B.GENÇER

ANALYSIS OF SELF-PROCESSING MECHANISM OF GALACTOSE OXIDASE BY SITE-DIRECTED MUTAGENESIS AND HETEROLOGOUS EXPRESSION IN ESCHERICHIA COLI

BURÇAK GENÇER

DECEMBER 2005

METU 2005

ANALYSIS OF SELF-PROCESSING MECHANISM OF GALACTOSE OXIDASE BY SITE-DIRECTED MUTAGENESIS AND HETEROLOGOUS EXPRESSION IN *ESCHERICHIA COLI*

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

BY

BURÇAK GENÇER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOTECHNOLOGY

DECEMBER 2005

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan ÖZGEN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assoc. Prof. Candan GÜRAKAN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Michael J. McPHERSON	Prof. Dr. Zümrüt Begüm ÖGEL
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Haluk HAMAMCI	(METU, FDE)	_____
Prof. Dr. Zümrüt Begüm ÖGEL	(METU, FDE.)	_____
Prof. Dr. Cumhur ÇÖKMÜŞ	(A.Ü, BIO)	_____
Prof. Dr. Hüseyin Avni ÖKTEM	(METU, BIO)	_____
Assoc. Prof. Candan GÜRAKAN	(METU, FDE)	_____

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 Gençer

Signature :

ABSTRACT

ANALYSIS OF SELF-PROCESSING MECHANISM OF GALACTOSE OXIDASE BY SITE-DIRECTED MUTAGENESIS AND HETEROLOGOUS EXPRESSION IN *ESCHERICHIA COLI*

Gençer, Burçak
M.S., Department of Biotechnology
Supervisor: Prof. Dr. Zümrüt Begüm Ögel
Co-Supervisor: Prof. Dr .Michael J. McPherson

December 2005, 153 pages

In this study, self-catalytic maturation of heterologously expressed pro-galactose oxidase was analysed in *E.coli* by altering some amino acids which were supposed to play a crucial role in pro-peptide removal. Galactose oxidase (GOase; EC 1.1.3.9) from *Fusarium graminearum*; having a molecular mass of 68kDa, is a monomeric, copper containing enzyme with an unusual thioether bond. The enzyme is produced as a precursor with an additional 8 amino acid pre- and a 17- amino acid pro-sequence at the N terminus. Previous work has shown that the pre-peptide is removed possibly by a protease during secretion, whereas the 17 amino acid pro-peptide is removed autocatalytically by the aerobic addition of Cu^{2+} to the precursor, preceding the formation of the thioether bond at the active site. The *pro-gao* gene was on ProGON1 and ProGOMN1 constructs which were previously established on pET101/D/*lacZ* vector in England by directed evolution. ProGON1 contains silent mutations at the N-terminus different from native galactose oxidase whereas

ProGOMN1 has six further mutations within the mature enzyme, providing high expression. The cleavage site mutations R-1P/A1P, R-1X/A1X, S2A, and the H522A mutation just against the cleavage site in the three dimensional configuration, were carried out by site-directed mutagenesis. Those and some extra mutations were confirmed by DNA sequence analysis. Next, mutant galactose oxidases were expressed in *E. coli* BL21 Star (DE3), and were purified by *Strep*-Tactin® Sepharose® column, operating on the basis of affinity chromatography. Subsequently, SDS-PAGE was performed to analyze self-processing by detecting molecular mass difference of protein bands resulting from pro-sequence removal or existence. When the bands obtained in SDS-PAGE were compared, it was seen that the products of original recombinant plasmids, i.e. ProGON1, ProGOMN1; and the mutational variants showed no difference in band size, all slightly above 70kDa; indicating pro-sequence presence on all constructs. Non-mutants and some of the mutants showed galactose oxidase activity, signifying proper active site construction by thioether bond formation. ProGOMN1 was submitted for N-terminal amino acid sequencing to be able to assert that a size above 70kDa is not solely due to the existence of a 1 kDa *Strep*-tag II at C-terminus. Sequencing data affirmed the presence of both the pre-peptide and the pro-preptide showing that processing has not occurred at the N-terminus. Accordingly, in this study, it was shown for the first time that the existence of a pre-pro-peptide at the N-terminus of galactose oxidase does not prevent thioether bond formation at the active site. Furthermore, since the pro-peptide is cleaved autocatalytically, the lack of removal of the pre-peptide in *E.coli* in the presence of Cu^{2+} and oxygen is very likely to be the cause of lack of pro-peptide cleavage. In future studies the region corresponding to the pre-peptide will be deleted to prove this hypothesis.

