DELETION MUTAGENESIS OF *FUSARIUM GRAMINEARUM* NRRL 2903 GALACTOSE OXIDASE IN *ESCHERICHIA COLI* TO STUDY STRUCTURE-FUNCTION RELATIONS AND FOR PROSPECTIVE BIOSENSOR APPLICATIONS

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

BY

BURÇAK KOCUKLU

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

SEPTEMBER 2013

Approval of the thesis:

DELETION MUTAGENESIS OF *FUSARIUM GRAMINEARUM* NRRL 2903 GALACTOSE OXIDASE IN *ESCHERICHIA COLI* TO STUDY STRUCTURE-FUNCTION RELATIONS AND FOR PROSPECTIVE BIOSENSOR APPLICATIONS

submitted by **BURÇAK KOCUKLU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biotechnology Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Nesrin Hasırcı _____
Head of Department, **Biotechnology**

Prof. Dr. Zümrüt Begüm Ögel _____
Supervisor, **Food Engineering Dept., METU**

Prof. Dr. Ufuk Bölükbaşı _____
Co-Supervisor, **Chemical Engineering Dept., METU**

Examining Committee Members

Prof. Dr. Gülay Özcengiz _____
Biology Dept., METU

Prof. Dr. Zümrüt Begüm Ögel _____
Food Engineering Dept., METU

Assoc. Prof. Sertaç Önde _____
Biology Dept., METU

Prof. Dr. Cumhur Çökmüş _____
Biology Dept., Ankara University

Assist. Prof. Dr. Gökhan Duruksu _____
KOGEM, Kocaeli University

Date: 17.09.2013

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 : Burçak KOCUKLU

Signature :

ABSTRACT

DELETION MUTAGENESIS OF *FUSARIUM GRAMINEARUM* NRRL 2903 GALACTOSE OXIDASE IN *ESCHERICHIA COLI* TO STUDY STRUCTURE-FUNCTION RELATIONS AND FOR PROSPECTIVE BIOSENSOR APPLICATIONS

Kocuklu, Burçak
Ph.D., Department of Biotechnology
Supervisor : Prof. Dr. Zümrüt Begüm Ögel
Co-Supervisor : Prof. Dr. Ufuk Bölükbaşı

September 2013, 227 pages

Galactose oxidase (GAO; EC 1.1.3.9) from *Fusarium graminearum*; is a 68kDa, monomeric extracellular enzyme having an unusual thioether bond with a single copper ion at its active site. The enzyme is produced as a precursor with a 25 amino-acid leader peptide, consisting of a 17-amino acid pro- and an 8-amino acid putative pre-peptide, in addition to a signal peptide. Based on previous studies, the 17-amino acid pro-peptide is removed autocatalytically by the aerobic addition of Cu^{2+} to the precursor *in vitro*, preceding the formation of the thioether bond at the active site. In our previous study, it was discovered that self-processing of the pro-peptide did not take place when GAO was expressed heterologously in *E.coli* in its pre-pro- precursor form. In this study, the question whether the 8-amino acid pre-peptide is hindering the self- cleavage of heterologously expressed pre-pro-galactose oxidase in *E.coli* is tried to be answered by deleting the 8 amino-acid pre-peptide via site-directed mutagenesis.

pPre-ProGON1 (encoding wild- type GAO together with the leader peptide with N-terminal silent mutations) and pPre-ProGOMN1 (encoding a variant of GAO together with the leader peptide, with silent mutations at the N-terminus different from native *gao* and six further mutations within the mature enzyme) which were developed by directed evolution, were used as the plasmid constructs in developing deletion mutations in this thesis study. The pre-peptide deletion was carried out by *QuikChange® Site-Directed Mutagenesis* approach and confirmed by DNA sequencing. The newly generated pProGON1 and pProGOMN1 constructs were expressed in *E. coli* BL21 Star (DE3) through pET expression system either by IPTG induction or by auto-induction. Following purification trials of mutant galactose oxidases by different affinity chromatography matrices, SDS-PAGE analysis was performed to figure out whether the pre-deleted galactose oxidase variants recover the ability of self-processing. Comparison of the SDS-PAGE bands substantiated that the pro-peptide is still present at the N-terminus. Accordingly; detection of GAO activity and molecular weight derived from SDS-PAGE also indicated that pro-peptide has no detrimental effect on the proper thioether bond formation at the active site.However, specific activities of Pro-GAOs

and Pre-Pro-GAOs are *c.*13 fold and *c.*1,8 fold lower than MatGOMN6, respectively, indicating pre-peptide effect on active site conformation. In addition, the prevention of the autoprocessing of the pro-peptide is not due to the presence of the pre-peptide at the N-terminus in *E.coli*. Moreover, this discovery underlines that the self-catalytic maturation of the galactose oxidase may not be only copper and oxygen dependent but any other mechanism present or absent in *E.coli* such as posttranslational modifications prevents autoprocessing. It also appears from this study that pro-peptide self-cleavage is independent of the primary amino acid sequence of GAO.

On the second part of this study, domain deletions of a variant of mature galactose oxidase, namely MatGOMN6 (with silent mutations at the N-terminus different from native *gao* and six further mutations within the mature enzyme), were aimed for novel biosensor applications. After His-tag had been substituted for *Strep*-tag successfully on the construct encoding mature GAO (pMatGOMN6), in order to facilitate downstream processes, deletion and expression studies of domain I/II and domain III were carried out on both derivatives as outlined above. Further optimization and purification studies are underway as a part of another project.

Keywords: galactose oxidase, auto-induction, pre-peptide, pro-peptide, self-processing, autoprocessing, site-directed mutagenesis, deletion, biosensor.

ÖZ

FUSARIUM GRAMINEARUM NRRL 2903 GALAKTOZ OKSİDAZININ YAPI-İŞLEV İLİŞKİLERİNİN VE İLERİYE DÖNÜK BİYOSENSÖR UYGULAMALARININ ARAŞTIRILMASI İÇİN *ESCHERICHIA COLI*'DE DELESYON MUTAJENEZİ ÇALIŞMALARI

Kocuklu, Burçak
Doktora, Biyoteknoloji Bölümü
Tez Yöneticisi : Prof. Dr. Zümrüt Begüm Ögel
Ortak Tez Yöneticisi: Prof. Dr. Ufuk Bölükbaşı

Eylül 2013, 227 sayfa

Fusarium graminearum galaktoz oksidazı (GAO; EC 1.1.3.9); 68kDa, aktif bölgesinde tek bir bakır iyonu içeren alışılmadık bir tiyoeter bağına sahip monomerik ekstraselüler bir enzimdir. Enzim sinyal peptide ek olarak, 17 amino asitlik pro- ve 8 amino asitlik muhtemel pre-peptidden oluşan 25 amino asitlik bir lider peptid ile bir öncü olarak üretilir. Daha önceki çalışmalara dayanarak, bakırın aerobik olarak öncüye *in vitro* ilavesiyle 17-amino asitlik pro-peptid, aktif bölgedeki tiyoeter bağının oluşmasının öncesinde otokatalitik olarak uzaklaştırılmaktadır. Daha önceki çalışmamızda, GAO'nun pre-pro- öncü formunda *E.coli'*de heterolog olarak ifade edildiğinde pro-peptidinin kendini işlemesinin gerçekleşmediği ortaya çıkarılmıştır. Bu çalışmada, 8 amino asitlik pre-peptidin *E.coli'*de heterolog olarak ifade edilen pre-pro-galaktoz oksidazın kendi kendini kesimini engelleyip engellemediği sorusu, 8 amino asitlik pre-peptid alan hedefli mutajenez yoluyla delesyona uğratılarak cevaplanmaya çalışılacaktır.

Bu tez çalışmasında, delesyon mutasyonları geliştirmek için plasmid yapıları olarak yönlendirilmiş evrim ile oluşturulan pPre-ProGON1 (N-terminalinde sessiz mutasyonlar ve lider peptidi ile birlikte yaban tipte GAO kodlamaktadır) ve pPre-ProGOMN1 (yerli *gao*'dan farklı olarak N-terminalinde sessiz mutasyonlar ve işlenmiş enzim kısmında altı mutasyona daha sahip olan, lider peptidi ile birlikte GAO varyantı kodlamaktadır) kullanılmıştır. Pre-peptid delesyonu *QuikChange® Alan-Hedefli Mutajenez* yaklaşımı ile gerçekleştirilmiş ve DNA dizi analizi ile doğrulanmıştır. Yeni oluşturulmuş pProGON1 ve pProGOMN1 yapıları pET ekspresyon sistemi yoluyla hem IPTG indüklemesi , hem de otoindüksiyonla *E. coli* BL21 Star (DE3)'de ifade edilmiştir. Mutant galaktoz oksidazların farklı afinite kromatografi matriksleri ile saflaştırma denemelerini takiben, pre-delesyona uğrayan galaktoz oksidaz varyantlarının kendini işleme yeteneğini yeniden kazanıp kazanamadığının anlaşılması için SDS-PAGE analizi gerçekleştirilmiştir. SDS-PAGE bantlarının karşılaştırılması pro-peptidin hala N-terminus'ta yer aldığını göstermektedir. Buna göre; GAO aktivitesinin tespiti ve SDS-PAGE'den elde edilen moleküler ağırlık da aktif bölgede tiyoeter bağının uygun bir

biçimde oluşumunda pro-peptidin zararlı bir etkiye sahip olmadığına işaret etmektedir. Bununla birlikte, Pro-GAO ve Pre-Pro-GAO'ların spesifik aktiviteleri MatGOMN6'dan sırasıyla *c.*13 kat ve *c.*1,8 kat daha düşüktür, ki bu aktif bölge konformasyonunda pre-peptidin etkisine dikkat çekmektedir. Buna ilaveten, *E.coli*'de pro-peptidin kendini işlemesinin engellenmesi N-terminus'taki pre-peptid varlığından kaynaklanmamaktadır. Üstelik, bu keşif galaktoz oksidazın kendi kendini keserek olgunlaşmasının yalnızca oksijen ve bakıra bağlı değil, ancak kendi kendine kesimi engelleyen post-translasyonal modifikasyonlar gibi *E.coli*'de var olan ya da olmayan başka bir mekanizmaya da bağlı olabileceğinin altını çizmektedir. Ayrıca, bu çalışmayla pro-peptid kesiminin galaktoz oksidazın primer amino asit sekansından bağımsız olduğu da gösterilmiştir.

Çalışmanın ikinci kısmında, yeni geliştirilen biyosensör uygulamaları için MatGOMN6 olarak adlandırılan işlenmiş galaktoz oksidaz varyantında (yerli *gao*'dan farklı olarak N-terminalinde sessiz mutasyonlar ve işlenmiş enzim kısmında altı mutasyona daha sahip) domain delesyonları amaçlanmıştır. İşlenmiş GAO kodlayan yapı üzerinde (pMatGOMN6), downstream prosesleri kolaylaştırmak için *Strep*-tag yerine His-tag değişimi başarılı bir şekilde gerçekleştirildikten sonra yukarıda özetlendiği üzere; her iki türevi üzerinde de domain I/II ve domain III için delesyon ve ekspresyon çalışmaları gerçekleştirilmiştir. Daha ileri optimizasyon ve saflaştırma çalışmaları başka bir projenin parçası olarak devam etmektedir.

Anahtar Kelimeler: galaktoz oksidaz, otoindüksiyon, pre-peptid, pro-peptid, kendini işleme, alan-hedefli (yönlendirilmiş) mutajenez, delesyon, biyosensör.

Dedicated to My Family & My Husband…

ACKNOWLEDGEMENTS

First and foremost, I would like to express my special thanks to my supervisor Prof. Dr. Zümrüt Ögel for her support, especially at the very end, her motivating and encouraging attitude towards me and my study. Knowing her that believing in me accomplishing this PhD work made me feel better when I was falling on hard times.

I owe my deepest thanks to Prof. Dr. Michael J. McPherson from University of Leeds, School of Molecular and Cellular Biology, who acted as my co-supervisor for accepting me working in his lab again after my MSc. This great opportunity, which facilitated the process that would be harder to finalize within a short time while working and his invaluable opinions with his immense consultancy guided me in framing this thesis. I also would like to thank Dr. Sarah E. Deacon and Thembaninkosi Gaule for their help and useful discussions, during those hard working days in Leeds.

I would like to extend my sincere thanks to my managers at Ministry of Forestry and Water Affairs, particularly to wonderful people Akil Yıldız and Hüsniye Kılınçarslan for their help by allowing me to create a very precious time within working days, which facilitated doing my PhD.

I would like to indicate that the financial support of TÜBİTAK is worth taking its place among my earnest thanks. The scholarship provided by BAYG throughout my PhD and particularly the bursary supplied by BİDEB for going abroad to University of Leeds made it easier to realize my endeavours.

My friends and lab mates whose names must appear in these lines are Sümeyra Gürkök, Gökhan Duruksu, Betül Söyler, Abduvali Valiev for their kind help, useful comments and intellectual conversations; Yonca Yüzügüllü, Özlem Cesur especially for sharing those nice and hard days in UK; Menekşe Şenyiğit, Evrim Oktay, Burcu İlçe for their support and unflagging friendship; Tülay Bal, who has the greatest potential to put a large smile on my face, for the boost she gave me notwithstanding the situation she is in, for sharing ups and downs. I am very happy to be a friend of you.

My sincere thanks go to my intimate friend Eda Alagöz, for her encouragement, help and understanding. Walking on the same rough road together with such a diligent and honest person made this long way more bearable. I wish our discussions and your valuable advice along with delightful breakfasts in the early morning would last beyond memories.

I want to give my deepest gratitude to my family for making everything easier for me when I was unable to work out. I am very lucky to feel your endless love and support anytime I need. It is wonderful to know that I have you and knowing that you believe in me.

Last, but not the least, my beloved husband deserves my thanks from the bottom of my heart for his great patience, boost, care and everlasting love. My mere expression of thanks does not suffice to tell how it makes me happy to feel him beside me for all the times. I love you more than you will ever know...

To the people mentioned above... without your support, this thesis would have remained a dream.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

A	adenine
C	cytosine
G	guanine
T	thymine
U	unit
bp	base pairs
kb	kilobase
gao	galactose oxidase gene
gla	glucoamylase
OD	optical density
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
BSA	bovine serum albumin
SDM	sample denaturing mixture
SDS	sodiumdodecylsulfate
APP	amyloid beta precursor protein
PrP	prion protein
ABTS	2,2'-Azino-bis(3-Ethylbenz-thiazoline-6-suphonic acid)
EDTA	ethylenediaminetetraacetate
EtBr	ethidium bromide
TEMED	N,N,N',N', Tetramethylethylenediamine
dNTP	deoxynucleosidetriphosphate
DNase	deoxyribonuclease
RNase	ribonuclease
F	forward
GAOX	galactose oxidase
R	reverse
dsDNA	double-stranded DNA
HF	high-fidelity
min	minute
s	second
h	hour
QCSDM	QuikChange®Site-Directed Mutagenesis
ml	mililiter
µl	microliter
PBS	phosphate-buffered saline
GAO	galactose oxidase
PVDF	polyvinylidene difluoride
CV	column volume

HABA	2- [4'-hydroxy-benzeneazo] benzoic acid
APS	ammonium persulfate
GB	gel buffer (separating)
SGB	stacking gel buffer
TBS	tris buffered saline
TBS-T	tris buffered saline-tween 20
IgG	immunuglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside
HRP	horseradish peroxidase
NaPi	sodium phosphate buffer
Pfu	*Pyrococcus furiosus*
Taq	*Thermus aquaticus*
aa	amino acid
IB	inclusion body
n.a	not applied
Pro-GAO	pro-galactose oxidase
Pre-Pro-GAO	pre-pro-galactose oxidase
TMB	3,3',5,5'-Tetramethylbenzidine
Gal-GalNAc	D-galactose-β[1,3]-*N*-acetylgalactosamine (T-antigen)
NAF	nipple aspirate fluid
SRP	signal recognition particle
RER	rough endoplasmic reticulum
TGN	trans-Golgi network
ER	endoplasmic reticulum
orf	open reading frame
RBS	ribosome binding site
D	domain
c-AMP	cyclic-AMP (adenosine monophosphate)
CRP	cyclic AMP receptor protein
CAP	catabolite activator protein
PTS	phosphoenolpyruvate/carbohydrate transferase system
EI	enzyme I
EII	enzyme II
EII^{Glc}	glucose transporter
lac A	β-galactosidase transacetylase
lac I	lac repressor
lac Y	lactose permease
lac Z	β-galactosidase

CHAPTER 1

INTRODUCTION

1.1 Galactose Oxidase

1.1.1 General View

Galactose oxidase (D-galactose: oxygen 6-oxidoreductase, GAO; EC 1.1.3.9) is a copper containing enzyme produced by various fungal species (Cooper *et al*., 1959; Amaral *et al*., 1963).

Galactose oxidase from *Fusarium graminearum* (NRRL 2903) having a relative molecular mass of 68kDa is an extracellular monomeric copper metalloenzyme (Kosman *et al.*,1974).

The reaction catalyzed by galactose oxidase is the two-electron oxidation of the wide range of primary alcohols (position C-6) to corresponding aldehydes with the concomitant two-electron reduction of dioxygen to H_2O_2 (Avigad *et al*., 1962; Tressel & Kosman, 1982).

$$RCH_2OH + O_2 \bullet\bullet\bullet\bullet\blacktriangleright RCHO + H_2O_2$$

Galactose oxidase contains two cofactors, a single copper at the active site and an unusual amino acid-derived tyrosine which constitutes the unusual thioether bond with cysteine residue at position 228, later identified to form a radical following the involvement of oxygen (Whittaker & Whittaker, 1988a; Whittaker *et al*., 1989; Ito *et al*., 1991, 1994; and Rogers *et al*., 2008).

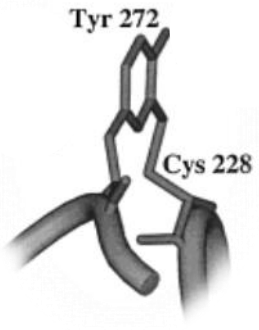


Figure 1.1 The thioether bond between Cys228 and Tyr272 (adopted from Rogers *et al*., 2000)

The isoelectric point of galactose oxidase is above pH 10 which leads to drive the enzyme to to be out of its character in physical measurements and enzyme-activity assays (Kosman, 1974)

1.1.2 Taxonomic Classification History of *Fusarium graminearum* GAO and Its Natural Producers

Galactose oxidase was first reported to be isolated from the mold *Polyporus circinatus* (Cooper *et al*., 1959). It was reclassified as *Dactylium dendroides* with a strain designation number NRRL 2903 (American Type Culture Collection -ATTC- 46032) following the isolation as a mycoparasite of the *Polyporus circinatus* in Southern Brazil, by Nobles and Madhosingh (1963) who suggested that this was a misnomer. *Dactylium dendroides* is called as *Cladobotryum dendroides* as well and *Hypomyces rosellus* is also used by NCBI database for sequence and structure enquiries.

Dactylium dendroides GAO was used as a model for many studies on synthesis, secretion, production and characterization of the enzyme. Even though the name was used in litrature for long years, molecular studies clarified the taxonomic classification fault. (Barbosa-Tessmann *et al.*, 2001).

Ögel *et al*.,(1994) reported the sporulation of the *Cladobotryum (Dactylium) dendroides* (NRRL 2903) for the first time indicating the similarity between the conidia and the conidiophores of this fungus with the ones of *Fusarium chlamydosporum* (Booth, 1971).The comparison of isolates of *Hypomyces sp.* and *F. chlamydosporum* were carried out by molecular techniques beside GAO production determination, and were resulted in reclassification of the organism as a *Fusarium sp* (Ögel *et al*.,1994). However, these differences impaired the classification of NRRL 2903 as *F. chlamydosporum* until O'Donnell *et al* (2000) demonstrated that NRRL 2903 is within the lineage I of the *F. graminearum* clade by phylogenetic, micotoxin, and pathogenicity studies on various *F. graminearum* (a plant pathogen) strains (Barbosa-Tessmann *et al.*, 2001). Within the following study of the O'Donnell *et al* (2004), with the organization of the *F. graminearum* species complex, this isolate was recognized as lineage 1 of this complex or *Fusarium austroamericanum*.

Re-identification of the galactose oxidase-producing fungus *Dactylium dendroides* as a *Fusarium* species was done by detecting mycotoxin production. Growth of the fungus on different media and conditions for verification of the mycotoxin and GAO production is to detect whether the fungus could produce *Fusarium* toxins. The *Fusarium* toxins deoxynivalenol, 3-acetyldeoxynivalenol and zearalenone were detected in the fungal culture medium used, which revealed that this result is coherent with the hypothesis that the fungus is in fact a *Fusarium* species (Machado & Kemmelmeier, 2001).

The previous wrong identification of *Dactylium dendroides* as a galactose oxidase producer was also supported by not detecting any enzyme activity in this species (Aisaka & Terada, 1981).

Besides *F.graminearum, Gibberella fujikuroi* (Aisaka & Terada, 1981), complex isolates (such as *Fusarium moniliforme* f.sp. *subglutinans, telemorph of Gibberella fujikuroi*) and *F. acuminatum* (Barbosa-Tessmann *et al*.,2001) were also reported to produce GAO.

Furthermore, there are also other fungal species reported to have galactose oxidase activity. These are; *Alternaria sp.*, *Helminthosporum sp.* and *Penicillium album* (Gancedo *et al.*, 1967), *Absidia butleri*, *Aspergillus japonicus*, *Aspergillus niger*, *Daedaleopsis styracina*, *Neurospora crassa* and *Penicillium sp.* (Aisaka & Terada, 1981). However, it is not clear that these activities belong to the same type of enzyme as *Fusarium sp*. have. At present, there is no evidence for the presence of GAO except in fungi. However, a prokaryotic homologue of *gao, fbfB* gene of *Stigmatella aurantiaca* with limited identity, was identified on a nucleotide sequence basis (Silakowski *et al.*, 1998) and appears to be involved in the developmental biology of the myxobacteria (Dworkin, 1996). The protein has a molecular mass of 57,8 kDa, the copper ligands and some active site residues show high similarity to that of GAO (McPherson *et al.*, 2001; Whittaker, 2003). Besides, sequence database searches in *Streptomyces coelicolor A3* and *Arabiodopsis thaliana* gave rise to limited identity although it is not clear that these encode similar enzymes (McPherson *et al.*, 2001).

1.1.3 Biological Function

Although the physiological function of GAO is unclear, it catalyses the oxidation of a broad range of primary alcohols to their corresponding aldehydes (Avigad *et al*., 1962) and its unique properties such as structural simplicity, stability, stereoselectivity (Ito *et al*., 1992) and self-processing ability are remarkable points with respect to biotechnological applications and catalytic mechanism.

It may take place in plant pathogenesis although this does not seem likely (Pallas & McPherson, 1997). Although there is not definitive evidence that *Fusaria* have lignin or cellulose degrading peroxidases, H_2O_2 as a final product can act as a substrate for such enzymes, besides providing an antibiotic defence within the rhizosphere. Instead, GAO may alter the fungal cell wall properties by modifying cell surface sugars or it can act to provide saprophytic or necrophytic growth (McPherson *et al.*, 2001).

The discovery of the similar enzyme glyoxal oxidase, also a hydrogen peroxide producer, has supported the idea that galactose oxidase may play a role in lignin or cellulose degradation, although there is no evidence. As the alcohol substrates of galactose oxidase are accounted for lignin degradation intermediates this theory is supported further (Whittaker, 1994,1996).

1.1.4 Substrate Specificity and Selectivity

The broad range of GAO substrates differs from a wide range of primary alcohols (i.e glycerol, benzyl alcohol and allyl alcohol) to D-galactose, galacto- pyranosides, oligo- and polysaccharides, although the biologically relevant substrate of GAO is not known (Cooper, 1959; Schlegel *et al*., 1968; Avigad, 1985; Bretting & Jacobs, 1987; Mendonca & Zancan, 1987).

Although the use of D-galactose is most common, the most efficient substrate is dihydroxyacetone that enzyme exhibits a fourfold greater k_{cat}/K_m for this substrate than for D-galactose (Hamilton *et al*., 1973; Zancan & Amaral, 1970). D-galactose itself may not be the natural substrate of GAO, however it shows much more affinity to galactosides and polysaccharides with galactose end groups (Avigad *et al*.,1962).D-galactose is oxidized at C-6 position rather than C-1 position by GAO, forming D-*galacto*-hexodialdose rather than the corresponding hexonic acid (Avigad *et al*., 1961).

D-glucose and L-galactose are not substrates of GAO (Ito *et al.*, 1992**)** which indicates that the enzyme has a strict stereo- and regio substrate specificity feature (Firbank *et al*., 2003). Besides its ability of oxidizing alcohols, the enzyme further converts aldehydes to the corresponding carboxylates (Kelleher & Bhavanandan, 1986).

An inducer of the GAO is not known (Ögel *et al.*, 1994). Meanwhile L-Sorbose is routinely used in GAO production and purification trials, although there is no information about the role of L-Sorbose on the GAO production mechanism whether it works on an induction or derepression basis (Ögel & Özilgen, 1995).

Substrate specificity alteration has a great importance for many potential applications of enzymes. Deacon and her colleagues (2004) investigated whether the GAO variants expressed in *P.pastoris* affect the oxidation of the alternative substrates D-glucose and D-fructose. The study has been focused on D-fructose due to the low activity against D-glucose. It has been reported that R330A shows higher catalytic efficiency when compared to mature GAO. The most active variant against D-fructose was found to be R330K, with a k_{cat}/K_M value of 75.1 $M^{-1}s^{-1}$ which is higher (8.2-fold increase) than that of mature GAO with the same substrate.

Additionally some GAO mutants were developed based on the ones generated by Sun *et al*. (2001, 2002). A library was generated by subjecting M_3 variant (carries the additional mutations Trp290Phe, Arg330Lys, Gln406Thr - all located at the active site - to M_1 variant carrying mutations Ser10Pro, Met70Val, Pro136, Gly195Glu, Val494Ala, Asn535Asp) to more rounds of error-prone PCR. As a result, these mutants show good activity and high enantioselectivity with (R)-1-phenylethanol (Figure 1.2). These variants indicate the improved enantioselective substrate specificity of GAO (secondary alcohol oxidation) by directed evolution (Escalettes & Turner, 2008).

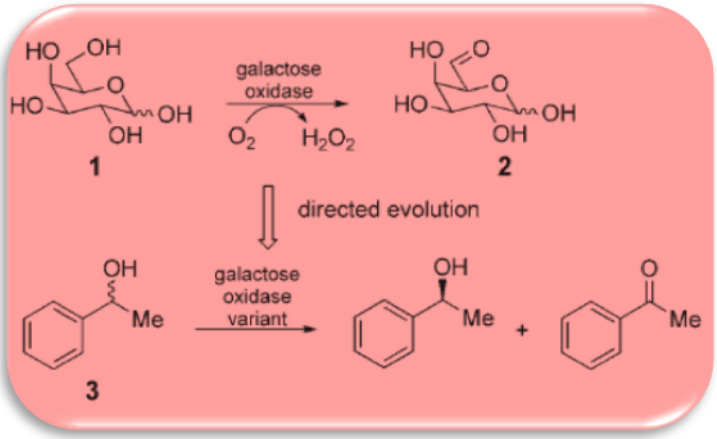


Figure 1.2 Reaction catalysed by wild-type galactose oxidase (above) and enantioselective oxidation of 1-phenylethanol by its variant generated by directed evolution. **1,** D-Galactose, **2,** Aldehyde, **3,** 1-phenylethanol (Escalettes & Turner, 2008)

1.1.5 Galactose Oxidase Applications in Different Areas

Galactose oxidase is important for different application areas due to its unique features such as high degree of substrate specificity and stability, production of H_2O_2, solely copper and dioxygen need for the formation of its organic cofactor (Deacon & McPherson, 2011).

1.1.5.1 Medical Usage and Clinical Assays

GAO-catalyzed oxidation of cell surface polysaccharides is an essential step in the radiolabelling of membrane bound glycoproteins (Calderhead & Lienhard, 1988; Gahmberg & Tolvanen, 1994). The enzyme can be used to detect a disaccharide tumor marker D-galactose-β[1,3]-*N*-acetylgalactosamine (Gal-GalNAc, also known as the Thomsen–Friedrich antigen or T-antigen) (Springer, 1997) in colon cancer diagnosis and precancer conditions by reacting with the terminal galactose residues of the TF antigen to oxidize the C-6 alcohol groups to aldehydes (Yang & Shamsuddin, 1996; Said *et al*., 1999;Vucenik *et al*., 2001). Novel quantitative galactose oxidase–schiff reaction (Evelegh, 2002) was also used in NAF (nipple aspirate fluid) samples from women with breast carcinoma in order to distinguish the affected breast from its healthy contralateral counter-part, in which further exploration is needed (Chagpar *et al*., 2004).

GAO is also used for induction of interferon in human lymphocyte culture (Dianzani *et al*.,1979).

Methods have also been developed to detect galactose and other galactose oxidase substrates in biological fluids such as blood and urine (Deacon, 2008). These methods, both measure the formation of H_2O_2, they are carried out by using UV-vis spectrometry (Roth *et al*., 1965) or a biosensor (will be detailed in Section 1.1.5.4), using a thin layer electrochemical cell in which galactose oxidase was immobilised (Johnson *et al*., 1982; Johnson *et al*., 1985).

GAO is also used as a part of a novel antibacterial system for an *in vitro* anti-plaque system. GAO-GBD (glucan binding domain) fusion protein is expressed in *E.coli* and GBD part of the coupling acts as targetting component whereas GAO plays an indirect role which does not have a bactericidal effect but has an effector activity that stimulates the local production of hydrogen peroxide onto the addition of D-galactose (Lis & Kuramitsu, 1997). Hydrogen peroxide is utilized by lactoperoxidase in saliva, to oxidize added iodide to hypoiodite (Majerus & Courtois, 1992; McFaul, 1986) which is an effective bactericidal agent acting on cariogenic microorganisms on tooth surfaces. This may represent a potential system for controlling dental plaque formation in humans. Lis & Kuramitsu (1997) also reported that this strategy could be further improved to fuse GAO to other targetting domains and use in other parts of the body as an antimicrobial agent.

1.1.5.2 Industrial Applications: Food, Drug and Material

Aldehyde derivatives, the oxidation products of GAO are used as strength additives in the paper industry (Hartmans, 2004; Frollini 1995).A natural polymer, guar which is present in the seeds of the guar plant and other sources can give a final product of an aldehyde-bearing polymer called oxidized guar. This resulting product is a valuble additive used for increasing the paper strength (Chiu *et al.*, 1996; Dasgupta, 1996; Brady & Leibfried, 2000).

Unmodified guar gum is used as a thickening or gelling agent, for making food appealing to the consumer and maintaining turbidity in soft drinks and juices (Wielinga, 2000). According to the Delegrave and colleagues (2001), the activity of GAO towards guar needs to be paid special attention and the enzyme was aimed to be improved by directed evolution to increase its specific activity, which is insufficient to act economically at an industrial scale when it is isolated from its native host.

Galactose oxidase is used as a biocatalyst for the production of nonnutritive sugar substitutes in food. (Mazur, 1991). Enzymatic carbohydrates synthesis by GAO does not require the protection of the hydroxyl groups (Root *et al.*, 1985; Mazur & Hiler, 1997; Liu & Dordick, 1999).

Aldehyde derivatives of several structurally different polysaccharides like O-acetyl galactoglucomannan, guar gum and tamarind seed xyloglucan were produced and high oxidation degrees were attained for understanding the relationship between structural and quantitative properties of subtrate and enzyme couple used and the products formed (Parikka *et al.*,2010).

Under benign environmental conditions, non-natural carbohydrate polymer construction that is closely related to natural products is a synthetic use for galactose oxidase (Andreana *et al.*, 2002).

Galactose oxidase has been engineered by directed evolution methods to possess glucose 6-oxidase activity (Sun *et al.*, 2002). Until now, oxidizing glucose at the 6-hydroxy group generating aldehyde is not found in any described oxidase. Three mutations within galactose oxidase, certifies low but significant glucose oxidase activity. This acquired ability provides

a good improvement basis for the new enzyme activity, which may supply potential benefits for other industries such as food, pharmaceutical and material. (Rogers & Dooley, 2003).

By directed evolution method, the structure of the active site of galactose oxidase was manipulated and the generated variants were able act on carbohydrates other than galactose, including mannose and N-acetylglucosamine (Rannes *et al*., 2011). This strategy allows the broad investigation of glycoproteins with a variety of carbohydrate structures.

GAO is used in hair dyes (Sørensen, 2002) and lipid labeling (Radin & Evangelatos, 1981) as well.

1.1.5.3 Taxonomic Studies

Since only *Fusarium* species are reported to be a GAO producer, (taking into consideration that the other *Fusarium* sp. are not tested or the methodology was not suitable for them) and it is not clear that the other other fungal species proposed to have GAO activity belong to the same type of enzyme as *Fusarium sp*. have described in Section 1.1.2; GAO production can be used as an important taxonomical marker in future studies as indicated in the study of Barbosa-Tessmann *et al.*(2001).Additionally these findings together with micotoxin production (Machado & Kemmelmeier, 2001) may represent a model for molecular taxonomy studies.

A PCR protocol amplifying the 3'coding region of the *gao*A gene coding for galactose oxidase enzyme was developed by Biazio *et al* (2008), which works on a low homology basis with the same region of GAO genes among other fungi. Among the tested 17 strains, the ones morphologically and molecularly identified as *Fusarium* spp. gave positive results by this PCR protocol with the designed primer set and they were able to produce GAO. PCR analysis with the genomic DNA of other *Fusarium* species and species of other fungi genera gave negative results with this PCR assay. This proposed PCR protocol which seems to be specific for *gao*A gene can be used as a molecular tool for *Fusarium graminearum* identification, providing an advance for head blight diagnosis and food safety assessment (Biazio *et al*., 2008).

1.1.5.4 Biosensors

Especially clinically, galactose detection and concentration determination have a great importance due to its vitality for human health. It is generally used for the preliminary diagnosis of galactosemia and galactose intolerance (Beutler, 1991; Sharma *et al*., 2004a; Szabó *et al*., 1996; Segal *et al*., 2006; Berry *et al*, 2004; McGlothlin &Purdy, 1977; Çevik *et al.*, 2010). Gas or liquid chromatography and enzymatic assays (electrochemical, spectrophotometric and fluorometric) with galactose oxidase, the most commonly used enzyme for galactose determination, were the methods used for galactose detection (Tkac *et al*., 1999; Brahim *et al*., 2002; Sharma *et al*., 2006; Suang & Bae, 2006; Yokoyama *et al*., 1997; Shaolin, 1994; Bílková *et al*., 2002; Yang *et al*., 1998).

Galactose oxidase has numerous applications in biosensors. Sensors incorporating GAO have been used to measure D-galactose, lactose and other GAO substrate concentrations (Vega *et al*., 1998; Tkac *et al.*, 1999), in process monitoring (Szabó *et al*., 1996), blood samples (Vrbova *et al*., 1992), quality control in the dairy industries (Adanyi *et al*., 1999; Mannino *et al*., 1999) and other biological fluids (Johnson *et al.*,1982).

Immobilized galactose oxidase was used in various amperometric biosensors (Taylor *et al*., 1977; Manowltz *et al*.,1995; Suang & Bae, 2006; Sharma *et al*., 2004b; Ekinci & Pasahan, 2004) where GAO was immobilized on the surface of various membranes and films such as acetyl cellulose, nylon mesh, polypyrrole, polyaniline, polyvinyl formal (Sharma *et al*., 2007) polyvinylferrocene, poly (glycidyl methacrylate-co-vinylferrocene) (Çevik *et al*, 2010) and poly(N-glycidylpyrrole-co-pyrrole) (Şenel *etal*., 2011) or trapped between the membrane and electrode. Additionally, a new galactose biosensor was developed based on functionalized zinc oxide (ZnO) nanorods (Khun *et al*, 2012) (Figure 1.3).

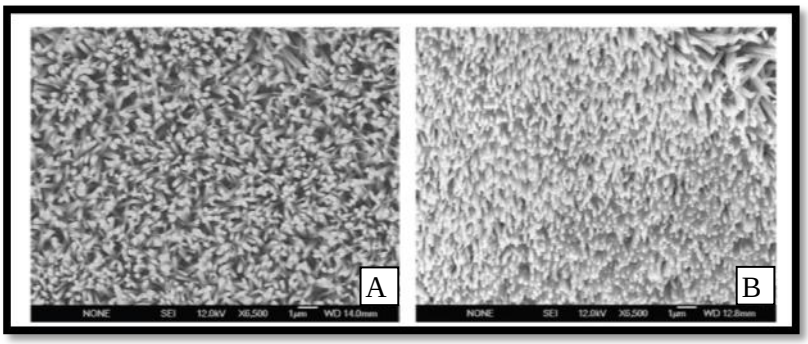


Figure 1.3 SEM images of ZnO nanorods grown on glass substrates coated with gold **(A)** The ZnO nanorods only and **(B)** ZnO nanorods with immobilized GAO enzyme (Khun *et al*, 2012).

1.1.5.4.1 Biomolecule Detection

Currently, antibody-based biomolecule detection systems are widely used. They are large, complicated, highly specific and have various binding characteristics, while they are relatively unstable molecules whose production and performance are not particularly reproducible. On the other hand, small, robust, stable antibody-mimetic proteins are used and they can be produced by simple, bacterial recombinant protein expression by which they can be purified reproducibly and in high yield.

1.1.5.4.2 Principle of the Split-GAO Amplification Systems

A new project about development of split enzyme amplification systems for electronic label-free biosensors is in operation at University of Leeds. It has been attempted to improve the biosensors for the biomolecules developed in University of Leeds by enhancing their sensitivity and specificity through engineering a signal amplification mechanism.

According to this project, two antibody-mimetic proteins which bind non-overlapping epitopes on the target biomolecule will be immobilized onto an electrode surface. Each binding protein (B1 and B2) will be coupled, to the electrode surface and to one half of a split enzyme (Ea and Eb) via flexible linker sequences (Figure 1.4A). No GAO activity will be detected if the two halves of the enzyme, here Ea and Eb, do not associate (neither half is functional alone) and this will only occur once both binding proteins, B1 and B2, have bound to the same target molecule, bringing the enzyme fragments close enough to function. When there is substrate, catalytic turnover generates products that can be detected electrically. In response to the detection of a single target molecule enzymatic activity is stimulated. This leads to multiple turnover events of the enzyme which means the amplification of the initial binding signal. This split-enzyme system leads to an important improvement in label-free detection technologies by increasing sensitivity and specificity.

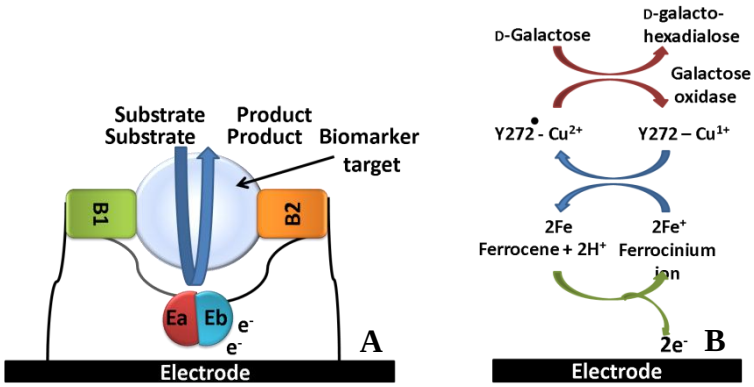


Figure 1.4 Principle of the GAO split-system based biosensor approach. (**A)** Target molecule detection by two binding proteins (B1 and B2) each linked to a part of split galactose oxidase which can only come together when brought into proximity by the binding events. Newly regenerated enzyme provides improved signalization by multiple catalytic turnover events. (**B**) Schematic representation of the reaction for amperometric detection of electrons via GAO enzyme activity, based on the well characterised glucose oxidase example.The two electrons were transferred to the electrode surface via an enzyme-tethered mediator such as ferrocene (Willner, *et al*.,1993).

1.1.5.4.3 A Split Redox Enzyme Formation by Dissection of Galactose Oxidase

In this thesis study, preliminary work will be put into practice by protein engineering approach as a part of this new project in operation at University of Leeds. Two separate part of the enzyme, which are domain I/II and domain III, will tried to be produced and detected by deletion studies over mature GAO construct. Molecular design of the mutations and region targeted to dissect the GAO into two separate molecules will be given in detail in Chapter 3, Results and Discussion.

1.2 Features of GAO on a Structural and Functional Basis

1.2.1 Properties of *gao* Gene

Galactose oxidase encoding gene *(gaoA)* from *Dactylium dendroides*, reclassified as *Fusarium sp.* (NRRL 2903) by Ögel and colleagues (1994), has been cloned and its DNA sequence has been determined. The *gaoA* gene does not contain any introns, in opposition to the most other filamentous fungal genes (Gurr *et al.*, 1987). The whole coding region together with 120 bp N-terminal leader peptide is ~2.0 kb. The mature enzyme as a final translation product consists of 639 amino acids (McPherson *et al.*, 1992; Ito *et al.*, 1991).

1.2.2 Glycosylation of GAO

In general, *O*-and/or *N*-glycosylation of eukaryotic extracellular proteins is carried out during the travel through the endoplasmic reticulum and Golgi, where the protein undergoes higher glycosylation levels in extracellular medium when compared to its intracellular forms.

Conversely, studies (Mendonca & Zancan, 1987,1988) show that galactose oxidase is likely to lose its carbohydrate moiety during secretion. Intracellular galactose oxidase with about 9% carbohydrate content is found to be more stable and to have more restricted substrate range than its extracellular form which contains only 2% carbohydrate. On the other hand, Kosman and colleagues (1974) reported that the enzyme contains 1% by weight of neutral carbohydrate.

X-ray structural analysis of extracellular GAO does not reveal any carbohydrate (Ito *et al*., 1991). However, this does not preclude its presence. Undetectability may result from the poorly ordered surface carbohydrate due to mobility and heterogeneity effects. When the translated protein sequence was analysed it has been seen that GAO is only *O*-glycosylated, because of the absence of consensus Asn-X-Ser/Thr sequences for *N*-glycosylation (Kornfield & Kornfield, 1985). This observation is contrary to the report that the N-glycosylation inhibitor, tunicamycin, affects the addition of [^{14}C] glucosamine and the rate of the migration of the intracellular galactose oxidase in SDS-PAGE (Mendonca & Zancan, 1989). However, McPherson and colleagues (1992) reported that the mobility difference is not probably due to inhibition of *N*-glycosylation, but may indicate an indirect effect on the normal processing of the enzyme and non-covalent interactions with sugars. On the other hand, it cannot be overlooked that *D. dendroides* might be using a different and unprecedented sequence motif for *N*-glycosylation.

Later in 2001, the structural studies of Firbank and colleagues revealed that the precursor GAO contains an additional sugar at the surface of domain II. According to the electron density, it is at least a disaccharide; with the first ring stacking almost parallel to the ring of Tyr-484, however it has not been identified and its biological role is uncertain due to the distance of sugar to the active site.

1.2.3 GAO Enzyme: Domains and Ligands

The detailed crystal structure of the enzyme revealed that the mature enzyme consists of three predominately β-sheet domains with only a single α helix (Ito *et al.*, 1991,1994), which are Domain I, II and III (Figure 1.5). It is probable that this preponderance of β structure of the domains plays a role in the remarkable stability of the enzyme, which is active in 6 M urea (Kosman *et al.*, 1974).

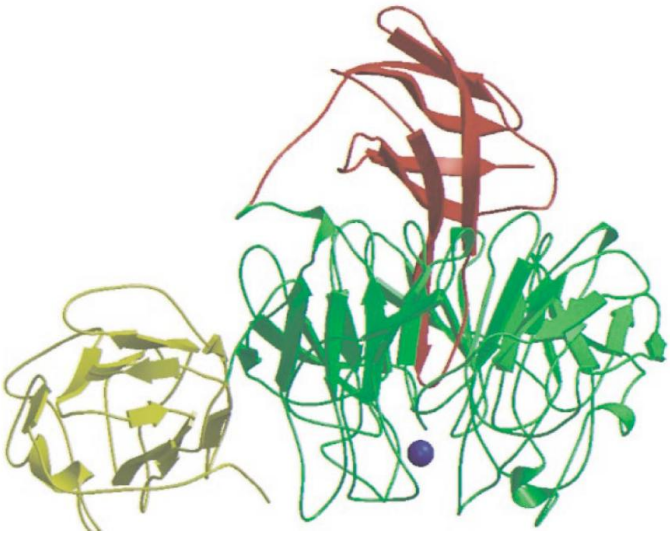


Figure 1.5 Schematic representation of galactose oxidase. The complete molecule is coloured according to domains (the first domain in yellow, the second green and the third red) (McPherson *et al.*, 2001).

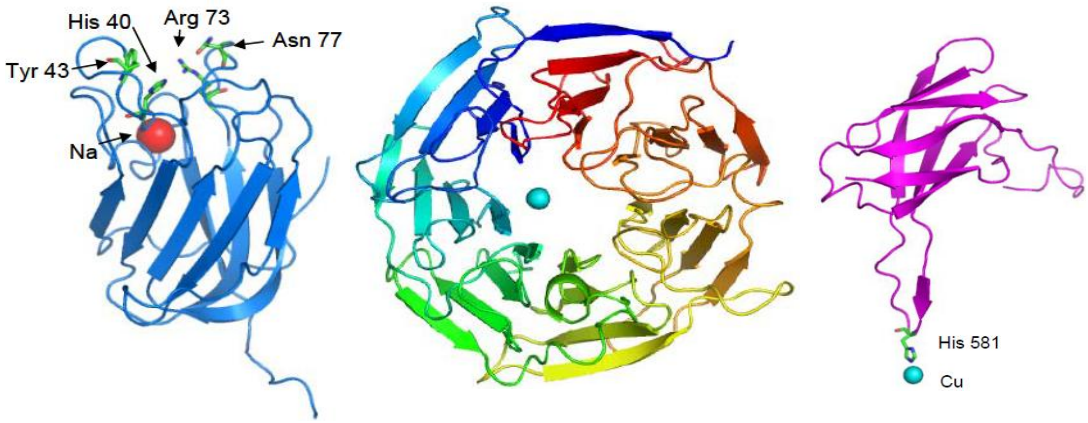


Figure 1.6 Separate display of domains of galactose oxidase **A.**The fold of domain I (Residues 1 - 155) is represented as a ribbon diagram shown in blue. Residues implicated in carbohydrate binding are shown in green, a bound sodium ion is shown as a red sphere.**B.** Folding of domain II (residues 155-532), viewed through the pseudo sevenfold axis, coloured from N-terminus blue to C-terminus red. **C.** The fold of domain III (residues 533-639), the copper ligand at the tip of the loop passing through the domain II, His 581, is shown in green.The copper is indicated as a cyan sphere (Deacon, 2008).

Domain I is composed of eight β-strands in a jelly-roll motif with a five-stranded antiparallel β-sheet facing a three-stranded antiparallel β-sheet, which is described as a "β sandwich" (Firbank *et al*., 2001). This domain contains a cabohydrate binding site and potentially hydrogen bonding residues with D-galactose, 2-deoxygalactose, D-fucose and D-xylose are all highlighted in Figure 1.6 A (Firbank, 2002). This site may allow adhesion of the enzyme to polysaccharides found in the cell walls of plants (like lignin) or to the cell surface of potential bacterial pathogens in cell protection.Carbohydrate binding site is not close to the active site, and since then is not thought to be crucial for enzyme activity.

There is also a non-spesific small metal-binding site assigned as a sodium ion that might play a role in protein folding. An ordered polypeptide chain links the domain I and domain II (Deacon, 2008).

Domain II is the largest and it provides three of the four protein ligands to the copper, Tyr-272, Tyr-495, and His-496.The seven-bladed β-propeller fold surrounds a central cavity (Figure 1.6 B) and the active site of the mature protein is situated at the surface of this domain, with the copper ion lying close to the central pseudo 7-fold axis (Firbank *et al*., 2001). The sequence forming the β-propeller structure in GAO has been described as the kelch motif (Bork & Doolittle, 1994). Additionally, closing of the β-propeller is likened to 'Velcro' fastening; this structure is thought to provide rigidity of the enzyme via the hydrophobic interactions between the β-strand of the N-terminus interacting with the last β-strand of the C-terminus (see Figure 1.6 B).

About 50% of the β-propeller proteins reported are enzymes and they are responsible for the catalysis of various reactions. The β- propeller motifs can act as scaffolds that can acquire specific activities by decoration of flexible loops with functional amino acids (Deacon, 2008).

Domain III is a bundle of seven, mostly antiparallel, β-strands surrounding a hydrophobic core. One long antiparallel β-ribbon passes through the central cavity of the domain II propeller, and there is His-581 residue, the fourth protein ligand to the copper at the tip of this loop axis (Firbank *et al*., 2001). The gap between this loop and domain II is occupied by some well-ordered 40 water molecules. It is thought that domain III has a structural key role, which is stabilising domain II (Deacon, 2008).

There are two disulfide bonds which includes the residues close to each other in the amino acid alignment. Cys18 and Cys27 are in domain I whereas Cys515 and Cys518 are in the terminatory part of domain II and reside close to the active site (McPherson, *et al*. 2001).

1.2.4 Locations and Interactions of the Residues of the Pro-sequence

Pro-GAO was expressed and produced in *A. nidulans* and then purified and crystallized in copper-free conditions (Rogers, 2000). The structure of the precursor enzyme was rebuilt and refined to a 1.4 Å by molecular replacement using the structure of the mature enzyme as a model (Firbank *et al*., 2001).

The 17-aa N-terminal pro-sequence resides between domains I and II of GAO. Residues –1 to –12 are visible in the final electron-density maps. These residues form a loop that turns back at the residues -5/-6 directing the pro-sequence away from the main body. The remaining residues (–13 to –17) and the side chain of the twelfth residue (Ile -12), are not visible (Firbank *et al*., 2001).

The main chain of the pro-sequence makes a number of hydrogen bonds to the mature enzyme sequence, to itself especially at turn residues and side chains also form hydrogen bonds to several residues (see Figure 1.7). Apart from these, there are 14 hydrogen bonds to water molecules (Firbank *et al*., 2003).

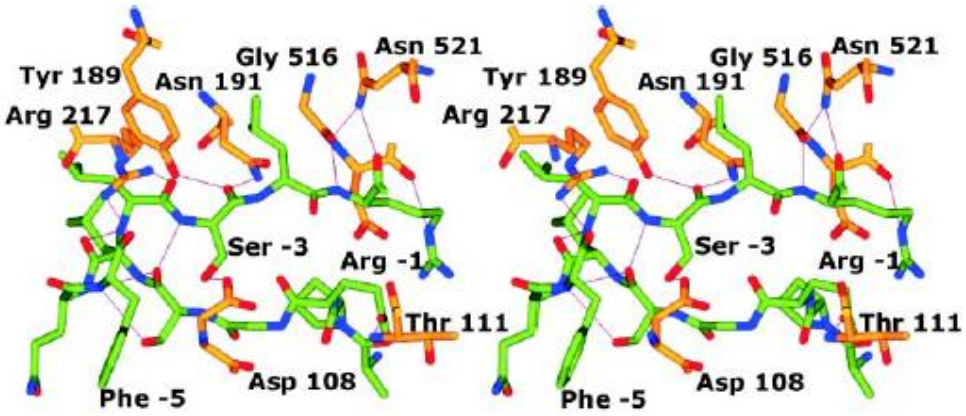


Figure 1.7 3-D structural representations of the pro-sequence residues with the amino acids which they form hydrogen bonds. The carbon atoms of the pro-sequence are green, whereas mature sequence carbon atoms are yellow (Firbank *et al.*, 2001).

1.2.5 Structural Differences between the Mature and the Precursor Enzyme

Although the overall structure of the copper-free precursor is similar to the mature enzyme (Protein Data Bank ID code 1GOG), local differences are remarkable between them .While the overall similarity (without pro-sequence) show a rmsd of 0.7 Å, domain II exhibits the highest deviation. Residues colored yellow in Figure 1.8 which include the five regions of the precursor main chain (1–5, 189–200, 216–225, 255–261, and 290–296) show a rmsd of 5.1Å. Important rearrangements were detected at some of the side chains and adjacent main-chain regions of the active site residues. Moreover, the position of Ala-1 differs by over 4 Å from its original location when the pro-sequence is present (Firbank *et al.*, 2001).

One of the five regions that change position in the precursor form is the one consists of the residues from 1-5. The presence of the pro-sequence prevents the second loop (189-200) from taking its original position in the mature protein. The third region (216-227) significantly differs between the two forms, differing by up to 6.1Å from the mature protein at main-chain positions. The differences including hydrogen bonding result in a change in the position of the fourth region, corresponding to the adjacent β-strand (residues 254-261).The residues in this strand section are not directly interacting with the pro-sequence.

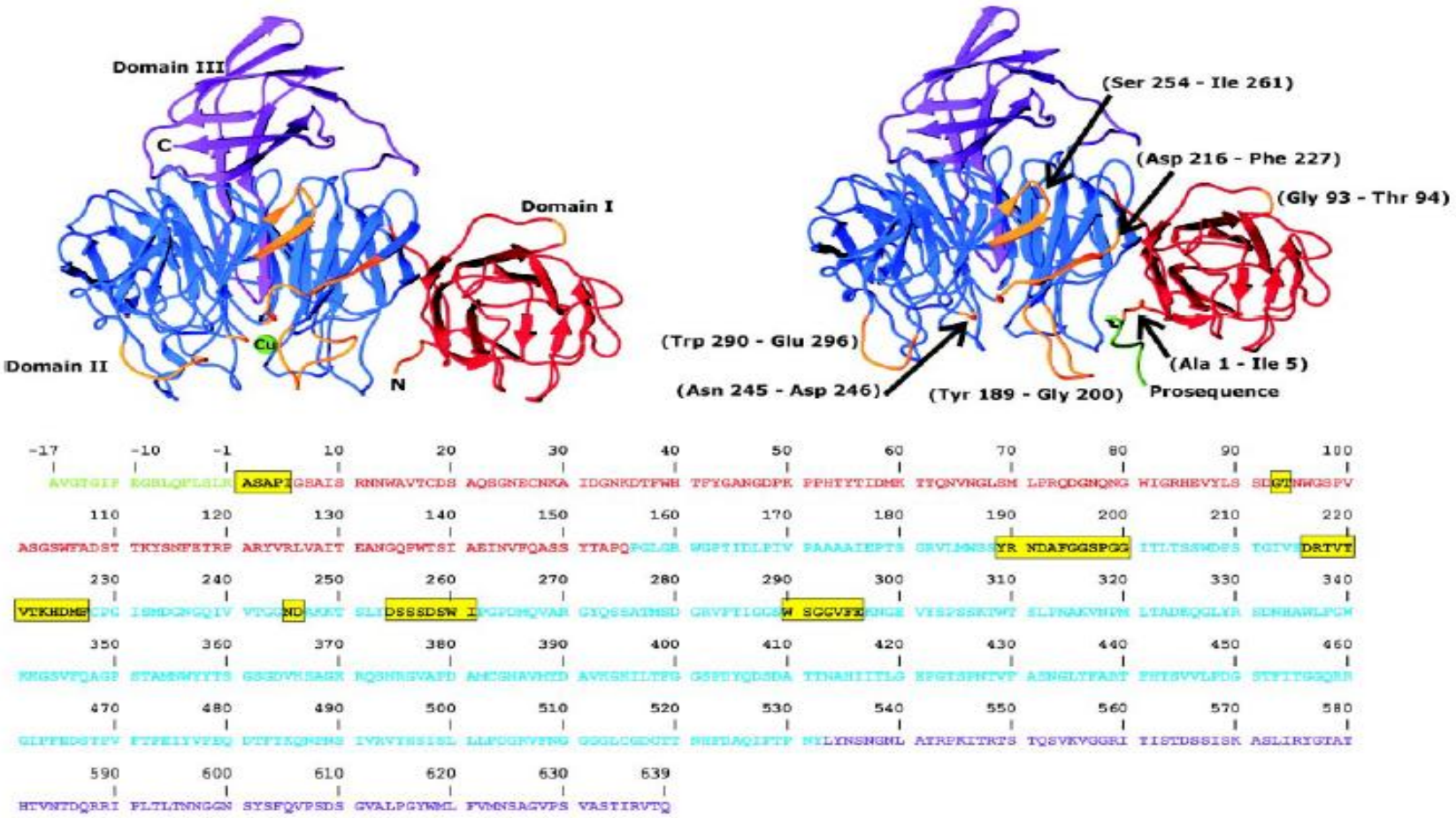


Figure 1.8 Structure of the mature form of galactose oxidase (*Upper Left*) and the pro-galactose oxidase (*Upper Right*). Domain I is coloured with red, domain II with blue, and domain III with purple. In the precursor form the N-terminal pro-sequence is green, and residues that deviate from the mature form by more than 2 Å structurally are yellow. The amino acid sequence of galactose oxidase (McPherson *et al.*, 1992) is shown (*Lower*) and coloured in compliance with the figure above. The pro-sequence residues are numbered -17 to -1. (Firbank *et al.*, 2001)

The fifth region forms a loop (residues 290–296) containing the active site residue Trp-290, stacking directly on top of Tyr-272 and the thioether bond. The C_α backbone between Trp-290 and Glu-296 differs by up to 8.0 Å (suggesting high mobility or disorder). The active site is more accessible due to the loop rearrangement affecting the stacking Trp-290 in the precursor protein (Figure 1.9). The thioether bond forming residues, Cys-228 and Tyr-272 in the mature enzyme, are exposed because of the absence of Trp-290. However, the thioether bond is not present in the precursor form, the side chains of both residues have quite different rotamers from their mature form. Main-chain movements of these two are smaller, with C_α of Tyr-272 and Cys-228 moving by 1.2 and 1.4Å, respectively.

The other three protein ligands to the copper, Tyr-495, His-496, and His-581, have main-chain C_α movements of less than 0.4 Å. In the precursor, the electron density at Cys-228 shows additional electron density, suggesting oxidation to an S-OH group (Figure 1.9 C). It is not determined yet that whether this apparent oxidation is an intermediate product in thioether bond formation (tyrosine might attack onto an electrophilic sulfur) or is an artifact as a result of crystallization or radiation damage (Firbank *et al.*, 2001).

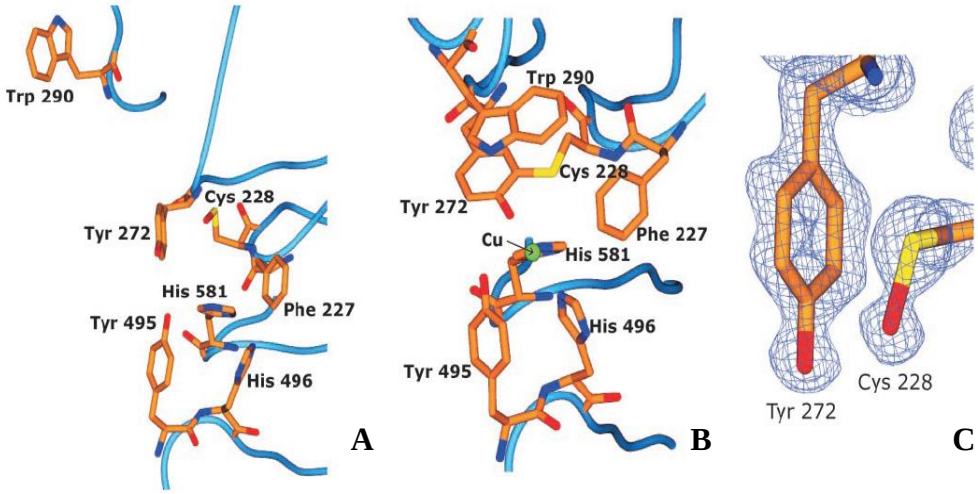


Figure 1.9 Stereoview of active site residues in precursor galactose oxidase **(A)** and in the mature protein **(B)** including the copper ligands (Tyr-272, Tyr-495, His-496, and His-581), thioether bond partners, Cys-228 and Tyr-272, the stacking Trp-290, and Phe-227. The sulfur of the cysteine appears to have been oxidized to a sulfenic group in the precursor form (**C**).

1.3 Protein-derived Cofactors in Enzymes

Various enzymes use varieties of cofactors generated by complex multistep biosynthetic pathways in order to exhibit their catalytic powers. In cofactor biosynthesis, a less commonly encountered one, but with an increasing frequency, is the posttranslational modification of the endogenous amino acids in the enzyme. Most part of the protein-derived cofactors remained unrecognized until recent years since they cannot be identified from the gene sequence and for most of the known protein-derived cofactors, there is no known sequence motif which can be used for the prediction of the modification site (Davidson , 2007). By the discovery of these cofactors, their formation mechanisms have taken great attention. These

modifications can be classified in two groups. One of these posttranslationally modified protein groups, cofactors are generated by one-electron oxidations of amino acids yielding amino acid radicals on tyrosine, glycine, tryptophan and cysteine residues (Stubbe & van der Donk, 1998). The second group undergoes more extensive posttranslational modifications that involve covalent modification of a residue, peptide cleavage and new bond-forming reactions (Rogers *et al*., 2008). Posttranslational modifications occur either by autocatalytic processes or by involvement of accessory proteins (Xie & van der Donk, 2001). The most frequently modified residue appears to be tyrosine which becomes part of cross-links or modified as quinones. These cofactors are generally catalytically active and associated with metal centers.

Examples carrying posttranslationaly modified cofactors include; histidine decarboxylase, cytochrome *c* oxidase, catalase HPII, tyrosinase, catechol oxidase, haemocyanin, phenylalanine ammonia lyase, histidine ammonia lyase, sulphatase, methylamine dehydrogenase, copper amine and lysyl oxidases, and galactose and glyoxal oxidases (Okeley & Van Der Donk, 2000).

Including Tyr-Cys thioether cross-link in galactose oxidase (Rogers, 2000), in copper amine oxidases, 2,4,5-trihydroxyphenylalanine quinone (TPQ)(Janes *et al*., 1990; Cai & Klinman, 1994**)**) in phenylalanine ammonia lyase the MIO structure (Schwede *et al*., 1999), and the chromophore in green fluorescent protein (Heim *et al*., 1994) are produced by autocatalytic processes whereas tryptophan tryptophyl quinone (TTQ) in methylamine dehydrogenase (Graichen *et al*.,1999) and formylglycine in sulfatases (Szameit, *et al*.,1999) are generated by accessory proteins . Histidine at Y-H bond of cytochrome c oxidase at the CuB site in both bacteria (Ostermeier *et al.*,1997) and mammals (Yoshikawa *et al.*, 1998), acts as a a copper ligand, and a tyrosyl radical at this location has been implicated to take part in catalysis (Proshlyakov *et al*., 2000). In *Escherichia coli* catalase HPII, His and Tyr residues also form a cross-link, however the bonding is between the C_β of tyrosine and N_δ of histidine (Bravo *et al.*, 1997). The Y69-C161 cross-link of *Mycobacterium tuberculosis* sulfite reductase (Schnell *et al*., 2005) is reported to play a role in catalysis whereas cysteine-tyrosine cross-link of mouse (McCoy *et al*., 2006) and rat (Simmons *et al*., 2006) cysteine dioxygenase are proposed to act in enzymatic turnover. Glyoxal oxidase also contains a Cys-Tyr bond. KatG has an autocatalytically and oxidatively (via compound I) formed M-Y-W cross-link adjacent to the heme cofactor (Yamada, *et al*., 2002) that forms during a turnover reaction (Ghiladi, *et al*., 2005) conferring to catalase activity to the catalase-peroxidase enzyme (Smulevich, *et al*., 2006). The most complex cofactor described up to now is the unusual cysteine tryptophylquinone that is encaged by three novel Cys to Asp or Glu thioether linkages found in a quinohaemoprotein amine dehydrogenase (Satoh *et al.*, 2002).The mechanisms of the formation of the posttranslationally modified cofactors have not been fully elucidated yet. A better understanding of the mechanisms of the biogenesis of these cofactors would make it possible to have an idea on introducing catalytic or redox centers within an existing protein to form protein- derived cofactors through amino acid sequence design or modification regardless of additional exogenous cofactor requirement (Davidson, 2007).

1.4 Cofactor Formation and Self-Catalytic Maturation Capability of the GAO

1.4.1 Active Site and Generation of the Thioether Bond

Galactose oxidase is the best-characterized member of a family of enzymes known as radical copper oxidases (Whittaker, 2003) which contains a single copper atom at its active site (Amaral *et al*., 1963; Ito *et al*.,1991). This metal ion is bound by four amino acid side chains: two tyrosines (Tyr272 and Tyr495) and two histidines (His496 and His581) which are ligands to the copper (Ito *et al*., 1991) (Figure 1.10). X-ray crystallographic studies on galactose oxidase have revealed that the protein contains a novel cofactor constituted between Cε carbon of the phenolic side chain of Tyr272 and the Sγ sulfur of Cys228, forming tyrosyl–cysteine (Tyr–Cys) cross-link (Ito *et al*.,1991). Galactose oxidase is unusual among metalloenzymes in terms of catalyzing two-electron redox chemistry at a mononuclear metal ion active site (Kosman, 1985; Ettinger and Kosman, 1982; Hamilton, 1982). The second redox active centre required for the reaction was found to be located at this Tyr–Cys site and the cross-linked tyrosine is oxidized to a stable free radical upon mild oxidation (Whittaker & Whittaker, 1990). Cys-Tyr$^{\cdot}$ cofactor abstracts a hydrogen atom from the substrate interacting with the copper, however less is known about the role of the thioether bond (Whittaker & Whittaker, 1990; Babcock *et al*., 1992; Gerfen *et al*., 1996). Besides affecting the reactivity of the protein it contributes to the rigidity of the active site (Whittaker, 2005).The function of this bridge has been attributed to lowering the tyrosine redox potential to 400 mV (Johnson *et al*., 1985) from ~ 1000 mV for an unmodified tyrosine residue.

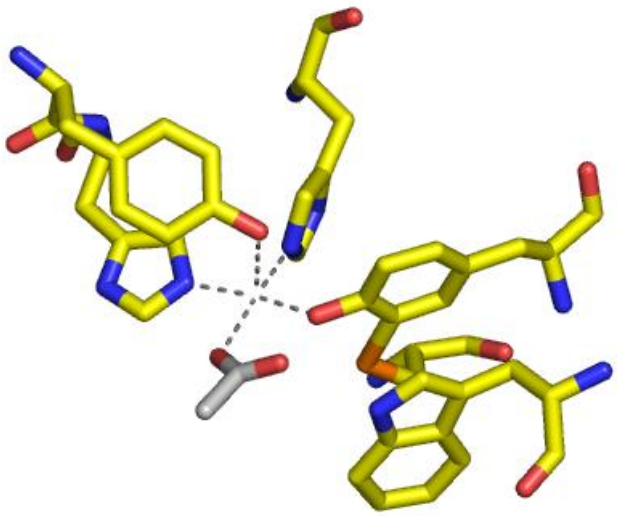


Figure 1.10 A diagram showing the four ligands and the stacking tryptophan to the copper in a square pyramidal geometry of galactose oxidase. The acetate ion from the crystallisation buffer (Ito *et al*., 1991) is thought to indicate the binding of the substrate position.

When the thioether bond formation mechanism is considered, it was initially proposed that thioether bond was formed by involvement of initial coordination of Tyr272 to Cu^{2+} and subsequent formation of tyrosine radical (Firbank *et al*, 2001). However, later studies on this mechanism carried out by the *P.pastoris*-expressed mature GAO under strict copper-free conditions revealed that the mechanism rather prefers the initial coordination of Cys228 and Cu^{2+}, indicating the anaerobic formation of the thioether bond. It seems more likely that

thioether bond formation occurs in the trans-Golgi network and it is not oxidized and transformed into the tyrosyl radical until it reaches to the more oxidizing environment (Rogers *et al*., 2008). On the other hand, the second postulated mechanism over the early oxidation of Cys228, attributed to the sulfenate modification (see Section 1.2.5), is not compatible with the anaerobic processing, supporting the suggestion that occurrence of electron density is an artifact due to X-ray exposure (Rogers *et al*., 2008).

1.4.2 Oxidation States and Putative Catalytic Mechanism of the GAO

Galactose oxidase is the first representative of a new class of copper active sites, incorporation of the protein in active redox chemistry. In GAO, the number of metal ions involved in the reaction is not compatible with the number of electrons transferred. This paradox has been solved with the identification of a tyrosyl free radical directly participating into the redox unit during the catalytic cycle. It has three possible redox states in order to carry out the two electron oxidation of alcohols to aldehydes as shown in Figure 1.11 (Whittaker & Whittaker, 1988b).

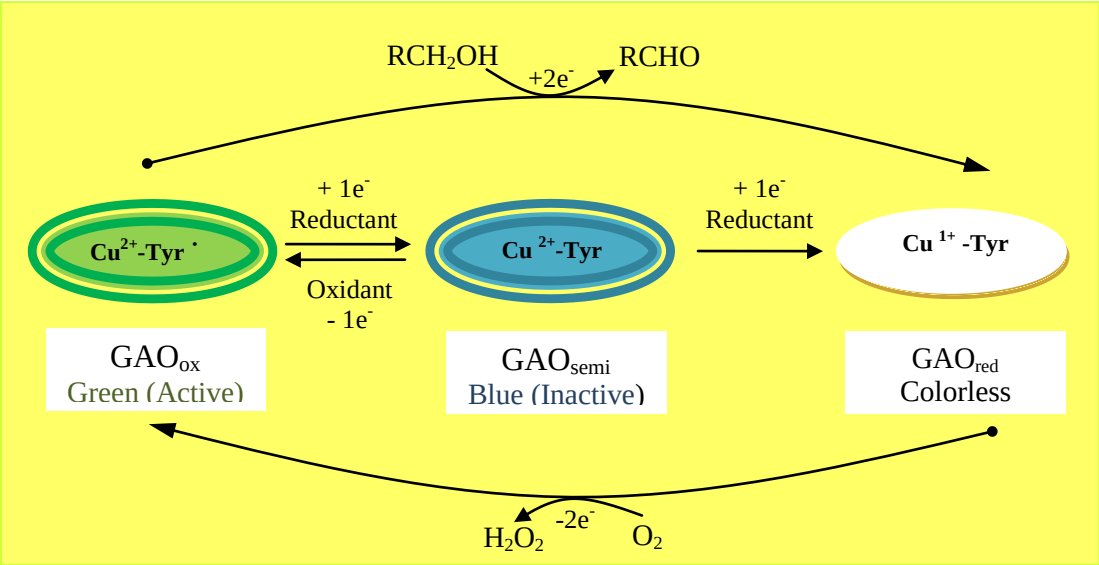


Figure 1.11 Schematic diagram representing the redox states of galactose oxidase. GAO_{ox} represents the fully active form of the enzyme, GAO_{semi} is a partially oxidised (semi-reduced) but inactive form and GAO_{red} is the reduced form of the enzyme, e^- represents an electron (adopted from Whittaker &Whittaker, 2000).

Oxidation step is necessary for catalytic activity, and is the basis of two-electron redox activity (Whittaker &Whittaker, 1988a). Although the GAO_{semi} form is not involved in the catalytic cycle, it is important for a full insight of the properties of the enzyme (Messerschmidt, 1994).

GAO is obtained as a mixture of semi-reduced and oxidized (only a small amount) form when it is isolated. Incubation with a mild oxidant such as ferricyanide yields fully oxidised

galactose oxidase (Whittaker & Whittaker,1988a) whereas incubation with the reductant ferrocyanide results in the reduced or semi-reduced forms of galactose oxidase.

Although the mechanism of catalysis is not completely understood, a bi-bi ping-pong mechanism has been proposed by early studies on galactose oxidase (Hamilton *et al.*, 1973).

The analysis of the reaction is complicated by occurrence of two distinct half reactions in a ping-pong kinetics following enzyme, each involving distinct bond-making and bond-breaking processes (Messerschmidt, 1994; Whittaker, 2005) (see Figure 1.12).

