTGGCCAAGTCTGGAGTCGCCAGCTACCACCTCAGAGGCTAC	248		
Query 267	ATAGACCTCCCAATGTCATCTCCTCAGTGCCTCTTCAATCTCGTGTCTTTTCCGTTAAA	326		
Sbjct 249	ATAGACCTCCCAATGTCATCTCCTCAGTGCCTCTTCAATCTCGTGTCTTTTCCGTTAAA	308		
Query 327	ACTCCGTTCCGTTTACCCTATACTGCCCTGGTTGTGCAGCTCTTACCACCTCGCGCCGC	386		
Sbjct 309	ACTCCGTTCCGTTTACCCTATACTGCCCTGGTTGTGCAGCTCTTACCACCTCGCGCCGC	368		
Query 387	TACTATCCGTAGTGGTCGAGCCGCATCAATATCACGTTGAAATAGAATAACTCCCTACAA	446		
Sbjct 369	TACTATCCGTAGTGGTCGAGCCGCATCAATATCACGTTGAAATAGAATAACTCCCTACAA	428		
Query 447	AAGCCGCACGCAACCATCAAATCTATATAAGGAACCTCAAATATCTAGCAACATCTTTTC	506		
Sbjct 429	AAGCCGCACGCAACCATCAAATCTATATAAGGAACCTCAAATATCTAGCAACATCTTTTC	488		
Query 507	AATTTACTACAACATATTCGTTAATCATCAATCAATTAGCTAGTACACAACAATGAGATT	566		
Sbjct 489	AATTTACTACAACATATTCGTTAATCATCAATCAATTAGCTAGTACACAACAATGAGATT	548		
Query 567	TCCTTCAATTTTACTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAA	626		
Sbjct 549	TCCTTCAATTTTACTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAA	608		
Query 627	CACTACAACAGAAG-ATGAAACGGCACAAATCCGGCTGAAGCTGTCATCGGTTACTCAG	685		
Sbjct 609	CACTACAACAGAAGAATGAAACGGCACAAATCCGGCTGAAGCTGTCATCGGTTACTCAN	668		
Query 686	-AATTAGAAGGGGA-TTTC-GATGTTG-CTGTTT-GCC-ATTTT-CCAACAGCACAAAT	738		
Sbjct 669	AATTTAGAAGGGGAATTTCCGATGTTGGCTGTTTTGCCCATTTNCCAACAGCACAAA	728		
Query 739	-AA-CGGG-TT-ATTG-TTTATAAA-TAC-TAC	764		
Sbjct 729	NAAACGGGGTTTATTGGTTTATAAAATACCTAC	761		



Model 310 73_F1-29-15-9-00 PM.ab1 Signal G:187 A:157 T:168 C:154 Page 2 of 3
Version 3.7 DT310POP6(BDV3)V1.mob Paz, Oub 01, 2015 5:17 PM
Basecaller-310PO73_F BIG DYE V3.1.mtx Per, Oca 29, 2015 9:00 PM
BC 1.3.0.0 Cap 7 Points 1452 to 11480 Pk 1 Loc: 1452 Spacing: 13.46{13.46





Forward sequencing with “new end forward” primer of PDC starts approximately at the end of the promoter before α -MF in order to confirm the proper consolidation of promoter and SP:

Range 1: 17 to 787 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1267 bits(686)	0.0	750/786(95%)	15/786(1%)	Plus/Plus
Query 42	TATTCGTTAATCATCAATCAATTAGCTAGTACACAACAAATGAGATTTCCCTTCAATTTT			101
Sbjct 17	TATTCGTT-AT-NTCCATCCATTAGCTAGTACACAACAAATGAGATTTCCCTTCAATTTT			74
Query 102	CTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAG			161
Sbjct 75	CTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAG			134
Query 162	ATGAAACGGCACAAATTCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATT			221
Sbjct 135	ATGAAACGGCACAAATTCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATT			194
Query 222	TCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATA			281
Sbjct 195	TCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATA			254
Query 282	CTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTG			341
Sbjct 255	CTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTG			314
Query 342	AAGCTGAATTCACCATCACCATCACCATATTGAAGGGAGATTTCCCAACTATACCACTAT			401
Sbjct 315	AAGCTGAATTCACCATCACCATCACCATATTGAAGGGAGATTTCCCAACTATACCACTAT			374
Query 402	CTCGTCATTTCGATAACGCTATGCTTCGTCATCGCTTTCATCAGCTGGCCTTTGACA			461
Sbjct 375	CTCGTCATTTCGATAACGCTATGCTTCGTCATCGCTTTCATCAGCTGGCCTTTGACA			434
Query 462	CCTACCAGGAGTTTGAAGAAGCCTATATCCCAAAGGAACAGAAGTATTCAITTCCTGCAGA			521
Sbjct 435	CCTACCAGGAGTTTGAAGAAGCCTATATCCCAAAGGAACAGAAGTATTCAITTCCTGCAGA			494
Query 522	ACCCCCAGACCTCCCTCTGTTTTCTCAGAGTCTATTCCGACACCCCTCCAACAGGGAGGAAA			581
Sbjct 495	ACCCCCAGACCTCCCTCTGTTTTCTCAGAGTCTATTCCGACACCCCTCCAACAGGGAGGAAA			554
Query 582	CACAACAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGC			641
Sbjct 555	CACAACAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGC			614
Query 642	TGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCCAACAGCCCTGGTGTACGGCGCCTCTG			701
Sbjct 615	TGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCC-ACAGCCTGGTGNACGGCGCCTCTG			673
Query 702	ACAGCAACGTCTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGA			761
Sbjct 674	ACNN-AACGTCTATGACCTCCTAANGGACNTA-AGGAAG-CATCCAN-CGCTTATGGGGA			729
Query 762	GGCTGGAAGATGGCAGCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCCG			821
Sbjct 730	NGCTGGAN-ANGGCANCCCCNG-ANTGGNN-ANCTTCA-GCAAAC-TAC-GCAGGTT-G			782
Query 822	ACACAA 827			
Sbjct 783	AC-CAA 787			

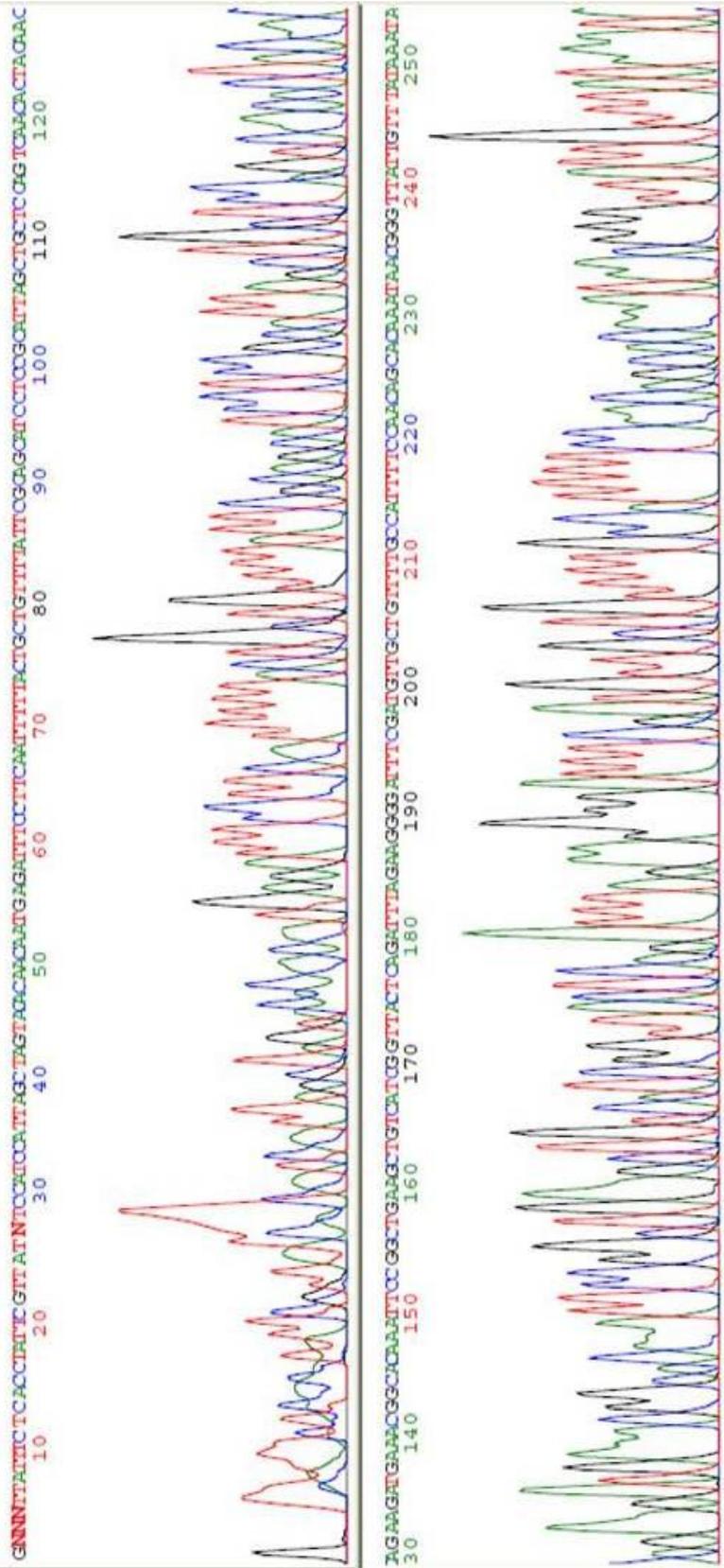
Model 310
Version 3.7
Basecaller-310POI73
BC 1.3.0.0

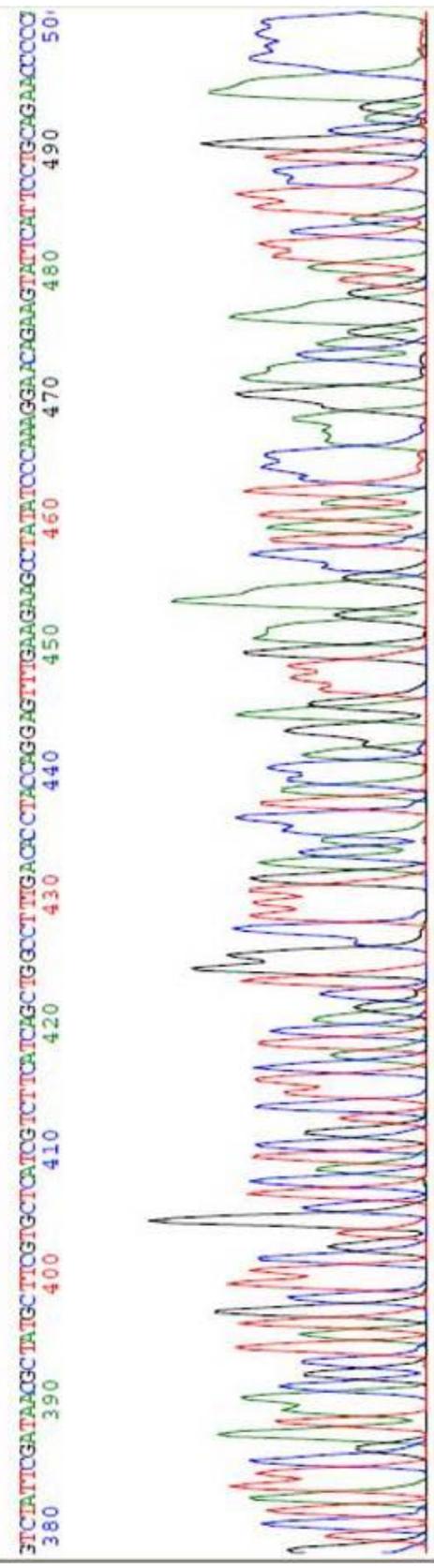
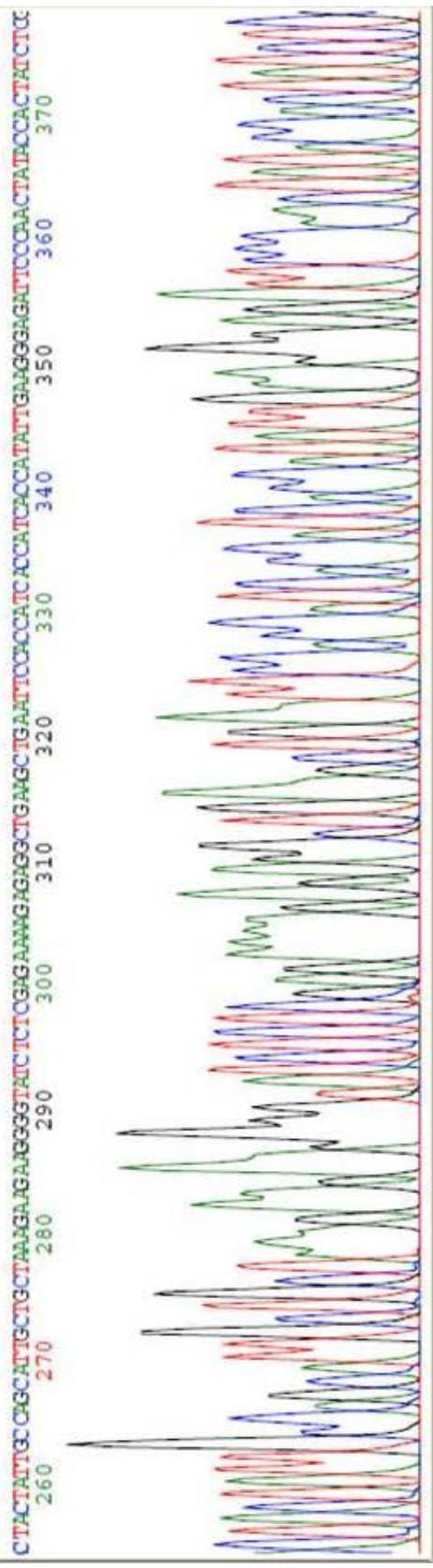
733-24-15-12-30 AM.ab1
Signal G:472 A:344 T:294 C:274
DT310POP6{BDV3}v1.mob
BIG DYE V3.1.mtx
Points 1332 to 10200 PK 1 Loc: 1332

Cap 4



Page 2 of 3
Sat, Mar 24, 2015 11:20 AM
Sat, Mar 24, 2015 12:30 AM
Spacing: 11.13{11.13}

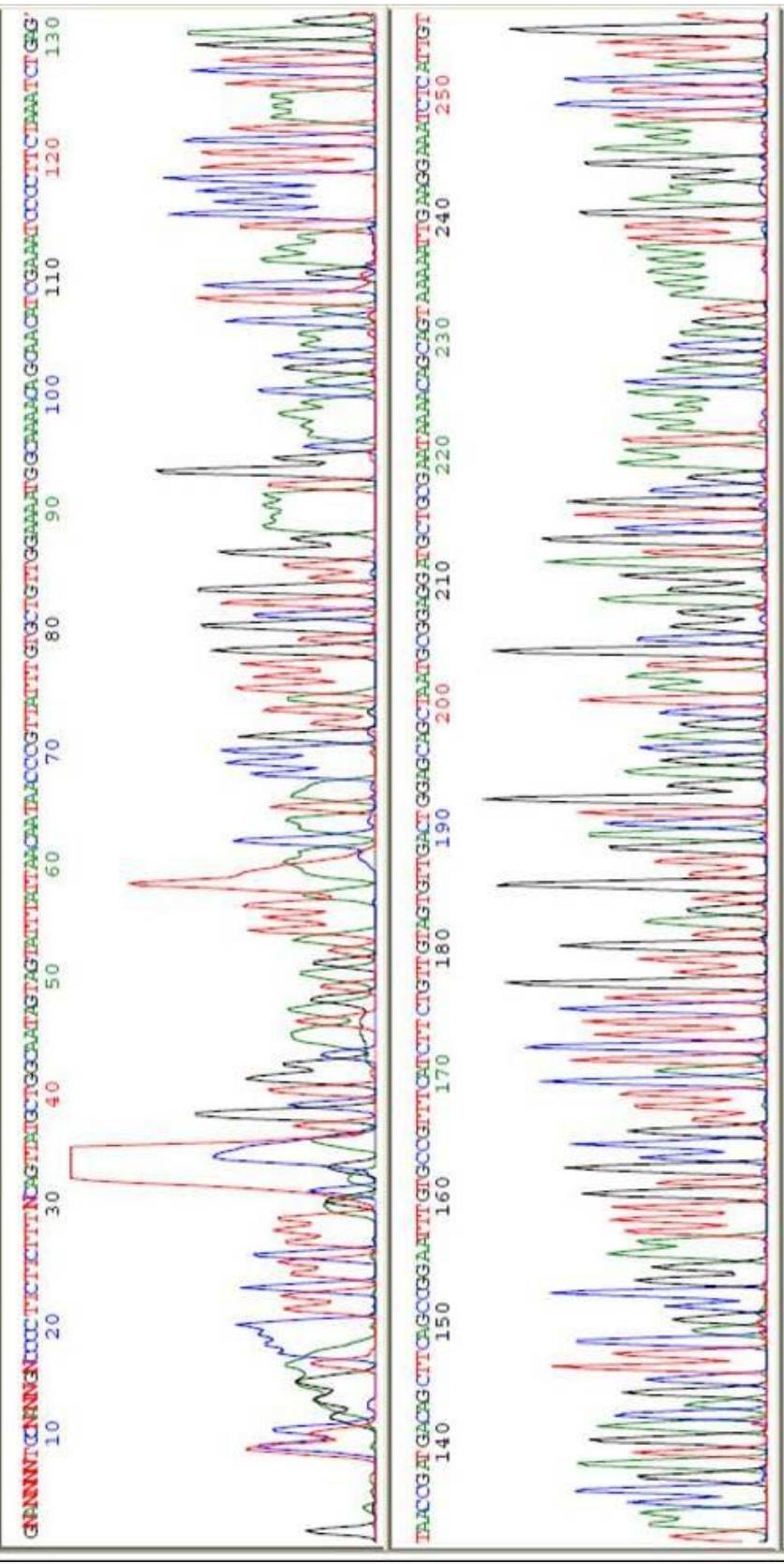


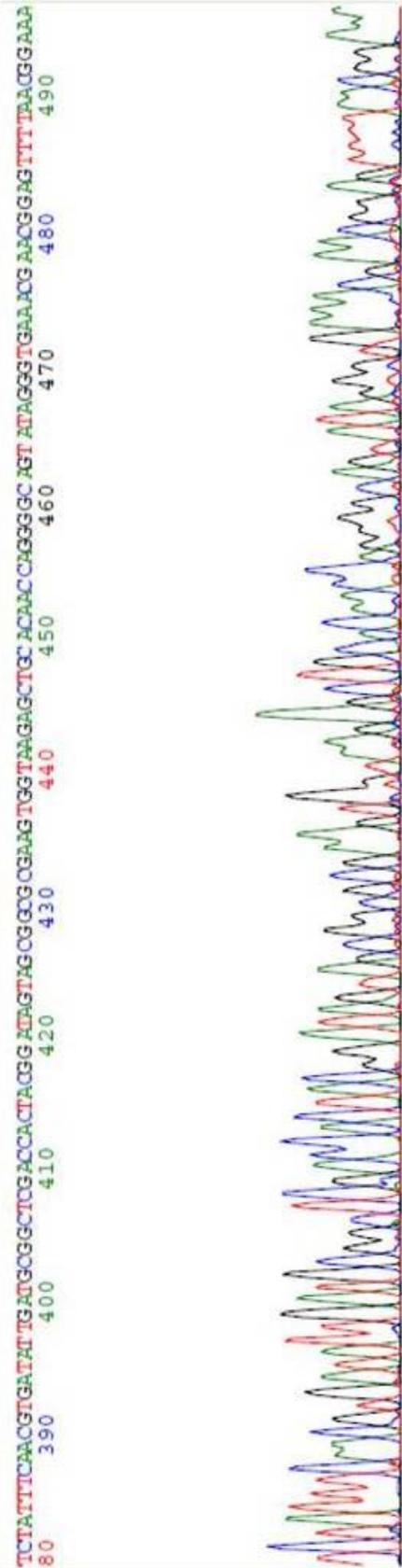
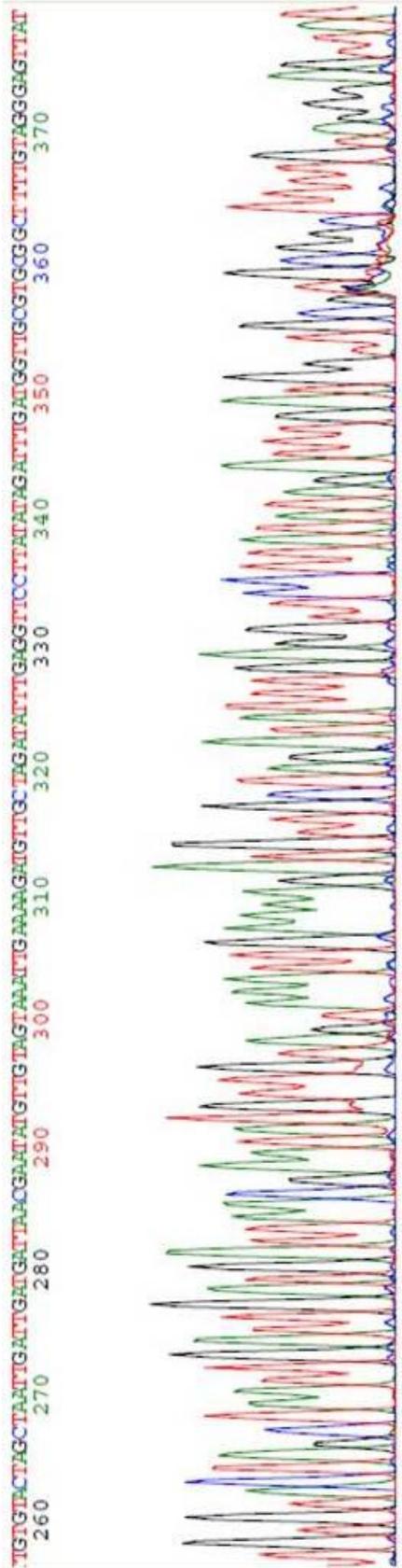


Reverse sequencing with promoter specific reverse primer:

Range 1: 17 to 640		Graphics		▼ Next Match ▲ Previous Match	
Score	Expect	Identities	Gaps	Strand	
1044 bits(565)	0.0	605/625(97%)	10/625(1%)	Plus/Minus	
Query	181	TTTTAA-GTTCGCTATT-AATC-ACTCA-GAATGGTC-TCACCTTGCTATTGGCCAAGTC			235
Sbjct	640	TTTTAAGGTTTCGCTATTGANNCAATTCAGGAATGGTCTTCNCCTTGNTATTGGCCAAGTC			581
Query	236	TGGAGTCGCCAGCTACCACCTCAGAGGCTACATAGACCTCCC-AA-TGT-CATCTCCTCA			292
Sbjct	580	TGGAGTCGCCAGCTACCACCTCAGAGGCTACATAGACCTCCCCAATTGTCCATCTCCTCA			521
Query	293	GTGCGCTCTTCAATCTC-GTGTCTTTCCGTTAAAACCTCCGTTTCGTTTCACCTATACTG			351
Sbjct	520	GTGCGCTCTTCAATCTCNGTGTCTTTCCGTTAAAACCTCCGTTTCGTTTCACCTATACTG			461
Query	352	CCCCTGGTTGTGCAGCTCTTACCACTTCGCGCCGCTACTATCCGTAGTGGTCGAGCCGCA			411
Sbjct	460	CCCCTGGTTGTGCAGCTCTTACCACTTCGCGCCGCTACTATCCGTAGTGGTCGAGCCGCA			401
Query	412	TCAATATCACGTTGAAATAGAATAACTCCCTACAAAAGCCGCACGCAACCATCAAATCTA			471
Sbjct	400	TCAATATCACGTTGAAATAGAATAACTCCCTACAAAAGCCGCACGCAACCATCAAATCTA			341
Query	472	TATAAGGAACCTCAAATATCTAGCAACATCTTTCAATTTACTACAACATATTCGTTAAT			531
Sbjct	340	TATAAGGAACCTCAAATATCTAGCAACATCTTTCAATTTACTACAACATATTCGTTAAT			281
Query	532	CATCAATCAATTAGCTAGTACACAACAATGAGATTTCCCTTCAATTTTACTGCTGTTTTA			591
Sbjct	280	CATCAATCAATTAGCTAGTACACAACAATGAGATTTCCCTTCAATTTTACTGCTGTTTTA			221
Query	592	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCA			651
Sbjct	220	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCA			161
Query	652	CAAATCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTCGATGTTGCT			711
Sbjct	160	CAAATCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTCGATGTTGCT			101
Query	712	GTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAAATACTACTATTGCC			771
Sbjct	100	GTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAAATACTACTATTGCC			41
Query	772	AGCATTGCTGCTAAAGAAGAAGGGG	796		
Sbjct	40	AGCATAACTGN-AAAGAAGAAGGGG	17		

Model 310 73_R1-30-15-12-00 AM.ab1 Signal G:148 A:95 T:85 C:57 Page 2 of 3
 Version 3.7 DT310POP6(BDv3)v1.mob Paz, ub 01, 2015 5:17 PM
 Basecaller-310PO73_R BIG DYE V3.1.mtx Cum, Oca 30, 2015 12:00 AM
 BC 1.3.0.0 Cap 8 Points 1446 to 11480 Pk 1 Loc: 1446 Spacing: 12.34{12.34}