Keywords: Galactose Oxidase, pro-peptide, self-processing, autocatalytic cleavage, site-directed mutagenesis.

ÖZ

GALAKTOZ OKSİDAZIN KENDİNİ İŞLEME MEKANİZMASININ ALAN-HEDEFLİ MUTAJENEZLE ANALİZİ VE *ESCHERICHIA COLI* 'DE HETEROLOG EKSPRESYONU

Gençer, Burçak
Yüksek Lisans, Biyoteknoloji Bölümü
Tez Danışmanı: Prof. Dr. Zümrüt Begüm Ögel
Yardımcı Tez Danışmanı: Prof. Dr. Michael J. McPherson

Aralık 2005, 153 sayfa

Bu çalışmada *E.coli*'de heterolog olarak ifade edilen pro-galaktoz oksidazın kendi kendini katalizleyen olgunlaşma mekanizması, pro-peptidin uzaklaştırılmasında önemli rolü olduğu düşünülen bazı amino asitler değiştirilerek analiz edilmiştir. *Fusarium graminearum* kökenli galaktoz oksidaz (GOase; EC 1.1.3.9) 68kDa moleküler kütleye sahip, monomerik, alışılmadık bir tiyoeter bağına sahip bakır içeren bir enzimdir. Enzim N-terminalinde 8 amino asitlik ilave pre-peptide ve 17 amino asitlik pro-sekansa sahip bir öncü olarak üretilir.Daha önceki çalışmalar, pre-peptidin salgılanma sırasında muhtemelen bir proteazla uzaklaştırıldığını, buna karşılık 17 amino asitlik pro-peptidin de bakırın aerobik olarak öncüye ilavesiyle aktif bölgede tiyoeter bağı oluşumu öncesinde otokatalitik olarak ayrıldığını göstermektedir. *pro-gao* geni daha önceden İngiltere'de yönlendirilmiş evrim ile pET101/D/*lacZ* vektöründe oluşturulan ProGON1 ve ProGOMN1 yapıları üzerindedir. ProGOMN1 yüksek ekspresyonu sağlamak üzere işlenmiş enzimde ileri

altı mutasyona daha sahipken, ProGON1 yaban galaktoz oksidazdan farklı olarak yalnızca N-terminalinde anlamsız mutasyonlar içerir. Kesim bölgesi mutasyonları R-1P/A1P, R-1X/A1X, S2A; ve üç boyutlu konfigürasyonda kesim bölgesinin tam karşısında bulunan H522A mutasyonu, alan-hedefli mutajenez ile gerçekleştirilmiştir.Bunlar ve meydana gelen bazı ekstra mutasyonlar DNA dizi analiziyle doğrulanmıştır.Daha sonra, *E.coli* BL21 Star (DE3)'de ifade edilen mutant galaktoz oksidazlar afinite kromatografisi esasına göre çalışan *Strep*-Tactin® Sepharose® kolonu ile saflaştırılmıştır. Ardından, pro-sekans varlığı ya da yokluğundan kaynaklanan protein bantlarının moleküler kütle farkı tespit edilerek, kendi kendine işlenme analiz edilmiştir. SDS-PAGE sonucu elde edilen bantlar karşılaştırıldığında, orijinal rekombinant plazmitler, yani ProGON1, ProGOMN1; ve mutasyonal varyant ürünlerinin 70kDa'nın biraz üzerinde bulunup tüm yapılarda pro-sekans varlığına işaret ederek bant boyutlarının hiçbir fark göstermediği görülmüştür. Mutant olmayanlar ve bazı mutantlar, tiyoeter bağı oluşumuyla uygun aktif bölgenin oluştuğuna dikkat çekerek galaktoz oksidaz aktivitesi göstermiştir.70 kDa'nın üstündeki bir boyutun yalnızca C-terminaldeki 1 kDa'lık *Strep*-tag II 'den kaynaklanmış olamayacağını kesinlikle belirtebilmek için ProGOMN1 N-terminal amino asit dizilemesine gönderilmiştir. Dizileme verileri, N-terminalde işlenme gerçekleşmediğini gösteren pre-peptid ve pro-peptid varlığını onaylamıştır.Bu nedenle, bu çalışmada galaktoz oksidazın N-terminalindeki pre-pro-peptid varlığının aktif bölgede tiyoeter bağı oluşumunu engellemediği ilk kez gösterilmiştir. Üstelik, pro-peptid otokatalitik olarak kesildiğinden, Cu^{2+} ve oksijen varlığında *E.coli*'de pre-peptidin uzaklaştırılamaması, pro-peptid kesiminin gerçekleşmemesi için çok olası bir sebeptir. İleriki çalışmalarda bu hipotezi kanıtlamak için pre-peptid bölgesi çıkarılacaktır.