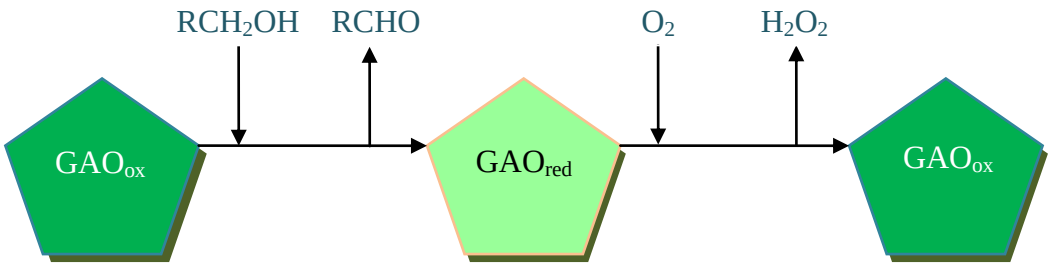


Figure 1.12 Schematic drawing of two half reactions of galactose oxidase reacting with primary alcohols. GAO_{ox} and GAO_{red} represents oxidized and reduced galactose oxidase, respectively (adopted from Deacon, 2008)

The proposed catalytic mechanism for GAO is shown in Figure1.13 in a stepwise manner (Whittaker, 1998). For proceeding of the catalysis, the enzyme must contain both the radical and Cu^{2+} (fully oxidised form). First of all, it is thought that substrate binds to the copper, with displacement of the equatorial solvent ligand (e.g. H_2O). Then the proton is transferred from the hydroxyl group of the alcohol substrate to the phenolate oxygen of Tyr 495, acting as a catalytic base (step (1) to (2) in Figure 1.13). Wild-type enzyme and mutagenesis studies provide strong evidence for this uptake (Whittaker & Whittaker, 1993; Reynolds *et al.*, 1995). Next, hydrogen atom is thought to be abstracted from the substrate intermediate by the Tyr 272$^{\bullet}$, resulting in quenching of the tyrosine radical and formation of substrate radical (step 2 to 3 in Figure 1.13).The abstracted hydrogen has been shown to be the *pro*-S hydrogen (Minasian *et al.*, 2004). One electron transfer to the copper center leads to the oxidation of the substrate-derived ketyl radical, generating Cu(I) and aldehyde product (steps 3 to 4 in Figure 1.13). Product aldehyde is released and molecular oxygen is bound to the Cu^{1+} in an endon coordination (steps 4 to 5 in Figure 1.13).Electron transfer gives rise to superoxide adduct formation (steps 5 to 6 in Figure 1.13) and oxidation of Cu^{1+} to Cu^{2+}. Following the electron and hydrogen atom transfer reforming the radical on Tyr 272 and generation of a peroxide anion species, proton is transferred from Tyr 495 to the peroxide anion and hydrogen bond isformed between Tyr 495 and Cu^{2+} (step 6 in Figure Figure 1.13), the resulting product hydrogen peroxide is released and the enzyme is ready for the new cycle (step 6 back to 1 in Figure 1.13) (Deacon, 2008). The metal ion, acting as a one-electron storage site, and Tyr–Cys cofactor, together form an efficient two-electron catalytic complex that can be turned on and off by reversible one-electron reduction, which controls the enzyme activity in the extracellular space (Whittaker, 2005)

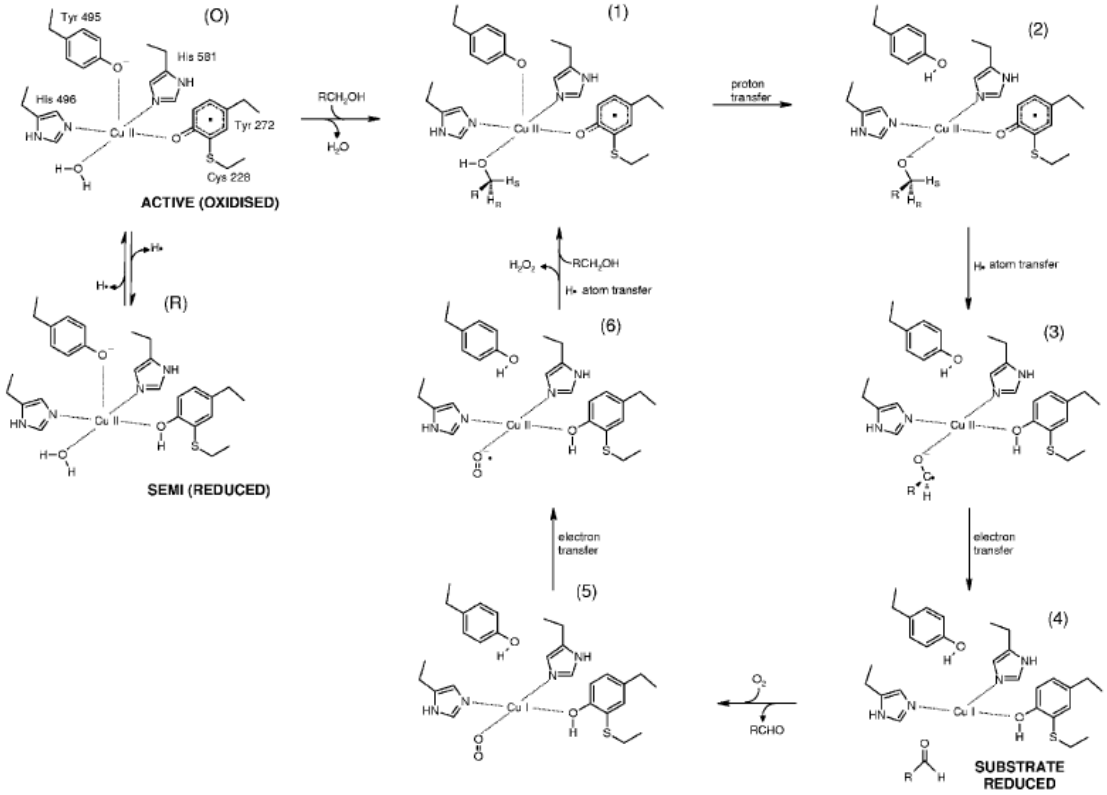


Figure 1.13 Schematic representation of the proposed catalytic cycle of galactose oxidase (McPherson *et al*.,2001)

1.4.3 Post-translational Processing Steps Leading to Maturation

Galactose oxidase belongs to the class of proteins that generate their protein-derived instrinsic cofactors by self-assembly as indicated in Section 1.3.

Galactose oxidase is unique among this group is because its conformational travel through the primary amino acid sequence to the mature and oxidized form involves at least four processing events:

1. Cleavage of a secretion signal sequence
2. Cleavage of a 17-aa N-terminal pro-sequence
3. Formation of the Cys228-Tyr272 thioether bond (a two-electron process)
4. Further one electron oxidation yielding Y272$^{\cdot}$ radical (McPherson *et al.*, 1992; Firbank *et al.*, 2001).

The latter three events have been shown to occur by a self-processing mechanism *in vitro* on the addition of copper and oxygen, even in the presence of protease inhibitors, indicating that not involving any additional processing enzymes. It was shown that pro-peptide cleavage and thioether bond formation do not proceed when the protein is heterologously expressed in

Aspergillus nidulans and purified under strictly metal-free conditions (Rogers *et al.*, 2000). Later, Rogers (2008) has shown that molecular oxygen is only required for the radical formation, not for the formation of the thioether bond.

Heterologous expression of galactose oxidase in copper-limited conditions, gives rise to three forms of the protein which can be identified as distinct bands on SDS-PAGE (Figure 1.14). The lower band belongs to the mature galactose oxidase and the presence of the thioether bond introduces an intramolecular loop between Tyr228 and Cys272 which forces the protein to run anomalously on SDS-PAGE (65.5 kDa) (McPherson *et al.*, 1993; Baron *et al.*,1994) .The middle band (68.5 kDa) is an incompletely processed form of the protein mirrored by the variant C228G, which is not able to generate a thioether bond (Baron *et al*, 1994). The resulting protein is in a fully extended, linear conformation of the galactose oxidase primary structure migrating with the expected apparent molecular mass. The upper band (70.2 kDa) is a precursor of galactose oxidase, carrying the additional N-terminal 17 amino acid pro-sequence as revealed by N-terminal amino acid sequencing . In terms of order of processing, results from SDS-PAGE analysis suggest that pro-sequence cleavage may occur before thioether bond formation since a band corresponding to a protein with both the pro-sequence and thioether bond has not been identified yet (Rogers *et al.*, 2000).

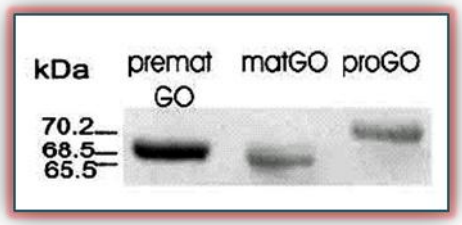


Figure 1.14 SDS-PAGE mobility assay results of three types of GAO. **prematGO**, (no pro-sequence, no thioether bond); **matGO,**(no pro-sequence, thioether bond formed); **proGO,** pro-sequence present, no thioether bond) (Firbank *et al.*, 2003)

Based on the reason that will be detailed in Section 1.6.2.1, the maturation process of the GAO may contain one more proteolytic additional step which includes the removal of 8-aa pre-peptide.

1.5 N-terminal Endoproteolytic Processing Mechanism of Fungal Precursor Enzymes

Post-translational modifications of one form or another such as phosphorylation, glycosylation and sulphation of amino acid side-chains, as well as α-amidation of C-terminal residues is central to the synthesis of the fully active form of almost all proteins (Eipper *et al*., 1992). Another fundamental example is protein processing by limited proteolysis, which can be classified into three groups: (1) signal peptide removal of preproteins by signal peptidase in the RER; (2) limited proteolysis of proteins after their release from the cell of origin, as a classic example being the activation of zymogens (Neurath,1957); and (3) intracellular processing of proproteins (Halban & Irminger, 1994).

Secretory proteins and peptides are mostly synthesized as inactive precursors, usually as pre-pro-proteins that undergo posttranslational processing to become biologically active mature proteins (Duckert *et al*.,2004). Transient peptides at the N-terminus may involve only a signal peptide or one or more additional peptides. A single additional peptide is called a pro-peptide, however if there are two additional peptides, they are called pre-peptide and pro-peptide, respectively. Such N-terminal transient peptides are often referred to as a leader peptide (Özöğür *et al*., 2009; Ögel *et al.*, unpublished data).

Although the process of the targeting and translocation of newly synthesized proteins is not fully elucidated and may vary between different cell types and organelles it is, in general dependant on whether the individual proteins have peptide signals within their structure (Ögel *et al.*, unpublished data). This signal is mostly in the form of an N-terminal peptide known as the signal peptide which facilitates translocation of proteins through hydrophobic membrane barriers; specifically the endoplasmic reticulum (ER) in eukaryotic cells and the cytoplasmic membrane in prokaryotes.The interaction of the signal peptide with its cognate protein complex (SRP or signal recognition particle) on the ribosome assures that secretory proteins are synthesized on the RER and translocated into the lumen of this organelle, usually co-translationally (Blobel & Dobberstein,1975; Walter & Lingappa, 1986; Meyer, 1991; Lingappa, 1991; Rapoport , 1992). Then the cleavage of the signal peptide by signal peptidase occurs rapidly. The secretory proteins become folded into their tertiary structure in the RER, accompanied by chaperone proteins (Rose & Doms, 1988; Landry & Gierasch, 1991; Hurtley & Helenius, 1989; Helenius *et al*.,1992; Rothman, 1989; Kelley and Georgopoulos, 1992). In addition to oligomerization (Hurtley & Helenius, 1989), glycosylation is also initiated within the RER and proceeds up to the trans-Golgi although glycosylation is not mandatory for every secretory protein (a non-glycosylated secretory protein proinsulin).It is the trans-Golgi network (TGN) that sorts out the proteins to be secreted via the regulated pathway or the constitutive pathway. Studies showed that passage through the trans-Golgi was an obligatory step in secretion (Tartakoff & Vassalli, 1978; Ravazzola *et al*., 1981). These observations were affirmed by the morphological studies monitoring the intracellular trafficking of regulated and constitutive secretory proteins (Orci, *et al*.,1987; Tooze *et al*.,1987). The glycosylation pattern of the proteins also supports that the TGN is the last compartment common to both secretory pathways (Moore & Kelly, 1985; Huttner, 1988).Then these proteins are transported to the outer surface via secretory vesicles (Kelly, 1985).

Recently, cloning and sequencing studies of filamentous fungal genes encoding extracellular proteins brings the role of transient peptides and particularly those at the N-terminus to light since they can act on the destiny of the secretory route of proteins, stability, assistance for proper folding (Ögel *et al.*, unpublished data) as well as functional effects such as inhibiton.

1.5.1 Pre-peptide Processing

Signal peptides, the N-terminal pre-regions direct the precursors to the appropriate cellular compartments (Duckert, *et al*., 2004).

Most of the proteolytic processing events take place at the N-terminal part of proteins. Among those, signal peptide cleavage is possibly the most well-known, for which software programmes are established to predict cleavage sites precisely, such as the SignalP Server (Nielsen *et al*., 1997).

Sequence comparisons of characterized signal peptides have failed to identify homology at the level of their primary structures. Nonetheless common three regions with distinct characteristics have consistently been described both in eukaryotic and prokaryotic signals. These are n-region, h-region and c-region which include a basic N-terminus of one to five residues (precore), a hydrophobic core of at least eight residues, and a neutral but polar C-terminal region of variable length (postcore). Residues at positions -3 (must not be aromatic, charged or large and polar) and - 1 (must be small; Ala, Ser, Gly, Cys, Thr, or Gln) relative to the cleavage site appear to be most important as structural determinants of the signal cut site (Perlman & Halvorson, 1983; von Heijne, 1985,1986a).

1.5.2 Pro-peptide Processing

Although alluded to earlier (Li, *et al*.,1965), since Steiner and colleagues provided the first direct evidence for the processing of a higher-molecular-mass precursor, proinsulin, into insulin which is a smaller, biologically active peptide hormone (Steiner *et al*., 1967; Steiner & Oyer; 1967); it has become apparent that the vast majority of regulatory peptides are derived from a higher-molecular-mass precursor, or pro- protein, by limited proteolysis. This process is now known as pro-protein conversion (Halban & Irminger, 1994).

At their N-terminus transient peptides may contain only a signal peptide or may also have one or more additional peptides, called as pro-peptide(s) between the signal peptide and mature protein. In light of the widely accepted passive bulk flow hypothesis, proteins translocated into the ER are secreted if they do not possess additional signals for targeting to other cellular compartments (Wieland *et al.*, 1987). Since an N-terminal signal peptide is necessary only for passage into the ER, this should normally be sufficient for export. However, more than one step seems to be required for various extracellular proteins.

The nature of residues at the leader peptide-mature protein junction is much more effective than the length of the leader peptide on the presence of a pro-peptide in filamentous fungal extracellular proteins. This has been the case particularly in those proteins where a basic residue is found preceding the cleavage site at the mature N-terminus. None of the 161 eukaryotic and 36 prokaryotic signal sequences analysed by von Heijne (1986a) contains a basic residue at position -1, relative to the cleavage site. When no better cleavage site is available in the vicinity, unusual residues can be used in position -1 (von Heijne, 1986a), however in almost all of the proteins with a basic residue preceding the cleavage site a favourable signal sequence cut site is found further upstream, indicating additional proteolytic processing (Ögel *et al.*, unpublished data).

Pro-peptides do not exist only in secreted proteins but also in some of the proteins that do not pass through the endoplasmic reticulum and relatively less is known about the function and

mechanism of the removal of pro-peptides. Özöğür and colleagues (2009) have developed an application that helps prediction of the pro-peptide cleavage site of fungal extracellular proteins which display mostly a monobasic or dibasic processing site.

Proteolytic processing events in cell biology attract more attention due to recent findings related to their critical functions including gene expression (Sakai *et al*., 1998), cell cycle (Wilkinson, 1995), embryogenesis (Cui *et al*., 1998; Dubois *et al*., 1995; Peschon *et al*., 1998; Roebroek *et al*., 1998), apoptosis (Weidemann *et al*., 1999), cell trafficking (Springael *et al*.,2002), human diseases such as Alzheimer's (Evin *et al*., 2003), and endocrine/neural functions (Rouillé *et al*., 1995; Seidah *et al*., 1994; Nakayama, 1997; Creemers *et al*.,1998; Seidah *et al*., 1998).

1.5.2.1 Monobasic Processing

Possession of a monobasic cleavage site at their leader-mature protein junction is common among a considerable number of filamentous fungal extracellular proteins. Based on the comparison of the precursors of the fungal proteins polygalacturonase II of *Aspergillus tubigensis* and *Aspergillus niger*, α-sarcin of *Aspergillus giganteus*, and cellobiohydrolase II of *Trichoderma reesei*; Bussink et *al.*, (1991) suggested that a pro-hexapeptide is present at monobasic processing sites of filamentous fungal extracellular proteins with the consensus sequence ser-PRO-leu-GLU-ala-ARG, where residues in upper case are completely conserved. A comparison of the pro-peptide regions of proteins with a monobasic cleavage site at their leader-mature protein junction suggests that the proposed pro-hexapeptide is not completely conserved. Considering the putative pro-peptides of proteins that are subject to monobasic processing, a common sequence motif does not exist, with the exception of a proline that is consistently present and frequently a Leu or Ile is adjacent to this residue. The fact that the pro-peptides contain both hydrophilic and hydrophobic residues and the absence of common sequence motif or similarity could either indicate processing by different proteases or the importance of conformational determinants for cleavage. In the latter case the presence of a proline may be highly important. There are no examples where proline is present immediately before or after the basic residue at the cleavage site in filamentous fungi. Nevertheless, since proline is suggested to take a role at the level of three-dimensional structure, rather than the primary sequence (Schwartz, 1986), a similar function can still be attributed to the proline residues within the structure of pro-peptides of filamentous fungi which undergo monobasic processing.

1.5.2.2 Dibasic Processing

Among leader peptides where bipartite processing takes place, pro-peptide cleavage following two basic residues appears to be most common among glucoamylases and lignin peroxidases. In the majority of dibasic processing sites 'Lys-Arg' pair is the most frequently encountered one at the cleavage sites, whereas 'Lys-Lys' and 'Arg-Arg' pairs are less common (Ögel ,1993). Dibasic processing of pro-peptides is likely to be a conserved feature rather than a similarity due to high homology and no correlation has been found between protein function and pro-peptide presence. Among the proteases, neutral protease II (NpII), a

zinc metalloprotease of *A. oryzae*, has been shown to have a pro-peptide that is cleaved by dibasic processing (Tatsumi *et al.*,1991). Auto-processing of pro-peptides are common among proteases (Silen & Agard, 1989; Egnel & Flock, 1991, 1992).However, purified NpII was processed into its mature form *in vitro* by the addition of $ZnCl_2$ following the Thr (+1) residue rather than two Arg residues, yielding Glu(2) as the N-terminal of the mature protein. This shift in pro-peptide processing might be due to a conformational change in the purified enzyme (Tatsumi *et al.*,1991). At the same time it may indicate the autocatalytic cleavage is likely to be dependent on structural motif of the pro-region rather than conserved residues in the primary sequence (Ögel *et al.*, unpublished data).

It is clear from previous studies that the mature part of the protein has also significant effect in the export process (Li *et al*., 1988). The net charge or hydrophobicity of filamentous fungal exported proteins at their N-terminus could likewise influence translocation across the cell membrane or the cell wall. It has been expressed that there is usually a net charge of zero or less in the region surrounding the signal peptide cleavage site and the first N-terminal five residues of the mature protein (von Heijne, 1986b; Li *et al*., 1988; Boyd & Beckwith, 1990). At the first position following the cleavage, Ala (+1) appears to be the most abundant amino acid. Unlike the negatively charged amino acids Asp and Glu, none of the positively charged residues appear at position +1 (Ögel *et al.*, unpublished data). Hydrophobic regions of the membrane spanning domains in integral membrane proteins are also important for the export process since the signal peptide has an essential hydrophobic core. It is not known whether hydrophobicity at the N-terminus of mature proteins is affecting the translocation of proteins into the extracellular medium (Ögel, 1993).

1.5.3 Autocatalytic N-terminal Processing

Up to date, post-translational processing of proteins, including processes such as activation of proenzymes by partial proteolysis, amino acid sidechain modifications by phosphorylation, alkylation, or acylation, and conjugation with prosthetic groups, was supposed to be catalyzed by other proteins with the evidence of characterization of several enzmes modifying proteins. However, at least five types of self-catalyzed protein rearrangements that do not require the intervention of other enzymes, have been discovered, including protein splicing and autoproteolysis of hedgehog proteins, glycosylasparaginases pyruvoyl enzyme precursors and galactose oxidase (Perler *et al*.,1997; Rogers, 2000).

The most complex of these modifications is protein splicing, which requires the co-ordinated cleavage of two peptide bonds that flank the protein splicing element (the intein) and the ligation of the external protein domains (the exteins) to form a functional protein whereas autoprocessing of hedgehog proteins, glycosylasparaginases, pyruvoyl enzyme precursors, and galactose oxidase involves the cleavage of a single peptide bond. The latter three have N-terminal modifications.

As it is observed in GAO, cleavage of pro-sequence occurred in spite of the protease inhibitor cocktail only by utilizing available reagents such as copper and dioxygen (Rogers *et al.*, 2000). Among those above; in addition to GAO, pyruvoyl enzyme precursors (histidine

decarboxylase) generate their cofactor by posttranslational modification involving an autocatalytic cleavage of a peptide bond. However, for glycosylasparaginases, little is known about the catalytic groups other than His150 that are required for this process (Guan *et al*., 1996).

Histidine decarboxylase consists of two subunits that originate from the self-processing of an inactive pro-enzyme. During the autocatalytic cleavage, an essential pyruvoyl group is formed at the amino terminus of the α-subunit that is derived from Ser82 of the pro-enzyme (van Poelje & Snell, 1990). However, in GAO, pro-sequence cleavage and cofactor formation must be separate processes because an intermediate form without the N-terminal pro-peptide and crosslink has been identified (Rogers *et al.*, 2000). The mechanisms of both modifications require more investigations for being elucidated in future. Although the site of pro-sequence cleavage was clear by the structural data, it does not reveal any obvious mechanism. An interesting point of pro-GAO processing is the mechanism by which the pro-sequence is cleaved in a copper-dependent manner, in which copper may bind transiently, near the cleavage site (Firbank *et al.*, 2003). It is also intriguing that site-specific cleavage events have been observed as copper binds close to the site of a disulphide bridge in the amyloid precursor protein implicated in Alzheimer's disease (Multhaup *et al.*, 1998). The site of cleavage in GAO is also near the disulphide bond formed between Cys515 and Cys518, and experiments are underway to assign any role for this site in pro-sequence cleavage (Firbank *et al.*, 2003).

It is known that addition of copper and oxygen leads to the *in vitro* processing of galactose oxidase (Rogers *et al.*,2000). Copper which is a highly toxic transition metal presents *in vivo* as less than one free copper ion per cell (Rae *et al.*, 1999) and intracellularly, it is associated with metallochaperone proteins that are responsible for its delivery, by direct interaction, to target copper-requiring proteins (Rae *et al.*, 1999; Harrison *et al.*, 2000; Rosenzweig & O'Halloran, 2000). Since galactose oxidase can be expressed heterologously in other organisms it seems unlikely that these hosts contain a suitable copper chaperone to facilitate intracellular copper delivery to GAO. Based on this, it seems possible that processing occurs outside the cell where higher free copper levels are present.

On the other hand, amyloid beta precursor protein (APP), the major components of amyloid plaques in Alzheimer's disease and prion protein (PrP) are cell membrane elements implicated in eurodegenerative diseases. Both proteins undergo endoproteolysis and the process in the two proteins could be related in terms of their functions and their distributions. A radically new function is suggested that PrP catalyses its own cleavage; the C-terminal fragment functions as an α secretase which releases anticoagulant and neurotrophic ectodomains from APP while the N-terminal segment chaperones the active site. The prosegment, may act as a chaperone, as in many pro-enzymes (Abdulla, 2001). It is only known about the enzyme responsible for the endoproteolytic cut is its inhibition by metal chelators (Jimenez-Huete *et al.*,1998).

1.5.4 Pro-peptides and Their Functions

Proprotein processing leads to the generation of at least two, and frequently many more, smaller peptides from limited proteolysis of a single precursor (Halban & Irminger, 1994).

For a better understanding of some of the peptides generated by conversion that do not have any known biological function as secretory products, a role has been attributed by intracellular sorting and conversion events for these peptides (while they are still retained within the precursor molecule). The roles of the pro-region of prosomatostatin in targeting to the regulated pathway (Stoller & Shields, 1989) and that of the connecting peptide (C-peptide) of proinsulin in proinsulin conversion (Gross *et al*., 1989) are the two examples of postulated involvement. Additional functions suggested for pro-regions include helping the precursor to fold into its correct three-dimensional structure, or simply as a spacer peptide assuring the minimum length believed to be required for a preprotein penetration into the lumen of the RER (Docherty & Steiner, 1982).

For example, pro-peptides have consistently been identified in proteases where they serve to maintain an intracellular inactive configuration or as it is indicated above some prodomains act as intramolecular chaperones that mediate correct folding of the newly synthesized proteins (Silen & Agard, 1989; Zhu *et al.*, 1989; Simonen & Palva, 1993). In mammalian cells, pro-regions are shown to be processed to produce alternative forms of secretory hormones (Benoit *et al.*, 1987).Other prodomains are only indirectly involved in folding and have other functions such as transport and localization, oligomerization, regulation of activity (Shinde & Inouye, 2000) and quality control of folding (Bauskin *et al*., 2000).

Limited endoproteolysis of the corresponding precursor frequently occurs at motifs containing multiple basic residues, arginine or lysine, immediately C-terminally of the basic amino acid motifs (Seidah & Chretien, 1999).

In Ögel and colleagues' unpublished study, analysis of transient peptides of filamentous fungal extracellular proteins and their potential roles were also carried out. Importance of long pro-peptides (50 to over 100 residues) of bacterial proteases for proper folding and activation of the enzyme (Silen & Agard, 1989; Simonen & Palva, 1993) could also be the case for filamentous fungal proteases. The role of short pro-peptides (6-9 residues) is not clear, which could be involved in secretion process (can enhance the level of secretion) (Eder & Fersht, 1995; Baardsnes *et al*., 1998) such as targeting, passage through the outer membrane, the release of proteins into the extracellular medium, in maintaining an inactive configuration or in protein folding. Some evidence for the latter arises from heterologous expression studies on active cutinase of *Fusarium solani* in *E. coli* (Soliday *et al.*, 1984; Martinez *et al.*, 1992). Conversely in some cases, pro-peptides could also be involved in retarding the folding process. The studies in prokaryotes indicated that an early function of the leader sequence is to slow down the folding process allowing interaction with the chaperone. Since for a soluble protein to be transferred across a membrane should not be in its final thermodynamically stable folded form, this process seems to be essential for export. (Hardy & Randall, 1992).

Lamy & Davies (1991) has suggested that the leader peptide protects producing strains from suicide by keeping the enzyme in an inactive state until processing is carried out in the Golgi system. Such short pro-peptides in *Bacillus* species are unlikely to act on secretion directly but one possibility is that they might stabilize the secreted proteins by allowing folding into a protease-resistant conformation (Simonen & Palva, 1993). Pro-peptide region deletions of certain mammalian proteins have not affected targeting of the mature protein to the regulated pathway, even though relatively short pro-peptides, often at the N-terminus, are mostly responsible for intracellular targeting (Chidgey, 1993).

1.5.4.1 Putative Functional Roles of GAO Pro-peptide

Although there is overall structural similarity between pro-GAO (retaining a 17-residue N-terminal pro-peptide extension and lacking the Tyr–Cys site) with mature GAO, significant alterations have been reported in main chain position and some active-site-residue side chains before and after the removal of the pro-peptide. The open structure of the active site of pro-GO, secreted by metal-starved *Aspergillus* cultures suggests a role that the prosequence acts as an intramolecular chaperone for positioning of active-site residues, including copper ligands, to facilitate the chemical processes required for thioether bond formation, including efficient initial copper binding (Firbank *et al.*, 2001,2003). Furthermore, studies of Seneviratne (1993) and McPherson *et al*., (1993) supported the idea that pro-sequence may be important for proper protein folding. However, this proposal is at odds since galactose oxidase is efficiently expressed in its mature, functional form from chimeric fusion protein precursors in *P.pastoris* containing a different leader sequence, glucoamylase signal peptide of *A. niger* (Whittaker , 2005; Whittaker & Whittaker, 2000). Furthermore, heterologous expression of functional mature GAO was accomplished in the absence of the pro-sequence including directed evolution studies in *E.coli* (Deacon *et al*., 2004; Lis & Kuramitsu, 1997; Rogers *et al*., 2008; Sun *et al.*, 2001). Similarly, the role of the pro-sequence was investigated in studies using an N-terminally processed (i.e., no pro-sequence) recombinant galactose oxidase precursor protein without tyrosine-cysteine cross-link. Mature galactose oxidase formation *in vitro* from that precursor by addition of O_2 and Cu^{1+} is also supports that the pro-sequence is unlikely to be acting as an intramolecular chaperone (Whittaker & Whittaker, 2003).

The N-terminal part of cloned *gaoA* gene includes a 17 amino acid pro-sequence which has been proposed to be associated with secretion (McPherson *et al*., 1993) and which is cleaved in a copper-mediated auto-processing reaction (Rogers *et al*., 2000).

The short length feature of the galactose oxidase pro-sequence (17-aa) resembles to that of fungal cutinase (Longhi & Cambillau; 1999) than to proteases. This prevents it from acting as a direct steric inhibitor of the active site without ruling out an inhibitory role through affecting the structure of the active site (Firbank, 2001).

It is known that at least some of the secreted proteins may pass through the vacuole before being transported to the extracellular medium (Peberdy, 1994). The study on the regulation of galactose oxidase synthesis and secretion states that the export of the enzyme to the

extracellular medium is a regulated process affected by pH and culture density. It was suggested that below pH 7.0 the enzyme is stored in the vacuole of the cell and secreted when the pH is raised to 7.0, accompanied by a decrease in vacuolization (Shatzman and Kosman, 1977). The filamentous fungi vacuole may indeed take place in the storage and/or processing of certain secreted proteins prior to export and certain pro-peptides could have a role in the sorting stage (Ögel *et al.*, unpublished data).

1.6 Use of Different Heterologous Expression Systems for Production and Processing of GAO

Several expression studies of *Fusarium graminearum* (NRRL 2903) *gaoA*, including site-directed mutagenesis applications in order to understand the functional features of different residues, especially around active site, together with the thioether bond constituents and to alter the catalytic efficiency and substrate range were carried out in different hosts such as *P.pastoris* (Deacon *et al*., 2004; Hartmans *et al.*, 2004; Rogers *et al*., 2008), *A. nidulans* (Baron *et al*,1994; Rogers *et al.*, 2000; Firbank *et al.*, 2001) and *E.coli* (Seneviratne, 1993; McPherson *et al*.,1993; Lis & Kuramitsu, 1997; Delagrave *et al.*, 2001; Sun *et al.*, 2001; Escalettes & Turner, 2008).

1.6.1 Expression of GAO by Directed Evolution in *E.coli*

In addition to the the hosts as explained above a few expression trials were carried out in *E.coli*, however functional enzyme was succeeded to be obtained only as a *lacZ* fusion (Lis & Kuramitsu, 1997).Two constructs were obtained as derivatives of pUC118, in which GAO starts from the second codon of the mature protein, one of them terminates at its original termination codon whereas the other is followed by an in-frame GBD coding sequence.

Functional expression in a host organism that permits creation and rapid screening of mutant libraries is a prerequisite to enzyme modification by powerful directed evolution methods (Arnold, 1998; Petrounia & Arnold, 2000). *E. coli* is an excellent host for directed evolution studies, however many important eukaryotic enzymes have not been functionally expressed in *E.coli*. Until the biochemical characterization of the *E.coli*-expressed GAO by directed evolution was reported by Sun *et al.*(2001), all have been performed on the enzyme obtained from its natural source or from fungal (McPherson *et al*., 1993; Xu *et al.*, 2000) and yeast (Whittaker & Whittaker, 2000) expression systems which are not suitable for directed evolution.

Directed evolution studies based on error-prone PCR have been reported so as to improve the activity of GAO towards appropriate substrates (Delagrave *et al.*, 2001). A digital imaging system, known commercially as ‘Kcat’, which relies on simple, widely known colorimetric activity assays was used to screen microcolonies for activity against guar and methylgalactose. The pBAD-myc/his vectors (Invitrogen) were used for recombinant expression of entire GAO orf in *E.coli*. Error-prone PCR was performed for single mutations and then double and triple mutants were constructed from these single mutants obtained by error-prone PCR using standard molecular cloning techniques, which is called “manual recombination”.

According to the second directed evolution study of Sun *et al.*, (2001) to increase the total activity of GAO in *E. coli*, random mutagenesis was applied to the entire mature *gaoA* gene also to just the region of the gene which is responsible for catalytic activity (McPherson *et al*., 1993). In this case, cultivation, lysis and assays in microtiter plates were carried out by individual bacterial colonies for screening the libraries and further rounds of random mutagenesis to these variants resulted in increased activity. The variant A3.E7 was the highest activity showing mutant identified which carries S10P, M70V, P136 (silent mutation), G195E, V494A, and N535D mutations.These mutations provide advantageous effects in terms of thermostability, enhanced expression and reduced pI. It was shown that the effects of mutations were not cumulative. This best variant displayed a 30-fold increase in total activity relative to wild-type, which reflects an 18-fold increase in GAO expression and a 1.7 fold increase in catalytic efficiency. The broad substrate specificity of wild type GAO was retained in the enzyme variants obtained by directed evolution (Sun *et al.*, 2001). Oxidized form of the enzyme produced in *E. coli* had redox potential and stability comparable to the *Fusarium* GAO (Baron *et al.*, 1994; Reynolds *et al*., 1997). The advantageous effects of GAO achieved in *E. coli* facilitate protein purification and characterization.

Deacon and colleagues (Deacon, 2008; Deacon *et al.*, 2011) generated the mutations (S10P, M70V, P136 (silent mutation), G195E, V494A, and N535D) reported by Sun *et al.*, (2001) in pPICαproGO by *QuikChange Site-Directed Mutagenesis* approach resulting in clone pPICαproGO-M1. The PCR-amplified fragments from pPICαproGO and pPICαproGO-M1 were subcloned into pET101D vector. Additionally, three N-terminal silent mutations were also introduced to the mature GAO decided by ProteoExpert (Roche Applied Science) software. Three N-terminal silent mutations were introduced into the Pre-Pro-GAO construct as well, not by ProteoExpert but by base substitutions. The pre-pro- and mature GAO carrying constructs were either carrying the silent mutations or the mutations of Sun *et al.*,(2001) or a combination of both.

In this thesis, Pre-Pro-GAO constructs (Deacon, 2008), namely Pre-ProGON1 and Pre-ProGOMN1 will be used for heterologous expression of GAO in *E.coli* and details of these constucts will be given in Section 2.1.4.

1.6.2 Pro-peptide Processing in Recombinant GAO

1.6.2.1 GAO Pro-peptide Processing in Fungi

Aspergillus nidulans was used in various expression studies of GAO. In the study of Baron and colleagues (1994) wild-type enzyme was produced from the original *Fusarium spp.* NRRL 2903 strain (native wild-type enzyme) and heterologous expression of *Fusarium graminearum* GAO (coding region) protein and mutational variants (C228G,W290H and Y272F-unable to be purified-) was performed in *Aspergillus nidulans*. Crystal structure of the *Aspergillus*-expressed wild-type GAO and that of mutational variants were obtained. It was found to be the crystal structures of the native wild-type enzyme (Ito *et al.*, 1991,1994) and the *A. nidulans* expressed wild-type were essentially identical. C228G variant also

provide insight into the mobility difference between the variants on SDS-PAGE, which results from the thioether bond formation.

According to the studies of Ögel (1993) based on predictive algorithm of von Heijne, the first 16-18 amino acids of GAO can act as a signal peptide for ER translocation, the remaining 23-25 residues of the leader sequence must act as a pro-peptide, removed at a later stage of secretion process. In addition to crystallographic studies, the most direct approach in order to provide evidence for the two-stage processing of galactose oxidase leader peptide would be the isolation of pro-galactose oxidase and submission to N-terminal amino acid sequencing. At the leader-mature enzyme junction of galactose oxidase, at the monobasic cleavage site, Arg-Ala pair was substituted into Phe-Phe to prevent processing of the putative pro-peptide. This mutation resulted in an enzyme that was correctly processed but displayed a significantly low efficiency of export compared to heterologously expressed wild type galactose oxidase in *A. nidulans* (Ögel, 1993). The location of the mutant enzyme was close to the outer surface, either the cell membrane or the cell wall. It is not clear that the membrane or cell wall retention was caused by the pro-peptide that was not cleaved but cleaved during purification of the enzyme, or whether this retention was indeed due to a single basic amino acid difference at the N-terminus (Ala1Phe). Phe residue at the beginning of mature galactose oxidase may influence export by altering the hydrophobicity balance at the N-terminus (Ögel, 1993).

In later studies, this proposed two-step cleavage of GAO and the presence of the pro-sequence by Ögel (1993) was affirmed by heterologous expression of *gaoA* under copper limited and metal-free growth conditions (Rogers *et al*., 2000) and by the crystallographic studies (Firbank *et al.*, 2001) in *Aspergillus nidulans*, as explained before in Section 1.2.4 and 1.4.3. This heterologous expression of *Fusarium graminearum* GAO protein in *A. nidulans* under copper-limited conditions gave rise to the appearance of multiple protein forms (see Figure 1.14, Table 1.1). These different species of GAO undergo similar posttranslational modifications as it does in its native host *Fusarium ssp* (Baron *et al*., 1994).

Table 1.1 Galactose oxidase forms classified according to their possession of pro-sequence or thioether bond (Rogers *et al.*, 2000).

	M_r (kDa)[a]	N-terminal sequence[b]	Is the pro-sequence present?	Is the thioether bond formed?
pro-sequence form	70.2	AVxxxIPEG	yes	no
unmodified form	68.5	ASAPIGxAI	no	no
thioether form	~65.5	ASAPIGS?AI	no	yes

[a] Estimated from SDS-PAGE. [b] Sequencing was performed by standard automated Edman procedures. "x" is an unidentified amino acid. These results are consistent with the nucleotide sequence (Genbank M86819).

Accordingly, the 41 amino acid leader peptide is likely to consist of a signal peptide, pro-peptide and an additional putative pre-peptide between these. Based on the above results both on signal peptide cleavage and crystallographic studies revealing a 17-aa pro-peptide, there appears to be a pre-peptide cleavage site after the His-Lys at position -19 and -18 After the removal of the signal peptide by signal peptidase, this pre-peptide is likely to be cleaved by a monobasic/dibasic processing protease during secretion (Figure 1.15). The autocatalytic pro-peptide cleavage is likely to take place following pre-peptide cleavage possibly by proteases (Gençer (Kocuklu), 2005; Ögel, personal communication, 2005).

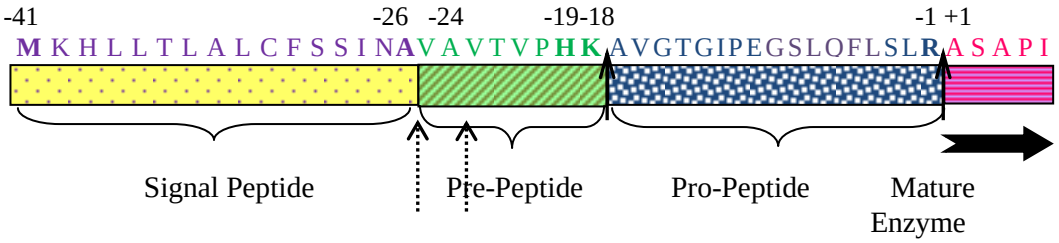


Figure 1.15 A diagram showing N-terminal part of GAO indicating cleavage sites for different processing events (dashed arrows show the possible cleavage sites for signal peptide- Ögel, 1993)

1.6.2.2 GAO Pro-peptide Processing in Yeast

There are two best representatives where pro-form of GAO is used and processing of the pro-sequence is subjected and revealed as consequences of the studies in which *Pichia pastoris* is used as an expression host.

In the study of Wilkinson *et al.*, (2004), galactose oxidase from *Fusarium spp.* (NRRL 2903) has been cloned and expressed in *Pichia pastoris* X-33. The constructs used in this work were regenerated according to the variants identified by Delagrave and colleagues (2001) by *QuikChange Multi Site-Directed Mutagenesis Kit* (Stratagene) and investigated by generating a series of single and double mutants comprising V494A, C383S and Y436H for comparison with the triple mutant clone, C383S/Y436H/V494A. The variants C383A and Y436A were also studied. These mutants were all tested in terms of catalytic efficiency and molecular masses. C383S/V494A is found to be the most efficient variant against 1-methyl-α,D-galactose and also against guar with $k_{cat}(s^{-1})$ values of 1495 ± 114 and 58 ± 4, respectively. The constructs contained the 17-aa N-terminal pro-sequence of the enzyme in-frame with the α-mating factor as a secretion signal, and are accordingly referred to as Pro-GAO.

According to the mass spectrometry data, the major form of the GAO protein produced in *P. pastoris* displayed a mixture of N-terminal extended species. It contains additional residues, Ser-3, Leu-2 and Arg-1 at the N-terminal end (Wilkinson *et al.*, 2004). This finding contrasts with the production of recombinant enzyme from a pro-sequence containing construct in the filamentous fungus *Aspergillus nidulans* or of the native enzyme from *Fusarium graminearum* where in both cases there is a single mature

enzyme species starting with Ala +1 (Baron, *et al*., 1994; McPherson *et al*., 1992; Rogers *et al.*, 2000). This result is consistent with a previous report of mixed N-terminal sequences from constructs expressing GAO with a native 41-aa leader peptide in *P. pastoris* (Whittaker & Whittaker, 2000). In Whittakers' study galactose oxidase cDNA has been cloned for expression in *Pichia pastoris* both as the full-length native sequence and as a fusion with the glucoamylase signal peptide sequence (*glaA*) of *A. niger.* Expression of the full-length native sequence results in a mixture of mature and partly processed galactose oxidase. The reason for the differences in N-terminal processing by *P. pastoris* in these two expressions is unclear; however, the recombinant enzymes prepared from *P. pastoris* and *A. nidulans* exhibit similar properties in terms of kinetics, UV/vis spectra and three dimensional structure (Deacon *et al.*, 2004).

The N-terminus of the full-length GAOX cDNA expression product is highly heterogeneous, with multiple residues appearing in each sequence cycle. The ragged end of the polypeptide is consistent with inefficient processing of the pro-peptide and partial digestion of the presequence linker. The sequencing results also indicate one of the possible signal peptidase cleavage sites (AVA-VTV) as shown in Table 1.2. The clean and correct N-terminal sequence of the protein expressed as a gla fusion construct (Table 1.2) confirms that the gla pro-sequence is efficiently processed to generate the authentic mature protein. Furthermore, it can be stated that the gla signal sequence is recognized by the secretory machinery of the yeast based on this study and previous expression studies by glucoamylase signal peptide in *P.pastoris* (Fierobe *et al.*, 1997; Whittaker *et al.*, 1999; Whittaker & Whittaker, 2000).

Table 1.2 N-terminal sequencing for recombinant galactose oxidase expressed in P. pastoris (Whittaker & Whittaker, 2000).

Expressed protein	Sequence
1. glaGAOX	−20 −10 −1 M H F R S L L A L S G L V C T G L A N V I S K R A S A P I G Cycle number 1 2 3 4 5 6 Results A S A P I G
2. GAOX	−40 −30 −20 −10 −1 M K H L L T L A L C F S S I N A V A V T V P H K A V G T G I P E G S L Q F L S L R A S A P I Cycle number 1 2 3 4 5 Results (A V G S R) (S T P L) (A V P K R) (P I G) (I K D N E)

The formation of two major protein products during expression of full-length GAOX suggests that *Pichia* is unable to efficiently process the native signal peptide, perhaps due to the presence of a suboptimal cleavage site in pro-sequence. The signal sequence analysis points out that correct processing would rather prefer an isolated Arg residue for cleavage to the favoured Lys-Arg dibasic recognition site for Kex2 cleavage, resulting in a protein product that is heterogeneous and exhibits low specific activity (Whittaker & Whittaker, 2000). A dibasic Lys-Arg motif requirement for Kex2-like maturation of recombinant secretory proteins in *Pichia* has previously been reported (Martı´nez-Ruiz *et al.*, 1998).

In this study we suggest that the above interpretations of Whittaker are incomplete since the the signal sequence is likely to be cleaved following Ala (-26). This is because Ögel (1993) based on the weight matrix approach of von Heijne (1986a), explained that this site gives the highest score, which does not rule out that Ala (-24) is a strong competitor for the signal peptide cleavage (Ögel, 1993; (Whittaker & Whittaker, 2000). The remaining peptide is supposed to be cleaved as explained before in Section 1.4.3 and 1.6.2.1 (Gençer (Kocuklu), 2005).

1.6.2.3 GAO Pro-peptide Processing in Bacteria

Expression of pro-GAO has been attempted in *E.coli* several times. The extracellular trials include the use of native galactose oxidase secretion signal, together with the *lac* promoter and use of alygnate lyase signal peptide with alyA strong bacterial promoter of this protein. Low levels of functional GAO was only obtained by the native signal peptide construct, however the protein could not be secreted (Seneviratne, 1993; McPherson *et al*., (1993). Together with the cytoplasmic expression studies, it was suggested that *E. coli*-expressed GAO is a cytoplasmic protein even when the leader sequence was present. The majority of the produced GAO was inactive; GAO activity had only been observed when the leader sequence was present, emphasizing that it might be important for correct protein folding in *E. coli.* Similarly, heterologous expression of active cutinase of *Fusarium solani* could be achieved in *E. coli* only in the presence of 15-aa long putative pro-peptide which is a part of a leader sequence of 32 residues (Soliday *et al.*, 1984; Martinez *et al.*, 1992). Cutinase activity was not detected without this 15-aa pro-peptide. Since the pro-peptide was not efficiently processed in *E. coli*, procutinase was used in X-ray crystallography, although no density was observed for the 16 amino acids at the N-terminus, which indicates a disordered structure (Martinez *et al.*, 1992). Alternatively, the leader sequence may lead the inefficient export of GAO to the periplasm of *E. coli*, resulting in low levels of active enzyme that arise due to the formation the thioether-bridge. It was thought that, bacterial expression system is not convenient for the purification of GAO variants (Seneviratne,1993; McPherson *et al*., (1993).

For cytoplasmic expression of GAO (Seneviratne,1993; McPherson *et al.*, 1993) two constructs were generated to express galactose oxidase with and without its pro-sequence. The expression was carried out in under the *lac* promoter, using pRH1090 (Baldwin *et al.*, 1990) as a vector and XL1 Blue (Stratagene) as the expression strain. No functional galactose oxidase expression was reported (might be due to sufficient copper unavailability, not clear) inspite of a band migrating at the expected molecular weight in western blotting.

In another study, enhanced production of galactose oxidase in *Fusarium graminearum* was reported by use of pET21a vector and IPTG induction, both with and without its pre-pro-sequence. However, the GAO protein belongs to another strain, F. *graminearum* PH-1 with different amino acid sequence at the N-terminal. Moreover, the study does not focus on the processing of the pro-peptide, it is not clear whether pro-peptide is cleaved in experiments performed with pro-peptide containing constructs. SDS-PAGE results of recombinant GAO without pre-pro-sequence is given, but GAO with pre-pro-sequence is absent (Choosri *et al.*, 2011).

In our previous study (Gençer (Kocuklu), 2005), four different mutations were designed in order to analyse the autocatalytic cleavage mechanism of GAO pro-sequence in *E.coli* with Pre-ProGON1 and Pre-ProGOMN1 constructs as indicated in Section 1.6.1.Three mutations were at the cleavage site (R-1P/A1P, R-1X/A1X, S2A) and one was close to the cleavage site (H522A) in the three dimensional structure of the enzyme. 13 different mutational variants were obtained resulted from both desired and undesired extra mutations. Following heterologous expression in the presence of protease inhibitor and copper loading under aerobic conditions, it was reported that none of the mutant proteins underwent self-cleavage. All the mutational variants, including the products of original recombinant plasmids, Pre-ProGON1 and Pre-ProGOMN1, showed no difference in band size, all slightly above 70kDa; indicating pre-pro-sequence presence on all constructs, which was confirmed by N-terminal sequencing (Pre-ProGOMN1 only). Non-mutants and SDS-PAGE detectable all mutants showed galactose oxidase activity, signifying proper active site construction by thioether bond formation. Since the pro-peptide is cleaved autocatalytically, the lack of removal of the putative pre-peptide in *E.coli* in the presence of Cu^{2+} and oxygen, which might be cleaved by a protease in fungus, is supposed to be the cause of lack of pro-peptide cleavage. Therefore, the region corresponding to the 8-aa pre-peptide is proposed to be deleted to prove this hypothesis (Gençer (Kocuklu), 2005).

In this thesis, pro-peptide processing will be investigated in the absence of 8-aa pre-peptide in *E.coli* by Pre-Pro-GAO constructs as well.

1.7 Protein Expression Approach

1.7.1 IPTG Induction

Today, DNA sequencing is a very important tool that provides coding sequences of many proteins of a huge number of organisms. Recombinant DNA technology renders it possible to clone these sequences into proper vectors for expression of the new protein in new hosts. pET expression system carries a *lacUV5* and T7 promoter which are inducible with IPTG to express clones effectively in *E.coli.* pET expression system is explained in Section 2.2.6.2 in detail.

BL21 Star (DE3) is found to be the best expression host with the experiments carried out by MatGOMN6, encoding the mature galactose oxidase, which showed the highest level of protein expression (Deacon *et al.*, 2011). In addition to the analysis of host strains, Deacon and colleagues (2011) have also analysed the effect of induction temperature and harvest time. Accordingly, optimum temperature was found as 25°C, for maximum soluble expression for all strains. As it is seen in Figure 1.16, the acivity is getting higher for almost all of the hosts when the harvest time is 16-22 hours.

IPTG induction method will be detailed in Section 2.2.6.3.1.

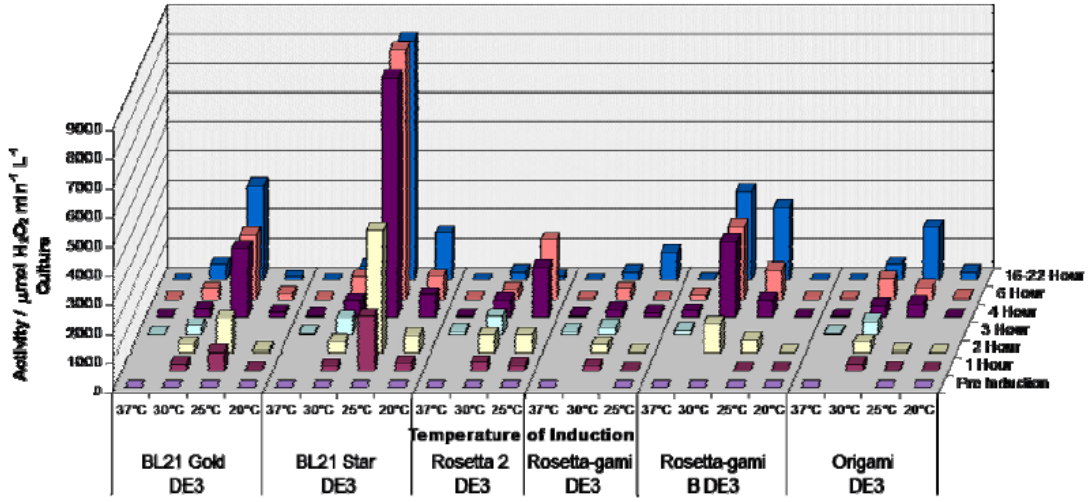


Figure 1.16 Screening optimum GAO expression parameters in different expression strains at different induction temperatures (20-37°C) at different harvest times changing between 0 - 22 hours (Deacon, 2008).

1.7.2 Auto-induction

Auto-induction is a kind of protein expression method reported by Studier (2005) for target gene expression in pET T7 polymerase-based vectors based on an optimized expression media (Studier & Moffatt, 1986; Studier *et al.,* 1990).The system uses three main carbon sources; glucose, glycerol and lactose and due to the catabolite repression system of *E.coli,* monitoring of growth and induction at a specified cell density is not required when compared to the IPTG induction. When there is D-glucose in the medium, upon the initial growth the expression of target gene was repressed by D-glucose, a catabolite repression actor. This process is regulated at different levels:

A) *E.coli* rather prefers the metabolism of glucose, since catabolism of another carbon source would bring an extra burden to the cell, such as production of large numbers of enzymes for utilization.

B) The phosphoenolpyruvate/carbohydrate transferase system (PTS) of *E.coli, w*hich is composed of at least three proteins enzyme I (EI), histidine protein (HPr) and enzyme II (EIIA), the latter is carbohydrate specific (Görke & Stülke, 2008) which transports and phosphorylates carbohydrates (Deutscher *et al*., 2006). Following the phosphorylation of glucose transporter $EIIA^{Glc}$ it binds to adenylate cyclase and cyclic AMP is synthesised. This affects the activation of promoters of catabolic genes over the formation of cAMP-CRP complexes. In the presence of the substrate (i.e glucose) the phosphate group of the PTS protein is transferred to the sugar and $EIIA^{Glc}$ becomes mainly unphosphorylated. This type of $EIIA^{Glc}$ inhibits the transport of various sugars into the cell by binding to the non-PTS sugar permeases. Hence, when both glucose and lactose are present, the inhibition of lactose

transport for the activation of *lac* operon by the transport of glucose is called *inducer exclusion* (Görke & Stülke, 2008).

C) When the glucose is present in the medium, transcription of the intrinsic *lac* operon of *E. coli* is down regulated. The *lac* repressor (*lac I*) is bound to the *lac* operator and prevents transcription of the *lac* operon genes encoding, β-Galactosidase (*lac Z*), lactose permease (*lac Y*) and β-galactosidase transacetylase (*lac A*) (see Figure 3.18). Following the depletion of glucose within the cell, lactose can be utilised by the cells and converted to allolactose. This binds to the *lac* repressor protein and results in a conformational change preventing the binding to the *lac* operator and rendering the access of RNA polymerase possible to the promoter. (Figure 3.18). c-AMP is also required to bind to the catabolite activator protein (CAP) to the promoter. The level of c-AMP is inversely proportional to the level of glucose, when glucose is depleted, the adenylate cyclase is activated and c-AMP is formed (Figure 1.17).

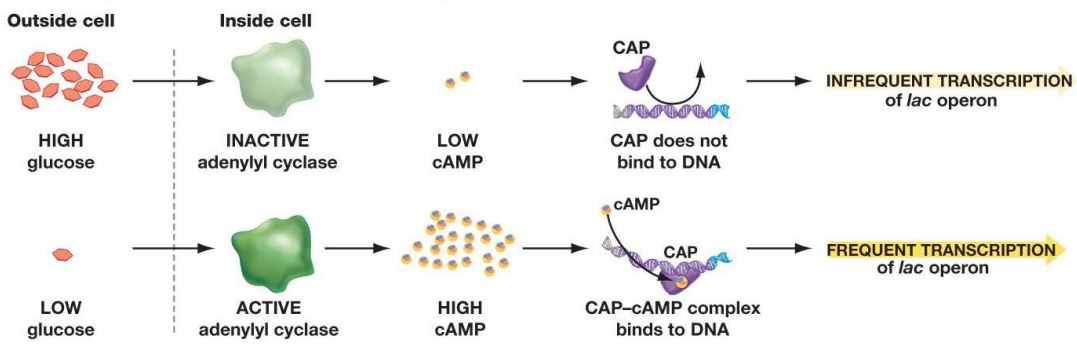


Figure 1.17 The relationship between the concentrations of c-AMP and glucose affecting the transcription of *lac* operon (Pearson Education, Inc., 2011).

Binding of c-AMP to the CAP activates the transcription by increasing formation of the closed RNA polymerase complex. Then transcription of the genes of the *lac* operon by RNA polymerase is carried out thereby further maintaining the induced gene expression by increasing the uptake of lactose (Deacon,2008). It appears that in *E.coli* the products of carbohydrate metabolism themselves are responsible for switching between glucose and lactose use (Brown *et al.*, 2008) (Figure 1.18).

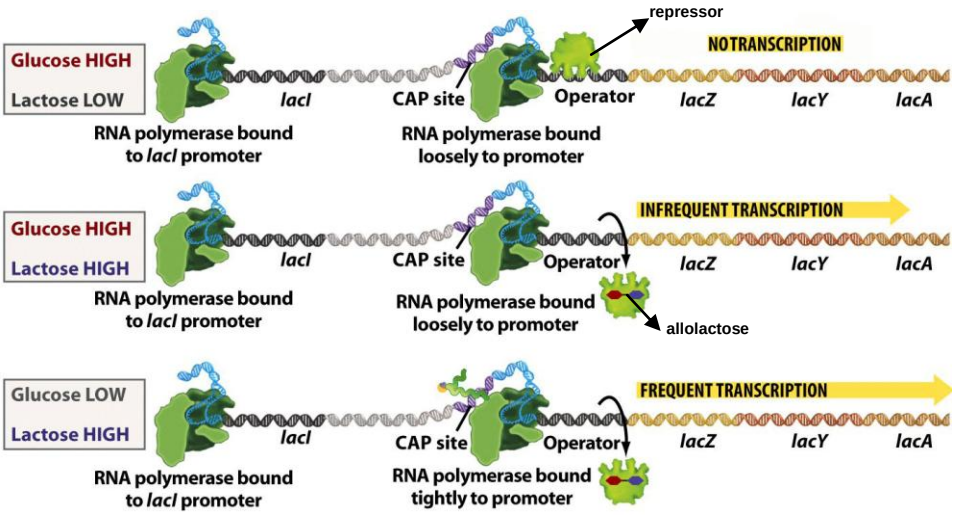


Figure 1.18 Catabolite repression in the inducible *lac* operon (adopted from Pearson Prentice Hall,Inc., 2005).

4. Similarly production of T7 RNA polymerase in DE3 lysogen cells is controlled by the *lacUV5* promoter, thereby when glucose is present transcription of the T7 RNA polymerase gene is inhibited since the *lac* repressor protein binds to the *lac* operator. Upon the depletion of glucose the cells begin to utilise lactose and transcription of the T7 RNA polymerase is initiated allowing transcription of the target gene. Additionally the target gene is under the control of the T7 promoter which is also regulated by binding of the *lac* repressor. Glycerol is added to the growth media during auto-induction, which can be used as a carbon source in the absence of lactose, but it does not interfere with the uptake and metabolism of lactose required for protein expression (Deacon,2008).

Induction by lactose is prevented by glucose, amino acids (during log-phase growth) and high rates of aeration (low lactose concentrations). Auto-induction allows efficient screening of many clones in parallel in terms of expression and solubility, as cultures have only to be grown to saturation following inoculation.

During the work of auto-induction (Studier, 2005) a number of parameters that helps improving the yields of proteins were identified, including the use of a lower concentration of buffer, altering the concentrations of glucose, glycerol and lactose, and using richer broth. Therefore, Deacon (2008) tested a number of these parameters for GAO production; like IPTG induction, Overnight Express Auto-Induction System (Novagen), ZY-LAC auto-induction media comparison, low salt buffer (50mM NPSC) and high salt buffer (100 mM NPS), increasing glucose, glycerol and lactose concentrations, sodium succinate effect, temperature and comparisons of lysis procedures. The optimum conditions for obtaining high levels of functional galactose oxidase tested for MatGOMN6 construct were found to be growth in 8ZY-4 x LAC - 25 mM sodium succinate media for 48 hours (Deacon, 2008) as indicated in Table 1.3. As it was reported by Deacon (2008), the use of auto-induction media

developed by Studier (2005) gave rise to an overall enhancement of 120000 fold in soluble functional galactose oxidase. By this expression system, the yield was 240 mg/L compared to 2 µg/L for wild-type GAO expression, which was more efficient than previously reported methods with *Pichia pastoris* yielding 40-50 mg/L (Deacon *et al.*, 2004), *Aspergillus nidulans* 30-50 mg/L (Baron *et al.*, 1994) and *E.coli* 10.8 mg/L (Sun *et al.*, 2001).

Table 1.3 Comparison of MatGOMN6 enzyme activity and soluble protein expression levels in various auto-induction media. Expression conditions of the wild- type GAO and the optimum expression conditions for MatGOMN6 are highlighted in red (Deacon, 2008).

Construct	Media & Induction	Temp / °C	Activity / µmol H_2O_2 min^{-1} L^{-1}	Estimated Yield / mg L^{-1}	Fold Increase / WT	Cost Per L of Media	Cost Per mg Enzyme
WT	LB/IPTG	37	2	0.0021	1	£1.9	£930
MN6	LB/IPTG	37	220	0.079	40	£1.9	£23
MN6	LB/IPTG	25	9100	10	5100	£1.9	£0.18
MN6	Overnight Express	37	250	0.27	140	£50	£180
MN6	ZY + LAC	37	26	0.029	15	£1.0	£35
MN6	ZY + LAC	25	19000	21	11000	£1.0	£0.048
MN6	8ZY + 2LAC	25	100000	120	57000	£5.8	£0.050
MN6	8ZY + 4LAC + Succinate	25	220000	240	120000	£5.9	£0.025

1.8 Aims of the Study

As far as it is known, GAO is produced as a precursor protein with 8- aa putative pre-peptide, 17- aa pro-peptide in addition to an N-terminal signal sequence. It is cleaved in a self-processing manner, requiring only the participation of copper and dioxygen. At the same time, GAO is a member of fast growing class of proteins bearing their instrinsic cofactors. It has a thioether bond at the active site and the tyrosine residue acts as a second redox center. It is believed that the removal of 17-amino acid pro-peptide is preceding the formation of the thioether bond at the active site since a form with both pro-peptide and thioether bond has not been detected yet, thus the order of events for maturation of the protein has not yet been definitively determined.

In our previous study, when the GAO was expressed in *E.coli* with its 8-aa putative pre- and 17-aa pro-peptide, it was found out that the pre-pro- peptide was not processed on addition of copper and oxygen in the presence of protease inhibitors.

Now, in this thesis the primary aim is to analyze the pro-peptide autoprocessing of recombinant GAO in the absence of its 8-aa pre-peptide which may hinder the self-cleavage

in *E.coli.* The deletion of 8-aa pre-peptide will be carried out by *QuikChange® Site-Directed Mutagenesis* approach. Following expression and purification studies, the protein bands will be characterised by SDS-PAGE in parallel with enzymatic activity analysis to find out whether the pro-peptide is able to be processed or not when the putative short pre-peptide is removed.

Clarification of the mechanism is of particular importance, since this will give insight to such cross-linked cofactor biogenesis in other enzymes, providing a solution for modelling and theoretical studies. A greater understanding of these modifications would allow incorporation of similar centres into designed or adapted enzymes (Firbank *et al.*, 2001), and also would provide a proper and useful model for copper-dependent proteolytic events that can occur in certain disease states, such as Alzheimer's disease (Firbank *et al.*,2003). Furthermore, the ability of self-processing may play a key role in the evolution of enzymes (Firbank *et al.*, 2001).

On the other hand, the second aim of this thesis is performing fundamental studies for foundation of a split-GAO system by deletion mutagenesis studies. GAO will be tried to be split into two parts by *QuikChange® Site-Directed Mutagenesis* approach and will be expressed heterologously in *E.coli* in order to be used as a biosensor for a prospective project at the University of Leeds in the future.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals, Enzymes and Equipments

All chemicals, molecular size markers, amplification and restriction enzymes, dyes and chromogenic substrates used in this study were supplied from Sigma-Aldrich (Germany), AppliChem (Germany), Fluka (Switzerland), Riedel-de Haën (Germany), Merck (Germany), Invitrogen-Life Sciences (USA), Fermentas, (now is a part of Thermo Fisher Scientific, USA), Stratagene (USA), Genemark (Taiwan), Biobasic (Canada), Novagen (Germany), New England Biolabs (USA), Promega (USA), Expedeon (UK) and Epicentre, Illumina Inc.(USA). Plasmid Isolation Kits were from Qiagen (Netherlands) and Fermentas (USA). In addition to the use of Eppendorf Mastercycler® (Germany) and MWG Primus thermocyclers, G-Storm GS2, (UK) gradient thermal cycler was also used for DNA amplification. Nucleic acid gel electrophoresis was performed by BlueMarine 100 Horizontal Electrophoresis Unit (SERVA Electrophoresis GmbH, Germany) and horizontal gel electrophoresis units of Scie-Plas Green Range™ (UK). The agarose gels were monitored by Bio-Rad Gel Doc™ XR (Bio-Rad, USA). NanoDrop® Spectrophotometer ND-1000 and its operating software v3.7 (Thermo Fisher Scientific, USA) was used for measuring DNA concentration. Centrifugation was done either by benchtop centrifuges (Eppendorf MiniSpin® Plus, Germany; Sigma, Germany) or Sorvall centrifuge (Thermo Fisher Scientific, USA). ProteoSpin™ Inclusion Body Isolation Micro Kit (Norgen, Biotek Corporation, Canada) was used for IB detection. Labsonic®P was used for disruption of the cells (Sartorius, Germany). Amicon® Ultra-4 and -15 Centrifugal Filter Units (Millipore®, Germany) as centrifugal concentrators and Whatman® membrane filters (GE Healthcare Life Sciences, USA) were used to concentrate proteins and to filter sterilize buffers for purification, respectively.Concentrating the proteins was also performed by Concentrator 5301 (Eppendorf, Germany). For degassing the chromatography solutions Nalgene® (USA) filtering system was used. GradiFrac (Pharmacia Biotech, Sweden- now is a part of GE Healthcare Life Sciences, USA) was used as FPLC purification system and the *Strep*-Tactin® Sepharose® (Amersham Pharmacia Biotech, Sweden- now is a part of GE Healthcare Life Sciences, USA) and Macro-Prep® (Bio-Rad Laboratories GmbH, Ltd.,USA) columns were supplied from IBA GmbH's (Germany). For detection of proteins SDS-PAGE was performed by using Blue Vertical 102 Electrophoresis System (SERVA Electrophoresis GmbH, Germany) and Mini-PROTEAN 3 System (Bio-Rad, USA). XCell SureLock™ Mini-Cell (Invitrogen-Life Sciences, USA) was used for electroblotting.Western Blotting antibodies were supplied from Novagen (Germany) and Alpha Diagnostic International (USA) for *Strep*-tag® and from Sigma-Aldrich (Germany) for

His-tag.Enzyme activity measurements were carried out by temperature-controlled UV Vis spectrophotometer (UV 1700) (Schimadzu, Japan).

2.1.2 Growth Media, Buffers and Solutions

The composition and preparation of the growth media, buffers and solutions are given in Appendix A.

2.1.3 Vectors and Molecular Markers

pET101/D-TOPO® (Invitrogen) and pET-28c(+) (Novagen) vectors are illustrated in Appendix C.

DNA and protein molecular markers are given in Appendix E.

2.1.4 Constructs

The deletion studies of pre-peptide on *gao* gene (Appendix B) was carried out by pPre-ProGON1 (essentially wild-type *gao gene*), and pPre-ProGOMN1 constructs (*gao* gene mutated according to the directed evolution results) constituted on pET101/D-TOPO® derived vector (Appendix C). pMatGOMN6 construct, which consists of the mature galactose oxidase coding region was also used for comparison (Figure 2.1). These constructs were kindly provided by Dr. Sarah E. Deacon who cloned the *gao* gene into the pET101/D-TOPO® together with an in-frame $(Gly)_4Ser$ linker followed by a Strep-*tag* II coding sequence and introduced the mutations giving S10P, M70V, P136 silent, G195E, V494A and N535D one at a time by *QuikChange® Site-Directed Mutagenesis* strategy as previously performed by directed evolution, by F. Arnold and colleagues (Sun *et al.* 2001). All constructs also carried N-terminal silent mutations, which provided high-level of expression.

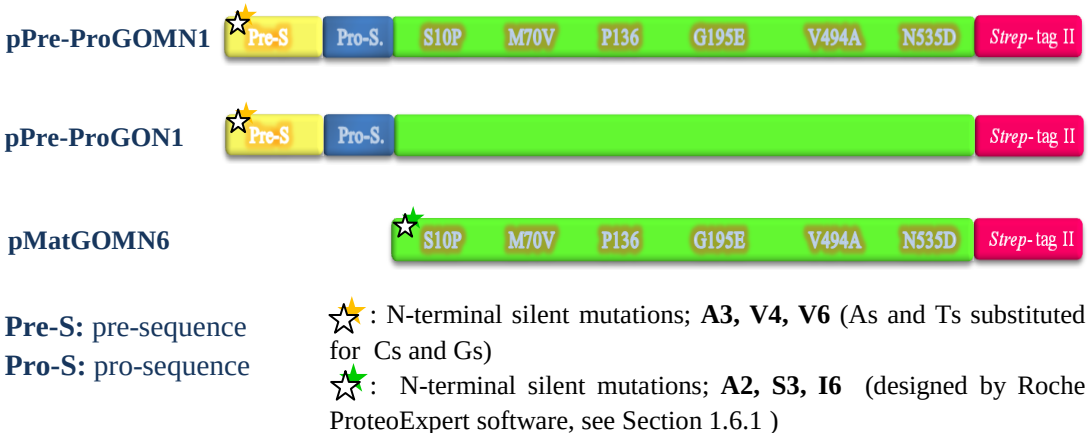


Figure 2.1 Schematic representation of the constructs used for pre-peptide and domain deletion studies.

Purification tag exchange and domain deletion studies of mature *gao* gene (MatGOMN6) were conducted on a construct with pET-28c(+) vector (pET28cStrepTEVMN6, shown in Appendix C). This construct was kindly provided by Prof. Michael J. McPherson.

2.1.5 Bacterial Strains

For propagation and maintenance of the plasmids *E.coli* strain XL1-Blue (Stratagene, USA) was chosen. Another *E.coli* strain, BL21 Star (DE3) (Invitrogen) was used for high-level heterologous expression.

Table 2.1 Bacterial strains and their genotypes

STRAIN	GENOTYPE
XL1-Blue	*rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, lac, [F'*pro*AB,*lac*IqZ.M15, Tn10 (*tet*r)] (Bullock *et al*., 1987)
BL21 Star (DE3)	F^- *ompT* $hsdS_B$ ($r_B^- m_B^-$) *gal dcm rne131* (DE3)

2.1.5.1 XL1-Blue

XL-1 Blue strain is chosen for many all-purpose cloning experiments. It allows blue-white color screening, single-strand rescue of phagemid DNA and preparation and maintenance of high-quality plasmid DNA. In addition, transformation efficiency is high ($\geq 1 \times 10^9$ transformants/μg DNA).

2.1.5.2 BL21 Star (DE3)

BL21(DE3) strain is one of the most widely used expression hosts for the production of recombinant target proteins. It has the advantage of being naturally deficient in both *lon* and *ompT* proteases, which results in a higher yield of intact recombinant proteins. The host is a lysogen of bacteriophage DE3, a lambda derivative that has the immunity region of phage *21* and carries a DNA fragment containing the *lacI* gene, the *lacUV5* promoter under the control of isopropyl-β-D-thiogalactopyranoside (IPTG), and the gene for T7 RNA polymerase (Studier and Moffatt, 1986; Novy and Morris, 2001). Such strains having T7-regulated genes are used to induce high-level protein expression.

The BL21 Star (DE3) *E. coli* strain is derived from the BL21(DE3). In addition to the λ DE3 lysogen, the BL21 Star strains also contain the *rne131* mutation. *rne131*gene encodes the RNase E enzyme, an essential, 1061 amino acid *E. coli* endonuclease which is involved in rRNA maturation and mRNA degradation as a component of a protein complex known as a degradosome (Grunberg-Manago, 1999; Lopez *et al*., 1999).

The N-terminal portion of RNase E (approximately 584 amino acids) is reported to be necessary for rRNA processing and cell growth as the C-terminal part of the enzyme (approximately 477 amino acids) is required for mRNA degradation (Kido *et al*., 1996; Lopez *et al*., 1999). The *rne131* mutation (present in the BL21 Star strains) encodes a truncated RNase E which lacks the C-terminal 477 amino acids of the enzyme required for mRNA degradation (Kido *et al*., 1996; Lopez *et al*., 1999). Thus, mRNAs expressed in the RNase E-defective BL21 Star strains exhibit increased stability in addition to enhanced yields of recombinant proteins with more mRNA available for protein translation.

2.2 Methods

2.2.1 Cultivation, Maintenance and Storage of the Strains

For maintenance of plasmids, *E.coli* XL1-Blue cells were inoculated into 5ml LB Broth containing ampicillin and carbenicillin or kanamycin with a final concentration of 50 µg/ml. The cells were grown overnight at 37 ˚C at 200 rpm in an orbital shaker, whereas the plates were grown in a stationary incubator. *E.coli* cells were incubated at 37 ˚C unless otherwise is stated.

For short time culture storage, each strain with appropriate plasmids was kept on LB agar plates containing proper antibiotics (Appendix A) at 4˚C, and the plates were refreshed periodically.

In order to store cultures for long term, glycerol stocks were prepared both for XL1-Blue and BL21 Star (DE3) strains carrying all the *gao* gene variants including pre-peptide and domain deletions and were stored at -80 ˚C following cryogenic freezing in liquid nitrogen. For 1 ml glycerol stock; 800µl of the overnight culture was mixed with 200 µl previously sterilized 100% glycerol in a cryotube.

The growth conditions for expression of *gao* variants in *E.coli* BL21 Star (DE3) cells will be given in Section 2.2.6.3 in detail.