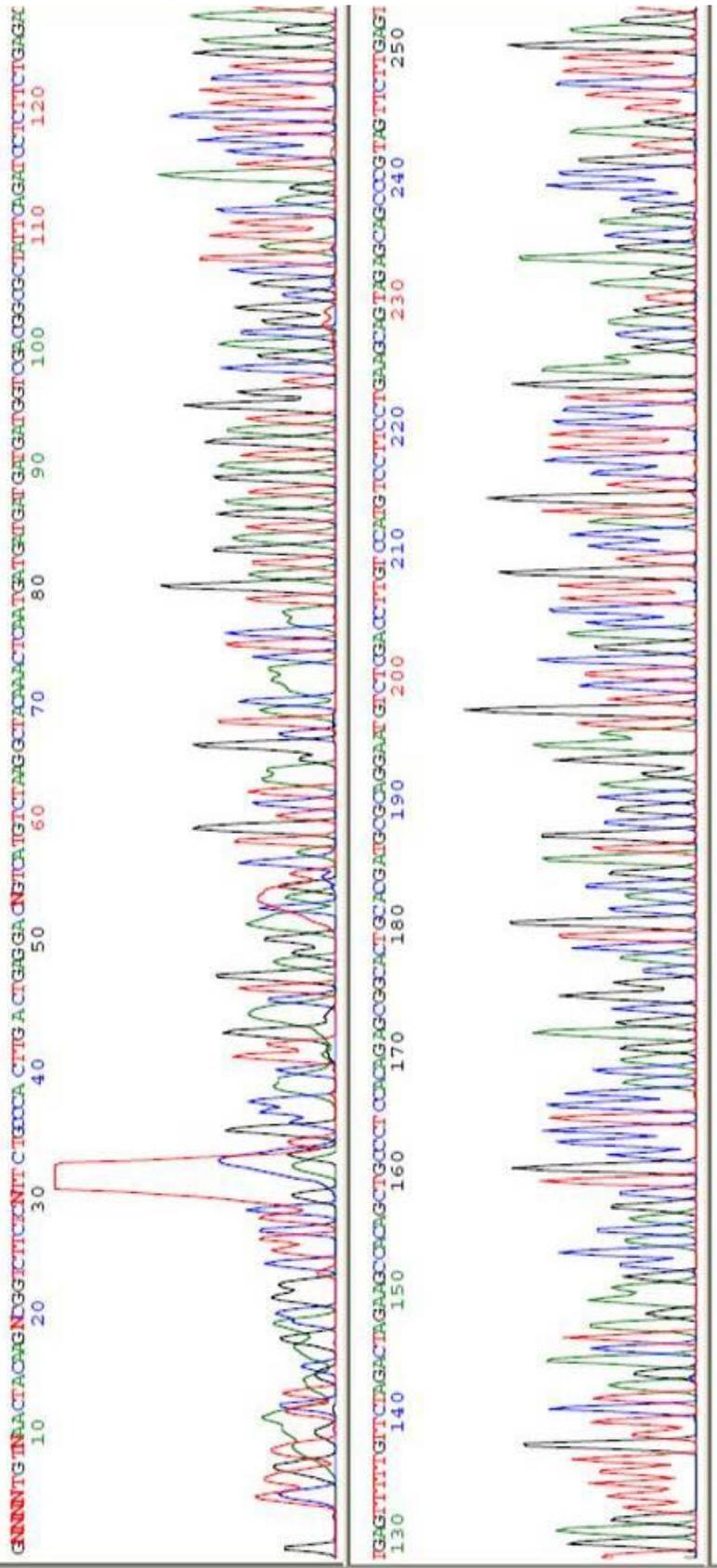


Reverse sequencing with AOX reverse primer:

Range 1: 20 to 751		Graphics		▼ Next Match ▲ Previous Match	
Score	Expect	Identities	Gaps	Strand	
1184 bits(641)	0.0	703/736(96%)	16/736(2%)	Plus/Minus	
Query	831	TCACCAT-ATTGAA-GGGAGATT-CCCAACTA-TA-CCACTATCTC-GTCTA-TTCGAT-	882		
Sbjct	751	TCACCATAATGGAANGGNANAITCCCAANNANTACCCNCTAICTCGGTCTATTTTCGATA	692		
Query	883	AACGCTAT-GCTTCGTGCTCATCGICTTCATC-AGCTGGCCTTTGACACCTIACCAGGAGT	940		
Sbjct	691	AACGNATATGGTTTCGTGNTCAICGICTTTCATCCAGCTGGNCCTTGNACCTNCCAGGAGT	632		
Query	941	TTGAA-GAAGCCTATATCCCAAGGAACAGAAGTATTCATTCCCTGCAGAACCCCCAGACC	999		
Sbjct	631	TTGAAINGAAGCCTATATCCCAAGGAACAGAAGTATTCATTCCCTGCAGAACCCCCAGACC	572		
Query	1000	TCCCTCTGTTTCTCAGAGTCTATTCCGACACCCT-CCAACAGGGAGGAAACACAACAGAA	1058		
Sbjct	571	TCCCTCTGTTTCTCAGAGTCTATTCCGACACCCTTCCAACAGGGAGGAAACACAACAGAA	512		
Query	1059	ATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGT	1118		
Sbjct	511	ATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGT	452		
Query	1119	GCAGTTCCTCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGT	1178		
Sbjct	451	GCAGTTCCTCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGT	392		
Query	1179	CTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGA	1238		
Sbjct	391	CTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGA	332		
Query	1239	TGGCAGCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAAACTC	1298		
Sbjct	331	TGGCAGCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAAACTC	272		
Query	1299	ACACAACGATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACAT	1358		
Sbjct	271	ACACAACGATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACAT	212		
Query	1359	GGACAAGGTCGAGACATTCTCGGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGTGG	1418		
Sbjct	211	GGACAAGGTCGAGACATTCTCGGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGTGG	152		
Query	1419	CTTCTAGTCTAGAACAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACcatc	1478		
Sbjct	151	CTTCTAGTCTAGAACAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATC	92		
Query	1479	atcatcatcatcatTIGAGTTTGTAGCCTTAGACATGACTGTTCCCTCAGTTCAAGTTGGGC	1538		
Sbjct	91	ATCATCATCATCATTGAGTTTGTAGCCTTAGACATGACNGT-CCTCAGT-CAAGT-GGGC	35		
Query	1539	ACTTACGAGAAGACCG 1554			
Sbjct	34	AGA-ANGAGAAGACCG 20			



Model 310 73_AR1-30-15-3-00 AM.ab1 Signal G:275 A:140 T:134 C:132 Page 2 of 2
Version 3.7 DT310POP6(BDV3)v1.mob Paz, 01, 2015 5:17 PM
Basecaller-310POI73_AR BIG DYE V3.1.mtx Cum, 03, 2015 3:00 AM
BC 1.3.0.0 Cap 9 Points 1538 to 11480 Pk 1 Loc: 1538 Spacing: 12.49(12.49)



K.3. pGAPZ23A::hGH

Sequencing was conducted with “new GAP forward” and AOX reverse primers specified in red highlights (available in Appendix I); the location of the primers can be seen in following sequence.

```
AGATCTTTTTGTAGAAATGCTCTGGTGTCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCA
ACTCCGAACGACCTGCTGGCAACGTAATAATTCTCCGGGGTAAACTTAAATGTGGAGTAATGGAACCAGAAACGTC
TCTTCCCTTCTCTCTCCTTCCACGCGCCGTTACCGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTACGG
CCCCCTTGAGCAATGCTCTTCCAGCATTACGTTGCGGGTAAACGGAGGTCTGTACCCGACCTAGCAGCCCAG
GGATGGAAGTCCCGGCGTCTGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGAAACCACCAGA
ATCGAATAATAAAGGCGAACACCTTTCCCAATTTGGTTTCTCTGACCCAAAGACTTTAAATTTAATTTATTTGT
CCCTATTCAATCAATTGAACAACATATTCGAAACGatgaaaatattaagtgcattgcttctcttttacgttgcccttgcT...AATTCAC
CATCACCATCACCATATTGAAGGGAGAttcccaactataaccactatctcgtctattcgataacgctatgcttcgtg
ctcatcgtcttcatcagctggcctttgacacctaccaggagtttgaagaagcctatatcccaaaggaacagaagta
ttcattcctgcagaacccccagacctccctctgtttctcagagtctattccgacacctccaacagggaggaaaca
caacagaaatccaacctagagctgctccgcatctccctgctgctcatccagtcgtggctggagcccgtgcagttcc
tcaggagtgtcttcgccaacagcctggtgtaecggcgctctgacagcaacgtctatgacctcctaaaggacctaga
ggaaggcatccaaacgctgatggggaggctggaagatggcagcccccgactgggcagatcttcaagcagacctac
agcaagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggctgctctactgcttcaggaagg
acatggacaaggtcgagacattcctgcgcatcgtgcagtgccgctctgctggagggcagctgtggcttcTAGT...CTA
GAACAAAACCTCATCTCAGAAGAGGATCTGAATAGCGCGTCCGACCATCATCATCATCATATTGAGTTTCTAGCC
TTAGACATGACTGTTCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGTAGATTCTAATCAAGAGGAT
GTCAGAATGCCATTTC
```

Forward sequencing with “new *GAP* forward” primer. The term “query” has been attributed to the expected sequence.

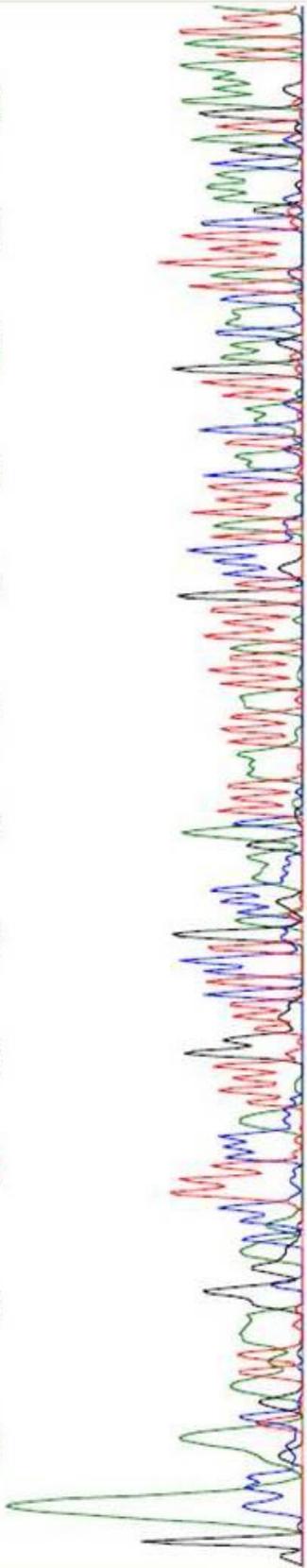
Range 1: 17 to 744 Graphics				▼ Next Match	▲ Previous Match
Score	Expect	Identities	Gaps	Strand	
1271 bits(688)	0.0	715/730(98%)	7/730(0%)	Plus/Plus	
Query	45	AAAGGCGAACACCTTTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTT			104
Sbjct	17	AAAGGCG-AC-CCTTTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTT			74
Query	105	ATTTGTCCTTATTTCAATCAATTGAACAACCTATTTGAAACGATGAAAAATTAAGTGCA			164
Sbjct	75	ATTTGTCCTTATTTCAATCAATTGAACAACCTATTTGAAACGATGAAAAATTAAGTGCA			134
Query	165	TTGCTTCTTCTTTTTACGTTGGCCTTTGCTGAATTCACCATCACCATCACCATATTGAA			224
Sbjct	135	TTGCTTCTTCTTTTTACGTTGGCCTTTGCTGAATTCACCATCACCATCACCATATTGAA			194
Query	225	GGGAGATTCCCAACTATACCACTATCTCGTCTATTGATAACGCTATGCTTCGTGCTCAT			284
Sbjct	195	GGGAGATTCCCAACTATACCACTATCTCGTCTATTGATAACGCTATGCTTCGTGCTCAT			254
Query	285	CGTCTTCATCAGCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTATATCCCAAAG			344
Sbjct	255	CGTCTTCATCAGCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTATATCCCAAAG			314
Query	345	GAACAGAAGTATTCAATTCCTGCAGAACCCCGACCTCCCTCTGTTTCTCAGAGTCTATT			404
Sbjct	315	GAACAGAAGTATTCAATTCCTGCAGAACCCCGACCTCCCTCTGTTTCTCAGAGTCTATT			374
Query	405	CCGACACCCCTCCAACAGGGAGGAAACACAACAGAAATCCAACCTAGAGCTGCTCCGCATC			464
Sbjct	375	CCGACACCCCTCCAACAGGGAGGAAACACAACAGAAATCCAACCTAGAGCTGCTCCGCATC			434
Query	465	TCCTTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCC			524
Sbjct	435	TCCTTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCTCANGAGTGTCTTCGCC			494
Query	525	AACAGCCTGGTGACGGCGCCTCTGACAGCAACGTCTATGACCTCCTAAAGGACCTAGAG			584
Sbjct	495	AACAGCCTGGTGACGGCGCCTCTGACAGCAACGTCTATGACCTCCTAAAGGACCTAGAG			554
Query	585	GAAGGCATCCAACGCTGATGGGGAGGCTGGAAGATGGCAGCCCCCG-GACTGGGCAGAT			643
Sbjct	555	GAAGGCATCCAACGCTGATGGGGAGGCTGGAAGATGGCAGNCCCCGNGACTGGGCAGAT			614
Query	644	CTTCAAGCAGACCTACAGCAAGTTCGACACAACTCACACAAC-GATGACGCACTACTCA			702
Sbjct	615	CTTCAAGCAGACCTACAGCAAGTTCGACACAACTCACACAACGGATGACGCACTACTCA			674
Query	703	AGAACTACGGGCTGCTCTACTGC-TTCAGGAAGGACATGG-ACAAGGT-CGAGACATTC			759
Sbjct	675	AGAACTACGGGCTGCTCTACTGGNTTCAGGAANGACATGGNACAANGNCCGAGACATTNN			734
Query	760	TGCGCATCGT	769		
Sbjct	735	TGCGCATCGT	744		



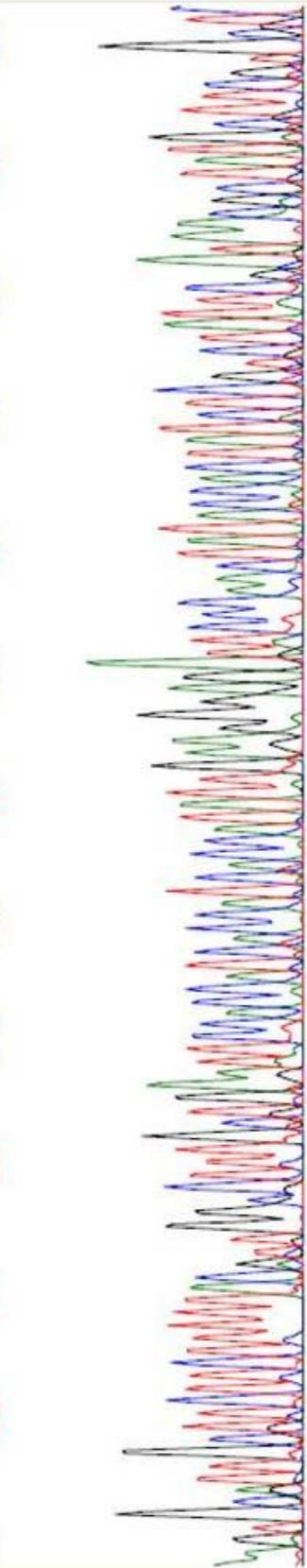
Model 310 20_GAP3-18-15-10-56 PM.abSignal G:317 A:221 T:189 C:218
Version 3.7 DT310POP6{BDV3}v1.mob
Basecaller-310PO20_GAP BIG DYE V3.1.mtx
BC 1.3.0.0 Cap 3 Points 1294 to 10200 Pk 1 Loc: 1294

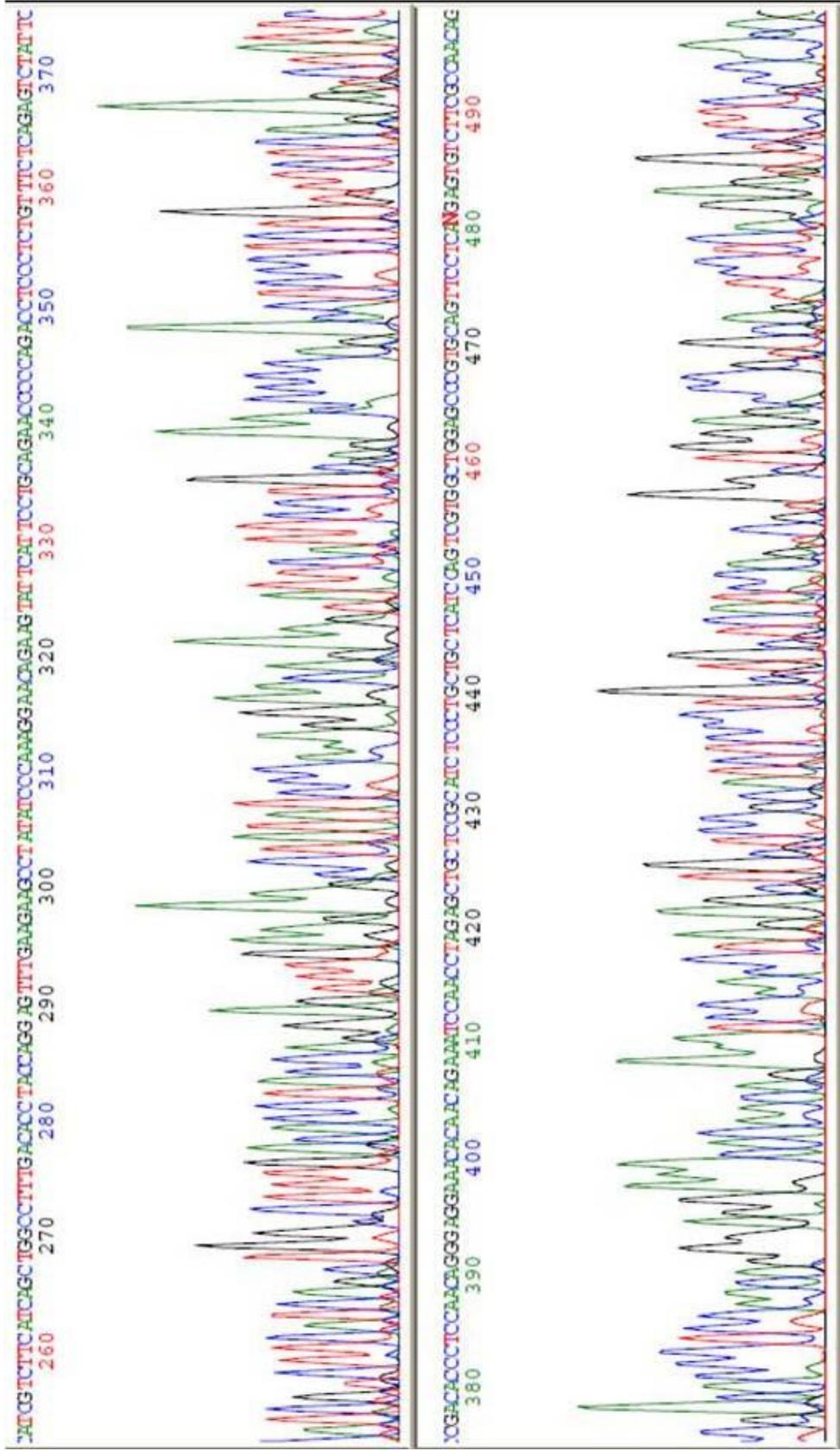
Page 2 of 3
Per, Mar 19, 2015 10:59 AM
Çar, Mar 18, 2015 10:56 PM
Spacing: 12.49(12.49)

GGAAACCAAAAGATTAAGGGGACCCCTTTTCCCATTTTGGGTTCTCCCTGGCCCAAGGACTTTTAAATTTTAAATTTTGGCCCTATTTTCAATCCATTTGAGACACTATTTTTCGAAACGATCGAAATATTTT
10 20 30 40 50 60 70 80 90 100 110 120



AGTGGATGCTTCCTTTTAAAGTTGGGCTTTTGGTTGAATCCACCATCCACATTTGAGGGGAGATTCOCACACTATACCCACTATCTGGICTATTCGATTAACGTTAAGCTTTCGGTCTC
130 140 150 160 170 180 190 200 210 220 230 240 250





Reverse sequencing with *AOX* reverse primer:

Range 1: 18 to 740		Graphics		▼ Next Match ▲ Previous Match	
Score	Expect	Identities	Gaps	Strand	
1271 bits(688)	0.0	716/728(98%)	10/728(1%)	Plus/Minus	
Query	213	CACCATATTGAAGGGAGATTCCC-AACTA-TACCACTATCTCGTCTATTGATAA-CGCT			269
Sbjct	740	CACCATATTGAAGGGAGATTCCCCANCTATTCCCACTATCTCGTCTATTGATAANC			681
Query	270	ATGCTTCGTGCTCATCGTCTTCAT-CAGCTGGCCTTTGACACCTACCAGGAGTTGAAGA			328
Sbjct	680	ATGCTTCGTGCTCATCGTCTTCATNCAGCTGGCCTTTGACACCTACCAGGAGTT-GAAGA			622
Query	329	AGCCTATAATCCC-AAAGGAACAGAAGTATTCATTCCCTGCAGAACCCCGACCTCCCTCT			387
Sbjct	621	AGCCTATAATCCCCAAGGAACAGAAGTATTCATTCCCTGCAGAACCCCGACCTCCCTCT			562
Query	388	GTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAACAGAAATCCAACC			447
Sbjct	561	GTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAACAGAAATCCAACC			502
Query	448	TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC			507
Sbjct	501	TAGAGCTGCTCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTC			442
Query	508	TCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGACC			567
Sbjct	441	TCAGGAGTGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGACC			382
Query	568	TCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGCAGCC			627
Sbjct	381	TCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGCAGCC			322
Query	628	CCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAACTCACACAACG			687
Sbjct	321	CCCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAACTCACACAACG			262
Query	688	ATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGG			747
Sbjct	261	ATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGG			202
Query	748	TCGAGACATTCCCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGTGGCTTCTAGT			807
Sbjct	201	TCGAGACATTCCCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGTGGCTTCTAGT			142
Query	808	CTAGAACA AAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACcatcatcatcat			867
Sbjct	141	CTAGAACA AAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCAT			82
Query	868	catcatTGAGTTTGTAGCCTTAGACATGACTGTTCCCTCAGTTCAGTTGGGCACCTACGA			927
Sbjct	81	CATCATTGAGTTTGTAGCCTTAGACATGACTGT-CCTCAGT-CAAGT-GGGCAC-TACGA			26
Query	928	GAAGACCG 935			
Sbjct	25	GAAGACCG 18			

L. Amino acid and nucleotide sequence of the corresponding proteins of utilized SPs.

The grey highlighted nucleotides are SPs determined by prediction softwares.

L.1. Corresponding protein of SP13 (gi 254567645)

Cell wall protein with similarity to glucanases with chromosomal location of PAS_chr2-1_0052.

Amino acid sequence:

MLSTILNIFILLFLFIQASLQAPIPVVTKYVTEGIAVVTETNVRVVTKTIPIVQVLI SDGATYTHLTLTVSTAEENGN
FQPIITTSIVNKEVVVPTSVFPNTQQTRPTQVDTTQNNADTPAAPTPSPPTSSNNGVFTTYSTTRSVVTSV VVVVGPD
GSPIENTGQTANPTTAPTSTTAARTSSSTSTPTASSTPGGNHPRSIVYSPYSDSSQCKDATTIETDLEFIASKG
ISAVRIYGNDCNYLTVVLPKASLGLKVNQGFWIGPSGVDSIDDAVQEFIQAVNGNNGFNWDLFELITVGN EAI SAG
YVSASSLISKI KEVSSILSSAGYTGPIITTAEPNVYEDYGLDLCSTDVMSIVGVNAHSYFNTLFAASDSGSFVKSQIE
VVQKACRSRSDITIIETGYPSQGATNGKNVPSKENQKTAIFSI FEVVGTDVITLSTYDDLWKDFPGPYGIEQFFGAIDL
FS

Nucleotide sequence:

ATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGCATCCCTACAGGC
TCCAATACCTGTAGTTACCAAATATGTTACCGAAGGTATTGCCGTTGTTACGGAGACCAACGTCCGCGTTGTA ACTA
AAACAATTCCAATCGTTCAAGTCCTAATCTCCGATGGGGCAACTTACACTCACACACTCACTACCGTTAGCACC GC
GAAGAGAACGGAAATTTTCAACCGATCACAACGACATCCATCGTGAATAAAGAGGTTGTTGTGCCAACTTCTGT TAC
CCCGAACACCCAACAACACGACCAACGACGAGTGGATACAACCTCAAAACAACGCTGACACTCCTGCCGCTCCTACTC
CGTCCCTACTACGTCACGCAATAACGGCGTATTCACTACCTACAGTACTACTCGTAGTGTAGTGACGCTGTG TAGTG
GTAGTTGGTCTGATGGGTCCCTATTGAAAACACTGGACAACTGCAAATCCAACAACAACCGCCCTACTACTTC
GACTACGGCCGCTCGTACTACAAGTCAACATCAACCACTCCTACCCGATCTAGTACCCCTGGTGGTAACCATCCAC
GAAGCATTTGTTACAGTCCATATTCGACTCAAGTCAATGCAAAGATGCCAGACTATTGAACTGATTTGGAATTT
ATTGCTCTAAAGGAATTTCTGCTGTAAGAATTTACGGTAATGATTGCAACTATCTCACGGTTGTGCTTCCCAAGTG
TGCTTCTTTGGGTCTAAAGGTTAACCAAGGTTTCTGGATTGGTCCATCTGGTGTGACTCTATAGATGATGTGTCC
AGGAATTTATCAAGCTGTGAATGGAAACAATGGGTTAATTGGGATCTCTTTGAGCTCATTACTGTGGGCAATGAA
GCTATCAGCGCAGGTTACGTTTCAGCTTCTAGTCTAATTTCAAAGATCAAAGAAGTCTCGTCCATATTTATCAAGTGC
AGGATACACTGGTCCAATTACTACTGCTGAGCCACCAAAATGTTTATGAAGATTACGGAGATCTATGTTCCACGGATG
TGATGTCTATCGTTGGAGTTAATGCTCATTCTACTTCAACACCTTGTGTTGCTGCTTCTGATFCCGTTCTTTTGTC
AAAAGCCAGATTGAGGTCGTCCAAAAGCATGTTCCCGATCAGACATCACCATTATAGAAAACAGGATATCCAAGCCA
AGGAGCAACTAACGGAAAGAATGTTCTTAGCAAAGAAAATCAAAAACCTGCAATATTTCCATTTTGGAGGTAGTGG
GAACAGATGTCACAATCCTGAGTACCTACGATGACCTATGAAAAGACCCAGGCCGATGGCATTGAGCAGTTTTTC
GGAGCTATCGATCTGTTTAGTTAA

L.2. Corresponding protein of SP23 (gi 254565391)

Protein disulfide isomerase (PDI), multifunctional protein resident in the endoplasmic reticulum lumen with chromosomal location of PAS_chr1-1_0160.