Anahtar Kelimeler: Galaktoz Oksidaz, pro-peptide, kendi kendine işlenme, otokatalitik kesim, alan- hedefli mutajenez.

“I’ve learned that we should be grateful to God

also for not giving us all the things we want…”

Dedicated to My Family…

ACKNOWLEDGEMENTS

From the very beginning, I would like to extend my sincere thanks to my supervisor Prof. Dr. Zümrüt Ögel for her encouraging and enlightening attitude towards me. Her guidance with a glittering smile on her face always made me work more willingly even in hopeless and exhausting situations. I'm very glad that I felt her approval and support during realization of this thesis.

I would like to give my special thanks to my co-supervisor Prof. Dr. Michael J. McPherson from Leeds University, Faculty of Biological Sciences, School of Biochemistry and Microbiology for his invaluable collaboration and financial support making it possible for me to carry out the foundation of my thesis in UK.

I also wish to thank to Sarah Deacon for her guidance in experiments at MJM laboratories, my instructors, technicians and lab mates for their cooperation.

I would like to express my thankfulness to TÜBİTAK both for providing me a master scholarship by BAYG and a financial support by TBAG for our project by facilitating all scientific applications throughout the whole study.

Among my friends, my warmest thanks go to Sümeyra Gürkök, Gökhan Duruksu, Eda Alagöz and Menekşe Karahasanoğlu for their kind helps, useful advices and their pleasant presence and particularly to Tülay Bal, with whom I can find a reason to smile; for making me feel her invaluable boost everytime and for raising my spirits.

I want to extend my deepest thanks to my intimate friend Gülden, for her encouragement, endless friendship and especially sharing worried and enjoyable days together.

Last, but not the least my family deserves my heartfelt thanks for their love, support, patience and consolation. I know how lucky I am by being a member of you.

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
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
MOPS	4-Morpholinepropanesulfonic acid
TEMED	N,N,N',N', Tetramethylethylenediamine
dNTP	Deoxynucleosidetriphosphate
DNase	Deoxyribonuclease
RNase	Ribonuclease
FGGO	*Fusarium graminearum* galactose oxidase
F	Forward
GAOX	Galactose Oxidase
GOase	Galactose Oxidase

GO	Galactose Oxidase
HrP	Horseradish peroxidase
KAC	Potassium acetate
NaAC	Sodium acetate
Napi	Sodiumphosphate
N-RPAP 1-4	ProGON1 R-1P/A1P clone 1-4
N-RXAX 1-4	ProGON1 R-1X/A1X clone 1-4
N-S2A 1-4	ProGON1 S2A clone 1-4
N-H522A 1-4	ProGON1 H5222A clone 1-4
MN-RPAP 1-4	ProGOMN1 R-1P/A1P clone 1-4
MN-RXAX 1-4	ProGOMN1 R-1X/A1X clone 1-4
MN-S2A 1-4	ProGOMN1 S2A clone 1-4
MN-H522A 1-4	ProGOMN1 H5222A clone 1-4
R	Reverse

CHAPTER 1

INTRODUCTION

1.1 Galactose Oxidase

Galactose oxidase (GOase; EC 1.1.3.9) from *Fusarium graminearum* (NRRL 2903) is a member of cupredoxins and is an extracellular monomeric enzyme, with a relative molecular mass of 68kDa (Kosman *et al.*,1974).