2.2.2 Competent *E .coli* Cell Preparation

2.2.2.1 Heat Shock Competent *E. coli* Cell Preparation with $CaCl_2$

A single colony was picked from a freshly streaked agar plate of the desired strain and was used to inoculate 5 ml of LB broth, and was incubated overnight at 37˚C at 200 rpm in an orbital incubator. 1 ml of overnight culture was used to inoculate 100 ml LB broth and cells were grown at 37°C to OD_{550} of 0,4-0,5. The culture was dispensed into two 50 ml conical centrifuge tubes. After chilling on ice for 10 minutes, the tubes were centrifuged at 6000 rpm for 5 minutes at 4˚C. The supernatant was discarded, then the cells in each tube were resuspended in 25 ml of ice-cold solution I (total volume of 50 ml), (Appendix A) and kept on ice for 15 minutes. Following centrifugation at 6000 rpm for 5 minutes at 4°C, the pellet in each tube was resuspended in 3,5 of ml ice-cold solution I (total volume of 7 ml) (Appendix A). Then, glycerol was added to a final concentration of 20%, and 300 µl aliquots were dispensed into microcentrifuge tubes and stored at -80°C, until use.

2.2.2.2 Heat Shock Competent *E. coli* Cell Preparation with $RbCl_2$

Either *E.coli* XL1-Blue or BL21 Star (DE3) was streaked from glycerol stock onto a LB agar plate (Appendix A) and was grown overnight at 37°C. After inoculation of a single colony into 5 ml of SOB (Appendix A) (LB can be used instead) cells were grown overnight at 37°C at 200 rpm in an orbital incubator. 1 ml overnight culture was used to inoculate 50 ml pre-warmed fresh SOB media (or LB) and the culture was grown under same conditions until OD_{595} reached 0,4 (2-3 hrs). Cells were transferred into a 50 ml conical centrifuge tube and chilled on ice for 5 minutes, then centrifuged (Hermle rotor) at 3000 rpm for 10 minutes. The supernatant was discarded and the pellet was resuspended in 20 ml of ice-cold Tfb1 buffer and chilled on ice for 5 minutes. Following the centrifugation of cells at 3000 rpm for 10 minutes and discarding the supernatant, the cells were resuspended in 2 ml of Tfb 2 buffer and incubated on ice for 15 minutes. 250 µl aliquots were dispensed into sterile microcentrifuge tubes, competent cells were frozen in liquid nitrogen and stored at –80°C.

2.2.3 Mutagenesis Procedures

Two types of techniques were used for deletion studies, which were actually based on *QuikChange® Site-Directed Mutagenesis* approach.

2.2.3.1 QuikChange® II Site-Directed Mutagenesis Kit

In vitro site-directed mutagenesis is a very important technique in order to understand and characterize the structure-function relationships and to study gene expression. *QuikChange® Site-Directed Mutagenesis Kit* (Stratagene) is used to delete or insert single or multiple amino acids, make point mutations and switch amino acids. *QuikChange® II Site-Directed Mutagenesis Kit* is preferred since acccuracy is much more essential during deletion studies in which a high fidelity (HF) DNA polymerase is used to modify a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest by two synthetic oligonucleotide primers with desired mutations (each complementary to opposite strands of the vector). The procedure is simply based on the extension of these primers during thermal cycling giving a mutant plasmid with staggered nicks, treatment of the product with *Dpn*I endonuclease and repair of the nicked plasmid following transformation into XL1-Blue cells. The basic procedure is demonstrated in Figure 2.2 ("Stratagene QuikChange® II", 2007).

It is even easier to use the *QuikChange II* kit for large deletions since it renders the initial PCR reaction and product purification unnecessary.This protocol, with minor alterations, was also used by utilizing KOD Hot Start DNA polymerase (Novagen) for domain deletion studies.

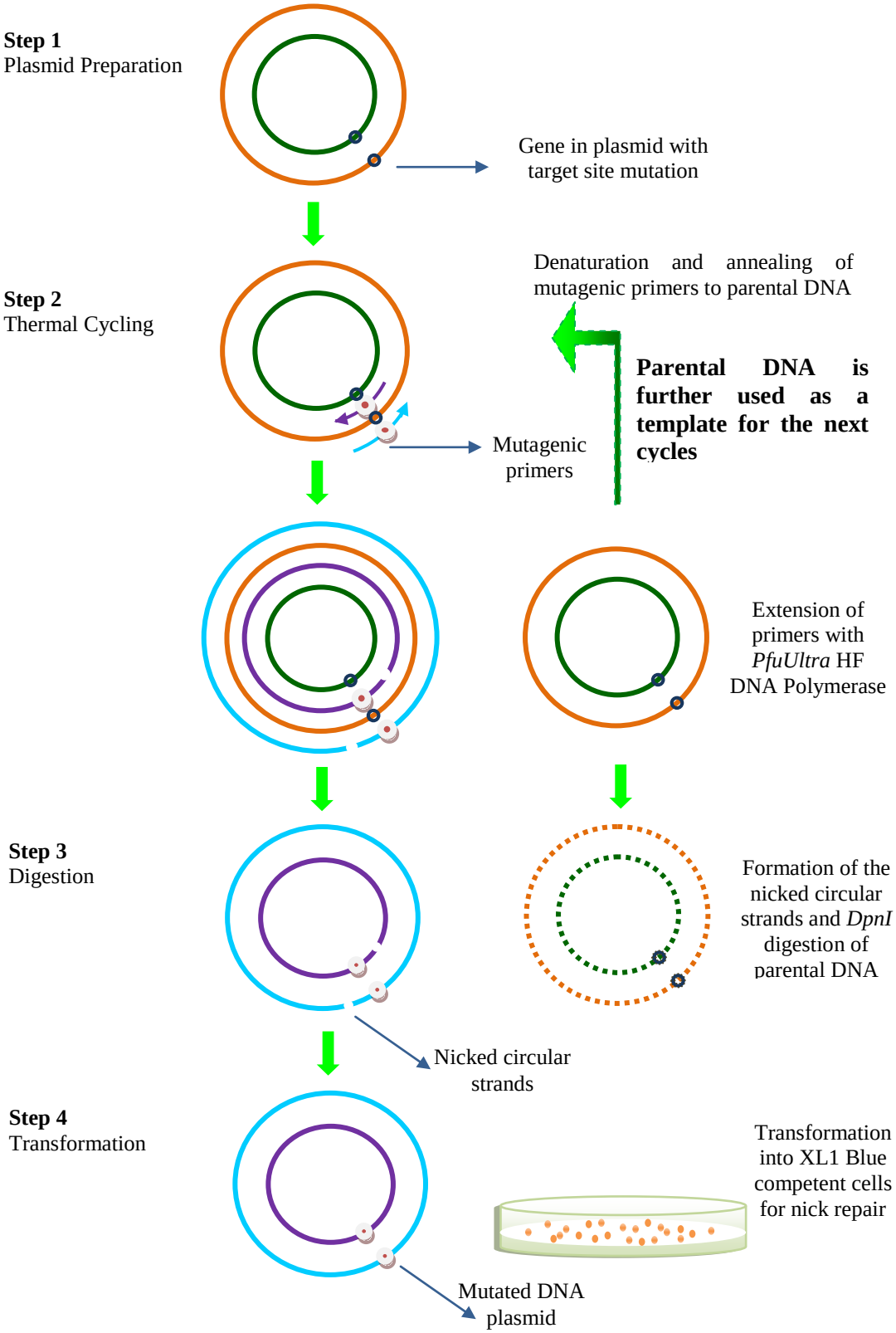


Figure 2.2 Overview of the QuikChange® II Site-Directed Mutagenesis procedure.

In this thesis, both pre-peptide deletion on pPre-ProGON1 and pPre-ProGOMN1 and Strep-tag /His tag exchange and domain deletion studies on pET28cStrepTEVMN6 constructs were done by using *QuikChange® II Site-Directed Mutagenesis* (Stratagene) approach by following the four major steps detailed below (for schematic illustration see Figure 2.2):

1. Plasmid preparation
2. Mutant strand synthesis with mutagenic primers (thermal cycling)
3. *Dpn*I digestion of parental DNA
4. Transformation and repair of the newly synthesized mutated DNA

2.2.3.1.1 Mutagenic Primer Design

Primer designation which is the critical point in *QuikChange®* was as follows: All of the primer pairs were designated meticulously containing the desired mutations and each annealed to the same sequence on opposite strands of the plasmid. The lengths of the primers were between 20 and 45 bases with a melting temperature (*T*m) of 78°C or higher,calculated by Stratagene QuikChange Primer T_m Calculator (www.stratagene.com/QPCR/tmCalc.aspx) or http://depts.washington.edu/bakerpg/primertemp/primertemp.cgi. Primers carry the desired mutation (deletion or insertion) at the centre of the in the middle with 14–16 bases of correct sequence on the flanking region. Optimal GC content of the primers should be 40% and higher and there should be one or more C or G bases at both ends. ("Stratagene QuikChange® II", 2007). For larger domain deletions 20 to 25 bp is left on either of the sides to bind the template DNA upstream or downstream of the region to be deleted. However, the largest oligo required for most of the deletions should not exceed 60 bp (Wang, 2010). It is recommended to purify the primers in order to prevent unwanted mutations, and to increase mutation efficiency.

Initially primers were resuspended in deionised water to 100µM . Then, to make them ready-to-use, the concentrations were adjusted to 125 ng/µl, as indicated in *QuikChange® II Site-Directed Mutagenesis Kit* manual and primers were kept at -20 °C, until use.

2.2.3.1.2 Thermal Cycling with Mutagenic Primers

During mutant strand synthesis with mutagenic primers, the 50 µl reaction was set up according to the manufacturer's instructions by slight modifications as follows:

		Final concentration
• 36 µl	Sterile ddH_2O	
• 5 µl	10 x Reaction Buffer	1x
• 5 µl	dNTP mix	0.2mM
• 1 µl	dsDNA template	0,4 ng/µl
• 1 µl	5' oligonucleotide primer	2,5 ng/µl
• 1 µl	3' oligonucleotide primer	2,5 ng/µl
• 1 µl	*PfuUltra* High-Fidelity DNA Polymerase	0,05 U/µl

$MgSO_4$ was not used since 10 x reaction buffer supplied with *PfuUltra* High-Fidelity DNA Polymerase contains sufficient amount of Mg^{2+}. Besides its high accuracy (exhibits an average error rate three-fold lower than *PfuTurbo* DNA polymerase and 18-fold lower than *Taq* DNA polymerase, making it the highest fidelity enzyme available) *PfuUltra* High-Fidelity DNA Polymerase renders robust amplification of long genomic targets. It also includes a key component; ArchaeMaxx polymerase-enhancing factor which has the capability for greater target length, eliminates a PCR inhibitor and promotes shorter extension times ("Agilent Technologies, Inc.", 2009).The ArchaeMaxx factor achieves the higher yield of products by overcoming dUTP poisoning, which is a result of dUTP accumulation during PCR through dCTP deamination (Hogrefe *et al*., 2002).

Both *Taq* DNA Polymerase (Genemark) and *Pfu* DNA Polymerase (Genemark, and Biobasic, Canada) were used by adding 1mM MgSO4 for optimization studies before using *PfuUltra* High-Fidelity DNA Polymerase.

This protocol, with minor alterations, was also used by utilizing KOD Hot Start DNA polymerase (to a final concentration of (1U/ µl) (Novagen (Germany) with 1mM $MgSO_4$, for purification tag exchange and domain deletion studies.

Mineral oil was not added to prevent the evaporation since the thermal cycler had a hot-top assembly. Each reaction was optimized and performed by using the cycling parameters outlined in Table 2.2 below.

Table 2.2 Thermal cycling parameters for *QuikChange® II Site-Directed Mutagenesis* method

Segment	Cycles	Temperature	Time
1	1	94˚C	30 seconds
2	25	94˚C	30 seconds
		55˚C	1 minute
		68˚C	8 minutes (1 minute/kb of plasmid length)

For purification tag exchange and domain deletion studies, thermal cycling parameters were the same except annealing temperature. Temperatures ranging between 53,1˚C and 61,2 ˚C were chosen and tried to catch the optimum annealing temperature for expected band sizes by a gradient thermal cycler for 24 cycles (G-Storm GS2).

2.2.3.1.3 Template DNA Elimination by *DpnI* Digestion

After thermal cycling the amplification products were treated with *Dpn*I restriction enzyme which is used for eliminating the parental DNA, to attain only the mutation carrying newly synthesized DNA. It acts only on the methylated and hemimethylated DNA (target sequence: 5'-Gm^{6}ATC-3') isolated from *E. coli*, however, it cannot digest the unmethylated DNA synthesized by thermal cycling, *in vitro*.

1 µl of the *Dpn*I restriction enzyme (10 U/µl) was directly added to each amplification reaction and thoroughly mixed by pipetting the solution up and down gently several times. Reaction mixtures were centrifuged for 1 minute and then incubated at 37°C for 1 hour to digest the parental (i.e., the nonmutated) supercoiled dsDNA. Following *Dpn*I treatment, the nicked vector DNA was transformed into *E. coli* in order to repair the nick.

2.2.3.2 Two-Step QuikChange® Site-Directed Mutagenesis Approach

The protocol of Wang and Malcolm (1999), a two-stage procedure based on the *QuikChange®Site-Directed Mutagenesis* Protocol, was followed with some modifications.
In stage one, two extension reactions are performed in separate tubes; one containing the forward primer and the other containing the reverse primer in order to prevent the primer dimerization. Subsequently, the two reactions are mixed in a single tube, and the QCSDM procedure is carried out.
Each single primer extension reaction was set up in a 0,5 ml thin wall PCR tube according to the components and amounts given below, using Mastercycler® personal (Eppendorf). Before the addition of 1 µl *Pfu Ultra* High-Fidelity DNA Polymerase (2,5 U/µl) (Stratagene) in the 10 x reaction buffer containing 49 ml reaction mixture, the tubes were preheated to 95°C for 3 min for initial denaturation. Then 1 µl *PfuUltra* High-Fidelity DNA Polymerase, was added into each reaction tube and mixed well. Reactions were initiated by 6 cycles of 95°C for 30 s, 55°C for 1 min and 68°C for 15 min. The reaction tubes could be kept at 4°C. Following the completion of single-tube extension reactions, each 50 ml pair was mixed in one tube and subjected to same conditions as in single primer extension reaction for 17 cycles.

	Final concentration	Tube 1	Tube 2
• Sterile ddH_2O		35µl	35µl
• 10 x Reaction Buffer	1x	5µl	5µl
• dNTP mix	0.2mM	5µl	5µl
• dsDNA template	1 ng/µl	2,5µl	2,5µl
• 5' oligonucleotide primer	3,75 ng/µl	1,5µl	-
• 3' oligonucleotide primer	3,75 ng/µl	-	1,5µl
Final volume		49µl	49µl

Table 2.3 Comparison of the single-primer and double-primer QCSDM reactions (adopted from Edelheit *et al.*,2009)

	Single-primer QCSDM	**Double-primer QCSDM**
Number of reactions	2 (1 for each primer) +1	1
Plasmid template	~50 ng	~20 ng
Primers	Forward or reverse (each in a **separate** tube)	Forward and reverse **together**
Primer-primer annealing during thermal cycling	-	+
Number of cycles	6+17	25
Mutation probability during thermal cycling	Low	High
DNA polymerase	*PfuUltra* High-Fidelity DNA Polymerase	*PfuUltra* High-Fidelity DNA Polymerase

The products can be visualized by agarose gel electrophoresis and following *Dpn*I digestion can be transformed into *E.coli* cells.

2.2.4 Transformation Procedure of *E.coli* XL1-Blue Cells

XL1-Blue competent cells (Stratagene) were thawed on ice gently. For (-) control and each sample reaction to be transformed, 50 µl of the competent cells was aliquoted to prechilled microcentrifuge tubes. Except the (-) control, separate aliquots of the competent cells were treated with 1 µl of the *Dpn*I-treated DNA (20 ng/µl). The competent cells with/without DNA were gently mixed and incubated on ice for 30 minutes. They were heat-shocked at 42°C for 45 seconds and immediately incubated on ice for 2 minutes. 500 µl of preheated SOC medium (Appendix A) was added to the transformation reactions and tubes were placed into the orbital incubator carefully and grown at 37°C for 1 hour with shaking at 200–225 rpm. 150-200 µl of each transformation reaction was aliquoted onto an LB agar plate containing the appropriate antibiotic (Appendix A), the rest of the culture was centrifuged at 6000 rpm for 2-3 minutes and following resuspension of the pellet in 250 µl fresh SOC medium, cells were plated onto another plate with proper antibiotic. The transformation plates were incubated at 37°C for >16 hours.

2.2.5 DNA Isolation and Analysis Techniques

2.2.5.1 Plasmid DNA Isolation with QIAprep® Spin Miniprep Kit

In order to obtain plasmid of interest QIAprep® Spin Miniprep Kit was used by following the supplied protocol ("Qiagen QIAprep®", 2006) based on modified alkaline lysis method. All steps of the protocol were conducted at room temperature.

For harvesting the cells carrying the desired plasmids; 5ml overnight culture with appropriate antibiotic was centrifuged at 8000 rpm for 3 minutes in a conventional microcentrifuge. RNase A was added to Buffer P1. The cells were resuspended in 250 µl Buffer P1 including LyseBlue reagent (for avoiding common handling errors that cause inefficient cell lysis and incomplete precipitation of SDS, cell debris and genomic DNA).

250 µl Buffer P2 was added and the tube was inverted gently 4-6 times and mixed thoroughly until a homogeneously blue colored suspension is obtained and the reaction was not allowed to proceed for more than 5 minutes . After the addition of Buffer P2, 350 µl N3 buffer was added, and mixed gently 4-6 times until a colourless suspension was obtained, indicating effective SDS precipitation. Then tubes were centrifuged at 13000 rpm for 10 minutes. The supernatants were applied to QIAprep spin columns by decanting or pipetting. After centrifugation for 30-60 seconds flow-through was discarded. Then column was washed by 0.75 ml ethanol containing PE Buffer, and centrifuged for 30-60 seconds. After discarding flow-through, an additional 1-minute centrifugation was carried out for residual wash buffer removal which might have interfered with subsequent enzymatic reactions due to the ethanol present. Column was placed into a clean 1.5 ml centrifuge tube. To elute DNA 50 µl ddH_2O was added to the centre of the column, let stand for 1 minute and then centrifuged for 1 minute at 13000 rpm.

2.2.5.2 Plasmid DNA Isolation with GeneJET™ Plasmid Miniprep Kit

GeneJET™ Plasmid Miniprep Kit ("Fermentas", 2008) works on a SDS/alkaline lysis principle to liberate the plasmid DNA by binding to the silica membrane in the spin column. For low-copy plasmids it is suggested to use up to 10 ml of overnight culture with appropriate antibiotic. Prior to first use, RNase A solution and ethanol (96-100%) should be added and mixed to the Resuspension Solution and Wash Solution, respectively. The pelleted cells were resuspended completely in 250 µl of the Resuspension Solution by vortexing or pipetting up and down in a microcentrifuge tube. Lysis Solution (250 µl) was added and mixed thoroughly by inverting the tube 4-6 times until the solution became viscous and slightly clear (not incubated for more than 5 minutes to avoid denaturation of supercoiled plasmid DNA). Then 350 µl of the Neutralization Solution was added and the tube was inverted 4-6 times immediately. For elimination of cell debris and chromosomal DNA the tube was centrifuged for 5 min. The supernatant was transferred to the supplied GeneJET™ spin column without disturbing or transferring the white precipitate. Following 1 min. centrifugation the flow-through was discarded and the column was placed back into the same collection tube. After adding 500 µl of the Wash Solution to the GeneJET™ spin column, it was centrifuged for 30-60 seconds and the flow-through was discarded. Washing step was repeated by using additional 500 µl of the Wash Solution to avoid residual ethanol in plasmid preps. The column was transferred into a fresh 1,5 ml microcentrifuge tube. For the elution of plasmid DNA, 50 µl of the Elution Buffer was added to the centre of the column membrane. After incubation at room temperature for 2 minutes, the column was centrifuged for 2 minutes. An additional elution step (optional) with Elution Buffer, carried out for recovery of the residual DNA, increased the overall yield by 10-20%. Purified plasmid DNA was stored at -20°C.

2.2.5.3 Plasmid DNA Isolation with Alkaline Extraction Procedure

Centrifugation of 5 ml overnight culture of bacterial cells carrying plasmid of interest was done at 6500 rpm at 4°C for 5 minutes. Following resuspension of cells in 200 µl Solution I and incubation at room temperature for 15 minutes (Appendix A), 200 µl Solution II was

added and the cells were gently inverted 7-8 times (Appendix A) prior to incubation on ice for precisely 5 minutes. After addition of Solution III, the mixture was slowly and thoroughly inverted 7-8 times and incubated on ice for 15 minutes. The tube was centrifuged at 13000 rpm at 4°C for 10 minutes. Supernatant was transferred into a new 1.5 ml microcentrifuge tube without disturbing the white precipitate. 2 volume of cold ethanol (96%) was added and kept at -80°C for nearly 1 hour (can be longer). Supernatant was discarded after centrifugation at 13000 rpm at 4°C for 10 min. Then the pellet was resuspended in 200 µl NE Buffer and incubated on ice for 1 hour. At this step, 5 µl plasmid DNA was loaded onto agarose gel (optional) for detection. After the suspension was centrifuged at 13000 rpm at 4°C for 15 minutes, the supernatant was transferred into a new 1.5 ml microcentrifuge tube. Following the addition of 2 volumes of cold ethanol (96%) it was kept at -20°C for approximately 30 minutes. Mixture was centrifuged at 13000 rpm at 4°C for 10 minutes and the supernatant was discarded. The pellet was resuspended in 20-30 µl ddH_2O after being air dried for 5-10 minutes. RNase treatment was carried out either after plasmid isolation (1/2-1 hour at 37 °C) or during the Solution I incubation (Appendix A).

2.2.5.4 Visualization of DNA by Agarose Gel Electrophoresis

For the investigation of DNA samples; after plasmid isolation, DNA manipulation (deletion) and restriction enzyme digestion, proper amount of agarose depending on the interest (1-5 %) was dissolved in 1 x TAE Buffer (Appendix A) by boiling the mixture in a microwave oven until homogeneous form was achieved (~3 minutes). In order to make DNA visible under UV light, after cooling the gel to 50-60 °C either ethidium bromide solution (Appendix A) or SYBR® Safe DNA Gel Stain (Invitrogen) was added to a final concentration of 0,5 µg/ml or 1x, respectively. The gel was poured into a tray, then a comb of proper size was placed and allowed to solidify for nearly 15 minutes. As it became completely polymerized, it was placed into an electrophoresis tank filled with 1 x TAE Buffer and combs were removed. Along with DNA size markers, the DNA samples were loaded into the wells of the agarose gel by mixing with 6 x loading dye (supplied with markers). SERVA BlueMarine 100 Horizontal Electrophoresis Unit (SERVA Electrophoresis GmbH) and horizontal gel electrophoresis units of Scie-Plas Green Range™ were used. Following separation of the DNA fragments by electrophoresis at 100 V for nearly 45 minutes, gels were visualized by gel imaging system, Gel Doc™ XR (Bio-Rad) and analysed by Bio-Rad Quantity One® software.

Table 2.4 Concentration of agarose gels used for size resolution of linear DNA molecules (adopted from Molecular Cloning: A Laboratory Manual 3rd Edition, Sambrook & Russel, 2001).

Agarose concentration (%, w/v)	DNA fragment range (kb)
0,3%	5-60
0,6%	1-20
0,7%	0,8-10
0,9%	0,5-7
1,2%	0,4-6
1,5%	0,2-3
2,0%	0,1-2

2.2.5.5 Determination of DNA Concentration

The below formula was used to calculate the DNA concentration by agarose gel electrophoresis:

DNA concentration: (DNA quantity in band) x (concentration of marker) x (volume of marker used) x (intensity ratio of bands) x (1/ loaded quantity of DNA)

On the other hand, the concentration of the plasmid DNA for sequencing submission was calculated by NanoDrop® Spectrophotometer ND-1000 and its operating software v3.7 (Thermo Fisher Scientific, USA) at Leeds University, UK ("ThermoScientific", 2008).

2.2.5.6 Restriction Enzyme Digestion

Restriction enzyme digestion was done for the detection of right clone after transformation; for the constructs pProGON1 and pMatGOMN6.

With *Nco*I, single enzyme digestions were performed.

Sample DNA : 20 µl plasmid DNA (14 ng/µl)
Buffer : 3 µl (10 x Buffer Tango, MBI Fermentas)
Restr.Enzyme : 0,5 µl *Nco*I (10U/µl)
Total volume : 23,5 µl

In *Pst*I and *Eco*RI double digestion, nearly 600-900 ng of pure plasmid DNA was needed, but less was used. 10 x Buffer O^+ was added to a final concentration of 1x.

Sample DNA : 18 µl plasmid DNA (30 ng/µl)
Buffer : 2,1 µl (10 x Buffer O^+, MBI Fermentas)
Restr.Enzyme : 1 µl (0.5 µl from each; *Pst*I and *Eco*RI) (10U/µl)
Total volume : 21,1 µl

In order to detect deletion, *Eco*RI single enzyme digestion and *Eco*RI and *Nco*I double digestion were performed.

Sample DNA : 23 µl plasmid DNA (30 ng/µl)
Buffer : 2,6 µl (10 x Buffer O^+, MBI Fermentas)
Restr.Enzyme : 0.5 µl *Eco*RI (10U/µl)
Total volume : 26,1 µl

Sample DNA : 23 µl plasmid DNA (30 ng/µl)
Buffer : 6 µl (10 x Buffer Tango, MBI Fermentas; 2x)
Restr.Enzyme : 1 µl (0,5 µl from each; *Eco*RI and *Nco*I) (10U/µl)
Total volume : 30 µl

Reaction mixtures were incubated overnight at 37°C. Then they were spun down by a microcentrifuge and analysed by properly prepared agarose gel electrophoresis.

For domain deletion studies, restriction enzyme digestions were made using *Cla*I and also *Pst*I seperately as a single cutter and with *Nco*I & *Nhe*I for double restriction enzyme digestions to test the main vector.

Sample DNA : 5 µl plasmid DNA (~870 ng/µl)
Buffer : 1 µl (10 x NEBuffer 4, New England Biolabs)
BSA : 1 µl (10 x BSA, (Appendix A)
Restr.Enzyme : 0,5 µl *Cla*I (5000U/ml)
ddH_2O : 2,5 µl
Total volume : 10 µl (1,5 h at 37°C)

Sample DNA : 5 µl plasmid DNA (~870 ng/µl)
Buffer : 2 µl (10 x NEBuffer 3, New England Biolabs)
BSA : 2 µl (10 x BSA)
Restr.Enzyme : 1 µl *Pst*I (5000U/ml)
ddH_2O : 10 µl
Total volume : 20 µl (1h at 37°C)

Sample DNA : 5 µl plasmid DNA (~870 ng/µl)
Buffer : 5 µl (10 x NEBuffer 2, New England Biolabs)
BSA : 5 µl (10 x BSA)
Restr.Enzyme : 2 µl (0.5 µl from each; *Nco*I & *Nhe*I) (5000U/ml)
ddH_2O : 33 µl
Total volume : 50 µl (Overnight at 37°C)

2.2.5.7 DNA Sequencing for Determination of GAO Deletion Mutants

Following the completion of transformation of pre-peptide deleted plasmids obtained by *QuikChange® II Site-Directed Mutagenesis* to the XL1-Blue cells, chosen colonies were grown in 5 ml LB with proper antibiotic at 37°C overnight and the plasmids isolated were submitted to REF-GEN Gene Research and Biotechnology, METU Technopolis, Ankara, Turkey for sequencing.

For the domain deletion studies carried out in U.K, plasmid samples concerning Domain I/II deletion and Domain III deletion were all sequenced at Beckman Coulter Genomics, Essex, UK.

Sequence alignment was done by using WSClustalW2/ClustalW2 services (http://www.ebi.ac.uk), chromatogram visualization and editing in addition to alignment were also performed by utilizing ChromasPro v 1.5 and Bioedit programs.

2.2.6 Expression of the Deletion Mutants of GAO in *E .coli*

2.2.6.1 Transformation Procedure of Expression Host *E.coli* BL21 Star (DE3)

BL21 Star (DE3) (Invitrogen) competent cells were left on ice to be thawed gently. 50 µl of the competent cells were transferred to a prechilled microcentrifuge tube for each transformation reaction. 1 µl of the *Dpn*I-treated plasmid DNA (20 ng/µl) was transferred to separate aliquots of the competent cells, except the (-)ve control. The transformation reactions were mixed gently and thoroughly and incubated on ice for 30 minutes. Then they were incubated at 42°C water bath for exactly 30 seconds and quickly placed on ice and kept for 2 minutes. 450-500 µl of SOC medium preheated to 42°C was added to the each transformation reaction and they were incubated at 37°C at 200–225 rpm for 1 hour. 150-200 µl of each transformation reaction was aliquoted onto an LB agar plate containing the appropriate antibiotic (Appendix A), the remaining transformation reaction was spun down at 6000 rpm for 2-3 minutes and plated onto a different plate with proper antibiotic in 250 µl fresh SOC medium in order to use at least two different volumes for ensuring well-spaced cells. The transformation plates were inverted and incubated at 37°C for >16 hours.

2.2.6.2 pET System : Regulation of T7 Promoter Based Protein Expression

The pET expression vectors, derived from plasmid pBR322 are one of the most powerful and most widely used systems yet developed for cloning and heterologous expression in *E. coli* (Rosenberg *et al*., 1987; Studier & Moffatt, 1986; Studier *et al*., 1990). The system is based on bacteriophage T7 elements and the expression of a recombinant protein is controlled by a strong T7 promoter which is modified to carry a *lac* operator sequence.

T7 promoter cannot be recognized by the RNA polymerases of the host itself, it specifically and selectively recognizes the T7 RNA polymerase.

Following the transformation of the plasmid in a non-expression host, a recombinant plasmid is especially transformed into an *E.coli* strain which carries the chromosomal copy of the T7

RNA polymerase gene. These λDE3 lysogen hosts (i.e BL21 Star DE3) have the immunity region of phage *21*, *lacUV5* promoter, *lacI* gene and T7 RNA polymerase gene which is regulated by the addition of IPTG (Isopropyl-β-thiogalactopyronoside) for expression of target proteins ("Novagen pET System", 2011).

It is necessary to deliver the T7 RNA polymerase to the cells by inducing the expression of polymerase in order to regulate the expression of the gene of interest. When the *lacUV5* promoter which is inducible by IPTG is activated, T7 RNA polymerase is expressed by the BL21 Star (DE3) and the polymerase binds to the T7 promoter on the vector and the expression of the target gene is initiated (Figure 2.3) ("Invitrogen Champion™ pET", 2010).

Studies have shown that, even in the absence of IPTG there is always some basal expression of T7 RNA polymerase from the *lacUV5* promoter in λDE3 lysogens (Studier & Moffatt, 1986). Although this is not problematic generally, in some cases the expressed recombinant protein may be toxic to the *E. coli* host and this basal expression may interfere with the growth and plasmid establishment. pET system has a powerful tool in order to prevent this toxicity. *lac* operator sequence placed downstream of the T7 promoter serves as a binding site for the lac repressor (encoded by the *lacI* gene) and functions to further repress T7 RNA polymerase-induced basal transcription of the gene of interest. On the other hand, low basal expression levels in the pET system can be maintained by supplementing the medium with glucose (Grossman *et al.*, 1998). When the cultures reach stationary phase, firstly glucose is consumed. Utilization of an alternate carbon source increases cyclic AMP (cAMP) levels, which is sufficient for stimulation of transcription from the *lac*UV5 promoter and subsequent expression of T7 RNA polymerase in λDE3 lysogens. Consequently, this stimulation leads to the increase of T7 promoter regulated target gene expression (Grossman *et al*., 1998; Kelley, 1995; Novy *et al*., 2001; Pan *et al*., 2000). Thus, based on the characteristics of the target protein to be expressed, addition of glucose to the medium plays an important role in keeping the basal transcription from *lac*UV5 promoter at low levels.

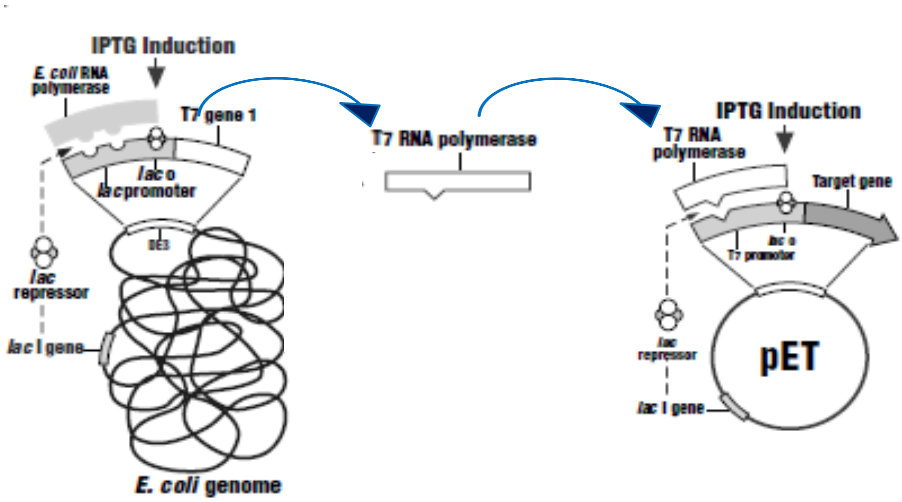


Figure 2.3 Illustration of the host and vector elements available for control of T7 lac promoter of the pET system for target gene expression ("Novagen pET System", 2011).

2.2.6.3 Heterologous Expression Studies in *E.coli* BL21 Star (DE3)

2.2.6.3.1 Expression by IPTG Induction

Day 1

An LB plate containing 50 μg/ml ampicillin was streaked from glycerol stock of BL21 Star (DE3) strain carrying gene of interest.

Day 2

5 ml LB with 50 μg/ml ampicillin was inoculated by a single colony picked from freshly streaked plate in a 50 ml conical centrifuge tube. Cells were grown at 37˚C at 200 rpm during the day. After 6 hours 25 μl small culture was used to inoculate fresh 25 ml LB with proper amount of antibiotic in a 100 ml flask and was grown overnight at 37˚C at 200 rpm.

Day 3

5 ml of the overnight culture was transferred into 250 ml LB containing 50 μg/ml ampicillin in a 1L flask and was incubated at 37˚C at 225 rpm to an OD_{595} of 0,4-0,5 for about 2 hours. As soon as the OD_{595} value was attained, culture was transferred to 25˚C. The OD_{595} was kept monitoring and when the value of 0.6-0.7 was reached the culture was induced by 1M IPTG (Appendix A) to a final concentration of 1 mM. Then the cells were grown at 25˚C at 225 rpm overnight. 1 ml pre-induction sample and 1 ml samples at 1, 2, 4 and 6 hour time-points were taken and kept at -20˚C so as to compare pre-induction and post-induction conditions.

Day 4

Cells were harvested after 16-21 hours by transferring into 250 ml Sorvall tubes and centrifugation at 6000 rpm at 4˚C for 10 minutes.

2.2.6.3.2 Expression by Auto-induction

Day 1

An LB plate containing 50 μg/ml ampicillin and 50 μg/ml carbenicillin (can be used together with ampicillin or instead of it, since it is much more stable than ampicillin) was streaked from glycerol stock of BL21 Star (DE3) strain carrying the gene of interest.

Day 2

A single colony picked from freshly streaked plates was initially grown in 2mL of ZYP-0.8G (Appendix A) containing ampicillin and carbenicillin (50 μg/ml each), at 37˚C and 220 rpm, for 6 hours, during the day. After 6 hour-incubation, 200 μl of culture was used to inoculate 400 mL of 8ZY-4LAC–25mM Succinate (Appendix A) in a 2 L baffled flask (generates more agitation and more aeration) and cells were grown overnight at 25^{o}C and 250 rpm.

Day 3

It was decided to perform a time course of 72 hours to monitor the cultures in order to detect the optimum growth time. 2 x 1 ml samples were taken from each flask twice per 24 hours. With one of the 1 ml sample, the optical density was measured and recorded at 595 nm. If

the measurements went above the range, 1/10 dilution of the culture was prepared by using 8ZY-4LAC–25mM Succinate medium. The other 1 ml sample was used to analyse expression on a small scale. Samples were kept as pellets at -20 ^{o}C after being centrifuged at 8000 rpm for 3 minutes.

Day 4
For the next 24 hours, the cultures were monitored at OD_{595},the values were recorded and1ml samples were collected for further analysis.

Day 5
As the previous two days, the cultures were kept monitoring and recording at OD_{595} and 1ml samples were collected for the last 24 hours. All the cells were harvested at the end of the 72 hours by centrifugation in Sorvall SLA3000 rotor at 5000 rpm at 4˚C for 15 minutes.
Auto-induction method details and the reasons how this media components were decided were given in detail in Section 1.7.2.

2.2.6.4 Lysis of the Bacterial Cells

The frozen pellets were thawed slowly at room temperature. Complete/Complete Mini, EDTA-free protease inhibitor cocktail tablet (ROCHE) or ProteoBlock™ Protease Inhibitor Cocktail (Fermentas), prepared according to the instructor's manual, was added to the resuspension buffers for each disruption procedure (Appendix A).

2.2.6.4.1 Sonication
The pellet was resuspended in 5 ml 1x PBS. After being resuspended and aliquoted into 50 ml conical centrifuge tubes the cells were disrupted by Sartorius ultrasonicator using 10 seconds on 40 seconds off for 7 cycles by placing the tubes on ice for cooling. Sonicated cells were centrifuged at 11000 rpm at 4˚C for 5 minutes (Sorvall). Supernatant was collected and transferred into clean 15 ml conical centrifuge tube.

2.2.6.4.2 Lysis Buffer

The thawed cell pellets of 400 ml culture were resuspended in 20 ml of Triton X 100 lysis buffer into which 20 mg lysozyme and 1 µl OmniCleave™ Endonuclease (Epicentre, Illumina Inc.,USA) and appropriate amount of protease inhibitor were added freshly before use. After complete resuspension, the mixture was gently mixed at 4˚C for 30 minutes on a roller mixer SRT-1 (Stuart, Bibby Scientific, UK). Soluble and insoluble fractions were separated by centrifugation at 16000 rpm at 4˚C for 45 minutes. If small scale expression analysis is in question, the mixing time is 20 minutes and centrifugation condition is 13000 rpm at 4˚C for 20 minutes.

2.2.6.5 Inclusion Body Isolation and Solubilization Techniques

2.2.6.5.1 Denaturation by Urea

Following the cell growth as described in Section 2.2.6.3.1 by IPTG induction (non-induced control was used), 50 ml of the culture was centrifuged at 6000 rpm at 4˚C for 10 minutes. Harvested cells were resuspended in 2 ml of binding buffer containing different molarities of urea (2M, 4M, 6M & 8M). Firstly, the samples were frozen at -80°C for 1 hour, then dissolved and vortexed. After that they were frozen at -80°C overnight and following dissolving they were vortexed. The cells were sonicated by bursts of 10 seconds on, 40 seconds off for 6 cycles. Cell extracts were centrifuged at 17000 rpm at 4˚C for 15 minutes. Supernatants were saved, pellets were resuspended in 500 µl binding buffer and centrifuged at 17000 rpm for 30 minutes. All collected supernatants were loaded on SDS-PAGE and fractions were analysed (adopted from Ayalew *et al.*, 2008).

2.2.6.5.2 ProteoSpin™ Inclusion Body Isolation Micro Kit

ProteoSpin™ Inclusion Body Isolation Micro Kit (NORGEN, Biotek Corporation, Canada) provides essential reagents for cell disruption, inclusion body solubilization and purification using spin column chromatography with silicon carbide (SiC) as an ion exchanger. Inclusion body isolation and purification steps were carried out according to the manufacturer's instructions as described below:

A. Cell Lysis and Isolation of Inclusion Bodies

All centrifugation steps are carried out at 13000 rpm in a benchtop microcentrifuge at room temperature.

At the end of the protein induction period, 1,5 ml of the bacterial culture was centrifuged for one minute in a microcentrifuge tube and supernatant was discarded. The pellet was frozen at -20˚C or flash frozen with liquid N_2. Then it was thawed at room temperature or at 37˚C to improve lysis efficiency. 200 µl of Cell Lysis Reagent was added to the bacterial pellet. The needle was assembled with 1 ml syringe (provided).The bacterial pellet was carefully disrupted by drawing it through the needle and ejecting the suspension back into the microcentrifuge tube 15 to 20 times. The crude extract was centrifuged for 10 minutes and supernatant was discarded without disturbing the pellet (The supernatant was saved in a fresh microcentrifuge tube for comparative analysis of the soluble proteins present in this fraction.). With the same needle-and-syringe assembly, again 200 µl of Cell Lysis Reagent was added to the tube and the pellet was carefully resuspended with a few passes through the needle. About 1,5 ml 10-fold dilution of the Cell Lysis Reagent was prepared with sterile deionized or MilliQ water. 600 µl of this solution was added to the 200 µl previously prepared suspension. This mixture was passed through the needle a few times and after centrifugation for 10 minutes supernatant was discarded. 800 µl of the diluted Cell Lysis Reagent was added to the pellet and resuspended using the needle-and-syringe, until homogeneity. Following 10 minutes centrifugation, the supernatant was discarded and it was ensured that the pellet was relatively dry by tapping.

B. Solubilization of Inclusion Bodies

50 µl of IB Solubilization Reagent was added to pelleted inclusion bodies and the pellet was dissolved by pipetting and vigorous vortexing. Once the pellet was dissolved (extra IB Solubilization Reagent could be added), a volume of sterile deionized water equal to the amount of IB Solubilization Reagent used, was added and mixed by vortexing. The recombinant deletion mutants of GAO were tried to be purified using the acidic procedure (it is assumed to be acidic below pI 8 according to the manual) depending on the pI of the recombinant GAO which is 7.76 for ProGON1 and 6.91 for ProGOMN1 (http://web.expasy.org/compute_pi/).

C. Purification of Acidic Proteins

Sample Preparation

50 µl of the dissolved protein sample was transferred to a fresh microcentrifuge tube and 200 µl deionized or Milli-Q water was added. The protein sample was prepared by adding 7,5 µl of pH Binding Buffer (acidic) to the sample and mixed by vortexing. This step was carried out to bring the pI of the sample to 4,5.

Column Activation

250 µl of Column Activation and Wash Buffer (acidic) was added to the column and the cap was closed. Following centrifugation for one minute at room temperature at maximum speed the flowthrough was discarded. Column activation was completed by repeating the centrifugation by additional washing step with same amount of buffer.

Protein Binding

During the binding step, centrifugation time provides sufficient contact time for the protein to bind the resin. 257,5 µl of prepared protein sample was applied onto the activated column and centrifuged for one minute. Flowthrough was the discarded part, however it was saved to check for binding.

Column Wash

250 µl of the Column Activation and Wash Buffer (acidic) was applied to the column and centrifuged for one minute. The flowthrough was discarded and the spin column was reassembled with its collection tube. Additional washing step and centrifugation was done so as to complete the column wash.The column was inspected to ensure that all liquid has passed through the resin and into the collection tube. In case there was liquid in the column, the column was spun down for an additional 1 minute to dry.

Protein Elution and pH Adjustment

5 µl of Neutralizer was added to a fresh 1,7 ml microcentrifuge tube and the spin column, from the Column Wash procedure, was transfer into the microcentrifuge tube. 25 µl of Elution Buffer was added to the column and centrifuged for one minute to elute the bound protein. Another 25 µl of Elution Buffer was added and centrifuged for one minute into the same microcentrifuge tube. Protein samples were analyzed by SDS-PAGE according to the protocol described in Section 2.2.8.4 ("NORGEN ProteoSpin™" ,2007).

2.2.7 Protein Isolation and Purification

After obtaining the supernatants of the mutant samples by lysis and centrifugation, supernatants were dialysed against two changes of 1L 1 x PBS buffer at 4°C for 3 hours, then they were left overnight for dialysis. Dialysis tubing (10k MWCO, Sigma-Aldrich) was prepared by boiling in dH_2O twice for 3 minutes and waited for cooling down. Then it was stored in 20% ethanol at 4°C.

2.2.7.1 Purification of GAO Deletion Mutants by *Strep*-tag®/*Strep*-Tactin® Affinity Chromatography

The next step after dialysis was the isolation and purification of the recombinant proteins intact and without activity loss. In order to gain the purified mutant GAOs with desired qualifications, *Strep*-tag®/*Strep*-Tactin® affinity chromatography (IBA GmbH's) developed for rapid one-column protein purification was used which works on the principle of affinity basis.

The system relies on the streptavidin-biotin interaction. The *Strep*-tag II is a short 8 amino acid peptide (WSHPQFEK), which binds with high selectivity to an engineered streptavidin, *Strep*-Tactin. Due to its small size, *Strep*-tag generally does not interfere with the folding and bioactivity of the fusion partner. Thus, removal of the tag becomes superfluous. Therefore, the removal of the tag was not performed in this study.

After application of the crude extract on a *Strep*-Tactin column and a short washing step, a particularly mild elution of the protein from the column is performed by specific competition with desthiobiotin, a biotin analog. Yellow azo dye HABA is added in excess to displace desthiobiotin for regenerating the column (Figure 2.4).

Throughout this thesis, all buffers for purification were prepared without EDTA, since recombinant GAO is a metalloprotein.

2.2.7.1.1 Purification by *Strep*-Tactin® Sepharose® Gravity Flow Columns

Strep-Tactin® (IBA GmbH's) Sepharose® (GE Healthcare) gravity flow columns were used which were imported by the financial support of TÜBITAK Scientific Research Group (TBAG).

All operations were performed at 4 °C for the stability of the recombinant GAO. To achieve optimal purification results, specified volumes and their ratios were applied (column bed, washing volumes etc.).

The column was equilibrated by adding 2 CV of Buffer W (Appendix A). Then the cell extract with a volume of 0.5-10 CV (2.5-50 ml) was added to the column (concentrated cell extracts are preferable). Flow through fraction was collected. If frozen cell extracts were used, before applying them to the column they were centrifuged in a benchtop microcentrifuge (Eppendorf) at 14000 rpm at 4°C for 5 minutes so as to get rid of any precipitates. After the cell extract had completely entered the column, the column was washed 5 times with 1 CV (5 x 5 ml) of Buffer W. The eluate was collected to load on an analytical SDS gel. Recombinant GAO was eluted by the addition of 6 times 0.5 CV (6 x 2.5 ml) Buffer E (Appendix A) containing desthiobiotin, a reversibly binding competitor. The purified protein was collected in 6 fractions, and it was generally eluted in the 3rd to 5th fractions.

The column was regenerated by the addition of 3 times 5 CV (3 x 25 ml) Buffer R (Appendix A) by which the colour of the matrix changed from yellow to red indicating the regeneration process was complete. The column was equilibrated by the addition of 2 times 4 CV Buffer W to make the column ready for the next purification run. The column was stored at 4°C overlaid with 2 ml of Buffer W or Buffer R ("IBA *Strep* -tag® Purification", 2002) . All the eluates were stored at -20 °C for further analysis.

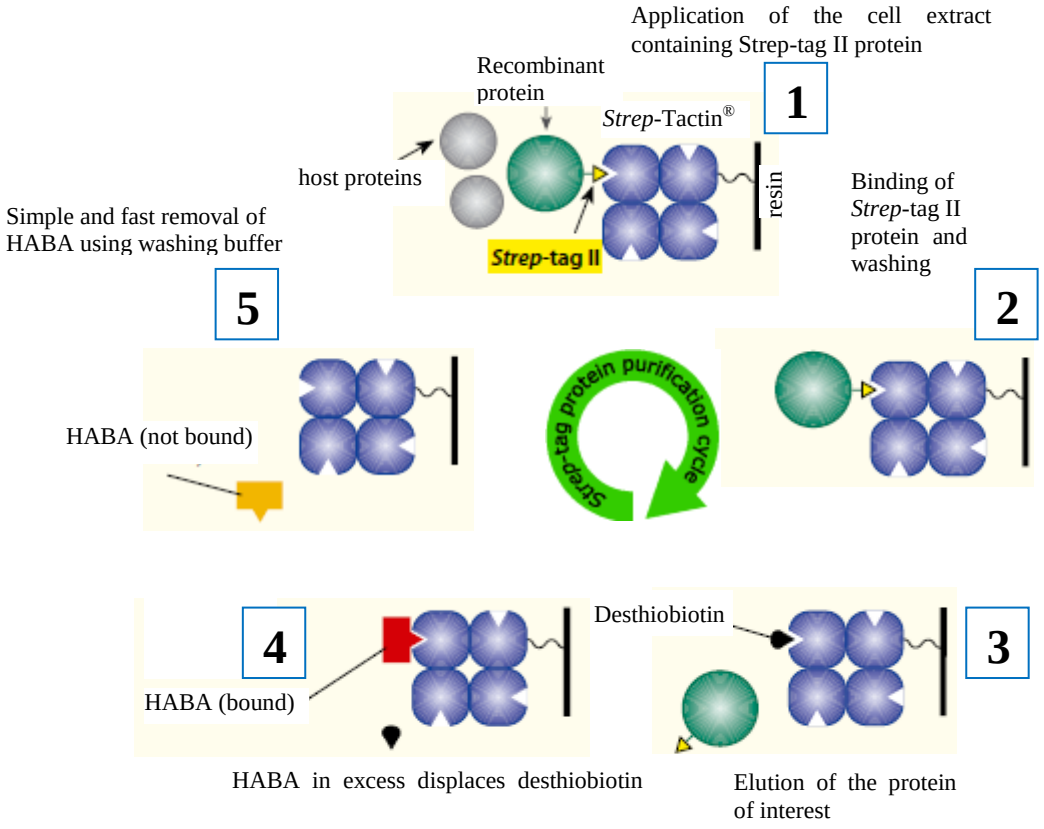


Figure 2.4 *Strep*-tag® II protein purification cycle ("IBA Protein TAGnologies").

2.2.7.1.2 FPLC Purification by Using *Strep*-Tactin® MacroPrep® Cartridges

1 ml *Strep*-Tactin® (IBA GmbH's) MacroPrep® (Bio-Rad Laboratories, GmbH, Ltd.) cartridge was used and connected to automated low pressure chromatography system; GradiFrac™ (Pharmacia Biotech, Sweden) for purification. Flow rate was adjusted as 1ml/min. The supernatants were passed through a 0.45µm syringe filter before loading to the column. All the buffers were filtered by membrane filters (Whatman® GE Healthcare Life Sciences) and degassed with Nalgene® filtering system before use. The amounts of the buffers required to enter the system for purification were adjusted according to the CV of the cartridge in compliance with the protocol in Section 2.2.8.1.1 and either the manual or programmed run was carried out. Putative purified protein was collected in 0,5 ml fractions.

2.2.8 Protein Analysis

2.2.8.1 Copper Loading

After pooling the fractions containing pure protein, they were brought together and dialysed against 3L 20 mM PIPES and 1mM $Cu(NO_3)_2$ (pH 6.1) for at least 6 hours (generally overnight) to ensure full processing of the enzyme. The protein was then dialysed against100 mM sodium phosphate buffer (pH 7.0), except those proteins used for crystallization studies.

2.2.8.2 Concentrating Protein Samples

Centrifugal concentrators or vacuum concentrator was used for protein concentration. Large volumes between 5-80 ml samples were concentrated by using Amicon® Ultra-4 and -15 Centrifugal Filter Units (Millipore®). For smaller volumes (< 1500 µl) Vacuum Concentrator 5301 (Eppendorf) was used at room temperature. These devices were used according to the manufacturer's instructions.

2.2.8.3 Determination of Protein Concentration by Bradford's Dye Binding Assay

In order to determine the amount of the protein content of samples Bradford method was followed (Bradford, 1976). For BSA standard solution the amounts given in the table below were prepared.

Table 2.5 The amounts required for BSA standard solution preparation.

Tube number	BSA standard(µl)	Distilled water (µl)	Bradford solution(ml)	Protein (mg/ml)
1	-	500	5	0
2	5	495	5	0.01
3	10	490	5	0.02
4	15	485	5	0.03
5	20	480	5	0.04
6	25	475	5	0.05

Distilled water was added to the tubes at proper amounts as indicated in Table 2.5. Then BSA protein standard was dispensed into each tube appropriately. The tubes were vortexed and mixed well following the addition of the Bradford's solution which must be at room temperature before addition. These tubes were kept in dark for 10 minutes and then the absorbance at OD_{595} was measured within 1 hour. The absorbance result of each BSA standard was plotted against its concentration. In order to determine the concentration of the protein of interest, different amounts of the sample were taken and measured until the result that ranges between the values of the standard was achieved. Triplicate assays were set up for all unknowns. The unknown protein concentration was calculated according to the below formula:

$$\text{Protein (mg /ml)} = \frac{\Delta OD_{595}}{\text{Slope}} \text{ x Dilution Factor}$$

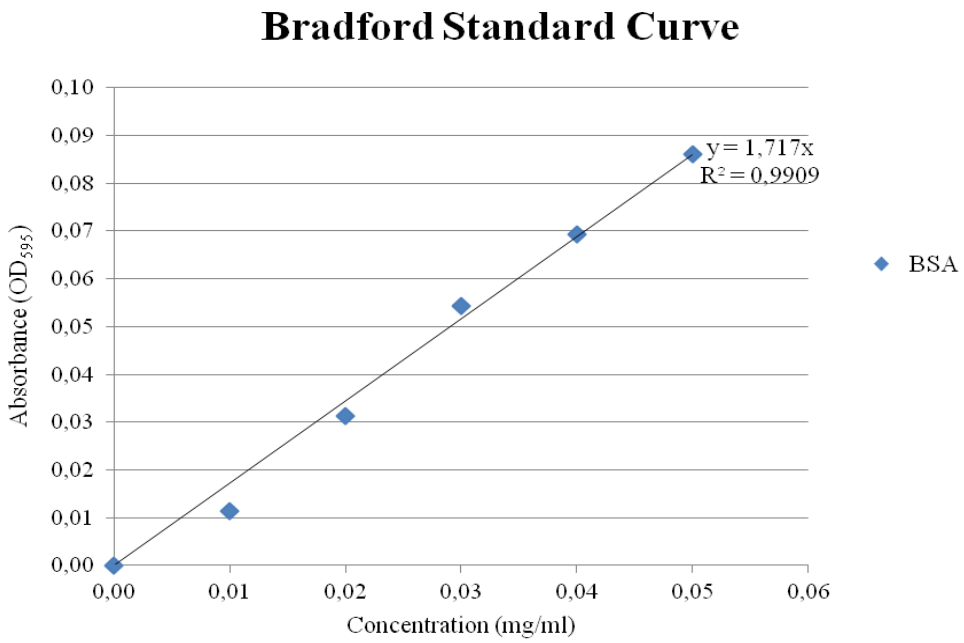


Figure 2.5 Standard curve for calculating concentration of GAO derivatives following with *Strep*-tag II column purification.The concentrations of the samples were determined in order to be used in enzyme activity assays.

Quick Start™ Bradford Protein Assay (Bio-Rad) was also used. It provides ready-to-use dye reagent at 1x concentration, two protein assay standards at seven prediluted concentrations and 2 mg/ml BSA. The dilutions were obtained by the 2 mg/ml BSA supplied according to the amounts below:

Table 2.6 Prepartion of 1 ml standard assay according to the Quick Start™ Bradford Protein Assay (Bio-Rad, USA).

Tube number	Standard Volume (µl)	Source of Standard	Diluent Volume (µl)	Final Protein (µg/ ml)
1	70	2 mg/ml stock	0	2.000
2	75	2 mg/ml stock	25	1.500
3	70	2 mg/ml stock	70	1.000
4	35	Tube 2	35	750
5	70	Tube 3	70	500
6	70	Tube 5	70	250
7	70	Tube 6	70	125
8 (blank)	-	-	70	0

The measurements were carried out according to the standard protocol of the manufacturer's instruction manual ("Bio-Rad Quick Start™ ").

2.2.8.4 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was performed via either Blue Vertical 102 Electrophoresis System together with Blue Power 500 (SERVA Electrophoresis GmbH) or Mini-PROTEAN 3 System (Bio-Rad, USA).

Throughout the study, manufacturer's instructions were followed for casting of either of the electrophoresis system. When Mini-PROTEAN 3 System was in use Bio-Rad multi-casting chamber was used for multiple gel pouring.

12,5% or 15% separating gel and 7,5% stacking gel was prepared by combining the components in Table 2.7 and Table 2.8, respectively. (Buffer compositions are given in Appendix A). TEMED and freshly made APS (Appendix A) were added just before the gel was ready to pour in order to avoid polymerization prior to pouring. Before pouring the gel mixture, ddH_2O was poured between gel plates up to the top to test any leakage. Following the pouring of the gels in an order, combs were inserted carefully to avoid bubble formation and allowed to set (for separating gel 1 hour and for stacking gel nearly 30 minutes).

Table 2.7 Component composition of separating gel preparation.

Separating gel	Volume (ml)							
	12,5%				15%			
	(x1)	(x2)	(x7)	(x20)	(x1)	(x2)	(x7)	(x20)
GB	1,2	2,4	8,4	24	1,2	2,4	8,4	24
30% Acrylamide Stock	2,1	4,2	14,7	42	2,6	5,2	18,2	52
H_2O	1,7	3,4	11,9	34	1,2	2,4	8,4	24
APS (25%)	0,03	0,06	0,21	0,6	0,03	0,06	0,21	0,6
TEMED	0,003	0,006	0,021	0,06	0,003	0,006	0,021	0,06
(Total)	(5)	(10)		(100)	(5)	(10)		(100)

Table 2.8 Component composition of stacking gel preparation.

Stacking gel	Volume (ml)			
	7,5%			
	(x1)	(x2)	(7x)	(x20)
SGB	0,6	1,2	4,2	12
30% Acrylamide Stock	0,4	0,8	2,8	8
H_2O	1,4	2,8	9,8	28
APS (25%)	0,02	0,04	0,14	0,4
TEMED	0,002	0,004	0,014	0,04
(Total)	(2,4)	(4,8)	(16,8)	(48)

As soon as the gels were polymerised and after the placement of running unit into the tank and removal of the combs, the appropriate volume of 1 x running buffer (from 10x stock) (Appendix A) was added to the upper and lower chambers. Meanwhile, samples were prepared including molecular weight markers. In order to denaturate the proteins, sample denaturing mix was made by adding 1:4 β-mercaptoethanol to the sample buffer (Appendix A). 1:3 volume sample denaturing mix (4x) was added to each sample to a final concentration of 1x. Samples were boiled for 5 minutes and allowed to cool. Then they were centrifuged at 12500 rpm at room temperature for 2-3 minutes in a benchtop microcentrifuge (Eppendorf). If there were wells left unloaded, in order to maintain the uniform electrical resistance across the gel they were filled with 1:1 dH_2O:sample denaturing mix in order to prevent "smiling" of outer samples. Constant voltage of 150 V was applied with 12,5% separating gels and 200 V was applied for 15% separating gels. The gel was run for about 60-120 minutes.

Size separation	%age gel
70-200 kDa	5.0%
40-150 kDa	7.5%
20-100 kDa	10.0%
10-70 kDa	2.5%
8-50 kDa	15.0%

Optimal protein loading

Individual polypeptide	0.5-5 μg
Complex mixture	25-50 μg

2.2.8.4.1 Coomassie Blue Staining

After the SDS-PAGE was completed, if Coomassie staining was decided, either Coomassie Blue R250 or InstantBlue™ (“Expedeon”, UK) was used. When the Coomassie Blue R250 was used the gel was fixed with 20% trichloroacetic acid (Appendix A) for ½ hour at room temperature preceding staining. Then it was rinsed with dH_2O 2-3 times for 3 minutes. The gel was transferred into 50-100 ml staining solution (Appendix A) and stained for overnight. After replacement of staining solution with 100 ml destaining solution (Appendix A) the gel

was destained with 3 changes of destaining solution, each for 15 minutes. After the photograph of the gel was taken, the gel was kept in preserving solution (Appendix A).

During the use of InstantBlue™ which is a ready-to-use, safe, sensitive and non-toxic proprietary Coomassie® stain, the protein gels were transferred directly into the InstantBlue staining solution without the need of washing or fixing. Typically ~20 ml was needed to cover the gel to facilitate diffusion. After 15 minutes incubation at room temperature with gentle shaking, coloured protein bands started to develop immediately. The gels were photographed when the required intensity had been achieved. Gels can be kept in staining solution, alternatively can be stored in ddH_2O after staining for 1 hour in InstantBlue.

2.2.8.4.2 Silver Staining

Silver staining is a much more sensitive method of staining than Coomassie staining, which is rather preferred when the protein of interest is hardly detectable due to the low amount or purity of the protein in question.

During the procedure, the gel was placed in plastic boxes with lids on and all the steps were performed on a constantly shaking platform shaker at room temperature. Silver nitrate solution was freshly prepared. When handling gels, gloves were worn at all stages since silver staining would detect skin and sweat proteins on fingers.

After electrophoresis, the gel was immediately fixed in fixer solution overnight (should be >1h). It was soaked into pre-treatment solution for exactly 1 minute after washing with 50% ethanol for 3 x 20 minutes. After pre-treatment the gel was rinsed with dH_2O for exactly 3 x 20 seconds. Following the draining of water, the gel was immersed into silver nitrate solution for 20 minutes. Then the gel was rinsed with dH_2O for exactly 2 x 20 seconds. The gel was transferred into developing solution and kept in this solution nearly 2 minutes (which can be adjusted according to the colour development by adding dH_2O into the developing solution to decrease the reaction rate and can take longer). As the desired intensity of the bands was attained, the gel was washed with dH_2O for 2 x 2 minutes to remove the developing solution and developing was terminated by placing the gel into the stop solution for >10 minutes.

2.2.8.5 Wet Electroblotting

Transfers of the proteins were performed by using XCell II™ Blot Module (Invitrogen) which is a semi-wet transfer unit. Before the transfer blotting membranes, filter papers and blotting pads were prepared, while the SDS-PAGE was running or just continuing at a low voltage of 5 V before finishing. Blotting membrane and filter paper sandwiches were cut according to the size of the gel (8 cm (W) x 7.3 cm (H)). PVDF membrane was activated in 100% methanol for 30 seconds and briefly rinsed with dH_2O. Then it was soaked into transfer buffer (Appendix A) for 3 minutes in a shallow dish. The filter papers were soaked into the transfer buffer with the blotting pads as well; pads were squeezed for releasing the bubbles that might prevent the transfer of the proteins. Pre-soaked 2 filter papers were placed on top of the 2 blotting pads. The gel was placed on top of the filter papers. The surface of

the gel was wet with the transfer buffer and pre-soaked transfer membrane was laid over the gel. Any corner of membrane can be cut to record orientation; or prestained markers might also help orientation after transfer. Then 2 filter papers (any trapped air bubbles were removed by rolling a glass pipette) and the 2 pads were placed on top of this assembly to make a "sandwich". This sandwich was placed onto the cathode such that the gel was closest to the cathode plate, then anode was placed against it. This module was slided into the lower chamber slowly and placed into the XCell SureLock™ Mini-Cell (Invitrogen) as described in user manual ("Invitrogen", 2009).The blot module was filled with transfer buffer until the sandwich was covered in transfer buffer and the outer buffer chamber was filled with ~ 650 mL dH_2O water to dissipate heat produced during the run. The transfer was operated at room temperature at a constant voltage of 15V for 75 minutes. At the end of the transfer the gel was stained by InstantBlue™ to detect how successful the transfer was.

2.2.8.6 Western Blotting

After electroblotting the membrane was blocked by 3% BSA in 1 x TBS-T (Appendix A) at 4°C overnight. The membrane was incubated with 1° antibody; *Strep*-Tag II Monoclonal Antibody (200 µg/ml) (Novagen, Germany) in 1 x TBS-T + 3% BSA with a dilution of 1:5000 at room temperature for 1 hour for samples carrying *Strep*-Tag II . After draining the solution, the membrane was treated with 2° antibody; goat anti-mouse IgG (H+L)-HRP (0,5-1 mg/ml) (Alpha Diagnostic International, USA) in 1 x TBS-T + 3% BSA with a dilution of 1:2500 at room temperature for 1 hour. The membrane was washed in TBS-T (Appendix A) for 3 x 10 minutes and then briefly rinsed with dH_2O for just 2 minutes. The excess liquid was drained off. 1-2 ml TMB Stabilized HRP substrate (Promega) was applied to the membrane to cover it and left for colour development for 4 minutes (The bands were started to appear within 1-2 minutes). The reaction was stopped by dH_2O. Detection was done by camera directly and by scanner after putting into a freezer bag. The membrane was kept at 4°C after air-drying and foiling for a while without fading.

For the detections of samples with His-tag, the membrane was incubated directly with Monoclonal Anti-polyHistidine Peroxidase Conjugate (5–11 mg/ml) (Sigma-Aldrich) in 1 x TBS-T + 3% BSA with a dilution of 1:10000 at room temperature for 1 hour.

2.2.8.7 Mass Spectrometry

The samples were dialysed against 50mM ammonium acetate buffer (pH 6.8) at 4˚C overnight. After concentrating the samples by Amicon® Ultra-4 Centrifugal concentrator and determining the amounts of proteins by Quick Start™ Bradford Protein Assay as described in Section 2.2.9.2.1 (50µl of 20-50 µM sample is required), they were submitted to Mass Spectrometry Facility, Astbury Centre for Structural Molecular Biology, Institute of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds.

2.2.9 Enzymatic Analysis by ABTS-HRP Coupled Assay

The standard coupled assay system was used for the calculation of protein samples (Amaral *et al.*, 1963). The assay is based on the oxidation of chromogenic reagent ABTS via catalysis of galactose oxidase-derived H_2O_2 by horseradish peroxidase (HRP).

GAO assay solution including the substrate ABTS was prepared (Appendix A). Since solutions containing ABTS are light sensitive GAO assay solution was stored in a dark foiled bootle at 4˚C.

2.2.9.1 Qualitative GAO Activity Assay

To observe the GAO activity in the crude extracts, 50 µl supernatant assumed to be containing the enzyme of interest and 100 mM $CuSO_4$ were added into 200 µl assay solution to a working concentration of 50 µM (excess copper was used instead). Colour development was monitored. Conversion of colourless transparent solution into green indicates the activity of the enzyme.

The pellets also could be resuspended in 1 x PBS containing 50 µM $CuSO_4$ for further protein analysis. This assay was performed with pure samples as well after being dialysed against copper as indicated in Section 2.2.8.1 (450 µl assay solution + 5 µl pure enzyme).

2.2.9.2 Quantitative GAO Activity Assay and Specific Activity of GAO

The change of absorbance showing the rate of reaction was measured using UV-1700 PharmaSpec UV-visible Spectrophotometer (Shimadzu) which was connected to a computer and operated with a UVProbe Ver.2.21 software.

All assays were carried out at 25˚C using disposable plastic cuvettes. The enzyme was added as a drop to the edge of the cuvette and reaction was initiated by inverting the cuvette 1-2 times immediately. Absorbance measurements were carried out at 414 nm. All readings were done in triplicates to obtain an average value.

The molar extinction coeffiecient of oxidized ABTS at 414 nm was known as 31300 M^{-1} cm^{-1}. Also as it is known, one international unit of activity corresponds to the amount of enzyme catalyzing oxidation of 1 µmol of substrate for 1 minute.

In order to calculate the activity of the enzyme; i.e. the number of moles of H_2O_2 $min^{-1}mg^{-1}$ of enzyme, Beer-Lambert Equation was used;

$$\text{no. of } \mu\text{moles } H_2O_2/\text{min/ml enzyme} = \frac{\Delta A_{414\ nm}.\text{min}^{-1}.V_T.\ D\ .\ 10^6.\ 0{,}5}{\varepsilon_{414nm}.\ L\ .V_S.}$$

V_T :Total volume (ml)
V_S :Sample volume (ml)
$\varepsilon_{414\ nm}$:Extinction coefficient of oxidized ABTS ($M^{-1}cm^{-1}$)
D :Dilution factor
L :Length (1 cm)

no. of μmoles H_2O_2/min/mg enzyme = $\frac{\text{u moles } H_2O_2\text{/min/ml enzyme}}{\text{mg /ml}}$

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Overall Strategy and Structure of the Study

Galactose oxidase (GAO) of *Fusarium graminearum*, which requires four post-translational modification steps in order to gain its mature active form, is a unique enzyme (Section 1.3.4). Its leader peptide consists of 41 amino acids. According to studies of Ögel (1993) based on the weight matrix approach of von Heijne (1986a) and the SignalP 4.1 (Petersen *et al.*, 2011), Ala (16) has the highest score and is the possible signal peptide cleavage site. Its 17- aa long pro-peptide, which was revealed by crystallographic studies (Firbank *et al.*, 2001) and N-terminal sequencing is cleaved autocatalytically when treated with copper and molecular oxygen (Rogers *et al.*, 2000).The remaining 8 residues between the signal peptide and the pro-peptide appear to be a pre-peptide which is possibly cleaved by a fungal monobasic or dibasic protease (Özöğür -Akyüz *et al.*,2009). Based our former study, where the recombinant precursor GAO expressed in *E. coli*, was left unprocessed, even in the presence of copper and molecular oxygen, (Gençer (Kocuklu), 2005), it was hypothesized that this putative pre-peptide, is not cleaved in *E. coli*, possibly due to the lack of this fungal protease and that this was responsible from the lack of autocatalytical cleavage. In order to test this hypothesis, it was decided to delete the N-terminal 8-aa long putative pre-peptide region of the, so called, pPre-Pro-GAO by site-directed mutagenesis.

In the second part of this thesis, mutagenesis studies on GAO, related to the preparation of split enzyme systems for biosensor applications, were started as initial steps of a comprehensive and a big project. Accordingly, domain I/II and domain III deletions of GAO were performed, again using *E. coli* as the heterologous expression system, this time replacing the *Strep*-tag II with His-tag.

All the mutant constructs of GAO were subjected to further gene and protein analysis, to detect and analyse the mutants, using suitable molecular techniques. The experimental strategy of this thesis is shown in Figure 3.1.

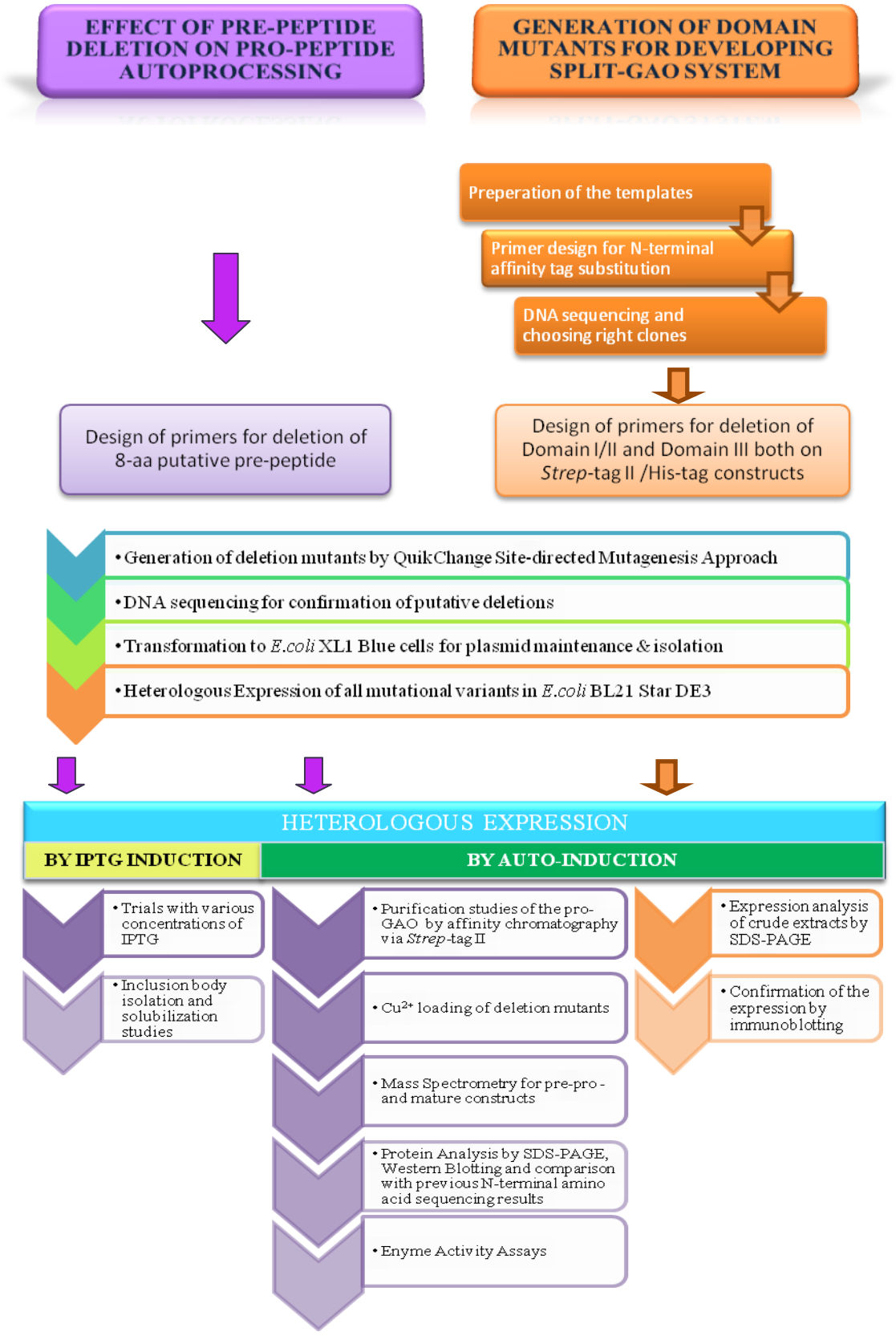


Figure 3.1 Overall experimental strategy of this thesis.