Amino acid sequence:

MKILSALLLFTLAFAEVIELTNKNFDDVVLKSGKYTLVKFYADWCSHCKRMNPEYEKLAELKPKSDLIQIAAIDA
NKYSKYMKVYDIDGFPTMKLFTPKDISHPIEFSGSRDSESLNLFLESTTGLKLLKKAEVNEPSLVQSIDDSTIDDLV
GKDRFIAVTASWCGYCKRLHPEWEKLAKEAFGNDDIVIGNVVTDVVEGENIKAKYKVSFPITLYFTAGSDEPIRYES
PDRTEVEGLVKFVNEQAGLFRDPDGTLEFNAGLIPGVSDKLTNYIKEKQSLLESTLDLLSNHEHIKDKFVSKYHKVK
IEKLLKGENEFLNNEVERLSKMLNTKLSANNSDSVIKRLNLRNFIEAKTESKPLLHQEL

Nucleotide sequence:

ATGAAATATTAAGTGCATTGCTTCTTCTTTTTACGTTGGCCTTTGCTGAGGTCATTGAGCTGAC
CAACAAGAACTTTGATGACGTGGTTCTAAAGTCCGAAAGTACACCTTAGTGAAGTTTTATGCCGATTGGTGTTCGC
ATTGCAAGCGAATGAATCCAGAGTATGAAAAGCTGGCCGAAGAACTGAAGCCAAAGAGTGATCTGATCCAGATTGCC
GCCATTGATGCTAACAAATACTCAAAGTACATGAAGGTGTACGATATTGATGGATTTCCGACGATGAAATTGTTTCC
ACCAAGGACATATCTCATCCGATCGAATTTCTGGATCAAGAGACAGTGAAAGCTTTTGAACTTTTGGAGTCAA
CTACTGGTTTGAAGTTGAAGAAGAAGGCCGAAGTAAATGAGCCTTCGTTAGTTCATCAATTGATGATTCAACAATA
GATGACCTTGTGGGAAGGACAGGTTTATGTCAGTACTGCTTCGTGGTGTGGATATTGCAAAAGATTGCATCCTGA
ATGGGAGAAGTTAGCCAAAGCTTTGGCAATGACGATATTGTTCATCGGAAACGTTGTTACCGATGTTGTGGAAAGGTG
AGAATATTAAGCGAAGTATAAAGTTCAATCTTCCCGACTATCCTGTACTTCACAGCAGGCTCAGATGAACCAATA
AGATATGAATCTCCAGATAGAAGTGTGAAGGTTTGGTTAAATTTGTCAATGAACAAGCTGGCTTATTTCTGTATCC
AGATGGAAGTTTGAATTTCAACGCCGGTCTAATTCAGGAGTGAGTGATAAAGTTACCAATTACATTAAGGAAAAAG
ACCAAGTTTATGGAGTCAACGTTAGACTTGCTAAGCAACCATGAACATATCAAGGACAAATTCAGTGTCAAATAC
CACAAGAAGGTCATAGAAAAGTTGTTGAAGGGAGAGAATGAATTCCTCAACAATGAAGTTGAGAGGCTATCAAAAAT
GCTGAATACAAAGCTATCGGCAACAATTCAGACTCTGTGATTAAGAACTGAATATCTTGGCAATTTTATTGAGG
CCAAAAGTGAAGTCAAAACCCAGTTATTACCAAGAGCTATAA

L.3. Corresponding protein of SP24 (gi 254572565)

Hypothetical protein with chromosomal location of PAS_c131_0001.

Amino acid sequence:

MKVSTTKFLAVFLVRLVCAVDSGQLRIGITRKPVPDECVQKTSQSDTVAIHYEGSLEDGTFDSSYERDQPLEFVL
GSQVIRGWDQGLQNMCTGEQRKLTIPDLGYGSRGIGPIPANAVLGESI IWIDLKKTNSASFQGRIGGYQEKR

Nucleotide sequence:

ATGAAAGTTTCTACGACCAAATTTCTGGCTGTGTTCTTATTAGTTAGACTCGTTTGCCTGT
TGATTCAGGTCAATTAAGGATTGGTACGTAGTACTTGAGTTCCACAATTACGTGTGCTAATAAGTGTTTTTAGGAA
TCACTAGGAAAGTACCACCAGATGAGTGTGTCCAAAAGACTCAATCAGGAGACACAGTTGCTATACACTATGAAGGG
TCTCTGGAGGACGGCACCATTTTGGATTCATCTTATGAACGTGATCAACCATTGGAGTTGTTTTAGGATCTGGACA
AGTTATCAGAGGTTGGGATCAAGGTTTACAAAAGTACGTGATTTTGGAGCTTCTTTCTTCTCAACTTTTTTACTGA
ATACTAATATTAAGCATGTGTATTGGAGAGCAGAGAAAGCTAACAATTCCTCCTGATTTAGGTTACGGCAGCCGCGG
AATAGGCCCAATTCAGCAAATGCTGTACTAGGTGAGTCTATTATTGGATTGATTTGAAAAAGATATACTAAGTCCG
CTAGTTTTCAAGGCAGAATTGGTGGATATCAAGAAAAGAGATGA

L.4. Corresponding protein of SP26 (gi 254572672)

Hypothetical protein with chromosomal location of PAS_chr4_0040.

Amino acid sequence:

MWSLFISGLLIFYPLVLGDDPKVEVDTELNDASPPKMSWQSWHMKSEHDIDGHDTDAVFKLHDYRHVNSLNHDDILR
MYGLLREEVVGKGDGTGGHDESEGISQKVVDKIVISTVFGLIDNSDGEISFDEWQKFSKGGELPDMGVGVGHLELDF
EAEYERHHWLKYHAENDPDVTIQHPEDIAHELLHHEHEVEHDETIANKRVIDGFKETVVLIEINIPDLFRSK

Nucleotide sequence:

ATGTGGTCGCTGTTTCATATCTGGACTATTAATCTTCTATCCTTTGGTCCTTGGAGATGACCCA
AAGGAGGTGGATACTGAGCTGAATGACGCTTCTCCTCCAAAAGGCATGTCCTGGCAATCGTGGCATAATGAAATCAGA
ACATGATATCGACGGACATGATACTGACGCTGTCTTCAAGCTGCATGACTATAGACATGTTAATCTTTGAACCACG
ACGATATCCTTAGGATGTACGGCCTTTTAAGGGAAGAAGTAGTTGGAAAGGGAGACGGAACCTGGTGGTCATGATGAG
AGTGAAGGGATATCCAGAAAAGTTAAAGACAAAAGTTATTTCCACTGTTTTGGATTGATAGACTCTAATAGTGACGG
CGAAATATCTTTTGATGAATGGCAGAAATTTCTGCAAAAAGGAGGAGCTTCCAGACATGGGTGTTGGAGTGGGGC
ATGAGTTGGACTTTGAAGCAGAAATATGAGAGACATCACTGGCTTAAATACCATGCCGAAAAATGATCCCGATGTTACA
ATTCACATCCAGAAGATATAGCTCATGAACCTTTTACATCACGAGCATGAAGTGGACATGACGAAACGATTGCCAA
TAAACGAGTTATTGATGGTTTTAAAGAGACGGTCGTCTAATTGAAAACATTCCTGATCTTTTCAGGTCAAATAG

L.5. Corresponding protein of SP34 (gi 254565679)

Cell wall protein that functions in the transfer of chitin to β - (1-6) glucan
with chromosomal location of PAS_chr1-1_0293.

Amino acid sequence:

MRPVLSELLLLASSVLADEVI ECDADNKC PEDKPCCSQYGVCGTGVNCLGGCDPRHSFNASACLMPVCRDVLKAST
DAFEIDTNYLGDANETDWWVYNGYLIDYDDSVLLAMPKESYGTVVSSTFYVWYVKITATLKT SRGAGVVT SFILFSNV
HDEIDWFEVGYNLSQVETNYYQGVLYNTNGRNVSL EEDVNSFEYFHDYEIDWKEDVITWSIDGDVVRTLKKEDTYN
ETTDKYMFPQPSRVQLSIWPAGAESNAIGTVSWAGGNVDWSEDIQDPGYFYTLKELTVECYDVPDGTEDGELA
YYFKESDAFDQGDIIITNNSTKIKSLDDTGFDPDEDDDDDESSSSSSSSSSSSSSSSSSRTGSSSTSSATSTSTNSND
DDDNNSPTTSSGTSASAAGSFVNMSQTSGSSSATSNNAASLSAGFLTITISFFASVLGFL

Nucleotide sequence:

ATGAGACCAGTGCCTTTCGTTATTACTCTTGCTGGCTTCTTCGGTACTCGCTGACGAGGTCATTG
AATGCGATGCTGACAACAAGTGCCCTGAAGATAAACCATGTTGTAGTCAATATGGAGTATGTGGGACTGGTGTGAAT
TGCTTAGGAGGATGTGACCCTAGACATCTTTCAACGCGTCCGCATGTTTACCCATGCCAGTTTGTTCGTGACGTGGA
TTTGAAGGCTAGCACTGATGCTTTTCGAAATTGATACCAACTACCTTGGTGATGCTAATGAAACCGACTGGGTTTACA
ACGGTTATCTAATAGATTATGACGACTCAGTATTGTTGGCCATGCCAAGGAGTCTTATGGTACCGTTGTGTCTTCC
ACCTTCTATGTGTGGTACGGAAAAATCACAGCTACTCTTAAGACTTCCAGAGGTGCCGGTGTGTGCACGTCTTTCAT
TTTATCTCCAATGTGCATGATGAAATCGATTGGGAATTTGTTGGTTATAATCTAAGTCAAGTCGAAACGAACTACT
ACTACCAGGTGTCTTGAACATATACCAACGGTCGTAATGTTTCCCTAGAGGAGGACGTC AATTCCTTGAATACTTC
CACGATTATGAAATTGACTGGAAGGAAGACGTCATAACCTGGTCTATTGACGGCGACGTTGTCAGGACACTGAAAA
GGAAGACACTTATAACGAAACTACAGACAAATACATGTTTCCCTCAAACCTCCATCGAGGGTT CAGTTATCCATCTGGC
CGGCTGGTGTGAATCCAATGCGATAGGTACAGTTTCTTGGGCTGGTGGAAACGTTGACTGGGATTCGAAGACATC
CAAGACCCAGGATACTTTTACTACAGTTGAAAGAGCTGACTGTTGAGTGTATGACGTCCCAGATGGTACCAGGA
AGATGGTGAATTGGCATATTACTTCAAAGAGTCCGATGCTTTTGACCAAGGGGATATCATAATTACTAATAACAGCA
CCAAAATAAAGTCTTTAGACGACACTGGCTTTGACCCCGATGAAGACGATGACGATGACGAATCCAGCAGTAGCAGC
AGCAGCAGCAGCAGATCGTCTTCAAGTCTTCAAGGACCCGAGTTCATCTACATCCAGTGTACTTCCACTTCCAC
TTCTAACTCAAATGATGATGATGATAACAATAATTCGTCTCCTACTACTTCGTCTGGCATTCTCCCGCAGCTTCCG
GATTCGTTCAAATATGAGCCAACTT CAGGCTTTCATCAGCAACGTCAAATAATGCCGCTGCCTCTCTTTCAGCT
GGTTTCTTACTACTATTTCTTTTTTGCATCTGTATTAGGATTCCTTTAG

L.6. Corresponding protein of *S. cerevisiae* α -MF

Pheromone (mating factor)

Amino acid sequence (<http://www.uniprot.org/uniprot/P01149>):

```
MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKE  
EGVSLDKREAEAWHLQLKPGQPMYKREAEAEAWHLQLKPGQPMYKREADAEAWHLQLKPGQPMYKREADAEAWH  
WLQLKPGQPMY
```

Nucleotide sequence (www.ncbi.nlm.nih.gov):

```
ATGAGATTTCTTCAATTTTTACTGCTGTTTTATTTCGCAGCATCCTCCGCATTAGCTGCTC  
CAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCGGCTGAAGCTGTCATCGGTTA  
CTCAGATTTAGAAGGGGATTTGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAAC  
GGGTTATTGTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTC  
TCGAGAAAAGAGAGGCTGAAGCTTGGCATTGGTTGCAACTAAAACCTGGCCAACCAATGTACAAGAGAGAA  
GCCGAAGCTGAAGCTTGGCATTGGCTGCAACTAAAGCCTGGCCAACCAATGTACAAAAGAGAAGCCGACGCTGAAGC  
TTGGCATTGGCTGCAACTAAAGCCTGGCCAACCAATGTACAAAAGAGAAGCCGACGCTGAAGCTTGGCATTGGTTGC  
AGTTAAAACCCGGCCAACCAATGTACTAA
```

M. Nucleotide sequence of the utilized promoters.

M.1. PYRK promoter

The nucleotide sequence of the pyruvate kinase (PAS_chr2-1_0769) in GS115 strain:

```
>gi|254570207:c1466838-1465321 Pichia pastoris GS115  
chromosome 2, complete sequence
```

```
ATGTCAACTTCTTCAAACCTGGATGGCTATCCAACTGGATGTATCATCCACTCCAGAGCGCAATCTTCGCCGATC  
TTCCATTATTTGGTACGATTGGTCCCTAAGACTAACAGTCCAGAACTACTTGTCTAGTCTGAGACAGGCTGGGCTTAATA  
TTGTTAGAATGAACCTCTCTCATGGTCTTACGAGTACCATCAATCAGTTGTGGACAATGCCAGAAAATCAGAGGAA  
ATATATCCTGGTAGACCGCTTGCAATTGCTTTGGACACCAAGGGTCTGAAATCAGAACAGGAACTACCAAAGGTGA  
AACAGATTATGCAATTCCTATGGGACATGAAATGATATTCACCACAGATTTGTCTTTGCGCAAAGTCTAGCGATGACA  
AAGTGATGTTTCAATTGATTATAAGAACATTACAAAGGTTATTGAACCTGGAAAGATTATTTATGTGGATGATGGTGT  
CTTTTCAATTTGAAGTTTGTAGAAGTTCGTGATGAAAACACTCTCAAAGTCAGATCTATCAATGCTGGTGAATTTTCATC  
CCATAAAGGTGTCAATTTACCTAACACAGACGTGGACTTGCCAGCTCTTAGTGAAAAAGACAAGCAAGATCTGAGGT  
TTGGTGTGAAAACAAAGTCAACATGGTATTTGCATCCTTCATTCGTTGTGCTAATGATATCAAGGAGATTCGCCAT  
GTTCTCGGAGAAGATGGCAAACAGATCCAGATTTTGTCTAAGATTGAGAACCAGCAAGGAGTTAACAATTTTGTATGA  
AATCCTGGAAGTCACTGATGGTGTCTATGGTGTGCTAGAGGAGATCTAGGTATTGAAAATCCCTGCACCTCAGGTATTTG  
TTGTTCAAGAAGCAATTGATTGCCAAGTGAATCTGGCAGGTAACCCGTAATCTGTGCTACCCAAATGTTAGAATCT  
ATGACTTACAACTCCTAGGCCAACAGAGCCGAGGTTTCGGACGTTGGAAATGCTATTTTGGATGGCGCCGATTTGTGT  
GATGTTATCCGGGAAACTGCCAAGGAACTACCCCTCATGAGGCTGTTGCTATGATGCACCACACTGCTCTAATTG  
CAGAGTCAGCTATTGCTTACCTCCACACTACAACGAAATCAAGGATCTTGCTCGTGGTCTTATTAACACAGTTGAA  
ACTATTGCTATTGCCGCTGTTTCTGCTCACTTTGAACAAAATGCCAAGGCCATTGTTGTGCTTTCTACTTCAGGAAC  
TTCAGCAAGAAATGATTTCTAAGTATAGACCGAATTGCCCAATCCTTATGGTAACCAGAAATGATGAGGCAGCAAGAT  
ATTCTCATCTCTATCGTGGAGTATATCCATTCATCTATAAACAGGAAGTTAATGATAACTGGCAACAAGATGTTGAA  
GAACGTTTACAATATGCCATCACTGAAGCATTGGCATGGGAATATTGAAAAAAGGTTGATGCTATCGTTGCAGTACA  
GGGGTGGACCAAGGGACTGGGTACACCAATACTATGCGCGTTGTTTTTGTCTAA
```

The nucleotide sequence of the PAS_chr2-1_0770 in GS115 strain:

```
>gi|254570207:c1468638-1467184 Pichia pastoris GS115  
chromosome 2, complete sequence
```

```
ATGAAACAATCAAGACTGAGACACACTGATACCATACTCCTCAATTCCAATCTTTTTGTAATTTCAAACGACTTCAA  
CGAGCCAGCATCTGGAACCTCATCTCAATATATCCGCCGTTGGACAGAGCCCTATATATAGGTGTTGCCGGAACCT  
CGGGGCTCGGTAAGACCAGTGTGCTAAGCATATTGTGAAGGCATCAATCAACCCGACAGTTGATTTGTCTCTG  
GATAACTTCTACAAGGTATTGACCCCGGAGCAACATATTTGGCGGAGCATGCGCAGTATGATCTTGATTCCGCAAC  
CGCTTTGGATTTTGAATTAATGCTTCTGTTGATTGGAGACCTTAAAACCTGAAAAACCAACGCAACTGCCGGTGTATG  
ACTTTTGACCCATTTCCCGTACAGAAAAGACTACAACGATTTATGGAGCATCTGTCAATGTTGGTTGAAGGTTTGTG  
GCACCTCCATCATGGACAATTAAGTGGATTTAATGGATACAAAAGTTTTGTGGATACCTGATCTAGATAATATGCATGGC  
TCGAAGGGTTAAAAGAGACCTGATTGAAAGGGGTAGAGATTTGGAAGGCATCCTTGACCAGTGGGATCGACATGTGA  
AGCCTAACACAATTCGGTATGTGATCCCCAGCTCCAAGAATGCGGATTTGATCCTACCTCGCAGCACTGATAATAAA  
ATTGCACTTGATATGATTTTCGCCATATCAATAACAGTTGGAACAAAAGTCAATGGTTCACCTGAAAAGACTTCA  
AGAGCTGGGGCAGATATCTAACGATGAGACTCTCATGAACCGGATAGCACGTTTGCCGCTAACAAAATCAGTTAAAAT  
GTATAAGTACCATTCTTTTTGACAGAGAACTTCTCGTACAGAGTTCATCTTTTACTTTGATCGGGTTGCTAACATG  
CTGATCCATCTGGCATTGGAACAGGTAGAGTTCGGACCCCTCGCAAGATGAGGTATTGACCCGCAATACCATTGCCT  
AACTGATCGGATACGACCGTTACAATCGGTTGTGCTGTTGACTATGGTACGGACAGGTGATGATTTTATGAATTCAA  
TCAGAAAACCTATTCCAGATGTAAGATTGGTAAGTTGCTAATTTCAATCAGACCTAATTACAGGCGAACCTCAATG  
CATACAAAGTCGCTGCCTCCATGTGAACAAAACCTACCAAGCTACTATTATTCGATGCGCACATTATATCGGGGGCCGC  
AGCAATTATGGGCATTCAGTACTTCTGGACCATGGTATTGAAGAAGGTAATATCGTATTGTAAGTTATCTGCGAG  
AAGAAGCTGGCCTACGTCGATACTGAACGCTTTCCAAAAGGTTACTATTATCGTAGGCTTATCCTCTGGGAGGATG  
ACCTCATTATTGAAAGAGCCAATGTTTCGTACACGGTTCATCGACGATTACTACTTCGGCAGTACGTAG
```

The region between two genes, i.e., PAS_chr2-1_0769 and PAS_chr2-1_0770 in the sequenced strand of chromosome 2:

```
5' ...TGGATTGATTTTCAGTTAAAAAAGTCTAAAAACAATTTACTGTACATGAACGGTCGCCTGCTAGGTTATATATA  
CCTATTACAGCAAAGTTGGATGCCATTAATTTTGTATCAGGGGACGACATTTCTAGGGGACCGTTAAGAAGCGTCCG  
TCCCCACGCCCTATACAATCTCAGCACAAAAGGTTTCATTCGCAAAATGATGGTGGGAGACCGGAGTAATAGTTTAC  
TCGTGGAAAACCCCTCATCCTTGAATGTGAAAAACGTACGGAACCTCTGATCCGATTTGGTTTAGTTTAAACAATTC  
TTTGAACCCCCGAGCGGAATACAGTATCCCGCACACTGAA...3'
```

The complementary sequence on the desired (complementary) strand:

```
3' ...ACCTAACTAAAGTCAATTTTTTTTTCAGATTTTGTAAATGACATGTACTTGCCAGCGGACGATCCAATATATAT  
GGATAATGTCGTTTCAACCTACGGTAATTTAAACTAGTCCCTGCTGTAAGATCCCCTTGCCAATCTTCGCAGGC  
AGGGGTGCGGGATATGTTAGAGTCGTGTTTCCAAAGTAAGCGCTTTACTACCACCTCTGGCCTCATTATCAAGTG  
AGCACCTTTTGGGAGTAGGAACCTTACACTTTTTCATGCCTTGGAGACTAGGCTAAACCAAAATCAAATTTGTTAAAG  
AAATCTTGGGGCTCGCCTTATGTCATAGGGCGTGTGACTT...5'
```

The putative promoter region represented in Figure 4.12 is the above sequence but in 5' → 3' direction.