The reaction catalyzed by galactose oxidase is the oxidation of primary alcohols to their corresponding aldehydes, coupled to the two-electron reduction of O_2 but with only a single copper at the active site (Tressel and Kosman,1982).

$$RCH_2OH + O_2 \longrightarrow RCHO + H_2O_2$$

Galactose oxidase includes an unusual thioether bond between a tyrosine and a cysteine (Tyr-272 and Cys-228) that plays a role as an intrinsic cofactor.

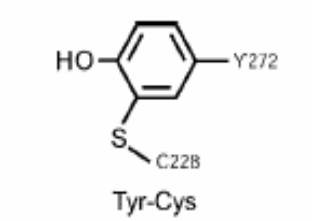


Figure 1.1 The thioether bond (Whittaker and Whittaker, 2003)

1.1.1 Subsrate Specificity

The biologically relevant substrate of GOase is not known, as the enzyme exhibits broad substrate specificity from small alcohols through sugars to oligo- and polysaccharides (Avigad *et al*., 1962; Maradufu *et al*., 1971; Maradufu and Perlin, 1974; Mendonca and Zancan, 1987).

The most commonly used experimental substrate is D-galactose, which is converted from an alcohol into an aldehyde by oxidation at the C-6 position. There is no known inducer of the enzyme, including D-galactose, which is probably not the natural substrate of the enzyme (Ögel *et al.,* 1994). L-Sorbose is routinely used in the production and purification of galactose oxidase, with no information on the mechanism of induction or derepression (Ögel and Özilgen, 1995).

A wide range of primary alcohols, including terminal D-galactose residues of oligo- and polysaccharides, can act as substrates for the enzyme, but GOase displays strict stereo- and regio specificity (Firbank *et al.,* 2003). Among those, D-glucose and L-galactose are not substrates (Ito *et al.,* 1992**).** In addition to catalyzing the oxidation of alcohols, the enzyme further converts aldehydes to the corresponding carboxylates (Kelleher and Bhavanandan, 1986).

1.1.2 Application Areas of Galactose Oxidase

Galactose oxidase is important for variable applications since it shows remarkably high degree of specificity for galactose and its derivatives among hexose sugars.

1.1.2.1 Medical Applications and Clinical Assays

Galactose oxidase has numerous applications in biosensors, chemical synthesis and diagnostics (Sun *et al.,* 2001). Sensors incorporating GOase have been used to measure D-galactose, lactose and other GOase substrate concentrations (Vega *et al*., 1998; Tkac *et al.,* 1999), blood samples (Vrbova *et al.,* 1992) and other biological fluids (Johnson *et al.,*1982). GOase-catalyzed oxidation of cell surface polysaccharides is an essential step in the radiolabelling of membrane bound

glycoproteins (Calderhead and Lienhard, 1998; Gahmberg and 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 and precancer (Yang and Shamsuddin, 1996; Said *et al*., 1999). It is also used for induction of interferon in human lymphocyte culture (Dianzani *et al.*,1979).

1.1.2.2 Food, Drug and Material Industry

Galactose oxidase is widely used in process monitoring (Szabo *et al.,* 1996), quality control in the dairy industries (Adanyi *et al.,* 1999; Mannino *et al.,* 1999) and enzymatic synthesis for the production of nonnutritive sugar substitutes in foods (Mazur, 1991). Enzymatic synthesis of carbohydrates by GOase circumvents the requirement for protecting the hydroxyl groups (Root *et al.,* 1985; Mazur and Hiler, 1997; Liu and Dordick, 1999).