PART I

EFFECT OF PRE-PEPTIDE DELETION ON PRO-PEPTIDE SELF-PROCESSING OF HETEROLOGOUSLY EXPRESSED GALACTOSE OXIDASE IN *E.coli*

3.2 Preparation of Plasmids Carrying the *gao* Gene

Overall, three constructs, which were previously developed based on directed evolution in the University of Leeeds, UK (Deacon, 2008, Deacon *et al*, 2011) were used in this study. These were pPre-ProGON1 and pPre-ProGOMN1(renamed again) and pMatGOMN6 (Section 2.1.4). The N- terminal region of GAO and the constructs is shown in Figure 3.2. As shown, the putative 8-aa long pre-peptide is followed by the 17-aa pro-peptide, after removal of the signal peptide.

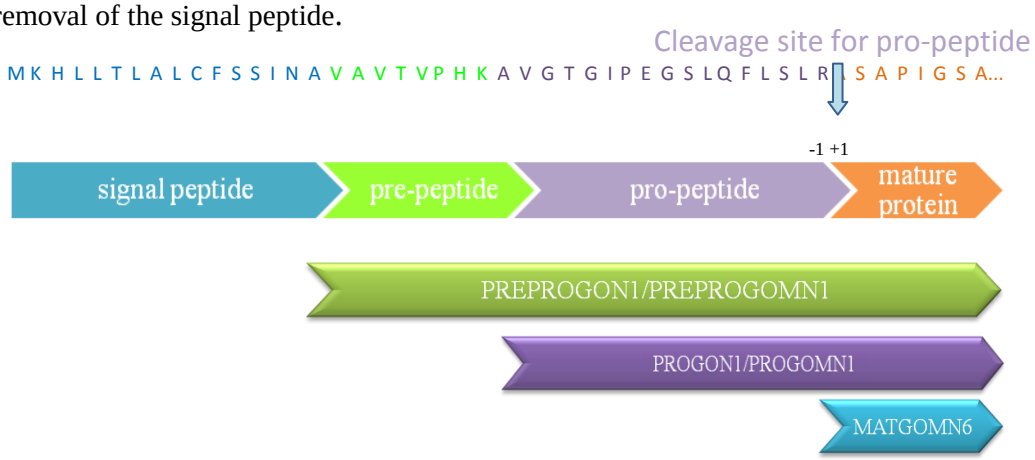


Figure 3.2 Schematic diagram showing the nature of the N-terminus of GAO and GAO-derived constructs.

The constructs were transformed into *E.coli*. and isolated as described in Section 2.2.5.3 . Further, restriction enzyme digestion was carried out (Section 2.2.5.6).Following the plasmid restriction digestion of pMatGOMN6, the expected band sizes were; 7031 bp, 699 bp by *Nco*I and 4550 bp,1776 bp, 860 bp, 544 bp by *Eco*RI and *Pst*I.

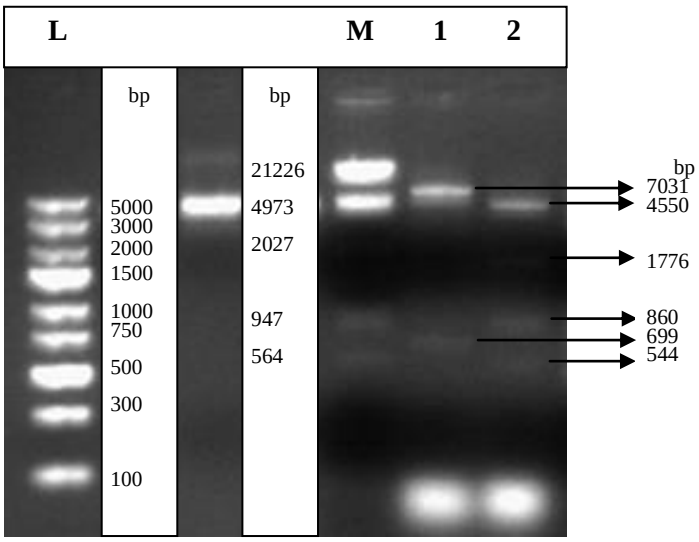


Figure 3.3 Single and double restriction enzyme digestion results of pMatGOMN6. **L,** GeneRuler™ Express DNA Ladder; **M,** λ DNA/*Eco*RI+*Hin*dIII; **1,** *Nco*I digestion; **2,** *Eco*RI & *Pst*I digestion.

The expected band sizes of pPre-ProGON1 by single and double restriction enzyme digestion were; 7031 bp, 774 bp by *Nco*I and 4550 bp,1776 bp, 860 bp, 559 bp, 60 bp by *Eco*RI and *Pst*I.

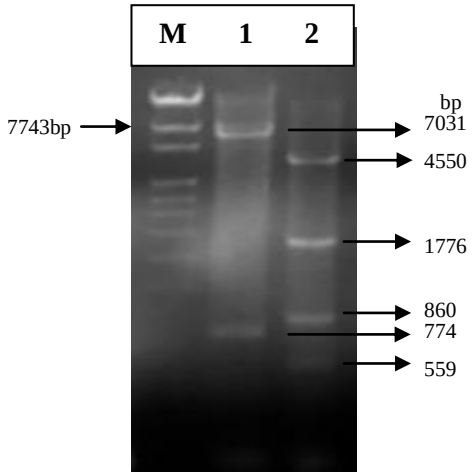


Figure 3.4 Restriction enzyme results of pPre-ProGON1. **M**, λ-pUC Mix; **1,** *Nco*I digestion; **2,** *Pst*I & *Eco*RI digestion.

pPre-ProGOMN1 was already used and checked in our previous studies (results not given),therefore it is not subjected to restriction digestion again here. The total sizes and maps of the plasmids are given in Appendix C.

3.3 Application of *Site-Directed Mutagenesis* Approach for the Deletion of the Pre-peptide Region of the GAO Constructs

As indicated in Section 3.1, in our previous study, it was discovered that GAO, when expressed in its pre-pro-form in *E. coli*, is left unprocessed, despite of the presence of copper and molecular oxygen, which were previously reported to be sufficient for pro-peptide self-processing in GAO, *in vitro* (Rogers *et al.*, 2000) . However, interestingly, in *E.coli*, the 25-aa leader peptide was still being kept at the N- terminus of the purified GAO preparations (Gençer (Kocuklu), 2005) . Inspired by a previous study by (Ögel,1993), it was hypothesized that the pre-pro-peptide of GAO is likely to undergo a two-step cleavage operation, namely, first an 8 aa pre-peptide is cleaved by a fungal monobasic/dibasic processing protease after the His-Lys monobasic /dibasic site, between the K(-18)-A(-17); second, the remaining 17 aa pro-peptide is cleaved autocatalytically. According to this hypothesis, the lack of the specific protease in *E.coli* prevents pre-peptide cleavage, hence changes the conformation of the pro-peptide and prevents autocatalytic cleavage, which likely requires a certain pro-peptide conformation.

3.3.1 Use of Previously Designed Primers

For pre-peptide deletion on pPre-ProGOMN1 and pPre-ProGON1 constructs, previously designed (Öreroğlu, unpublished data) 32 bp complementary primer pairs (Table 3.1) were used. The expected size of the mutant fragment was 7781bp. However, these primers failed to amplify a fragment of the expected size.

Only non-specific amplification products were observed. Several parameters, as summarized below, were checked, however, all these attempts failed in amplification. Thus it was decided to design new primers.

I. Increased amount of primers
II. Use of other working primers
III. High amount of template DNA
IV. New high-quality enzyme, *Pfu Ultra* High Fidelity DNA Polymerase (Stratagene)
V. Use of different thermocyclers
VI. Use of amplification facilitators such as glycerol and DMSO
VII. Increasing the denaturation temperature
VIII. Increasing Mg^{2+} ion concentration
IX. Decreasing the annealing temperature
X. Modification of cycling parameters
XI. Applying 2-Step Site-Directed Mutagenesis Method (each primer into different tubes)
XII. Transformation of the putative amplified DNA into the *E.coli* even it was not detected on the agarose gel

3.3.2 Modification of the Deletion Primer

New primer pairs were designed by simply deleting 2 bases from the 3' end of the forward, and 5' end of the backward primer (Table 3.1). The T_m of the 30 bp primers was calculated as 85°C by Stratagene QuikChange Primer T_m Calculator (www.stratagene.com/QPCR/tmCalc.aspx).

Table 3.1 Comparison of the former and the modified primer pairs.

Pre-designed Primer Pairs	**Length**	**Tm**	**GC%**
FORWARD : 5' GGAGCCCTTCACCATGGCCGTAGGAACTGG**AA** 3 ' **REVERSE** : 5' **TT**CCAGTTCCTACGGCCATGGTGAAGGGCTCC 3'	32 bp	84,8	59,37
New Primer Pairs			
DEL_PROGAO F : 5'GGAGCCCTTCACCATGGCCGTAGGAACTGG 3' **DEL_PROGAO R :** 5'CCAGTTCCTACGGCCATGGTGAAGGGCTCC 3'	30bp	85	63,33

Based on primer design rules, primers should terminate in one or more C or G bases. Thus, the two adenines were discarded from the sequence.

3.3.3 Generation of pProGON1 and pProGOMN1 From pPre-ProGON1 and pPre-ProGOMN1

3.3.3.1 Application of *QuikChange® Site-Directed Mutagenesis*

QuikChange® Site-Directed Mutagenesis Kit II was used for deletion mutagenesis (as described in Section 2.2.3.1). All reaction and thermal cycling parameters were adjusted as given in Section 2.2.3.1.2 . This time, a band with an expected size of 7781 bp was detected on the agarose gel, with the modified primers.

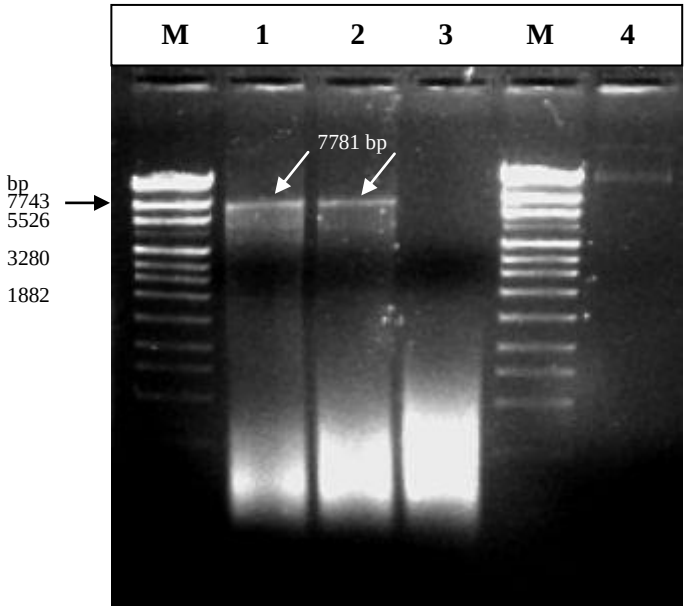


Figure 3.5 Detection of putative ProGON1 and ProGOMN1 fragments by DEL_PROGAO primer pairs and *QuikChange® Site-Directed Mutagenesis Kit II* on agarose gel. **M,** λ-pUC Mix; **1,** pProGON1; **2,** pProGOMN1; **3,** (-) control; **4,** pPre-ProGOMN1

It has been demonstrated that when the proper and precise amount of primers were used, the 7781 bp band of interest could be observed even at the first trial, using standart amplification conditions. 1µl *PfuUltra* High-Fidelity DNA Polymerase was used with primer pairs of precise concentration (125 ng/ µl), without using extra Mg^{2+}, since it was sufficiently supplied in the 10 x reaction buffer. The reaction conditions were 30 seconds at 94 °C for enzyme activation, 30 seconds at 94 °C for denaturation, 1 minute at 55°C for annealing and 8 minute at 68°C for extension, cycling for 25 times, which were slightly modified from *QuikChange® II Site-Directed Mutagenesis* kit as given in Section 2.2.3.1.2.This result clearly demonstrates the importance of the sequence of primers, and especially, the design of primers by taking the general design rules into consideration, in the success of modificational DNA amplification.

*Dpn*I digestion of the amplified fragment was performed, as explained in Section 2.2.3.1.3 for eliminating the parental DNA, in order to attain only the mutated newly synthesized DNA.

Next, the putative pProGON1 and pProGOMN1 constructs were transformed into XL1-Blue competent cells after *Dpn*I digestion (Section 2.2.4).

3.3.3.2 Identification of pProGON1 and pProGOMN1

Following the transformation of plasmids into XL1 Blue cells, 6 colonies from each of the constructs were inoculated into proper antibiotic containing medium and plasmid isolation was carried out, as described in Section 2.2.5.3. These plasmids were tagged and glycerol stocks were prepared for further analysis.

In order to detect pre-peptide deletion, first, *Eco*RI single enzyme digestion and *Eco*RI and *Nco*I double digestion were also applied. The restriction digestion products cover the sequence that would be deleted; therefore the band sizes at the end of the digestion of the unprocessed pre-pro-enzyme and putative pro-enzyme were anticipated to differ 24 bp. When single and double digestions were applied as explained in Section 2.2.5.6, the expected band sizes were;

***Eco*RI**		***Eco*RI and *Nco*I**	
6326 bp		6326 bp	
1419 bp		732 bp	
60 bp	→ 36 bp (if deleted)	687 bp	
		42 bp	→ 18bp (if deleted)
		18 bp	

It was also thought that the detection of the deletion may be been much easier with double digestion since 42 bp band would disappear and instead 18 bp band, obtained by the deletion of 24 bp, would appear. with a higher intensity due to the already present 18 bp. 5 % agarose) was used in order to visualize the small sized fragments together with the help of NoLimits™ 15bp DNA Fragment (Fermentas) and low range DNA ladder (Figure 3.6).

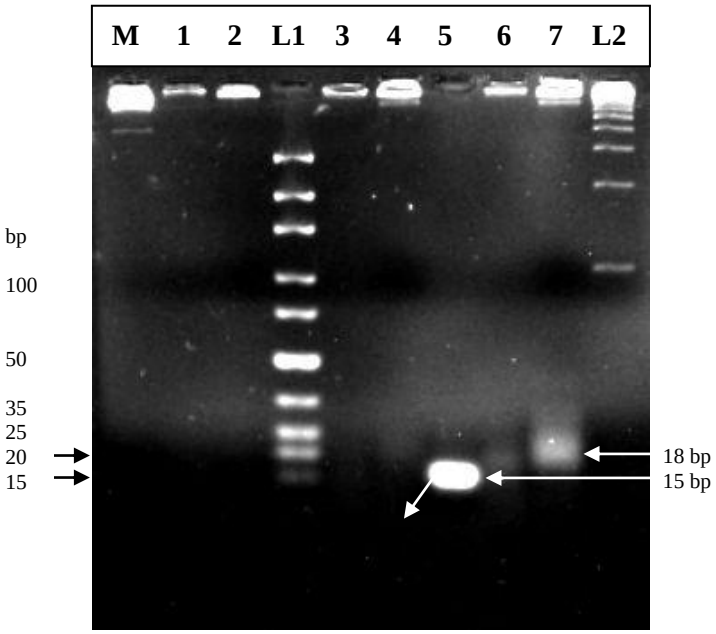


Figure 3.6 Double restriction enzyme digestion to detect 8-aa pre-peptide deletion by visualizing intense 18 bp fragment on putative pProGOMN1 and pProGON1 constructs. **M**, λ-pUC Mix; **L1,** GeneRuler™ DNA Ladder, Ultra Low Range; **1**, pProGON1_1; **2,** *Eco*RI single digestion of pProGON1_1; **3,** pProGON1_4; **4,** *Eco*RI *&* *Nco*I digestion of pProGON1_4; **5,** NoLimits™ 15bp DNA Fragment; **6,** pProGOMN1_4; **7,** *Eco*RI *&* *Nco*I digestion of pProGOMN1_4; **L2,** O'Range Ruler Ladder (100 bp+ 500 bp)

According to the Figure 3.6, the 18 bp band in question seemed to be detected only on pProGOMN1_4, among the 12 plasmid samples. The possibility of the presence of RNA was eliminated by RNase treatment. There is a faint band around 18 bp in lane 4, indicating the possible deletion of 8-aa pre-peptide in pProGON1_4 whereas single enzyme digestion of pProGON1_1 did not reveal any product.

To confirm the deletion of the 8-aa pre-peptide region, pProGOMN1 and pProGON1 samples were further analysed by DNA sequencing.

3.3.3.3 DNA Sequencing for Confirmation of Pre-Peptide Deletion

Following plasmid isolation by GeneJET™ Plasmid Miniprep Kit (Section 2.2.5.2) 10 samples were submitted to DNA sequencing, as it was explained in Section 2.2.5.7. Plasmid isolation results for four of pProGOMN1 samples were given in Figure 3.7.

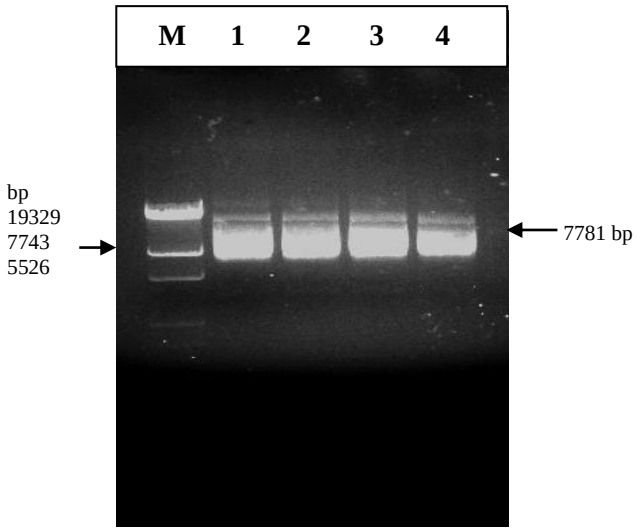


Figure 3.7 Plasmid isolation results of putative pProGOMN1 samples submitted for sequencing. **M,** λ-pUC Mix; **1- 4,** pProGOMN1_2,3,6 and 7.

Among the 10 samples sequenced (3 pProGON1, 7 pProGOMN1, each numbered from 1-7) 1 of the pProGON1 and 4 of the pProGOMN1 constructs have been detected to carry the desired 8-aa deletion. pProGON1_4 and pProGOMN1_2 were chosen particularly, the ones which gave rise to a longer and precise reading frame.

The DNA sequencing results revealed that while pProGON1_4 had the expected deletion, pProGON1_1 did not have a pre-peptide deletion as anticipated from the above restriction digestion results in Figure 3.6. On the other hand, the pProGOMN1_4 did carry the desired deletion , however the vector sequence did not seem to be accurate, containing additional nucleotides.

Figure 3.8 DNA and amino acid sequence alignments of ProGON1_4 and ProGOMN1_2 indicating pre-peptide deletion in comparison with pre-peptide carrying constructs Pre-ProGON1 and Pre-ProGOMN1, respectively. WSClustalW2 (http://www.ebi.ac.uk) was used for sequence analysis.

These samples will be referred as ProGON1 and ProGOMN1 hereinafter unless anything else was indicated.

All the nucleotide and amino acid sequence alignments of ProGON1 and ProGOMN1 are provided in the Appendix D.

3.4 Heterologous Expression of Pro-GAO Constructs

3.4.1 Selection of Expression Host, Induction Temperature and Duration

Based on the findings as indicated in Section 1.7.1, BL21 Star (DE3) was selected for the expression of Pro-GAO constructs in this study. Accordingly, 25°C and 21-22 hours were chosen as optimum induction temperature and harvest time, respectively (Deacon *et al.*(2011).

Inspite of the higher levels of expression in other strains at 30°C and 37°C, it was not chosen since at higher temperatures the rate of translation may be fast and this may lead to improper folding and aggregation of proteins. It was noted that there is a ~130-fold improvement in expression when the temperature is reduced from 37°C to 25°C (Deacon; 2008). Lower temperatures would decrease the rate of translation allowing recruitment of the rare tRNAs and this also helps slowing down the folding process allowing correct folding of the protein. Additionally, BL21 Star (DE3) carries a mutant *rne131*gene encoding a truncated RNase E involved in mRNA degradation. This also allows the accumulation of mRNA and makes high level of templates ready for later stages of induction.

3.4.2 Transformation of GAO Constructs into BL21 Star (DE3)

Following the verification of the pre-peptide deletion by DNA sequencing, the selected constructs, namely pProGOMN1 and pProGON1 were transformed into BL21 Star (DE3) (Invitrogen) cells as explained in Section 2.2.6.1. BL21 Star (DE3) which enhances the capability of recombinant protein expression, by its protease and RNase deficient nature, allows high-level expression of T7-regulated genes with IPTG induction.

3.4.3 Expression of pProGON1 and pProGOMN1 Constructs by IPTG Induction

After transformation, 6 single colonies were picked from each of the transformation plates as representatives of ProGON1 and ProGOMN1 and were tagged from A to F. These cells were inoculated into the proper growth media containing the appropriate antibiotic. At the end of the incubation, glycerol stocks were prepared (Section 2.2.1).

Cells taken from representatives of pProGON1 and pProGOMN1 plates together with the pPre-ProGOMN1 were then grown according to the experimental procedure indicated in Section 2.2.6.3.1 for the expression of mutant GAOs. When the cells were harvested, cell extracts were obtained by ultrasonication (Section 2.2.6.4.1) in the presence of protease inhibitor (Complete Mini, EDTA-free, ROCHE) to prevent any potential protease activity. An experiment without protease inhibitor was also set during expression trials.

Qualitative GAO assay was carried out by crude extracts of ProGON1_A and ProGOMN1_A, as described in Section 2.2.9.1. However, colour change was not detected, except for the Pre-ProGOMN1 control.

According to these results, the Pro-GAO constructs showed no activity, suggesting either lack of expression of the recombinant *pro-gao* genes, or, more likely, a problem related to post-translational processing or protein structure, such as, inability of proper folding.

A list of possible causes of the lack of GAO activity is given below:

- Different IPTG concentrations ranging between 0,05 mM - 0,5 mM were used for expression induction during cell growth, however, activity could not be detected with the crude extracts of ProGOMN1_A samples by qualitative activity assay.

- There may be been a frameshift within the reading frame of the constructs. However, DNA sequencing was performed twice, once at REF-GEN Gene Research and Biotechnology, METU Technopolis, Ankara and once again, at Beckman Coulter Genomics, Essex, UK as indicated in Section 2.2.5.7 . Possibility of an error in the DNA sequence was eliminated by the results (Appendix D).

- The plasmids obtained from BL21 Star (DE3) cells were also sequenced again to detect whether there occurred a problem with the sequence following the transformation into the expression host (Section 2.2.5.2 and 2.2.5.7). The sequencing results eliminated this possibility.
- Inclusion body (IB) formation is a typical problem of heterologous expression in *E.coli*. Aggregation of the protein may be a significant problem due to the high level of expression of the recombinant protein. However, all attempts to identify IB in the cells, by microscopical techniques, have failed to identify any IBs. IB analysis was continued by isolation of IB, as described in Section 3.4.3.3 below.It was decided to do new expression trials with ProGON1_B and ProGOMN1_B,C,D and E, which were not tested before.

At the end of the cell growth after 21-22 hours, following induction with 0,1mM IPTG as indicated in Section 2.2.6.3.1 (protease inhibitor was not used) qualitative enzyme activity assay was conducted with the cell extracts. As seen in Figure 3.9, a pale and slowly developing green colour was observed at the end of 15 minutes, only with the ProGOMN1 samples, although was relatively low as compared to the control.

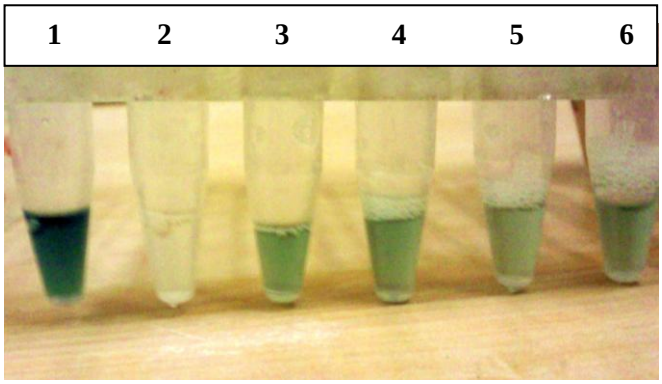


Figure 3.9 ABTS-HRP coupled assay results with crude extracts of Pro-GAO samples. **1,** Pre-ProGOMN1 (+) control; **2,** ProGON1_B; **3-6** ProGOMN1_B,C,D,E, respectively.

Pre-ProGOMN1 shows higher activity when compared to Pre-ProGON1, due to the point mutations acquainted by directed evolution (Deacon, 2008; Sun *et al.*,2001). Thus lower activity in the ProGON1 sample is not surprising. Trials with different ProGOMN1 constructs, MatGOMN6 and the negative control, at different time intervals are shown in Figure 3.10.

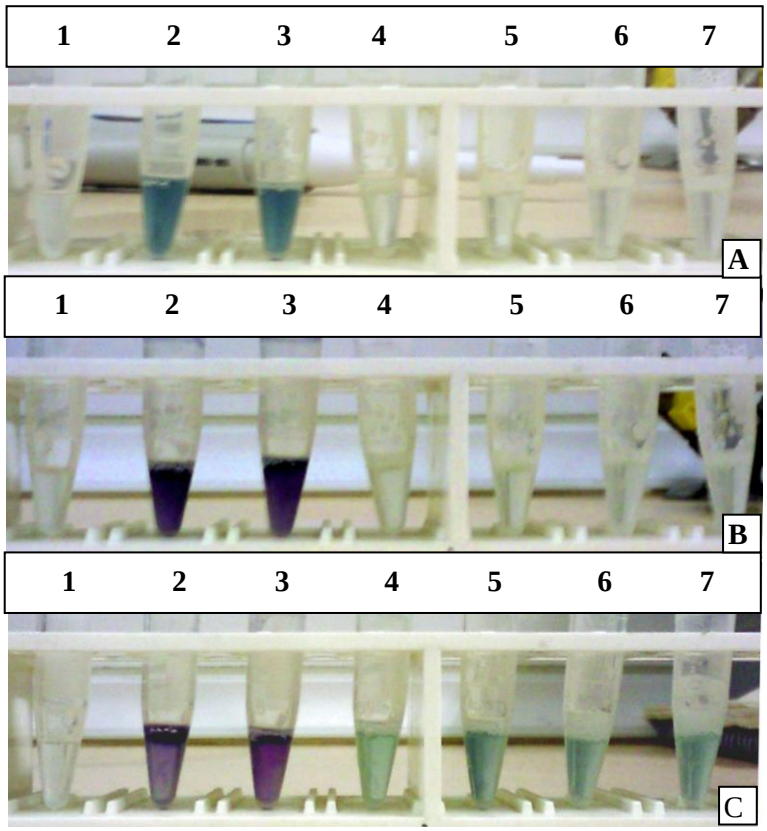


Figure 3.10 Monitoring green colour development in crude extracts of ProGOMN1 samples by ABTS-HRP coupled assay. **(A)** indicates the rapid colour development with related samples; **(B)** denotes that no colour change has been occurred even after 4 minutes; in **(C)** colour change has been detected 12-15 minutes later following involvement of additional crude extract **1,** ABTS-HRP assay solution (-) control; **2,** Pre-ProGOMN1 (+) control; **3,** MatGOMN6; **4,** ProGOMN1_F; **5,** ProGOMN1_3; **6,** ProGOMN1_6; **7,** ProGOMN1_7

As indicated in Figure 3.10, other ProGOMN1 constructs also show very low activity. These results may be due to the conformational changes that may have occurred as a result of pre-peptide deletion, or following potential pro-peptide processing, slower folding of GAO.

3.4.3.1 Purification of Pro-GAO by *Strep*-Tactin® Sepharose® Column and Activity Analysis

Both ProGON1 and ProGOMN1 samples were used in the purification of pro-galactose oxidase using the *Strep*-tag II method, described before ("IBA Protein TAGnologies"). Each dialysed cell extract of pre-peptide deleted mutational variants of GAO was loaded to the equilibrated *Strep*-Tactin® Sepharose® column. Following the anticipated adsorption of the *Strep*-tag II fused GAO to the column, the column was washed. Then the recombinant GAO enzyme was eluted in 6 fractions. Subsequently, regeneration and equilibration was performed to make the column ready for the next run.

Purification trials were performed according to the protocol in Section 2.2.7.1.1 with ProGON1_A, B; ProGOMN1_A,B,C,D and E and ProGOMN1_3 samples.

After purification of the samples, GAO activity was performed by ABTS-HRP coupled assay. Results are shown in Table 3.2.

Table 3.2 Summary of constructs indicating affinity chromatography purification and activity analysis.

Sample	Column	Protease Inhibitor	Activity (Crd. Ext.)	Activity (Pure)	Activity (30x Conc.)
Pre-ProGOMN1	A	(-)	(+)	(+)	n.a
ProGOMN1_A	B	(+)	(-)	(-)	n.a
ProGOMN1_B	B	(-)	(+)	(-)	(+)
ProGOMN1_C	A	(-)	(+)	(-)	(+)
ProGOMN1_D	A	(-)	(+)	(-)	(+)
ProGOMN1_E	B	(-)	(+)	(-)	(+)
ProGOMN1_F	A	(+)	(+)	n.a	n.a
ProGOMN1_3	B	(+)	(+)	(+)	n.a
ProGOMN1_6	n.a	(+)	(+)	n.a	n.a
ProGOMN1_7	n.a	(+)	(+)	n.a	n.a
ProGON1_A	A	(+)	(-)	(-)	n.a
ProGON1_B	n.a	(-)/(+)	(-)/(+)	(-)/n.a	n.a

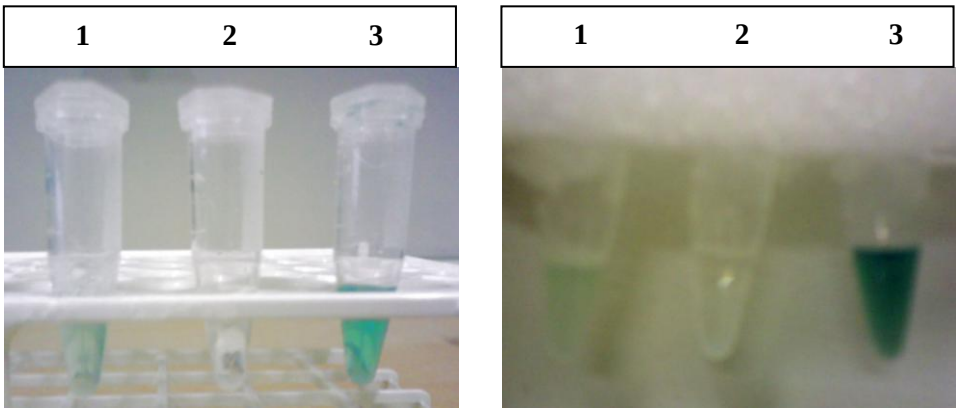


Figure 3.11 Detection of activity after purification, bringing together and 30x concentrating of ProGOMN1_B and C samples with qualitative ABTS-HRP coupled assay. **1,** ProGOMN1_B and C; **2,** Filtrate (flow through) of ProGOMN1_B and C (-) control; **3,** Pre-ProGOMN1 (+) control

Also activity was monitored in ProGOMN1_D and E separately after concentrating the samples ~ 30x without bringing together that time by vacuum concentrator.

As it is indicated in Figure 3.11 slight colour change was monitored following concentrating the purified samples. According to this result, it has been reassured that 8-aa pre-peptide deleted pro-GAO constructs are being expressed as confirmed via activity assays carried out by crude extracts and this also proves that *Strep*-tag is present and the absence of the enzyme activity after purification was not due to the affinity tag.

Colour detection with GAO ABTS-HRP coupled assay after concentrating the putative pure enzyme samples, which had not been able to be observed before, emphasizes the possible low level expression or partial degradation.

In order to solve the possible low expression problem, one of the colour giving sample ProGOMN1_E was chosen and expression was done with 3-fold cell culture of this construct in the presence of EDTA-free protease inhibitor, together with samples Pre-ProGOMN1, MatGOMN6 and ProGON1_B. However, no activity was detected with crude extracts (300 µl ABTS Assay Solution was used).

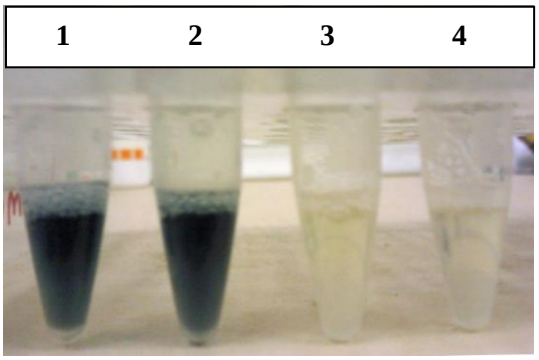


Figure 3.12 GAO ABTS-HRP coupled assay results with purified ProGOMN1_E samples of 3-fold cell culture. **1**, Pre-ProGOMN1; **2,** MatGOMN6; **3,** ProGON1_B; **4**, ProGOMN1_E

Similarly, as it is demonstrated in Figure 3.12, enzyme activity analysis after large scale growth of the ProGOMN1_E did not result in colour change following purification.

Following the expression studies with other transformants (ProGOMN1_3, ProGOMN1_6 and ProGOMN1_7), purification was decided to be performed with best colour giving sample ProGOMN1_3 among the others (see Figure 3.10). Colour development was also detected after purification (result not given here).

The main problem still seems to be the low level expression of the pre-peptide deleted constructs or low amount of solubilised protein of interest due to the development of lighter green colour change when compared to Pre-ProGOMN1 and MatGOMN6. On the other hand, the expressed protein may have undergone a conformational change hindering or decreasing the activity of the enzyme.

Meanwhile, SDS-PAGE was done in order to understand whether there was protein or not in inactive samples. In case there was protein, this might give some idea on inactivity of the samples. Also, SDS-PAGE analysis may provide to have insight about the autoprocessing of the pre-peptide within active pure samples.

3.4.3.2 Copper Loading and SDS-PAGE Analysis of Pro-GAO

The putative purified pro-GAO samples were also incubated with copper for the full processing of the enzyme as indicated in Section 2.2.8.1 in order to assure the cleavage of the pro-peptide.

SDS-PAGE was carried out for ProGOMN1_A and ProGON1_A samples (EDTA-free protease inhibitor was used), in order to observe the protein bands in case the protein might be detected although activity is not monitored due to a conformational change in the enzyme. However, no protein band could be detected belonging to the purification fractions of the protein of interest (result not given), like in ~30x concentrated green colour giving samples.

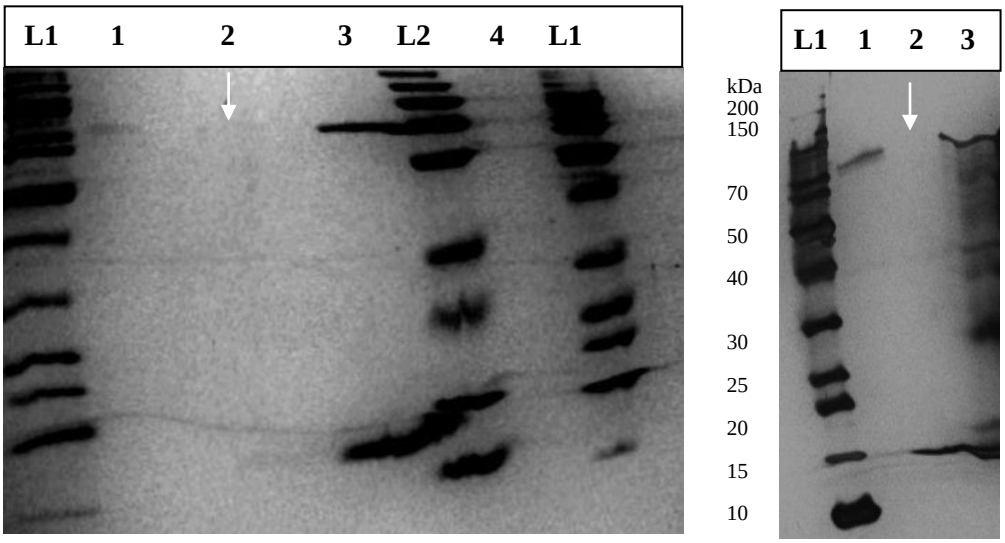


Figure 3.13 SDS-PAGE results with purified 3-fold cell cultures of ProGOMN1_E samples after Cu^{2+} treatment. **(A)** Coomassie Blue staining, **(B)** Silver staining. **L1,** PageRuler Unstained Protein Ladder**; L2,** PageRuler Plus Prestained Protein Ladder; **1,** Pre-ProGOMN1; **2,** ProGOMN1_E (3-fold); **3,** MatGOMN6; **4,** ProGON4_B.

Since slight colour change was detected with the concentrated purified samples of ProGOMN1, it is decided to be used in order to come over low expression or protein degradation problem.Enzyme dosage was increased by increasing the culture volume for purification and further concentration, using ProGOMN1_E, which gave positive results with the activity assay.

As seen in Fig. 3.13 a faint band of the expected size was observed on the coomassie blue stained gel, but not on the silver stained gel. Interestingly, however, on the silver- stained gel, the degradation product appeared in a much more distinct manner. Unless it is resulted from contamination, these observations, suggest that the Pro-GAO form is expressed, but is rapidly degraded due to instability of the molecular structure. Thus, the 8 aa pre-peptide, is likely to be essential for the stable folding and maintenance of the GAO protein structure.

SDS-PAGE was also carried out for samples obtained from ProGOMN1_F and ProGOMN1_3 samples, following copper treatment (Fig. 3.14). Accordingly, a degradation product at c. 40 kDa was observed in both pre-pro- and the Pro-GAO samples, but not the mature GAO, suggesting, the pre-pro- and the Pro-GAO forms to be less stable than the mature GAO.

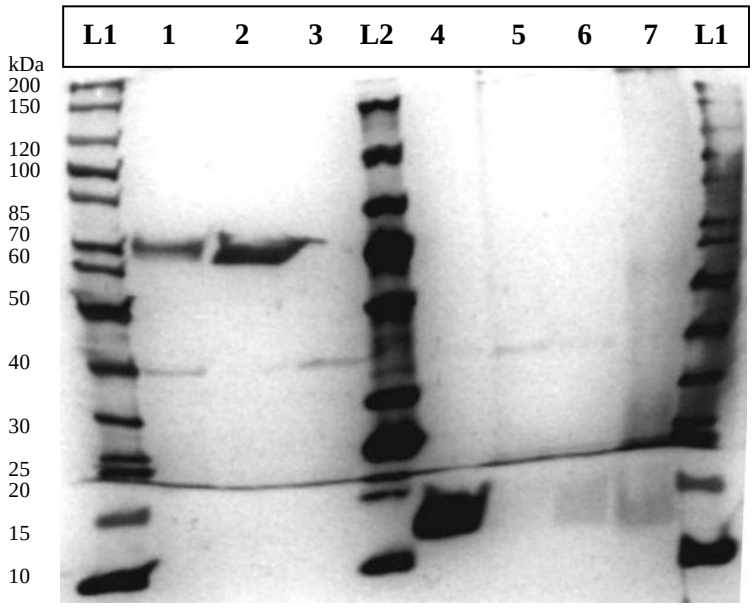


Figure 3.14 Detection trials of purified activity showing ProGOMN1_3 before and after Cu^{2+} treatment by SDS-PAGE and silver staining. **L1,** PageRuler Unstained Protein Ladder**;** **L2,** PageRuler Plus Prestained Protein Ladder; **1,** Pre-ProGOMN1 (+) Cu^{2+}; **2,** MatGOMN6 (+) Cu^{2+}; **3,** ProGOMN1_F; **4,** ProGOMN1_3 (+) Cu^{2+}; **5-7,** ProGOMN1_3 (-) Cu^{2+} fractions 3-5, respectively.

3.4.3.3 Inclusion Body Isolation and Solubilisation Studies of Pro-GAO Samples

Variable numbers of protein expression systems renders the efficient bacterial production of most proteins possible. Yet, heterologous overexpression of proteins occasionally leads to aggregation and accumulation in inclusion bodies (IB). Inclusion body formation can be observed in various host systems; regardless of intrinsic properties, such as molecular weight, hydrophobicity, folding pathway of a certain protein; exceptionally in the case of disulfide bonded proteins. Inclusion body formation can be anticipated if the protein is produced as a cytoplasmic protein in bacteria, since disulfide bonds usually do not form in this reducing compartment. Consequently, the result is aggregation of improperly folded proteins (Lilie e*t al.*,1998).

In this study, even though the activity was observed within the crude extracts and pure samples, the enzymes could not be detected by SDS-PAGE, even with silver staining, which is the most sensitive colorimetric method. Besides, the possibility of protein instability, another explanation could be the fact that most GAO is entrapped in inclusion bodies, thus becoming insoluble and unavailable for protein purification and activity analysis. In order to test this possibility, IB isolation and protein solubilisation studies were carried out.

3.4.3.3.1 Denaturation and Solubilisation of Inclusion Bodies by Various Urea Concentrations

Expression experiments were designed with Pre-ProGOMN1, MatGOMN6 and ProGOMN1_E. Growth conditions and induction type were explained in Section 2.2.6.3.1. Another ProGOMN1_E culture was also prepared and not induced with IPTG to use as a negative control for further comparison in order to detect inclusion body formation according to the differences between induced and non-induced samples. ProGOMN1_E sample was resuspended in different concentration of urea, ranging between 2 M - 8 M following centrifugation.

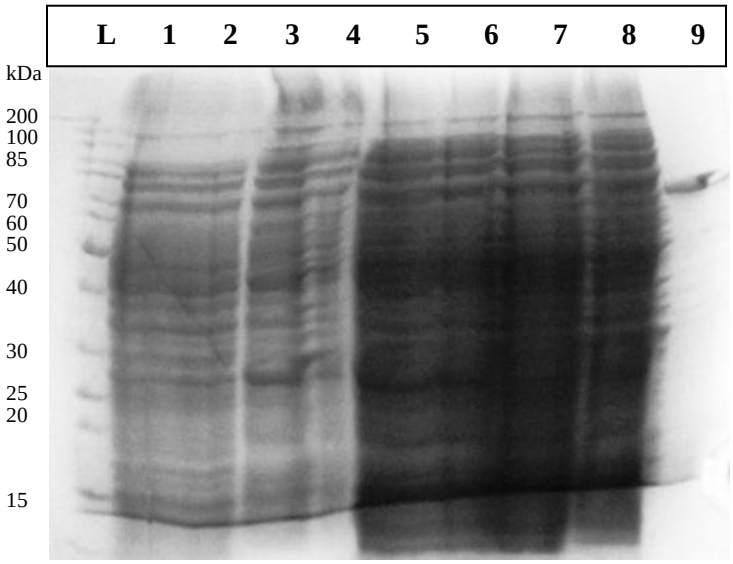


Figure 3.15 Effect of urea at different concentrations on induced ProGOMN1_E sample and comparison with its non-induced forms. **L,** PageRuler Unstained Protein Ladder; **1,2,** ProGOMN1_E non-induced (-) resuspended in PBS; **3,4,** ProGOMN1_E induced (+) resuspended in PBS; **5,** ProGOMN1_E (-) in 2 M urea ; **6,** ProGOMN1_E (+) in 2 M urea; **7,** ProGOMN1_E (-) in 4 M urea; **8,** ProGOMN1_E (+) in 4 M urea; **9,** MatGOMN6 +Cu^{2+}.

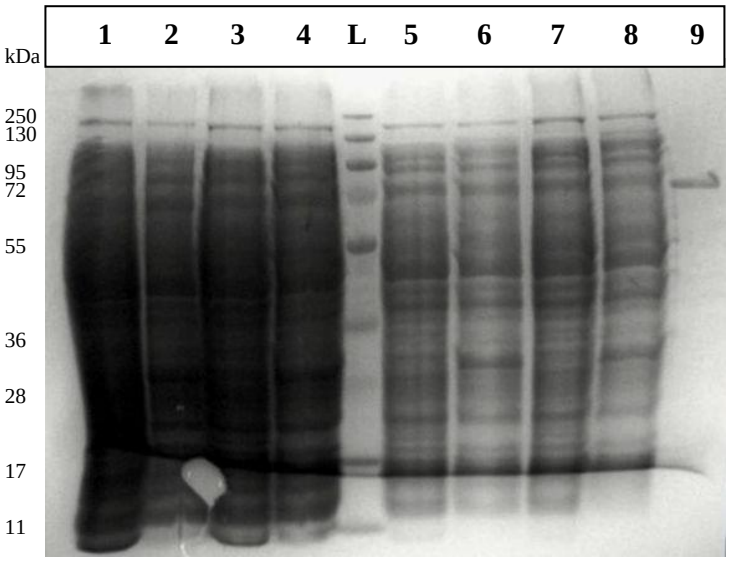


Figure 3.16 SDS_PAGE analysis showing the effect of different molarities of urea on the soluble and insoluble parts of induced and non-induced ProGOMN1_E sample. **L,** PageRuler Plus Prestained Protein Ladder; **1,** ProGOMN1_E (-) resuspended in 6 M urea; **2,** ProGOMN1_E (+) in 6 M urea; **3,** ProGOMN1_E (-) in 8 M urea; **4,** ProGOMN1_E (+) in 8 M urea; **5,** ProGOMN1_E (-) pellet in 6 M urea; **6,** ProGOMN1_E (+) pellet in 6 M urea; **7,** ProGOMN1_E (-) pellet in 8 M urea; **8,** ProGOMN1_E (+) pellet in 8 M urea; **9,** MatGOMN6 +Cu^{2+}.

As it is demonstrated in Figure 3.15 and 3.16 application of different urea concentrations did not result in a different pattern between induced and non-induced samples around the expected band size.

3.4.3.3.2 Purification of the Inclusion Bodies by Using Inclusion Body Isolation Micro Kit

The protocol of ProteoSpin™ Inclusion Body Isolation Micro Kit (Norgen) for acidic proteins was used in order to isolate and purify in case the protein of interest was aggregated as IBs (Section 2.2.6.5.2).

A heterologous expression experiment was designed for the samples ProGOMN1_E, ProGOMN1_3, ProGON1_B with the working constructs pPre-ProGOMN1, pPre-ProGON1and pMatGOMN6. The inclusion bodies are thought to be inert aggregates of improperly folded insoluble proteins. However, it has been reported that physiological aggregation in bacteria might only lead to a moderate loss of biological activity and that inclusion bodies can be used as efficient catalysts in reaction mixtures (García-Fruitós *et al.*, 2005).

After the qualitative enzyme activity assay was done it was understood that the protein was expressed and IB isolation kit was decided to be used. 1,5 ml of the culture was separated before the sonication of the cells and used for the kit. The supernatant obtained by

centrifugation, following the addition of Cell Lysis Reagent, was also kept in order to analyse whether there was soluble fraction of the protein of interest.

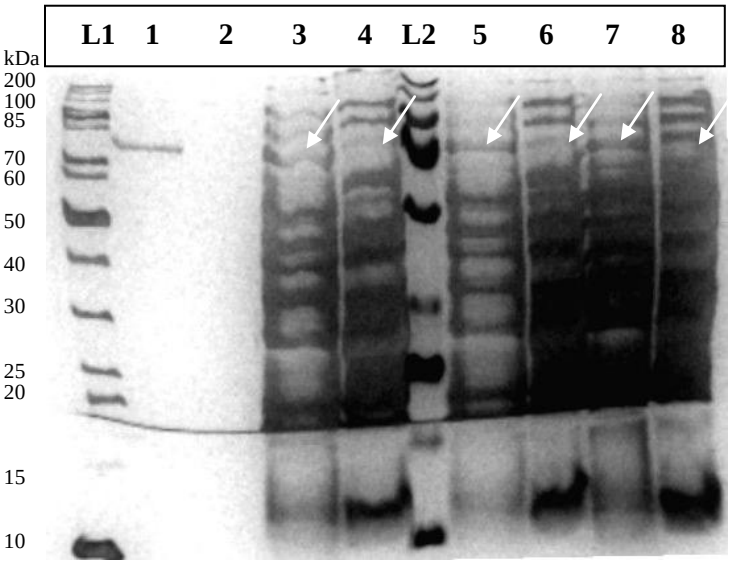


Figure 3.17 Analysis of IB formation in ProGOMN1_E and ProGOMN1_3 samples by using IB isolation kit. **L1,** PageRuler Unstained Protein Ladder; **L2,** PageRuler Plus Prestained Protein Ladder; **1,** Pre-ProGOMN1 (-) Cu^{2+}; **2,** ProGOMN1_3 (-) Cu^{2+}; **3,** ProGOMN1_3 after IB kit use; **4,** ProGOMN1_3 supernatant after cell lysis; **5,** ProGOMN1_E (IB); **6,** ProGOMN1_E (IB supernatant); **7,** ProGON1_B (IB); **8,** ProGON1_B (IB supernatant).

As it was demonstrated in Figure 3.18, the IB isolation eluates of ProGOMN1_E and ProGOMN1_3 samples were brought together and since they were solubilised it was attempted to purify and obtain the bands around 70 kDa by *Strep*-Tactin® affinity chromatography columns. After separating small amounts of protein without copper treatment, copper loading was performed for the rest of the sample.

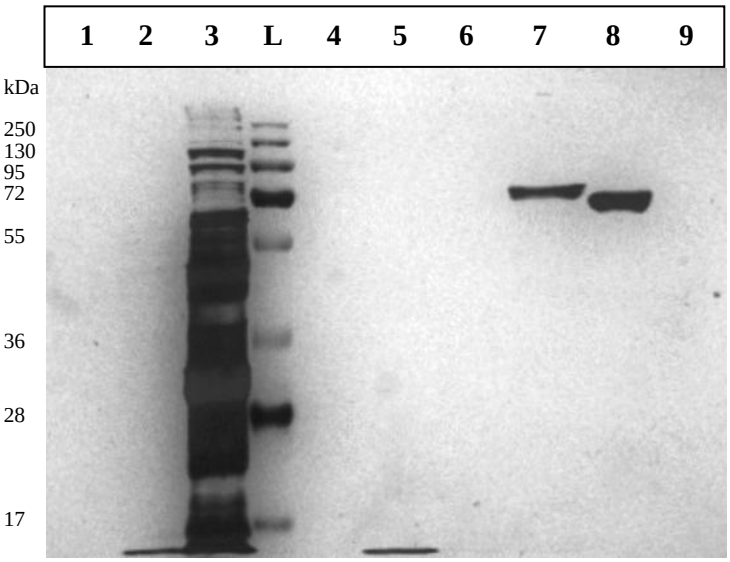


Figure 3.18 *Strep*-Tactin® affinity chromatography purification of putative IBs obtained from ProGOMN1_E and ProGOMN1_3 after IB isolation kit application. **L,** PageRuler Plus Prestained Protein Ladder; **1,** ProGOMN1_3 (-) Cu^{2+}; **2,** ProGOMN1_E (IB); **3,** ProGOMN1_E (IB supernatant); **4,** ProGOMN1_E and ProGOMN1_3 (IB) after *Strep*-Tactin purification (-) Cu^{2+}; **5,** Flow through (*Strep*-Tactin); **6,** ProGOMN1_E and ProGOMN1_3 (IB) (*Strep*-Tactin) (+) Cu^{2+}; **7,** Pre-ProGOMN1 (-) Cu^{2+}; **8,** MatGOMN6 (+) Cu^{2+}; **9,** ProGOMN1_3 (+) Cu^{2+}.

When the lanes were analysed in Figure 3.17, especially the elutions after IB isolation and supernatants resulted from cell lysis (lanes 4&5; 7&8; 9&10) it could be noted that the band around 70 kDa belonging to IB isolation kit results did not appear in the supernatant part, which is indicated by arrows. This could have been interpreted as the protein which was unable to be observed after the cell lysis in the soluble fraction, was solubilised and appeared after IB kit treatment. Therefore, this finding leads to the purification of the samples via *Strep*-Tactin® column. However, it was assessed that the multi-band profile and intensity of bands could not be regarded as an inclusion body pattern (McPherson, personal communication). The purification results given in Figure 3.18 supported this remark as well.

3.4.4 Expression of pProGON1 and pProGOMN1 Constructs by Auto-Induction

At the University of Leeds, an alternative induction method, auto-induction was used, other than IPTG induction with pET vector system, for expression studies of ProGAO samples during bacterial growth. Auto-induction allows slow production of even fragile proteins and therefore helps proper folding. Furthermore, yields of target protein are typically considerably several folds higher than that obtained by IPTG induction via richer medium (Studier, 2005).

Based on the abovementioned reasons and experimental data in Section 1.7.2, the experimental precedure with pre-peptide deleted samples was set up according to these conditions.

pPre-ProGOMN1, pPre-ProGON1, pMatGOMN6, pProGOMN_1 and pProGON_1 were transferred into the BL21 Star (DE3) cells. These transformed cells were inoculated into 400 ml 8ZY-4LAC–25mM succinate medium. Bacterial cell growth was monitored at 25°C at 250 rpm as described in Section 2.2.6.3.2 (Figure 3.19). Auto-induction system working at 25°C produces higher levels of soluble proteins than the IPTG induction system rendering translation slower and correct folding possible. Moreover, at 25°C oxygen becomes more soluble to attain high saturating cell densities, subsequently in order to enhance the induction by lactose limited oxygen is needed (Deacon, 2008).

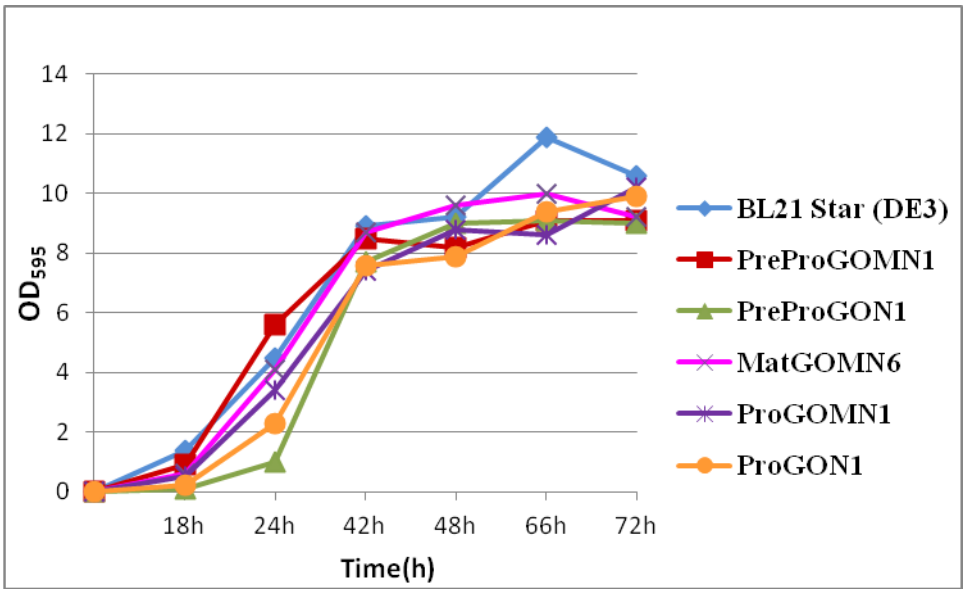


Figure 3.19 Investigation of optimum time course for bacterial growth by auto-induction media 8ZY-4LAC–25mM succinate at 25°C at 250 rpm by multiple GAO samples.

As it is derived from the graph above 48-hour growth is sufficient for achieving high cell densities.

3.4.4.1 Small Scale Bacterial Lysis and Analysis by Preliminary Auto-Induction Trials

Although 48-hour growth was sufficient, it was decided to monitor the cultures for 72 hours at the beginning. At the end of the expression studies, 1 ml cultures were lysed with 250µl lysis buffer as it is explained in Section 2.2.6.4.2. Small volumes of soluble and insoluble parts were resuspended in equal volume of lysis buffer so that it is possible to comment on soluble and insoluble parts directly when loaded on an SDS gel beside each other.

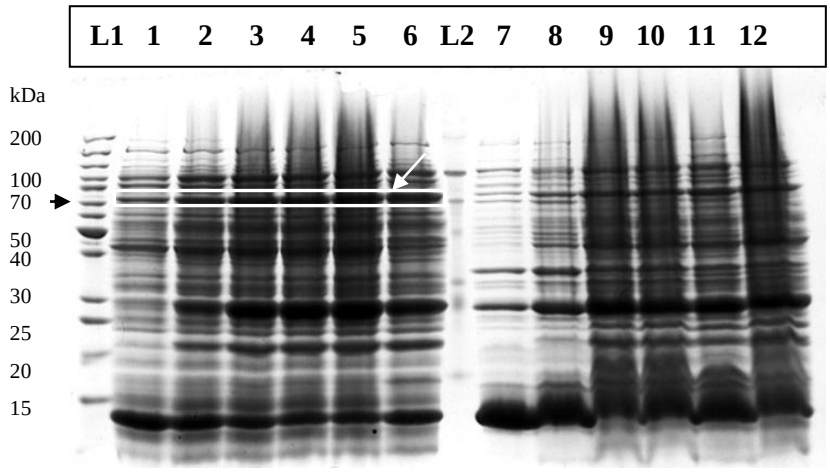


Figure 3.20 Detecting Pre-ProGOMN1 expression by auto-induction method and comparison of soluble and insoluble parts. **L1**, PageRuler Unstained Protein Ladder; **1-6 Soluble parts,** 16, 24, 40, 48, 64 and 72 hours, respectively; **L2** Prestained Protein Marker (NEB); **7-12 Insoluble parts,** 16, 24, 40, 48, 64 and 72 hours, respectively.

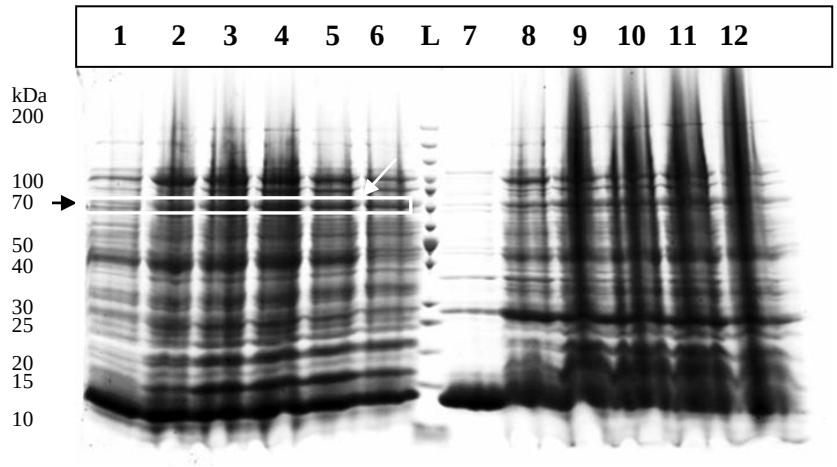


Figure 3.21 SDS-PAGE of soluble and insoluble parts of BL21 Star (DE3) crude extract as a (-) control after expression by auto-induction method for analysis and comparison with pre-peptide deleted GAO carrying constructs. **L**, PageRuler Unstained Protein Ladder; **1-6 Soluble parts,** 16, 24 ,40, 48, 64 and 72 hours, respectively; **7-12 Insoluble parts,** 16, 24, 40, 48, 64 and 72 hours, respectively.

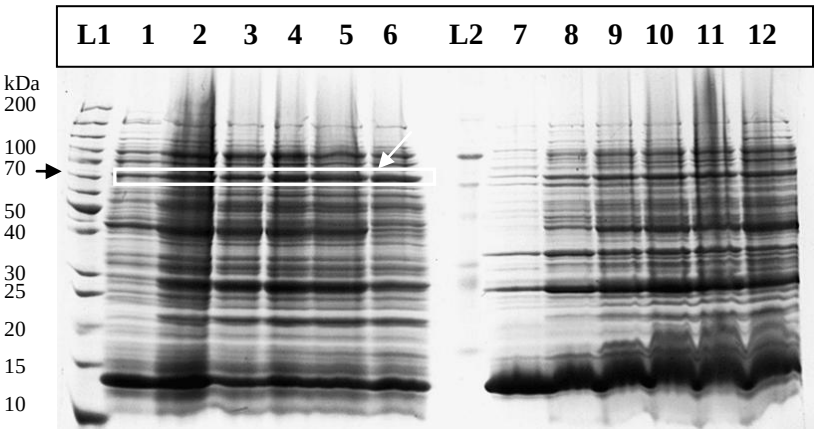


Figure 3.22 SDS-PAGE analysis of certain time interval fractions of ProGOMN1 and comparison of soluble and insoluble parts. **L1**, PageRuler Unstained Protein Ladder; **L2,** Prestained Protein Marker NEB; **1-6 Soluble parts,** 16, 24, 40, 48, 64 and 72 hours respectively; **7-12 Insoluble parts,** 16, 24, 40, 48, 64 and 72 hours, respectively.

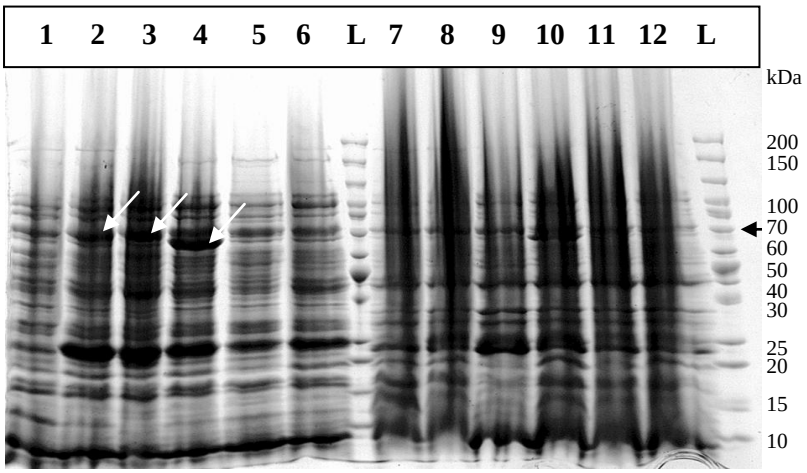


Figure 3.23 SDS-PAGE analysis of expression of pre-peptide deleted GAO by auto-induction method through comparing the 72-hour sample crude extracts of each variant. **L,** PageRuler Unstained Protein Ladder; **Soluble parts, 1,** BL21 Star DE3 (-) control; **2,** Pre-ProGOMN1**; 3,** Pre-ProGON1; **4,** MatGOMN6; **5,** ProGOMN1; **6,** ProGOMN1_3; **Insoluble parts, 7-12** in the same order as soluble parts.

Together with the positive controls that is known to be expressed in *E.coli*; MatGOMN6, Pre-ProGOMN1 and Pre-ProGON1 (which have a relatively lower expression level when compared with MatGOMN6); BL21 Star (DE3) fractions were used as a (–) control for comparison. However, as it is seen in Figure 3.21 unfortunately BL21 Star (DE3) has a same sized band around 70kDa, which may mask the insight of whether expression is carried out properly or not in constructs ProGOMN1 and ProGOMN1_3; also in ProGON1, in further trials. Purification studies should be carried out in order to solve this issue

3.4.4.2 Purification of ProGON1 and ProGOMN1 by *Strep*-Tactin® MacroPrep® Cartridges

Following auto-induction, as explained in Section 2.2.6.3.2, the cells were disrupted by lysis buffer containing Complete Protease Inhibitor Cocktail Tablet (Roche) and proteins were pooled by using *Strep*-Tactin® MacroPrep® Cartridges connected to automated low pressure chromatography system; GradiFrac™ (Pharmacia Biotech, Sweden), as described in (Section 2.2.7.1.2).

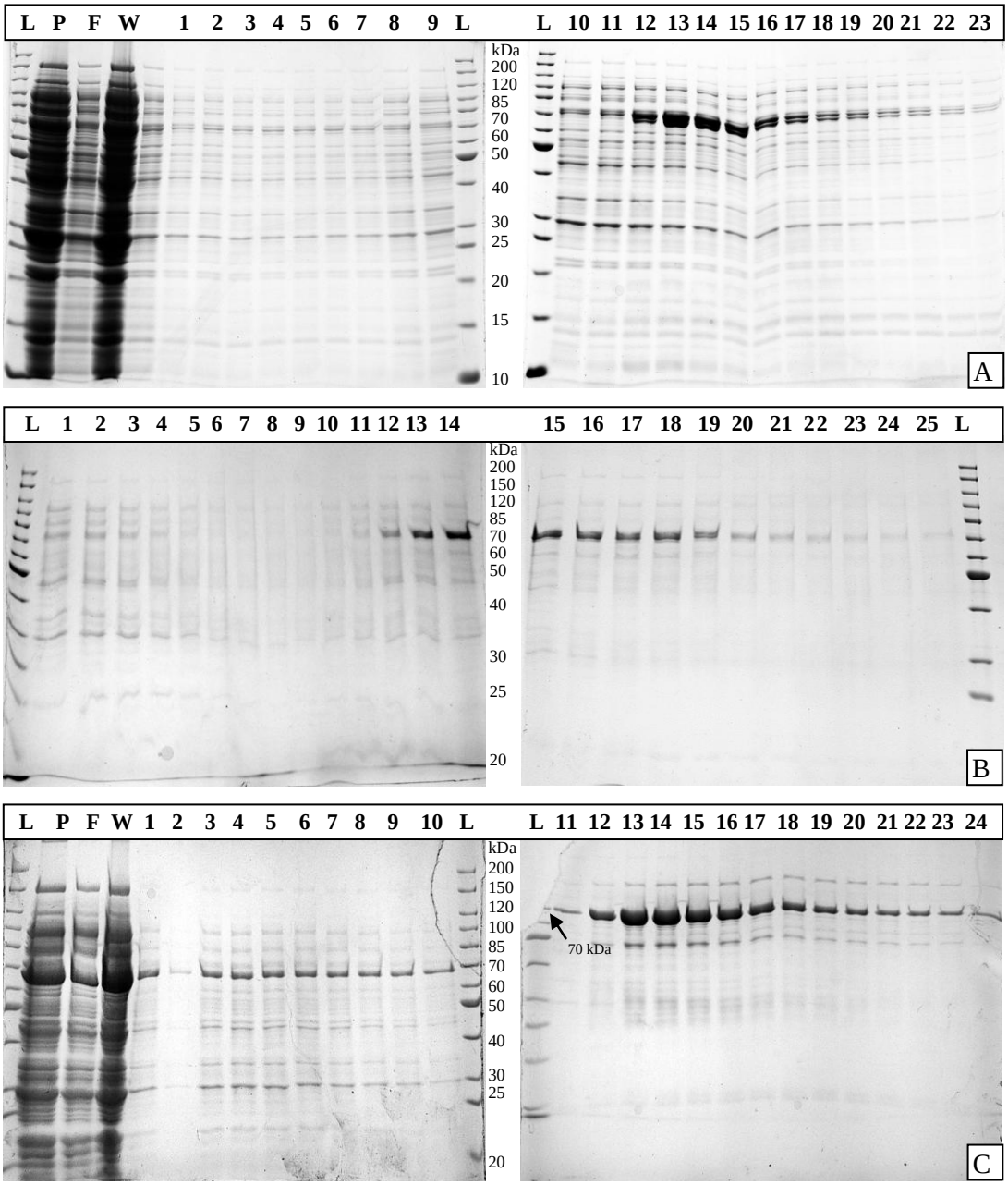


Figure 3.24 SDS-PAGE results showing column purification fractions from Pre-ProGON1 **(A)**, Pre-ProGOMN1 **(B)** and MatGOMN6 **(C)**. **L,** PageRuler Unstained Protein Ladder; **P,** Pre-load; **F;** Flow through; **W,** Washing; **1-23;** Fractions

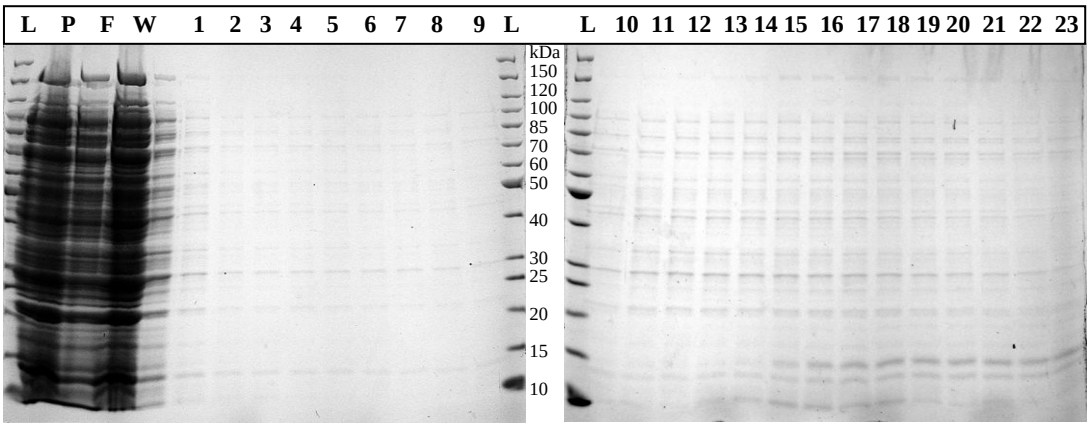


Figure 3.25 Purification of ~7 fold concentrated Pro-GAO from ProGOMN1 in 23 fractions. **L,** PageRuler Unstained Protein Ladder; **P,** Pre-load; **F;** Flow through; **W,** Washing; **1-23,** Fractions

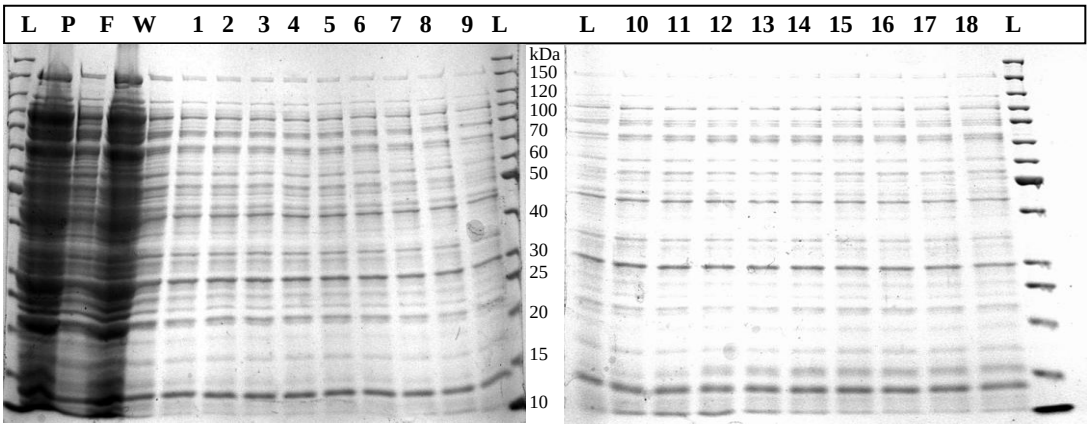


Figure 3.26 Column fractions of Pro-GAO from ProGOMN1_3. **L,** PageRuler Unstained Protein Ladder; **P,** Pre-load; **F;** Flow through; **W,** Washing; **1-18;** Fractions

As given in Figure 3.24-3.26, the fractions of the related samples do not seem to be properly purified. Even Pre-Pro-GAO from Pre-ProGOMN1, Pre-ProGON1 and GAO from MatGOMN6, which were purified as a single band in previous studies with *Strep*-Tactin® Sepharose® column, were not efficiently purified. However, as seen in in Figure 3.24, Pre-Pro-GAO and the mature GAO are detected easily as a distinct band, and can be distinguished from the contaminants. On the other hand, Pro-GAO does not appear as a distinct band (Fig. 3.25), in contrast to the Pre-Pro-GAO and GAO fractions. This suggests that the auto-induction system has not contributed much to the expression and production of Pro-GAO.

The purification problem may have been resulted from;

- Loading more than column capacity (10 ml): More than 10 ml cell extract was loaded in order to increase the amount of low level expressed protein, especially in pre-peptide deleted constructs.
- Concentrated cell extract: Cell extract was concentrated by ultrafiltration (Millipore Amicon Ultra Centrifugal Filter Units), to have more protein in a certain amount of volume proper for the column capacity
- Inadequate washing: Even 5X 1CV (5 ml) is enough according to the protocol the elution peak chart was not compatible with this amount indicating need of more washing. 30-40 ml extra washing was done.

The purification was also repeated with ProGON1 and ProGOMN1 constructs with a new expression experiment set up in order to sort out these points. Although the contaminants were washed out by subsequent excessive washings until the baseline was achieved in elution peak chart, there still seems to be an impurity problem. Moreover, the protein could not be detected and distinguished among other slight bands because of impurity.