M.2. PDC promoter

The nucleotide sequence of the pyruvate decarboxylase (PAS_chr3_0188) in GS115 strain:

>gi|254572604:c383969-382287 Pichia pastoris GS115 chromosome 3, complete sequence

```
ATGGCTGAAATAACACTAGGAACCTACATCTTCGAACGCTCTGAAGCAAATCGACGTCAAGACTATCTTTGGTGTTC  
AGGAGATTTCAACTTGGCTTTGTTGGATCACATCTATGAAGTTGAGGGCATGAGATGGGCAGGAAACGCCAATGAAT  
TGAACGCTGCTTACGCTGCAGATGGTTATTTCCAGAGTGAAGTCAATGGCTGCTTTGATCACCACCTTCGGTGTCCGGT  
GAGCTCAGTGCCGTCAATGGTATCGCTGGTTCCCTTGGCCGAGCATGTCGGTCTGCTGCATATTGTCGGTGTCCCAGC  
TATCTCTTCCCAGGAGAAGAAGCTGTGGCTTACCACACCCTCGGTAACGGAGACTTCCGGTGTCTTCAAGCGTGTCT  
TCAAGAAGCTCTCCAAGTCTGCCAAGTTTATCTCTGATATCAACGAGGCTCAAGATATGATCGACGGTGCATCCGT  
GAAGCTTTCATTTACCAACGTCCTATCTACCTTGGACTGCCAACAACCTTGTGAGATGAAGGTTGACAGAACTCG  
TTTGAACACCCCAATTGACTTGAAGCCAGTTCCAACCCAGTCGAGGCTGAGGAGGAGGCATTACAGTCCATCTTGG  
AACTGATCTCAAAGGCTTCAAAGCCAGTTATTTCTGGTTGATGCCTGTGCATCCAGACACTTCTGTCAAAAAGAGGTT  
GACCAACTCATTGACGTCACCAACTTCCAGTCTTGTGCTCACTCCAAATGGGTAAGGTGGTGTGACGAACAAAAGCC  
ACAGTTTGGTGGTGTCTTATGTGGGTTCTCTGTCTAACCCAGACATTACAGAGTTCGTGGAGAAGGCTGATCTTGTCA  
TCTCTGTTGGTGTCTTCTGTCTGACTTCAACACTGGATCTTTCTCTTACTCCCACTCCAAGAATGTTGTTGAGTTC  
CACTCTGACTACACTCAAATCCGTAGTGCAAGTTCCTCAAGGCGTTCAGATGAAGGCCCTATTGCAAAAAGCTGTTACC  
ACTTGTGTCGGTAAGGCATCTAAGCACATTACTGCTCAAGTTCCTCCAAAGATTGCTCCTCCTATCGAAAAGGGTGCCA  
GCCAAGATTTGACCCAGGACTGGTGTGGAGCAACATCAGCAAGTTCCTCAGAGCTGGTGACGTCGTCATCACAGAG  
ACCGGAACCTCTGCTTTCCGGTATCGTTCAATCAAGATTCCTCAACCCACGTATCTGGTATTTCCCAGGTCCTGTGGGG  
TTCCATTGGTTTCTCCGTCGGTGTACCTTAGGAGCTGTCTGTGCTACTGAGGAGCTCGACCCCAACAGAAGAGTCA  
TCCTGTTTGGTGGTGTGATGGTTCTTTGCAACTGACCGTCCAGGAGATTTCCACTATGGCTCGCTGGAACCTTGAAGCCA  
TACCTGTTTCATCCTTAATAACGATGGTTACACTATCGAGAGATTGATTACGGTGAGAAGGCTTCTTATAACGATAT  
TCAACCATGGGACCACCTTAAGCTGCTAGACACCTTTAAGGCCAAGAACTCTGAAAGTACTAGAGTCTCTACTGTGCG  
GAGAGTTGACCAAAATGTTCAAGGACCAAGGTTTCAACAAGCCTGACAAGATCCGTGTCATCGAGATCATGTTGGAG  
ACTATGGATGCTCCAATTTCTTAGTCAAGCAAGCTGAAATCACTGCCAAGACCAACGCAGCTTAA
```

The nucleotide sequence of the PAS_chr3_0189 in GS115 strain:

```
>gi|254572604:c386051-384516 Pichia pastoris GS115 chromosome  
3, complete sequence
```

```
ATGGCCTACAACAACTCAACCAATGGCGAACGACTGTTGAATACGTCAGACGCCCTGTTGAATGAAGATATAGGCGC  
CTTAGGGGTGTTTGCCTCTCATATCAACTCCTCTCGGCTAGTGACGCCGAAACGAAACTCCAGGCTTGACAGTTTAG  
ATTCTGTATCAAAGCCAGCAGCGGTGGACGAAATTACACCACCGACTCGGACAAGGAAGGTGAAATTCAGAAAAT  
AACCAAGCTTGGAAAGACTCAGAATGCCCCAGCTTGTATCCCCAGATCACATGCAGCATCCAGCAATGACCAACA  
CCTCCAAAACGGAAATGGCTTTTCTCATATAAACAACAATACGAACACTCTTAAGCCATTTAATCTCCCCCAATAT  
CTCCATCACCCTACCAATGAATGGCGGCAGCCTGAGAAAACATAAACGGCGTACTTAATGATCTTCAGCAGTAACC  
GACAAACAGCCAACAGCTACCACAACATCGCCAGGGTATCAGCATTACACTTTTGGCAGCAATATTGGAACAAC  
TGGAGGCAAAGACAGACTCGGAAAGGTCATACAATATGGGCTCAGAAATCTGCTAATTTATTCCAAGAGATCATCAA  
CATTTTTAAGCACAGTTTACGAAGTTCCGATGCAAGTCTGAACAAGTCCGGAGGGAGGTTTTCTTTACTGGTCTCT  
CTTCTGAGAAGGCCTGAACCTTTGTGTATTGGTTCCCTTAAGCAGTTTGAGAGAAGGCCACGGGTCTATCTCCGG  
CCTCTCAATGTATCGTCAAAATGCTACGTGCAGGTAACACTCCTTTCCGGCTAGTCAATCTCGGTTACAAGATTAATG  
ATACCTTCCAAGTTTTCAAGAAAGAAAGGAGTAGACGATGGCATTGTTTACTTCCGCAAGAATGGGTCAATGAGAAG  
ACTGTCACTGAACTGGCACAACCTGTATTATAGTTGGTTTGATGAGTCGTTACTATGTTTTAAGCTAAAGATGATAGA  
CGACCTAAATATCGGGAAATGCCACTAGACAGGAGCGCCTGGCATGGCTTCCACCATTCTCCTAATCTAAAGA  
AGAATATTTCAAATACCAAGCAAATAAACGCAAAGGCAACATATTAGGATAAACCAACAGGTTAAGTATAGAGCT  
AAAGCATTTGCTGAGCAATATTTCGGATCAACATCAACAGAATAGATCGTCACCACGAAGTATATCCAATCTCC  
AAGAATCCAGGCTTATCCCTCAAGAATGGCCTCTCCATCTCCTTCAATGAAGATGGCATTGACAACTCCGTCAGCT  
TGCAGGATACAATGGATGCGGAACTAGAAGCGTTGGCAAACGAACAATATTACGTCATGCTAGATATCCTTAAGGGA  
TTTTGCGATCTTTCAATTTGACATGGTAAACATATTTCTATCCAGATCCCTGAGGTATTACATCTCCTCTTCGGCTT  
TGGTGCAGGGTCACTGGCCTTAACGGGAGTCTGATCCAAAACGAGACAAGAACTAATAAATAAACAGATATAA
```

The region between two genes, i.e., PAS_chr3_0188 and PAS_chr3_0189 in the sequenced strand of chromosome 3:

```
5'...TGTTGTGTAAGTACTAGTAATGATTGATGATTAACGAATATGTTGTAGTAAATGAAAAGATGTTGCTAGATATTT  
GAGGTTCTTATATAGATTTGATGGTTGCGTGGCGCTTTTGTAGGGAGTTATTCTATTTCAACGTGATATTGATGCG  
GCTCGACCACTACGGATAGTAGCGGCGCAAGTGGTAAGAGCTGCACAACCAGGGGCGAGTATAGGGTGAAACGAACG  
GAGTTTTAACGGAAAAGACACGAGATTGAAGAGCGCACTGAGGAGATGACATTGGGAGGTCTATGTAGCCTCTGAGG  
TGGTAGCTGGCGACTCCAGACTTGGCCAATAGCAAGGTGAGACCATCTGAGTGATTAATAGCGAACTAAAATTAC  
TAGGTTCTGGTTAAATCCACGACTTGTGGGACGATTATGATAACCAAAAAGTGGACTACAGATGTCCATTCCGAAAA  
ACCTGGGAAGAAAATTCGGTAAATATGGCATAAACACGGAACAAAATTTGGTGCATATCAAGTGATTGGGTAATCGTG  
TGCTTGTC...3'
```

The complementary sequence:

```
3'...ACAACACATGATCGATTAACAACTACTAATTGCTTATACAACATCATTTAACTTTTCTACAACGATCTATAAA  
CTCCAAGGAATATATCTAAACTACCAACGCACGCCGAAAACATCCCTCAATAAGATAAAGTTGCACTATAACTACGC  
CGAGCTGGTGATGCCATCATCGCCGCGCTTACCATTCTCGACGTGTTGGTCCCGTCATATCCCACTTTGCTTGC  
CTCAAAATTCGCTTTTCTGTGCTCTAACTTCTCGCGTACTCCTCTACTGTAACCCCTCCAGATACATCGGAGACTCC  
ACCATCGACCGCTGAGGTCTGAACCGTTATCGTTCCACTCTGGTAAGACTCACTAATATCGCTTGAATTTAATG  
ATCCAAGACCAATTTAGGTGCTGAACCCCTGCTAATACTATTGGTTTTTCCACTGATGCTACAGGTAAGCCTTTT  
TGGACCTTCTTTTAAAGCAATTTATACCGTATTTGTTGCTTGTGTTAACCACGTATAGTTCACTAACCATTAGCAC  
ACGAACAGGG...5'
```

The putative promoter region represented in Figure 4.14 is the above sequence but in 5' → 3' direction.

M.3. TH13 promoter

The nucleotide sequence of PAS_chr4_0065 in GS115 strain:

>gi|254574547:129794-130816 Pichia pastoris GS115 chromosome 4, complete sequence

```
ATGTCTACTAACAAGATCACTTTCTGTCTTAACTGGCAAGCTGCCCCATACCATGCCCAATCTACCTGGCCCCAAA  
ATTGGGCTACTTCAAGGATGAGGGTCTTGATATTGCTATCTTGGAAACCAGGTAACCCATCTGATGTCACCTGA  
TTGGCTCTGAAAAGGTTGACATGGGTCTTAAGGCTATGATTCATACCTTTGGCTGCCAAGGCTCGTGGTTCC  
ACATCTGTGGTTCTCTTTTGGATGAGCCCTTACCAGGAAATTTGTATCTCGAGAGCTCCGGTATCACTGATTT  
ATCCCTCAAGGGAAGAGAATTGGTTACGTTGGTGAATTCGAAAAATCAATTAGATGAGCTGACCAAGCATTAT  
GAATGACTCCTGATGACTACACTGCTGTTAGAAAGCGGGATGAACGTTGCTAGAGAGATAATCAACGGTAACAT  
GCTGGTATGGTATTGAATGTGTCAGCAAGTTGAGTTGGAGGAGTATCTGAGATCCCAAGGAAAGGACGTTGAT  
GGCTAAAATGCTCAGAATTGACAAGCTGGCCGAGCTAGGATGCTGCTGTTTCTGTACCGTCTTGTACATTGTCA  
ACAAGTTCTTGGCTGCTAACCCAGATAAGGTTCAGAAAGTTTATGAGTGTGTTAAGAGAGCTACTGACTATGTT  
CAAAAGCCAGCTGAGGCATACGCTGATTTCAATTGAGATTAAGCCACTTATGGGAACCTCTGAACTACAAGATTT  
CAAAGAAGTTATGCCTACTTTTCTGAGTCTCTTACAACGTCACAGAGACTGGAACAGGTCATGCCTACGGTA  
AGAGATTGATGGTTTGGCAGAAGACTTTAAGGCCAATTATACCAACGAGTACCTATCTTGGCCAGAACCAAGG  
GTCTCGGATCCACTGGAGGCTCAAAGAAGTGAACATCCACCAAGAACAATGCAAGTGAATCCATCTTCAAAG  
ACTGGCTCTCACTGGTCTTTAG
```

The nucleotide sequence of the PAS_chr4_0064 in GS115 strain:

>gi|254574547:c129335-128559 Pichia pastoris GS115 chromosome 4, complete sequence

```
ATGTCAGGAATAAATTTGCAGATTAACGATGAATTTAAGCCTATCATCTCCCCTTTTTATGGAGTTATCATACGG  
TCTCTTTGATGCCAATGGTATCAAAATCAATTTTCAAGCAAGGAAAGCTCCAGTGGAGAATTCAAAGACAATGG  
AATTCACAATTTATTGAGTCTCAAATAACCCCTGTTGGAAGGGTAAACTTAGAGCTCCTTTACCTTGATAACT  
AATTTATCTGAAAAATTAATAATAGGTGCTCCTATTGTTGGTGGGCTGGACACTAATCAATTTACGGCGTTAT  
AGACAAATTCATAATTTTGGAGCTCTTGATTTTGGACACTCAAACACTTGTTCGAGGCATGTATTCTAAACA  
ATGCCATTGTGGGGTTTGGACCTCCTTCAAAAAATTCAGCTTGAAGATCATGGCATGCGACTTAAGAGTTCTA  
GTAAGAACAAGGTTGTTAGATTCCAAAGTTGACTTGGTCTGTGGGCTGAGAAATATCCTGACCGATGAAGCT  
GATTAACCTCTGCTTGAATTTGCGCAATTTGAACAAATATTATCAGACCCGTTAGATACACTTGGAGATATT  
TTGACCGAAGCTGATTTGACTTTCAACTAAGAATTTAGAAAGAAGTTTCCCTACTTTGTGAAATCAAAAAT  
AAACACTCTGCAGTCATTGACTACCAACAATTCAAATCATTCCAAGATGAAGCAGGTGGATTTCACCGCATAT  
TGAATAA
```

The region between two genes, i.e., PAS_chr3_0064 and PAS_chr4_0065 in the sequenced strand of chromosome 4:

```
5' ...GTCTTTGTAATAGTTATAGTTCAGAACTGGAATTGAGCTCAAAAACTGGAATCGAGCGGATATTTGAAGAT  
TGATGCCTTACTCATGAATTGATTGATAAGAGCTCCGCTGATTCCTGTCATGATTACCCCTCTCCTACCCGATT  
TGGGACTTTTCTTCAGTCTTGGGGACTTTTTTCATATGACTTGACCTTGCTTTCCCAATAGGGAAGGACTCACC  
ATGGATGATTAAGTTTGGATTACTCGTTTAGGAAATAGTAGCCATGAATCAATTTGAATCATACCATCATGAAAT  
GGTTAGGCTGTAATGCCTCAAAAATGGCTCTTGGGCTGGATTTTGGGTATTGGAATGTTGGTAGCAATGGTAT  
AAAAGGCCATTTGATTTCACTTTTTTGTCTTCATACTTTACTCTTCTCAACTTTGAAACTTCAATAAATCATC...  
3'
```

The putative promoter region represented in Figure 4. 17 is the above sequence.

M.4. GAP promoter

5' ...AGATCTTTTTGTAGAAATGCTTGGTGTCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTG
GCTCCGTTGCAACTCCGAACGACCTGCTGGCAACGTAAAATTCTCCGGGGTAAAACCTAAAT
GTGGAGTAATGGAACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACCGCCCGTTACCGT
CCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTT
CCCAGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGG
AAAAGTCCCGGCCGTCGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAA
ACCACCAGAATCGAATATAAAAAGGCGAACACCTTCCCAATTTGGTTTCTCCTGACCCAA
AGACTTTAAATTTAATTTATTTGTCCCTATTTCAATCAATTGAACAACCTAT...3'

M.5. SDH promoter

5' ...AAGTTGTATATTATTAATGGCGGGGCAGAATGCTCTCAGTCTCAACTTTTGATGATTG
CGCGGGATTAAAAAATCATGCACACCGGAAAAACTTATCCCAAATGTATGTACCCACGTA
ATCCGTTGTAGTTCCACCAAAGAATTTAAAACGTACTTCCCGCCCAATAATCTGTTCCCTT
TCCCGCTTCATTACACTCACTATTATCCAACATG...3'

N. Nucleotide sequence of constructs of selected promoters.

The deigned promoter constructs include cleavage sites of *NsiI* and *EcoRI* restriction enzymes in addition to the selected “putative promoter region”. Green highlights are the selected promoters; blue highlights are the cleavage sites of the *NsiI* and *EcoRI* restriction enzymes. Yellow highlights are α -MF secretion signal peptide.

N.1. *NsiI* restriction site + *PDC* promoter + α -MF + *EcoRI* restriction site

```
5' ...ATGCATGAGATCGGGACAAGCACACGATTACCCAATCACTTGATATGCACCAATTTGTTCCGTTGTTTATGCCA  
TATTTACCGAATTTTCTTCCCAGGTTTTTCCGAATGGACATCTGTAGTCCACTTTTTGGTTATCATAATCGTCCCAC  
AAGTCGTGGATTTAACCAGAACCTAGTAATTTAAGTTCGCTATTAATCACTCAGAATGGTCTCACCTTGCTATTTGG  
CCAGGCTCGGAGTCGCCAGCTACCACCTCAGAGGCTACATAGACCTCCCAATGTCATCTCCTCAGTGGCGCTTTCAA  
TCTCGTGTCTTTTCCGTTAAAACTCCGTTTCGTTTACCCTATACTGCCCTGGTTGTGCAGCTCTTACCCTTCGGG  
CCGCTACTATCCGTAGTGGTCGAGCCGCATCAATATCAGGTTGAAATAGAATAACTCCCTACAAAAGCCGCACGCAA  
CCATCAAATCTATATAAGGAACCTCAAATATCTAGCAACATCTTTCAATTTACTACAACATATTCGTTAATCATCA  
ATCAATTAGCTAGTACACAACAATGAGATTTCCCTCAATTTTTACTGCTGTTTTATTCCGAGCATCCTCCGCATTAG  
CTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAATTTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTA  
GAAGGGGATTTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTATTGTTTATAAATACTACTAT  
TGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTC...3'
```

N.2. *NsiI* restriction site + *PYRK* promoter + α -MF + *EcoRI* restriction site

```
5' ...ATGCATGAGATCTTCAGTGTGCGGGATACTGTATTCCGCTCGGGGTTCTAAAGAAATGTTTAAACTAAACCAA  
ATCGGATCAGAGGTTCCGTACGTTTTTCACATTCAAGGATGAGGGTTTTCCACGAGTGAACATACTCCTCGGCTCC  
CACCATCATTTGCGGAATGAAACCTTTTGTGCTGAGATTGTATAGGGCGTGGGGACGGACGCTTCTTAACCGTTCCC  
CTAGAATGTCGTCCCTGATCAAAATTTAATGGCATCCAACCTTTGCTGTAATAGGTATATATAACCTAGCAGGCGAC  
CGTTCATGTACAGTAAATGTTTTAGACTTTTTTTAACTGAAATCAATCCAATGAGATTTCCCTCAATTTTTACTG  
CTGTTTTATTCCGAGCATCCTCCGATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAATTTCCG  
GCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAA  
TAACGGGTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAG  
AGGCTGAAGCTGAATTC...3'
```

N.3. *NsiI* restriction site + *THI3* promoter + α -MF + *EcoRI* restriction site

```
5' ...ATGCATGAGATCGTCTTTGTAATAGTTATAGTTTCAGAACTGGAATTGAGCTCAAAAACTGGAATCGAGCGG  
ATATTTGAAGATTGATGCCTTACTCATGAATTGATTGATAAGAGCTCCGTGATTCACTCTGTCAATGATTACCCCTC  
TCTACCCGATTTGGGACTTTTTCTTCAGTCTTTGGGACTTTTTTTCATATGACTTGACCTTGCTTTCCCAATAGGG  
AAGGACTCAACCATGGATGATTAAGTTTGGATTACTCGTTTAGGAAATAGTAGCCATGAATCAATTTGAATCATACC  
ATCATGAAATAGGGTTAGGCTGTAATGCCTCAAAAATGGCTCTTGAGGCTGGATTTTTGGGTATTGGAATGTTGGT  
AGCAATTGGTATAAAAGGCCATTTGTATTTCACTTTTTTGTCTTCATACTTTACTCTTCTCAACTTTGGAACTTC  
AATAAATCATCATGAGATTTCCCTCAATTTTTACTGCTGTTTTATTCCGAGCATCCTCCGCATTAGCTGCTCCAGTC  
AAGACTACAACAGAAGATGAAACGGCACAATTTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTT  
CGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTATTGTTTATAAATACTACTATTGCCAGCATTG  
CTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTC...3'
```

P. Typical results of single-digestion (with NsiI restriction enzyme) prior to Pichia transfection.

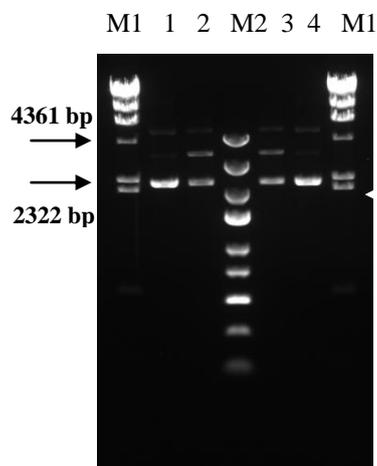


Figure P.1. Single-digestion results of pGAPZ23A::hGH and pGAPZ24A::hGH plasmids. M1: DNA marker λ DNA/HindIII, 1: circular pGAPZ23A::hGH, 2: 1/3 dilution of the eluted (purified) single-digested pGAPZ23A::hGH, M2: GeneRuler™ express DNA ladder, 3: 1/3 dilution of the eluted (purified) single-digested pGAPZ24A::hGH, 4: circular pGAPZ24A::hGH.

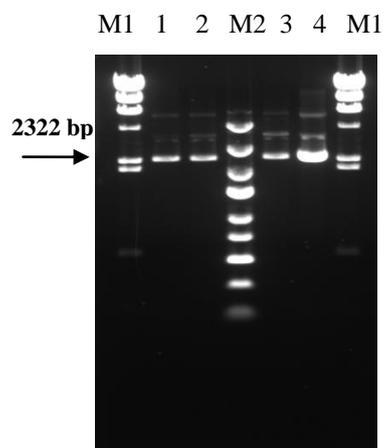


Figure P.2. Single-digestion results of pGAPZ α A::hGH and pPDCZ α A::hGH. M1: DNA marker λ DNA/HindIII, 1: circular pGAPZ α A::hGH, 2: 1/3 dilution of the eluted (purified) single-digested pGAPZ α A::hGH, M2: GeneRuler™ express DNA ladder, 3: 1/3 dilution of the eluted (purified) single-digested pPDCZ α A::hGH, 4: circular pPDCZ α A::hGH.

Q. Tabulated results of the qPCRs related with copy number determination.

Table Q.1. The processed data related with SPs group in order to calculate average copy numbers.

Sample	CT	Copy No.	Avg. copy	Variance	SD (CV%)	Sample	CT	Copy No.	Avg. copy	Variance	SD (CV%)	Absolute copy number
109-ARG	19,80	111645	123594	6.48E+07	8052	109-hGH	18,63	68201	69699.167	2.68E+07	5181	
109-ARG	19,54	132514			7	109-hGH	18,57	71140				
109-ARG	19,35	128969				109-hGH	18,39	64245				0.56
109-ARG	19,39	125123				109-hGH	18,40	63854				
109-ARG	19,94	127272				109-hGH	18,69	76235				
109-ARG	20,08	116041				109-hGH	18,72	74520				
4-ARG	20,62	65217	76642.833	6.11E+07	7818	4-hGH	19,31	44046	44917.833	5.73E+06	2393	
4-ARG	20,54	68962			10	4-hGH	19,26	45473				
4-ARG	20,06	80615				4-hGH	19,00	43524				0.59
4-ARG	20,05	81386				4-hGH	19,07	41450				
4-ARG	20,71	78387				4-hGH	19,43	47817				
4-ARG	20,57	85290				4-hGH	19,45	47197				
33-ARG	20,28	81335	91088.833	7.62E+07	8726	33-hGH	18,96	55381	57873.5	3.76E+07	6131	
33-ARG	20,29	81225			10	33-hGH	18,99	54204				
33-ARG	19,70	101997				33-hGH	18,68	53302				0.64
33-ARG	19,78	97087				33-hGH	18,64	54885				
33-ARG	20,38	96297				33-hGH	19,07	60027				
33-ARG	20,51	88592				33-hGH	18,84	69442				

Table Q.1. The processed data related with SPs group in order to calculate average copy numbers (Continued).

Sample	CT	Copy No.	Avg. copy	Variance	SD (CV%)	Sample	CT	Copy No.	Avg. copy	Variance	SD (CV%)	Absolute copy number
53-ARG	19,25	160598	185172.33	2.76E+08	16608	53-hGH	18,01	102375	108029.83	1.08E+08	10391	
53-ARG	19,18	168383			9	53-hGH	17,94	106535			10	
53-ARG	18,75	190355				53-hGH	17,76	96621				0.58
53-ARG	18,73	193367				53-hGH	17,69	101089				
53-ARG	19,22	201120				53-hGH	17,94	121824				
53-ARG	19,25	197211				53-hGH	17,97	119735				
64-ARG	19,92	103570	110184.5	2.74E+08	16557	64-hGH	18,45	76827	73838.5	1.85E+07	4307	
64-ARG	20,00	98033			15	64-hGH	18,53	73074			6	
64-ARG	19,37	126641				64-hGH	18,29	68895				0.67
64-ARG	19,36	127828				64-hGH	18,29	68779				
64-ARG	20,54	87033				64-hGH	18,68	76517				
64-ARG	20,06	118002				64-hGH	18,63	78939				
81-ARG	20,16	88164	99104.667	1.09E+08	10461	81-hGH	18,91	57239	59336	1.67E+07	4086	
81-ARG	20,20	86193			11	81-hGH	18,85	59341			7	
81-ARG	19,65	105600				81-hGH	18,64	54803				0.60
81-ARG	19,62	107440				81-hGH	18,60	56150				
81-ARG	20,38	96255				81-hGH	18,97	63872				
81-ARG	20,16	110976				81-hGH	18,95	64611				

Table Q.2. The processed data related with promoters group in order to calculate average copy numbers.