In a fascinating study, galactose oxidase has been engineered by directed evolution methods to possess glucose 6-oxidase activity (Sun *et al.,* 2002). Enzymatic ability to oxidize glucose at the 6-hydroxy group generating aldehyde is not found in any currently described oxidase. Galactose oxidase, with three mutations, manifests low but significant glucose oxidase activity, which provides a good basis for further refinement of the new enzyme activity, potentially providing benefits for the food, pharmaceutical and material industries (Rogers and Dooley, 2003).

1.1.2.3 Carbohydrate Polymer Construction

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

1.1.3 Structural and Functional Features of Galactose Oxidase

The gene encoding galactose oxidase (gaoA) of Dactylium dendroides, reclassified as Fusarium spp. (NRRL 2903) (Ögel, et al., 1994), has been cloned and sequenced

(McPherson *et al.*, 1992).Unlike most other filamentous fungal genes (Gurr *et al.*, 1987) the *gaoA* gene is intronless. The entire coding region is ~2.0 kb with 120 bp at the N-terminus corresponding to a leader sequence. The translation product of the mature enzyme coding region consists of 639 amino acids (McPherson *et al.*, 1992; Ito *et al.*, 1991). A detailed crystal structure of the enzyme has been established (Ito *et al.*, 1991**)** (Figure 1.2).

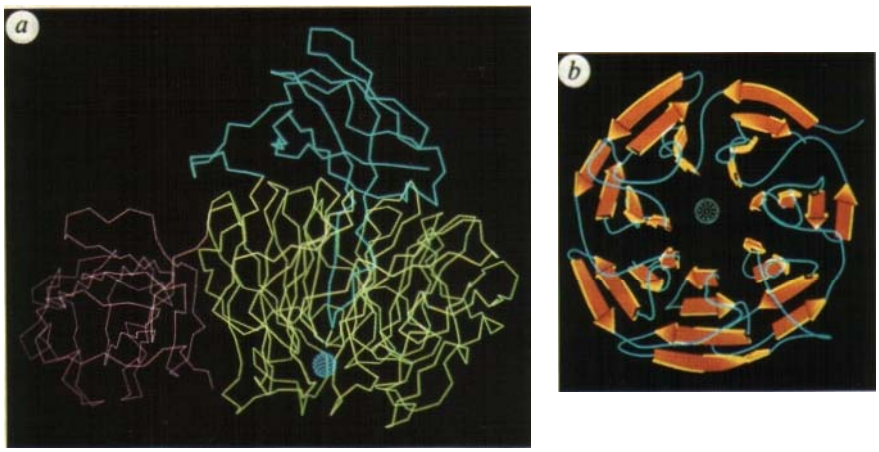


Figure 1.2 a. Cα backbone of galactose oxidase. The complete molecule is coloured according to domains (the first domain in red, the second yellow and the third blue). **b.** Computer-generated ribbon diagram (Pristle, 1988) of the second domain, looking along the pseudo-sevenfold axis.

The mature enzyme comprises three predominately β-sheet domains with only a single α helix (Ito *et al.*, 1991 and 1994) (Figure 1.2). This preponderance of β structure both within and between domains probably contributes to the remarkable stability of the enzyme, which is active in 6 M urea (Kosman *et al.*, 1974). Domain I consists of eight β-strands in a jelly-roll motif with a five stranded antiparallel β-sheet facing a three-stranded antiparallel β-sheet. Domain II is the largest, with a seven-bladed β-propeller fold surrounding a central cavity. 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. Three of the four protein ligands to the copper, Tyr-272, Tyr-495, and His-496, are provided by domain II. Domain III is a bundle of seven, mostly antiparallel, β-strands surrounding a hydrophobic core. One

long antiparallel β-ribbon penetrates into the central cavity of the domain II propeller, and at the tip of this loop is the fourth protein ligand to the copper, His-581. The overall structure of the copper-free precursor is similar to the mature enzyme, but with significant local differences (Figure1.4).The structure of the precursor also reveals the presence of an additional sugar at the surface of domain II. The electron density suggests at least a disaccharide, with the first ring stacking almost parallel to the ring of Tyr-484, but its identity has not been determined. As the sugar is some distance from the active site its biological relevance is uncertain.