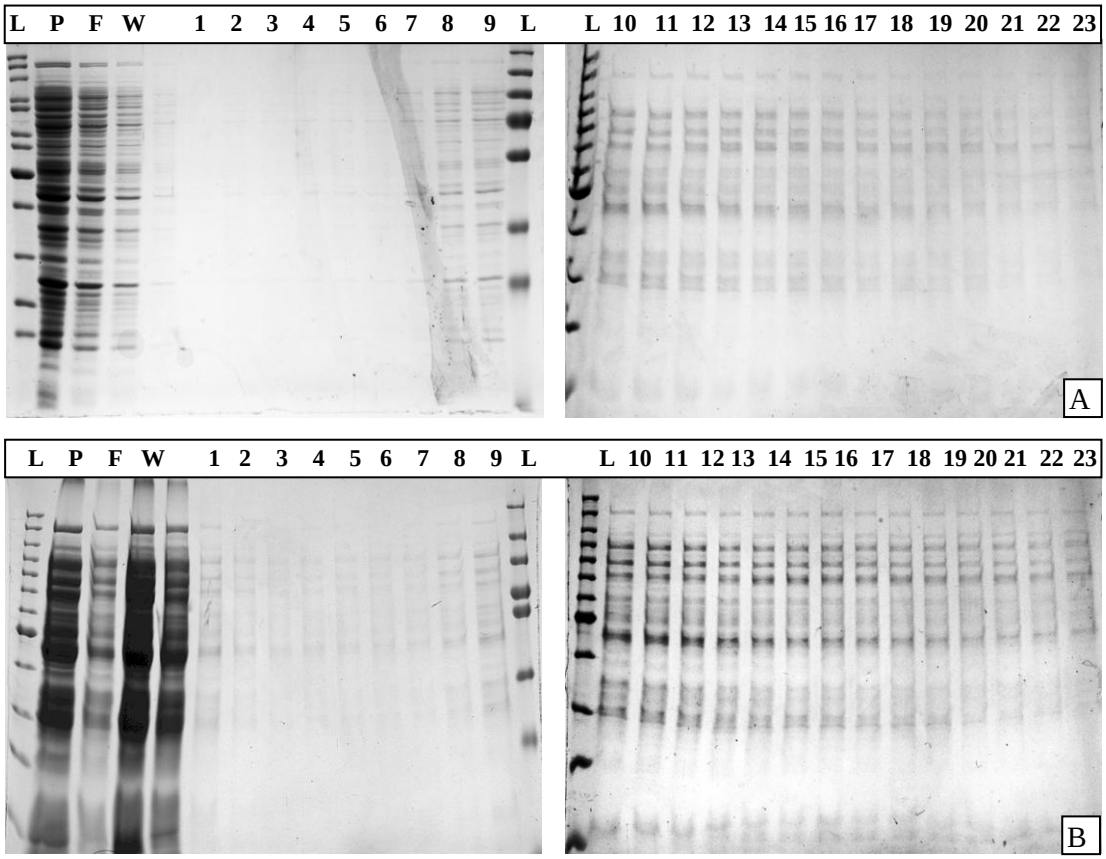


Figure 3.27 SDS-PAGE results of Pro-GAO from ProGON1 **(A**) and ProGOMN1 **(B)** in 23 fractions achieved by automated chromatography system. **L,** PageRuler Unstained Protein Ladder; **P,** Pre-load; **F;** Flow through; **W,** Washing; **1-23;** Fractions

Moreover, it was also indicated that purification trials of pro-forms of GAO by *Strep*-Tactin® MacroPrep® HP-R Cartridge may have some problems like precipitation, the reason for this is very likely to be the column support itself, polymethacrylate, different than that of *Strep*-Tactin® Sepharose®, which is agarose (Deacon, 2011, personal communication).

3.4.4.2.1 Effect of Copper Loading on GAO Samples

It is known that copper with the molecular oxygen are the only requirements for Pro-GAO to be processed autocatalytically without any extra accessories (Rogers, 2000). Therefore, after bringing the positive fractions together, the proteins were loaded by copper as it is described in Section 2.2.8.1 and SDS-PAGE was done in order to detect proteins.

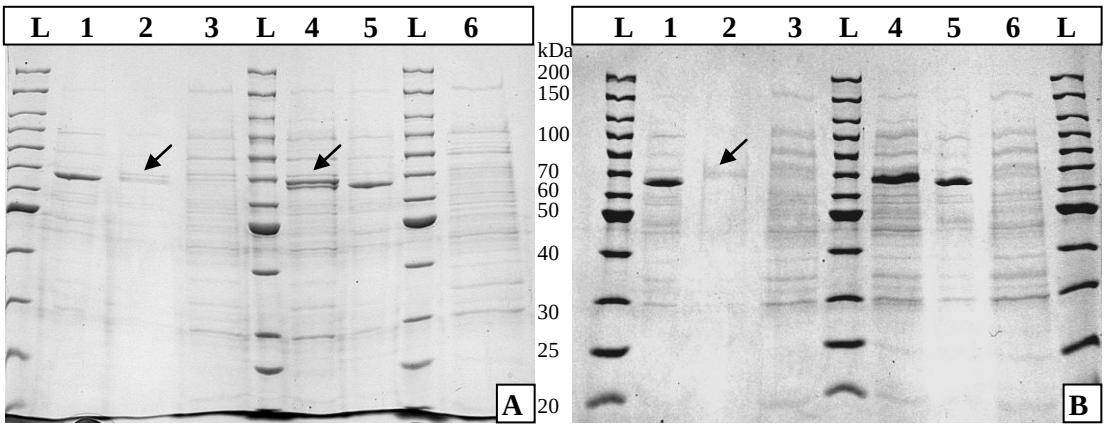


Figure 3.28 SDS-PAGE analysis of GAO constructs in terms of processing before **(A)** and after **(B)** Cu^{2+} treatment on separate gels. **L**, PageRuler Unstained Protein Ladder; **1,** MatGOMN6; **2,** Pre-ProGOMN1; **3,** ProGOMN1; **4,** Pre-ProGON1; **5,** MatGOMN6; **6,** ProGOMN1_3.

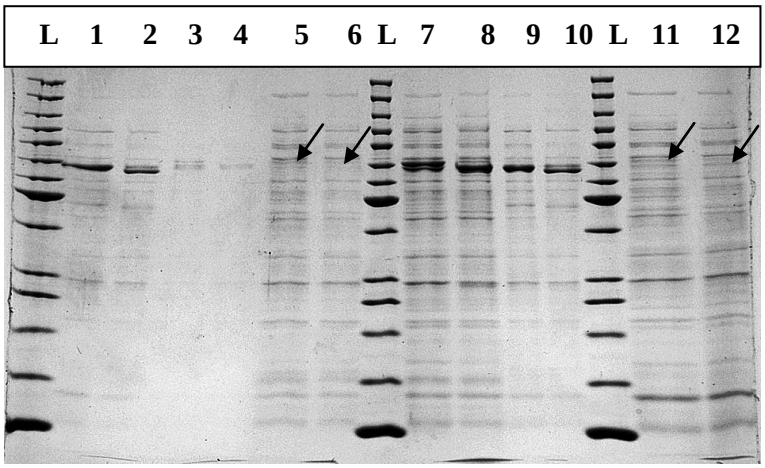


Figure 3.29 Comparison of GAO constructs before and after Cu^{2+} treatment to assess the effect of copper on self-processing by SDS-PAGE analysis **L**, PageRuler Unstained Protein Ladder; **1,** MatGOMN6 (-) Cu^{2+}; **2,** MatGOMN6 (+) Cu^{2+}; **3,** Pre-ProGOMN1 (-) Cu^{2+}; **4,** Pre-ProGOMN1 (+) Cu^{2+}; **5,** ProGOMN1 (-) Cu^{2+}; **6,** ProGOMN1 (+) Cu^{2+}; **7;** Pre-ProGON1(-) Cu^{2+}; **8,** Pre-ProGON1 (+) Cu^{2+} **9,** MatGOMN6 (-) Cu^{2+}; **10,** MatGOMN6 (+) Cu^{2+}; **11,** Pre-ProGOMN1_3 (-) Cu^{2+}; **12,** Pre-ProGOMN1_3 (+) Cu^{2+}.

As it is derived from Figure 3.28 and Figure 3.29; both the mature and the pre-pro-peptide carrying constructs were all processed after Cu^{2+} treatment indicating Cys-Tyr thioether bond formation. Taking into account that the only difference between the mature and the pre-pro-constructs was the existence of pre-pro-peptide, the motility difference on SDS-PAGE can only be attributed to the unprocessing of the pro-peptide. Based on previous studies (Gençer, (Kocuklu), 2005), especially the N-terminal amino acid sequencing result of Pre-ProGOMN1 see (Figure 3.30), it can be strongly asserted that the pre-peptide was still there as well as the pro-peptide.

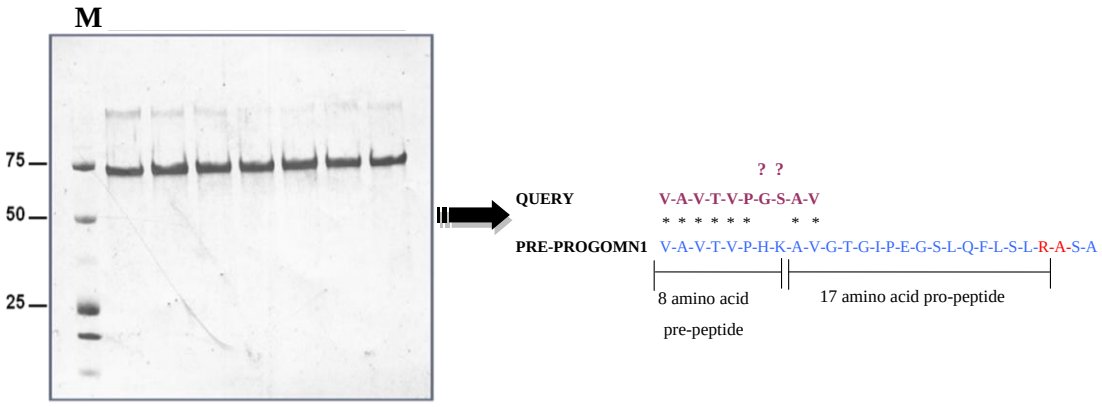


Figure 3.30 PVDF electroblot sample of Pre-ProGOMN1 (all lanes except molecular marker) and N-terminal amino acid sequencing results derived from the electroblot sample.

In terms of ProGOMN1 and ProGOMN1_3, according to the Figure 3.29 it is very hard to detect and characterise the bands due to the uncertainty resulted from low amount of enzyme and impurity of the samples.

3.4.4.2.2 Electroblotting and Western Blot Analysis of Pro-GAO

Auto-induction method followed by protein analysis did not yield any distinct product for Pro-GAO. It was decided that large scale-growth followed by western analysis may solve the problem, due to the sensitivity of the method to very low quantities of protein.

Western Blotting was performed by using *Strep*-Tag antibody as a primary antibody and goat anti-mouse IgG–HRP as a secondary antibody with TMB chromogenic substrate for detection (Section 2.2.8.6).

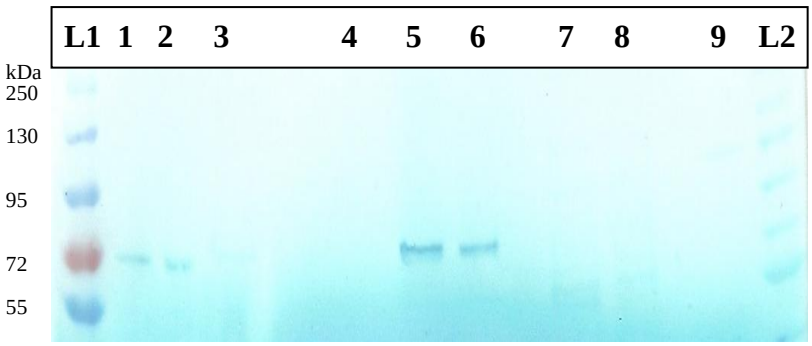


Figure 3.31 Western blotting for detection of ProGOMN1 before and after Cu^{2+} treatment **L1,** PageRuler Prestained Protein Ladder; **1,** MatGOMN6 (-) Cu^{2+}; **2,** MatGOMN6 (+) Cu^{2+}; **3,** Pre-ProGOMN1 (-)Cu^{2+}; **4,** Pre-ProGOMN1 (+) Cu^{2+}; **5,** Pre-ProGON1 (-) Cu^{2+}; **6,** Pre-ProGON1 (+) Cu^{2+}; **7;** ProGOMN1 (-) Cu^{2+}; **8,** ProGOMN1 (+) Cu^{2+}; **9,** His-tagged CATPO enzyme (-) control; **L2**, PageRuler Unstained Protein Ladder (contains an integral *Strep*-tag® II sequence).

Although the experiment was successful, yielding clear bands for Mature GAO and Pre-Pro-GAO, unfortunately, Pro-GAO was not apparent, indicating lack of production of even small quantities of the enzyme. These results were clearly unexpected, and difficult to explain indeed.

3.4.4.2.3 Mass Spectrometry Analysis

In order to understand whether pro-peptide processing on Pro-GAO samples was realized or not by comparing the results of mass data with Pre-Pro-GAO and mature-GAO, the samples were submitted to Mass Spectrometry Facility, University of Leeds following dialysis against ammonium acetate as indicated in Section 2.2.8.7. The copper treated and untreated original samples (Pre-ProGOMN1, Pre-ProGON1 and MatGOMN6) were submitted first in order to have an idea on whether mass spectrometry could give reliable results in this heterogeneity.

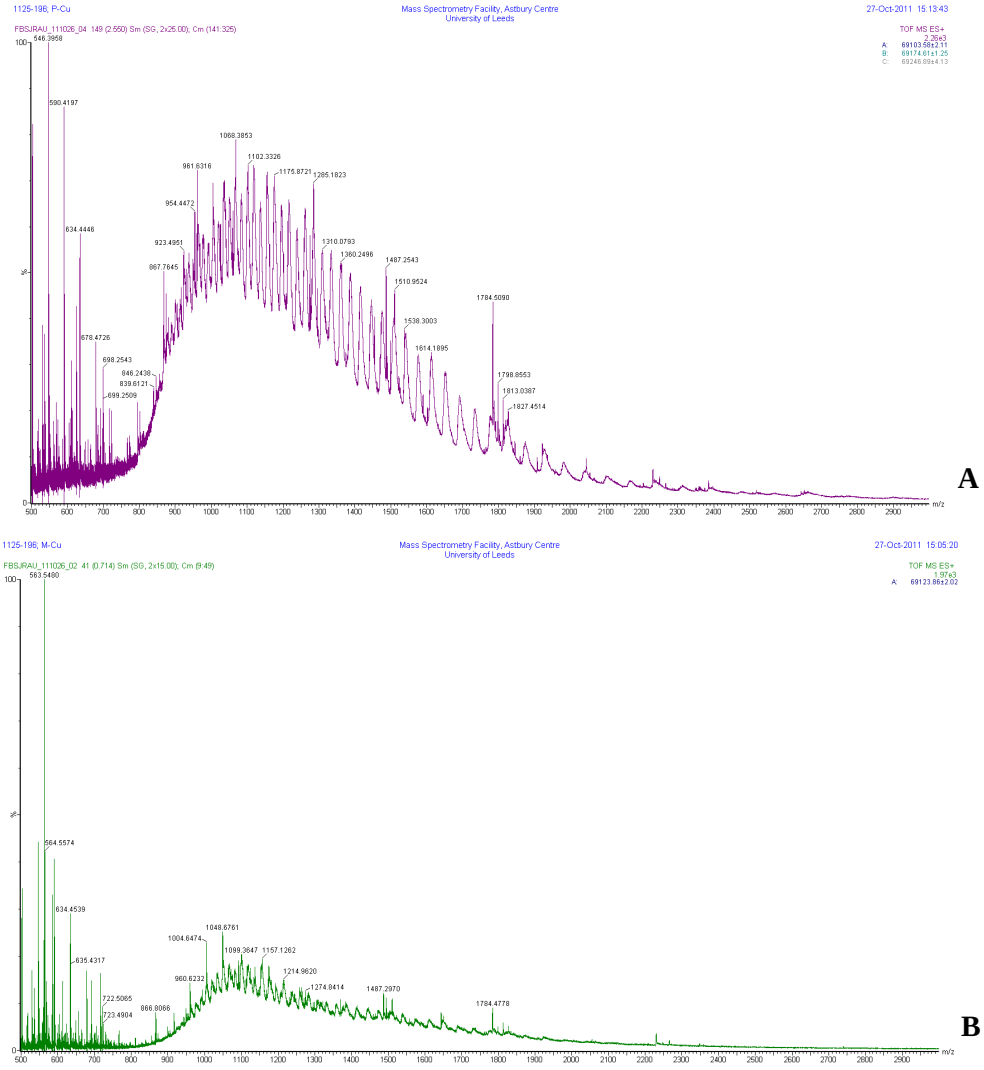


Figure 3.32 Mass spectrometry analysis results of Pre-Pro-GAO from Pre-ProGON1 (-) Cu **(A)**; and Mature GAO from MatGOMN6 (-) Cu **(B).**

Only the mass data of Pre-ProGON1 looks compatible (69,1 kDa) (Figure 3.32 A) with the theoretical calculations (69,4 kDa) at a certain degree (all data are not given) . Since the samples are not pure, the molecular mass of the protein of interest could not be derived and interpreted from the mass data. Therefore, ProGON1 and ProGOMN1 were not submitted to mass spectrometry.

3.4.4.2.4 Qualitative GAO Activity Assays with Partially Purified ProGON1 and ProGOMN1

GAO activity of all samples was monitored by ABTS coupled assay following copper treatment (Section 2.2.9.1).

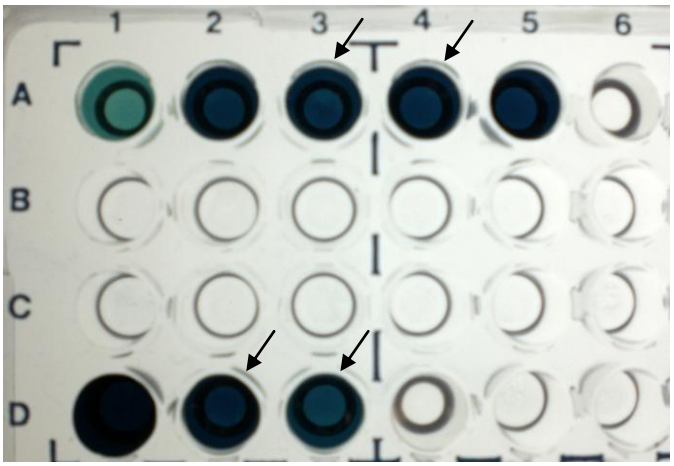


Figure 3.33 Qualitative enzyme activity analysis for detection of pre-peptide deleted GAO samples after Cu^{2+} treatment **A1,** Pre-ProGOMN1(protein loss in purification); **A2,** Pre-ProGON1; **A3,** MatGOMN6; **A4,** ProGOMN1; **A5** ProGOMN1_3; **A6,** (-) control; **D1;** Pre-ProGOMN1; **D2**; ProGOMN1; **D3,** ProGON1, **D4,** (-) control (Results in Lane D belong to second expression and purification trials)

As it is given in Figure 3.33, the products of all constructs were giving the green colour indicating the activity of the enzyme, thus formation of the thioether bond.

By this assay it has been approved that pre-peptide deleted constructs are being expressed and showing activity in spite of their low expression level and undetectability in SDS-PAGE Although the bands of ProGOMN1 and ProGON1 could not be detected either with SDS-PAGE or by western blotting, according to the results obtained from DNA samples which were submitted for DNA sequencing twice, both in Turkey and UK, supported with the ABTS-coupled assay results; specify that the deletion of the 8-aa pre-peptide is successfully carried out and there is not any problem with the reading frame.

Then it can be strongly asserted that the reason of the undetectability of the pre-peptide deleted samples is due to the low level of expression or instability of the protein.

Additionally, Cannarozzi and colleagues (2010) suggest that the pattern of codon usage can control ribosome speed and translation fine-tuning , which increase the efficiency of protein synthesis.The use of identical codons at the N-terminus consecutively makes translation faster than when they are arranged alternatively. Similarly, in Pre-ProGOMN1 and Pre-ProGON1, N-terminal silent mutations within first 6 codons (As and Ts for Gs and Cs) which introduced for high level of expression (Deacon, 2008) leads to the occurance of repeatative valine codons (Figure 3.34). Removal of the 8-aa pre-peptide covering these two substitutions might have influenced the translation and protein synthesis (Fredrick & Ibba, 2010).

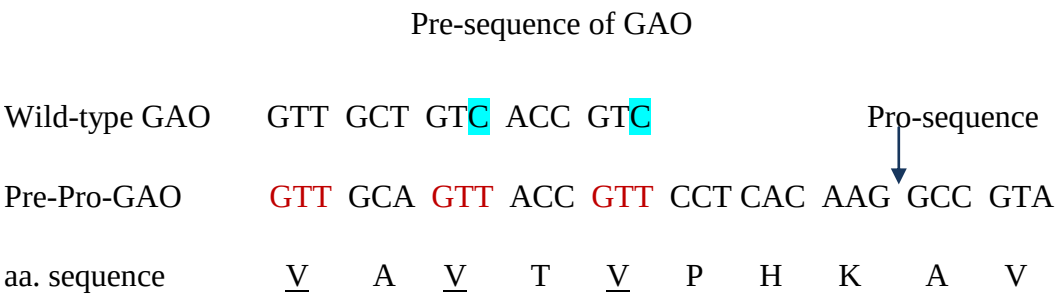


Figure 3.34 N-terminus of wild-type GAO and Pre-Pro-GAO indicating consecutive codon occurance of valine as a result of base substitutions.

3.4.4.2 Purification of ProGON1 and ProGOMN1 by *Strep*-Tactin® Sepharose® Column

Although auto-induction did not highly affected the expression of pro-GAO (the putative bands in Section 3.4.4.2 could not be attributed to pro-GAO due to impurities resulted from purification by *Strep*-Tactin® MacroPrep® Cartridges), the colour development after purification is improved relative to the assay results of IPTG-induced expression (Figure 3.33).

By taking into consideration the advantages of the production of a recombinant protein in *E.coli* by auto-induction growth method and best-result giving *Strep*-Tactin® Sepharose® gravity flow column affinity purification, the combination of these two methods was decided to be used in order to obtain the pre-peptide deleted samples as indicated in Section 2.2.6.3.2 and Section 2.2.7.1.1.

Following expression of pre-pro, pro and mature GAO samples by auto-induction in BL21 Star DE3, the samples were disrupted by sonication and purified by *Strep*-Tactin® Sepharose® gravity flow column. Then, the concentrations of the enzymes were calculated.

Table 3.3 Concentrations of the GAO variants calculated based on Bradford's dye binding assay.

GAO Variants	Concentration
Pre-ProGOMN1	113,57 µg/ml
Pre-ProGONI	26,694 µg/ml
MatGOMN6	582,411 µg/ml
ProGOMN1	16,178 µg/ml
ProGON1	18,119 µg/ml

3.4.4.2.1 Copper Loading and SDS-PAGE Analysis to Investigate the Fate of the Pro-Sequence

The collected eluates were treated with copper as indicated in Section 2.2.8.1 (Baron *et al.*, 1994) in order to assure the cleavage of the pro-peptide. 1 ml of each sample was separated before copper loading and SDS-PAGE was carried out with those samples for detection of ProGON1 and ProGOMN1.

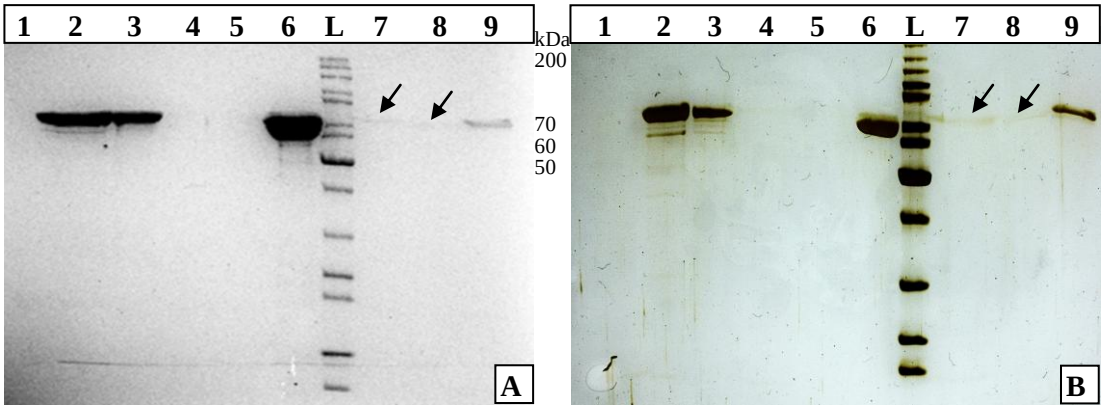


Figure 3.35 SDS-PAGE analysis for detection of 8-aa pre-peptide deleted samples ProGON1 and ProGOMN1 after growth by auto-induction method and *Strep*-Tactin® Sepharose® gravity flow column affinity purification. **(A)** Coomassie staining and **(B)** Silver staining in which lesser amounts were (1:4- 1:7) loaded for lanes 2, 3 and 6. **L,** PageRuler Unstained Protein Ladder **1,** BL21 Star DE3 (-) control **2,** Pre-ProGOMN1; **3,** Pre-ProGOMN1 (+) Cu^{2+}; **4,** ProGOMN1; **5,** ProGOMN1 (+) Cu^{2+}; **6,** MatGOMN6 (+) Cu^{2+}; **7,** ProGON1 ; **8,** ProGON1 (+) Cu^{2+}; **9,** Pre-ProGON1 (+) Cu^{2+}.

This time, fortunately, Pro-GAO was detected for the first time. Yet, the resolution of the gel was not sufficient for the examination of the size differences between the different forms of GAO. Nonetheless, the product of Pro-GAO appeared to be somewhat bigger than the mature GAO, suggesting the lack of pro-peptide cleavage.

3.4.4.2.2 Concentrating the Proteins for Distinguishing the Size Differences

Pro-GAO was concentrated *c.* 10 fold by using Vacuum Concentrator 5301 (Eppendorf) at room temperature. In order to observe the size difference, pro-peptide samples were run for three hours, which is approximately 60-75 minutes longer than the usual, in order to achieve better separation (Blue Vertical 102 Electrophoresis System, SERVA).

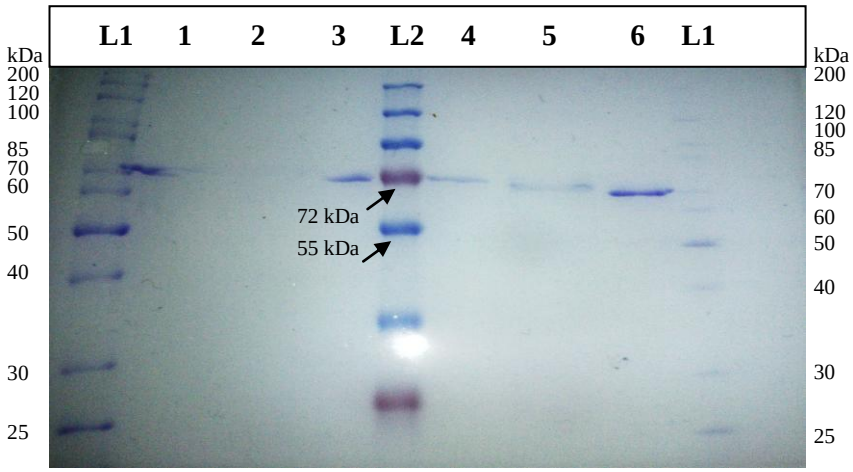


Figure 3.36 Comparison of GAO constructs in order to detect pro-peptide self-processing by SDS-PAGE analysis and Coomassie staining after Cu^{2+} treatment. **L1,** PageRuler Unstained Protein Ladder; **L2**, PageRuler Prestained Protein Ladder; **1,** MatGOMN6; **2,** ProGOMN1; **3,** Pre-ProGOMN1; **4,** Pre-ProGON1; **5,** ProGON1; **6,** MatGOMN6.

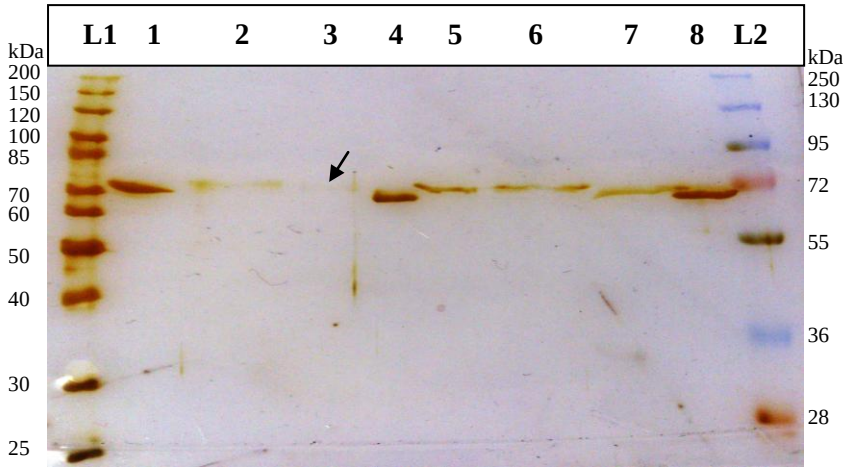


Figure 3.37 Silver staining of the SDS-PAGE of the GAO samples in order to detect ProGOMN1 and comparison of the Cu^{2+} treated and untreated samples for pro-peptide self-cleavage. **L1,** PageRuler Unstained Protein Ladder; **L2**, PageRuler Prestained Protein Ladder; **1,** Pre-ProGOMN1 (+) Cu^{2+}; **2,** ProGOMN1 (-) Cu^{2+}; **3,** ProGOMN1 (+) Cu^{2+}; **4,** MatGOMN6 (+) Cu^{2+}; **5,** Pre-ProGON1 (+) Cu^{2+}; **6,** ProGON1 (-) Cu^{2+}; **7;** ProGON1 (+) Cu^{2+}; **8,** MatGOMN6 (+) Cu^{2+}.

Based on the previous studies (Gençer (Kocuklu), 2005; Deacon, 2008), the expression of Pre-ProGOMN1 was higher than the Pre-ProGON1 which lacks the 6 further mutations. Similarly, here, it was anticipated that the expression of ProGOMN1 to be more than ProGON1 due to the positive cumulative effects of the N-terminal silent mutations and 6

mutations in the mature GAO on expression. However, after pre-peptide deletion, the expression level of ProGOMN1 looks lower than ProGON1.

As indicated in 3.4.4.3, the pro-forms of GAO could not be detected distinctly either by SDS-PAGE or enzyme activity assays (only slight colour change with crude extracts relative to pure samples) with IPTG-induced growth. However, when pProGON1 and pProGOMN1 were expressed by auto-induction, Pro-GAO forms began to appear in SDS-PAGE though the bands were faint and hardly detectable. Moreover, for Pro-GAO, significantly better qualitative activity results were achieved for the first time after purification, by auto-induction. Additionally, impurity problem was overcome by *Strep*-Tactin® Sepharose® gravity flow column. It also eliminates the potential precipitation problem of *Strep*-Tactin® MacroPrep® Cartridges with pro-GAO forms, which is possibly the reason for decreasing the amount of soluble protein fraction in the sample, which is already low. Thus, use of proper induction method for expression and affinity matrix for purification together gave rise to the visualization of the ProGOMN1 and ProGON1, which were not able to be detected before. Furthermore, use of sonication for cell disruption might have a facilitator effect on the detection of the Pro-GAO forms rather than lysis buffer usage.

According to Figure 3.36 and 3.37 it can be seen that there is a distinct size difference between PrePro-GAO, Pro-GAO and mature GAO, indicating strongly that the pro-peptide has not been processed. Thus, these results indicate that the lack of pre-pro-peptide cleavage in *E. coli*, is not due to the N-terminal 8 aa pre-peptide, which raises questions as to the existance of, or mechanism of, pro-peptide autocatalytic cleavage by GAO. Furthermore, the 8 aa long putative pre-peptide appears to be highly essential for the successful expression and production of GAO in *E. coli*.

Table 3.4 Molecular weight estimations of different forms of GAO according to the previous studies (in first two columns) and to this study based on Compute pI/Mw tool which allows the computation of the theoretical isoelectric point and molecular weight. The values in (-) Cu $^{2+}$ columns indicate only the estimated band sizes on SDS-PAGE, not the molecular weights ([1] Rogers *et al.*, 2000, [2] McPherson *et al.*, 1992 and [3] Baron *et al.*,1994).

	(-) *Strep*-tag®		(+)*Strep*-tag®	
ENZYME FORMS	(-) Cu $^{2+}$	(+) Cu $^{2+}$	(-) Cu $^{2+}$	(+) Cu $^{2+}$
Pre-Pro-GAO	N/A	N/A	72,4 kDa	69,4 kDa
Pro-GAO	70,2 kDa[1]	65,5 kDa[1]	71,6 kDa	68,6 kDa
Mature GAO	68,5 kDa[2,3]	65,5 kDa[3]	69,9 kDa	66,9 kDa

It was shown by Baron *et al.*, (1994) that the thioether bond produces a loop that prevents the protein from unfolding after SDS treatment, resulting in a migration of a ~3 kDa lagging band on SDS-PAGE and it was emphasized that this mobility shift is a useful tool for monitoring the thioether bond formation. In order not to attribute the retardation of the bands of ProGON1 and ProGOMN1 to the lack of thioether bond formation, the molecular weight calculations were performed. When Table 3.4 is examined carefully, if the ProGON1 and ProGOMN1 samples did not undergo a thioether bond formation and the pro-peptide was

self-processed, they would have run around 69,9 kDa (MatGOMN6 + 3 kDa) theoretically, above the Pre-ProGON1 and Pre-ProGOMN1 (i.e. the thioether bond was not formed and pro-peptide was cleaved). Furthermore, the motility patterns of pre-peptide deleted variants are in a good consistency with the estimated molecular weights (1,7 kDa for the pro-peptide and additional 0,8 kDa for 8-aa pre-peptide). Therefore, the three forms of the enzyme can be referred as Pre-Pro-GAO, Pro-GAO and mature-GAO.

Most importantly, GAO pro-peptide cleavage cannot be regarded as a self-processing event, when GAO is heterologously expressed in *E.coli.* An additional step making cleavage possible, when produced in its native host absent in *E.coli*, or some reaction or mechanism taking place in *E. coli*, highly likely inhibits the cleavage.

Post-translational modifications of enzymes take place in the ER or the Golgi, like methylation or glycosylation in eukaryotes while they are secreted from the cells. On the other hand, although it was believed that glycosylation is restricted to eukaryotes, glycosylation does take place in prokaryotes, particularly in cell surface- associated and secreted molecules (Sherlock *et al*., 2006). It has been dicovered that *Campylobacter jejuni* has a protein N-glycosylation machinery (Szymanski *et al.*, 1999; Young *et al.*, 2002). These N-glycans take place in a number of important biological recognition processes such as intracellular sorting, cell adhesion, host-pathogen interaction, and immune response (Dwek,1996; Helenius & Aebi, 2001). In *E.coli*, there are two glycoproteins characterised, which are AIDA-I and TibA adhesins associated with diarrhea-causing *E. coli* strains. Addition of the heptose residues by specific heptosyl transferases glycosylates the both proteins (Sherlock *et al*., 2006). Taking into consideration the points explained in Section 1.3.2, without ruling out the glycosylation motif differences between the hosts, self-cleavage of the pro-peptide in GAO in its native host may depend on glycosylation-like mechanism in which glycosylation tends to decrease from the intracellular medium to the extracellular medium. The bound sugars might have been cleaved within the compartments and the enzyme is directed to the next room. The reason for the lack of pro-peptide autoprocessing may be caused by compartmentalization deficiency in which desired modifications occur, (Rehemtulla *et al*., 1992) or recognition problem of unfamiliar motifs in *E.coli.* This is also compatible with the idea that the pro-peptide of GAO does not seem to act as an intramolecular chaperon or inhibitor as explained in Section 1.5.4.1.It may act in cell trafficking or in secretion.

3.4.4.2.3 Significance of the Detection of Pro-GAO

In addition to the SDS-PAGE results in Section 3.4.4.3.1, the activity detection in these samples, indicating the formation of thioether bond, underlines that the motility difference is due to the additional 17-aa pro-peptide. By this finding, the fifth form of the enzyme (Pro-GAO*) has been explored for the first time, in consistency with our previous study in which the Pre-Pro-GAO enzyme was captured (Gençer (Kocuklu), 2005). Pro-Mature enzyme, the previously found form, was renamed as Pre-Pro-GAO (see Figure 3.38).

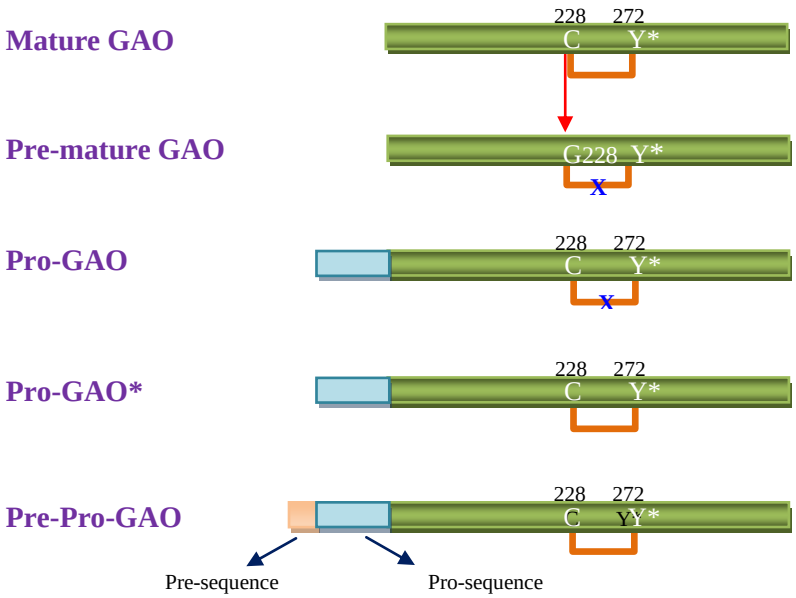


Figure 3.38 Different forms of GAO. * Fifth one explored in this study.

Until now, it has been suggested that pro-peptide cleavage precedes thioether bond formation, since a form which carries both the pro-peptide and the thioether bond was not detected. However, by this study, it is strongly asserted that thioether bond formation is independent of pro-peptide cleavage and the removal of the pro-peptide may take place after the thioether bond formation (see Figure 3.38), in contrast to the idea that gives priority to self-cleavage.

On the other hand, the presence of the pro-peptide at the N-terminus does not prevent the thioether bond formation and enzyme activity totally.

3.4.4.2.4 Comparison of the Activities of the GAO Variants: Mature GAO, Pre-Pro-GAO and Pro-GAO

In this study, it was possible, for the first time, to capture the Pro-GAO form of GAO. In our previous study (Gençer (Kocuklu), 2005), the Pre-Pro-GAO form was isolated. Thus, it was of interest to compare the activities of these three forms, in order to be able to analyse the influence of the presence of these peptide sequences on the activity of the enzyme.

All three forms, namely Mature GAO, Pro-GAO and Pre-Pro-GAO were heterologously produced in *E. coli* BL21 Star DE3, and were isolated using the *Strep*-tag system, as described. Equal quantities of the purified enzyme were used in the ABTS-HRP coupled assay, performed in triplicates. The activities of the GAO variants were determined by measuring the rate of increase in the absorbance at 414 nm (min^{-1}) at 25˚C (Figure 3.39)

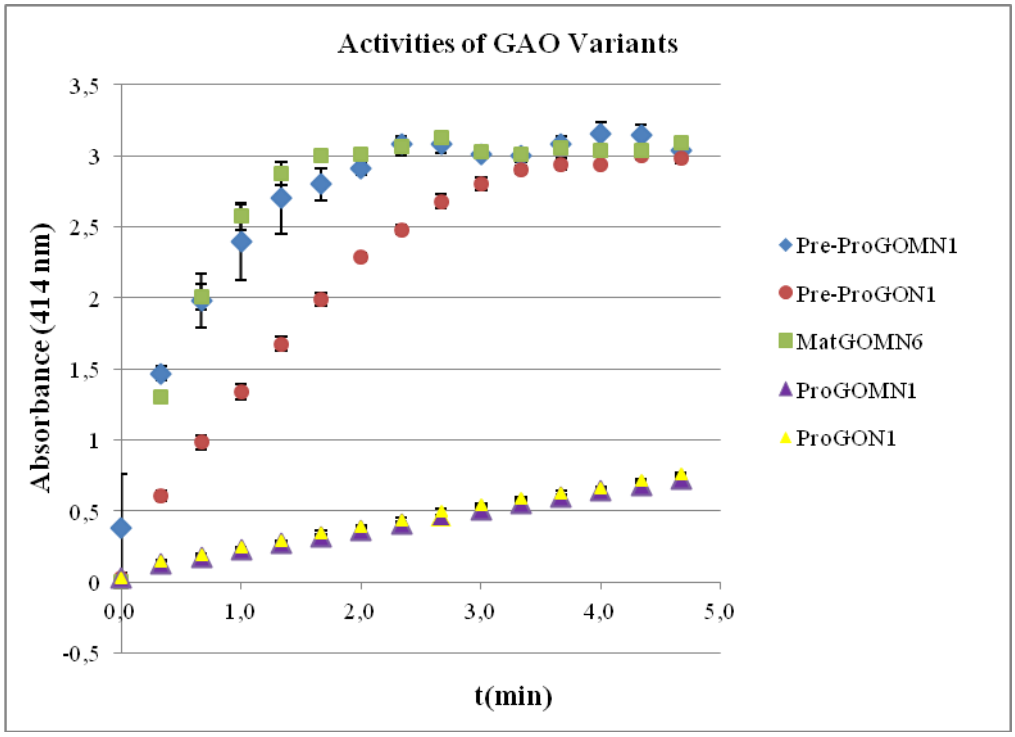


Figure 3.39 Absorbance at 414 nm versus time determined for the GAO variants.

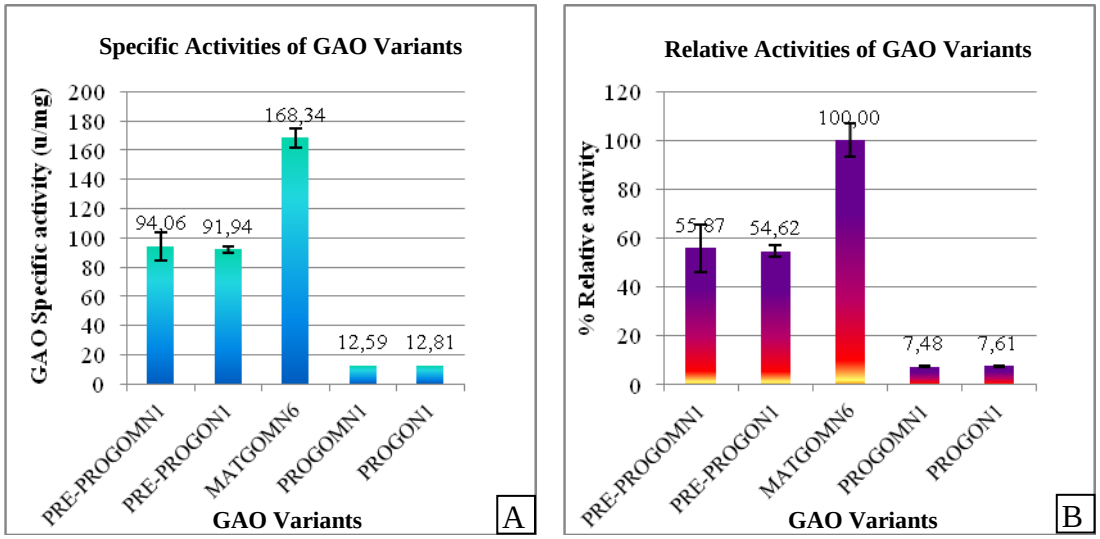


Figure 3.40 Specific **(A)** and relative activities **(B)** of purified GAO variants expressed in E. coli BL21 Star DE3 at 25˚C by auto-induction.

As it is derived from the activity graph of GAO variants above in Figure 3.40, MatGOMN6 shows the highest activity followed by Pre-ProGAO and ProGAO. Specific activity of MatGOMN6 is *c.*1,8 fold higher than Pre-Pro-GAOs and *c.*13 fold higher than Pro-GAOs. Similarly, there is a significant difference between the Pre-Pro-GAO samples and Pro-GAOs. The specific activity of Pre-Pro-GAO is *c.*7,5 fold higher than Pro-GAO. There is not

a significance difference between Pre-ProGOMN1 and Pre-ProGON1, like between ProGOMN1 and ProGON1, which indicates the importance of N-terminal mutations rather than six further mutations within the mature part of the enzyme on expression in pre-pro and pro-forms. Since the activities of the ProGON1 and ProGOMN1 are the lowest ones, it can be suggested that pro-peptide prevents the enzyme from folding into its native conformation.

Based on these results, a possible effect of the 8 aa pre-peptide on the formation of the native three-dimensional conformation of the enzyme (in the presence of pro-peptide), specifically the active site, cannot be ruled out.

The significantly low level of activity of the purified pro-GAO, may explain the low level of production of the enzyme, as misfolded proteins are likely to undergo rapid degradation in the cell.

Indeed, many abnormal proteins are degraded in *E.coli*. Such abnormal proteins include incompletely or incorrectly synthesized proteins, due to errors in transcription, translation or synthesis, misassembled or misfolded proteins, improperly transported proteins, and heat-denatured proteins (Gottesman & Maurizi, 1992). Briefly, they are the proteins that do not have a native conformation. Recombinant proteins may also fall under this catagory since they are regarded as unwanted by the cell (Rozkov, 2001).

Furthermore, the N-end rule (Bachmair *et al.*,1986), which relates to the in vivo half-life of a protein to the identity of its N-terminal residue, influences the stabilty of a protein, and this is not restricted to the eukaryotes (Tobias *et al.*,1991) and is present both in *S. cerevisiae* and *E.coli* (Varshavsky, 1997). While N-end recognition in eukaryotes is mediated through the ubiquitin system, it is recognized by the Clp protease in bacteria, which do not have a ubiquitin system (Gonda *et al*,1989; Tobias *et al*,1991; Varshavsky,1992). Although *lon* and *ompT* proteases are absent in *E. coli* BL21 Star DE3, the Clp protease still exists in the system (Rozkov ,2001).

N-terminal residues of Pre-Pro-GAO or Pro-GAO, are valine (V) and alanine (A), respectively, and both are categorized as stabilizing residues with an half-life of >10h. Although the N-teminus of Pro-GAO does not seem to be unstabilizing, since the N-end-recognizing system might be involved in screening proteins that have been improperly processed at the N-terminus (Bachmair *et al.*,1986), and thus N-terminal of Pro-GAO might be recognized as an unusual end since it is not familiar to *E.coli* and might undergo degradation.

Moreover, some motifs on proteins play an important role as a potential degradation signal. These protease recognition regions are inaccessible to the proteases unless the protein is partially unfolded or damaged by mutation and any other stress. On the other hand, protein-protein interaction is one of the most important issues that affect the rate of the degradation of a protein, which may serve directly or indirectly. A motif may improve protein-protein

interactions, and thus make degradation signals less accessible, whereas the motif itself, which is not accessible within a properly folded protein under certain conditions may be

recognized by the protease as degradation signals (Gottesman & Maurizi, 1992), like in the case, in which pre-peptide might prevent partial degradation by interacting with the pro-peptide which might involve a potential degradation signal.

Other possibilities to explain low level of expression may also include those related to mRNA secondary structure, that affects the efficiency of translation (Deacon, 2011).Considering that the Pre-ProGOMN1 and Pre-ProGON1 were carrying a pre-peptide with silent mutations at its N-terminus causing no detrimental effects in the final protein product, it may also be speculated that the deletion of this pre-peptide may likely show its impact at the post-transcriptional level. Since the N-terminal silent mutations introduced into the *pre-pro-gao* was A and T rich (Deacon, 2008), removal of pre-sequence may result in reduced amount of adenine and thymine, leading to the secondary structure formation of RNA, which may impair the translation initiation and efficiency (Laursen *et al.*, 2002).An 8-aa deletion at the N-terminus may result in some structural changes that influence the translational efficiency.

Based on the results, autocatalytic pro-peptide cleavage of GAO is likely to be independent of the primary amino acid sequence of the enzyme. This is supported by the following findings: First, mutations at the cleavage site do not influence processing of the pro-peptide (Ögel., 1993; Gençer (Kocuklu), 2005). Second, production of the enzyme with its native amino acid sequence in *E. coli* prevents autocatalytic cleavage; i.e. neither the pre-peptide, nor the pro-peptide are cleaved in *E. coli* despite of the presence of copper and oxygen. Thus, post-translational modifications that occur in the fungus, and that are absent in *E. coli*, may be required for autocatalytic cleavage. An alternative explanation could be lack of autocatalytic cleavage due to misfolding. Namely, Pre-Pro-GAO was not cleaved may be due to the pre-peptide and the Pro-GAO was not cleaved due to misfolding and lack of proper localization of copper. Nonetheless, the latter possibility seems less likely, because, Pre-Pro-GAO is almost as active as the Mature GAO, (*c*.1,8 fold decrease) suggesting lack of autocatalytic cleavage despite of proper folding. However, there is still *c*.1.8 fold difference in activity, and this possibility, therefore, cannot be ruled out.

Fungal mono- and dibasic processing of pro-peptides is a known phenomenon (Jalving, 2005). It can be suggested that the pre-peptide of GAO is cleaved by a fungal mono- or dibasic processing protease, which is absent in *E. coli*, thus explaining lack of processing of the pre-peptide. However, it is extremely interesting that the Pre-Pro-GAO is almost as active as the Mature GAO, but the Pro-GAO, in contrast, is nearly inactive. This finding clearly indicates that the pre-peptide is required for the formation of the active conformation.

In the fungal system, transportation to the extracellular medium requires passage through organelles like the ER and the golgi, which are absent in prokaryotes. Thus, it may be hypothesized that the pre-peptide at the N-terminus is used to initiate the folding process and is then cleaved by a protease, followed by the autocatalytic cleavage of the pro-peptide during the remaining steps of transportation. Thus, it may be argued that either, some post-translational modifications during transport or the formation of the proper conformation by copper binding are essential for the autocatalytic cleavage of the pro-peptide of GAO.

PART II

Domain Deletion Studies of *Fusarium graminearum* NRRL 2903 GAO in *E.coli* for Future Biosensor Applications

3.5 Split-GAO Approach for Use in Biosensor Applications

GAO consists of three predominately β-sheet domains, namely Domain I, II and III (Ito *et al.*, 1991,1994). Domain I contains a carbohydrate binding site and potentially forms hydrogen bonds with its substrates (Firbank, 2002). Domain II is the largest and it contains the active site providing Tyr-272, Tyr-495, and His-496 ligands to the copper. Hydrophobic interactions between the β-strands ('Velcro' fastening, see Section 1.3.3) provide rigidity to the enzyme. One of the β-ribbons passes through the Domain III and presents the fourth protein ligand His-581 to the copper. Domain III presumed to have a structural key role of stabilising domain II (Deacon, 2008). The detailed information about domains is given in Section 1.3.3.

In a biosensor system to be used, two antibody-mimetic proteins are used to bind the target molecule and each is coupled to an electrode surface and one-half of a split-enzyme. Since GAO activity will be utilized in signal amplification, it was planned to generate a split GAO system, which consists of two halves of the enzyme. Signal production is based on the association of the two-halves of the split-GAO as soon as the target biomolecule is bound to the antibody-mimetic proteins. At the end of the multiple turnover events, products of the catalytic activity are measured electrically, which means the amplification of the initial binding signal (Section 1.2.5.4.2 and 1.2.5.4.3).

In order to form such a split system, domain I/II (DI/II) deletion and domain III (DIII) deletions were planned on MatGOMN6 carrying constructs named pET28cStrepTEVMN6 and pET28cHisTEVMN6 for prospective biosensor studies as a part of a big project at University of Leeds. Domain I/II deleted construct (GAO with only domain III) was designed to start with Serine536 (Ser536) and domain III deleted (GAO with domains I and II) construct was designed to end with Asparagine531 (Asn531) (Figure 3.41).

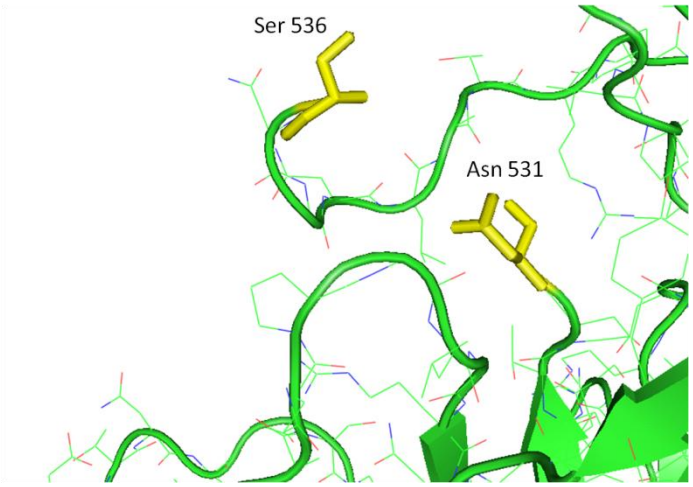


Figure 3.41 3-D structural representation of the residues in domain deletion studies of GAO

Figure 3.42 Amino acid sequence of MatGOMN6 carrying *Strep*-tag II and TEV protease cleavage site on pET28cStrepTEVMN6. The residues used to split the enzyme between DII and DIII is highlighted with green and yellow.

As it is shown in Figure 3.41 and Figure 3.42 the sites for deleting the domains occur within a surface loop region that joins DII and DIII. With regard to the missing residues in this loop, YLYD, There is no evidence for any biological function other than as a domain linker for this region which is on the opposite side of the enzyme from the active site and the substrate recognition region. Since it is tried to split GAO into two parts, the presence of these residues is of no consequence for the folding or assembly of the protein (McPherson, personal communication, 2011).

3.6 Generation of His-tag carrying GAO constructs by Site-Directed Mutagenesis

In order to perform domain deletion studies on both *Strep*-tag and His-tag carrying constructs to facilitate the further purification trials, *Strep*-tag / His-tag exchange was carried out at first.

After transformation of pET28cStrepTEVMN6 into *E.coli* XL1 Blue cells and incubating on LB agar plates containing proper antibiotic (Kan), plasmid isolation was carried out with QIAprep Spin Miniprep Kit (Qiagen) (Figure 3.43).

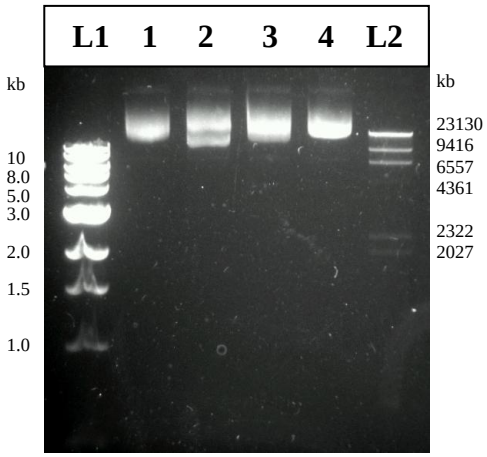


Figure 3.43 Plasmid isolation results of pET28cStrepTEVMN6. **L1**, 1kb ladder (NEB); **1-4,** pET28cStrepTEVMN6; **L2,** λ DNA *Hin*d III Marker.

Restriction enzyme digestions were made using *Cla*I and also *Pst*I seperately as a single cutter to test the main vector, pET28cStrepTEVMN6 (7241 bp) (Figure 3.44 and 3.45).Single enzyme restriction digestion with *Cla*I did not give a good result possibly due to the degradation of the enzyme.

Double restriction enzyme digestions with *Nco*I & *Nhe*I were also applied. According to the plasmid restriction digests the expected band sizes for *Nco*I & *Nhe*I were; 5898 bp, 747 bp and 596 bp (Figure 3.44).

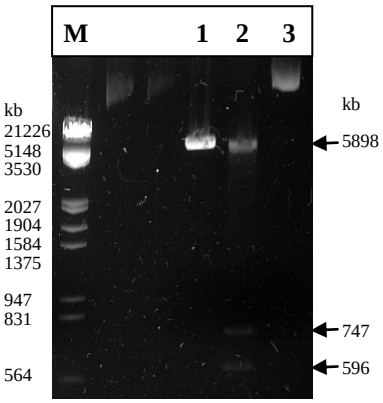


Figure 3.44 *Pst*I single restriction enzyme digest and *Nco*I & *Nhe*I double restriction enzyme digest results of pET28cStrepTEVMN6. **M,** λ DNA *Eco*RI+*Hin*d III Marker (Promega) **1**, pET28cStrepTEVMN6 *Pst*I digest; **2**, pET28cStrepTEVMN6 *Nco*I & *Nhe*I double digest **3**, pET28cStrepTEVMN6.

3.6.1 Design of *Strep*-tag / His-tag Exchange Primers

Since the expected sizes were detected, *Strep*-tag / His-tag exchange trials were carried out on pET28cStrepTEVMN6 construct by *QuikChange® Site-Directed Mutagenesis* Kit (Stratagene) as indicated in Section 2.2.3.1.2.with the primer pair given below:

GO Histag_F 5’catcatcaccatcaccaccatgaaaacctgtacttccagg 3’ 40bp
GO Histag_R 5’ggtgatggtgatgatggtgacccatggtatatctccttc 3’ 39bp

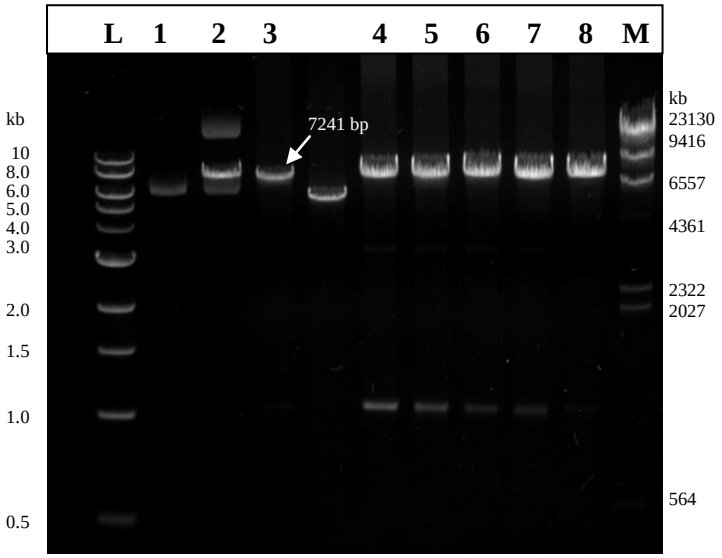


Figure 3.45 *QuikChange® Site-Directed Mutagenesis* for His-tag substitution for *Strep*-tag **L,** 1kb ladder (NEB); **1**, pET28cStrepTEVMN6; **2**, *Cla*I digest; **3**, pET28cHisTEVMN6 55,4˚C (50 µl reaction mixture); **4-8,** pET28cHisTEVMN6 53,1˚C, 55,4˚C, 57,5˚C, 59,7˚C, 61,2˚C (10 µl reaction mixture each); **M,** λDNA *Hin*d III Marker.

*Dpn*I digestion was done with pET28cHisTEVMN6 (55,4°C, 50 µl reaction mixture) at 37 °C for 1 hour to digest parental DNA.

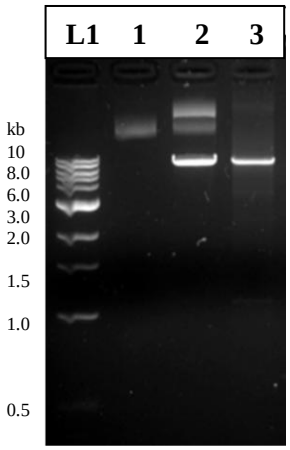


Figure 3.46 His-tag substitution on pET28cStrepTEVMN6 after *Dpn*I digestion **L1,** 1kb ladder (NEB); **1,** pET28cStrepTEVMN6; **2,** *Cla*I digest of pET28cStrepTEVMN6; **3,** pET28cHisTEVMN6 *Dpn*I digestion.

3.6.2 Determination of the His-tag Substitution by DNA Sequencing

Following transformation of the putative pET28cHisTEVMN6 samples into *E.coli* XL1-Blue cells, plasmid isolation was carried out. 3 of 4 samples (pET28cHisTEVMN6_2,3,4) submitted to DNA sequencing were carrying His-tag substitution. Domain deletion studies for His-tag construct were conducted on fully sequenced pET28cHisTEVMN6_3 construct as a template (Appendix D).

3.7 Generation of Domain Deletions both on *Strep*-tag and His-tag Constructs

3.7.1 Design of Mutational Primers for Domain I/II Deletion and Domain III Deletion

Domain deletions were designed on both constructs carrying *Strep*-tag and His-tag, which are pET28cStrepTEVMN6 and pET28cHisTEVMN6, respectively by *QuikChange® Site-Directed Mutagenesis* as indicated in Section 2.2.3.1.2 with the primer pairs given below (underlined residues indicate the complementary parts):

DOMAIN I/II deletion primers

DI/IIdel_Ser536_F	5’ agcaacggcaatctcgcgac 3’	20bp
DI/IIdel_R	5’ gtcgcgagattgccgttgctaccctggaagtacaggttttc 3’	41bp

DOMAIN III deletion primers

D3del_F	5’ tgatccgaattcgagctccg 3’	20bp
D3del_R	5’ cggagctcgaattcggatcagtttggcgtaaagatttgcgc 3’	41bp

3.7.2 Domain Deletion Studies on pET28cStrepTEVMN6 by Site-Directed Mutagenesis

Domain deletion studies were carried out on both constructs carrying *Strep*-tag and His-tag, which are pET28cStrepTEVMN6 (7241 bp) and pET28cHisTEVMN6 (7241 bp), respectively by *QuikChange® Site-Directed Mutagenesis* as indicated in Section 2.2.3.1.2 with the primer pairs given above. The expected band sizes of the constructs at the end of the deletions were;

Domain I/II deletion (-1605 bp): 5636 bp
Domain III deletion (-324 bp) : 6917 bp

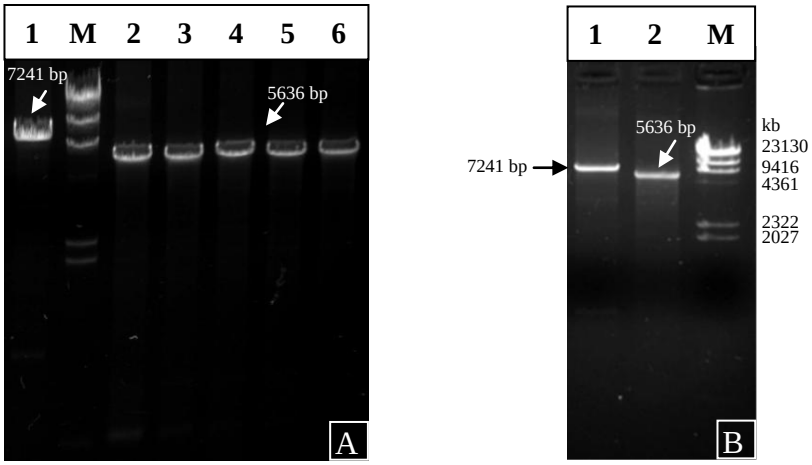


Figure 3.47 Domain I/II deletion on pETStrepTEVMN6 by *QuikChange® Site-Directed Mutagenesis* **(A)** and results after *Dpn*I digestion **(B)**. **M,** λ DNA *Hin*d III Marker **A1**, pET28cHisTEVMN6; **A2-6,** pET28cStrepTEVMN6DI/IIdel 53,1˚C, 55,4˚C, 57,5˚C, 59,7˚C, 61,2˚C (10 µl reaction mixture each); **B1,** pET28cHisTEVMN6 *Dpn*I digest; **B2,** pET28cStrepTEVMN6DI/IIdel *Dpn*I digest.

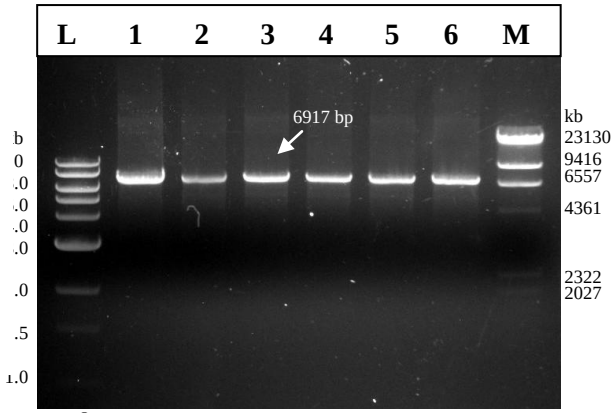


Figure 3.48 *QuikChange® Site-Directed Mutagenesis* results of domain III deletion on pET28cStrepTEVMN6 construct. **L,** 1kb ladder (NEB); **1,** pET28cStrepTEVMN6DIIIdel 55,4˚C; **2-6,** pET28cStrepTEVMN6DIIIdel 53,1 ˚C, 55,4 ˚C, 57,5 ˚C, 59,7 ˚C, 61,2 ˚C (10 µl reaction mixture each) ; **M,** λ DNA *Hin*d III Marker

With the domain deletion mutagenesis primers given in Section 3.7.1 the expected sizes of bands were obtained by using pET28cStrepTEVMN6 as a template. Domain I/II and domain III deleted mature GAO (MatGOMN6) seemed to be attained, which would be confirmed by DNA sequencing.

3.7.3 Domain Deletion Studies on pET28cHisTEVMN6 by Site-Directed Mutagenesis

Domain I/II and domain III deletions were carried out on pET28cHisTEVMN6 construct obtained after *Strep*-tag and His-tag substitution trials.

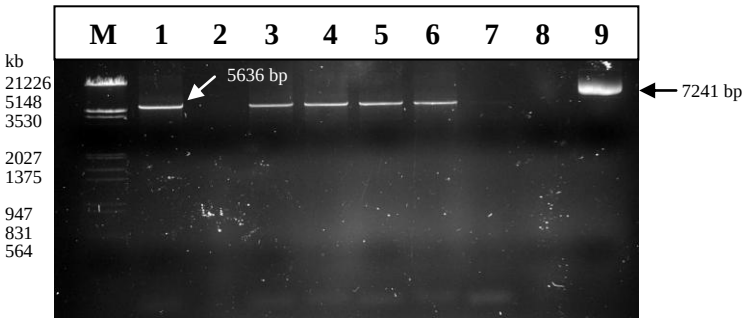


Figure 3.49 Domain I/II deletion results of pET28cHisTEVMN6 construct by *QuikChange*® *Site-Directed Mutagenesis* **M,** λ DNA *Hin*d III+*Eco*RI Marker (Promega); **1,** pET28cHisTEVMN6DI/IIdel 55,4˚C (50 µl); **2,** pET28cHisTEVMN6 DI/IIdel 55,4˚C (-) control **3-7,** pET28cHisTEVMN6DI/IIdel 53,1 ˚C, 55,4 ˚C, 57,5 ˚C, 59,7 ˚C, 61,2 ˚C (10 µl reaction reaction mixture each); **8,** pET28cHisTEVMN6DI/IIdel (-) control 55,4 ˚C (10 µl); **9,** pET28cHisTEVMN6.

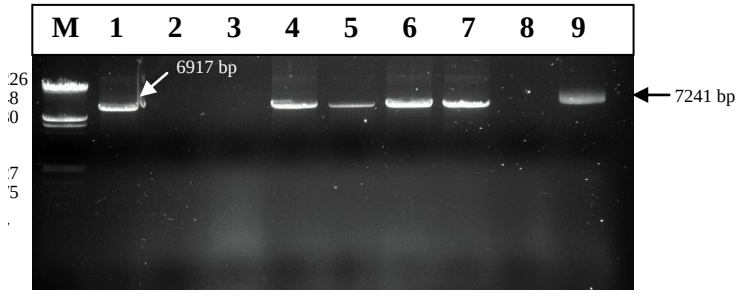


Figure 3.50 Domain III deletion results of pET28cHisTEVMN6 construct by *QuikChange*® *Site-Directed Mutagenesis* **M,** λ DNA *Hin*d III+*Eco*RI Marker (Promega); **1,** pET28cHisTEVMN6DIIIdel 55,4˚C (50 µl); **2,** pET28cHisTEVMN6 DIIIdel 55,4˚C (-) control **3-7,** pET28cHisTEVMN6DIIIdel 53,1 ˚C, 55,4 ˚C, 57,5 ˚C, 59,7 ˚C, 61,2 ˚C (10 µl reaction reaction mixture each); **8,** pET28cHisTEVMN6DIIIdel (-) control 55,4 ˚C (10 µl); **9,** pET28cHisTEVMN6.

3.7.4 Determination of the Domain Deletions by DNA Sequencing

Following the *Dpn*I digestion of the proper constructs, they were transformed into *E.coli* XL1-Blue cells. Plasmid isolation was carried out both for pET28cStrepTEVMN6 and pET28cHisTEVMN6 variants (Figure 3.51, all results are not given) and samples were submitted to DNA sequencing (Beckman Coulter Genomics, Essex, UK). DNA sequences of some of the right clones are given in Appendix D as representatives.

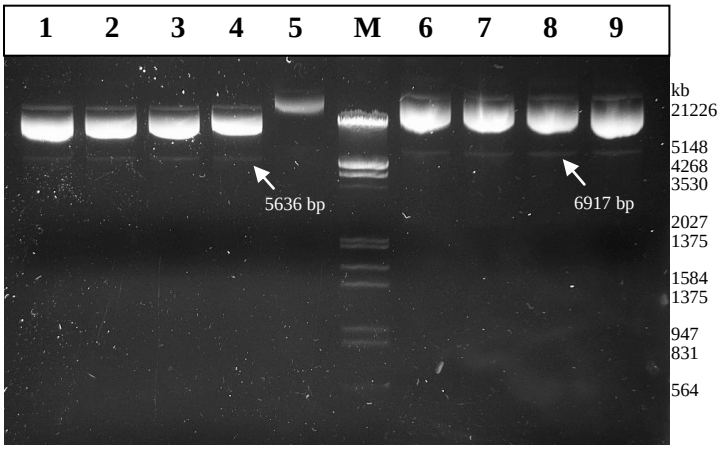


Figure 3.51 Plasmid isolation results following *QuikChange® Site-Directed Mutagenesis* for domain I/II and domain III deletions on pET28cHisTEVMN6 construct **M,** λ DNA *Hin*d III + *Eco*RI Marker (Promega); **1-4** pET28cHisTEVMN6DI/IIdel_1-4; **5,**pET28cHisTEVMN6 **6-9,** pET28cHisTEVMN6DIIIdel_1-4.

3.8 Heterologous Expression of Mutant GAO Constructs

Appropriate constructs carrying desired deletions were chosen according to the DNA sequencing results.

Then these constructs were transformed into BL21 Star DE3 cells for protein expression by auto-induction. As it is described in Section 2.2.6.3.1 and 3.4.4.1; samples were taken from cultures by regular time intervals and crude extracts were analyzed by SDS-PAGE for detecting the expression of the relevant mutants.

The products of pET28cStrepTEVMN6 and pET28cHisTEVMN6 will be named as StrepGAO and HisGAO hereafter.

3.8.1 SDS-PAGE Analysis of StrepGAO and HisGAO Variants

Samples were taken at regular time intervals and crude extracts were analyzed by SDS-PAGE for detection of expression and whether the expressed protein is soluble or insoluble.

3.8.1.1 *Strep*-tag / His-Tag Substitution

3 positive samples of His-tag substitution (Section 3.6.2) were analyzed by SDS-PAGE following expression in BL21 Star DE3.

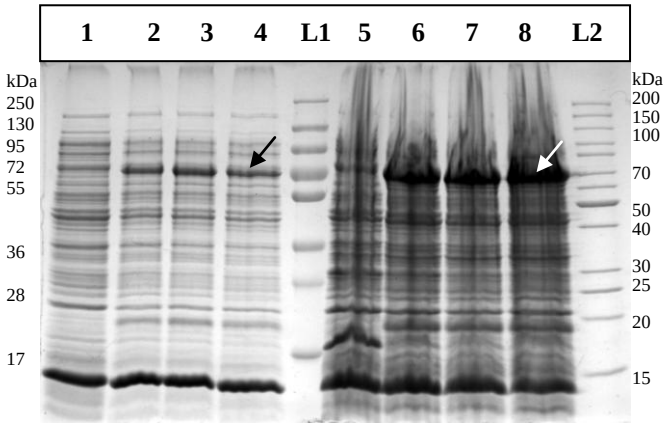


Figure 3.52 Expression detection for His-tag substitution with 48-hour soluble and insoluble fractions **L1,** PageRuler Prestained Protein Ladder (Fermentas) ; **L2,** PageRuler Unstained Protein Ladder; **Soluble parts 1,** BL21 Star DE3(-) control; **2,** HisGAO_2; **3,** HisGAO_3; **4,** HisGAO_4; **Insoluble parts 5-8,** samples in the same order as 48-hour soluble samples.