Sample	CT	Copy No.	Avg. Copy	Variance	SD (CV%)	Sample	CT	Copy No.	Avg. Copy	Variance	SD (CV%)	Absolute Copy Number
109-ARG	16,65	112203	110123.5	3.61E+07	6009.1	109-hGH	15,56	65793	59471.1667	1.53E+07	3917.0	
109-ARG	16,60	116001			5.5	109-hGH	15,80	56316			6.6	
109-ARG	21,04	104197				109-hGH	19,69	61896				0.54
109-ARG	21,05	103900				109-hGH	19,73	60404				
109-ARG	20,47	110358				109-hGH	19,09	56515				
109-ARG	20,42	114082				109-hGH	19,11	55903				
120-ARG	16,09	163788	171269.8333	3.45E+07	5877.3	120-hGH	15,03	92921	100151.5	4.74E+07	6882.6	
120-ARG	15,98	176995			3.4	120-hGH	14,78	109397			6.9	
120-ARG	20,34	166096				120-hGH	18,91	103542				0.58
120-ARG	20,33	166617				120-hGH	18,91	103814				
120-ARG	19,70	180592				120-hGH	18,23	98674				
120-ARG	19,76	173531				120-hGH	18,33	92561				
144-ARG	16,36	136907	121699.1667	1.53E+08	12376.3	144-hGH	15,46	70293	69793.8333	1.03E+07	3215.1	
144-ARG	16,41	132203			10.2	144-hGH	15,33	76523			4.6	
144-ARG	20,91	114017				144-hGH	19,52	69508				0.58
144-ARG	20,92	112776				144-hGH	19,49	70711				
144-ARG	20,32	121659				144-hGH	18,84	66391				
144-ARG	20,44	112633				144-hGH	18,87	65337				
160-ARG	17,14	80733	67209	8.50E+07	9221.5	160-hGH	16,20	43370	38057.5	1.91E+07	4370.0	
160-ARG	17,30	72155			13.7	160-hGH	16,12	45719			11.5	
160-ARG	21,85	61127				160-hGH	20,41	38672				0.57
160-ARG	21,82	62253				160-hGH	20,51	36102				
160-ARG	21,26	66740				160-hGH	19,89	33796				
160-ARG	21,42	60246				160-hGH	20,04	30686				

R. Selected qPCR reports.

R.1. Selected run for promoter group

ARG4

Quantitation Information

Threshold	0.0468
Left Threshold	1.000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 [^] (-0.294*CT + 9.947)
Standard Curve (2)	CT = -3,401*log(conc) + 33.826
Reaction efficiency (*)	0.96815 (* = 10 [^] (-1/m) - 1)
M	-3.40069
B	33.82575
R Value	0.99979
R ² Value	0.99958
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

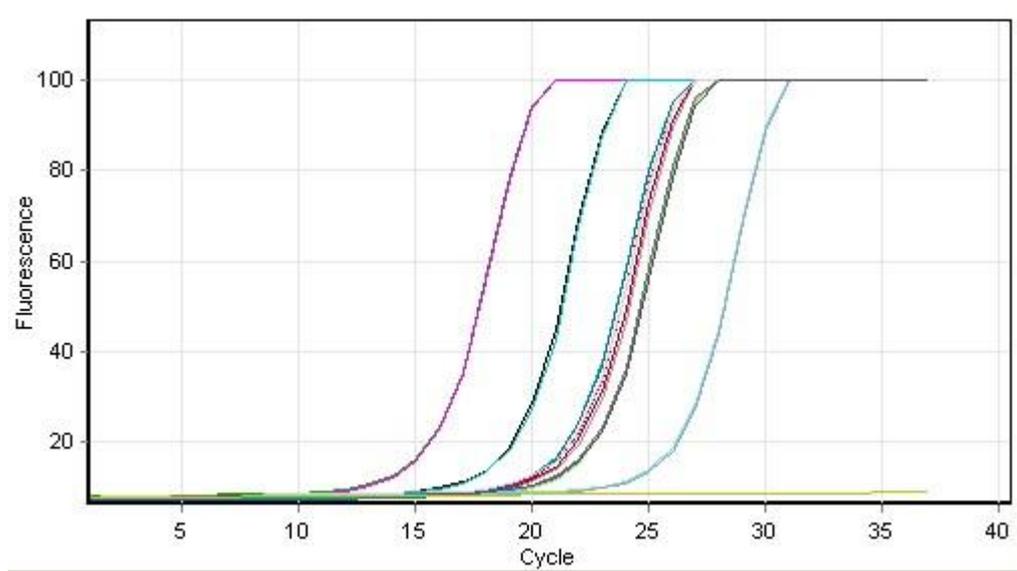
Messages

Operator skipped end of cycle for Cycling @ Repeat 38

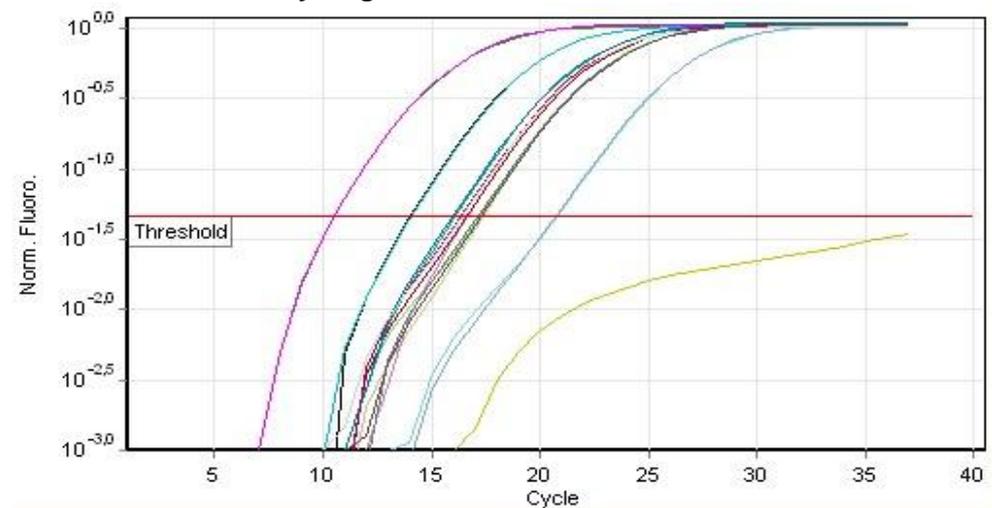
Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

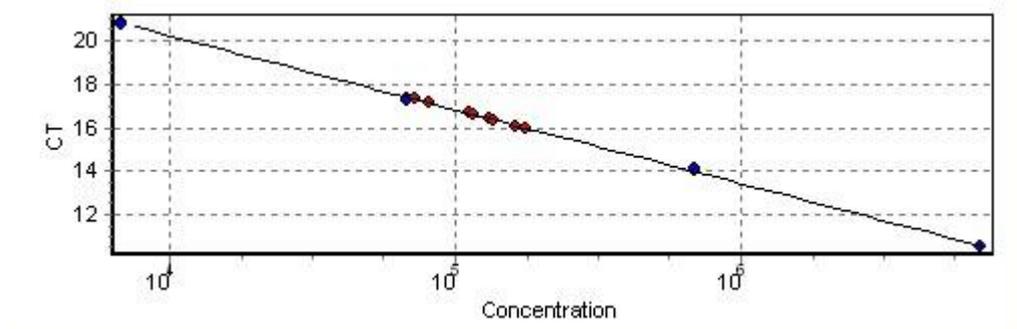
Raw Data for Cycling A.Green



Quantitation Data for Cycling A.Green



Standard Curve

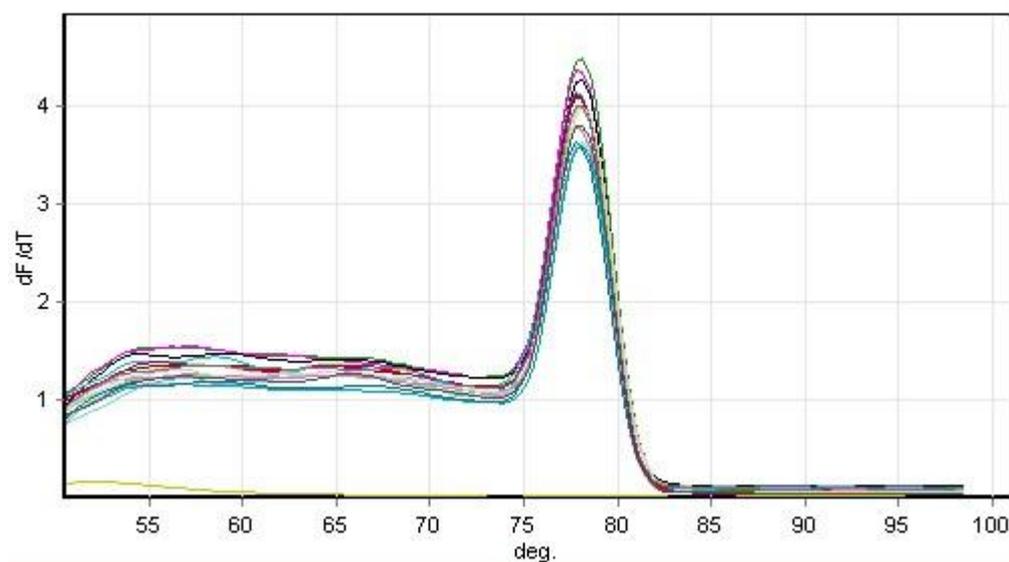


9		ARG-1	Standard	10.55	6820000	6972676	2.2%
10		ARG-1	Standard	10.53	6820000	7062725	3.6%
11		ARG-2	Standard	14.06	682000	650728	4.6%
12		ARG-2	Standard	14.12	682000	624641	8.4%
13		ARG-3	Standard	17.39	68200	68291	0.1%
14		ARG-3	Standard	17.24	68200	75141	10.2%
15		ARG-4	Standard	20.84	6820	6596	3.3%
16		ARG-4	Standard	20.77	6820	6908	1.3%
19		109-ARG	Unknown	16.65		112203	
20		109-ARG	Unknown	16.60		116001	
23		120-ARG	Unknown	16.09		163788	
24		120-ARG	Unknown	15.98		176995	
27		144-ARG	Unknown	16.36		136907	
28		144-ARG	Unknown	16.41		132203	
31		160-ARG	Unknown	17.14		80733	
32		160-ARG	Unknown	17.30		72155	
34		NTC-ARG	NTC				

Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

Melt Data for Melt A.Green



9	■	ARG-1	56.2	62.5	65.5	78.0	88.7	95.3	
10	■	ARG-1	57.3	61.2	78.0	90.3	95.0		
11	■	ARG-2	54.3	59.0	64.5	66.2	78.0	84.5	94.0
12	■	ARG-2	54.8	58.5	64.5	78.0	84.5	93.0	
13	■	ARG-3	57.2	64.5	78.0	88.5	96.0		
14	■	ARG-3	58.2	63.8	78.0	88.5	95.7		
15	■	ARG-4	56.0	63.5	78.0	90.5	95.0		
16	■	ARG-4	57.0	64.2	68.0	78.0	85.8	91.5	96.3
19	■	109-ARG	54.0	57.5	65.0	78.0	88.5	94.7	
20	■	109-ARG	59.2	64.5	78.0	87.5	90.5	97.2	
23	■	120-ARG	56.5	64.7	78.0	88.2	95.5		
24	■	120-ARG	54.0	58.5	63.5	66.7	78.0	89.0	95.3
27	■	144-ARG	55.5	58.2	65.5	78.0	94.3		
28	■	144-ARG	56.5	61.7	65.3	78.2	88.5	95.3	
31	■	160-ARG	57.7	64.0	68.0	78.0	86.2	92.0	
32	■	160-ARG	56.8	65.8	78.0	92.8			
34	■	NTC-ARG	52.2	70.3	76.8	80.5	87.5	93.3	

hGH

Quantitation Information

Threshold	0.0468
Left Threshold	1.000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 [^] (-0.283*CT + 9.217)
Standard Curve (2)	CT = -3.537*log(conc) + 32.606
Reaction efficiency (*)	0.91732 (* = 10 [^] (-1/m) - 1)
M	-3.53738
B	32.60556
R Value	0.99977
R ² Value	0.99955
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

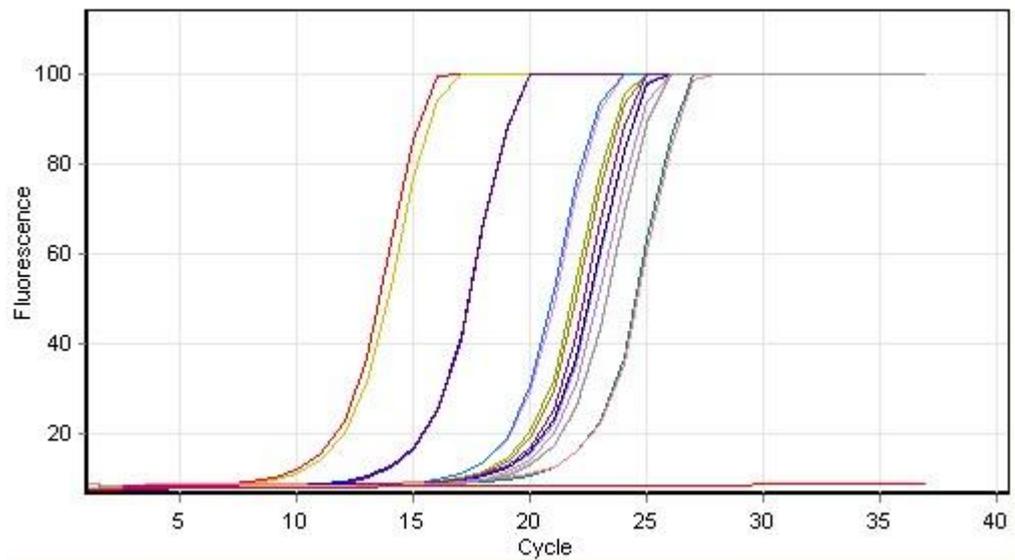
Messages

Operator skipped end of cycle for Cycling @ Repeat 38

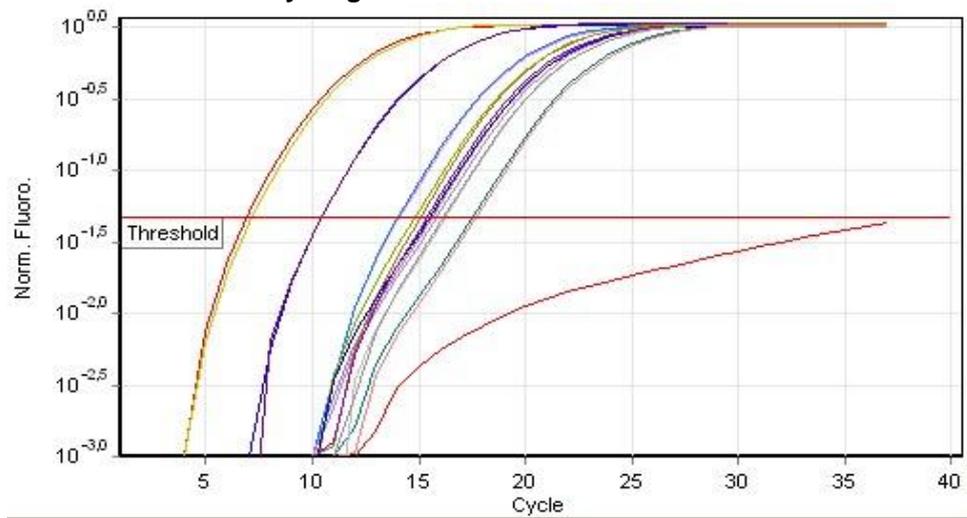
Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

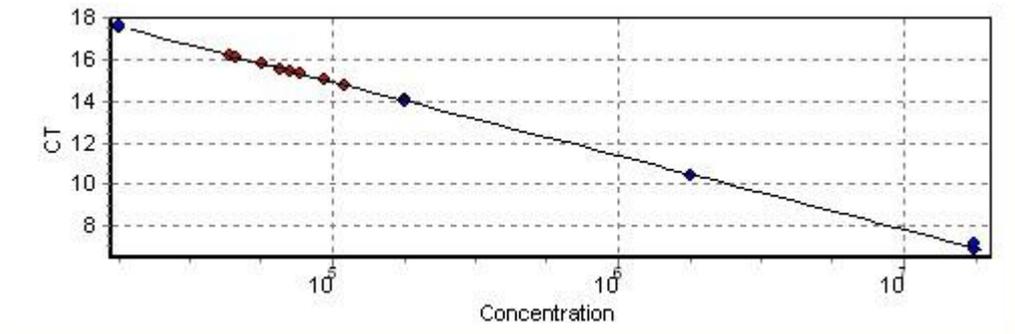
Raw Data for Cycling A.Green



Quantitation Data for Cycling A.Green



Standard Curve

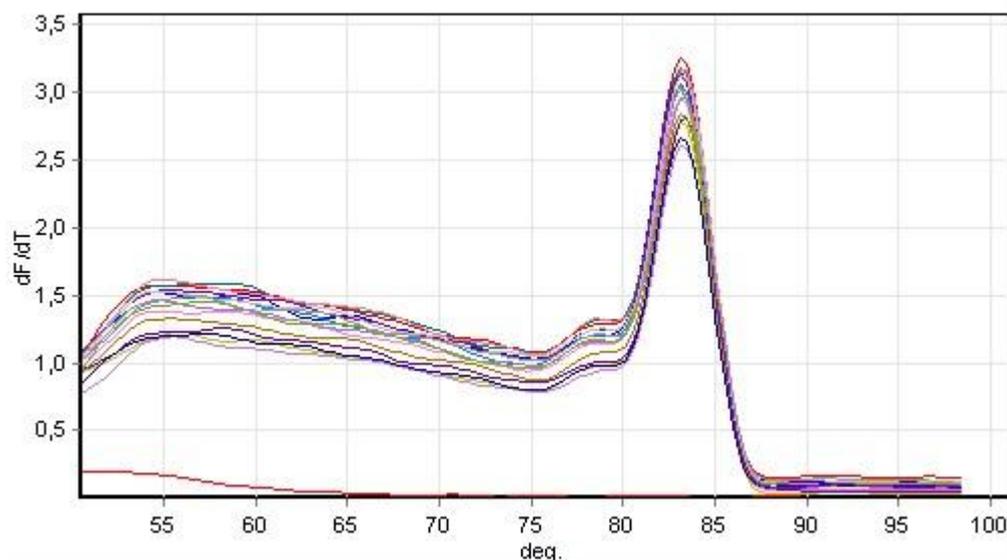


1		hGH-1	Standard	6.87	17820000	18890992	6.0%
2		hGH-1	Standard	7.12	17820000	16040208	10.0%
3		hGH-2	Standard	10.44	1782000	1846095	3.6%
4		hGH-2	Standard	10.41	1782000	1876376	5.3%
5		hGH-3	Standard	14.09	178200	171875	3.5%
6		hGH-3	Standard	14.03	178200	178706	0.3%
7		hGH-4	Standard	17.50	17820	18601	4.4%
8		hGH-4	Standard	17.64	17820	16956	4.8%
17		109-hGH	Unknown	15.56		65793	
18		109-hGH	Unknown	15.80		56316	
21		120-hGH	Unknown	15.03		92921	
22		120-hGH	Unknown	14.78		109397	
25		144-hGH	Unknown	15.46		70293	
26		144-hGH	Unknown	15.33		76523	
29		160-hGH	Unknown	16.20		43370	
30		160-hGH	Unknown	16.12		45719	
33		NTC-hGH	NTC				

Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

Melt Data for Melt A.Green



1	■	hGH-1	54.5	56.0	78.5	83.2	91.3	96.8	
2	■	hGH-1	55.8	59.7	78.8	83.3	88.8	92.0	97.2
3	■	hGH-2	54.7	59.3	65.0	78.5	83.2	92.0	96.5
4	■	hGH-2	55.7	58.3	83.5	88.7	94.0		
5	■	hGH-3	54.7	58.3	78.7	83.2	89.0	90.5	95.5
6	■	hGH-3	57.5	69.2	78.5	83.2	91.0	95.0	
7	■	hGH-4	58.2	78.7	83.2	88.8	92.0		
8	■	hGH-4	55.0	78.7	83.3	88.5	92.2		
17	■	109-hGH	54.8	73.0	78.5	83.2	89.0	92.0	98.0
18	■	109-hGH	55.5	83.3	88.8	95.3			
21	■	120-hGH	55.3	64.2	83.3	91.7	94.3		
22	■	120-hGH	57.2	64.0	78.5	83.0	93.3		
25	■	144-hGH	57.3	78.5	83.2	91.2	95.7		
26	■	144-hGH	55.2	78.7	83.2	92.0			
29	■	160-hGH	54.8	63.2	73.5	78.8	83.3	89.2	95.5
30	■	160-hGH	56.8	61.5	78.5	83.3	91.0	95.0	
33	■	NTC-hGH	51.3	70.5	75.5	80.2	82.8	88.3	95.3

R.2. Selected run for signal peptide group

ARG4

Quantitation Information

Threshold	0.2093
Left Threshold	1.000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 [^] (-0.285*CT + 10.688)
Standard Curve (2)	CT = -3.511*log(conc) + 37.522
Reaction efficiency (*)	0.92687 (* = 10 [^] (-1/m) - 1)
M	-3.51059
B	37.52239
R Value	0.99997
R ² Value	0.99993
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

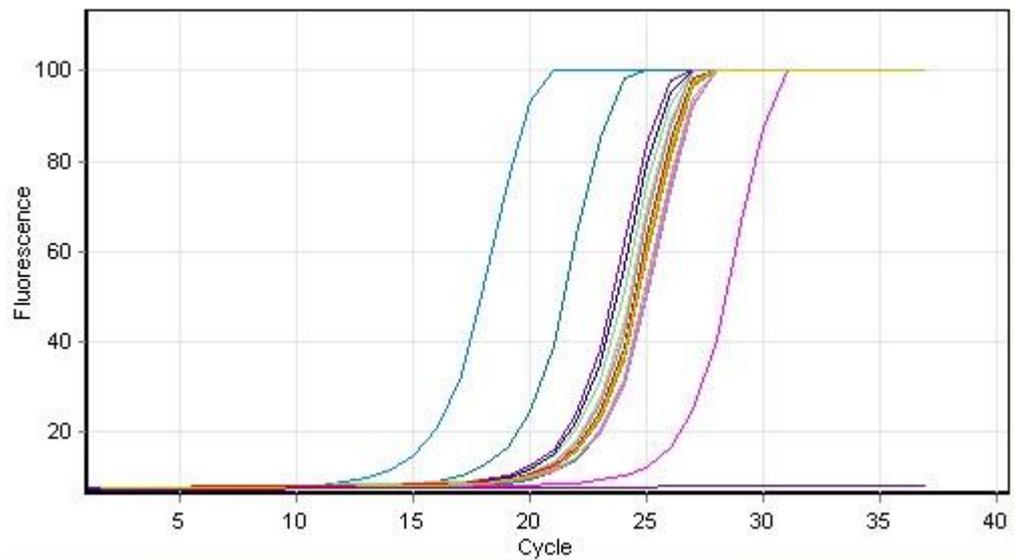
Messages

Operator skipped end of cycle for Cycling @ Repeat 38

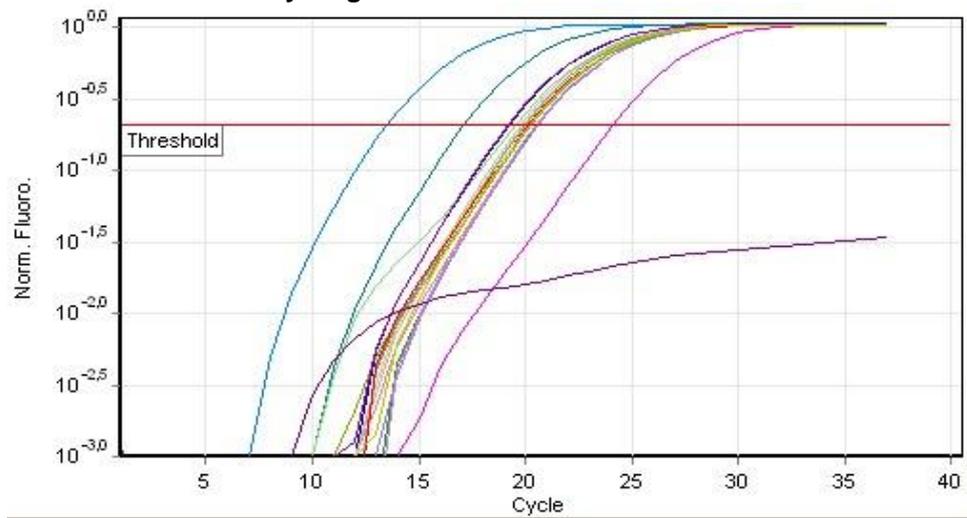
Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

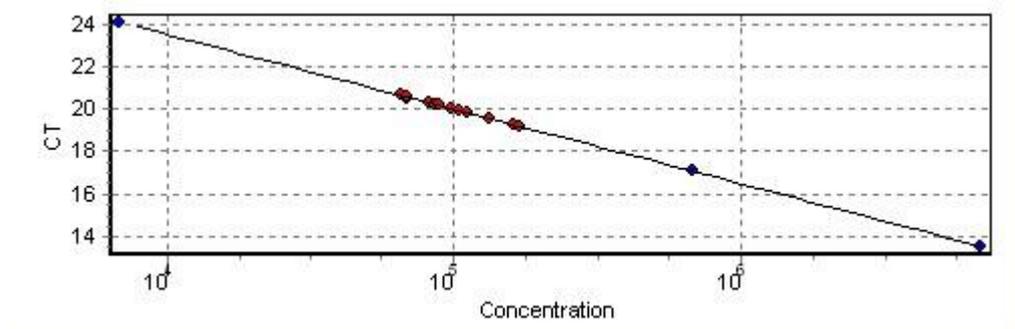
Raw Data for Cycling A.Green



Quantitation Data for Cycling A.Green



Standard Curve

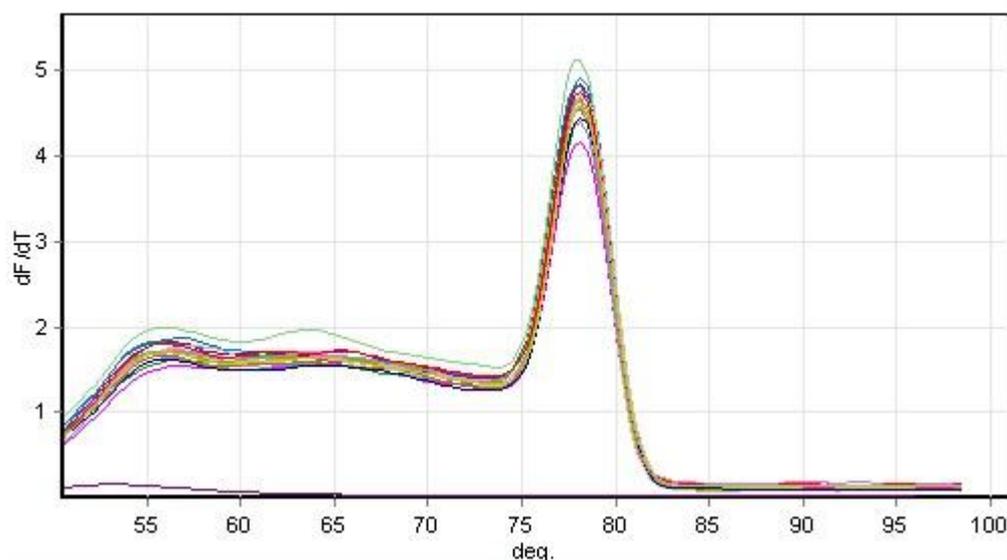


6		ARG-1	Standard	13.51	6820000	6923868	1.5%
7		ARG-2	Standard	17.08	682000	666701	2.2%
8		ARG-3	Standard	20.52	68200	69762	2.3%
9		ARG-3	Standard	20.59	68200	66675	2.2%
10		ARG-4	Standard	24.05	6820	6872	0.8%
13		109-ARG	Unknown	19.80		111645	
14		109-ARG	Unknown	19.54		132514	
17		4-ARG	Unknown	20.62		65217	
18		4-ARG	Unknown	20.54		68962	
21		33-ARG	Unknown	20.28		81335	
22		33-ARG	Unknown	20.29		81225	
25		53-ARG	Unknown	19.25		160598	
26		53-ARG	Unknown	19.18		168383	
29		64-ARG	Unknown	19.92		103570	
30		64-ARG	Unknown	20.00		98033	
33		81-ARG	Unknown	20.16		88164	
34		81-ARG	Unknown	20.20		86193	
36		NTC-ARG	NTC				

Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

Melt Data for Melt A.Green



6		ARG-1	55.5	78.2	91.7	96.5			
7		ARG-2	57.0	62.2	78.0	85.2	92.5	95.2	
8		ARG-3	56.3	63.7	78.0	89.0	93.0		
9		ARG-3	56.5	64.3	78.2	89.8	94.3		
10		ARG-4	56.5	64.3	78.0	86.5	90.8		
13		109-ARG	56.2	63.7	78.0	89.3	93.5	96.3	
14		109-ARG	56.0	63.7	78.0	87.5	90.7	93.2	
17		4-ARG	56.2	63.3	78.0	85.5	89.5	93.5	95.3
18		4-ARG	57.0	64.0	78.2	90.0	93.0		
21		33-ARG	56.0	63.5	67.2	78.0	92.0	95.2	
22		33-ARG	55.8	61.5	64.5	78.0	88.7	93.0	
25		53-ARG	56.0	64.2	78.2	83.8	89.7	94.7	
26		53-ARG	56.3	61.2	65.3	78.0	89.0	95.5	
29		64-ARG	57.3	64.7	78.3	90.7			
30		64-ARG	55.8	62.2	78.0	90.7	96.3		
33		81-ARG	56.0	61.8	65.5	78.0	89.8	95.0	
34		81-ARG	56.0	64.3	78.2	86.7	91.0		
36		NTC-ARG	53.2	69.5	75.2	77.3	83.3	88.5	92.5

hGH

Quantitation Information

Threshold	0.2093
Left Threshold	1.000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 [^] (-0.281*CT + 10.067)
Standard Curve (2)	CT = -3.560*log(conc) + 35.845
Reaction efficiency (*)	0.90924 (* = 10 [^] (-1/m) - 1)
M	-3.56049
B	35.84496
R Value	0.9999
R ² Value	0.9998
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

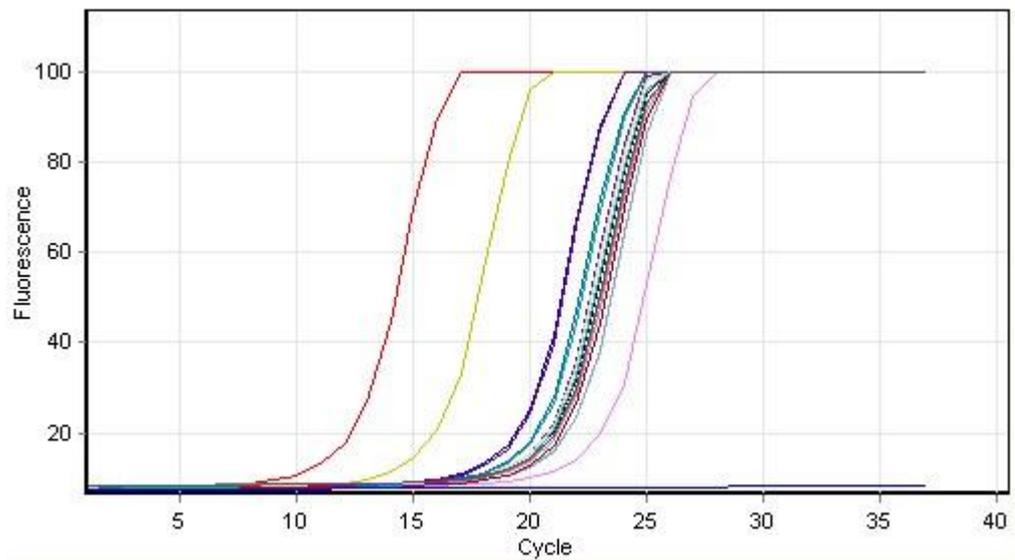
Messages

Operator skipped end of cycle for Cycling @ Repeat 38

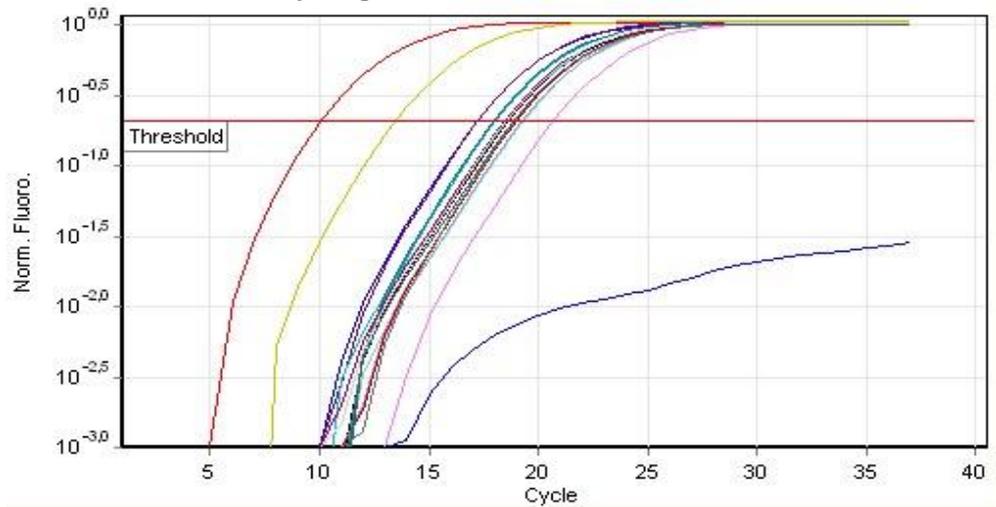
Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

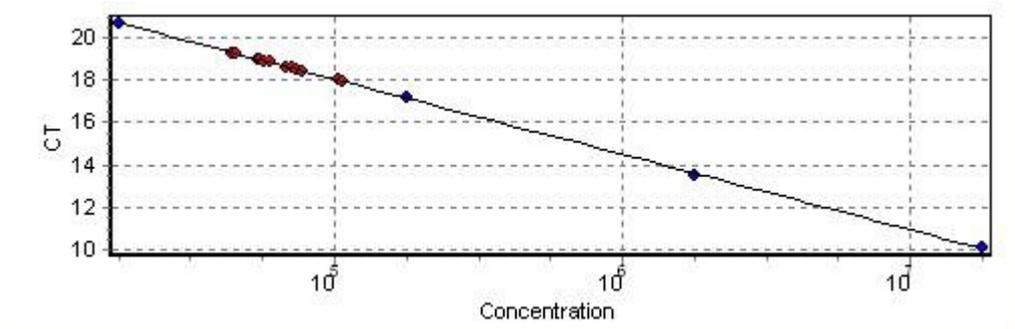
Raw Data for Cycling A.Green



Quantitation Data for Cycling A.Green



Standard Curve

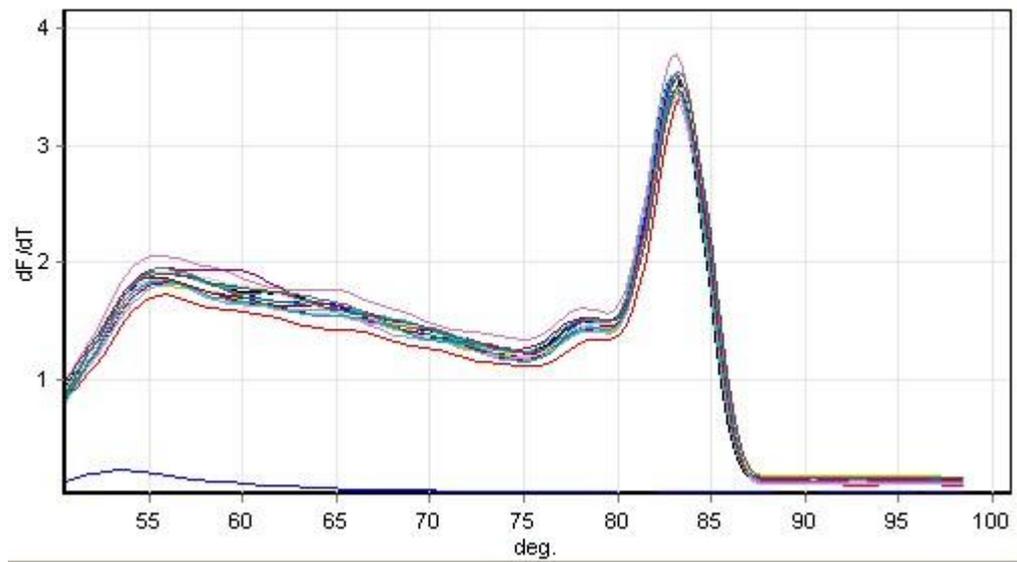


1		hGH-1	Standard	10.07	17820000	17393438	2.4%
2		hGH-2	Standard	13.50	1782000	1890731	6.1%
3		hGH-3	Standard	17.17	178200	176101	1.2%
4		hGH-3	Standard	17.20	178200	172257	3.3%
5		hGH-4	Standard	20.69	17820	18013	1.1%
11		109-hGH	Unknown	18.63		68201	
12		109-hGH	Unknown	18.57		71140	
15		4-hGH	Unknown	19.31		44046	
16		4-hGH	Unknown	19.26		45473	
19		33-hGH	Unknown	18.96		55381	
20		33-hGH	Unknown	18.99		54204	
23		53-hGH	Unknown	18.01		102375	
24		53-hGH	Unknown	17.94		106535	
27		64-hGH	Unknown	18.45		76827	
28		64-hGH	Unknown	18.53		73074	
31		81-hGH	Unknown	18.91		57239	
32		81-hGH	Unknown	18.85		59341	
35		NTC-hGH	NTC				

Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (40 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

Melt data for Melt A.Green



1	■	hGH-1	55.8	78.8	83.5	88.8	90.7	95.5		
2	■	hGH-2	55.8	78.5	83.3	89.0	90.7	94.5		
3	■	hGH-3	56.0	63.7	78.5	83.2	89.0	94.0		
4	■	hGH-3	56.0	78.8	83.2	88.8	94.5			
5	■	hGH-4	55.8	78.5	83.0	88.8	91.0			
11	■	109-hGH	55.5	78.3	83.0	92.0	96.5			
12	■	109-hGH	55.7	78.3	83.0	88.5	93.0			
15	■	4-hGH	55.5	78.5	83.2	88.7	91.5			
16	■	4-hGH	56.0	63.8	78.5	83.2	88.8	95.2		
19	■	33-hGH	55.5	64.5	78.3	83.0	88.5			
20	■	33-hGH	55.3	63.5	78.5	83.2	88.8	90.5		
23	■	53-hGH	55.7	78.5	83.2	90.2	94.5			
24	■	53-hGH	55.7	83.2	88.8	91.0	93.8			
27	■	64-hGH	55.8	59.2	78.2	83.2	88.8	93.7		
28	■	64-hGH	56.0	78.7	83.3	90.5	93.5			
31	■	81-hGH	55.7	61.7	78.7	83.3	89.0	92.8		
32	■	81-hGH	55.8	78.5	83.3	88.8	93.0	96.8		
35	■	NTC-hGH	53.5	69.2	74.7	78.5	83.7	87.0	92.0	94.5

R.3. Selected run for mRNA level measurement

Quantitation Information

Threshold	0.3283
Left Threshold	1.000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 [^] (-0.287*CT + 10.645)
Standard Curve (2)	CT = -3.489*log(conc) + 37.142
Reaction efficiency (*)	0.93463 (* = 10 [^] (-1/m) - 1)
M	-3.48922
B	37.14195
R Value	0.99995
R ² Value	0.9999
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

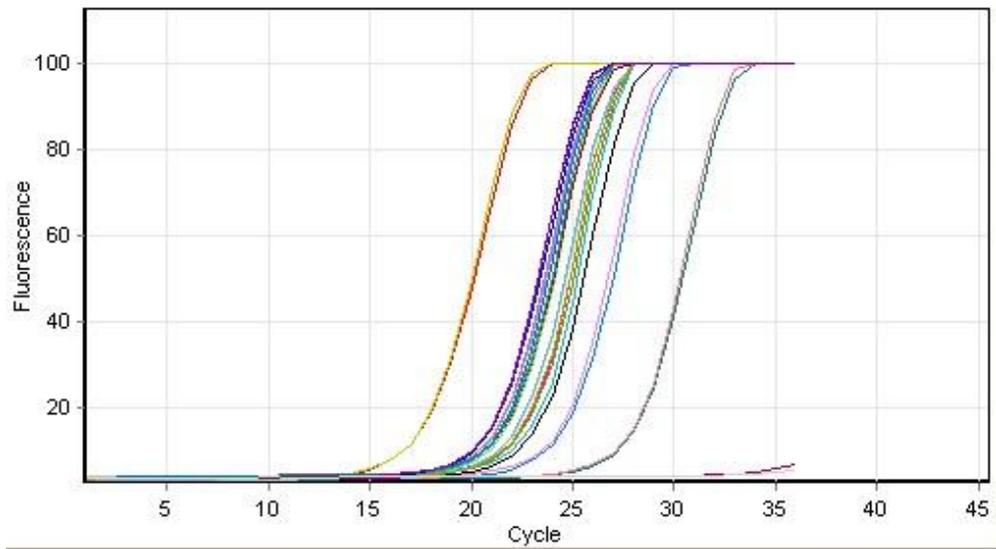
Messages

Operator skipped end of cycle for Cycling @ Repeat 37

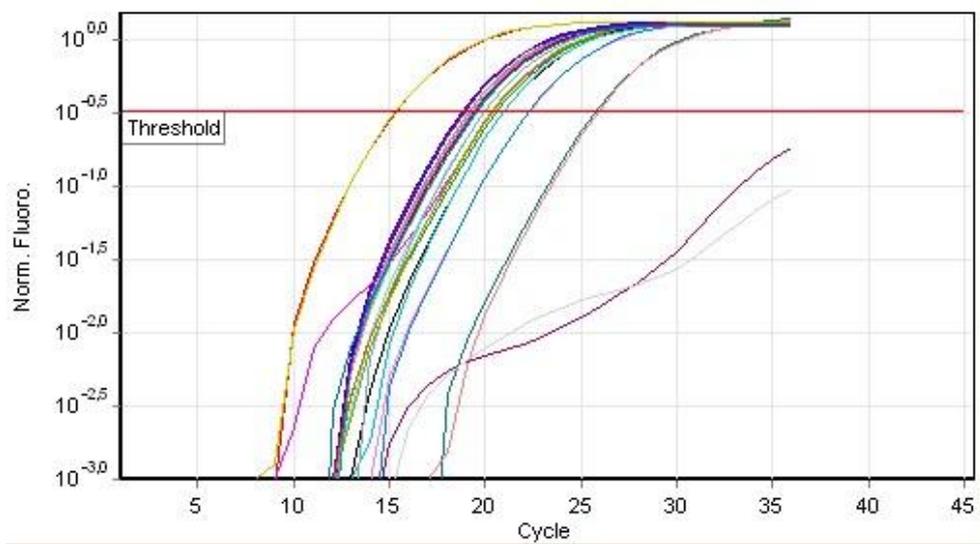
Profile

Hold @ 95°C, 10 min 0 secs	
Cycling (45 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

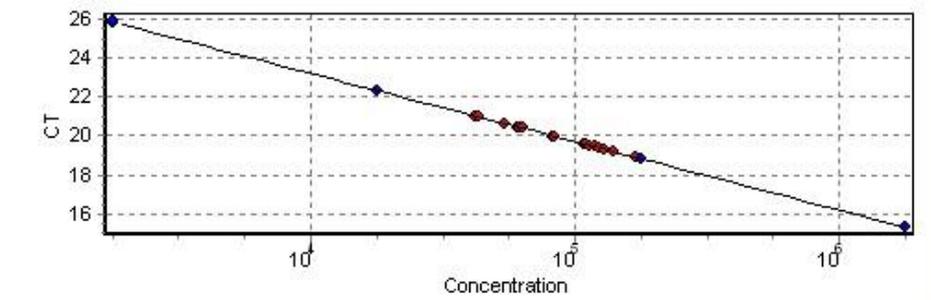
Raw Data for Cycling A.Green



Quantitation Data for Cycling A.Green



Standard Curve

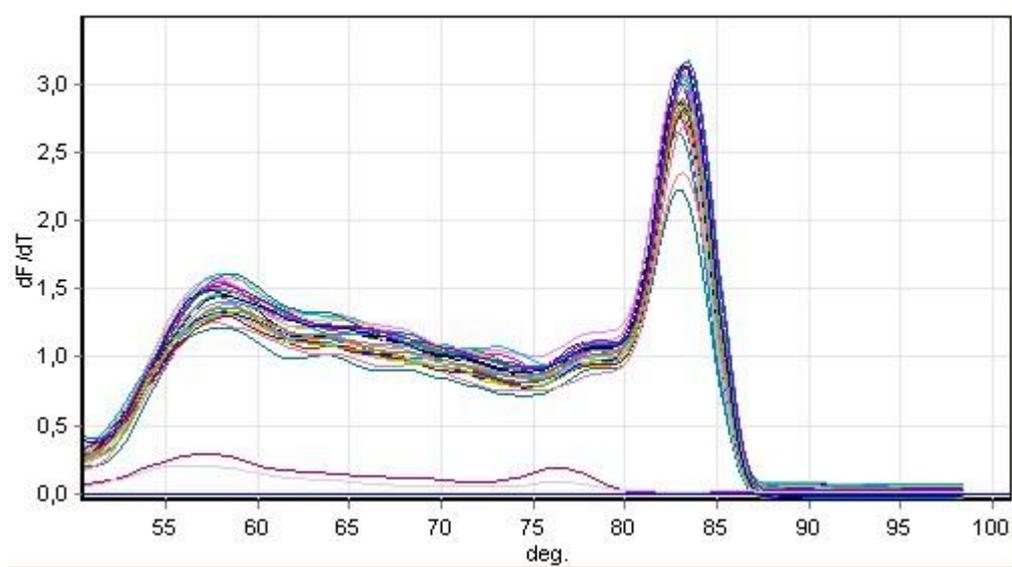


[Redacted Header]							
1		hGH-1	Standard	15.35	1782000	1762062	1.1%
2		hGH-1	Standard	15.32	1782000	1789915	0.4%
3		hGH-2	Standard	18.79	178200	181824	2.0%
4		hGH-2	Standard	18.86	178200	174010	2.4%
5		hGH-3	Standard	22.30	17820	17957	0.8%
6		hGH-3	Standard	22.28	17820	18183	2.0%
7		hGH-4	Standard	25.75	1782	1843	3.4%
8		hGH-4	Standard	25.88	1782	1693	5.0%
9		3	Unknown	20.63		53975	
10		3	Unknown	20.45		60629	
11		8	Unknown	21.00		42245	
12		8	Unknown	20.96		43497	
13		13	Unknown	19.49		114458	
14		13	Unknown	19.56		109067	
15		4	Unknown	19.98		82923	
16		4	Unknown	19.97		83487	
17		9	Unknown	19.17		141186	
18		9	Unknown	19.30		130182	
19		14	Unknown	19.37		123778	
20		14	Unknown	19.43		119365	
21		5	Unknown	20.44		61346	
22		5	Unknown	20.38		63763	
23		10	Unknown	19.55		110028	
24		10	Unknown	19.54		111055	
25		15	Unknown	18.90		169359	
26		15	Unknown	18.88		171446	
27		NTC	NTC				
28		NTC	NTC				

Profile

Profile	
Hold @ 95°C, 10 min 0 secs	
Cycling (45 repeats)	Step 1 @ 95°C, hold 10 secs
	Step 2 @ 55°C, hold 5 secs
	Step 3 @ 72°C, hold 10 secs, acquiring to Cycling A([Green][1][1])
Melt (50-99°C) , hold secs on the 1st step, hold 5 secs on next steps, Melt A([Green][1][1])	

Melt data for Melt A.Green



1		hGH-1	58.7	64.3	83.0	87.8	89.5	94.2	
2		hGH-1	58.0	78.0	83.0	87.7	94.7		
3		hGH-2	58.0	83.0	91.5	97.3			
4		hGH-2	58.3	64.2	78.0	83.0	87.8	89.7	94.3
5		hGH-3	57.5	63.2	72.7	83.0	94.7		
6		hGH-3	58.7	64.2	68.7	83.0	87.8	90.5	95.3
7		hGH-4	58.0	63.5	67.5	78.5	83.0		
8		hGH-4	58.0	62.2	78.5	83.2	89.0	90.7	96.2
9		3	58.0	71.5	78.5	83.3			
10		3	58.0	72.8	78.3	83.2			
11		8	58.2	64.8	78.5	83.0	90.8	94.8	
12		8	58.0	73.0	78.5	83.3	88.7		
13		13	58.5	64.5	78.5	83.5	89.0	97.8	
14		13	58.0	64.5	78.5	83.3	88.8	95.8	
15		4	58.0	63.0	78.2	83.0	87.7	93.2	
16		4	58.3	63.7	78.5	83.2	89.0	95.3	
17		9	58.0	62.8	78.8	83.2	89.0	90.2	96.5
18		9	57.7	66.5	78.5	83.0	90.5	94.0	
19		14	57.5	63.8	68.5	74.5	83.3		
20		14	58.2	78.5	83.2	88.5	95.0		
21		5	58.3	67.8	78.7	83.2	92.3	94.2	
22		5	58.8	66.0	78.5	83.2			
23		10	57.7	65.3	78.5	83.3	88.8	94.3	
24		10	58.5	83.5					
25		15	57.3	78.5	83.2				
26		15	58.0	78.5	83.3	89.0			
27		NTC	57.2	76.5	85.0	89.3	96.5		
28		NTC	56.2	71.5	76.5	86.0	92.7		

S. Complementary SDS-PAGE results related with shake flask bioreactor experiments.

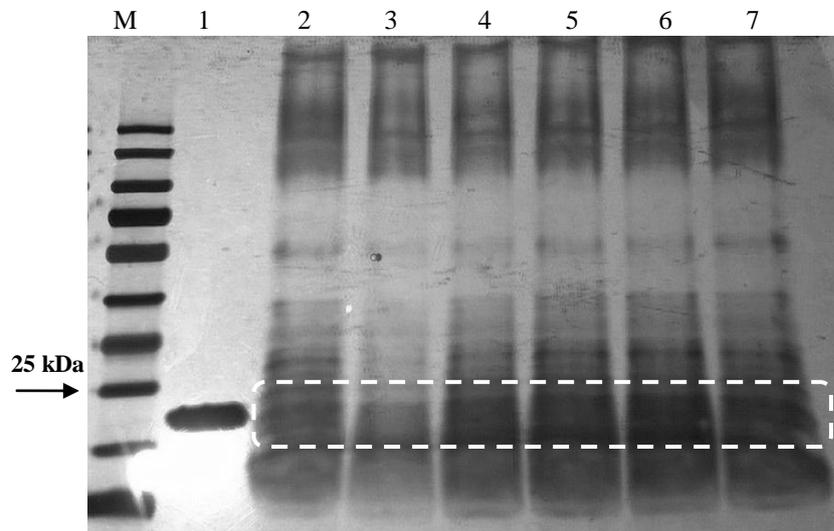


Figure S.1. SDS-PAGE results of the 1st shaker experiment at t= 32 h of production. M: protein marker. 1: hGH standard 0.05 g/L, cell-free supernatant of the r-*P.pastoris* strains of 2: pGAPZ α A::hGH, 3: pGAPZ13A::hGH, 4: pGAPZ23A::hGH, 5: pGAPZ24A::hGH, 6: pGAPZ26A::hGH, 7: pGAPZ34A::hGH.

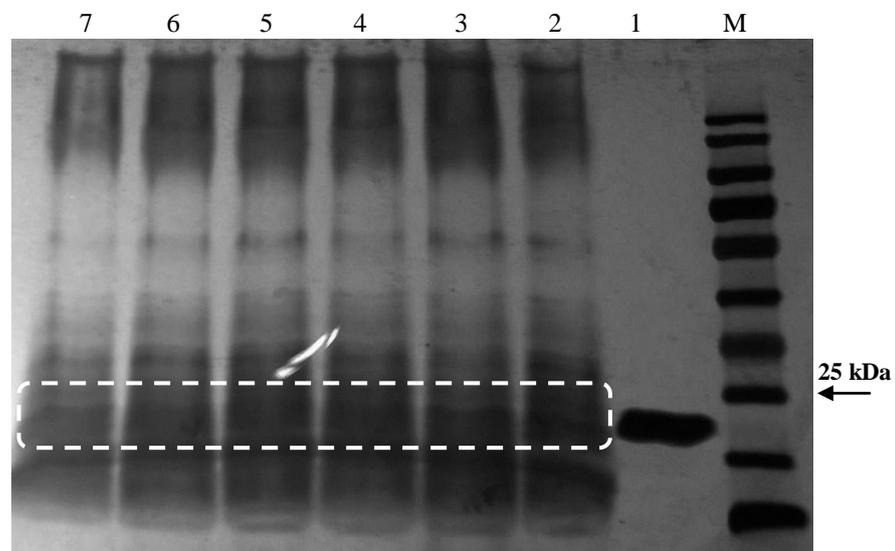


Figure S.2. SDS-PAGE results of the 1st shaker experiment at t=48 h of production. M: protein marker. 1: hGH standard 0.05 g/L, cell-free supernatant of the r-*P.pastoris* strains of 2: pGAPZ α A::hGH, 3: pGAPZ13A::hGH, 4: pGAPZ23A::hGH, 5: pGAPZ24A::hGH, 6: pGAPZ26A::hGH, 7: pGAPZ34A::hGH.