1.1.3.1 Location and Interactions of the Pro-Sequence

The three-dimensional crystal structure of pro-GOase was determined by Firbank *et al.,* (2001). The structure was solved by molecular replacement using the structure of the mature enzyme as a model. The precursor structure was rebuilt and refined to 1.4 Å to give a final model with good stereochemistry. The 17-residue N-terminal pro-sequence is present as a loop between domains I and II of GOase, with residues –1 to –12 visible in the final electron-density maps, implying that residues –13 to –17 are mobile. The main chain of the pro-sequence makes a number of hydrogen bonds to regions of the mature enzyme sequence, while side chains also form hydrogen bonds to several residues, and there are 14 hydrogen bonds to water molecules (Firbank *et al*., 2003).

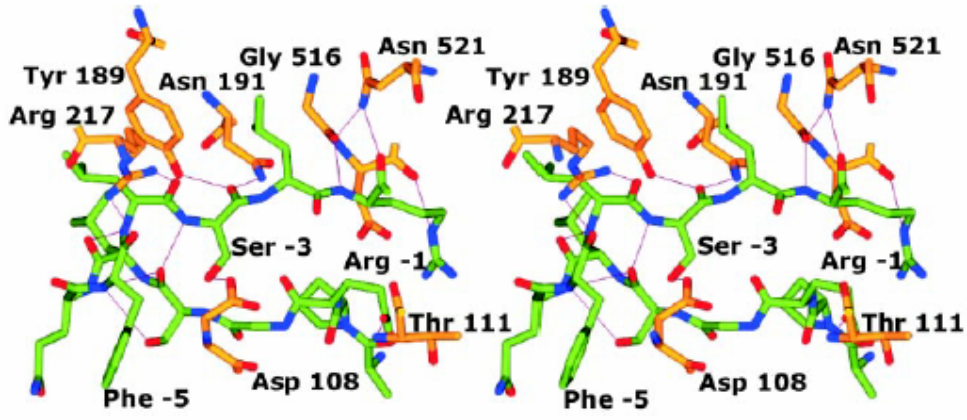


Figure 1.3 Stereoview of the residues of the pro-sequence and the amino acids with which they form hydrogen bonds. The carbon atoms of the pro-sequence are green, whereas those of the mature sequence are yellow (Firbank *et al.,* 2001).