As it is seen in Figure 3.52 bands of expected size were detected around 70 kDa. Most of the expressed protein is in the insoluble part in all samples, however, there is detectible sufficient amount of enzyme in the soluble part.

3.8.1.2 Domain Deletion Mutants of StrepGAO and HisGAO

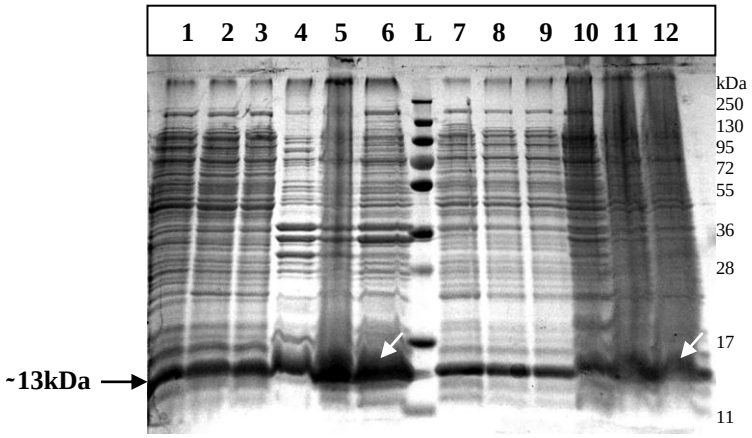


Figure 3.53 Analysis of 24 and 48-hour soluble and insoluble fractions for detection of StrepGAODI/IIdel expression **L,** PageRuler Prestained Protein Ladder; **Soluble parts (24h) 1,** BL21 Star DE3 (-) control; **2,** StrepGAO DI/IIdel _2; **3,** StrepGAO DI/IIdel _3; **Insoluble parts (24h) 4,** BL21 Star DE3 (-) control; **5,** StrepGAO DI/IIdel _2; **6,** StrepGAO DI/IIdel _3; **7-12,** 48-hour samples in the same order as 24 hour samples.

18 hour fractions of the samples are not given since detectable expression was not obtained.

Theoretical average mass of StrepGAODI/IIdel is 13112.72Da and that of StrepGAODIIIdel is 59054.25Da. Similarly, theoretical average masses of HisGAODI/IIdel and HisGAODIIIdel is 13169.71Da and 59111.23Da, respectively.

(http://web.expasy.org/peptide_mass/)

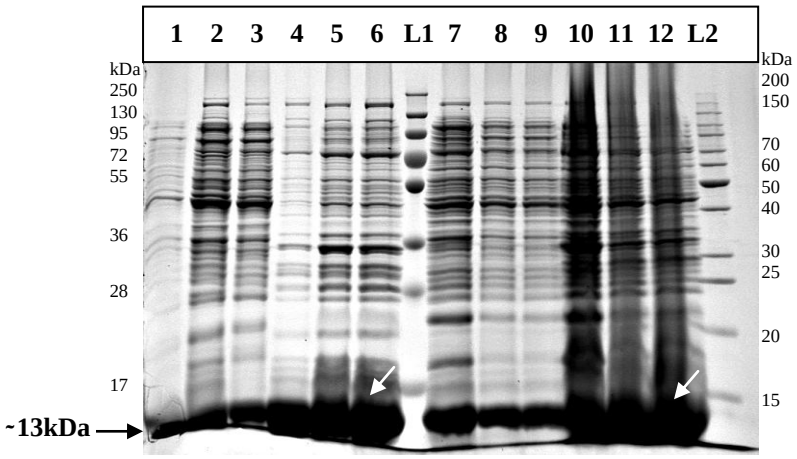


Figure 3.54 Expression detection of HisGAODI/IIdel with 24 and 48-hour soluble and insoluble fractions **L1,** PageRuler Prestained Protein Ladder; **L2,** PageRuler Unstained Protein Ladder; **Soluble parts (24h) 1,** BL21 Star DE3(-) control; **2,** HisGAODI/IIdel_2; **3,** HisGAODI/IIdel_3; **Insoluble parts (24h) 4,** BL21 Star DE3 (-) control; **5,** HisGAODI/IIdel _2; **6,** HisGAODI/IIdel _3; **7-12,** 48-hour samples in the same order as 24 hour samples.

Among the pET28cHisTEVMN6DI/IIdel and pET28cStrepTEVMN6DI/IIdel constructs giving the right DNA sequence for deletion of domain I/II, two of them were used in the expression studies. According to the SDS-PAGE results in Figure 3.53 and 3.54, there is high amount of protein around 13 kDa, which is the expected size of domain III of GAO. There seems to be other proteins of similar size when compared with the negative control, BL21 Star DE3. However, there is a distinct difference especially in the insoluble part of the crude extract of 24 hour samples.

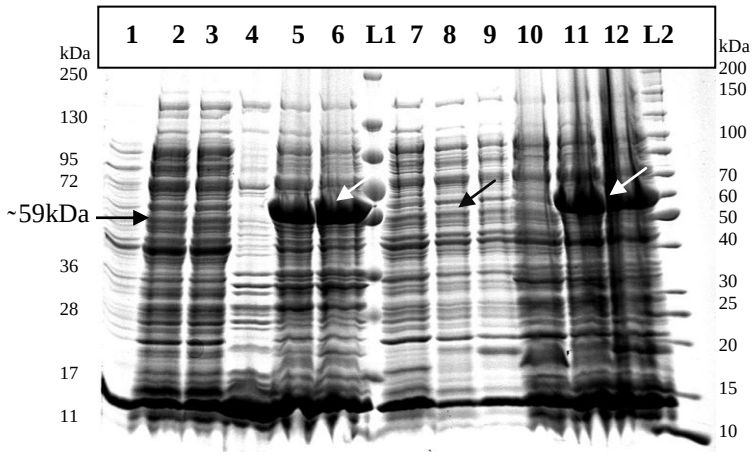


Figure 3.55 Analysis of 24 and 48-hour soluble and insoluble fractions for detection of StrepGAODIIIdel and HisGAODIIIdel expression **L1,** PageRuler Prestained Protein Ladder; **L2,** PageRuler Unstained Protein Ladder; **Soluble parts (24h) 1,** BL21 Star DE3(-) control; **2,** StrepGAODIIIdel; **3,** HisGAODIIIdel; **Insoluble parts (24h) 4,** BL21 Star DE3 (-) control; **5,** StrepGAODIIIdel; **6,** HisGAODIIIdel; **7-12,** 48-hour samples in the same order as 24 hour samples.

According to the SDS-PAGE results of domain III deletion trials on both constructs, almost all of the expressed protein is trapped as inclusion bodies in the insoluble part. Possibly low amount of enzyme is present in the soluble part which is not detected in the crude extract.

3.8.2 Confirmation of GAO Mutations by Western Blotting

3.8.2.1 Immunodetection of *Strep*-tag / His-tag Exchange

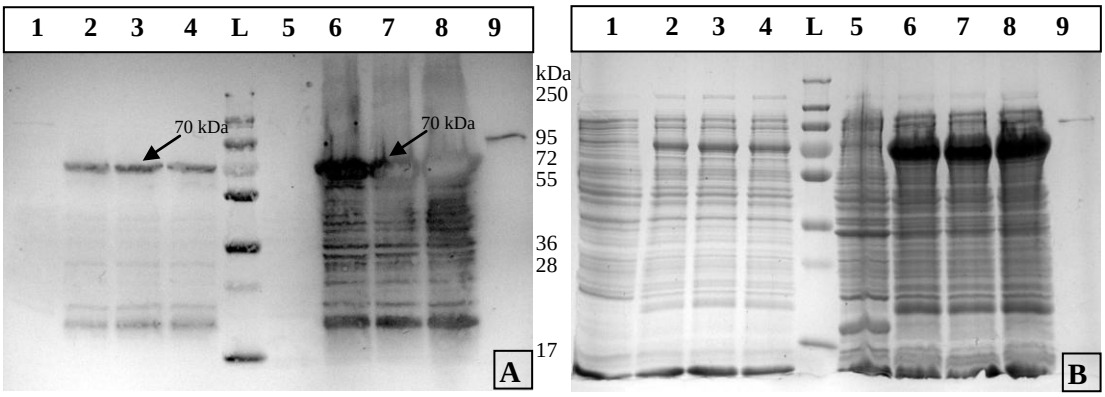


Figure 3.56 Western blot analysis of 48-hour soluble and insoluble fractions of HisGAO for His-tag substitution **(A)** and simultaneously run SDS-PAGE of the same samples **(B) L,** PageRuler Prestained Protein Ladder; **Soluble parts 1,** BL21 Star DE3 (-) control; **2,** HisGAO_2; **3,** HisGAO_3; **4,** HisGAO_4; **Insoluble parts 5,** BL21 Star DE3 (-) control; **6,** HisGAO_2; **7,** HisGAO_3; **8,** HisGAO_4; **9,** His-tag carrying protein, (+) control

Immunodetection of the His-tag was carried out by monoclonal anti-polyhistidine peroxidase conjugate (Sigma).

As it is derived from Figure 3.56 all the positive clones carrying His-tag are expressed, however most of the expressed protein is in insoluble part and some degradation products are also detected.

3.8.2.2 Immunodetection of Domain I/II and Domain III Deletions

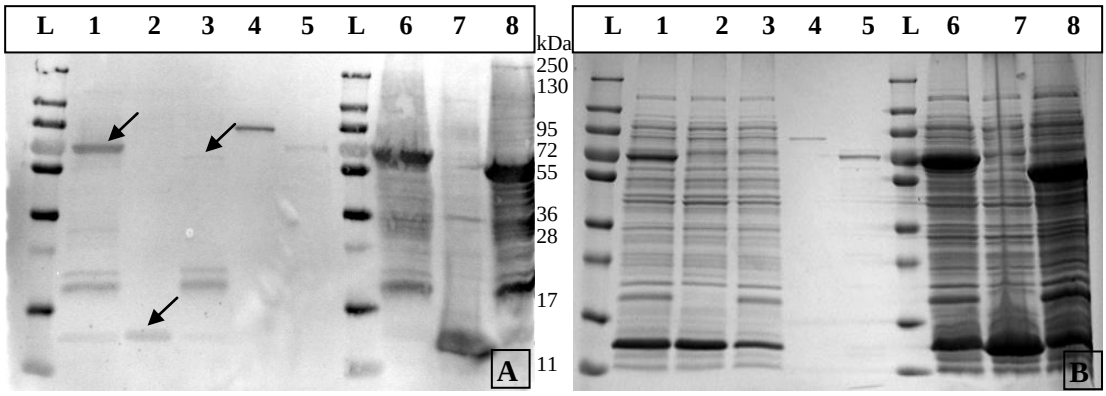


Figure 3.57 Western blotting results of HisGAO domain deletions with 48-hour soluble and insoluble fractions (A) and simultaneously run SDS-PAGE of the same samples (B) L, PageRuler Prestained Protein Ladder; Soluble parts 1, HisGAO; 2, HisGAODI/IIdel; 3, HisGAODIIIdel; 4, His-tag carrying protein, (+) control; 5, Strep-Tag carrying Pre-ProGOMN1 (-) control Insoluble parts 6, HisGAO; 7, HisGAODI/IIdel; 8, HisGAODIIIdel.

As expected theoretically, molecular weights of HisGAODI/IIdel and HisGAODIIIdel are around 13 kDa and 59 kDa, respectively. However, as detected by western blotting most of the expressed protein is trapped as inclusion bodies in the insoluble part .When soluble and insoluble fractions of the entire protein (HisGAO) is compared with those of the domain I/II and domain III deleted variants, it can be detected that there is more soluble protein in HisGAO. It can also be inferred that when the protein loses its integrity the solubility of the protein decreases.

In this part of the study, generation of preliminary split-GAO system constituents was aimed on a protein engineering basis. Domain deletions of galactose oxidase were performed by site-directed mutagenesis in order to split GAO into two parts. According to the SDS-PAGE and western blotting results it can be suggested that each part of GAO is expressed, however it looks as if entrapped in inclusion bodies. Before the purification studies for biosensor use, these results can be improved by changing different parameters such as temperature, shaking speed and harvest time. Additionally inclusion body solubilisation and refolding techniques might also be required to make expressed protein soluble for utilization.

This split-GAO system is a biosensor candidate to be used in different areas such as diagnostics, environment and security in the future, by utilizing its enzymatic properties for generating signals. It provides a simple approach to the production of a split redox enzyme, in contrast to glucose oxidase which is a much more complex system. This technology enables rapid *in vitro* detection of biomarkers at low cost and with no need for complicated assays with a high sensitivity.

CHAPTER 4

CONCLUSION

Heterologous expression of GAO in *E.coli* prevents autocatalytic cleavage of the pro-peptide, despite of the presence of Cu^{2+} and dioxygen.

The presence or absence of the pre-peptide does not influence lack of autocatalytic cleavage in *E.coli.*

In the absence of the pre-peptide, the resulting Pro-GAO is ~ 90 %less active than the mature GAO suggesting that the pre-peptide is necessary for the formation of the active conformation of GAO.

Pro-GAO has low activity, suggesting misfolding.

Autocatalytic cleavage of GAO pro-peptide is likely to be independent of the primary amino acid sequence of the enzyme.

Pre-Pro-GAO is highly active as compared to Pro-GAO, suggesting that the formation of the thioether bond is independent of the cleavage of the pro-peptide.

According to the SDS-PAGE data and the deduced molecular weight (68,6 kDa) the thioether bond is present in Pro-GAO, thus a new form of galactose oxidase is able to be observed for the first time.

Lack of autocatalytic cleavage in GAO expressed in *E.coli* suggests that autocatalytic cleavage is not only dependent on the presence of Cu^{2+} and dioxygen, but has additional requirements, such as the post-translational modifications of the enzyme.

Auto-induction is a better approach than IPTG induction, in expressing Pro-GAO in *E.coli.*

CHAPTER 5

RECOMMENDATIONS

In order obtain high amounts of pro-enzyme further optimization studies can be carried out for characterization studies. SUMO Fusion technology which is chosen for difficult-to-express proteins can also be used for the expression of Pro-GAO as a fusion protein in order to increase the expression.

As a further work; together with a strong assumption of GAO is an autoprocessing enzyme and *E.coli* is responsible for this unprocessing, alternative hosts can be used in which pro-peptide can be cleaved (similar to the native) and various point mutations can be carried out at/around the cleavage site as new trials for figuring out the self-cleavage mechanism of the pro-peptide.

On the other hand, crystallography studies can also be performed for the Pro-GAO obtained by heterologous expression in *E.coli*, even after aerobic incubation of copper *in vitro* with/without a substrate. Importantly, up to now the positions of the pro-peptide residues have only been able to be detected in copper-free environment (Firbank *et al*., 2001) whereas the mature form has been obtained in the presence of copper. *E.coli* produced Pro-GAO may make it possible to compare the pro- and mature forms under the same conditions with an accompanying thioether bond, removing the need to deal with laborious copper-free experiments for detection of the pro-sequence. This may allow the direct visualization of the copper, pro-sequence, thioether bond and their positions to each other and the conformational differences between the mature GAO and the Pro-GAO.Additionally, post-translational modification differences such as acetylation, glycosylation or phosphorylation can be detected between the enzyme types (pre-pro, pro and mature) expressed in *E.coli* and *Aspergillus nidulans* or its native host, *Fusarium graminearum*, which may inhibit self-processing.

For developing biosensor applications, in order to try to get rid of the inclusion bodies; the split GAO system can be developed by changing parameters of the experiments, i.e temperature, rpm, etc. and also purification, solubilization and refolding studies can be carried out.

Alternatively, instead of a system generating two different enzyme fragments of a protein, whole protein can be produced with a cleavage site inserted recognizing a proper protease. Those will help to use GAO in different applications as a biosensor candidate (McPherson, personal communication, 2011).

REFERENCES

Abdulla, Y.H.(2001). *BioEssays,* **23,** 456-462.

Adanyi, N., Szabo,E.E. & Varadi, M. (1999). *Eur. Food. Res. Technol.* **209,** 220-226.

Agilent Technologies, Inc. (2009). *Pfu* Ultra High Fidelity DNA Polymerase Instruction Manual, USA.

Aisaka, K.; Terada, O. (1981). *Agric. Biol. Chem.,* **45**, 2311-2316.

Amaral, D., Bernstein, L., Morse, D., & Horecker, B. (1963). *J. Biol. Chem J.* **238**, 2281-2284.

Andreana, P.R, Xie, W., Cheng, H.N., Qiao. L., Murphy, D.J., Gu, Q.M., Wang, P.G. (2002). *Org Lett.,* **4,** 1863-1866.

Arnold, F. H. (1998). *Acc. Chem. Res.* **31**, 125-131.

Avigad, G. (1985). *Arch. Biochem. Biophys.,* **239**, 531–537.

Avigad, G., Amaral, D., Asensio, C., & Horecker, B. L. (1962). *Journal of Biological Chemistry* **237,** 2736-2743.

Avigad, G., Asensio, C., Amaral, D., & Horecker, B. L., (1961). *Biochem. and Biophys. Research Communs.,* **4,** 474 .

Ayalew S., Confer, A.W., Payton, M.E., Garrels, K.D.,Shrestha, B., Ingram K.R., Montelongo M.A., Taylor, J.D. (2008). *Vaccine,* **26**, 4955–4961.

Baardsnes, J., Sidhu, S., MacLeod, A., Elliott, J., Morden, D., Watson, J. & Borgford, T. (1998). *J. Bacteriol.,* **180,** 3241-3244.

Babcock,G.T.,El-Deeb, M.K., Sandusky, P.O., Whittaker, M.M., Whittaker, J.W., (1992). *J. Am. Chem. Soc.,* **114,** 3727–3734.

Bachmair, A., Finley, D. & Varshavsky, A. (1986). *Science,* **234**, 179-186.

Baldwin, J., Blackburn, J., Schofield, C. & Sutherland, J. (1990). *FEMS Microbiology Letters,* **56**, 45-51.

Barbosa-Tessmann, I.P.; Da Silva, D.A.; Peralta, R.M.; Kemmelmeier, C. (2001). *J. Basic Microbiol.,* **41,** 143-148.

Baron, A. J., Stevens, C., Wilmot, C., Seneviratne, K. D., Blakeley, V., Dooley, D. M., Phillips, S. E. V., Knowles, P. F. & McPherson, M. J. (1994). *J. Biol.Chem.*, **269,** 25095–25105.

Bauskin, A.R., Zhang, H.P., Fairlie, W.D., He, X.Y., Russell, P.K., Moore, A.G., Brown, D.A., Stanley, K.K. & Breit, S.N. (2000). *EMBO J*., **19,** 2212-2220

Benoit, R., Ling, N. & Esch, F. (1987). *Science*, **238**, 1126-1128.

Berry, G. T., Moate, P. J., Reynolds, R. A., Yager, C. T., Ning, C., Boston, R. C., Segal, S. (2004). *Mol.Genet. Metab.* 81, 22-30.

Bertani, G. (1951). *J. Bacteriol.*, **62,** 293-300.

Bertani, G. (2004). *J. Bacteriol*, **186 (3),** 595–600.

Beutler, E., (1991). *Clini.Biochem.*, 24293.

Bílková, Z., Slováková, M., Lyčka, A., Horák, D., Lenfeld, J., Turková, J., Churáček, J. (2002). *J. Chroma. B:Anal.Tech.Biomedic.Life Sci*. **770**, 25.

BIO-RAD, Quick Start™ Bradford Protein Assay Instruction Manual.

Biazio, G.R., Leite, G.G.S., Tessmann, D.J., & Barbosa-Tessmann, I.P., (2008). *Brazilian Journal of Microbiology*, **39(3),** 554-560.

Blobel, G. & Dobberstein, B. (1975). *J. Cell Biol.* **67,** 835-851.

Booth C., (1971). Commonwealth Agricultural Bureaux, Kew, Surrey, England.

Bork, P. & Doolittle, R. F. (1994). *Journal of Molecular Biology*, **236**, 1277-1282.

Boyd, D. & Beckwith, J. (1990). *Cell*, **62**, 1031-1033.

Bradford, M.M., (1976). *Analytical Biochemistry*, **72,** 248-254.

Brady,R.L. & Leibfried,R.T. (2000). US Pat., 6 022 717.

Brahim, S.I., Maharajh, D., Narinesingh, D., Guiseppi-Elie, A.(2002). *Anal.Lett*. 35797.

Bravo, J. , Fita, I. , Ferrer, J. C. , Ens, W. , Hillar, A. , Switala, J. & Loewen, P. C. (1997). *Protein Sci.*, **6**, 1016-1023.

Bretting, H. & Jacobs, G. (1987). *Biochim. Biophys. Acta*, **913**, 342–348.

Brown, W., Ralston, A. & Shaw, K. (2008). Positive transcription control: The glucose effect. Nature Education 1(1)- http://www.nature.com/scitable/topicpage/positive-transcription-control-the-glucose-effect-1009 last accessed 25.05.2013

Bullock, W.O., Fernandez, J.M. & Short, J.M. (1987). *Biotechniques* **5(4),** 376-378.

Bussink, H. J. D., Buxton, F. P., & Visser, J. (1991). *Curr. Genet.* 19,467-474.

Cai, D., & Klinman, J. P. (1994). *J. Biol. Chem.*, **269**, 32039–32042.

Calderhead, D.M.& Lienhard, G.E. (1988). *J. Biol .Chem* ,**263,** 12171-12174.

Chagpar, A., Evelegh, M., Fritsche, H. A., Krishnamurthy, S., Hunt, K. K., & Kuerer, H. M. (2004). *Cancer*, **100(12),** 2549–54. doi:10.1002/cncr.20311

Chidgey, M.A. J. (1993). *BioEssays*, 15(5), 317-321.

Chiu, C.W., Jeffcoat,R., Henley,M. & Peek,L. (1996) US Pat., 5 554 745.

Choosri, W., Paukner, R., Wührer, P., Haltrich, D., & Leitner, C. (2011). *World Journal of Microbiology and Biotechnology*, **27(6),** 1349–1353. doi:10.1007/s11274-010-0585-2

Cannarozzi, G., Schraudolph, N.N., Faty, M., von Rohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., Barral, Y. A. (2010). *Cell.* **141(2),** 355-67. doi: 10.1016/j.cell.2010.02.036.

Cooper, J.A.D., Smithe,W., Bacila,M. & Medina,H. (1959). *J. Biol. Chem.*, **234**, 445–448.

Creemers, J. W., Jackson, R. S., & Hutton, J. C. (1998). *Semin. Cell Dev. Biol.*, **9,** 3–10.

Cui, Y., Jean, F., Thomas, G., & Christian, J. L. (1998). *EMBO J.*, **17**, 4735–4743.

Çevik, E, Şenel, M., Abasıyanık, M.F. (2010). *Current Applied Physics*, **10**, 1313-1316.

Dasgupta,S.P. (1996). US Pat., 5 502 091.

Davidson, V.L. (2007). *Biochemistry.*, **46(18)**, 5283-92.

Deacon, S. E. (2008). PhD Thesis, University of Leeds, Leed, UK.

Deacon, S. E., Mahmoud K., Spooner, R. K., Firbank, S. J., Knowles, P. F., Phillips, S. E. V., McPherson, M. J. (2004). *ChemBioChem*, **5,** 972–979.

Deacon, E.S. & McPherson, M.J. (2011). *ChemBioChem* **12**, 593-601.

Delagrave, S., Murphy, D.J., Pruss, J.L., Maffia, A.M. III, Marrs BL, Bylina EJ, Coleman WJ, Grek CL, Dilworth MR, Yang MM, Youvan DC. (2001). *Protein Eng.*, **14(4)**, 261-267.

Deutscher, J., Francke, C., Postma, P.W. (2006). *Microbiol Mol Biol Rev*. **70(4),** 939-1031.doi: 10.1128/MMBR.00024-06

Dianzani. F., Monahan, T. M., Scupham, A. & Zucca, M. (1979). *Infect. Immun.* **26**, 879-882.

Docherty, M. & Steiner, D. F. (1982). *Annu. Rev. Physiol*., **44**, 625-638.

Dubois, C. M., Laprise, M.-H., Blanchette, F., Gentry, L. E., & Leduc, R. (1995). *J. Biol. Chem*., **270**, 10618–10624.

Duckert, P., Brunak, S., & Blom, N. (2004). *Protein engineering, design & selection,* **17(1)**, 107–12. doi:10.1093/protein/gzh013.

Dwek, R.A., (1996). *Chem Rev*. **96**, 683–720.

Dworkin, M. (1996), *Microbiol. Rev.*, **60,** 70-102.

Edelheit, O.,Hanukoglu, A. & Hanukoglu, I. (2009). *BMC Biotechnology*, **9,** 61.

Eder, J. & Fersht, A. R.(1995.) *Mol. Microbiol.,* **16**, 609-614.

Egnell, P. & Flock, J.I. (1991). *Gene,* **97**, 49-54.

Egnell, P. & Flock, J.I. (1992). *Molecular Microbiology,* **6 (9),** 1115-1119.

Eipper, B. A., Stoffers, D. A. & Mains, R. E. (1992). *Annu. Rev. Neurosci.* **15,** 57-85.

Ekinci, E. & Pasahan, A. (2004). *Eur.Polymer J.* **40**, 1605.

Escalettes, F., & Turner, N. J. (2008). *Chembiochem : a European journal of chemical biology*, **9(6),** 857–860. doi:10.1002/cbic.200700689.

Ettinger, M.J.,& Kosman, D.J. (1982) *in Copper Proteins* (Spiro,T.G., ed) pp 219-261,Wiley Interscience, New York.

Evelegh, M.J. (2002). *J Clin Ligand Assay*. **25**, 91–94.

Evin, G., Zhu, A., Holsinger, R.M.D., Masters, C.L. & Li, Q-X. (2003). *Journal of Neuroscience Research*, **74**, 386-392.

Expedeon Instant*Blue*™ Manual For the Quickest, Sensitive and Safe Staining of Proteins.

Fermentas (2008) GeneJET™ Plasmid Miniprep Kit.

Fierobe, H.-P., Mirgorodskaya, E., Frandsen, T. P., Roepstorff, P., & Svensson, B. (1997). *Protein Expr. Purif.*, **9**, 159–170.

Firbank, S. (2002). *PhD Thesis,* School of Biochemistry and Molecular Biology, University of Leeds: Leeds.

Firbank, S.J., Rogers, M., Guerrero, R.H., Dooley, D.M., Halcrow, M.A., Phillips, S.E.V., Knowles, P.F., & McPherson, M.J. (2003). *Biochem. Soc. Trans*., **31**, 506-509.

Firbank, S. J., Rogers, M. S., Wilmot, C. M., Dooley, D. M., Halerow, M. A., Knowles, P. F.,McPherson, M. J. & Phillips, S. E. V. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 12932-12937.

Fredrick, Kurt; Ibba, M., (2010). *Cell*, **141(2)**, 227-229.

Frollini, E., Reed, W.F., Milas, M., Rinaudo, M. (1995). *Carbohydrate Polymers*, **27,** 129-135.

Gahmberg, C.G. & Tolvanen,M. (1994). *Methods Enzymol.*, **230,** 32-44.

Gancedo, J. M., C. Gancedo, & C. Asensio.(1967). *Arch. Biochem. Biophys*., **119**, 588-590.

García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R.M., Arís, A., Ventura, S. & Villaverde, A. (2005). *Microbial Cell Factories*, **4,** 27.

Gençer (Kocuklu), B**.** (2005). *M.Sc. Thesis*, Middle East Technical University, Ankara.

Gerfen, G.A., Bellew, B., Grin R., Singel D., Ekberg C.A. & Whittaker J.W., (1996). *J. Phys. Chem.*, **100,** 16739–16748.

Ghiladi, R. A., Knudsen, G. M., Medzihradszky, K. F., & de Montellano, P. R. O. (2005). *J. Biol. Chem.*, **280**, 22651–22663.

Gonda, D. K., Bachmair, A., Wunning, I., Tobias, J. W., Lane, W. S. & Varshavsky, A.(1989). *J. Biol. Chem.*, **264**, 16700-16712.

Gottesman, S. & Maurizi, M.R.(1992). *Microbiol Rev*. **56(4),** 592–621.

Görke, B. & Stülke, J. (2008). *Nature Reviews. Microbiology*, **6 (8)**, 613–24.

Graichen, M. E., Jones, L. H. , Sharma, B. V. , van Spanning, R. J. , Hosler, J. P. & Davidson, V. L. (1999). *J. Bacteriol.*, **181**, 4216-4222.

Gross, D. J., Villa-Komaroff, L., Kahn, C. R., Weir, G. C. & Halban, P. A. (1989). *J. Biol. Chem.*, **264**, 21486-21490.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., & Osburne, M.S. (1998). *Gene*, **209**, 95-103.

Grunberg-Manago, M. (1999). *Annu. Rev. Genet.*, **33**, 193-227.

Guan, C., Cui, T., Rao, V., Liao, W., Benner, J., Lin, C. L. & Comb, D. (1996). *J. Biol.Chem*., **271,** 1732- l737.

Gurr,S.J., Unkles, S.E., & Kinghorn, J.R. (1987). *in Gene Structure in Eukaryotic Microbes* (Kinghorn,J.R. ed.), 93-139, IRL Pres, Oxford.

Halban, P. A., & Irminger, J. C. (1994). *The Biochemical Journal*, **299**, 1–18.

Hamilton., G.A.(1982*). in Copper Proteins* (Spiro,T.G., ed) pp 193-218,Wiley Interscience, New York Chem., **100,**16739–16748.

Hamilton, G. A., DeJersey, J., & Adolf, P. K. (1973), in *Oxidases and Related Redox Systems*, King, T. E., Mason, H. S., & Morrison, M., Ed., Baltimore, Md., University Park Press, p 103.

Hamilton, G. A., Libby, R. D. & Hartzell, C. R. (1973). *Biochemical and Biophysical Research Communications*, **55**, 333-340.

Hardy, S.J.S. & Randall, L. L. (1992). *Antonie van Leeuwenhoek*, **61(2)**, 101-103.

Harrison, M. D., Jones, C. E., Solioz, M. & Dameron, C. T. (2000). *Trends Biochem. Sci.*, **25**, 29–32.

Hartmans, S., de Vries, H.T., Beijer, P., Brady, R.L., Hofbauer, M., Haandrikman, A.J. (2004). *Hemicelluloses: Science and Technology*, **864,** 360-371.

Heim, R. , Prasher, D. C. & Tsien, R. Y. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 12501-12504.

Helenius, A., Aebi, M. (2001). *Science*, **291**, 2364–2369.

Helenius, A., Marquardt, T. & Braakman, I. (1992). *Trends Cell Biol*., **2**, 227-231.

Hogrefe, H. H., Hansen, C. J., Scott, B. R. & Nielson, K. B. (2002). *Proc Natl Acad Sci USA*, **99(2)**, 596-601.

Hurtley, S. M. & Helenius, A. (1989). *Annu. Rev. Cell Biol.* **5**, 277-307.

Huttner, W. B. (1988). *Annu. Rev. Physiol.*, **50**, 363-376.

IBA (2002) *Strep* -tag® Purification Protocol Manual Purification of Strep-tag proteins with *Strep*-Tactin® matrices.

IBA Protein TAGnologies Newsletter Issue 9.

Invitrogen (2009) XCell II™ Blot Module User Manual Rev. Date: 14 December 2009 Manual part no. IM9051 MAN0000740.

Invitrogen (2010) Champion™ pET Directional TOPO®Expression Kits User Manual, Rev. Date 7 June 2010 Manual part no. 25-0400, MAN0000214.

Ito, N., Phillips, S.E.V., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D.S., Knowles, P.F., (1991). *Nature*, **350,** 87-90.

Ito, N., Phillips, S.E.V., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D.S., Knowles, P.F., (1992). *Faraday Discuss*, **93,** 75–84.

Ito, N., Phillips, S. E. V.,Yadav, K. D. S., & Knowles, P.F. (1994). *J.MoLBiol.*, **238,** 794-814.

Springael, J.Y., Nikko, E. André, B.& Marini, A.M. (2002). *FEBS Letters*, **517**, 103-109.

Jalving, R. (2005). *PhD Thesis*, Wageningen University, Netherlands.

Janes, S.M., Mu, D.,Wemmer, D., Smith, A. J., Kaur, S.,Maltby, D., Burlingame, A. L., & Klinman, J. P. (1990). *Science*, **248**, 981–987.

Jimenez-Huete ,A., Lievens, P.M.J., Vidal, R., Piccardo, P., Ghetti, B., Tagliavini, F., Frangione, B., Prelli, F.(1998). *Am J Path*; **153**,1561-1572.

Johnson, J. M., Halsall, H. B. & Heineman, W. R. (1982). *Analytical Chemistry*, **54**, 1394-1399.

Johnson, J. M., Halsall, H. B. & Heineman, W. R. (1985). *Biochemistry*, **24**, 1579-1585.Kelleher, F.M., & Bhavanandan, V.P.(1986). *J. Biol. Chem*., **261**, 11045-11048.

Kelley, K.C., Huestis, K.J., Austen, D.A., Sanderson, C.T., Donoghue, M.A. Stickel, S. K, Kawasaki, E.S., & Osburne, M.S. (1995). *Gene*, **156**, 33–36.

Kelley, W. L. & Georgopoulos, C. (1992). *Curr. Opin. Cell Biol*., **4**, 984-991.

Kelly, R.B. (1985). *Science*, **230**, 25-32.

Khun, K., Ibupoto, Z. H., Nur, O., & Willander, M. (2012). *Journal of Sensors*, **2012,** 1–7. doi:10.1155/2012/696247

Kido, M., Yamanaka, K., Mitani, T., Niki, H., Ogura, T., and Hiraga, S. (1996). *J. Bacteriol.* **178**, 3917-3925

Kornfield, R., & Kornfield, S. (1985). *Annu. Rev. Biochem.*, **54,** 631- 664.

Kosman, D.J. (1985). *Copper Proteins and Copper Enzymes*, **2** (ed Lontie, R.), (CRC Press, Boca Raton publishers, Boca Raton 1–26.

Kosman, D.J., Ettinger, M.J., Weiner, R.E. & Massaro, E.J. (1974). *Arch. Biochem. Biophys*, **165,** 456-467.

Lamy, B., & Davies, J. (1991). *Nucleic Acids Research*, **19(5),** 1001-1006.

Landry, S. J. & Gierasch, L. M. (1991). *Trends Biochem. Sci.*, **16,** 159-163.

Laursen, B.S., Steffensen, S., Hedegaard, J., Moreno, J.M., Mortensen, K.K., Sperling-Petersen, H.U. (2002). *Genes Cells*. **7**, 901–910.

Li, C. H., Barnaffi, L., Chretien, M. & Chung, D. (1965). *Nature (London)*, **208**,1093-1094.

Li, P., Beckwith, J. & Inouye, H. (1988). *Proc. Natl. Acad. Sci. U.S.A*, **85,** 7685-7689.

Lilie, H.,Schwarz, E. & Rudolph, R. (1998). *Current Opinion in Biotechnology*, **9,** 497-501

Lingappa, V. (1991).*Cell*, **65**, 527-530.

Lis, M., & Kuramitsu, H. K. (1997). *Antimicrobial Agents and Chemotherapy*, **41(5)**, 999–1003.

Liu, X. C. & Dordick, J. S.(1999). *J. Am. Chem. Soc.*, **121**, 466-467.

Longhi, S. & Cambillau, C. (1999). *Biochim. Biophys. Acta*, **1441,** 185–196.

Lopez, P. J., Marchand, I., Joyce, S. A., & Dreyfus, M. (1999). *Mol. Microbiol*. **33**, 188-199

Machado, L.C, Kemmelmeier C. (2001). *Mycopathologia*, **149(2),** 79-85.

Majerus, P. M. C. & Courtois P. A. P. (1992). *J. Biol. Buccale*, **20**, 241–245.

Manowltz, P., Stoecker, P.W., Yacynnych, A.M. (1995). *Biosens.Bioelectron.*,**10**, 359.

Mannino, S., Cosio,M.S. & Buratti,S. (1999). *Ital. J. Food. Sci.*, **11,** 57-65.

Martínez-Ruiz, A., del Pozo, A. M., Lacadena, J., Mancheño, J. M., Oñaderra, M., López-Otín, C. & Gavilanes, J. G. (1998). *Protein Expr. Purif.,* **12,** 315–322.

Martinez, C., De Gaus, P., Lauwereys, M., Matthyssens, G. & Cambillau, C.(1992). *Nature(London),* **356,** 615-618.

Mazur, A. W. (1991). *ACS Symposium Series,* **466**, 99-l10.

Mazur, A.W. & Hiler, G.D.(1997). *J. Org. Chem*., **62,** 4471-4475.

McGlothlin, C.D., Purdy, W.C. (1977). *Analytica Chimica Acta.,* **88,** 33.

McCoy, J. G., Bailey, L. J., Bitto, E., Bingman, C. A., Aceti, D. J., Fox, B. G., & Phillips, G. N. (2006). *Proc. Natl. Acad. Sci. U.S.A.,* **103,** 3084–3089.

McFaul, S. J., H. Lin, & J. Everse. (1986). *Proc. Soc. Exp. Biol. Med.*,**183,** 244–249.

McPherson, M.J., Ogel, Z.B., Stevens, C., Yadav, K.D.S., Keen, J.N. & Knowles, P.F., (1992). *J.Biol.Chem.*,**267**, 8146-8152.

McPherson, M. J.; Parsons, M. R.; Spooner, R. K.; Wilmot, C. R. In Handbook of Metalloproteins; Messerschmidt, A.; Huber, R., Wieghardt, K., Poulos, T., Eds.; Wiley: Hoboken, NJ (2001); pp 1272-1283.

McPherson, M.J., Stevens, C., Baron, A.J., Ogel, Z.B., Seneviratne, K.,Wilmot, C., Ito, N., Brocklebank, I., Phillips, S.E.V. & Knowles, P.F.(1993). *Biochem. Soc. Trans*., **21**, 752–756

Mendonca, M. H., & Zancan, G. T. (1987). *Arch. Biochem. Biophys.*, **252,** 507-514.

Mendonca, M. H., & Zancan, G. T. (1988). *Arch. Biochem. Biophys*., **266,** 427-434

Mendonca, M. H., & Zancan, G. T. (1989). *Arch. Biochem. Biophys*., **275**, 130-139

Messerschmidt, A. (1994). *Adv. Inorg. Chem.,* **40**, 121.

Meyer, D. I. (1991). *Trends Cell Biol.,* **1**, 154-159.

Minasian, S. G., Whittaker, M. M. & Whittaker, J. W. (2004). *Biochemistry,* **43**, 13683-13693.

Moore, H.-P. H. & Kelly, R. B. (1985). *J. Cell Biol.,* **101**, 1773-1781.

Multhaup, G., Ruppert, T., Schlicksupp, A., Hesse, L., Bill, E., Pipkorn, R.,Masters, C.L. & Beyreuther, K. (1998). *Biochemistry,* **37**, 7224–7230.

Nakayama, K. (1997). *Biochem. J.*, **327**, 625–635 12.

Neurath, H. (1957). *Adv. Protein Chem.*, **12**, 319-386.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). *Protein Engineering*, **10**, 1-6.

NORGEN (2007). ProteoSpin™ Inclusion Body Isolation Micro Kit Application Manual.

Novagen (2011). pET System Manual 11th Edition, User Protocol TB055 Rev. C 0611JN.

Novy, R., Drott, D., Yaeger, K., & Mierendorf, R. (2001). *Innovations*, **12,** 1–3.

O'Donnell, K., Kistler, H.C., Tacke, B.K., Casper, H.H. (2000). *Proc. Natl. Acad. Sci. USA,* **97,** 7905–7910.

O'Donnell, K., Ward, T.J., Geiser, D.M., Kistler, H.C., Aoki, T.(2004). *Fungal Genet. Biol*., **41**, 600–623.

Okeley, N.M. & van der Donk, W.A. (2000) *Chem. Biol* **.7,** R159-R171.

Orci, L., Ravazzola, M., Amherdt, M., Perrelet, A., Powell, S. K., Quinn, D. L. & Moore, H. P. H. (1987). *Cell*, **51**, 1039-1051.

Ostermeier, C., Harrenga, A., Ermler, U., & Michel, H. (1997). *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10547–10553.

Ögel, Z.B. (1993). *Ph.D.Thesis,* Univ. of Leeds, Leeds, U.K.

Ögel, Z.B., Brayford, D. & McPherson, M.J. (1994). *Mycol. Res*., **98,** 474–480

Ögel, Z. B. & Özilgen, M. (1995). *Enzyme and Microbial Technology*, **17**, 870-876.

Özögür-Akyüz, S., Shawe-Taylor , J., Weber, G.-W., Ögel, Z. B. (2009). *Discrete Applied Mathematics*,**157(10)**, 2388-2394.

Pallas, J. A. & McPherson, M. J. (1997). *Plant Physiology* **114**, 1165-1165

Pan, S. & Malcom, B.A. (2000). *BioTechniques*, 29, 1234–1238.

Parikka, K., Leppänen, A.-S., Pitkänen, L., Reunanen, M., Willför, S. & Tenkanen, M. **(**2010). *Journal of Agricultural and Food Chemistry*, **58 (1),** 262-271.

Pearson Education, Inc., 2011

Pearson Prentice Hall,Inc., 2005

Peberdy, J. F. (1994) *TIBTECH*, **12**, 50-57

Perlman, D. & Halvorson, H.O. (1983). *Journal of Molecular Biology,* **167***,* 391-409.

Peschon, J., Slack, J. L., Reddy, P., Stocking, K. L., Sunnarborg, S. W., Lee, D. C., Russell, W. E., Castner, B. J., Johnson, R. S., Fitzner, J. N., Boyce, R. W., Nelson, N., Kozlosky, C. J., Wolfson, M. F., Rauch, C. T., Cerretti, D. P., Paxton, R. J., March, C. J., & Black, R. A. (1998). *Science*, **282,** 1281–1284.

Petersen, T.N., Brunak S., von Heijne G. & Nielsen H.(2011). *Nature Methods*, **8,** 785-786.

Petrounia, I. P. & Arnold F.H.(2000). *Curr. Opin. Biotechnol*., **11,** 325-330.

Proshlyakov, D. A., Pressler, M. A., DeMaso, C., Leykam, J. F., DeWitt, D. L., & Babcock, G. T. (2000). *Science*, **290**, 1588–1591.

QIAGEN (2006) QIAprep® Miniprep Handbook.

Radin, N., & Evangelatos, G. (1981). *Journal Of Lipid Research*, **22(3)**, 536-541.

Rae, T. D., Schmidt, P. J., Pufahl, R. A., Culotta, V. C. & O'Halloran, T. V. (1999). *Science*, **284**, 805–808.

Rannes, J. B., Ioannou, A., Willies, S.C., Grogan, G., Behrens, C., Flitsch, S.L. & Turner, N. J. (2011). *Journal of the American Chemical Society*,**133 (22)**, 8436-8439.

Rapoport, T. A. (1992). *Science*, **258**, 931-936.

Ravazzola, M., Perrelet, A., Roth, J. & Orci, L. (1981). *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 5661-5664.

Rehemtulla, A., Dorner, A. J., & Kaufman, R. J. (1992). *Proc. Natl. Acad. Sci. U.S.A*, **89(17)**, 8235-8239.

Reynolds, M. P., Baron, A. J., Wilmot, C. M., Phillips, S. E. V., Knowles, P. F. & McPherson, M. J. (1995). *Biochemical Society Transactions*, **23**, 510S.

Reynolds,M. P., Baron, A. J., Wilmot,C. M., Vinecombe, E., Stevens,C., Phillips, S. E. V., Knowles, P. F. & McPherson, M. J.(1997). *J. Biol. Inorg. Chem*., **2**, 327-335.

Roebroek, A. J., Umans, L., Pauli, I. G., Robertson, E. J., van Leuven, F., Van de Ven, W. J., & Constam, D. B. (1998). *Development*, **125**, 4863-4876.

Rogers, M. S., Baron, A. J., McPherson, M. J., Knowles, P. F., & Dooley, D. M. (2000). *Journal of the American Chemical Society*, **122,** 990-991.

Rogers, M. S. & Dooley, D. M., (2003). *Current Opinion in Chemical Biology*, **7**,189-196.

Rogers, M. S., Hurtado-Guerrero, R., Firbank, S. J., Halcrow, M. A., Dooley, D. M., Phillips, S. E. V. , Knowles, P. F., McPherson M.J. (2008). *Biochemistry,* **47,** 10428–10439.

Root, R.L., Durrwachter, J.R. & Wong.C.H. (1985). *J. Am. Chem. Soc*., **107,** 2997-2999.

Rose, J. K. & Doms, R. W. (1988). *Annu. Rev. Cell Biol*. **4,** 257-288.

Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J., and Studier, F.W. (1987). *Gene*, **56**, 125–35.

Rosenzweig, A. C. & O'Halloran, T. V. (2000). *Curr. Opin. Chem. Biol.,* **4**, 140–147.

Roth, H., Segal, S. & Bertoli, D. (1965). *Analytical Biochemistry,* **10**, 32-52.

Rothman, J. E. (1989). *Cell*, **59**, 591-601.

Rouillé, Y., Duguay, S., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A., Jr., Chan, S., & Steiner, D. (1995). *Front. Neuroendocrinol.*, **16**, 322–361.

Rozkov, A, (2001). *MSc Thesis*, Department of Biotechnology, Royal Institute of Technology (KTH) S-10691 Stockholm, Sweden.

Sharma, S.K., Suman Pundir, C.S., Sehgal, N., Kumar, A., *Sens Actuators B Chem.* (2006). **119,** 15.

Sharma, S.K., Singhal, R., Malhotra, B.D., Sehgal, N., Kumar, A. (2004). *Electrochimica Acta*, **49,** 2479.

Sharma, S.K., Singh, S.K., Sehgal, N., Kumar, A. (2004). *FoodChem.*,**88**, 299.

Said, I.T., Shamsuddin,A.M., Sherief,M.A., Taleb,S.G., Aref, W.F. & Kumar, D. (1999). *Histol. Histopathol.*,**14,** 351-357.

Sakai, J., Rawson, R. B., Espenshade, P. J., Cheng, D., Seegmiller, A. C., Goldstein, J. L., & Brown, M.S. (1998). *Mol. Cell*, **2**, 505–514.

Sambrook, J. & Russel, D.W. (2001). *Molecular Cloning: A Laboratory Manual* 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

Satoh, A., Kim, J.K., Miyahara, I., Devreese, B., Vandenberghe, I.,Hacisalihoglu, A., Okajima, T., Kuroda, S., Adachi, O., Duine, J.A. *et al*. (2002). *J.Biol. Chem*., **277**, 2830–2834.

Schlegel, R.A., Gerbeck,C.M. & Montgomery,R. (1968). *Carbohydr. Res.,* **7,** 193–199.

Schnell, R., Sandalova, T., Hellman, U., Lindqvist, Y., & Schneider, G. (2005). *J. Biol. Chem.*, **280**, 27319–27328.

Schwartz, T. W. (1986). *FEBS Letters*, **200**, 1-10.

Schwede, T. F. , Rétey, J. & Schulz, G. E. (1999). *Biochemistry*, **38**, 5355-5361.

Scitable by Nature Education, The Information in DNA Determines Cellular Function via Translation, http://www.nature.com/scitable/topicpage/the-information-in-dna-determines-cellular-function-6523228, last accessed, 23.08.2013.

Segal, S., Wehrli, S., Yager, C., Reynolds, R. (2006). *Mol.Genet.Metab*., **87**, 92.

Seidah, N.G. & Chretien, M. (1999). *Brain Res.*, **848**, 45-62.

Seidah, N., Chrétien, M., & Day, R. (1994). *Biochimie (Paris)*,**76**, 197–209.

Seidah, N., Mbikay, M., Marcinkiewicz, M., & Chrétien, M. (1998). *in Proteolytic and Cellular Mechanisms in Prohormone Processing* (Hook, V., ed) pp. 49–76, RG Landes, Georgetown.

Seneviratne, K. (1993). *PhD Thesis*, School of Biochemistry and Molecular Biology: University of Leeds, UK.

Shaolin, M. (1994). *Journal of Electroanalytical Chemistry*, **370**, 135-139.

Sharma, S.K., Kumar A., Chaudhary, R. Pundir, S.& Pundir, C. S., Sehgal, N. (2007). *Artificial Cells, Blood Substitutes, and Biotechnology*, **35**, 421–430. doi: 10.1080/10731190701460309.

Shatzman, A.R. & Kosman, D. J. (1997). *Journal of Bacteriology*, **130**, 455-463.

Sherlock, O., Dobrindt, U., Jensen, J. B., Vejborg, R. M., & Klemm, P. (2006). *Journal of Bacteriology*, **188(5)**, 1798–1807. doi:10.1128/JB.188.5.1798

Shinde, U. & Inouye,M. (2000). *Semin. Cell Dev. Biol.*, **11**, 35-44.

SignalP 4.1 Server, http://www.cbs.dtu.dk/services/SignalP/, last accessed, 22.08.2013.

Silakowski, B. Ehret, H. & Schairer, U. *J. Bacteriol*., (1998). **180**, 1241-1247.

Silen, J.L. & Agard, D.A. (1989). *Nature*, **341**, 462-464.

Simmons, C. R., Liu, Q., Huang, Q. Q., Hao, Q., Begley, T. P., Karplus, P. A., & Stipanuk, M. H. (2006). *J. Biol. Chem.*, **281**, 18723–18733.

Simonen, M. & Palva, I. (1993). *Microbiological Reviews,* **57(1),** 109-137.

Smulevich, G., Jakopitsch, C., Droghetti, E., & Obinger, C. (2006). *J. Inorg. Biochem.,* **100**, 568–585.

Soliday, C. L., Flurkey, W. H., Okita, T.W. & Kolattukudy, P.E. (1984). *Proc. Natl. Acad. Sci. U.S.A* **81**, 3939-3943.

Sørensen N. H. & Novozymes A/S (Bagsvaerd, DK) PCT/DK2001/000166, 2002

Springer, G. F. (1997). *J. Mol. Med.,* **75,** 595–602.

Steiner, D. F., Cunningham, D., Spigelman, L. & Aten, B. (1967). *Science*, **157**, 697-699.

Steiner, D. F. & Oyer, P. E. (1967). *Proc. Natl. Acad. Sci. U.S.A.* **57**, 473-480.

Stoller, T. J. & Shields, D. (1989). *J. Cell Biol.,* **108**, 1647-1655.

Stratagene (2007). *QuikChange® II Site-Directed Mutagenesis* Kit Instruction Manual, USA.

Stubbe, J.& van der Donk, W.A.(1998). *Chem. Rev.* **98,** 705-762.

Studier, F. W. (2005). *Protein Expression and Purification,* **41**, 207-234.

Studier, F. W. & Moffatt, B. A. (1986). *Journal of Molecular Biology*, **189**, 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods in Enzymology* **185**, 60-89.

Sun, L. H, Bulter, T., Alcalde, M., Petrounia, I.P., Arnold, F.H. (2002). *ChemBioChem*, **3,** 781-783.

Sun, L. H., Petrounia, I.P., Yagasaki, M., Bandara, G.,Arnold, F.H., (2001). *Protein Engineering*, **14,** 699-704.

Suang, W.J., Bae, Y.H., (2006). *Sens Actuators B Chem*.,**114,** 164.

Szabó, E.E., Adányi, N., Váradi, M. (1996). *Biosensors and Bioelectronics,* **11(10)** 1051-1058.

Szameit, C. , Miech, C. , Balleininger, M. , Schmidt, B. , von Figura, K. & Dierks, T. (1999). *J. Biol. Chem.,* **274**, 15375-15381.

Szymanski, C.M., Yao, R., Ewing, C.P., Trust., T.J., Guerry, P. (1999). *Mol Microbiol.,* **32**, 1022–30.

Şenel, M., Bozgeyik İ.,Çevik E., Abasıyanık, M. F. (2011). *Synthetic Metals*, **161(5–6)**, 440-444.

Tartakoff, A. & Vassalli, P. (1978). *J. Cell Biol*. **79**, 694-707.

Tatsumi, H., Murakami, S.,Tsuji, R.F., Ishida,Y., Murakami, K., Masaki, A., Kawabe, H., Arimura, H., Nakano, E. & Motai, H. (1991). *Molecular and General Genetics*, **228**, 97-103.

Taylor P.J., Kemetec, F., Johnson, M.J. (1977). *Anal.Chem.*, **49,** 789.

ThermoScientific (2008). NanoDrop 1000 Spectrophotometer V3.7 User's Manual, USA.

Tkac, J., Gemeiner,P. & Sturdik,E. (1999) *Biotechnol. Tech,* **13,** 931-936.

Tobias, J. W., Shrader, T. E., Rocap, G., & Varshavsky,A. (1991). *Science*, **254**,1374-1376.

Tooze, J., Tooze, S. A. & Fuller, S. D. (1987). *J. Cell Biol*., **105**, 1215-1226.

Tressel, P.S., & Kosman, D.J.(1982) *Methods Enzymol*. **89**, 163-171.

van Poelje, P. D. & Snell, E. E. (1990). *Annu. Rev. Biochem.* **59**, 29-59.

Varshavsky, A.(1992). *Cell*, **69,** 725-735.

Varshavsky, A. (1997). *Genes to Cells*, **2,** 13–28.

Vega, F.A., Numez, C.G., Weigel, B., Hitzmann, B. & Ricci, J.C.D. (1998). *Anal. Chim. Acta,* **373,** 57-62.

Von Heijne, G. (1985). *Journal of Molecular Biology*, **184**, 99-105.

Von Heijne, G. (1986a). *Nucleic Acids Research,* **14(1),** 4683-4690.

Von Heijne, G. (1986b). *Journal of Molecular Biology,* **189,** 239-242.

Vrbova, E., Peckova,J. & Marek, M. (1992). *Collect. Czech. Chem. Commun.*, **57,** 2287-2294.

Vucenik, I., Gotovac, J., Druzijanic, N. & Shamsuddin, A. M. (2001). *Anticancer Res.*, **21,** 1247–1255.

Walter, P. & Lingappa, V. R. (1986). *Annu. Rev. Cell Biol.* **2**, 499-516.

Wang W. (2010). Large Insertions: Two Simple Steps Using *QuikChange II Site-Directed Mutagenesis Kits,* Application note, Agilent Technologies, Inc. , USA.

Wang, W. & Malcolm, B. (1999). *Biotechniques,* **26**, 680-682.

Weidemann, A., Paliga, K., Dürrwang, U., Reinhard, F.B.M., Schuckert, O., Evin, G.& Masters, C.L. (1999). *The Journal of Biological Chemistry* **274 9 (26),** 5823-5829.

Whittaker, M. M. & Whittaker, J. W. (1988a). *Journal of Biological Chemistry,* **263**, 6074-6080.

Whittaker, M. M. & Whittaker, J. W. (1988b). *Abstract Papers American Chemical Society* **195**, 553-INOR.

Whittaker,M.M., Devito,V.L., Asher,S.A. & Whittaker, J.W. (1989) *J. Biol. Chem.,* **264**, 7104–7106.

Whittaker, M.M. & Whittaker, J.W. (1990) *J. Biol. Chem*. **265**, 9610–9613.

Whittaker, M. M. & Whittaker, J. W. (1993). *Biophysics Journal,* **64**, 762-772.

Whittaker, J. W. (1994). *Metal Ions in Biological Systems, Metalloenzymes Involving Amino Acid- Residue and Related Radicals*, H. Sigel, A. Sigel (Eds.), vol. 30, Marcel Dekker, New York, p. 315.

Whittaker, M. M., Kersten, P. J., Nakamura, N., Sanders-Loehr, J., Schweizer, E.S. & Whittaker, J. W. (1996). *Journal of Biological Chemistry,* **271**, 681-687.

Whittaker, M. M., Ballou, D. P. & Whittaker, J. W. (1998). *Biochemistry* **37**, 8426-8436.

Whittaker, J.W. & Whittaker, M.M. (1998). *Pure & Appl. Chem.*, **70 (4),** 903-910.

Whittaker, M. M., Kersten, P. J., Cullen, D., & Whittaker, J. W. (1999). *J. Biol. Chem.*, **274**, 36226–36232.

Whittaker, M. M., & Whittaker, J. W. (2000). *Protein expression and purification*, **20(1**), 105–111. doi:10.1006/prep.2000.1287.

Whittaker, J.W. (2003). *Chem. Rev.*, **103**,2347-2363.

Whittaker, M. M., & Whittaker, J. W. (2003). *J. Biol. Chem.,* **278**, 22090-22101

Whittaker, J.W. (2005). *Archives of Biochemistry and Biophysics,* **433**, 227–239.

Wieland, F. T., Gleason, M. L., Serafini, T. A. & Rothman, J. E. (1987). *Cell,* **50,** 289-300.

Wielinga, W. C. (2000). *Galactomannans. In Handbook of Hydrocolloids*. Phillips, G. O.,Williams, P. A., Eds.;Woodhead Publishing: Cambridge, UK, pp 137-154.

Wilkinson, D., Akumanyi, N., Hurtado-Guerrero, R., Dawkes, H., Knowles, P. F., Phillips, S. E. V., & McPherson, M. J. (2004). *Protein Engineering, Design & Selection*, **17(2),** 141–148. doi:10.1093/protein/gzh018.

Wilkinson, K. (1995). *Annu. Rev. Nutr*. **15**, 161–189.

Willner, I.,Riklin, A., Shoham, B., Rivenzon, D. & Katz, E. (1993). *Adv. Mater.,* **5(12),** 912-915.

Xie, L. & van der Donk, W.A.(2001). *Proc. Natl. Acad. Sci. USA,* **98**, 12863-12865.

Xu, F., Golightly, E., Schneider, P., Berka, R., Brown, K. Johnstone, J., Baker, D., Fuglsang, C., Brown, S., Svendsen, A. & Klotz, A. (2000). *Appl. Biochem. Biotechnol.*, **88**, 23-32.

Yamada, Y., Fujiwara, T., Sato, T., Igarashi, N., & Tanaka, N. (2002). *Nat. Struct. Biol.*, **9**, 691–695.

Yang, Y., Mu, S.& Chen, H. (1998), *Synth.Met.*, 92173.

Yang, G.Y. & Shamsuddin, A.M. (1996). *Histol. Histopathol.*, **11,** 801-806.

Yokoyama, K., Kawada M., Tamiya E. (1997). *Journal of Electroanalytical Chemistry,* **434**, 217-224.

Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M. J., Libeu, C. P., Mizushima, T., Yamaguchi, H., Tomizaki, T., & Tsukihara, T. (1998). *Science*, **280**, 1723–1729.

Young, N. M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., Michael, F.S., Aberg, E. & Szymanski, C.M. (2002). *J Biol Chem.* **277**, 42530-42539.

Zancan, G. T., & Amaral, D. (1970). *Biochim. Biophys. Acta*, **198,** 146.

Zhu, X., Ohta, Y., Jordan, F. & Inouye, M. (1989). *Nature* **339,** 483-484.

APPENDIX A

PREPARATIONS OF GROWTH MEDIA, BUFFERS AND SOLUTIONS

Agarose (1%)

1 g agarose is dissolved in 100 ml 1 x TAE buffer by strring and dissolving.

Ammonium Acetate (Mass Spectrometry)

50 mM ammonium acetate

Ampicillin (50 mg/ml stock)

500 mg ampicillin is dissolved in 10 ml ddH_2O, filter sterilised (0,2 um filter), aliquoted into ~400 µl and stored at -20°C. It is used at a dilution of 1/1000 (400 µl in 400 ml) to give 50 $\mu g.ml^{-1}$ working concentration.

APS (25%)

1 g ammonium persulphate is dissolved in 4 ml dH_2O. It can be stored fresh at 4°C for about one week.

Binding Buffer (500 ml) (pH 7,9)

1 M NaCl
20mM Tris–HCl
5mM imidazole
8M urea

29,22 g NaCl and 0,17 g imidazole is dissolved in 20mM Tris-HCl. Then to 200 µl of this buffer 96 g urea was added to a final concentration of 8M. This solution is diluted as 6M, 4M, 2M with the remaining buffer.

10x BSA (100 µl)

10 x NEBuffer 4	10 µl
100 x BSA	10 µl
ddH_2O	80 µl

Bradford's Solution (1x) (1L)

100 mg Coomassie blue G250
50 ml 95% ethanol
100 ml phosphoric acid

After the dye is completely dissolved, it is diluted to 1 L with dH_2O and filtered through 3MM filter paper.

BSA Protein Standart (1 mg.ml^{-1})

104.17 mg BSA (96%) is dissolved in 0.9% NaCl and 0.05% sodium azide (100 ml). This solution is stored at 4°C.

Buffer P1 (Resuspension Buffer)

50 mM Tris·Cl, pH = 8.0
10 mM EDTA
100 µg/ml RNase A

Buffer P2 (Lysis Buffer)

200 m M NAOH
1% SDS

Buffer E (Elution buffer) (1L)

100 mM Tris-Cl pH 8.0
150 mM NaCl
2.5 mM desthiobiotin

Reagents are dissolved in ¾ final volume of deionised water, made up to required volume then aliquoted into 250ml and autoclaved at 121°C 15psi for 20minutes. Just before use 0,268g of desthiobiotin is dissolved in 250 ml of buffer, once dissolved, pH is adjusted to 7.4 with 1M NaOH and filtered through a 0,22 µm filter.

Buffer R (Regeneration buffer) (500 ml)

100 mM Tris-Cl pH 8.0
150 mM NaCl
1 mM HABA (hydroxy-azophenyl- benzoic acid)

Reagents are dissolved in ¾ final volume of deionised water, pH to 8.0 with concentrated HCl, then made up to required volume with deionised water. Buffer is filtered through a 0,22 μm filter before passing through column .

Buffer W (Washing buffer)

100 mM Tris-Cl pH 8.0
150 mM NaCl

Carbenicillin (50 mg/ml stock)

500mg carbencillin is dissolved in 10ml ddH_20, filter sterilised (0.2μm filter), and aliquoted into ~500 μl and stored at -20^{o}C. It is used at a dilution of 1/1000 (500 μl in 500 ml) to give 50 μg/ml working concentration.

Coomassie blue Staining Solution (1 L)

Methanol	450 ml	(45%)
Acetic acid	70 ml	(7%)
dH_2O	480 ml	
Coomassie blue R250	2.5 g	(0.25%)

Coomassie blue is dissolved in methanol, acetic acid and H2O are added to 1 L. It is stored in dark bottle at room temperature, recycled if necessary.

$Cu(NO_3)_2$ (1 mM)

0.188 g $Cu(NO_3)_2$ is dissolved in 1L dH_2O.

$CuSO_4$ (100 mM)

15.95 g is dissolved in 1L dH_2O.

Destaining Solution (2 L)

Methanol	500 ml	(25%)
Acetic acid	150 ml	(7.5%)
dH_2O	1350 ml	

After mixing all, it is stored in dark bottle at room temperature, recycled if necessary.

Developing Solution (Silver Staining)

9g potassium carbonate, 300 μl 37 % formaldehyde and 8 ml previously kept pretreatment solution are mixed and volume is completed to 400ml with dH_20.

DNase-free RNase

RNase A is dissolved in 0.01 M NaAc (pH = 5.2) to give a final concentration of 10 mg/ml. The solution is heated to 100°C for 15 minutes in a boiling water bath for the inactivation of DNase. It is cooled slowly to room temperature. 0,1 volume of 1 M Tris-HCl (pH = 7.5) is added until the pH of the solution is 7.0. The solution is dispensed into aliquots and stored at -20 °C.

EDTA (0.5 M, pH = 8.0)

186,1 g of ethylenedinitrilotetraacetic acid disodium salt dihydrate is added to 800 ml of distilled water. It is stirred vigorously on a magnetic stirrer while the pH is adjusted to 8.0 with NaOH pellets. The solution is dispensed into aliquots and sterilized by autoclaving.

Ethanol (20%) (500 ml)

Absolute ethanol	100 ml
H_2O	400 ml

Ethanol (70%) (100 ml)

Absolute ethanol	70 ml
H_2O	30 ml

Ethidium Bromide Solution (10 mg/ ml)

EtBr	0.2 g
H_2O	20 ml

EtBr is dissolved carefully by stirring several hours, stored by wrapping in foil/in dark bottle at room temperature. Also a 10 mg ethidium bromide tablet can be dissolved in 1 ml distilled water

Fixer Solution (Silver Staining)

150 ml methanol, 36 ml acetic acid and 150 µl of 37 % formaldehyde is mixed and completed to 300 ml with dH_2O.

Fixing Solution (20%)(100 ml)(Coomassie Staining)

20 g trichloroacetic acid

After dissolving 20 g trichloroacetic acid in ¾ of volume of deionised water dH2O is added to 100 ml. It is kept at room temperature.

40% glucose

	100ml	300ml
Glucose	40g	120g

Glucose is added to ¾ required volume of water and allowed to dissolve.Then it is made up to required volume and filter sterilised through a 0.22 μm filter.

Galactose Oxidase Assay Solution (ABTS-HRP coupled assay)

D galactose	5.4 g
ABTS	22 mg
HRP (90 U/mg)	8.25 mg
100 mM NaPi, pH 7	50 ml

All reagents are mixed and dissolved. It is stored by wrapping in foil/in dark bottle at 0˚C.

Gel Buffer (GB) (pH 8.9) (100 ml)

Tris	18.5 g	(1.5 M)
SDS	0.4 g	(0.4%)

Volume is made up to 100 ml by dH_2O. pH is adjusted by HCl (~2 ml). It is filtered and stored at 4°C.

IPTG (1 M)

2,4 g IPTG is dissolved in 10 ml dH_2O, filter sterilized, dispensed in to aliquots and stored at -20°C.

Kanamycin (50 mg/ml stock)

1g kanamycin is dissolved in 20ml ddH_20, filter sterilised (0.2μm filter), and aliquoted into 1 ml and stored at -20^{o}C. It is used at a dilution of 1/1000 (500 μl in 500 ml) to give 50 μg/ml working concentration.

50 x LAC

	100ml	1L
Glycerol	25g	250g
Glucose	2.5g	25g
α-Lactose	10g	100g

All are dissolved in ¾ of volume of deionised water, it takes long time for lactose to dissolve, brief heating in microwave may be helpful.

LB (Luria-Bertani broth) Ampicillin Agar (per Liter)

10 g of NaCl
10 g of tryptone
5 g of yeast extract
20 g of agar (2%)

Deionized H_2O is added to a final volume of 1 litre. After adjusting pH to 7.2 with 10N NaOH the medium is autoclaved.1 ml of 50 mg/ml ampicillin stock is added when it cools down to nearly 55˚C to give a final concentration of 50 μg/ml. Then it is poured into the petri dishes (20-25 ml/ 90 mm plate).They are sealed with parafilm and stored at 4˚C.

LB * (Luria-Bertani) Medium (per Liter)

10 g of NaCl
10 g of tryptone
5 g of yeast extract

To reach a final volume of 1 liter deionized H_2O is added. After adjusting pH to 7.2 with 10 N NaOH the medium is autoclaved. If it is desired, 1 ml of 50 mg/ml ampicillin stock is added when it cools down to nearly 55˚C to give a final concentration of 50 μg/ml. It is stored at 4˚C.

* The acronym was actually intended to use for "Lysogeny Broth" which was misinterpreted as **"**Luria-Bertani" or "Luria Broth" as reported by its creator (Bertani,1951, 2004)

LB Tetracycline Agar (per Liter)

10 g NaCl
10 g tryptone
5 g yeast extract
20 g agar

Final volume is adjusted to 1 liter with distilled water. After adjusting pHto 7.0 with NaOH, the medium is autoclaved. 1.5 ml 10 mg/ml tetracycline is added when it cools to 55°C, and poured to petri dishes. The plates are covered with parafilm and stored in dark at 4°C.

Loading dye (6x)

0.2% bromophenol blue
0.2% xylene cyanol FF
60% glycerol
60 mM EDTA

Lysis Buffer (pH 8.0):

50 mM HEPES
25% (W/V) sucrose
5mM $MgCl_2$
1% Triton X100

After mixing the contents 2µl of omnicleave endonuclease and 1mg/ml lysozyme is freshly added before use in 40ml of lysis buffer.Additionally 1,6ml EDTA-free protease inhibitor (25x) which is prepared as 1 tablet/2 ml is also added.

$MgCl_2$ (2M)

19 g $MgCl_2$ is dissolved in 90 ml dH_2O and the volume is made up to 100 ml with dH_2O, sterilized by autoclaving.

1M $MgSO_4$

	100ml
$MgSO_4.7H_2O$	24.65g

It is dissolved in ¾ of volume of deionised water, then made up to required volume and filter sterilised through a 0.22µm filter.

NaOH (10 N)

20 g NaOH pellets are dissolved in 50 ml dH_2O and store in plastic bottle.

NE Buffer

0.3 M NaAC (pH 7.0)

20 x NPS

	100ml	1L	
$(NH4)_2SO_4$	6.6g	66g	0.5M
KH_2PO_4	13.6g	136g	1M
Na_2HPO_4	14.2g	142g	1M

All are dissolved in ¾ of volume of deionised water and pH is checked (~6.75). Deionised water is added to a final volume.

20 x NPSC

	100ml	1L	
NH_4Cl	5.35g	53.5g	1M
Na_2SO_4	3.22g	32.23g	0.1M
KH_2PO_4	6.8g	68g	0.5M
Na_2HPO_4	7.1g	71g	0.5M

All are dissolved in ¾ of volume of deionised water and and pH is checked (~6.75). Deionised water is added to a final volume.

PBS (1 x) (pH=7.4)

NaCl	8 g
KCl	0.2 g
Na_2HPO_4 (anhy)	1.44 g
KH_2PO_4	0.24 g

Distilled H_2O is added to 1000 ml. If necessary, pH is adjusted to 7.4 by phosphoric acid to make more acidic/ 1 M NaOH to make more alkali. It is sterilized by autoclaving.

Phosphate buffer (100 mM, pH 7.0) (2L)

NaH_2PO_4	12.17 g
Na_2HPO_4	17.3 g

From table ratio of NaH_2PO_4 to Na_2HPO_4 to give pH 7.0 was 39 g and 61 g

[NaH_2PO_4 = 0.39 x 2 x 0.1 x 156.01g =12.17 g
Na_2HPO_4 = 0.61 x 2 x 0.1 x 141.96 g= 17.3 g]

PIPES (20mM, pH 6.1)

6.048 g PIPES is dissolved in 1L dH_2O, it is noted that PIPES does not dissolve until close to pH 6.1

Preserving Solution (1L)

Acetic acid	100 ml
Glycerol	100 ml
dH_2O	800 ml

Pretreatment Solution (Silver Staining)

0.08 g sodiumthiosulfate ($Na_2S_2O_3$. 5 H_2O) is dissolved in 400 ml dH_2O and mixed by a glass rod and 8 ml is separated for further use in developing solution preparation.

Running buffer (10x) (pH 8.3) (1 L)

Tris	30 g	(0.25 M)
Glycine	144. g	(1.91 M)
SDS	10 g	(1%)

All the reagents are dissolved in dH_2O, volume is made up to 1L with dH_2O and stored at $4^{\circ}C$.

Sample buffer (4 x) (4ml)

20 % SDS	2.0 ml
1 M Tris-HCl, pH7	1.0 ml
Glycerol	1.0 ml
Bromophenol blau	few grains

All are mixed, stored at room temperature.

SDM (5x)

Sample buffer (4 x)	4.0 ml	(80%)
Mercaptoethanol	1.0 ml	(20%)

It is prapered just before use (freshly made).

SDS (20 %)

20g SDS is dissolved 100ml dH_2O carefully by wearing a mask. It is autoclaved and stored at room temperature.

SOC Medium (1L)

20 g bactotryptone
5 g bactoyeast extract
0.5 g NaCl

They are dissolved in 950 ml dH_2O. After addition of 10 ml 250 mM KCl, pH is adjusted to 7.0 with 5N NaOH (~0.2 ml). The volume is made up to 1L with dH_2O and the medium is autoclaved. When it is cooled to $60^{\circ}C$, 20 ml sterile 1 M glucose is added. Also, before use 5 ml of sterile 2 M $MgCl_2$ is added.

Sodium Acetate (3 M, pH = 5.2)

408,1 g of sodium acetate is dissolved in 800 ml distilled water. The pH of the solution is adjusted to 5.2 with glacial acetic acid. The volume is adjusted to 1 liter with distilled water. The solution is sterilized by autoclaving.