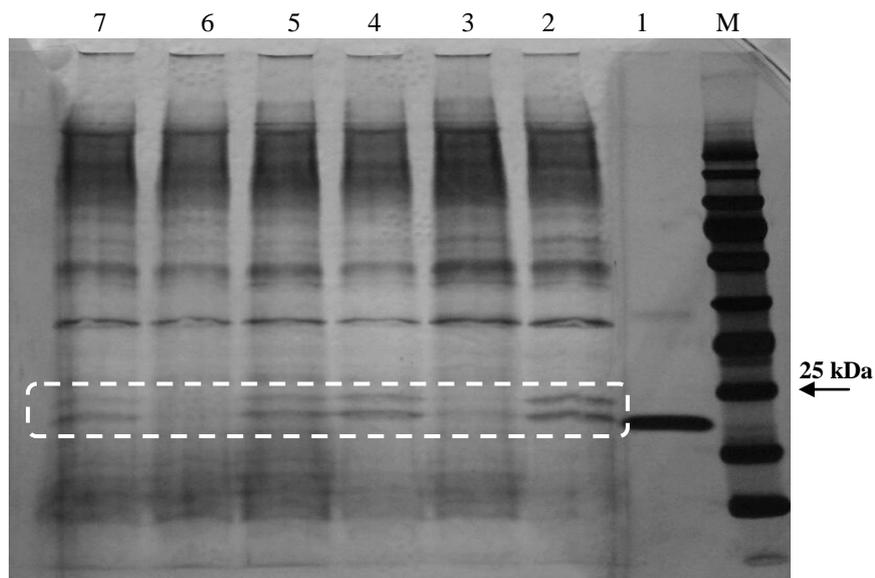


Figure S.3. SDS-PAGE results of the 2nd shaker experiment at t=24 h of production. M: protein marker. 1: hGH standard 0.05 g/L, cell-free supernatant of the r-*P.pastoris* strains of 2: pGAPZ α A::*hGH* , 3: pGAPZ13A::*hGH*, 4: pGAPZ23A::*hGH* , 5: pGAPZ24A::*hGH*, 6: pGAPZ26A::*hGH*, 7: pGAPZ34A::*hGH*.

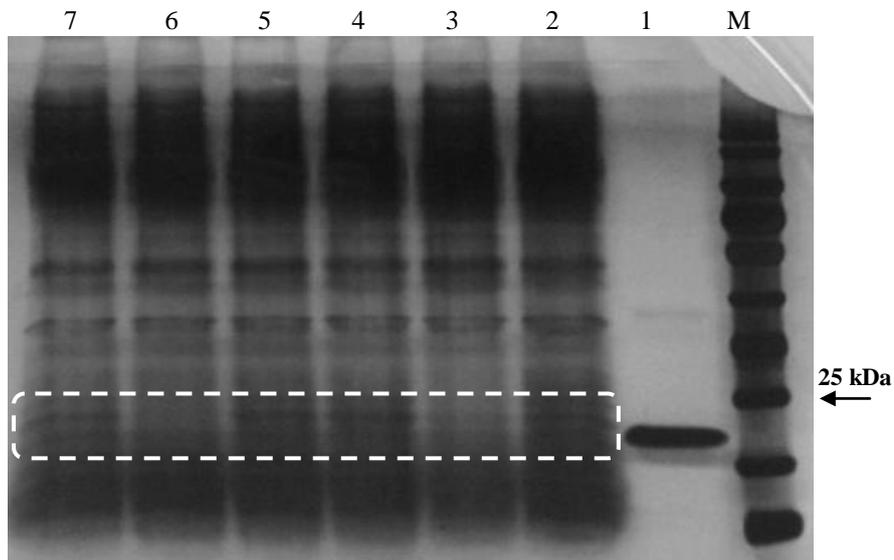


Figure S.4. SDS-PAGE results of the 3rd shaker experiment at t=24 h of production. M: protein marker. 1: hGH standard 0.05 g/L, cell-free supernatant of the r-*P.pastoris* strains of 2: pGAPZ α A::*hGH* , 3: pGAPZ13A::*hGH*, 4: pGAPZ23A::*hGH* , 5: pGAPZ24A::*hGH*, 6: pGAPZ26A::*hGH*, 7: pGAPZ34A::*hGH*.

T. A glimpse on selected experimental techniques.

T.1. Agarose gel electrophoresis

This method is used for separation, visualization and quantification of the different DNA fragments or purification of desired DNA fragments. Electrophoresis is the migration of the dispersed (charged) particles relative to a stationary fluid medium under the exertion of a uniform external electric field. DNA and RNA molecules based on their negative charge of phosphate backbone can easily move from negatively-charged cathode toward positively-charged anode upon applying electric field. However, separation of the molecules in fluid is not possible due to the convective flow brought about by movement of the particles and, thus, by absence of solid medium, the migration will be independent of the molecule/particle length. Therefore, using a gel matrix not only eliminates problems related with convective flow but also provides a polymeric sieve in order to separate particles according to their length by impeding their movement. Agarose gel is one of the most convenient gels which is used in this technique and its inter-connected and randomly-distributed pores provide sieving mechanism (Ogston theory). Larger molecules travel slower than smaller molecules due to the continuous collisions with the gel matrix and, in consequence, separation is achieved. The separated DNA molecules can be visualized under UV light by applying a staining compound; ethidium bromide (EtBr) as an intercalating agent can be used for nucleic acid staining.

The movement of the nucleic acids in porous medium of the gel is affected by several parameters like: gel concentration, size and conformation of DNA, applied voltage, and ionic strength of the electrophoresis buffer. More concentrated gels have a higher resolution power. However, very large DNA molecules cannot be separated by agarose gel electrophoresis due to the breakdown of the available relationship between the rate of DNA movement and the number of base pairs in DNA. Supercoiled DNA in comparison with relaxed DNA and, further, circular plasmid compared to

linear plasmid travel faster due to the more compact structure. At lower voltages the movement rate is proportional to the applied voltage and voltage increase can increase resolution; however, in higher voltages the resolution is lower. Therefore, an optimal voltage can be prescribed for each specific range DNA fragments. High voltages can also heat the gel and melt it. EtBr itself can also affect DNA molecule migration by changing its conformation and charge; so, its concentration in gel can also be considered as an important factor.

T.2. Polymerase chain reaction

Polymerase chain reaction (PCR) is the in vitro amplification of specific sequences of the nucleic acids without any non-specific by-products. PCR can be used for both control of the presence of a desired segment in a DNA sample and increase of the amount of a desired DNA fragment for further utilization in subsequent biological experiments. This process consists of three major steps: denaturation, annealing, and elongation where a single primary denaturation cycle and a single final extension is devised in almost all programs. In denaturation step, template DNA molecule is subjected to heat in order to be denatured and changing into single-strand form. Annealing step comprises binding of pairs of designed sequence-specific oligonucleotides (primers) to the sequences within desired DNA fragment in both strands of template DNA. Finally, elongation step is the extension of the primers by the help of a DNA polymerase enzyme to copy the desired fragment of the template DNA. The utilized DNA polymerase should be thermostable. *Taq* DNA polymerase is a widely used enzyme in PCR reactions which has been isolated from thermophilic bacterium *Thermus aquaticus*. However, it lacks 3'→5' proofreading ability and, thus, can make errors during amplifications. Therefore, substitute DNA polymerase enzymes have been introduced which are less prone to error during amplifications; *Pfu* DNA polymerase from hyperthermophilic archaeon *Pyrococcus furiosus* is one of the substitutes that is also more thermostable than *Taq* DAN

polymerase. Vent DNA polymerase is another example of superior DNA polymerase enzymes which has been isolated from *Thermococcus litoralis*. Furthermore, the polymerization rate of the polymerase enzyme is another factor that can have a great influence on required time for PCR.

All the components of the PCR reaction can play role in its yield and specificity. However, optimal design of the primers is the most critical parameter for a successful PCR. Poorly-designed primers lead to non-specific bands and also primer dimer formation. One of the parameters which strongly influences PCR specificity is the melting temperature (T_m) of the designed primers. Calculated T_m of the primer pair in each PCR should not differ by more than 2-3°C. In addition, GC content of the primer should also be taken into consideration.

The constituents of a normal PCR mixture are: PCR buffer containing Mg^{2+} and KCl, dNTP solution containing all four dNTPs, thermostable DNA polymerase, forward and reverse primers, template DNA, and dH_2O for finalizing the reaction volume to the desired amount. All these components are mixed in thin-walled PCR tubes and put into the thermal cycler. Thermal block cycler should be programmed with desired amplification protocol. After mixture preparation, a program should be decided for amplification of the desired nucleotide segment. In PCRs catalyzed by *Taq* DNA polymerase, denaturation is carried out at 94-95°C, which is the highest temperature that the enzyme can endure for 30 or more cycles without sustaining excessive damage. Too high annealing temperature leads to poor binding of the primers to the template. Too low annealing temperature leads to non-specific binding of the primers. Annealing is usually performed 3-5°C lower than the calculated T_m at which the primers dissociate from their templates. Elongation temperature is the optimum temperature for DNA synthesis which in the case of *Taq* DNA polymerase is 72-78°C. Elongation time will depend on the type of the DNA polymerase and the length of the segment to be amplified. The number of the cycles depends on the amount of the starting template, and efficiency of the primer extension and amplification. Generally

30 cycles are enough in PCRs containing 10^5 copies of the template and *Taq* polymerase efficiency of 0.7.

In the cases that the PCR results are the starting material of another experiment such as ligation, in order to remove remained enzyme and salts, PCR results should be subjected to purification.

T.3. Real-time polymerase chain reaction

In conventional PCR, the amount of the desired product is observed at the end of the reaction. However, in real-time PCR (qPCR) by progression of the reaction (amplification), the accumulated fragment can be detected by inclusion of a fluorescent molecule. By this method, progression of the PCR reaction can be monitored while the exponential phase of the reaction is recruited for calculation of the starting copy number of the template. This method is faster and cost-effective in comparison with DNA hybridization method. Furthermore, it possesses no safety problems.

This technique works based on the binding of a fluorescent dye to the double-stranded DNA (dsDNA); the fluorescence is increased sharply upon binding. Two major groups of the available fluorescent chemistries are: i) DNA-binding dyes (SYBR Green I) which bind non-preferentially to double-stranded DNA, and ii) sequence-specific probes (oligonucleotides) where bind to a specific region on the desired target. In present research, in order to track the amplification of the desired segment, DNA-binding dyes (SYBR Green I) was utilized; the fluorescence of the dye is increased 1000-fold when binds to dsDNA. Utilization of this dye makes it possible to create melt-curve analysis to verify the specificity of the amplification product. Amplification specificity can be checked by both melting curve and gel electrophoresis. The identity of the amplified fragment can also be confirmed by sequencing. It worth mentioning that selection of shorter amplicons will lead to amplification with higher efficiency.

After start of the reaction, the desired amplicon begins to accumulate, at the beginning the amount is low, however, gradually the product

accumulates and the fluorescent signal becomes detectable; Threshold cycle (C_T) is where this detection happens. Before C_T , the fluorescent level is not as high as to be detectable.

In each qPCR experiment, accuracy and reproducibility should be taken into account regarding the template quantification. Therefore, the qPCR should be optimized. The characteristics of an optimized qPCR analysis can be summarized as:

- Linear standard curve with regression coefficient of $R^2 > 0.980$
- Consistency during several similar experiments (replicates)
- High amplification efficiency (90-105%)

Above-mentioned standard curve can be produced either by template sample with known or unknown concentration. Standard curve is made by plotting C_T against the logarithm of the arbitrary starting quantity in each dilution. It should be noticed that the template concentration should be within the range which standard curve has been obtained for, else smaller or larger amounts of the sample should be used regarding to the situation. The unknown (target) samples should be assayed in the same run with standard samples; this will lead to accurate estimation since same reaction conditions have been used for both sample and standard. However, it should be kept in mind that the accuracy of the qPCR results is largely dependent on the factors such as quality of the standard and sample preparation.

Amplification efficiency (E) represents the amplification in PCR cycles; if suppose an ideal increase, product will be doubled during each cycle. It means reaction efficiency of 2.

$$E = 10^{-\frac{1}{a}} \quad (T.1)$$

$$\%E = (E - 1) \times 100 \quad (T.2)$$

Where, “ a ” is the slope of the standard curve.

Equation (T.1) implies that the optimum standard curve slope will be -3.32. The closer the efficiency to 100%, the more robust and reproducible the PCR

will be. The range of 90 – 105% should be expected in order to have an acceptable process.

In addition to the standards and desired samples there should be also no-template control in qPCR reaction in order to be confident about lack of any contamination. During qPCR, statistical variations can be reduced (having more consistent and reproducible results) by preparing master mix for replicate reactions. Triplicate assays are desirable and it would be beneficial to avoid frequent freezing and thawing of stocks.

Finally, it worth mentioning that qPCR quantification can either be absolute or relative. In absolute quantification the copy number (or virtually the amount of the nucleic acid) is obtained by interpolating of known standard samples regarding C_T values. In consequence, the quantity of the sample is obtained by comparing with standard curve. Therefore, having a reliable source of standard with known concentration is the major requisite of this method. In addition, the unknown test samples should be assayed along with standards in same run. On the other hand, in relative quantification the amount relative to a calibrator (strain with known copy number) is achievable. This is useful in order to analyse the fold of increase or decrease of the target gene relative to calibrator or “housekeeping gene”. Standard curve is derived by preparation of serial dilutions of DNA samples with known concentrations containing our desired genes (including target or reference gene). These samples are assayed in the same run with the unknown samples. Then, by the help of obtained relation between C_T and logarithm of the concentration and by taking into account the C_T of the unknown samples, the concentration (here, copy number) of the unknown samples can be determined.

T.4. SDS-PAGE

This method is generally regarded as an analytical method for small amounts of protein samples. The technique is most often performed in polyacrylamide gels (PAGE). During protein movement in electric field this

gel acts like a molecular sieve, similar to AGE, and affects the velocity of the proteins (proportional to charge-to-mass ratio) and, therefore, its electrophoretic mobility which is the ratio of the particle velocity in electric field to the electrical potential ($\mu=V/E$). Since electrophoretic mobility can also be considered as a function of protein net charge and its shape, while proteins preserve their native state and ordered structure, the separation can be analysed by inclusion of charge and shape. By addition of an anionic detergent, sodium dodecyl sulfate (SDS), which binds roughly one molecule for each two amino acids the proteins are negatively-charged (with their intrinsic charge insignificant) with similar charge-to-mass ratio and also denatured and become linear (Figure T.1). Therefore, their mobility in the gel upon exerting an external electric field is only dependent on their molecular weight. It should be mentioned that, the effective range of separation of SDS-PAGE depends on the concentration of the polyacrylamide used to cast the gel and on the amount of the cross-linking. Furthermore, the sieving properties of the gel are determined by the size of the pores, which is a function of the absolute concentration of the acrylamide and bisacrylamide used to cast the gel. Acryl amide concentration of 10% and 12% are suitable for linear range of separation of 20-80 kDa and 12-60 kDa respectively and hGH lies in both of these ranges. After finishing electrophoresis, the proteins may be visualized by utilizing a dye such as silver nitrate which just binds to the protein and not the gel. It should be reminded that in the case of proteins with several subunits, they are separated as the result of treatment by SDS and appear as separate bands on the gel. Using a marker containing proteins with known molecular weights conduces to finding the approximate molecular weight of an unknown protein. Furthermore, in the case of known proteins, standard solution of the protein of interest facilitates verifying the presence of the desired protein.

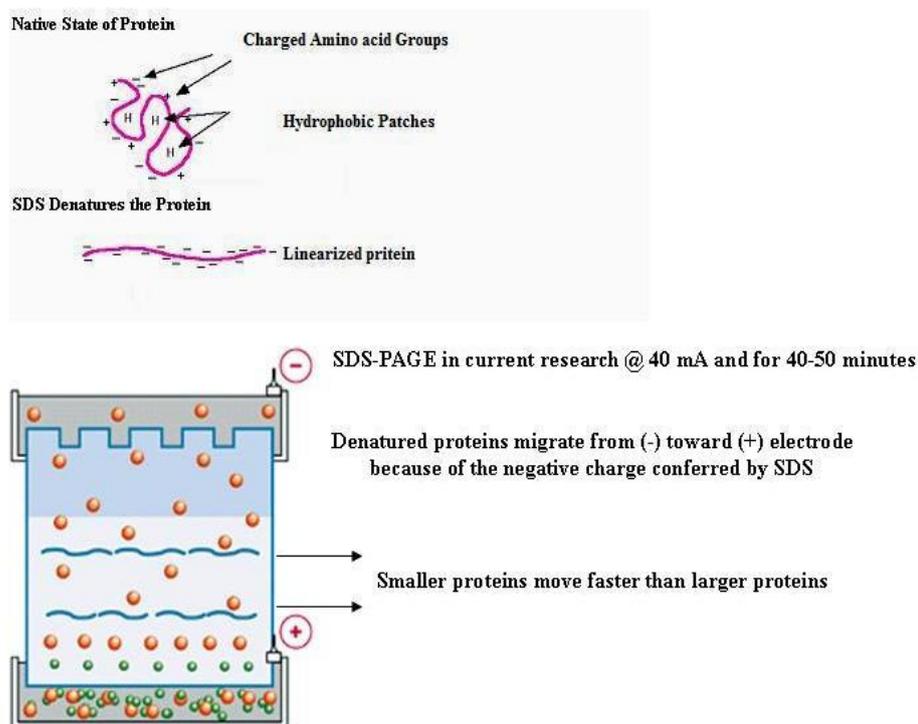


Figure T.1. SDS effect on protein and simple illustration of SDS-PAGE. Reproduced from (Nationa Diagnostics, accessed 14 June 2014, www.nationaldiagnostics.com).

T.5. Dot-blot

This technique is a kind of Western blot, namely the base of this method is to transfer the proteins into a membrane and, then, visualizing them by immunologic methods using antibodies. However, in contrast to Western blot, there is no need to run a gel and transfer the protein bands into a membrane. The protein-containing sample is directly loaded on the membrane. Membrane-bound desired protein first is located by primary antibody which is specific to that protein. Then, secondary antibody which has been conjugated with horseradish peroxidase (HRP) binds to primary antibody. By inclusion of the substrate of the HRP, 3,3'-Diaminobenzidine (DAB) which is a chromogenic substance and is oxidized into a dark-brown

precipitate in presence of hydrogen peroxide, the protein can be detected. The overall procedure has been summarized in Figure T.2.

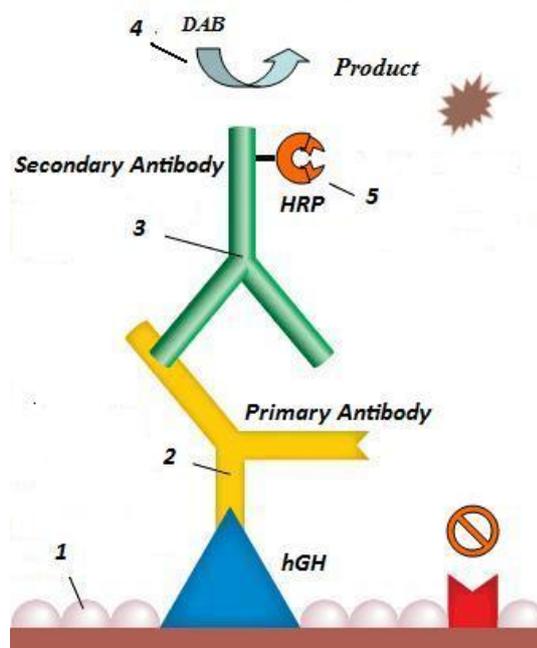


Figure T.2. Schematic representation of the dot-blot analysis. The numbered steps briefly are: 1) Blocking the membrane with blocking buffer. 2) Incubation with protein/antigen-specific primary antibody. 3) Incubation with enzyme-labeled secondary antibody. 4) Addition of the chromogenic substrate of the enzyme. 5) Enzyme catalyzes the reaction which leads to a dark brown precipitate on the membrane. The enzyme in the present research is HRP.

T.6. Cell cycle synchronization

Cell cycle or cell division cycle is a series of events that happen in a typical cell which result in the division and duplication of the cell. The cell cycle differs between eukaryotes and prokaryotes where binary fission led to cell division in prokaryotes. In contrast, in eukaryotes, so called cell cycle includes four well-defined stages: G1, S, G2, and M which can be seen in Figure T.3. In the S phase, DNA replication occurs. In the G2 phase synthesizing of new proteins is carried out and cell size becomes

approximately two-fold. In M phase (followed by cytokinesis), the original nuclear envelope breaks down (karyokinesis) and chromosomes go to the opposite poles. G1 is starting point of a new division where RNA and protein synthesis takes place without any DNA synthesis. However, after M phase, a cell can cease to divide and leave the cycle for a definite period or for whole lifetime, G0 phase. Combination of G1, S, and G2 phases is called interphase which is also called preparatory phase or intermitosis. Roughly, G1 phase lasts 6-12 hours. S phase lasts 6-8 hours. G2 phase lasts 3-4 hours. M phase lasts 1 hour. It should be mentioned that at the end of both G1 and G2 phases there are checkpoints in order to let the cell to proceed to the next stage. As in both G1 and G2, there is a Checkpoint in the middle of mitosis (Metaphase checkpoint). Briefly, Interphase (G1, S, G2) is a kind of preparation that cell grows and accumulates nutrients. In mitotic phase the cell splits into two daughter cells. Finally, cytokinesis is when the cell is completely divided (cytoplasmic division). It should be emphasized that striking phenotypic individuality can be observed within single cells in the cultures that are genetically homogenous. Among proven causatives, cell cycle, cell aging and epigenetic regulations can be mentioned. Thus, in order to prevent discrepancies in results of the fermentation experiments as a result of heterogeneous cell cycles, it was decided to perform cell cycle synchronization prior to production in bioreactor experiments related to promoter strains.

The aim of the cell cycle synchronization is to bring all the cells to the same developmental stage. In spite of having small genome, yeasts show highly conserved cellular processes (e.g. cell cycle control) in comparison with higher eukaryotes regarding genetic, biochemical, and molecular mechanisms. Two broad categories for cell cycle synchronization are chemical approach (induced synchronization) and physical fractionation (selection method synchronization). Where two generally used methods for yeast cell synchronization are: block and release and centrifugal elutriation. It should be reminded that the method of interest must not be harmful to the

cells, the metabolic obstacle should not be irreversible, synchrony must endure for more than one cell cycle, and the synchronized cells proportion must be high.

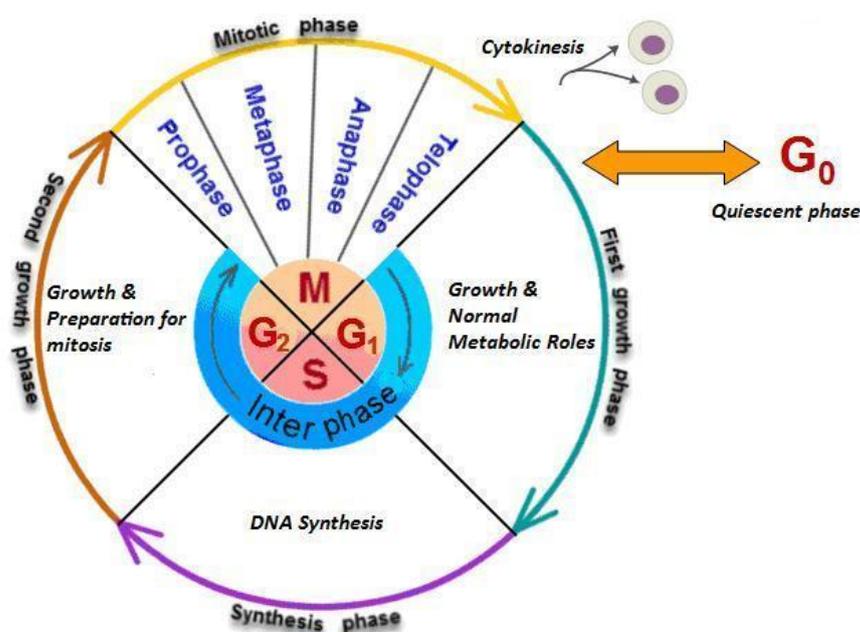


Figure T.3. Schematic of the cell cycle. Reproduced from (TutorVista, accessed 17 March 2016, <http://biology.tutorvista.com/>).

The used method in this study was a kind of block and release method where the base is exposing the cells to an interfering agent or to a condition that blocks specific metabolic reaction and, therefore, blocks cell cycle progression, e.g. hydroxyurea prevents replication of DNA. This method prevents cell cycle progress but not cell growth.

T.7. High pressure liquid chromatography

Chromatography, is a (set of) technique(s) which is used for separation of components of a sample by carrying the sample using a fluid, mobile phase, through a structure called support or stationary phase; the different speeds of

the components through the solid phase leads to the separation of the constituents.

High pressure liquid chromatography (HPLC) is an analytical technique which is used to separate, identify, and quantify the components in a mixture. It has been based on the passage of the components by the aid of a pressurized mobile phase through a solid stationary phase. Instead of moving downward a column under gravitational force, the liquid is forced to move through the column and, thus, it becomes faster. The components of the mixture are separated regarding their different interactions with the stationary phase. The mobile phase is pressurized using pumps; the difference from traditional liquid chromatography is the high pressure, 50-350 bar, of the liquid phase.

Reversed phase HPLC (hydrophobic chromatography) recruits a hydrophobic stationary phase and solutes in a polar solvent come to contact with it. In this case the hydrophobic patches of the solutes interact with hydrophobic support and binds to it. The retention is described by adsorption. Reversed-phase chromatography mechanism encompasses hydrophobic rather than coulombic or hydrogen bonding interactions between the solutes and the ligands in stationary phase. The process is a reversible adsorption/desorption related to the solute molecules, with changing degrees of hydrophobicity, to a hydrophobic matrix.