Solution I (Alkaline Lysis)

50 mM Glucose
25 mM Tris-HCl (pH 8.0)
10 mM EDTA

Solution II (Alkaline Lysis)

0.2 N NaOH
1% SDS

Solution III (Alkaline Lysis)

3 M NaAC (pH 4.8)

Solution I (Competent *E .coli* Preparation) (50 ml)

10 mM Tris-HCl (pH=8)
50 mM $CaCl_2$
20% Glycerol

500 µl 1M Tris-HCl (pH=8) and 2,5 ml 1 M $CaCl_2$ are added to 47 ml distilled and sterile water to reach a 50 ml final volume. Secondly, with the same concentrations to reach 7 ml final volume this time, 70 µl 1M Tris-HCl (pH=8), 350 µl 1M $CaCl_2$ and 1,4 ml glycerol are added to 5180 µl distilled and sterile water.

Silver Nitrate Solution (Silver Staining)

0.8 g silver nitrate is dissolved in 400 ml dH_20 and 300 µl 37 % formaldehyde is added.It is prepared freshly.

Stacking Gel Buffer (SGB) (100 ml, pH to 6.7)

Tris	5.1 g	(0.4 M)
SDS	0.4 g	(0.4%)

Volume is made up to 100 ml by dH_2O. pH is adjusted by HCl (~3 ml). It is filtered and stored at 4°C.

Stop Solution (Silver Staining)

200 ml methanol and 48 ml acetic acid are mixed and completed to avolume of 400 ml with dH_2O

0,5 M Succinate

	100ml	1L
Sodium succinate	8.1g	81.03g ($NaO_2CCH_2CH_2COONa.6H_2O$)

It is dissolved in ¾ of the required volume and made up to final volume with deionised water. Then it filter sterilised into desired aliquots.

TAE Buffer (50x) (1L)

242 g of Tris base is dissolved in 600 ml distilled water. The pH is adjusted to 8.0 with approximately 57 ml glacial acetic acid. Then 100 ml 0.5 M EDTA (pH 8.0) is added and the volume is adjusted to 1 liter.

10x TBS (pH 7.4)

500 mM Tris-HCl
1500mM NaCl

TBS-Tween 20 (0,1 %)

1ml Tween 20 is added into 1L TBS

Tetracycline (10 mg.ml^{-1} stock)
100 mg tetracycline is dissolved in 10 ml 50% ethanol. If possible (does not always dissolve fully) it is filter sterilised (0.2 μm filter), aliquoted and stored at -20^{o}C.

Tfb 1	**250ml**	**MW**
3 mM KAC	0.735g	98.14
100 mM $RbCl_2$	3.025g	121
10 mM $CaCl_2$	0.37g	147.02
50 mM MnCl	2.475g	197.91
Glycerol (15%)	37.5ml	

Before adding glycerol pH is adjusted to 5.8 with 0.2M acetic acid, then it is filter sterilised.

Tfb 2	**100ml**	**MW**
10 mM MOPS	0.21g	209.3
75 mM $CaCl_2$	1.1g	147.02
10 mM $RuCl_2$	0.12g	120.94
Glycerol (15%)	15ml	

Before the addition of glycerol pH is adjusted to pH 6.5 with KOH, then it is filter sterilised.

Trace Elements Solution (1L)
(1000x working concentration)

$FeSO_4.7H_2O$	1.0 g	
$ZnSO_4.7H_2O$	8.8 g	
$CuSO_4.7H_2O$	0.4 g	
$MnSO_4.4H_2O$	0.15 g	
$Na_2B_4O_7.10H_2O$	0.1 g	(disodiumtetraborate)
$(NH_4)_6MO_7O_{24}.4H_2O$	0.05 g	(ammonium molybdate)

Salts are added to 250 ml distilled water and swirled to gather the crystals and 0.2 ml concentrated HCl is added directly onto the crystals to dissolve them. Deionised water is added to make up to final volume, and dispensed in 500 ml volumes. It is autoclaved or filter sterilized and stored in the fridge.

Transfer Buffer

25mM Tris (pH 8.3)
192 mM glycine
20 % methanol v/v

Methanol is added following the adjustment of pH to 8.3.

Tris-HCl Buffer (1M, pH=8) 1L

121,1 g Tris base is dissolved in 800 ml of distilled water. pH is adjusted to the desired value with concentrated HCl. Then to achieve 1 liter volume certain amount of distilled water is added, sterilized by autoclaving.

Washing Solution (Silver Staining)

50 % Ethanol

1 L of 50 % ethanol was prepared.

ZY Component

	100 ml	200ml	1 l	2 L	(%)
Yeast extract	0.5g	1g	5g	10 g	(0.5 %)
Bactotryptose (tryptone)	1 g	2g	10g	20 g	(1 %)

All are mixed and autoclaved at 121^{o}C 15psi for 20 minutes.

8ZY Component

	100ml	200ml	1L	2L	
Yeast extract	4.0g	8.0g	40g	80g	(4%)
Bactotryptone	8.0g	16.0g	80g	160g	(8%)

All are mixed and autoclaved at 121^{o}C15psi for 20 minutes

ZYP-0.8G Medium

	50 ml	100ml	400ml	
ZY Component	46.5ml	93ml	372ml	
1M MgSO4	100µl	200µl	800µl	1mM
40% Glucose	1ml	2ml	8ml	0.8%
20 x NPS	2.5ml	5ml	20ml	1 x

$MgSO_4$ is added before adding NPS to avoid precipitate. Media is kept sterile and appropriate antibiotic is added.

8ZYM– 4 x LAC + 25mM Succinate Medium

	50ml	400ml
8ZY Component	46.5ml	372ml
1M $MgSO_4$	100µl	800µl
50 x LAC	4ml	32ml
20 x NPSC	2.5ml	20ml
Trace Elements	50µl	400µl
0.5M Succinate	2.5ml	20ml

$MgSO_4$ is added before adding NPSC to avoid precipitate. Media is kept sterile and appropriate antibiotic is added.

APPENDIX B

NUCLEOTIDE AND AMINO ACID SEQUENCE OF *FUSARIUM GRAMINEARUM GAO*

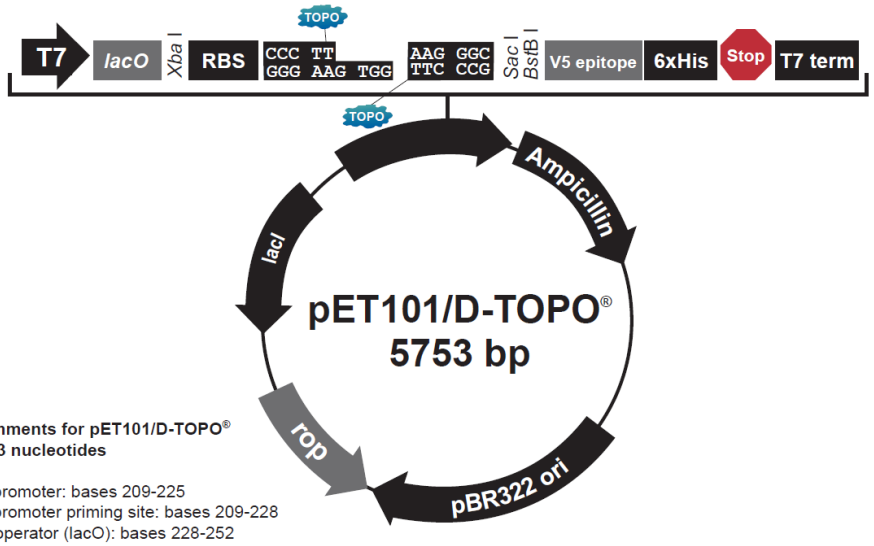
Fusarium graminearum gao with the encoded amino acid sequence

```
atgaaacacctttaacactcgctctttgcttcagcagcatcaatgctgttgctgtcaccgtc
M  K  H  L  L  T  L  A  L  C  F  S  S  I  N  A  V  A  V  T  V
cctcacaaggccgtaggaactggaattcctgaagggagtcttcagttcctgagccttcga
P  H  K  A  V  G  T  G  I  P  E  G  S  L  Q  F  L  S  L  R
gcctcagcacctatcggaagcgccatttctcgcaacaactgggccgtcacttgcgacagt
A  S  A  P  I  G  S  A  I  S  R  N  N  W  A  V  T  C  D  S
gcacagtcgggaaatgaatgcaacaaggccattgatggcaacaaggatacctttggcac
A  Q  S  G  N  E  C  N  K  A  I  D  G  N  K  D  T  F  W  H
acattctatggcgccaacggggatccaaagccccctcacacatacacgattgacatgaag
T  F  Y  G  A  N  G  D  P  K  P  P  H  T  Y  T  I  D  M  K
acaactcagaacgtcaacggcttgtctatgctgcctcgacaggatggtaaccaaaacggc
T  T  Q  N  V  N  G  L  S  M  L  P  R  Q  D  G  N  Q  N  G
tggatcggtcgccatgaggtttatctaagctcagatggcacaaactggggcagccctgtt
W  I  G  R  H  E  V  Y  L  S  S  D  G  T  N  W  G  S  P  V
gcgtcaggtagttggttcgccgactctactacaaaatactccaactttgaaactcgccct
A  S  G  S  W  F  A  D  S  T  T  K  Y  S  N  F  E  T  R  P
gctcgctatgttcgtcttgtcgctatcactgaagcgaatggccagccttggactagcatt
A  R  Y  V  R  L  V  A  I  T  E  A  N  G  Q  P  W  T  S  I
gcagagatcaacgtcttccaagctagttcttacacagccccccagcctggtcttggacgc
A  E  I  N  V  F  Q  A  S  S  Y  T  A  P  Q  P  G  L  G  R
tggggtccgactattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcg
W  G  P  T  I  D  L  P  I  V  P  A  A  A  A  I  E  P  T  S
ggacgagtccttatgtggtcttcatatcgcaatgatgcatttggaggatcccctggtggt
G  R  V  L  M  W  S  S  Y  R  N  D  A  F  G  G  S  P  G  G
atcactttgacgtcttcctgggatccatccactggtattgtttccgaccgcactgtgaca
I  T  L  T  S  S  W  D  P  S  T  G  I  V  S  D  R  T  V  T
gtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgta
V  T  K  H  D  M  F  C  P  G  I  S  M  D  G  N  G  Q  I  V
gtcacaggtggcaacgatgccaagaagaccagtttgtatgattcatctagcgatagctgg
V  T  G  G  N  D  A  K  K  T  S  L  Y  D  S  S  S  D  S  W
atcccgggacctgacatgcaagtggctcgtgggtatcagtcatcagctaccatgtcagac
I  P  G  P  D  M  Q  V  A  R  G  Y  Q  S  S  A  T  M  S  D
ggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaa
G  R  V  F  T  I  G  G  S  W  S  G  G  V  F  E  K  N  G  E
gtctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatg
V  Y  S  P  S  S  K  T  W  T  S  L  P  N  A  K  V  N  P  M
ttgacggctgacaagcaaggattgtaccgttcagacaaccacgcgtggctctttggatgg
L  T  A  D  K  Q  G  L  Y  R  S  D  N  H  A  W  L  F  G  W
aagaagggttcggtgttccaagcgggacctagcacagccatgaactggtactataccagt
K  K  G  S  V  F  Q  A  G  P  S  T  A  M  N  W  Y  Y  T  S
ggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgat
G  S  G  D  V  K  S  A  G  K  R  Q  S  N  R  G  V  A  P  D
gccatgtgcggaaacgctgtcatgtacgacgccgttaaaggaaagatcctgacctttggc
```

```
A  M  C  G  N  A  V  M  Y  D  A  V  K  G  K  I  L  T  F  G
ggctccccagattatcaagactctgacgccacaaccaacgcccacatcatcaccctcggt
G  S  P  D  Y  Q  D  S  D  A  T  T  N  A  H  I  I  T  L  G
gaacccggaacatctcccaacactgtctttgctagcaatgggttgtactttgcccgaacg
E  P  G  T  S  P  N  T  V  F  A  S  N  G  L  Y  F  A  R  T
tttcacacctctgttgttcttccagacggaagcacgtttattacaggaggccaacgacgt
F  H  T  S  V  V  L  P  D  G  S  T  F  I  T  G  G  Q  R  R
ggaattccgttcgaggattcaaccccggtatttacacctgagatctacgtccctgaacaa
G  I  P  F  E  D  S  T  P  V  F  T  P  E  I  Y  V  P  E  Q
gacactttctacaagcagaaccccaactccattgttcgcgtctaccatagcatttccctt
D  T  F  Y  K  Q  N  P  N  S  I  V  R  V  Y  H  S  I  S  L
ttgttacctgatggcagggtatttaacggtggtggtggtctttgtggcgattgtaccacg
L  L  P  D  G  R  V  F  N  G  G  G  G  L  C  G  D  C  T  T
aatcatttcgacgcgcaaatctttacgccaaactatctttacaatagcaacggcaatctc
N  H  F  D  A  Q  I  F  T  P  N  Y  L  Y  N  S  N  G  N  L
gcgacacgtcccaagattaccagaacctctacacagagcgtcaaggtcggtggcagaatt
A  T  R  P  K  I  T  R  T  S  T  Q  S  V  K  V  G  G  R  I
acaatctcgacggattcttcgattagcaaggcgtcgttgattcgctatggtacagcgaca
T  I  S  T  D  S  S  I  S  K  A  S  L  I  R  Y  G  T  A  T
cacacggttaatactgaccagcgccgcattcccctgactctgacaaacaatggaggaaat
H  T  V  N  T  D  Q  R  R  I  P  L  T  L  T  N  N  G  G  N
agctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttg
S  Y  S  F  Q  V  P  S  D  S  G  V  A  L  P  G  Y  W  M  L
ttcgtgatgaactcggccggtgttcctagtgtggcttcgacgattcgcgttactcagtga
F  V  M  N  S  A  G  V  P  S  V  A  S  T  I  R  V  T  Q  -
```

APPENDIX C

MAPS AND SEQUENCE DATA OF THE VECTORS AND CONSTRUCTS USED AND AMINO ACID SEQUENCE COMPARISONS WITH GAO



Comments for pET101/D-TOPO®
5753 nucleotides

T7 promoter: bases 209-225
T7 promoter priming site: bases 209-228
lac operator (lacO): bases 228-252
Ribosome binding site (RBS): bases 282-288, 292-296
TOPO® cloning site (directional): bases 297-310
V5 epitope: bases 333-374
Polyhistidine (6xHis) region: bases 384-401
T7 reverse priming site: bases 455-474
T7 transcription termination region: bases 416-544
bla promoter: bases 845-943
Ampicillin (*bla*) resistance gene (ORF): bases 944-1804
pBR322 origin: bases 1949-2622
ROP ORF: bases 2990-3181 (complementary strand)
lacI ORF: bases 4493-5584 (complementary strand)

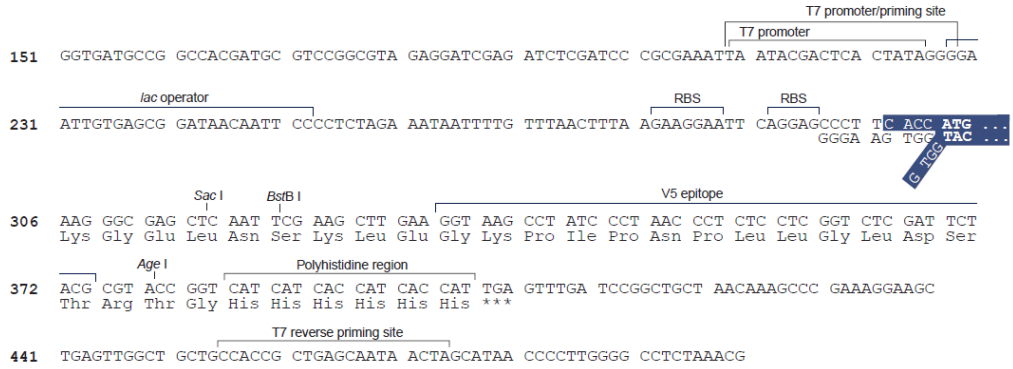


Figure C.1 The map of pET101/D-TOPO vector and multiple cloning site (Invitrogen).

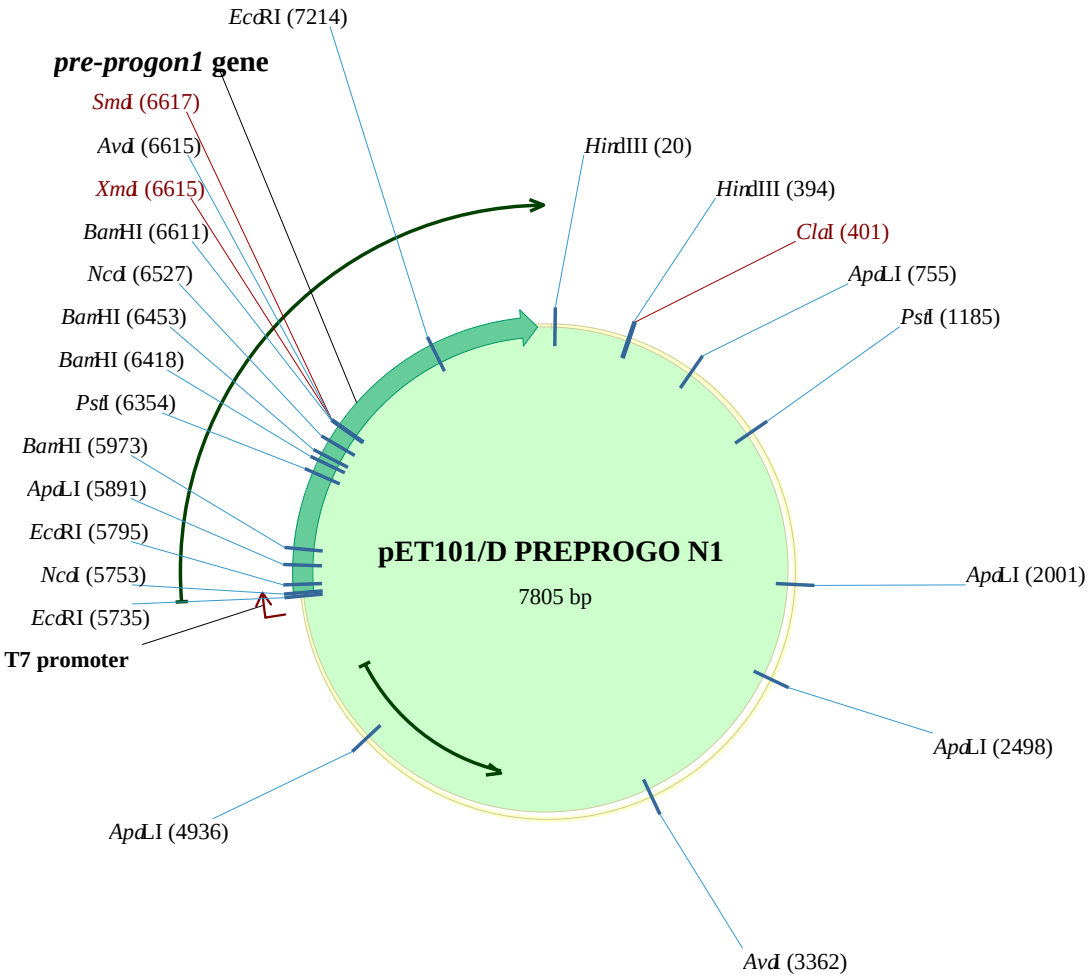


Figure C.2 The map of pET101/D Pre-ProGON1 vector.

***pre-progon1* gene sequence**

atggttgcagttaccgttcctcacaaggccgtaggaactggaattcctgaagggagtcttcagttcctgagccttcgagcctcagcacc
tatcggaagcgccatttctcgcaacaactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaacaaggccattgatgg
caacaaggatacctttggcacacattctatggcgccaacggggatccaaagccccctcacacatacacgattgacatgaagacaact
cagaacgtcaacggcttgtctatgctgcctcgacaggatggtaaccaaaacggctggatcggtcgccatgaggtttatctaagctcag
atggcacaaactggggcagccctgttgcgtcaggtagttggttcgccgactctactacaaaatactccaactttgaaactcgccctgct
cgctatgttcgtcttgtcgctatcactgaagcgaatggccagccttggactagcattgcagagatcaacgtcttccaagctagttcttaca
cagccccccagcctggtcttggacgctggggtccgactattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcgggg
acgagtccttatgtggtcttcatatcgcaatgatgcatttggaggatcccctggtggtatcactttgacgtcttcctgggatccatccactg
gtattgtttccgaccgcactgtgacagtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgtagtcac
aggtggcaacgatgccaagaagaccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgcaagtggctcgtgg
gtatcagtcatcagctaccatgtcagacggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaag
tctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagcaaggattgtaccgt
tcagacaaccacgcgtggctctttggatggaagaagggttcggtgttccaagcgggacctagcacagccatgaactggtactatacc
agtggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgatgccatgtgcggaaacgctgtcatg
tacgacgccgttaaaggaaagatcctgacctttggcggctccccagattatcaagactctgacgccacaaccaacgcccacatcatca
ccctcggtgaacccggaacatctcccaacactgtctttgctagcaatgggttgtactttgcccgaacgtttcacacctctgttgttcttcca
gacggaagcacgtttattacaggaggccaacgacgtggaattccgttcgaggattcaaccccggtatttacacctgagatctacgtcc
ctgaacaagacactttctacaagcagaaccccaactccattgttcgcgtctaccatagcatttccctttgttacctgatggcagggtattt
aacggtggtggtggtctttgtggcgattgtaccacgaatcatttcgacgcgcaaatctttacgccaaactatctttacaatagcaacggc
aatctcgcgacacgtcccaagattaccagaacctctacacagagcgtcaaggtcggtggcagaattacaatctcgacggattcttcga
ttagcaaggcgtcgttgattcgctatggtacagcgacacacacggttaatactgaccagcgccgcattcccctgactctgacaaacaat
ggaggaaatagctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatgaactcggccggtgt
tcctagtgtggcttcgacgattcgcgttactcagggcggtggcggttcttggagccatccgcagtttgagaaatgatgagcggccgcc
agc

The *pre-progon1* gene has a 4 x glycine + 1 x serine linker and an 8 amino acid *Strep*-tag II at its C-terminal.

Amino acid sequence alignment of GAO and Pre-proGON1

CLUSTAL W (1.82) multiple sequence alignment

```
GAO             MKHLLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 60
PREPREPROGON1   ---------------MVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 45
                               ****************************************

GAO             SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 120
PREPREPROGON1   SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 105
                ************************************************************

GAO             GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 180
PREPREPROGON1   GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 165
                ************************************************************

GAO             IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 240
PREPROGON1      IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 225
                ************************************************************

GAO             GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 300
PREPROGON1      GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 285
                ************************************************************

GAO             WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 360
PREPROGON1      WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 345
                ************************************************************

GAO             MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 420
PREPROGON1      MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 405
                ************************************************************

GAO             DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 480
PREPROGON1      DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 465
                ************************************************************

GAO             TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 540
PREPROGON1      TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 525
                ************************************************************

GAO             LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 600
PREPROGON1      LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 585
                ************************************************************

GAO             ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 660
PREPROGON1      ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 645
                ************************************************************

GAO             LFVMNSAGVPSVASTIRVTQ------------------- 680
PREPROGON1      LFVMNSAGVPSVASTIRVTQGGGGGSWSHPQFEK--AAAS 682
                ********************
```

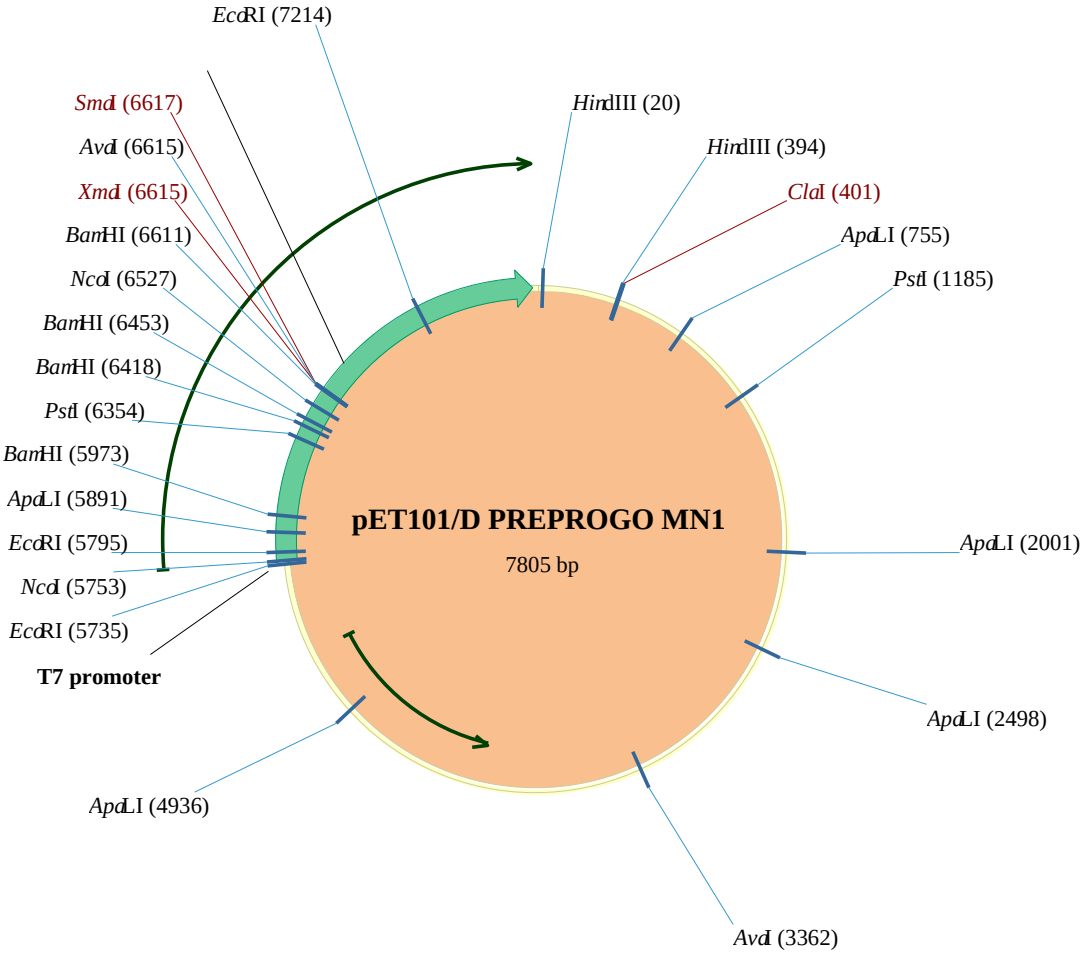


Figure C.3 The map of pET101/D Pre-ProGOMN1 vector.

***pre-progomn1* gene sequence**

atggttgcagttaccgttcctcacaaggccgtaggaactggaattcctgaagggagtcttcagttcctgagccttcgagcctcagcacc
tatcggaagcgccattcctcgcaacaactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaacaaggccattgatgg
caacaaggataccttttggcacacattctatggcgccaacggggatccaaagccccctcacacatacacgattgacatgaagacaact
cagaacgtcaacggcttgtctgtgctgcctcgacaggatggtaaccaaaacggctggatcggtcgccatgaggtttatctaagctcag
atggcacaaactggggcagccctgttgcgtcaggtagttggttcgccgactctactacaaaatactccaactttgaaactcgccctgct
cgctatgttcgtcttgtcgctatcactgaagcgaatggccagccctggactagcattgcagagatcaacgtcttccaagctagttcttaca
cagccccccagcctggtcttggacgctggggtccgactattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcgggg
acgagtccttatgtggtcttcatatcgcaatgatgcatttgaaggatcccctggtggtatcactttgacgtcttcctgggatccatccactg
gtattgtttccgaccgcactgtgacagtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgtagtcac
aggtggcaacgatgccaagaagaccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgcaagtggctcgtgg
gtatcagtcatcagctaccatgtcagacggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaag
tctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagcaaggattgtaccgt
tcagacaaccacgcgtggctctttggatggaagaagggttcggtgttccaagcgggacctagcacagccatgaactggtactatacc
agtggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgatgccatgtgcggaaacgctgtcatg
tacgacgccgttaaaggaaagatcctgacctttggcggctccccagattatcaagactctgacgccacaaccaacgcccacatcatca
ccctcggtgaacccggaacatctcccaacactgtctttgctagcaatgggttgtactttgcccgaacgtttcacacctctgttgttcttcca
gacggaagcacgtttattacaggaggccaacgacgtggaattccgttcgaggattcaaccccggtatttacacctgagatctacgtcc
ctgaacaagacactttctacaagcagaaccccaactccattgttcgcgcttaccatagcatttccctttgttacctgatggcagggtattt
aacggtggtggtggtctttgtggcgattgtaccacgaatcatttcgacgcgcaaatctttacgccaaactatctttacgatagcaacggc
aatctcgcgacacgtcccaagattaccagaacctctacacagagcgtcaaggtcggtggcagaattacaatctcgacggattcttcga
ttagcaaggcgtcgttgattcgctatggtacagcgacacacacggttaatactgaccagcgccgcattcccctgactctgacaaacaat
ggaggaaatagctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatgaactcggccggtgt
tcctagtgtggcttcgacgattcgcgttactcagggcggtggcggttcttggagccatccgcagtttgagaaatgatgagcggccgcc
agc

The *pre-progomn1* gene has a 4 x glycine + 1 x serine linker and an 8 amino acid *Strep*-tag II at its C-terminal.

Amino acid sequence alignment of GAO and Pre-proGOMN1

CLUSTAL W (1.82) multiple sequence alignment

```
GAO             MKHLLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 60
PREPROGOMN1     ---------------MVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAIPRNNWAVTCD 45
                               **********************************.*********

GAO             SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 120
PREPROGOMN1     SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQN 105
                **************************************************:*********

GAO             GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 180
PREPROGOMN1     GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 165
                ************************************************************

GAO             IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 240
PREPROGOMN1     IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPG 225
                ************************************************** ****

GAO             GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 300
PREPROGOMN1     GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 285
                ************************************************************

GAO             WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 360
PREPROGOMN1     WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 345
                ************************************************************

GAO             MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 420
PREPROGOMN1     MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 405
                ************************************************************

GAO             DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 480
PREPROGOMN1     DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 465
                ************************************************************

GAO             TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 540
PREPROGOMN1     TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYHSIS 525
                ********************************************************.*****

GAO             LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 600
PREPROGOMN1     LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSNGNLATRPKITRTSTQSVKVGGR 585
                **********************************:*************************

GAO             ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 660
PREPROGOMN1     ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 645
                ************************************************************

GAO             LFVMNSAGVPSVASTIRVTQ------------------- 680
PREPROGOMN1     LFVMNSAGVPSVASTIRVTQGGGGSWSHPQFEK--AAAS 682
                ********************
```

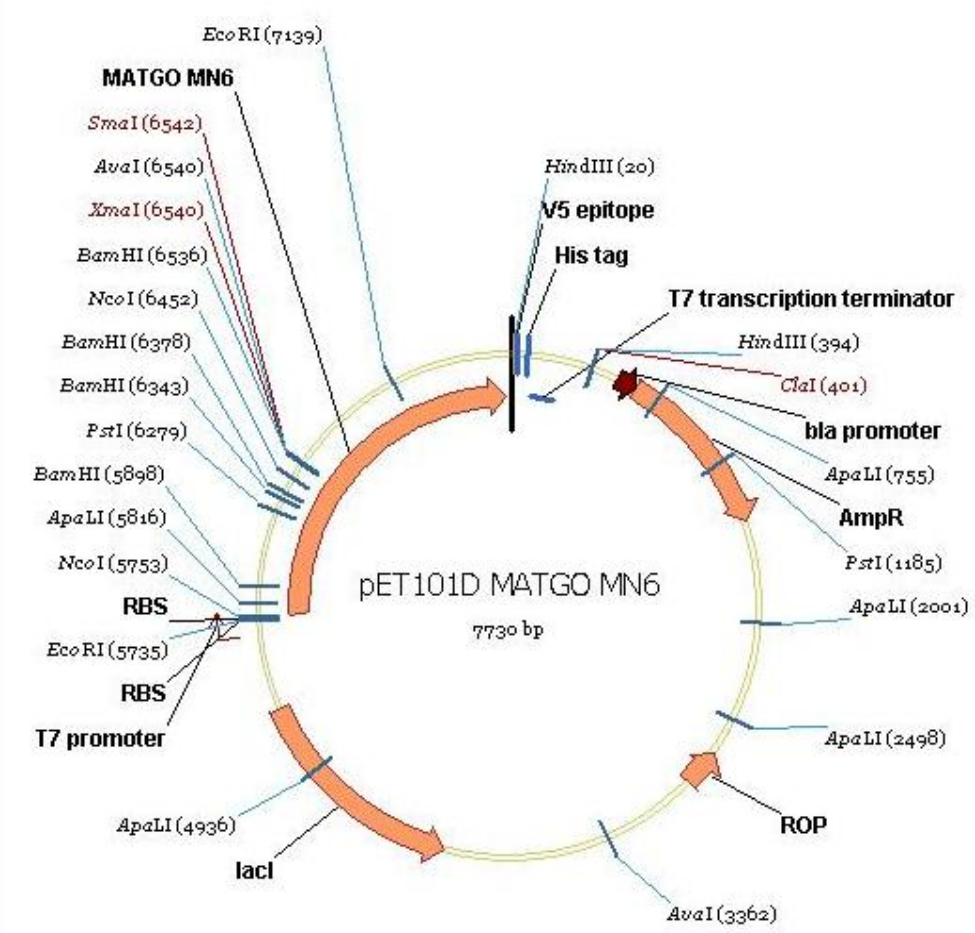


Figure C.4 The map of pET101/D MatGOMN1 vector

***matgomn6* gene sequence**

atggcatctgcacctattggaagcgccattcctcgcaacaactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaac
aaggccattgatggcaacaaggatacctttggcacacattctatggcgccaacggggatccaaagccccctcacacatacacgattg
acatgaagacaactcagaacgtcaacggcttgtctgtgctgcctcgacaggatggtaaccaaaacggctggatcggtcgccatgagg
tttatctaagctcagatggcacaaactggggcagccctgttgcgtcaggtagttggttcgccgactctactacaaaatactccaactttga
aactcgccctgctcgctatgttcgtcttgtcgctatcactgaagcgaatggccagccctggactagcattgcagagatcaacgtcttcca
agctagttcttacacagccccccagcctggtcttggacgctggggtccgactattgacttaccgattgttcctgcggctgcagcaattga
accgacatcgggacgagtccttatgtggtcttcatatcgcaatgatgcatttgaaggatcccctggtggtatcactttgacgtcttcctgg
gatccatccactggtattgtttccgaccgcactgtgacagtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtc
agatcgtagtcacaggtggcaacgatgccaagaagaccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgca
agtggctcgtgggtatcagtcatcagctaccatgtcagacggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaa
gaatggcgaagtctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagca
aggattgtaccgttcagacaaccacgcgtggctctttggatggaagaagggttcggtgttccaagcgggacctagcacagccatgaa
ctggtactataccagtggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgatgccatgtgcgg
aaacgctgtcatgtacgacgccgttaaaggaaagatcctgacctttggcggctccccagattatcaagactctgacgccacaaccaac
gcccacatcatcaccctcggtgaacccggaacatctcccaacactgtctttgctagcaatgggttgtactttgcccgaacgtttcacacc
tctgttgttcttccagacggaagcacgtttattacaggaggccaacgacgtggaattccgttcgaggattcaaccccggtatttacacct
gagatctacgtccctgaacaagacactttctacaagcagaaccccaactccattgttcgcgcttaccatagcatttccctttgttacctga
tggcagggtatttaacggtggtggtggtctttgtggcgattgtaccacgaatcatttcgacgcgcaaatctttacgccaaactatctttac
gatagcaacggcaatctcgcgacacgtcccaagattaccagaacctctacacagagcgtcaaggtcggtggcagaattacaatctcg
acggattcttcgattagcaaggcgtcgttgattcgctatggtacagcgacacacacggttaatactgaccagcgccgcattcccctgac
tctgacaaacaatggaggaaatagctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatga
actcggccggtgttcctagtgtggcttcgacgattcgcgttactcagggcggtggcggttcttggagccatccgcagtttgagaaatga
tgagcggccgccagc

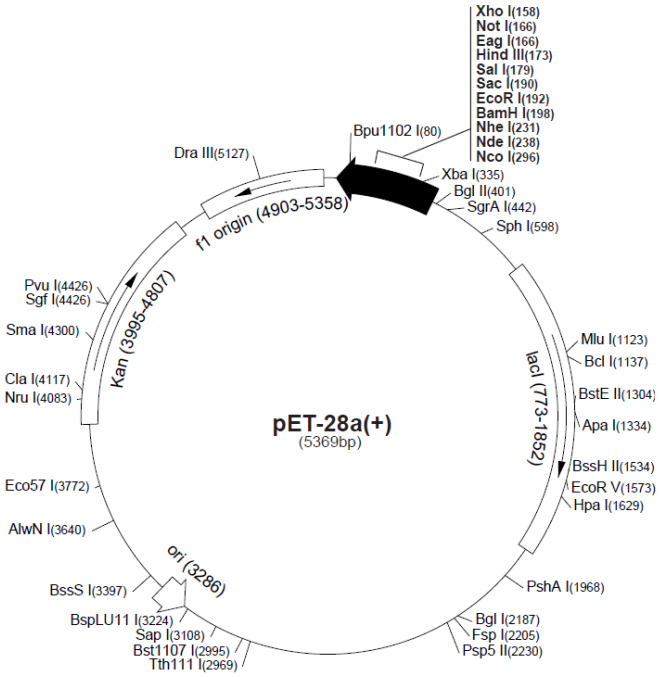
The *matgomn6* gene has a 4 x glycine + 1 x serine linker then 8 amino acid strep-tag II at C-terminal

Amino acid sequence alignment of GAO and MatGOMN6

```
CLUSTAL O(1.2.0) multiple sequence alignment


GAO           MKHLLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD
MATGOMN6      ----------------------------------------MASAPIGSAIPRNNWAVTCD
                                                       ********* *********

GAO           SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN
MATGOMN6      SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQN
              ************************************************:*********

GAO           GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS
MATGOMN6      GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS
              ************************************************************

GAO           IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG
MATGOMN6      IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPG
              ************************************************** ****

GAO           GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS
MATGOMN6      GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS
              ************************************************************

GAO           WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP
MATGOMN6      WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP
              ************************************************************

GAO           MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP
MATGOMN6      MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP
              ************************************************************

GAO           DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR
MATGOMN6      DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR
              ************************************************************

GAO           TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS
MATGOMN6      TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYHSIS
              ********************************************************.*****

GAO           LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR
MATGOMN6      LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSNGNLATRPKITRTSTQSVKVGGR
              ************************************:************************

GAO           ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM
MATGOMN6      ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM
              ************************************************************

GAO           LFVMNSAGVPSVASTIRVTQ--------------------
MATGOMN6      LFVMNSAGVPSVASTIRVTQGGGGSWSHPQFEK--AAAS
                ********************
```

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites (*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.



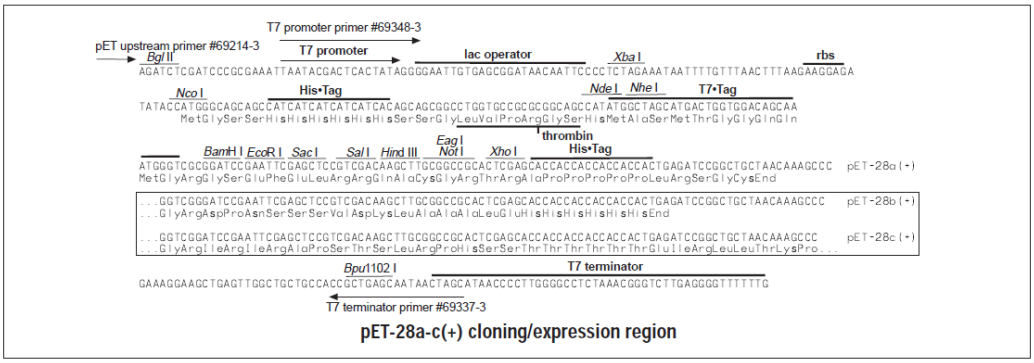


Figure C.5 The map of pET28a-c(+) vector and cloning /expression region (Novagen).

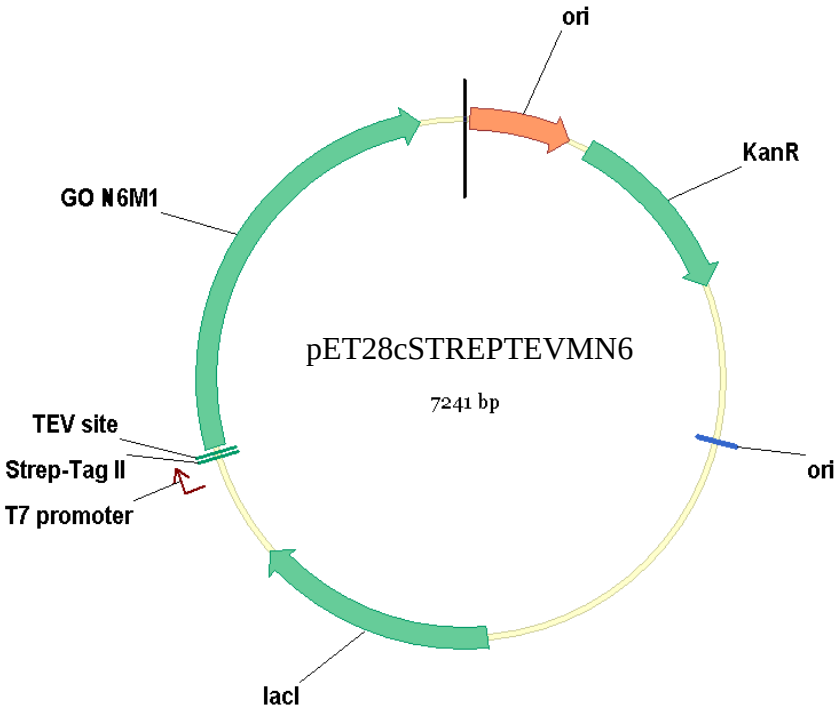


Figure C.6 The map of pET28c-StrepTEVMN6 vector.

StrepTEVMN6 nucleotide sequence on pET28c-StrepTEVMN6 vector

atgggt**tggagccatccgcagtttgagaaa**gaaaacctgtacttccagggtgcatctgcacctattggaagcgccattcctcgcaac
aactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaacaaggccattgatggcaacaaggatacctttggcacaca
ttctatggcgccaacggggatccaaagccccctcacacatacacgattgacatgaagacaactcagaacgtcaacggcttgtctgtgc
tgcctcgacaggatggtaaccaaaacggctggatcggtcgccatgaggtttatctaagctcagatggcacaaactggggcagccctg
ttgcgtcaggtagttggttcgccgactctactacaaaatactccaactttgaaactcgccctgctcgctatgttcgtcttgtcgctatcactg
aagcgaatggccagccctggactagcattgcagagatcaacgtcttccaagctagttcttacacagccccccagcctggtcttggacg
ctggggtccgactattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcgggacgagtccttatgtggtcttcatatcg
caatgatgcatttgaaggatcccctggtggtatcactttgacgtcttcctgggatccatccactggtattgtttccgaccgcactgtgaca
gtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgtagtcacaggtggcaacgatgccaagaaga
ccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgcaagtggctcgtgggtatcagtcatcagctaccatgtcag
acggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaagtctatagcccatcttcaaagacatgg
acgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagcaaggattgtaccgttcagacaaccacgcgtggctctttg
gatggaagaagggttcggtgttccaagcgggacctagcacagccatgaactggtactataccagtggaagtggtgatgtgaagtcag
ccggaaaacgccagtctaaccgtggtgtagcccctgatgccatgtgcggaaacgctgtcatgtacgacgccgttaaaggaaagatcc
tgacctttggcggctccccagattatcaagactctgacgccacaaccaacgcccacatcatcaccctcggtgaacccggaacatctcc
caacactgtctttgctagcaatgggttgtactttgcccgaacgtttcacacctctgttgttcttccagacggaagcacgtttattacaggag
gccaacgacgtggaattccgttcgaggattcaaccccggtatttacacctgagatctacgtccctgaacaagacactttctacaagcag
aaccccaactccattgttcgcgcttaccatagcatttccctttgttacctgatggcagggtatttaacggtggtggtggtctttgtggcgat
tgtaccacgaatcatttcgacgcgcaaatctttacgccaaactatctttacgatagcaacggcaatctcgcgacacgtcccaagattacc
agaacctctacacagagcgtcaaggtcggtggcagaattacaatctcgacggattcttcgattagcaaggcgtcgttgattcgctatgg
tacagcgacacacacggttaatactgaccagcgccgcattcccctgactctgacaaacaatggaggaaatagctattctttccaagttc
ctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatgaactcggccggtgttcctagtgtggcttcgacgattcgcgtt
actcag

WT MatGOMN6 sequence(starting from gca following TEV cleavage site)
Strep tag
TEV cleavage site
N6 mutations
M mutations

APPENDIX D

NUCLEOTIDE AND AMINO ACID SEQUENCE ALIGNMENTS OF MUTATIONAL VARIANTS

PROGON1: 8-AA PRE-PEPTIDE DELETION

PROGON1 & PRE-PROGON1 DNA Sequence Alignment/GAOstr Primer

```
PROGON1          ---------------------------------------------TGGGGCGGTAAATTC 15
PRE-PROGON1      CTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCC 5700
                                                             **     ** * *

PROGON1          CCTCTAGAA-TAATTTTGTTTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGG--- 71
PRE-PROGON1      CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGGTTG 5760
                 ********* ********************************************

PROGON1          ---------------------CCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC 110
PRE-PROGON1      CAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC 5820
                                      ************************************

PROGON1          TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA 170
PRE-PROGON1      TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA 5880
                 ************************************************************

PROGON1          CTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATA 230
PRE-PROGON1      CTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATA 5940
                 ************************************************************

PROGON1          CCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGA 290
PRE-PROGON1      CCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGA 6000
                 ************************************************************

PROGON1          TTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTA 350
PRE-PROGON1      TTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTA 6060
                 ************************************************************

PROGON1          ACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGG 410
PRE-PROGON1      ACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGG 6120
                 ************************************************************

PROGON1          GCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTCCAACTTTG 470
PRE-PROGON1      GCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTCCAACTTTG 6180
                 ************************************************************

PROGON1          AAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGCCAGCCTT 530
PRE-PROGON1      AAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGCCAGCCTT 6240
                 ************************************************************

PROGON1          GGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCCCCAGCCTG 590
PRE-PROGON1      GGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCCCCAGCCTG 6300
                 ************************************************************

PROGON1          GTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGCAGCAATTG 650
PRE-PROGON1      GTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGCAGCAATTG 6360
                 ************************************************************

PROGON1          AACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATTTGGAGGAT 710
PRE-PROGON1      AACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATTTGGAGGAT 6420
                 ************************************************************
```

```
PROGON1          CCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTA------------ 758
PRE-PROGON1      CCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGTTTCCGACC 6480
                 *********************************************

PROGON1          ------------------------------------------------------------
PRE-PROGON1      GCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGATGGTAACG 6540
```

PROGON1 & PRE-PROGON1 Amino acid sequence alignment

GeneWise (Sample; protein sequence)

```
PROGON1            12 EGIQEPFTM--------AVGTGIPEGSLQFLSLRASAPIGSAISRNNWA
                      EGIQEPFTM        AVGTGIPEGSLQFLSLRASAPIGSAISRNNWA
                      EGIQEPFTMVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWA
PRE-PROGON1      5730 ggacgctaagggagccagggagacggacctcaccgtgcagagatcaatg
                      agtaactcttctctcaactgcgtcaggtattgtgcccctggctcgaagc
                      aatggcccgtatcttcgcaatattagttgcgctacaatcaccttcccgc


PROGON1            53 VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
                      VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
                      VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
PRE-PROGO        5877 gatgagctgagtaagaggaagattcattggaggcacccataagaaaaca
                      tcgagcacgaagaactagaaactgactagcagacaccacactataccaa
                      ctcctaggataccgcttccgtctgcactcccgtagctcacgtcggatgc


PROGON1           102 VNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
                      VNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
                      VNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
PRE-PROGO        6024 gagttaccccggacagtagccggtcatggaatgacggtgattggtaaat
                      tagtcttcgaagaaaggtggaatatgcagcagggctccgggtcacccaa
                      cccgtggtagttcaccgctctgttacatcacgccttgattgcccttaac


PROGON1           151 SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
                      SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
                      SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
PRE-PRO          6171 tatgaccgctgccggaaggagcctaaaggaagtcgattagcccgcgctg
                      catacgccgatgttctcacagacgcgtcatattacgcacccacgtgggg
                      cctatcttcttttctctagtcgtgtctagccccatttcaccgtttacgt


PROGON1           200 PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTSSWDPST
                      PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTSSWDPST
                      PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTSSWDPST
PRE-P            6318 caagtcagcggggagcatgcgcattttcaggtggtcggaatatttgcta
                      cctatcttccccctacccggtttgccagaactggccggtctcccgaccc
                      gttcagtttgtaatagagaactggtatcttataactttctggtcgtact


PROGON1           249 G
                      G
                      G
PRE-PROGON1      6465 g
                      g
                      t
```

PROGON1 & PRE-PROGON1 DNA Sequence Alignment/BRCK1F primer

```
PROGON1          -------------------------------------CGTT-CGT-AGGTAGTTGGTTCG 21
PRE-PROGON1      TTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCG 480
                                                       *** *** **************

PROGON1          C-GACTCTACTACAAAATACTCCAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTG 80
PRE-PROGON1      CCGACTCTACTACAAAATACTCCAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTG 540
                 * ************************************************************

PROGON1          TCGCTATCACTGAAGCGAATGGCCAGCCTTGGACTAGCATTGCAGAGATCAACGTCTTCC 140
PRE-PROGON1      TCGCTATCACTGAAGCGAATGGCCAGCCTTGGACTAGCATTGCAGAGATCAACGTCTTCC 600
                 ************************************************************

PROGON1          AAGCTAGTTCTTACACAGCCCCCCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACT 200
PRE-PROGON1      AAGCTAGTTCTTACACAGCCCCCCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACT 660
                 ************************************************************

PROGON1          TACCGATTGTTCCTGCGGCTGCAGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGT 260
PRE-PROGON1      TACCGATTGTTCCTGCGGCTGCAGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGT 720
                 ************************************************************

PROGON1          CTTCATATCGCAATGATGCATTTGGAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCT 320
PRE-PROGON1      CTTCATATCGCAATGATGCATTTGGAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCT 780
                 ************************************************************

PROGON1          GGGATCCATCCACTGGTATTGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGT 380
PRE-PROGON1      GGGATCCATCCACTGGTATTGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGT 840
                 ************************************************************

PROGON1          TCTGCCCTGGTATCTCCATGGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATG 440
PRE-PROGON1      TCTGCCCTGGTATCTCCATGGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATG 900
                 ************************************************************

PROGON1          CCAAGAAGACCAGTTTGTATGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGC 500
PRE-PROGON1      CCAAGAAGACCAGTTTGTATGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGC 960
                 ************************************************************

PROGON1          AAGTGGCTCGTGGGTATCAGTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTG 560
PRE-PROGON1      AAGTGGCTCGTGGGTATCAGTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTG 1020
                 ************************************************************

PROGON1          GAGGCTCCTGGAGCGGTGGCGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAA 620
PRE-PROGON1      GAGGCTCCTGGAGCGGTGGCGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAA 1080
                 ************************************************************

PROGON1          AGACATGGACGTCCCTACCCAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAG 680
PRE-PROGON1      AGACATGGACGTCCCTACCCAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAG 1140
                 ************************************************************

PROGON1          GATTGTACCGTTCAGACAACCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCC 740
PRE-PROGON1      GATTGTACCGTTCAGACAACCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCC 1200
                 ************************************************************

PROGON1          AAGCGGGACCTAGCACAGCCATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGT 800
PRE-PROGON1      AAGCGGGACCTAGCACAGCCATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGT 1260
                 ************************************************************

PROGON1          CAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTG 860
PRE-PROGON1      CAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTG 1320
                 ************************************************************

PROGON1          TCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAG 920
PRE-PROGON1      TCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAG 1380
                 ************************************************************

PROGON1          ACTCT-------------------------------------------------------- 925
PRE-PROGON1      ACTCTGACGCCACAACCAACGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCA 1440
                 *****
```

PROGON1 & PRE-PROGON1 Amino acid sequence alignment

```
PRE-PROGON1      134 DSTTKYSNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQP
                     DSTTKYSNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQP
                     DSTTKYSNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQP
PROGON1           23 gtaaattatgaccgctgccggaaggagcctaaaggaagtcgattagccc
                     acccaacatacgccgatgttctcacagacgcgtcatattacgcacccac
                     cttaaccctatcttcttttctctagtcgtgtctagccccatttcaccgt


PRE-PROGON1      183 GLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTS
                     GLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTS
                     GLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPGGITLTS
PROGON1          170 gcgctgcaagtcagcggggagcatgcgcattttcaggtggtcggaatat
                     gtggggcctatcttccccctacccggtttgccagaactggccggtctcc
                     ttacgtgttcagtttgtaatagagaactggtatcttataactttctggt


PRE-PROGON1      232 SWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYD
                     SWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYD
                     SWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYD
PROGON1          317 ttgctagagtgcagagaacgattcgataggagcaggaggaggaaaattg
                     cgacccgttcagctctcaaattgcgtctagagatttcggaacaacgtaa
                     cgtacttttccctgaccgttgccttccgttctgcacatcctcggctgtt


PRE-PROGON1      281 SSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYS
                     SSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYS
                     SSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYS
PROGON1          464 ttagatacgcgacggcgtcttgaatggcgtaaggttagggtgaagggta
                     ccgaggtcgcatatcggaacccctcaggttctggcggggttaaagatag
                     atctcgcgatcgagttgtgaatcgacttttctaccgctcatggtcactc


PRE-PROGON1      330 PSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPST
                     PSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPST
                     PSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPST
PROGON1          611 cttaatatccagagacataggacgttctgacgtctgtaagtgtcggcaa
                     cccacgcctcacatacttccaaagtagcaaacgttggaagcttacgcgc
                     atagaggcactcgccagggtcgaagctacccggctagggtggcagatca


PRE-PROGON1      379 AMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGG
                     AMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGG
                     AMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGG
PROGON1          758 gaatttaagagggatggacctacgggcggatgaggatgggagaacatgg
                     ctagaacggggataccgagacaggtccactggacttaactagattctgg
                     cgcgctctatttggacaacgtcttacttcgcactcgccctaagcgctcc


PRE-PROGON1      428 SPDYQDS
                     SPDYQDS
                     SPDYQDS
PROGON1          905 tcgtcgt
                     ccaaaac
                     cattact
```

PROGON1 & PRE-PROGON1 DNA Sequence Alignment/ BRCK3R primer
(Reverse complement of sequencing results were taken)

```
PROGON1          CCAGAGCAAACGCACGCGTGGGGTTTGTTAGGAAAGGAAAGTAAAGGTGTTCGCATGTTT 1551
PRE-PROGON1      ------CAGACAACCACGCGTGGCTCTTT------GGATGGAAGAAGGGTTCG---GTGT 1197
                       ** **   * ** * ** *  **      ***  * * * * *****   ** *

PROGON1          GCGAAGCGGGGACCTAGCCACAGACCATGATACTGGTACTATACCAGTGGAAGTGGTGAT 1611
PRE-PROGON1      TCCAAGCGGG-ACCTAGC-ACAG-CCATGA-ACTGGTACTATACCAGTGGAAGTGGTGAT 1253
                  * ******* ******* **** ****** ****************************

PROGON1          GTGAAGTCAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGA 1671
PRE-PROGON1      GTGAAGTCAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGA 1313
                 ************************************************************

PROGON1          AACGCTGTCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGAT 1731
PRE-PROGON1      AACGCTGTCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGAT 1373
                 ************************************************************

PROGON1          TATCAAGACTCTGACGCCACAACCAACGCCCACATCATCACCCTCGGTGAACCCGGAACA 1791
PRE-PROGON1      TATCAAGACTCTGACGCCACAACCAACGCCCACATCATCACCCTCGGTGAACCCGGAACA 1433
                 ************************************************************

PROGON1          TCTCCCAACACTGTCTTTGCTAGCAATGGGTTGTACTTTGCCCGAACGTTTCACACCTCT 1851
PRE-PROGON1      TCTCCCAACACTGTCTTTGCTAGCAATGGGTTGTACTTTGCCCGAACGTTTCACACCTCT 1493
                 ************************************************************

PROGON1          GTTGTTCTTCCAGACGGAAGCACGTTTATTACAGGAGGCCAACGACGTGGAATTCCGTTC 1911
PRE-PROGON1      GTTGTTCTTCCAGACGGAAGCACGTTTATTACAGGAGGCCAACGACGTGGAATTCCGTTC 1553
                 ************************************************************

PROGON1          GAGGATTCAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACAAGACACTTTCTAC 1971
PRE-PROGON1      GAGGATTCAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACAAGACACTTTCTAC 1613
                 ************************************************************

PROGON1          AAGCAGAACCCCAACTCCATTGTTCGCGTCTACCATAGCATTTCCCTTT-GTA--CTGAT 2028
PRE-PROGON1      AAGCAGAACCCCAACTCCATTGTTCGCGTCTACCATAGCATTTCCCTTTTGTTACCTGAT 1673
                 ********************************************** **   *****

PROGON1          GAG------------------------------------------------------- 2031
PRE-PROGON1      GGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTCGAC 1733
                 *
```

PROGON1 & PRE-PROGON1 Amino acid sequence alignment

```
PRE-PROGON1       382 WYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPD
                      WYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPD
                      WYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPD
PROGON1           745 tttaagagggatggacctacgggcggatgaggatgggagaacatggtcg
                      gaacggggataccgagacaggtccactggacttaactagattctggcca
                      gctctatttggacaacgtcttacttcgcactcgccctaagcgctcccat


PRE-PROGON1       431 YQDSDATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTF
                      YQDSDATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTF
                      YQDSDATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTF
PROGON1           892 tcgtggaaagcaaacggcgatcaagtgaagtttgcatcatggccggaat
                      aaacacccacattctgacgcccacttcgagtatcgctacctttcaggct
                      tactccacccccccctacaatcctcttctggctcagtccttttacacgt


PRE-PROGON1       480 ITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSISLLL
                      ITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSISL
                      ITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSISLCT
PROGON1          1039 aaggcccgactggtacgtacgatgcgcgattacacatagcgtcaatcta
                      tcggagggtctaacccttccatatcaaactaaaacacttgtaagtctgc
                      taacaatatgcgtacgatatgccctaactccggccccttccctctcttt
```

```
PRE-PROGON1       529 PDGRVF
                      +G  F
                      DEGHFF
PROGON1          1186 gggctt
                      aagatt
                      tggttc
```

PROGON1 & PRE-PROGON1 DNA Sequence Alignment/ GAOend Primer

```
PROGON1          -----CACTC-ATTGTTCGCGTCTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGG 54
PRE-PROGON1      ACCCCAACTCCATTGTTCGCGTCTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGG 1680
                       **** **************************************************

PROGON1          TATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAA 114
PRE-PROGON1      TATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAA 1740
                 ************************************************************

PROGON1          TCTTTACGCCAAACTATCTTTACAATAGCAACGGCAATCTCGCGACACGTCCCAAGATTA 174
PRE-PROGON1      TCTTTACGCCAAACTATCTTTACAATAGCAACGGCAATCTCGCGACACGTCCCAAGATTA 1800
                 ************************************************************

PROGON1          CCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTT 234
PRE-PROGON1      CCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTT 1860
                 ************************************************************

PROGON1          CGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACC 294
PRE-PROGON1      CGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACC 1920
                 ************************************************************

PROGON1          AGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTC 354
PRE-PROGON1      AGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTC 1980
                 ************************************************************

PROGON1          CTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCG 414
PRE-PROGON1      CTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCG 2040
                 ************************************************************

PROGON1          GTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGGGCGGTGGCGGTTCTTGGAGCC 474
PRE-PROGON1      GTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGGGCGGTGGCGGTTCTTGGAGCC 2100
                 ************************************************************

PROGON1          ATCCGCAGTTTGAGAAATGATGAAAGGGCGAGCTCAATTCGAAGCTTGAAGGTAAGCCTA 534
PRE-PROGON1      ATCCGCAGTTTGAGAAATGATGAGCGGCCGCCAGC------------------------- 2135
                 **********************  ** **    *
```

PROGON1 & PRE-PROGON1 Amino acid sequence alignment

```
PRE-PROGON1       517 IVRVYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGN
                      IVRVYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGN
                      IVRVYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGN
PROGON1             6 agcgtcaatcttcggagtaggggctggtaaactggcatacatctaaaga
                      ttgtaagtctttcaggttaggggtggagccaatacattccaataagaga
                      ttccctctctgattcgatctttttcttcgttccgactgacttctccct


PRE-PROGON1       566 LATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQR
                      LATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQR
                      LATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQR
PROGON1           153 cgaccaaaaatacagagggaaaatagttaaagttactgagacagaagcc
                      tccgcatcgcccagtatgggtctccacctgaccttgagcccactacaag
                      cgatcgtcactagccgctcatacggttgtcggggtcttagacgtttcgc
```

```
PRE-PROGON1      615 RIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTIRV
                     RIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTIRV
                     RIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTIRV
PROGON1          300 caccacaaaggaatttcgcagtgggtcgttattgaatgggcaggtaacg
                     gtctctcaaggagactatcgacgtctcgagtttttaccgtcgtccctgt
                     ctcgtgactaatcttcattccttttgtccgggcggcgcttttgtggtct


PRE-PROGON1      664 TQ
                     TQ
                     TQ
PROGON1          447 ac
                     ca
                     tg
```

PROGOMN1: 8-AA PRE-PEPTIDE DELETION

PROGOMN1 & PRE-PROGOMN1 DNA Sequence Alignment/ GAOstr Primer

```
PROGOMN1        ----------------------------------------GGGGTCGCCTGCGTCATTC 19
PRE-PROGOMN1    CTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCC 5700
                                                       * *  **  *      * *

PROGOMN1        CCTCTAGAA-TAATTTTGTTTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGG--- 75
PRE-PROGOMN1    CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGGTTG 5760
                ********* **********************************************

PROGOMN1        ---------------------CCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC 114
PRE-PROGOMN1    CAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC 5820
                                     ************************************

PROGOMN1        TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCA 174
PRE-PROGOMN1    TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCA 5880
                ************************************************************

PROGOMN1        CTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATA 234
PRE-PROGOMN1    CTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATA 5940
                ************************************************************

PROGOMN1        CCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGA 294
PRE-PROGOMN1    CCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGA 6000
                ************************************************************

PROGOMN1        TTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTA 354
PRE-PROGOMN1    TTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTA 6060
                ************************************************************

PROGOMN1        ACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGG 414
PRE-PROGOMN1    ACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGG 6120
                ************************************************************

PROGOMN1        GCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTCCAACTTTG 474
PRE-PROGOMN1    GCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTCCAACTTTG 6180
                ************************************************************

PROGOMN1        AAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGCCAGCCCT 534
PRE-PROGOMN1    AAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGCCAGCCCT 6240
                ************************************************************

PROGOMN1        GGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCCCCAGCCTG 594
PRE-PROGOMN1    GGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCCCCAGCCTG 6300
                ************************************************************

PROGOMN1        GTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGCAGCAATTG 654
PRE-PROGOMN1    GTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGCAGCAATTG 6360
                ************************************************************

PROGOMN1        AACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATTTGAAGGAT 714
PRE-PROGOMN1    AACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATTTGAAGGAT 6420
                ************************************************************

PROGOMN1        CCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGTTTCCGACC 774
PRE-PROGOMN1    CCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGTTTCCGACC 6480
                ************************************************************

PROGOMN1        GCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGATGGTAACG 834
PRE-PROGOMN1    GCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGATGGTAACG 6540
                ************************************************************

PROGOMN1        GTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGATTCATCTA 894
PRE-PROGOMN1    GTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGATTCATCTA 6600
                ************************************************************

PROGOMN1        GCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTCATCAGCTA 954
PRE-PROGOMN1    GCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTCATCAGCTA 6660
                ************************************************************
```

```
PROGOMN1          CCATGTCAGACGGTCGTGTTTT-ACCATTGGAG-CTCCTGGAGCGGTGGCGTATTTGAGA 1012
PRE-PROGOMN1      CCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGTATTTGAGA 6720
                  ********************** ********** ************************

PROGOMN1          AGAATGGCGAAGTTTTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAATGCCAAA 1072
PRE-PROGOMN1      AGAATGGCGAAGTCT-ATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAATGCCAAG 6779
                  ************* * ****************************************

PROGOMN1          GTCCACCCAATGTTGA-AGCT-------------------------------------- 1092
PRE-PROGOMN1      GTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCACGCGTGG 6839
                  *** ************  ***
```

PROGOMN1 & PRE-PROGOMN1 Amino acid sequence alignment

GeneWise (Sample; protein sequence)

```
PROGOMN1            13 EGIQEPFTM--------AVGTGIPEGSLQFLSLRASAPIGSAIPRNNWA
                       EGIQEPFTM        AVGTGIPEGSLQFLSLRASAPIGSAIPRNNWA
                       EGIQEPFTMVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAIPRNNWA
PRE-PROGOMN1      5730 ggacgctaagggagccagggagacggacctcaccgtgcagagaccaatg
                       agtaactcttctctcaactgcgtcaggtattgtgcccctggctcgaagc
                       aatggcccgtatcttcgcaatattagttgcgctacaatcaccttcccgc


PROGOMN1            54 VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
                       VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
                       VTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQN
PRE-PROGOMN1      5877 gatgagctgagtaagaggaagattcattggaggcacccataagaaaaca
                       tcgagcacgaagaactagaaactgactagcagacaccacactataccaa
                       ctcctaggataccgcttccgtctgcactcccgtagctcacgtcggatgc


PROGOMN1           103 VNGLSVLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
                       VNGLSVLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
                       VNGLSVLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKY
PRE-PROGOMN1      6024 gagttgccccggacagtagccggtcatggaatgacggtgattggtaaat
                       tagtcttcgaagaaaggtggaatatgcagcagggctccgggtcacccaa
                       cccgtggtagttcaccgctctgttacatcacgccttgattgcccttaac


PROGOMN1           152 SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
                       SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
                       SNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTAPQPGLGRWG
PRE-PROGOMN1      6171 tatgaccgctgccggaaggagcctaaaggaagtcgattagcccgcgctg
                       catacgccgatgttctcacagacgcgtcatattacgcacccacgtgggg
                       cctatcttcttttctctagtcgcgtctagccccatttcaccgtttacgt


PROGOMN1           201 PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPGGITLTSSWDPST
                       PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPGGITLTSSWDPST
                       PTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPGGITLTSSWDPST
PRE-PROGOMN1      6318 caagtcagcggggagcatgcgcattttcaggtggtcggaatatttgcta
                       cctatcttccccctacccggtttgccagaactagccggtctcccgaccc
                       gttcagtttgtaatagagaactggtatcttataactttctggtcgtact


PROGOMN1           250 GIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSW
                       GIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSW
                       GIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSW
PRE-PROGOMN1      6465 gagtgcagagaacgattcgataggagcaggaggaggaaaattgttagat
                       gttcagctctcaaattgcgtctagagatttcggaacaacgtaaccgagg
                       tttccctgaccgttgccttccgttctgcacatcctcggctgttatctcg
```

```
PROGOMN1          299 IPGPDMQVARGYQSSATMSDGRVLPLELLERWRIEEWRSFIAHLQRHGR
                      IPGPDMQVARGYQSSATMSDGRV  + LLERWRI      IAHLQRHGR
                      IPGPDMQVARGYQSSATMSDGRVFTI!LLERWRI!RMAKSIAHLQRHGR
PRE-PROGOMN1     6612 acgcgacggcgtcttgaatggcgtaa5ccgctca5aagatagcccacgc
                      tcgcatatcggaacccctcaggttct ttagggt gtcactcatagagg
                      cgatcgagttgtgaatcgacttttct cggggtt agggtacttaatat


PROGOMN1          348 PYPMPKSTQC
                      PYPMP+STQC
                      PYPMPRSTQC
PRE-PROGOMN1     6763 ctcacatact
                      cactcgccag
                      ccagagacat
```

PROGOMN1 & PRE-PROGOMN1 DNA Sequence Alignment/ BRCK3R Primer (Reverse complement of sequencing results were taken)

```
PROGOMN1        ---------------GTATTGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGT 45
PRE-PROGOMN1    GGGATCCATCCACTGGTATTGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGT 840
                              ******************************************

PROGOMN1        TCTGCCCTGGTATCTCCATGGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATG 105
PRE-PROGOMN1    TCTGCCCTGGTATCTCCATGGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATG 900
                ************************************************************

PROGOMN1        CCAAGAAGACCAGTTTGTATGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGC 165
PRE-PROGOMN1    CCAAGAAGACCAGTTTGTATGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGC 960
                ************************************************************

PROGOMN1        AAGTGGCTCGTGGGTATCAGTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTG 225
PRE-PROGOMN1    AAGTGGCTCGTGGGTATCAGTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTG 1020
                ************************************************************

PROGOMN1        GAGGCTCCTGGAGCGGTGGCGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAA 285
PRE-PROGOMN1    GAGGCTCCTGGAGCGGTGGCGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAA 1080
                ************************************************************

PROGOMN1        AGACATGGACGTCCCTACCCAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAG 345
PRE-PROGOMN1    AGACATGGACGTCCCTACCCAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAG 1140
                ************************************************************

PROGOMN1        GATTGTACCGTTCAGACAACCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCC 405
PRE-PROGOMN1    GATTGTACCGTTCAGACAACCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCC 1200
                ************************************************************

PROGOMN1        AAGCGGGACCTAGCACAGCCATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGT 465
PRE-PROGOMN1    AAGCGGGACCTAGCACAGCCATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGT 1260
                ************************************************************

PROGOMN1        CAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTG 525
PRE-PROGOMN1    CAGCCGGAAAACGCCAGTCTAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTG 1320
                ************************************************************

PROGOMN1        TCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAG 585
PRE-PROGOMN1    TCATGTACGACGCCGTTAAAGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAG 1380
                ************************************************************

PROGOMN1        ACTCTGACGCCACAACCAACGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCA 645
PRE-PROGOMN1    ACTCTGACGCCACAACCAACGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCA 1440
                ************************************************************

PROGOMN1        ACACTGTCTTTGCTAGCAATGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTC 705
PRE-PROGOMN1    ACACTGTCTTTGCTAGCAATGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTC 1500
                ************************************************************
```

```
PROGOMN1        TTCCAGACGGAAGCACGTTTATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATT 765
PRE-PROGOMN1    TTCCAGACGGAAGCACGTTTATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATT 1560
                ************************************************************

PROGOMN1        CAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGA 825
PRE-PROGOMN1    CAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGA 1620
                ************************************************************

PROGOMN1        ACCCCAACTCCATTGTTCGCGCTTACCATA-CGCATTCCCTTTGT-ACCTGATGAAG--- 880
PRE-PROGOMN1    ACCCCAACTCCATTGTTCGCGCTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGG 1680
                ***************************** *   * ** ***** ********
```

PROGOMN1 & PRE-PROGOMN1 Amino acid sequence alignment

```
PRE-PROGOMN1    239 IVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWI
                    IVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWI
                    IVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWI
PROGOMN1          3 agtgcagagaacgattcgataggagcaggaggaggaaaattgttagata
                    ttcagctctcaaattgcgtctagagatttcggaacaacgtaaccgaggt
                    ttccctgaccgttgccttccgttctgcacatcctcggctgttatctcgc


PRE-PROGOMN1    288 PGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWT
                    PGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWT
                    PGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWT
PROGOMN1        150 cgcgacggcgtcttgaatggcgtaaggttagggtgaagggtacttaata
                    cgcatatcggaacccctcaggttctggcggggttaaagatagcccacgc
                    gatcgagttgtgaatcgacttttctaccgctcatggtcactcatagagg


PRE-PROGOMN1    337 SLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYT
                    SLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYT
                    SLPNAKVNPMLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYT
PROGOMN1        297 tccagagacataggacgttctgacgtctgtaagtgtcggcaagaattta
                    ctcacatacttccaaagtagcaaacgttggaagcttacgcgcctagaac
                    cactcgccagggtcgaagctacccggctagggtggcagatcacgcgctc


PRE-PROGOMN1    386 SGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPDYQDS
                    SGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPDYQDS
                    SGSGDVKSAGKRQSNRGVAPDAMCGNAVMYDAVKGKILTFGGSPDYQDS
PROGOMN1        444 agagggatggacctacgggcggatgaggatgggagaacatggtcgtcgt
                    ggggataccgagacaggtccactggacttaactagattctggccaaaac
                    tatttggacaacgtcttacttcgcactcgccctaagcgctcccattact


PRE-PROGOMN1    435 DATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTFITGG
                    DATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTFITGG
                    DATTNAHIITLGEPGTSPNTVFASNGLYFARTFHTSVVLPDGSTFITGG
PROGOMN1        591 ggaaagcaaacggcgatcaagtgaagtttgcatcatggccggaataagg
                    acccacattctgacgcccacttcgagtatcgctacctttcaggcttcgg
                    ccacccccccctacaatcctcttctggctcagtccttttacacgttaac


PRE-PROGOMN1    484 QRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYHSISL
                    QRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYH+ SL
                    QRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYHTHSL
PROGOMN1        738 cccgactggtacgtacgatgcgcgattacacatagcgtcactc
                    agggtctaacccttccatatcaaactaaaacacttgcaacact
                    aatatgcgtacgatatgccctaactccggccccttctctgtct
```

PROGOMN1 & PRE-PROGOMN1 DNA Sequence Alignment/ GAOend Primer

```
PROGOMN1        ----ACACTCAGTTGTTCGCGCTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGG 56
PRE-PROGOMN1    ACCCCAACTCCATTGTTCGCGCTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGG 1680
                      ****  ************************************************


PROGOMN1         TATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAA 116
PRE-PROGOMN1     TATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAA 1740
                 ************************************************************

PROGOMN1         TCTTTACGCCAAACTATCTTTACGATAGCAACGGCAATCTCGCGACACGTCCCAAGATTA 176
PRE-PROGOMN1     TCTTTACGCCAAACTATCTTTACGATAGCAACGGCAATCTCGCGACACGTCCCAAGATTA 1800
                 ************************************************************

PROGOMN1         CCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTT 236
PRE-PROGOMN1     CCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTT 1860
                 ************************************************************

PROGOMN1         CGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACC 296
PRE-PROGOMN1     CGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACC 1920
                 ************************************************************

PROGOMN1         AGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTC 356
PRE-PROGOMN1     AGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTC 1980
                 ************************************************************

PROGOMN1         CTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCG 416
PRE-PROGOMN1     CTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCG 2040
                 ************************************************************

PROGOMN1         GTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGGGCGGTGGCGGTTCTTGGAGCC 476
PRE-PROGOMN1     GTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGGGCGGTGGCGGTTCTTGGAGCC 2100
                 ************************************************************

PROGOMN1         ATCCGCAGTTTGAGAAATGATGAAAGGGCGAGCTCAATTCGAAGCTTGAAGGTAAGCCTA 536
PRE-PROGOMN1     ATCCGCAGTTTGAGAAATGATGAGCGGCCGCCAGC------------------------- 2135
                 **********************  ** **    *
```

PROGOMN1 & PRE-PROGOMN1 Amino acid sequence alignment

```
PRE-PROGOMN1     515 NSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSN
                     +S+VRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSN
                     HSVVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSN
PROGOMN1           2 ctggcgtcaatcttcggagtaggggctggtaaactggcatacatctgaa
                     acttgcaagtctttcaggttaggggtggagccaatacattccaataaga
                     cattctctctctgattcgatctttttttcttcgttccgactgacttctcc


PRE-PROGOMN1     564 GNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTD
                     GNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTD
                     GNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTD
PROGOMN1         149 gacgaccaaaaatacagagggaaaatagttaaagttactgagacagaag
                     gatccgcatcgcccagtatgggtctccacctgaccttgagcccactaca
                     ctcgatcgtcactagccgctcatacggttgtcggggtcttagacgtttc


PRE-PROGOMN1     613 QRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTI
                     QRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTI
                     QRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLFVMNSAGVPSVASTI
PROGOMN1         296 cccaccacaaaggaatttcgcagtgggtcgttattgaatgggcaggtaa
                     aggtctctcaaggagactatcgacgtctcgagtttttaccgtcgtccct
                     gcctcgtgactaatcttcattccttttgtccgggcggcgcttttgtggt
```

```
PRE-PROGOMN1      662 RVTQ
                      RVTQ
                      RVTQ
PROGOMN1          443 cgac
                      gtca
                      cttg
```

STREP-TAG /HIS-TAG SUBSTITUTION

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DNA Sequence Alignment/ T7 Primer

```
pETStrepTEV      TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT 60
pETHisTEV_3      ------TTTTTTTNGGGGGGGGGGNNCNNGCGGN------ATTCCCTCNANNAAAATTTT 48
                       : * : .. .****.. .. . ****       :* *****.*. **:*:***

pETStrepTEV      TGTTTAACTTTAAGAAGGAGATATACCATGGG-TTGGAGCCATCCGCAGTTTGAGAAAGA 119
pETHisTEV 3      G--TTAACTTTAAGAAGGAGATATACCATGGGGTCACCATCATCACCATCACCACCATGA 106
                    **************************** * . .. ****. **  :  * .*:**

pETStrepTEV      AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 179
pETHisTEV 3      AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 166
                 ************************************************************

pETStrepTEV      GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 239
pETHisTEV_3      GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 226
                 ************************************************************

pETStrepTEV      CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 299
pETHisTEV_3      CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 286
                 ************************************************************

pETStrepTEV      ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 359
pETHisTEV 3      ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 346
                 ************************************************************

pETStrepTEV      GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 419
pETHisTEV 3      GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 406
                 ************************************************************

pETStrepTEV      AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 479
pETHisTEV_3      AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 466
                 ************************************************************

pETStrepTEV      CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 539
pETHisTEV_3      CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 526
                 ************************************************************

pETStrepTEV      CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 599
pETHisTEV 3      CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 586
                 ************************************************************

pETStrepTEV      CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 659
pETHisTEV 3      CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 646
                 ************************************************************

pETStrepTEV      AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 719
pETHisTEV_3      AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 706
                 ************************************************************

pETStrepTEV      TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 779
pETHisTEV 3      TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 766
                 ************************************************************

pETStrepTEV      TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 839
pETHisTEV 3      TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 826
                 ************************************************************

pETStrepTEV      TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 899
pETHisTEV_3      TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 886
                 ************************************************************

pETStrepTEV      TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 959
pETHisTEV_3      TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 946
                 ************************************************************

pETStrepTEV      ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 1019
pETHisTEV 3      ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 1006
                 ************************************************************
```

```
pETStrepTEV      ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1079
pETHisTEV 3      ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1066
                 ************************************************************

pETStrepTEV      TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCA 1139
pETHisTEV_3      TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCA 1126
                 ************************************************************

pETStrepTEV      CGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGCCAT 1199
pETHisTEV_3      CGCGTGGCTCTTTGGATGGAAGAAGGGTCCGNTNTTCCAAGCGGGACCTAGCNANGCCAT 1186
                 *************************** **.*.***************** . *****

pETStrepTEV      GAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGTCAGCCGGAAAACGCCAGTCTAA 1259
pETHisTEV 3      GAACTGGTNCTATACCAGTGGAAGTGGNNATGTGAATTCANCCGGAAAACGCCAGNCTAA 1246
                 ******** ******************..******* ***.**************.****

pETStrepTEV      CCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCCGTTAAAGG 1319
pETHisTEV 3      CCGTGGNNGAACCCNTGAAGCCANNNGNN----------------------------GG 1277
                 ******.. *.*** ***:****...* .                           **

pETStrepTEV      AAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAACGC 1379
pETHisTEV_3      AAANNNTNN-------------------NNNNNNCNNAANNC------------------ 1300
                 ***. .  .                  . .. .*  .* .*
```

pET28cStrepTEVMN6 & pET28cHisTEVMN6 Amino acid sequence alignment

```
pETStrepTEV        1 MGWSHPQFEKENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAID
                     MG  H     ENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAID
                     MGHHHHHHHHENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAID
pETHisTEV_3       74 agccccccccgacttcggtgcagagaccaatggatgagctgagtaagag
                     tgaaaaaaaaaatatagcccctggctcgaagctcgagcacgaagaacta
                     gtcttctcctacgccgtatattaccttcccgcctcctaggataccgctt


pETStrepTEV       50 GNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGW
                     GNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGW
                     GNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGW
pETHisTEV_3      221 gaagattcattggaggcacccataagaaaacagagttgccccggacagt
                     gaaactgactagcagacaccacactataccaatagtcttcgaagaaagg
                     ccgtctgcactcccgtagctcacgtcggatgccccgtggtagttcaccg


pETStrepTEV       99 IGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAIT
                     IGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAIT
                     IGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAIT
pETHisTEV_3      368 agccggtcatggaatgacggtgattggtaaattatgaccgctgccggaa
                     tggaatatgcagcagggctccgggtcacccaacatacgccgatgttctc
                     ctctgttacatcacgccttgattgcccttaaccctatcttcttttctct


pETStrepTEV      148 EANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPT
                     EANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPT
                     EANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPT
pETHisTEV_3      515 ggagcctaaaggaagtcgattagcccgcgctgcaagtcagcggggagca
                     acagacgcgtcatattacgcacccacgtggggcctatcttccccctacc
                     agtcgcgtctagccccatttcaccgtttacgtgttcagtttgtaataga


pETStrepTEV      197 SGRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFC
                     SGRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFC
                     SGRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFC
pETHisTEV_3      662 tgcgcattttcaggtggtcggaatatttgctagagtgcagagaacgatt
                     cggtttgccagaactagccggtctcccgacccgttcagctctcaaattg
                     gaactggtatcttataactttctggtcgtacttttccctgaccgttgcc
```

```
pETStrepTEV      246 PGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSAT
                     PGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSAT
                     PGISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSAT
pETHisTEV_3      809 cgataggagcaggaggaggaaaattgttagatacgcgacggcgtcttga
                     cgtctagagatttcggaacaacgtaaccgaggtcgcatatcggaacccc
                     ttccgttctgcacatcctcggctgttatctcgcgatcgagttgtgaatc


pETStrepTEV      295 MSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQ
                     MSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQ
                     MSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQ
pETHisTEV_3      956 atggcgtaaggttagggtgaagggtacttaatatccagagacataggac
                     tcaggttctggcggggttaaagatagcccacgcctcacatacttccaaa
                     gacttttctaccgctcatggtcactcatagaggcactcgccagggtcga


pETStrepTEV      344 GLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNR
                     GLYRSDNHAWLFGWKKG +FQAGPS AMNW+YTSGS++V S+GKRQ+NR
                     GLYRSDNHAWLFGWKKGPxFQAGPSxAMNWxYTSGSxxVNSxGKRQxNR
pETHisTEV_3     1103 gttctgacgtctgtaagcNtcggcaNgaatttaagagNgatNgaccNac
                     gtagcaaacgttggaagcttacgcgactagNacggggataccgagacag
                     agctacccggctagggtgNcagatcNcgcgctctatNtgtacaacgtct


pETStrepTEV      393 GVAPDAMCG
                     +  ++A+++
                     xxTxEAxxx
pETHisTEV_3     1250 gNacggaNN
                     ggcNacNgg
                     NactacNNg
```

pET28cStrepTEVMN6 DOMAIN I/II DELETION

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DI/II Deletion DNA Sequence Alignment/ T7 Primer

```
pETStrepTEV         ----------------------------------------------ACCACG------- 6
StrepDI/II          NNNNCNAATNCCANTCTNNANNNTNNNNNNNNTGNANGNATGGNAATNCCTCNNAAANAT 60
                                                                    **:*.

pETStrepTEV         -----AATCATTTCGACG---------------------------CGCAAATCTT-TACG 33
StrepDI/II          TTTGTTTACTTTAAGAAGGAGATATACCATGGNGTTGGAGCCATCCGCAGTTTGAGAAAG 120
                         :::*:**:.**.*                          ****.:*  : :*.*

pETStrepTEV         CCAAACTATCTTTACG--ATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAA 91
StrepDI/II          AAAACCTGTACTTCCAGGGTAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAA 180
                    ..**.**.*. **.*.  .****************************************

pETStrepTEV         CCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTA 151
StrepDI/II          CCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTA 240
                    ************************************************************

pETStrepTEV         GCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCC 211
StrepDI/II          GCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCC 300
                    ************************************************************

pETStrepTEV         GCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTCCTAGCG 271
StrepGAO            GCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTCCTAGCG 360
                    ************************************************************

pETStrepTEV         ACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCGGTGTTC 331
StrepDI/II          ACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCGGTGTTC 420
                    ************************************************************

pETStrepTEV         CTAGTGTGGCTTCGACGATTCGCGTTACTCAGTGATCCGAATTCGAGCTCCGTCGACAAG 391
StrepDI/II          CTAGTGTGGCTTCGACGATTCGCGTTACTCAGTGATCCGAATTCGAGCTCCGTCGACAAG 480
                    ************************************************************

pETStrepTEV         CTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCC 451
StrepDI/II          CTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCC 540
                    ************************************************************

pETStrepTEV         CGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGG 511
StrepDI/II          CGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGG 600
                    ************************************************************

pETStrepTEV         GCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT----- 566
StrepDI/II          GCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCG 660
                    ************************************************************
```

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DI/II Deletion Amino acid sequence alignment

```
pETStrepTEV          2 GWSHPQFEKENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSSI
                       +WSHPQFEKENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSSI
                       xWSHPQFEKENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSSI
StrepDI/II          93 Ntaccctgagacttcgaagacgaccaaaaatacagagggaaaatagtta
                       gggacataaaatataggagatccgcatcgcccagtatgggtctccacct
                       tgctggtgaacgccgtccctcgatcgtcactagccgctcatacggttgt


pETStrepTEV         51 SKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGY
                       SKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGY
                       SKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGY
StrepDI/II         240 aagttactgagacagaagcccaccacaaaggaatttcgcagtgggtcgt
                       gaccttgagcccactacaaggtctctcaaggagactatcgacgtctcga
                       cggggtcttagacgtttcgcctcgtgactaatcttcattccttttgtcc
```

```
pETStrepTEV      100 WMLFVMNSAGVPSVASTIRVTQ
                     WMLFVMNSAGVPSVASTIRVTQ
                     WMLFVMNSAGVPSVASTIRVTQ
StrepDI/II       387 tattgaatgggcaggtaacgac
                     gtttttaccgtcgtccctgtca
                     gggcggcgcttttgtggtcttg
```

pET28cStrepTEVMN6 DOMAIN III DELETION

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DIII Deletion DNA Sequence Alignment/ T7 Primer

```
pETStrepTEV     TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT 60
StrepDIII       -------ANNNNNNNNNNGG------------------ANTNNCTNNNA-----ANNTT 29
                       *          **                  * *  *    *     *  **

pETStrepTEV     TGTTTAACTTTAAGAAGGAGATATACCATGGGTT-GGAGCCATCCGCAGTTTGAGAAAGA 119
StrepDIII       TNGTTNNCTTTAAGAAGGAGATATACCATGGGNTTGGAGCCATCCGCAGTTTGAGAAAGA 89
                *  **  ************************ * ************************

pETStrepTEV     AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 179
StrepDIII       AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 149
                ************************************************************

pETStrepTEV     GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 239
StrepDIII       GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 209
                ************************************************************

pETStrepTEV     CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 299
StrepDIII       CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 269
                ************************************************************

pETStrepTEV     ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 359
StrepDIII       ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 329
                ************************************************************

pETStrepTEV     GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 419
StrepDIII       GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 389
                ************************************************************

pETStrepTEV     AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 479
StrepDIII       AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 449
                ************************************************************

pETStrepTEV     CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 539
StrepDIII       CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 509
                ************************************************************

pETStrepTEV     CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 599
StrepDIII       CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 569
                ************************************************************

pETStrepTEV     CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 659
StrepDIII       CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 629
                ************************************************************

pETStrepTEV     AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 719
StrepDIII       AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 689
                ************************************************************

pETStrepTEV     TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 779
StrepDIII       TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 749
                ************************************************************

pETStrepTEV     TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 839
StrepDIII       TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 809
                ************************************************************

pETStrepTEV     TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 899
StrepDIII       TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 869
                ************************************************************

pETStrepTEV     TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 959
StrepDIII       TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 929
                ************************************************************
```

```
pETStrepTEV     ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 1019
StrepDIII       ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 989
                ************************************************************

pETStrepTEV     ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1079
StrepDIII       ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1049
                ************************************************************

pETStrepTEV     TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCA 1139
StrepDIII       TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCA 1109
                ************************************************************

pETStrepTEV     CGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGCCAT 1199
StrepDIII       CGCGTGGCTCTTTGGATGGAAGAANGGTTCGGTNNTCCAAGCGGGACCTAGNACAGCCAT 1169
                ************************ ********  **************** ********

pETStrepTEV     GAACTGGTACTATACCAGTGGAAGTGG-TGATGTGAAGTCAGCCGGAAAACGCCAGTCTA 1258
StrepDIII       GAACTGGTACTATACCAGTGGAAATGGGNNANNTGAANNCANCCGGAAANCGCCAGTCTA 1229
                ************************ ***   *  ****  ** ******* **********

pETStrepTEV     ACCGTGGTGTA-GCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCC-GTTAA 1316
StrepDIII       ACCNNGGNNNAANCCNNNNNTGCCANNNGCGGAANCCNNTTNATNNNNNANNNCCNTTNA 1289
                ***  **   *  **     *****   ****** *    * **     *   *  ** *

pETStrepTEV     AGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAA 1376
StrepDIII       AGGAAAANNNNTNAACNTTGGGGGNNCCCNNNNNNNN-AANNNNNNNNANCCNNNNNNAA 1348
                ******     * * * **** **  ***         **          **      **

pETStrepTEV     CGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAGCAA 1436
StrepDIII       NNNCCNNTTNNNC-CCNNNNNNNNAACCGGNNNNNNTNCCNNNNNNNNNTTNNNNNNAAN 1407
                   **   *   * **        * ****      * **         **       *

pETStrepTEV     TGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCACGTT 1496
StrepDIII       NNGNNNGNNNNTNNNNNNNNNNTTNANNNNNNNNNNNNNNNNNNNNNNNN--NCNNNTT 1465
                  *   *    *          **                                *   **

pETStrepTEV     TATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACC 1556
StrepDIII       NNNNNNNNNGGNNNNNNNNNG---------------GGNNTNNNNNNNNNNNNNNCNCN 1510
                          *         *               **  *              * *

pETStrepTEV     TGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCG 1616
StrepDIII       NGNNNNNNNNNNNNNNNNNNNNNNN-----NANNAN----------------------- 1542
                 *                           *  *
```

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DIII Deletion Amino acid sequence alignment

```
StrepTEVMN6       2 GWSHPQFEKENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDG
                    +WSHPQFEKENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDG
                    xWSHPQFEKENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDG
StrepDIII        61 gtaccctgagacttcggtgcagagaccaatggatgagctgagtaagagg
                    Nggacataaaatatagcccctggctcgaagctcgagcacgaagaactag
                    tgctggtgaacgccgtatattaccttcccgcctcctaggataccgcttc


StrepTEVMN6      51 NKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWI
                    NKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWI
                    NKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWI
StrepDIII       208 aagattcattggaggcacccataagaaaacagagttgccccggacagta
                    aaactgactagcagacaccacactataccaatagtcttcgaagaaaggt
                    cgtctgcactcccgtagctcacgtcggatgccccgtggtagttcaccgc


StrepTEVMN6     100 GRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITE
                    GRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITE
                    GRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITE
StrepDIII       355 gccggtcatggaatgacggtgattggtaaattatgaccgctgccggaag
                    ggaatatgcagcagggctccgggtcacccaacatacgccgatgttctca
                    tctgttacatcacgccttgattgcccttaaccctatcttcttttctcta
```

```
StrepTEVMN6      149 ANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTS
                     ANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTS
                     ANGQPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTS
StrepDIII        502 gagcctaaaggaagtcgattagcccgcgctgcaagtcagcggggagcat
                     cagacgcgtcatattacgcacccacgtggggcctatcttccccctaccc
                     gtcgcgtctagccccatttcaccgtttacgtgttcagtttgtaatagag

StrepTEVMN6      198 GRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCP
                     GRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCP
                     GRVLMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCP
StrepDIII        649 gcgcattttcaggtggtcggaatatttgctagagtgcagagaacgattc
                     ggtttgccagaactagccggtctcccgacccgttcagctctcaaattgc
                     aactggtatcttataactttctggtcgtacttttccctgaccgttgcct

StrepTEVMN6      247 GISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATM
                     GISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATM
                     GISMDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATM
StrepDIII        796 gataggagcaggaggaggaaaattgttagatacgcgacggcgtcttgaa
                     gtctagagatttcggaacaacgtaaccgaggtcgcatatcggaacccct
                     tccgttctgcacatcctcggctgttatctcgcgatcgagttgtgaatcg

StrepTEVMN6      296 SDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQG
                     SDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQG
                     SDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQG
StrepDIII        943 tggcgtaaggttagggtgaagggtacttaatatccagagacataggacg
                     caggttctggcggggttaaagatagcccacgcctcacatacttccaaag
                     acttttctaccgctcatggtcactcatagaggcactcgccagggtcgaa

StrepTEVMN6      345 LYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRG
                     LYRSDNHAWLFGWK+GS++QAGP+TAMNWYYTSG ++++++G+RQSN++
                     LYRSDNHAWLFGWKxGSxxQAGPxTAMNWYYTSG!xxxxxxGxRQSNxx
StrepDIII       1090 ttctgacgtctgtaagtgNcggcaagaatttaag4gNNaNNgacctacg
                     tagcaaacgttggaagcttacgcgcctagaacgg gataccgagacaNg
                     gctacccggctaggNtgNcagatNacgcgctcta NNgNacaNcgtcNN

StrepTEVMN6      394 VAPDAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTS
                      + +A++G+  + +    K+    G +  +   +  + ++    +P+
                     !xxxAxxGxxxxxxxxxRKxxxxWGxxxxxxxxxxxxxxxxxxxxPxxx
StrepDIII       1238 4NNNgaNgactaNNNctaaaNNctgNcNNNNNNNNNaNctNcNNNcgNN
                      cNNcNggNNttNaNNNgaNNaNggNcNNaNNNcNNaNNtNcNNacNNt
                      cNtcNcacNNNNNctagaNtatggcNNNaNNacNNNcNNcNNNagNNN

StrepTEVMN6      443 PNTVFASNGLYF
                     +   +   +  +
                     xxxxxxxxxxxx
StrepDIII       1386 cNNNtNNNgNNt
                     cNNNNNaNNgNN
                     NNNtNNaNNNNN
```

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DIII Deletion DNA Sequence Alignment/ T7REV Primer

```
pETStrepTEV          AACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGC 540
StrepDIII            AANTTNGAANCTNNCCCNNNNTNNNNNNTTNNGTTTNNNNNNNNCCATGAANNNNANGGC 125
                      ** ** *** **  ***           **    **        *  ****    * ***

pETStrepTEV         CAGCCC-TGGACTAGCATTGCAGAGATCAACGTCTTCC--AAGCTAGTTCTTACACAGCC 597
StrepDIII           CANNNCCTGGNNNACCATTGCNANNANNAANNNNTTTCCAANNTTAGTTTTNNNACNNNN 185
                    **   * ***   * ******    *  **    ** *  *   ***** *   **

pETStrepTEV         CCCC-AGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGC 656
StrepIII            CCCCCAGCCTGGTCTTGGACGCTGGGNNCCNNCTATTGACTTACCGATTGTTCCTGCGGC 245
                    **** ********************  **  ****************************

pETStrepTEV         TGCAGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGC 716
StrepDIII           TGCAGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTNTTCATATCGCAATGATGC 305
                    ************************************** ******************

pETStrepTEV         ATTTGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTAT 776
StrepDIII           ATTTGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTAT 365
                    ************************************************************

pETStrepTEV         TGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCAT 836
StrepDIII           TGTTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCAT 425
                    ************************************************************

pETStrepTEV         GGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTA 896
StrepDIII           GGATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTA 485
                     ************************************************************
pETStrepTEV         TGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCA 956
StrepDIII           TGATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCA 545
                    ************************************************************

pETStrepTEV         GTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGG 1016
StrepDIII           GTCATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGG 605
                    ************************************************************

pETStrepTEV         CGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACC 1076
StrepDIII           CGTATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACC 665
                    ************************************************************

pETStrepTEV         CAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAA 1136
StrepDIII           CAATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAA 725
                    ************************************************************

pETStrepTEV         CCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGC 1196
StrepDIII           CCACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGC 785
                    ************************************************************

pETStrepTEV         CATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGTCAGCCGGAAAACGCCAGTC 1256
StrepDIII           CATGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGTCAGCCGGAAAACGCCAGTC 845
                    ************************************************************

pETStrepTEV         TAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCCGTTAA 1316
StrepDIII           TAACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCCGTTAA 905
                    ************************************************************

pETStrepTEV         AGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAA 1376
StrepDIII           AGGAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAA 965
                    ************************************************************

pETStrepTEV         CGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAGCAA 1436
StrepDIII           CGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAGCAA 1025
                    ************************************************************

pETStrepTEV         TGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCACGTT 1496
StrepDIII           TGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCACGTT 1085
                    ************************************************************
```

```
pETStrepTEV         TATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACC 1556
StrepDIII           TATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACC 1145
                    ************************************************************

pETStrepTEV         TGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCG 1616
StrepDIII           TGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCG 1205
                    ************************************************************


pETStrepTEV         CGCTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGG 1676
StrepDIII           CGCTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGG 1265
                    ************************************************************

pETStrepTEV         TCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCT 1736
StrepDIII           TCTTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACT---- 1321
                    ************************************************************

pETStrepTEV         TTACGATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAG 1796
StrepDIII           ------------------------------------------------------------


pETStrepTEV         CGTCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTT 1856
StrepDIII           ------------------------------------------------------------

pETStrepTEV         GATTCGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGAC 1916
StrepDIII           ------------------------------------------------------------


pETStrepTEV         TCTGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGC 1976
StrepDIII           ------------------------------------------------------------


pETStrepTEV         TTTGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCGGTGTTCCTAGTGTGGCTTC 2036
StrepDIII           ------------------------------------------------------------

pETStrepTEV         GACGATTCGCGTTACTCAGTGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCAC 2096
StrepDIII           --------------------GATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCAC 1361
                                        ************************************

pETStrepTEV         TCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTG 2156
StrepDIII           TCGAGCACCACCACCACCACCNCT-AGATCCGG-TGNNANNA--GCCNNAAGNNNNNNNN 1417
                    ******************** ** ******** **  *  *  ***  **

pETStrepTEV         AGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG 2216
StrepDIII           N------------------------------------------------------------ 1418


pETStrepTEV         TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT 2258
StrepDIII           -------------------------------------------
```

pET28cStrepTEVMN6 & pET28cStrepTEVMN6 DIII Deletion Amino acid sequence alignment

```
pETStrepTEV        113 WGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTSIAEIN
                       ++ P  +G+ +     ++ +++  +    ++    +    +W +I+ ++
                       xxxPxxxGxxxxxFFQxxPxxxxxxxxxxxxxxxMxxxAxxWxTIxxxx
StrepDIII           10 NgNcNNNgaNtcNttcaNcatacNcNNNtNtNNNaaNagaNtNaagaaa
                       NgNcgNcgNNNcNttaNNcNNatcNNNNtgtNNctaNNcNcgNctcNNa
                       gNNtNtatNNNNNtcaNtatgNNcNtNNNtNNNcgNNgcNcgNctNNNN


pETStrepTEV        162 VFQASSYT-APQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRND
                        F    +   PQPGLGRW+++IDLPIVPAAAAIEPTSGRVLMW+SYRND
                       xFPxLVxxxxPQPGLGRWxxxIDLPIVPAAAAIEPTSGRVLMWxSYRND
StrepDIII_3        157 NtcatgtNcNcccgcgctgcNagtcagcggggagcatgcgcattttcag
                       NtcNtttNNNcacgtgggNcctatcttccccctacccggtttgNcagaa
                       NtaNatNaNccgtttacgNNttcagtttgtaatagagaactggtatctt
```

```
pETStrepTEV      210 AFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVV
                     AFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVV
                     AFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVV
StrepDIII        304 gtggtcggaatatttgctagagtgcagagaacgattcgataggagcagg
                     ctagccggtctcccgacccgttcagctctcaaattgcgtctagagattt
                     ataactttctggtcgtacttttccctgaccgttgccttccgttctgcac

pETStrepTEV      259 TGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSW
                     TGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSW
                     TGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGRVFTIGGSW
StrepDIII        451 aggaggaaaattgttagatacgcgacggcgtcttgaatggcgtaaggtt
                     cggaacaacgtaaccgaggtcgcatatcggaacccctcaggttctggcg
                     atcctcggctgttatctcgcgatcgagttgtgaatcgacttttctaccg

pETStrepTEV      308 SGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFG
                     SGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFG
                     SGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRSDNHAWLFG
StrepDIII        598 agggtgaagggtacttaatatccagagacataggacgttctgacgtctg
                     gggttaaagatagcccacgcctcacatacttccaaagtagcaaacgttg
                     ctcatggtcactcatagaggcactcgccagggtcgaagctacccggcta

pETStrepTEV      357 WKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVM
                     WKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVM
                     WKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPDAMCGNAVM
StrepDIII        745 taagtgtcggcaagaatttaagagggatggacctacgggcggatgagga
                     gaagcttacgcgcctagaacggggataccgagacaggtccactggactt
                     gggtggcagatcacgcgctctatttggacaacgtcttacttcgcactcg

pETStrepTEV      406 YDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYF
                     YDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYF
                     YDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYF
StrepDIII        892 tgggagaacatggtcgtcgtggaaagcaaacggcgatcaagtgaagttt
                     aactagattctggccaaaacacccacattctgacgcccacttcgagtat
                     ccctaagcgctcccattactccacccccccctacaatcctcttctggct

pETStrepTEV      455 ARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQ
                     ARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQ
                     ARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQ
StrepDIII       1039 gcatcatggccggaataaggcccgactggtacgtacgatgcgcgattac
                     cgctacctttcaggcttcggagggtctaacccttccatatcaaactaaa
                     cagtccttttacacgttaacaatatgcgtacgatatgccctaactccgg

pETStrepTEV      504 NPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPN
                     NPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPN
                     NPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPN
StrepDIII       1186 acatagcgtcaatcttcggagtaggggctggtaaactggcataca
                     acacttgcaagtctttcaggttaggggtggagccaatacattcca
                     ccccttctctctctgattcgatctttttttcttcgttccgactgac
```

pET28cHisTEVMN6 DOMAIN I/II DELETION

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DI/II Deletion DNA Sequence Alignment/ T7 Primer

```
pETStrepTEV          TCTGTTGTTCTTCCAGACGGAAGCACGTTTATTACAGGAGGCCAACGACGTGGAATTCCG 60
HisDI/II             ----------------------------------------------GGGNAATTCCN 11
                                                                    * * ******

pETStrepTEV          TTCGAGGATTCAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACAAGACACTTTC 120
HisDI/II             NNNNAATANTT---------TGTTNAC---------TTTA------AGAAGGAGATA----- 44
                         *  * *         * ** **         * **       ***  *** *

pETStrepTEV          TACAAGCAGAACCCCAACTCCATTGTTCGCGCTTACCATAGCATTTCCCTTTTGTTACCT 180
HisDI/II             TAC-----------------CATGGGTC------ACCAT--CATCACC-----ATCACC- 73
                     ***                  *** * **      *****  ***  **      * ***

pETStrepTEV          GATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCACGAATCATTTC 240
HisDI/II             ---------------------------------------------ACCATGA-------- 80
                                                                   **** **

pETStrepTEV          GACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAATCTCGCGACACGT 300
HisDI/II             -------AAACCTGTAC----------TTCCAGGGTAGCAACGGCAATCTCGCGACACGT 123
                            *** ** ***           *  * * ************************

pETStrepTEV          CCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCG 360
HisDI/II             CCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAATTACAATCTCG 183
                     ************************************************************

pETStrepTEV          ACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTT 420
HisDI/II             ACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGACACACACGGTT 243
                     ************************************************************

pETStrepTEV          AATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCT 480
HisDI/DII            AATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGAAATAGCTATTCT 303
                     ************************************************************

pETStrepTEV          TTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATG 540
HisDI/II             TTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATGTTGTTCGTGATG 363
                     ************************************************************

pETStrepTEV          AACTCGGCCGGTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGTGATCCGAATTC 600
HisDI/II             AACTCGGCCGGTGTTCCTAGTGTGGCTTCGACGATTCGCGTTACTCAGTGATCCGAATTC 423
                     ************************************************************

pETStrepTEV          GAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCC 660
HisD/II              GAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCC 483
                     ************************************************************

pETStrepTEV          GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACT 720
HisDI/II             GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACT 543
                     ************************************************************

pETStrepTEV          AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAAC 780
HisDI/II             AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAAC 603
                     ************************************************************

pETStrepTEV          TATATCCGGAT---------------------------------------------- 791
HisDI/II             TATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTG 663
                     ***********
```

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DI/II Deletion Amino acid sequence alignment

```
pETStrepTEV        1 MGWSHPQFEKENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSS
                     MG  H     ENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSS
                     MGHHHHHHHHENLYFQGSNGNLATRPKITRTSTQSVKVGGRITISTDSS
HisDI/II          49 agccccccccgacttcgaagacgaccaaaaatacagagggaaaatagtt
                     tgaaaaaaaaaatataggagatccgcatcgcccagtatgggtctccacc
                     gtcttctcctacgccgtccctcgatcgtcactagccgctcatacggttg


pETStrepTEV       50 ISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPG
                     ISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPG
                     ISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPG
HisDI/II         196 aaagttactgagacagaagcccaccacaaaggaatttcgcagtgggtcg
                     tgaccttgagcccactacaaggtctctcaaggagactatcgacgtctcg
                     tcggggtcttagacgtttcgcctcgtgactaatcttcattccttttgtc


pETStrepTEV       99 YWMLFVMNSAGVPSVASTIRVTQ
                     YWMLFVMNSAGVPSVASTIRVTQ
                     YWMLFVMNSAGVPSVASTIRVTQ
HisDI/II         343 ttattgaatgggcaggtaacgac
                     agtttttaccgtcgtccctgtca
                     cgggcggcgcttttgtggtcttg
```

pET28cHisTEVMN6 DOMAIN III

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DIII Deletion DNA Sequence Alignment/ T7 Primer

```
pETStrepTEV         TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT 60
HisDIII             -------------------------NNGNNNGGA-----ATTNCCNNNNAAAN----TTT 26
                                             *   ***     *** **    * *     ***

pETStrepTEV         TGTTTAACTTTAAGAAGGAGATATACCATGGGTTGGA-GCCATCCGCAGTTTGAGAAAGA 119
HisDIII             TGTTNN-CTTTANGAAGGAGATATACCATGGGNTCACCATCATCACCATCACCACCATGA 85
                    ****   ***** ****************** *      ****  **     *  * **

pETStrepTEV         AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 179
HisDIII             AAACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTG 145
                    ************************************************************

pETStrepTEV         GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 239
HisDIII             GGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAA 205
                    ************************************************************

pETStrepTEV         CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 299
HisGAODIII          CAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACAC 265
                    ************************************************************

pETStrepTEV         ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 359
HisDIII             ATACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACA 325
                    ************************************************************

pETStrepTEV         GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 419
HisDIII             GGATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCAC 385
                    ************************************************************

pETStrepTEV         AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 479
HisDIII             AAACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTC 445
                    ************************************************************

pETStrepTEV         CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 539
HisDIII             CAACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGG 505
                    ************************************************************

pETStrepTEV         CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 599
HisDIII             CCAGCCCTGGACTAGCATTGCAGAGATCAACGTCTTCCAAGCTAGTTCTTACACAGCCCC 565
                    ************************************************************

pETStrepTEV         CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 659
HisDIII             CCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTGC 625
                    ************************************************************

pETStrepTEV         AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 719
HisDIII             AGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCATT 685
                    ************************************************************

pETStrepTEV         TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 779
HisDIII             TGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTGT 745
                    ************************************************************

pETStrepTEV         TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 839
HisDIII             TTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGGA 805
                    ************************************************************

pETStrepTEV         TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 899
HisDIII             TGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATGA 865
                    ************************************************************

pETStrepTEV         TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 959
HisDIII             TTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGTC 925
                    ************************************************************
```

```
pETStrepTEV         ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 1019
HisDIII             ATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCGT 985
                    ************************************************************

pETStrepTEV         ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1079
HisDIII             ATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCAA 1045
                    ************************************************************

pETStrepTEV         TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACCA 1139
HisDIII             TGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCANACAACCA 1105
                    ************************************************** *******

pETStrepTEV         CGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGCCAT 1199
HisDIII             CGCGTGGCTCTTTGGATGGAANAAGGGTTCGNNGTTCCAAGCGGGACCTAGCACAGCCAT 1165
                    ******************** *********  **************************

pETStrepTEV         GAACTGGTACTATACCAGTGGAAGTGGTG-ATGTGAAGTCAGCC-GGAAAACGCCAGTCT 1257
HisDIII             GAACTGGTACTATACCAGTGGNAATGGNNGATGTGAANNCNNNCCGGAAANCGCCAGTCT 1225
                    ******************** * ***   *******  *   * ***** *********

pETStrepTEV         AACCGTGGTG-TAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGA-CGCCGTTA 1315
HisDIII             AANNNNGGNNGNANCCCNNGNTGCCATGGNCGGAAANNNNNNNANNNACNAACCCNNTTA 1285
                    **    **    * ***  * *******  ******       *   ** * * *  ***

pETStrepTEV         AAGGAAAGATCCTGA-CCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCC-ACAAC 1373
HisDIII             AAGGNAANNNCCTNNACTTNNGGGGGTCCCNNNNNNNNNNAAGANNNTGAANCCCNNANN 1345
                    **** **   ***   * *  ** **  **          ****   ***  **   *

pETStrepTEV         CAACGCCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAG 1433
HisDIII             NAANNCCCNANTNNNNNCCC--------CCCGGNNNA---------------------- 1374
                     **  ***   *     ***        *****   *

pETStrepTEV         CAATGGGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCAC 1493
HisDIII             ----------------ANCCGNNNNNTTCNNNNNNNNNNNNNT-TTNNNNANNNAGG--- 1414
                                   ***     ***             * **    *   * *

pETStrepTEV         GTTTATTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTAC 1553
HisDIII             ---------------------------------GNNNGNNNNTTNNNNNNNNNNTNNNN 1441
                                                     *   *    **          *

pETStrepTEV         ACCTGAGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGT 1613
HisDIII             NNNNNNNNNNNNNCNNNNNNNNNNNNGNNNNTNNNN---------------------- 1477
                                               *
```

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DIII Deletion Amino acid sequence alignment

```
pETStrepTEV       5 HPQFEKENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDGNKD
                      H     ENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDGNKD
                      HHHHHHENLYFQGASAPIGSAIPRNNWAVTCDSAQSGNECNKAIDGNKD
HisDIII            66 ccccccgacttcggtgcagagaccaatggatgagctgagtaagaggaag
                      aaaaaaaatatagcccctggctcgaagctcgagcacgaagaactagaaa
                      tctcctacgccgtatattaccttcccgcctcctaggataccgcttccgt

pETStrepTEV       54 TFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWIGRH
                      TFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWIGRH
                      TFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQNGWIGRH
HisDIII          213 attcattggaggcacccataagaaaacagagttgccccggacagtagcc
                      ctgactagcagacaccacactataccaatagtcttcgaagaaaggtgga
                      ctgcactcccgtagctcacgtcggatgccccgtggtagttcaccgctct
```

```
pETStrepTEV      103 EVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANG
                     EVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANG
                     EVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANG
HisDIII          360 ggtcatggaatgacggtgattggtaaattatgaccgctgccggaaggag
                     atatgcagcagggctccgggtcacccaacatacgccgatgttctcacag
                     gttacatcacgccttgattgcccttaaccctatcttcttttctctagtc

pETStrepTEV      152 QPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRV
                     QPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRV
                     QPWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRV
HisDIII          507 cctaaaggaagtcgattagcccgcgctgcaagtcagcggggagcatgcg
                     acgcgtcatattacgcacccacgtggggcctatcttccccctacccggt
                     gcgtctagccccatttcaccgtttacgtgttcagtttgtaatagagaac

pETStrepTEV      201 LMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGIS
                     LMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGIS
                     LMWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGIS
HisDIII          654 cattttcaggtggtcggaatatttgctagagtgcagagaacgattcgat
                     ttgccagaactagccggtctcccgacccgttcagctctcaaattgcgtc
                     tggtatcttataactttctggtcgtacttttccctgaccgttgccttcc

pETStrepTEV      250 MDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDG
                     MDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDG
                     MDGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDG
HisDIII          801 aggagcaggaggaggaaaattgttagatacgcgacggcgtcttgaatgg
                     tagagatttcggaacaacgtaaccgaggtcgcatatcggaacccctcag
                     gttctgcacatcctcggctgttatctcgcgatcgagttgtgaatcgact

pETStrepTEV      299 RVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYR
                     RVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYR
                     RVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYR
HisDIII          948 cgtaaggttagggtgaagggtacttaatatccagagacataggacgttc
                     gttctggcggggttaaagatagcccacgcctcacatacttccaaagtag
                     tttctaccgctcatggtcactcatagaggcactcgccagggtcgaagct

pETStrepTEV      348 SDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQS-NRGVA
                     S+NHAWLFGW+KGS FQAGPSTAMNWYYTS+++  +   + S  ++++
                     SxNHAWLFGWxKGSxFQAGPSTAMNWYYTSxNxxCExxPExASLxxxxx
HisDIII         1095 tNacgtctgtaagtNtcggcaagaatttaagagNtgNNcgagacaNggN
                     caaacgttggaagcNtacgcgcctagaacggagggaNNcaNcgtNNNNc
                     acccggctagNgtggcagatcacgcgctctNtNatacNgacctaNgNac

pETStrepTEV      396 PDAMCGNAVMYDAVKGKILTFGGSPDYQ
                     ++AM G+  +++ +K++ + ++GS+  +
                     xxAMxGxxxxxx!xKxxxxxxxGSxxxx
HisDIII         1242 cggaggaNNaNN4NagaNcNtNgtcNNN
                     NNctNgaNNNaa tagaNtatggcNNNN
                     NtcgcaNNNNca taNNcNcNggcNNNa
```

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DIII Deletion DNA Sequence Alignment/ T7 REV

```
pETStrepTEV      TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT 60
HisDIII          ------------------------------------------------------------


pETStrepTEV      TGTTTAACTTTAAGAAGGAGATATACCATGGGTTGGAGCCATCCGCAGTTTGAGAAAGAA 120
HisDIII          ------------------------------------------------------------


pETStrepTEV      AACCTGTACTTCCAGGGTGCATCTGCACCTATTGGAAGCGCCATTCCTCGCAACAACTGG 180
HisDIII          ------------------------------------------------------------


pETStrepTEV      GCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAAC 240
HisDIII          ------------------------------------------------------------


pETStrepTEV      AAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACA 300
HisDIII          ------------------------------------------------------------


pETStrepTEV      TACACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAG 360
HisDIII          ------------------------------------------------------------

pETStrepTEV      GATGGTAACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACA 420
HisDIII          ------------------------------------------------------------


pETStrepTEV      AACTGGGGCAGCCCTGTTGCGTCAGGTAGTTGGTTCGCCGACTCTACTACAAAATACTCC 480
HisDIII          ------------------------------------------------------------


pETStrepTEV      AACTTTGAAACTCGCCCTGCTCGCTATGTTCGTCTTGTCGCTATCACTGAAGCGAATGGC 540
HisDIII          -NNNNNNGGNNNNNNTNNNTTNNNNNNNTTC--CNTG-----------AANNNNAATGGC 46
                                 *       ***  * **           *    ******

pETStrepTEV      CAGCCCTGG-ACTAGCATTGCAGAGATCAACGTCTT-CCAAGCTAGTTCTTACACAGCCC 598
HisDIII          CNACCCNNGGNNNNNCATTNNNNNNNNNAANNNTTTTCCAAGGTAGTTNTTNNNAAGCCC 106
                 *  ***  *      ****         **    ** ***** ***** **    *****

pETStrepTEV      CCCAGCCTGGTCTTGGACGCTGGGGTCCGACTATTGACTTACCGATTGTTCCTGCGGCTG 658
HisDIII          CCNNGCCTNNNTTTGGACGCTGGGGTCCGGNNNTTGACTTACCGATTGTTCCTGCGGCNG 166
                 **  ****    *****************    ************************ *

pETStrepTEV      CAGCAATTGAACCGACATCGGGACGAGTCCTTATGTGGTCTTCATATCGCAATGATGCAT 718
HisDIII          CAGCAATTGAACCGACATCGGGACGAGTCCTTNNGNGGTNTTCATATNNCAATGATGCAT 226
                 ********************************  * *** *******  ***********

pETStrepTEV      TTGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTG 778
HisDIII          TTGAAGGATCCCCTGGTGGTATCACTTTGACGTCTTCCTGGGATCCATCCACTGGTATTG 286
                 ************************************************************

pETStrepTEV      TTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGG 838
HisDIII          TTTCCGACCGCACTGTGACAGTCACCAAGCATGATATGTTCTGCCCTGGTATCTCCATGG 346
                 ************************************************************

pETStrepTEV      ATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATG 898
HisDIII          ATGGTAACGGTCAGATCGTAGTCACAGGTGGCAACGATGCCAAGAAGACCAGTTTGTATG 406
                 ************************************************************

pETStrepTEV      ATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGT 958
HisDIII          ATTCATCTAGCGATAGCTGGATCCCGGGACCTGACATGCAAGTGGCTCGTGGGTATCAGT 466
                 ************************************************************

pETStrepTEV      CATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCG 1018
HisDIII          CATCAGCTACCATGTCAGACGGTCGTGTTTTTACCATTGGAGGCTCCTGGAGCGGTGGCG 526
                 ************************************************************
```

```
pETStrepTEV     TATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCA 1078
HisDIII         TATTTGAGAAGAATGGCGAAGTCTATAGCCCATCTTCAAAGACATGGACGTCCCTACCCA 586
                ************************************************************

pETStrepTEV     ATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACC 1138
HisDIII         ATGCCAAGGTCAACCCAATGTTGACGGCTGACAAGCAAGGATTGTACCGTTCAGACAACC 646
                ************************************************************

pETStrepTEV     ACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGCCA 1198
HisDIII         ACGCGTGGCTCTTTGGATGGAAGAAGGGTTCGGTGTTCCAAGCGGGACCTAGCACAGCCA 706
                ************************************************************

pETStrepTEV     TGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGTCAGCCGGAAAACGCCAGTCTA 1258
HisDIII         TGAACTGGTACTATACCAGTGGAAGTGGTGATGTGAAGTCAGCCGGAAAACGCCAGTCTA 766
                ************************************************************

pETStrepTEV     ACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCCGTTAAAG 1318
HisDIII         ACCGTGGTGTAGCCCCTGATGCCATGTGCGGAAACGCTGTCATGTACGACGCCGTTAAAG 826
                ************************************************************

pETStrepTEV     GAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAACG 1378
HisDIII         GAAAGATCCTGACCTTTGGCGGCTCCCCAGATTATCAAGACTCTGACGCCACAACCAACG 886
                ************************************************************

pETStrepTEV     CCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAGCAATG 1438
HisDIII         CCCACATCATCACCCTCGGTGAACCCGGAACATCTCCCAACACTGTCTTTGCTAGCAATG 946
                ************************************************************

pETStrepTEV     GGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCACGTTTA 1498
HisDIII         GGTTGTACTTTGCCCGAACGTTTCACACCTCTGTTGTTCTTCCAGACGGAAGCACGTTTA 1006
                ************************************************************

pETStrepTEV     TTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACCTG 1558
HisDIII         TTACAGGAGGCCAACGACGTGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACCTG 1066
                ************************************************************

pETStrepTEV     AGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCG 1618
HisDIII         AGATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCG 1126
                ************************************************************

pETStrepTEV     CTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTC 1678
HisDIII         CTTACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTC 1186
                ************************************************************

pETStrepTEV     TTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTT 1738
HisDIII         TTTGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACT------ 1240
                ***************************************************

pETStrepTEV     ACGATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCG 1798
HisDIII         ------------------------------------------------------------

pETStrepTEV     TCAAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGA 1858
HisDIII         ------------------------------------------------------------

pETStrepTEV     TTCGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTC 1918
HisDIII         ------------------------------------------------------------

pETStrepTEV     TGACAAACAATGGAGGAAATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTT 1978
HisDIII         ------------------------------------------------------------

pETStrepTEV     TGCCTGGCTACTGGATGTTGTTCGTGATGAACTCGGCCGGTGTTCCTAGTGTGGCTTCGA 2038
HisDIII         ------------------------------------------------------------

pETStrepTEV     CGATTCGCGTTACTCAGTGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTC 2098
HisDIII         ------------------GATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTC 1282
                                  ****************************************
```

```
pETStrepTEV      GAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG 2158
HisDIII          GAGCACCACCACCACCACCACTGAGATCCGGCTGCTANNA--GCCNNAAGANNNNNNN-- 1338
                 ************************************  *  ***  **

pETStrepTEV      TTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTC 2218
HisDIII          ------------------------------------------------------------


pETStrepTEV      TTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT 2258
HisDIII          ----------------------------------------
```

pET28cStrepTEVMN6 & pET28cHisTEVMN6 DIII Deletion Amino acid sequence alignment

```
pETStrepTEV      153 PWTSIAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVL
                     +  ++   ++FQ S++ AP+P  GRWGP +DLPIVPA+AAIEPTSGRVL
                     xxxxxxxxxxFQGSxxxAPxPxFGRWGPxxDLPIVPAxAAIEPTSGRVL
HisDIII           52 cgNNtNNNaNtcgattNgcNcNtgctgcgNgtcagcggggagcatgcgc
                     NgNctNNNNttaggNNaccNcNtggggcNtatcttcccccctacccggtt
                     NNNaNNNaNtcatttNaccgtNtacgtgNtcagtttgNaatagagaact


pETStrepTEV      202 MWSSYRNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISM
                     +++SY NDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISM
                     xxxSYxNDAFEGSPGGITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISM
HisDIII          199 NNtttNaggtggtcggaatatttgctagagtgcagagaacgattcgata
                     NgNcaNaactagccggtctcccgacccgttcagctctcaaattgcgtct
                     ggtatcttataactttctggtcgtacttttccctgaccgttgccttccg


pETStrepTEV      251 DGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGR
                     DGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGR
                     DGNGQIVVTGGNDAKKTSLYDSSSDSWIPGPDMQVARGYQSSATMSDGR
HisDIII          346 ggagcaggaggaggaaaattgttagatacgcgacggcgtcttgaatggc
                     agagatttcggaacaacgtaaccgaggtcgcatatcggaacccctcagg
                     ttctgcacatcctcggctgttatctcgcgatcgagttgtgaatcgactt


pETStrepTEV      300 VFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRS
                     VFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRS
                     VFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNPMLTADKQGLYRS
HisDIII          493 gtaaggttagggtgaagggtacttaatatccagagacataggacgttct
                     ttctggcggggttaaagatagcccacgcctcacatacttccaaagtagc
                     ttctaccgctcatggtcactcatagaggcactcgccagggtcgaagcta


pETStrepTEV      349 DNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPD
                     DNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPD
                     DNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAPD
HisDIII          640 gacgtctgtaagtgtcggcaagaatttaagagggatggacctacgggcg
                     aaacgttggaagcttacgcgcctagaacggggataccgagacaggtcca
                     cccggctagggtggcagatcacgcgctctatttggacaacgtcttactt


pETStrepTEV      398 AMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTV
                     AMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTV
                     AMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTV
HisDIII          787 gatgaggatgggagaacatggtcgtcgtggaaagcaaacggcgatcaag
                     ctggacttaactagattctggccaaaacacccacattctgacgcccact
                     cgcactcgccctaagcgctcccattactccacccccccctacaatcctc
```

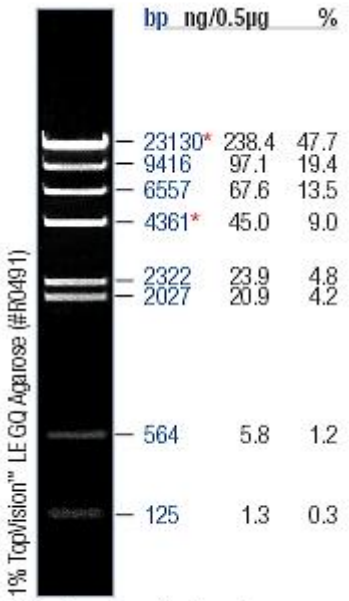
```
pETStrepTEV      447 FASNGLYFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVP
                     FASNGLYFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVP
                     FASNGLYFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVP
HisDIII          934 tgaagtttgcatcatggccggaataaggcccgactggtacgtacgatgc
                     tcgagtatcgctacctttcaggcttcggagggtctaacccttccatatc
                     ttctggctcagtccttttacacgttaacaatatgcgtacgatatgccct


pETStrepTEV      496 EQDTFYKQNPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQI
                     EQDTFYKQNPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQI
                     EQDTFYKQNPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTNHFDAQI
HisDIII         1081 gcgattacacatagcgtcaatcttcggagtaggggctggtaaactggca
                     aaactaaaacacttgcaagtctttcaggttaggggtggagccaatacat
                     aactccggccccttctctctctgattcgatctttttcttcgttccgac


pETStrepTEV      545 FTPN
                     FTPN
                     FTPN
HisDIII         1228 taca
                     tcca
                     tgac
```

Although the sequence data of all the mutational variants were analyzed, here only the sequence analysis of the mutant of interest is given.

APPENDIX E

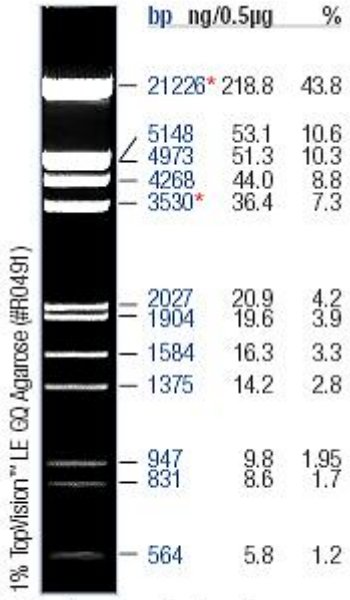
DNA AND PROTEIN MOLECULAR SIZE MARKERS

Lambda DNA/HindIII Marker



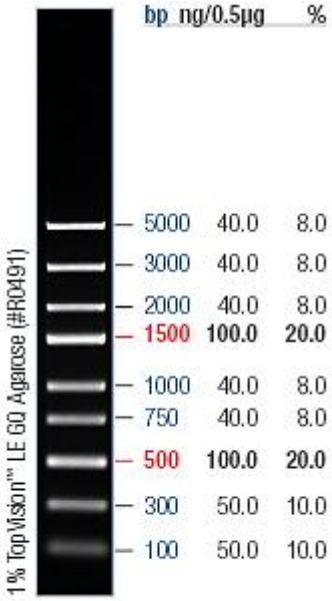
0.5µg/lane, 8cm length gel,
1X TAE, 7V/cm, 45min

Lambda DNA/EcoRI+HindIII Marker



0.5µg/lane, 8cm length gel,
1X TAE, 7V/cm, 45min

GeneRuler™ Express DNA Ladder



0.5µg/lane, 8cm length gel,
1X TAE, 7V/cm, 40min

NoLimits™ 15bp DNA Fragment

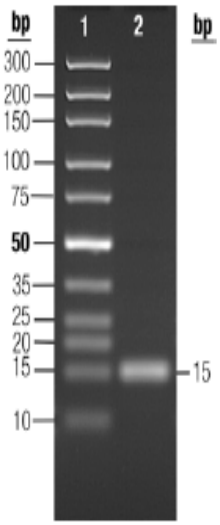
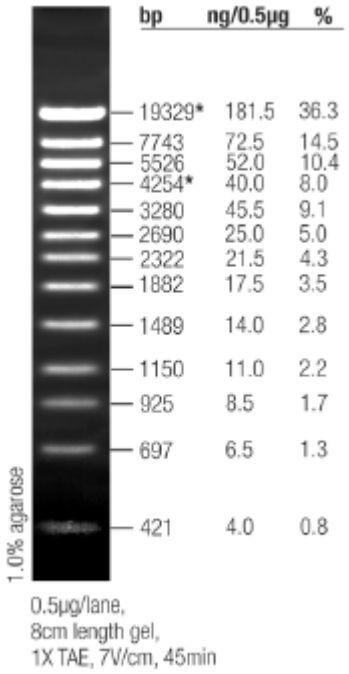
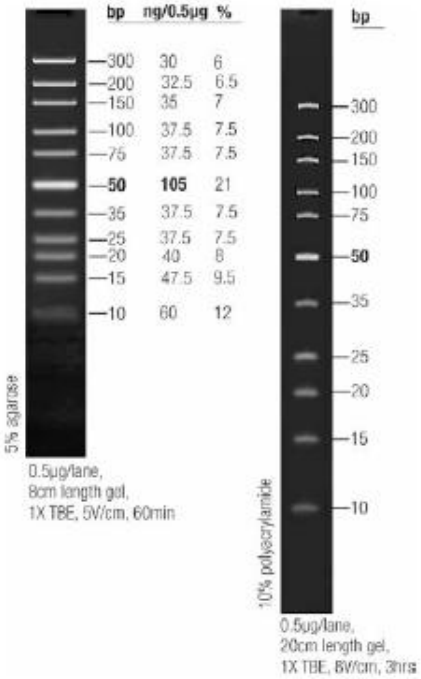


Fig.1. Analysis on agarose gel.
5%TopVision™ LE GQ Agarose (#R0491), 1X TBE, 7V/cm, 60min
1 – GeneRuler™ DNA Ladder, Ultra Low Range (#SM1223) – 1µg
2 – NoLimits™ 15 bp DNA Fragment – 0.5µg

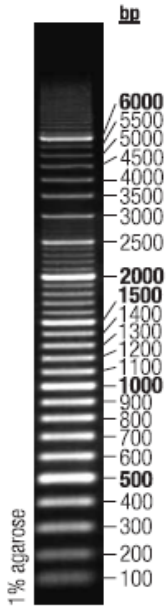
Lambda – pUC Mix Marker



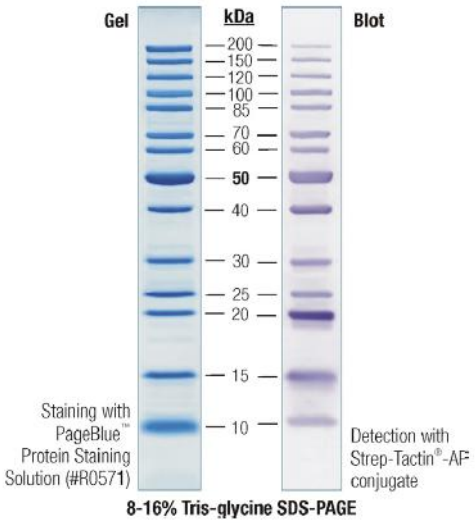
GeneRuler™ DNA Ladder, Ultra Low Range



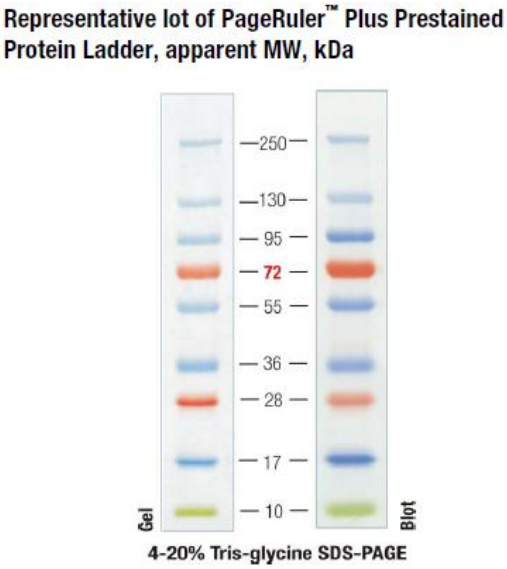
O'RangeRuler ™ 100bp+500bp DNA Ladder, ready to use



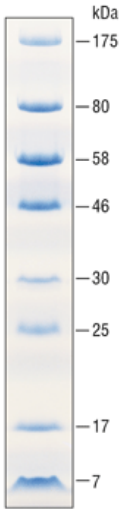
PageRuler™ Unstained Protein Ladder



PageRuler™ Plus Prestained Protein Ladder (improved)



Prestained Protein Marker, Broad Range (7-175 kDa) NEB P7708



PageRuler™ Plus Prestained Protein Ladder

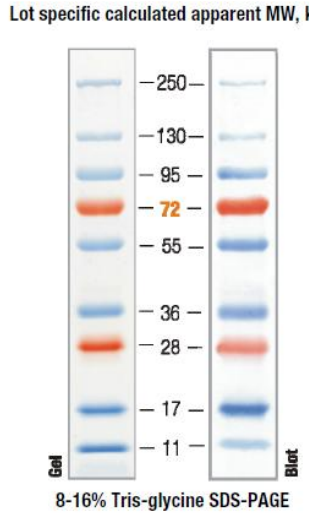


Figure E.1 DNA and protein molecular size markers.

APPENDIX F

SIGNAL PEPTIDE CLEAVAGE SITE OF GAO

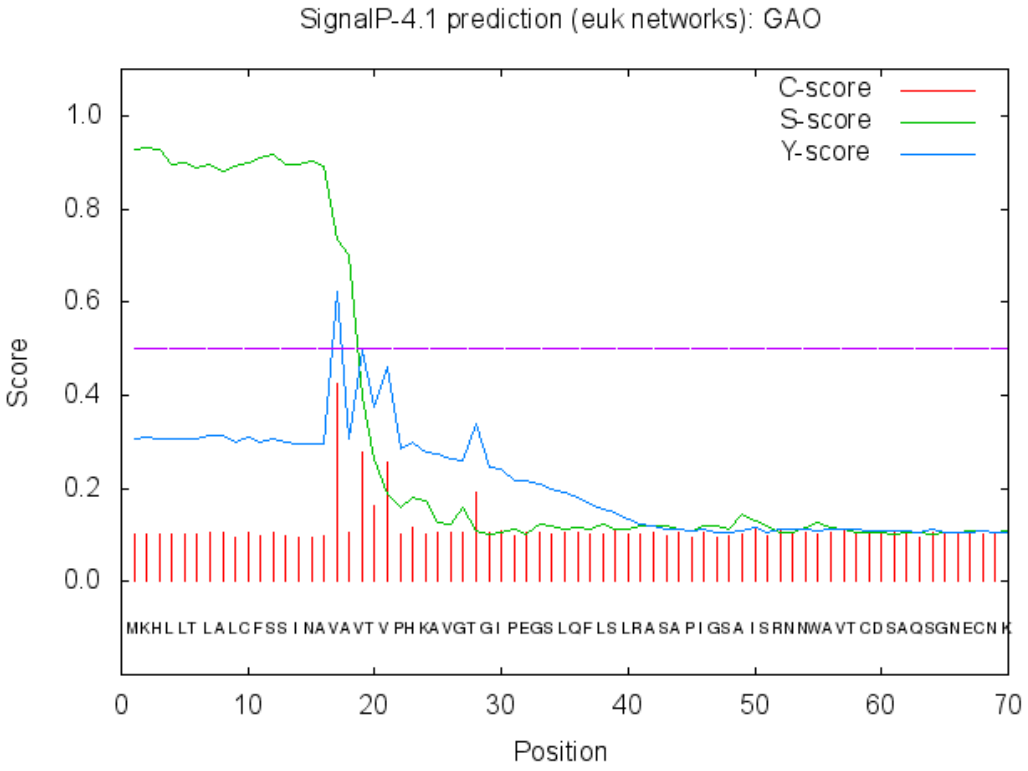


Figure F.1 SignalP 4.1 results for signal peptide cleavage site of nascent GAO (http://www.cbs.dtu.dk/services/SignalP/).

>GAO

# Name=GAO		Length=70				
#		SignalP-noTM		SignalP-TM		
# pos	aa	CS	SP	CS	SP	TM
1	M	0.101	0.926	0.102	0.865	0.071
2	K	0.103	0.932	0.102	0.893	0.071
3	H	0.101	0.926	0.102	0.881	0.081
4	L	0.102	0.895	0.102	0.889	0.119
5	L	0.103	0.899	0.103	0.884	0.107
6	T	0.103	0.889	0.103	0.904	0.114
7	L	0.105	0.894	0.104	0.904	0.101
8	A	0.106	0.882	0.105	0.909	0.087
9	L	0.096	0.890	0.098	0.916	0.084
10	C	0.105	0.898	0.105	0.903	0.070
11	F	0.097	0.909	0.099	0.916	0.084
12	S	0.104	0.918	0.104	0.927	0.080
13	S	0.097	0.896	0.099	0.906	0.076
14	I	0.095	0.897	0.098	0.913	0.089
15	N	0.096	0.902	0.100	0.942	0.081
16	A	0.097	0.891	0.099	0.948	0.099
17	V	0.425	0.733	0.398	0.825	0.164
18	A	0.104	0.703	0.104	0.771	0.191
19	V	0.277	0.396	0.228	0.500	0.209
20	T	0.161	0.265	0.144	0.354	0.212
21	V	0.255	0.188	0.250	0.247	0.231
22	P	0.101	0.160	0.103	0.249	0.200
23	H	0.117	0.182	0.118	0.226	0.160
24	K	0.103	0.172	0.103	0.213	0.139
25	A	0.107	0.128	0.105	0.177	0.130
26	V	0.105	0.122	0.106	0.165	0.128
27	G	0.106	0.160	0.103	0.154	0.114
28	T	0.193	0.109	0.178	0.133	0.123
29	G	0.107	0.103	0.106	0.124	0.121
30	I	0.110	0.107	0.108	0.124	0.114
31	P	0.098	0.111	0.101	0.127	0.111
32	E	0.105	0.101	0.104	0.109	0.116
33	G	0.106	0.122	0.104	0.116	0.121
34	S	0.103	0.120	0.102	0.117	0.123
35	L	0.105	0.114	0.104	0.115	0.142
36	Q	0.104	0.116	0.104	0.110	0.143
37	F	0.103	0.113	0.103	0.110	0.172
38	L	0.103	0.122	0.103	0.104	0.164
39	S	0.109	0.113	0.107	0.103	0.138
40	L	0.102	0.114	0.103	0.106	0.140
41	R	0.102	0.120	0.103	0.111	0.134

Name=GAO SP='YES' Cleavage site between pos. 16 and 17: INA-VA
D=0.773 D-cutoff=0.450 Networks=SignalP-noTM

APPENDIX G

THE GENETIC CODE AND SINGLE-LETTER AMINO ACID DESIGNATONS

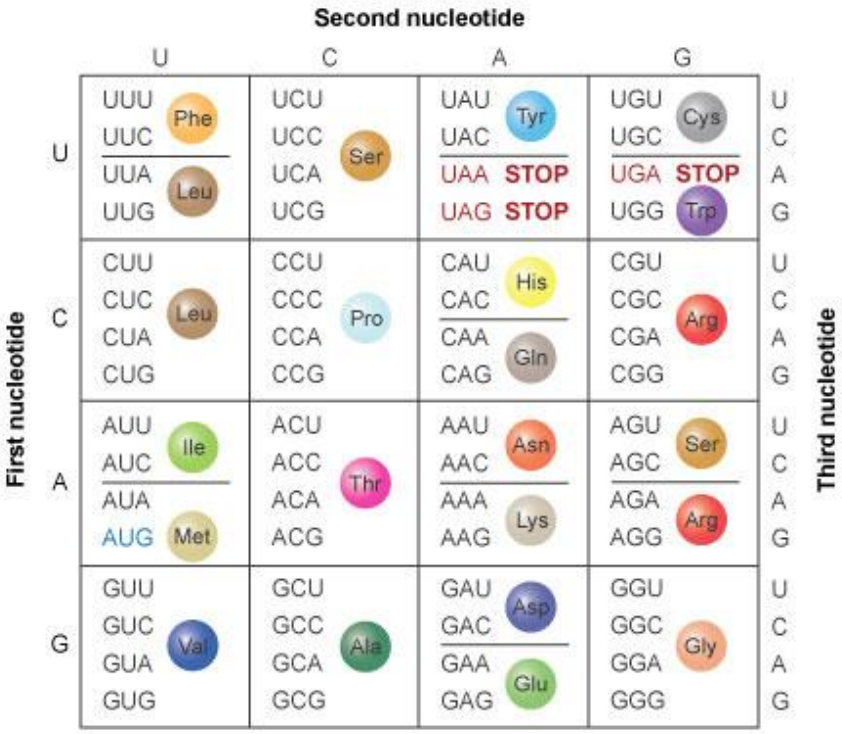


Figure G.1 The genetic code (http://www.nature.com/scitable/topicpage/the-information-in-dna-determines-cellular-function-6523228).

AMINO ACID			MASS[a]	pI [b]
Alanine	ALA	A	71.09	6.107
Arginine	ARG	R	156.19	10.76
Aspartic Acid	ASP	D	114.11	2.98
Asparagine	ASN	N	115.09	-
Cysteine	CYS	C	103.15	5.02
Glutamic Acid	GLU	E	129.12	3.08
Glutamine	GLN	Q	128.14	-
Glycine	GLY	G	57.05	6.064
Histidine	HIS	H	137.14	7.64
Isoleucine	ILE	I	113.16	6.038
Leucine	LEU	L	113.16	6.036
Lysine	LYS	K	128.17	9.47
Methionine	MET	M	131.19	5.74
Phenylalanine	PHE	F	147.18	5.91
Proline	PRO	P	97.12	6.3
Serine	SER	S	87.08	5.68
Threonine	THR	T	101.11	-
Tryptophan	TRP	W	186.12	5.88
Tyrosine	TYR	Y	163.18	5.63
Valine	VAL	V	99.14	6.002

[a] mass [dalton],

[b]The Merck Index, Merck & Co. Inc., Nahway, N.J., 11(1989); CRC Handbook of Chem.& Phys., Cleveland, Ohio, 58(1977)

Figure G.2 Single-letter and three letter amino acid designations with their mass and pI values.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Kocuklu, Burçak

Nationality: Turkish

Date and Place of Birth: 18 March 1979, Ankara

Marital Status: Married

Phone: +90 312 3436380

e-mail: brrrck@hotmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	METU- Biotechnology	2005
BS	Ankara University-Biology	2001
High School	Fethiye-Kemal Mumcu Anatolian HS	1997

WORK EXPERIENCE

Year	Place	Enrollment
8/2010-	Republic of Turkey Ministry of Forestry and Water Affairs Biotechnology Section	Expert
1/2007- 8/2010	Republic of Turkey Ministry of Environment and Forestry Biodiversity and Genetic Resources Section	Deputy Expert
7/2006-12/2006	Ankara University, Faculty of Medicine Molecular Biology and Technology R&D Unit	Researcher

PROFFESSIONAL EXPERIENCE

Year	**Place**
7/2004-9/2004	School of Biochemistry and Microbiology, Faculty of Biological Sciences, University of Leeds, UK- Visiting researcher
8/2011-12/2011	Institute of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, UK- Visiting researcher

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Kocuklu, B., Ögel Z.B., McPherson, M.J. "Analysis of Self-Catalytic Maturation Capability of Galactose Oxidase by Site-Directed Mutagenesis" FEBS Journal 277 (Suppl. 1) 37–271 (2010), 269-270

Gençer (Kocuklu), B. "Analysis of Self-Processing Mechanism of Galactose Oxidase by Site-Directed Mutagenesis and Heterologous Expression in *E. coli*" M.Sc. Thesis, Middle East Technical University, Ankara 2005

CONFERENCE PRESENTATIONS

Kocuklu, B., Ögel Z.B., McPherson, M.J. "Analysis of Self-Catalytic Maturation Capability of Galactose Oxidase by Site-Directed Mutagenesis" 35th FEBS Congress "Molecules of Life", 26 June –1 July, 2010, Gothenburg, Sweden (Poster presentation, also chosen for a short oral presentation).

Kocuklu, B., Ögel Z.B., McPherson, M.J. "Site-directed Mutagenesis Studies to Reveal the Self-Processing Mechanism of Galactose Oxidase" International Symposium on Biotechnology: Developments and Trends, 27-30 September 2009, Middle East Technical University, Cultural and Convention Center, Ankara, Türkiye
(Poster presentation)

Gençer (Kocuklu), B. Ögel, Z.B., McPherson, M.J. "Analysis of Self-Processing Mechanism of Galactose Oxidase by Site-Directed Mutagenesis in the Presence of Copper and Oxygen" XIV.National Biotechnology Congress, 31 August-2 September 2005, Eskişehir, Türkiye (Poster presentation)

FELLOWSHIPS AND SCHOLARSHIPS

8/2011-12/2011	Research Fellowship Abroad Programme for Ph.D. Students_ 2214 from TÜBİTAK (Scientific and Technological Research Council of Turkey - Scientist Supporting Department_BİDEB)
26/6–1/7 2010	FEBS Bursary (registration and travel) for the 35th FEBS Congress "Molecules of Life" in Gothenburg,Sweden
2006-2011	Ph.D. Programme Fellowship from TÜBİTAK (Scientific and Technological Research Council of Turkey -Scientist Supporting Department_BİDEB)
7-9 /2004	Research visit fellowship funded by School of Biochemistry & Molecular Biology, Faculty of Biological Sciences, University of Leeds, UK
2002-2004	M.Sc. Programme Fellowship from TÜBİTAK (Scientific and Technological Research Council of Turkey –Scientist Training Group_BAYG)

HOBBIES

Reading, foreign languages and grammar, handicrafts, volleyball, nature and nature conservation