In order to elute the molecules, a buffer with increasing hydrophobicity is used to overcome the interaction between solute and the matrix by interacting with solute. In brief, the starting mobile phase is aqueous (polar) but at the end the compounds are eluted by a high organic mobile phase such as methanol or acetonitrile.

In total, the procedure involves mainly four steps:

- Equilibration: equilibrating the column with the initial mobile phase
- Adsorption: applying the sample to the column;
- Desorption: changing the mobile phase polarity in a gradual state by an organic phase, e.g., acetonitrile as a polarity controller and

increasing the share of the organic phase to 100% in order to elute all the components adsorbed. Experimental conditions are designed to favour adsorption of the solute from the mobile phase to the stationary phase. Subsequently, the mobile phase composition is modified to favour desorption of the solute from the stationary phase back into the mobile phase

- Regeneration: equilibrating with the initial buffer for next round of the usage

A simplified representation of the procedure is available in (Figure T.4); Sampler enters the mixture into the mobile phase in order to be transferred to the column. Delivering the desired amount and flow of the mobile phase to the column is the responsibility of the pump. Finally, the detector, upon emergence of a component from the column, generates a signal which is proportional to the component concentration. Generally, silica or synthetic organic polymers such as polystyrene are utilized in columns which both mechanical and chemical stability should be taken into consideration in them. The most common columns are filled with silica. Silica is dissolved at high pH, so, pH should not exceed 7. The stationary phase is generally made up of hydrophobic alkyl chains ($-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3$) as ligand on the silica support that interact with the analyte. There are three common chain lengths, C4, C8, and C18. C4 is generally used for proteins and C18 is generally used to capture peptides or small molecules; columns packed with the octadecyl group bonded type silica gel (C_{18} , ODS) are the most widely employed. The particles are characterized by their size and pore. Particle sizes generally range between 3 and 50 μm , with 5 μm particles being the most popular for peptides. Higher separation efficiencies (resolutions) are achieved by smaller particles. The particle pore size is generally between 100-1000 A° . The most utilized pore size in the case of proteins and peptides is 300 A° and 100 A° is the most preferred size for small molecules.

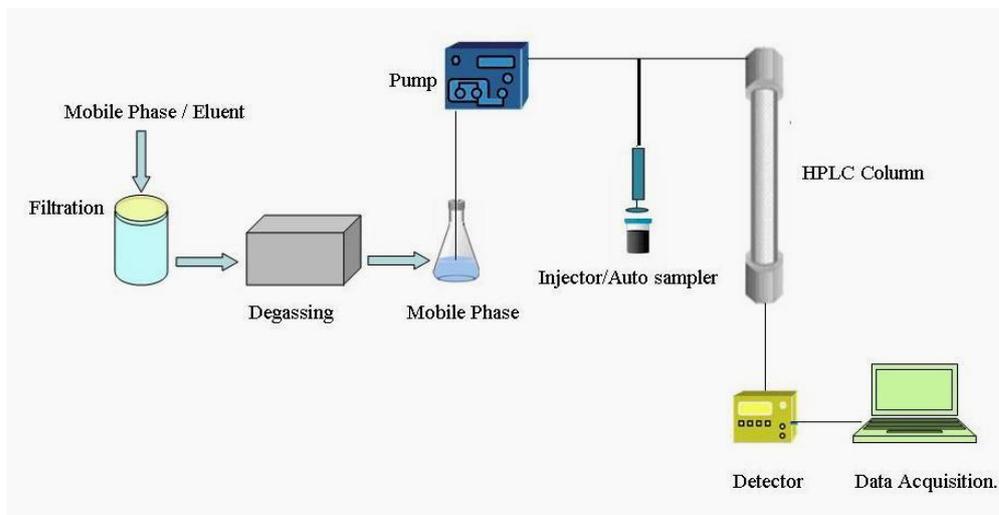


Figure T.4. Simple representation of an HPLC experiment components and their interaction.

U. Metabolic reactions for P. pastoris used in metabolic flux analysis.

MeOH Metabolism

1. MeOH \rightarrow FormAl
2. FormAl \rightarrow For + NADH
3. For \rightarrow NADH + CO₂
4. Xyl5P + FormAl + ATP \rightarrow ADP + 2 G3P

Sorbitol Metabolism

5. Sorb + ATP \rightarrow ADP + F6P + NADH

Glycolysis and Gluconeogenesis Pathway

171. Glc + ATP \rightarrow G6P + ADP
6. F6P \rightarrow G6P
7. G6P \rightarrow F6P
8. F6P + ATP \rightarrow 2 G3P + ADP
9. 2 G3P \rightarrow F6P + Pi
10. G3P + ADP + Pi \rightarrow 3PG + ATP + NADH
11. 3PG + ATP + NADH \rightarrow G3P + ADP + Pi
12. 3PG \rightarrow PEP
13. PEP \rightarrow 3PG
14. PEP + ADP \rightarrow Pyr + ATP
15. Pyr \rightarrow AcCoA_m + CO₂ + NADH_m

Pentose Phosphate Pathway

16. G6P \rightarrow R5P + 2 NADPH + CO₂
17. R5P \rightarrow Xyl5P
18. Xyl5P \rightarrow R5P
19. R5P \rightarrow Rib5P

20. Rib5P \rightarrow R5P
 21. Xyl5P + Rib5P \rightarrow S7P + G3P
 22. S7P + G3P \rightarrow Xyl5P + Rib5P
 23. Xyl5P + E4P \rightarrow F6P + G3P
 24. F6P + G3P \rightarrow Xyl5P + E4P
 25. G3P + S7P \rightarrow F6P + E4P
 26. F6P + E4P \rightarrow G3P + S7P

Branches from Glycolysis Pathway

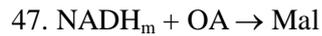
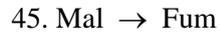
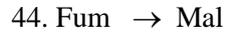
27. Pyr \rightarrow Acet + CO₂
 28. Acet \rightarrow Ac + NADPH
 29. Acet + NADH \rightarrow EtOH
 173. EtOH \rightarrow Acet + NADH
 30. Ac + 2 ATP \rightarrow AcCoA + 2 ADP + 2Pi
 31. AcCoA \rightarrow AcCoAm
 32. Pyr + NADH \rightarrow Lac
 33. Lac \rightarrow NADH + Pyr
 172. Pyr \rightarrow For + AcCoA
 176. For + AcCoA \rightarrow Pyr

Anaplerotic Reactions

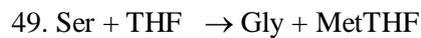
34. Mal \rightarrow Pyr + CO₂ + NADPH
 174. Pyr + CO₂ + NADPH \rightarrow Mal
 35. Pyr + CO₂ + ATP \rightarrow OA + ADP
 36. OA + ATP \rightarrow PEP + ADP + CO₂

TCA cycle

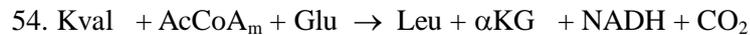
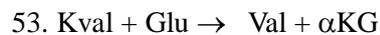
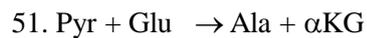
37. AcCoAm + OA \rightarrow Cit
 38. Cit \rightarrow ICit
 39. ICit \rightarrow α KG + CO₂ + NADH_m
 40. α KG \rightarrow SucCoA + CO₂ + NADH_m
 41. SucCoA + Pi + ADP \rightarrow Suc + ATP



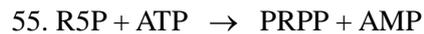
Biosynthesis of Serine Family Amino Acids



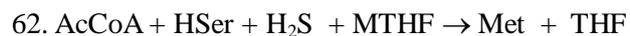
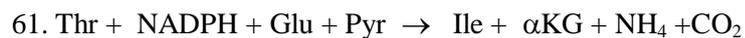
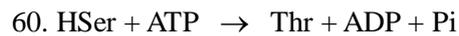
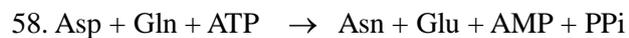
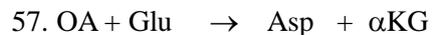
Biosynthesis of Alanine Family Amino Acids



Biosynthesis of Histidine



Biosynthesis of Aspartic Family Amino Acids



Biosynthesis of Aromatic Family Amino Acids

63. $2 \text{ PEP} + \text{E4P} + \text{ATP} + \text{NADPH} \rightarrow \text{Chor} + \text{ADP} + 4 \text{ Pi}$
64. $\text{Chor} + \text{Glu} \rightarrow \text{Phe} + \alpha\text{KG} + \text{CO}_2$
65. $\text{Chor} + \text{Glu} \rightarrow \text{Tyr} + \alpha\text{KG} + \text{NADH} + \text{CO}_2$
66. $\text{Chor} + \text{Gln} + \text{PRPP} + \text{Ser} \rightarrow \text{Trp} + \text{Glu} + \text{Pyr} + \text{G3P} + \text{CO}_2 + \text{PPi}$

Biosynthesis of Glutamic Family Amino Acids

67. $\alpha\text{KG} + \text{NH}_4 + \text{NADPH} \rightarrow \text{Glu}$
68. $\text{Glu} + \text{ATP} + \text{NH}_4 \rightarrow \text{Gln} + \text{ADP} + \text{Pi}$
69. $\text{Glu} + \text{ATP} + 2 \text{ NADPH} \rightarrow \text{Pro} + \text{ADP} + \text{Pi}$
70. $\text{Gln} + \text{CO}_2 + 2 \text{ ATP} \rightarrow \text{CaP} + \text{Glu} + 2 \text{ ADP} + \text{Pi}$
71. $2 \text{ Glu} + \text{AcCoA}_m + 4\text{ATP} + \text{NADPH} + \text{CaP} + \text{Asp} \rightarrow \text{Arg} + \alpha\text{KG} + 4\text{ADP} + \text{Fum} + 5\text{Pi}$
72. $2 \text{ Glu} + \text{AcCoA} + 3 \text{ ATP} + 2 \text{ NADPH} \rightarrow \text{Lys} + \alpha\text{KG} + \text{CO}_2 + 2\text{NADH}$

Catabolism of Amino Acids

73. $\text{Ala} + \alpha\text{KG} \rightarrow \text{Pyr} + \text{Glu}$
74. $\text{Arg} + \alpha\text{KG} + \text{NADPH} \rightarrow \text{Glu} + 2 \text{ NH}_4 + \text{CO}_2$
75. $\text{Asn} \rightarrow \text{Asp} + \text{NH}_4$
76. $\text{Asp} + \alpha\text{KG} + \text{NADH} \rightarrow \text{Glu} + \text{Mal}$
77. $\text{Cys} \rightarrow \text{Pyr} + \text{NH}_4 + \text{H}_2\text{S}$
78. $\text{Gln} \rightarrow \text{Glu} + \text{NH}_4$
79. $\text{Glu} \rightarrow \text{NH}_4 + \text{NADH} + \alpha\text{KG}$
80. $\text{Gly} + \text{MetTHF} \rightarrow \text{Ser} + \text{THF}$
81. $\text{His} + \text{THF} \rightarrow \text{Glu} + \text{F10THF} + \text{NH}_4$
82. $\text{Ile} + \alpha\text{KG} \rightarrow \text{Glu} + \text{FADH}_2 + 2 \text{ NADH} + \text{CO}_2 + \text{AcCoA} + \text{SucCoA}$
83. $\text{Leu} + \alpha\text{KG} + \text{ATP} \rightarrow \text{Glu} + \text{FADH}_2 + \text{NADH} + 2 \text{ AcCoA} + \text{ADP} + \text{Pi}$
84. $\text{Phe} + \text{NADH} \rightarrow \text{Tyr}$
85. $\text{Pro} \rightarrow \text{Glu} + \text{NADH} + \text{FADH}_2$
86. $\text{Ser} \rightarrow \text{Pyr} + \text{NH}_4$
87. $\text{Thr} \rightarrow \text{Gly} + \text{NADH} + \text{AcCoA}$

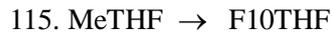
88. $\text{Trp} + \text{NADPH} \rightarrow 2 \text{AcCoA} + \text{Ala} + \text{CO}_2 + \text{NH}_4 + \text{For} + 2\text{NADH} + \text{FADH}_2$
89. $\text{Tyr} + \alpha\text{KG} \rightarrow \text{Glu} + \text{Fum} + 2 \text{AcCoA} + \text{CO}_2$
90. $\text{Val} + \alpha\text{KG} + \text{ATP} \rightarrow \text{Glu} + \text{FADH}_2 + 3 \text{NADH} + \text{CO}_2 + \text{SucCoA}$
91. $\text{Lys} + \text{AcCoA} + 2 \alpha\text{KG} \rightarrow 2 \text{Glu} + \text{NADH} + \text{CO}_2$

Biosynthesis of Nucleotides

92. $\text{PRPP} + 2 \text{Gln} + \text{Asp} + \text{CO}_2 + \text{Gly} + 4 \text{ATP} + \text{F10THF} \rightarrow 2 \text{Glu} + \text{PPi} + 4 \text{ADP} + 4 \text{Pi} + \text{THF} + \text{PRAIC} + \text{Fum}$
93. $\text{PRAIC} + \text{F10THF} \rightarrow \text{IMP} + \text{THF}$
94. $\text{IMP} + \text{Gln} + \text{ATP} \rightarrow \text{NADH} + \text{GMP} + \text{Glu} + \text{AMP} + \text{PPi}$
95. $\text{GMP} + \text{ATP} \rightarrow \text{GDP} + \text{ADP}$
96. $\text{ATP} + \text{GDP} \rightarrow \text{ADP} + \text{GTP}$
97. $\text{GTP} + \text{ADP} \rightarrow \text{ATP} + \text{GDP}$
98. $\text{NADPH} + \text{ATP} \rightarrow \text{dATP}$
99. $\text{NADPH} + \text{GDP} + \text{ATP} \rightarrow \text{dGTP} + \text{ADP}$
100. $\text{IMP} + \text{GTP} + \text{Asp} \rightarrow \text{GDP} + \text{Pi} + \text{Fum} + \text{AMP}$
101. $\text{AMP} + \text{ATP} \rightarrow 2 \text{ADP}$
102. $\text{PRPP} + \text{Asp} + \text{CaP} \rightarrow \text{UMP} + \text{NADH} + \text{PPi} + \text{Pi} + \text{CO}_2$
103. $\text{UMP} + \text{ATP} \rightarrow \text{UDP} + \text{ADP}$
104. $\text{UDP} + \text{ATP} \rightarrow \text{ADP} + \text{UTP}$
105. $\text{UTP} + \text{NH}_4 + \text{ATP} \rightarrow \text{CTP} + \text{ADP} + \text{Pi}$
106. $\text{CTP} + \text{ADP} \rightarrow \text{CDP} + \text{ATP}$
107. $\text{CDP} + \text{ATP} \rightarrow \text{CTP} + \text{ADP}$
108. $\text{CDP} + \text{ADP} \rightarrow \text{CMP} + \text{ATP}$
109. $\text{ATP} + \text{NADPH} + \text{CDP} \rightarrow \text{dCTP} + \text{ADP}$
110. $\text{UDP} + \text{MetTHF} + 3 \text{ATP} + \text{NADPH} \rightarrow \text{dTTP} + \text{DHF} + 3 \text{ADP} + \text{Ppi} + \text{Pi}$

Biosynthesis and Interconversion of One-carbon Units

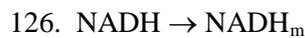
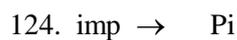
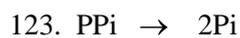
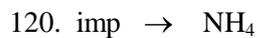
111. $\text{DHF} + \text{NADPH} \rightarrow \text{THF}$
112. $\text{Gly} + \text{THF} \rightarrow \text{MetTHF} + \text{NH}_4 + \text{NADH} + \text{CO}_2$
113. $\text{MetTHF} + \text{NADH} \rightarrow \text{MTHF}$



Oxidative Phosphorylation ($P/O = 2$)



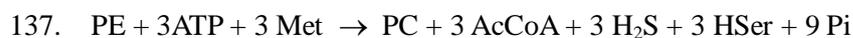
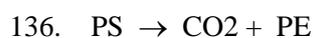
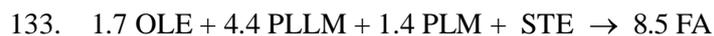
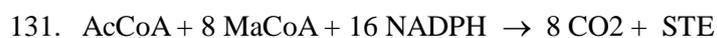
Transport Reactions



Biosynthesis of Carbohydrate

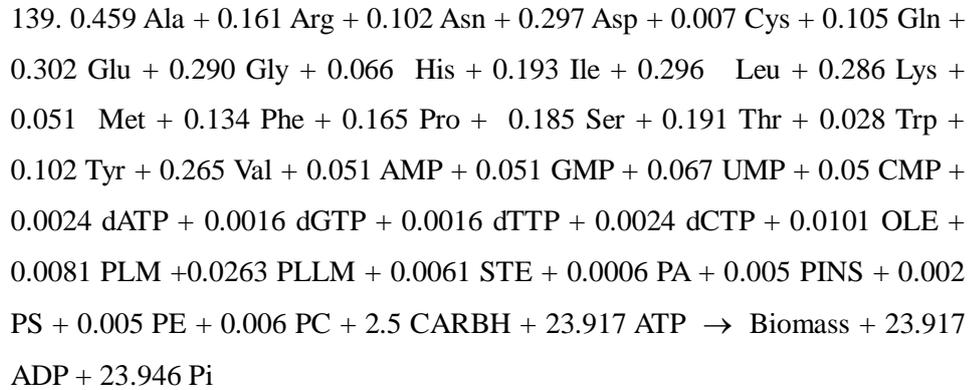


Biosynthesis of Lipids

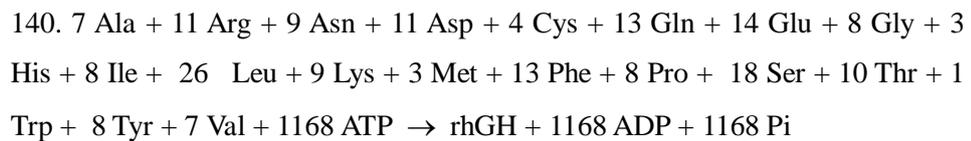




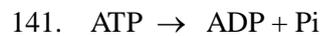
Biomass Synthesis



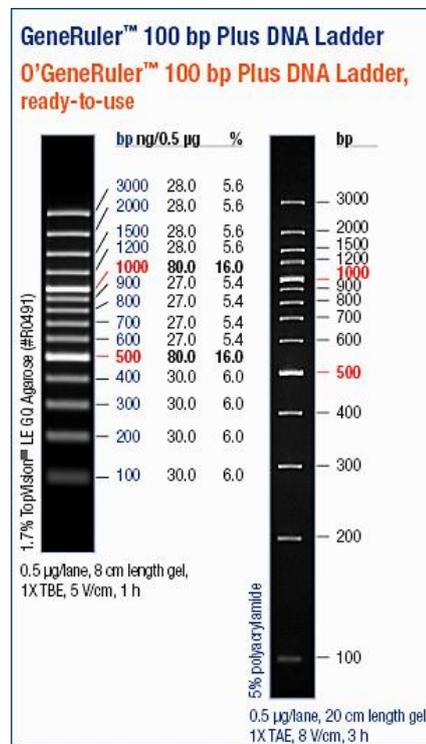
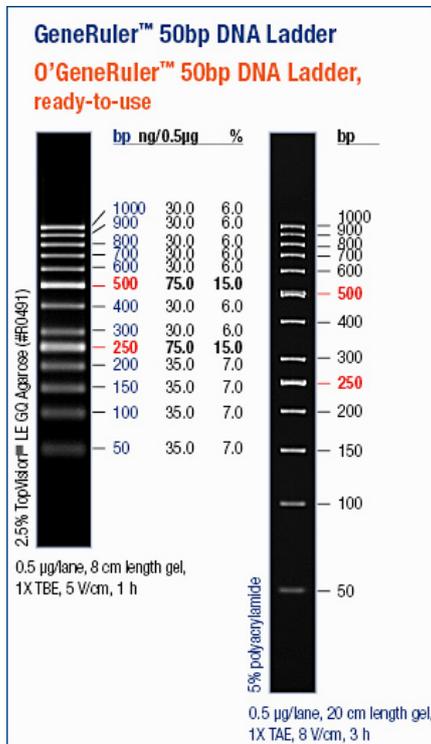
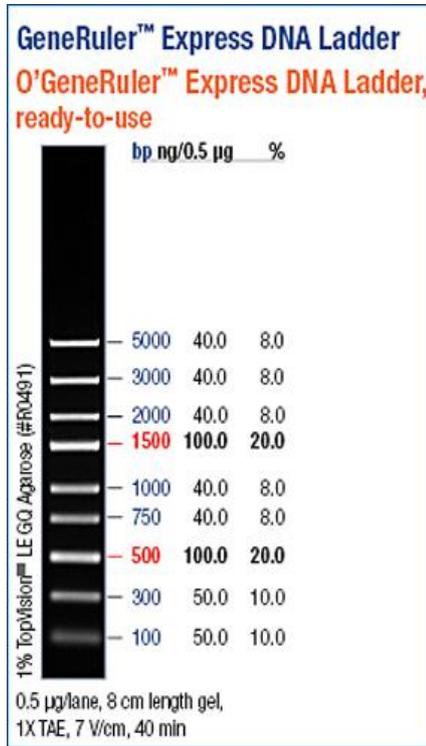
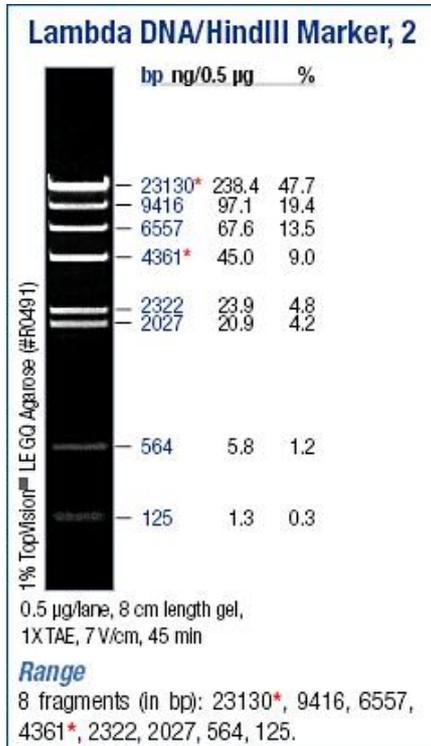
Human Growth Hormone Synthesis



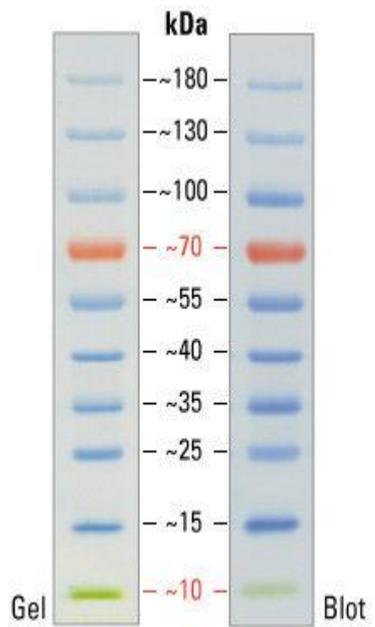
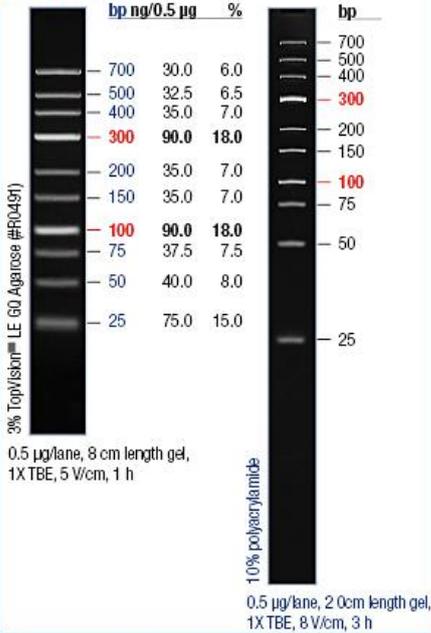
Maintenance Energy



V. Utilized DNA markers and protein marker.



GeneRuler™ Low Range DNA Ladder
O'GeneRuler™ Low Range DNA Ladder,
ready-to-use



PageRuler™ prestained protein ladder

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Massahi, Aslan
Nationality: Iranian
Date and Place of Birth: 13 August 1980, Maragheh (IRAN)
Marital Status: Married
Phone: +98 935 494 7577
E-mail: aslan.massahi@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
PhD.	Middle East Technical University- Ankara - TURKEY	2017
MSc.	Amirkabir University of Technology- Tehran - IRAN	2005
BSc.	Sharif University of Technology Tehran - IRAN	2002
High School	Nemooneh Azadegan Maragheh – IRAN	1997

WORK EXPERIENCE

Year	Place	Enrollment
Sep. 2000	Tabriz Petrochemical Complex– Polyethylene Unit	Internship
2003-2005	Amirkabir University of Technology- Tehran - IRAN	Research Assistant
2006-2008	Iran's Police Health Center Tehran - IRAN	Military Service
2013-2017	Middle East Technical University	Research Assistant

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

1- Massahi, A., Çalık, P., 2015. In-silico determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production. Journal of Theoretical Biology 364, 179-188.

2- Çalık,P., Ata,Ö., Güneş, H., Massahi, A., Boy,E .,Keskin,A., Öztürk,S., Zerze, G.H., Özdamar, T.H., 2015. Recombinant protein production in *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter: from carbon source to bioreactor operation parameters. Biochemical Engineering Journal 95, 20–36.

3- Massahi, A., Çalık, P., 2016. Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: From in-silico analysis to fermentation. Journal of Theoretical Biology 408, 22-33.

HOBBIES

Soccer, Music, Poem, Movies