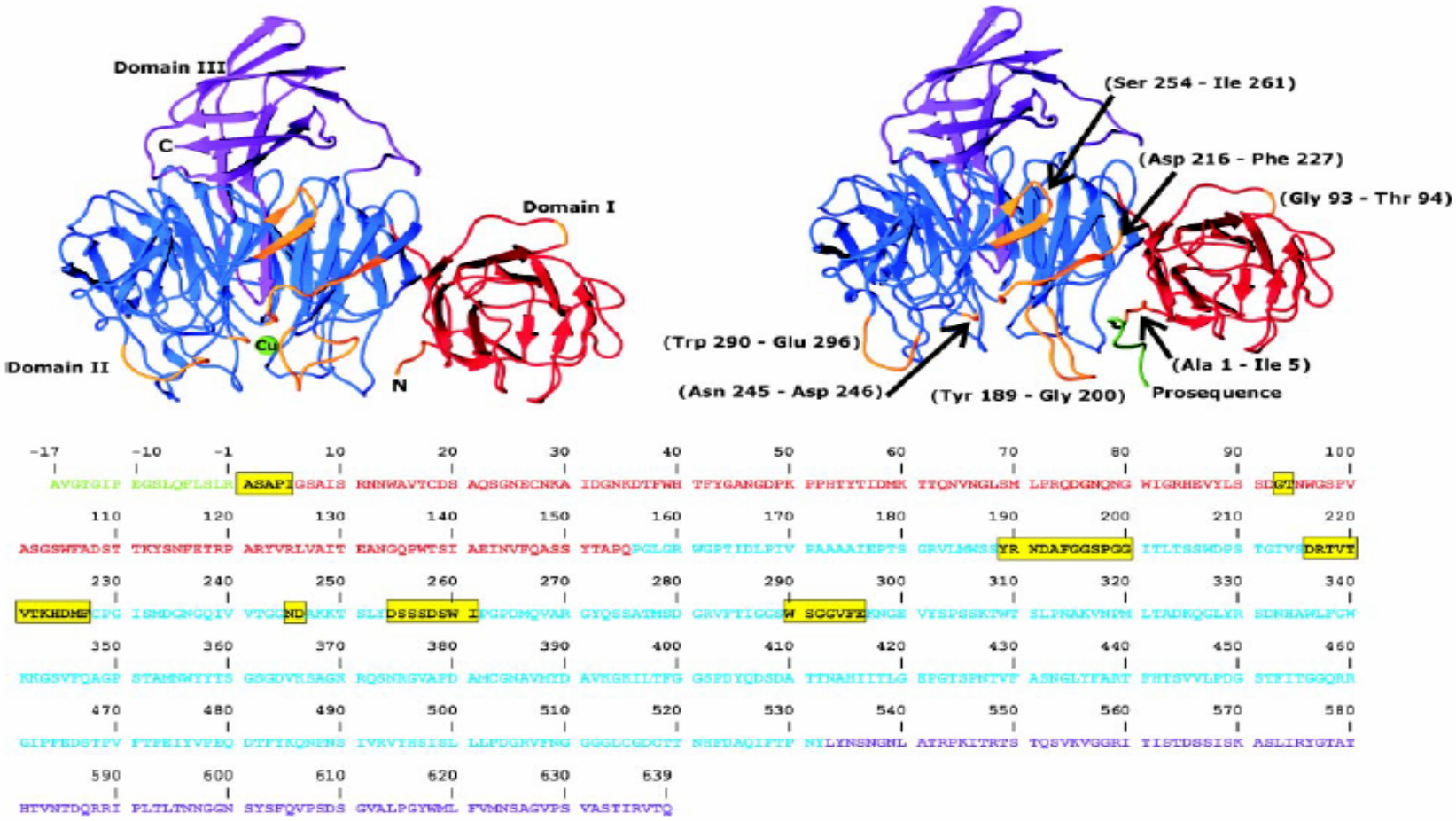


Figure 1.4 Structure of the mature form of galactose oxidase (*Upper Left*) and the precursor form (*Upper Right*). Domain I is red, domain II is blue, and domain III is purple. In the precursor form the N-terminal pro-sequence is green, and regions that differ from the mature structure by more than 2 Å are yellow. The sequence of galactose oxidase (McPherson *et al*., 1992) is shown (*Lower*) and coloured as above. The pro-sequence residues are numbered -17 to -1. (Firbank *et al*., 2001)

1.1.3.2 Structural Differences between Precursor and Mature Protein

The Cα positions of the precursor, excluding the pro-sequence, were aligned with those of the mature enzyme (Protein Data Bank ID code 1GOG). Whereas the majority of the two structures are very similar, overall (rmsd of 0.7 Å), residues colored yellow in Figure 1.4, 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 5.1 Å, with the greatest deviations found in domain II. Some residues of the active site show significant rearrangements, with movements of side chains and adjacent main-chain regions. In addition, the presence of the pro-sequence affects the position of Ala-1 moving it by over 4 Å from its location in the mature protein (Firbank *et al.*, 2001).

1.1.3.3 Cofactor Formation and Post-Translational Processing Steps

1.1.3.3.1 Tyrosyl–Cysteine Cross-Link and Catalytic Mechanism

Numerous enzymes use a variety of cofactors for achieving their impressive catalytic prowess. Generally, these cofactors are generated via complex multistep biosynthetic pathways involving many proteins. A less commonly encountered means of cofactor biosynthesis, but one that is found with increasing frequency, involves the posttranslational modification of the endogenous amino acids in the enzyme. These modifications can occur via autocatalytic processes or may be catalyzed by other auxiliary proteins (Xie and van der Donk, 2001).

Galactose oxidase is the best-characterized member of a family of enzymes known as radical copper oxidases (Whittaker, 2003). X-ray crystallographic studies on galactose oxidase have revealed the structural basis for the unusual reactivity of galactose oxidase: the protein contains a novel post-translational covalent modification, a cross-link between tyrosine and cysteine side chains, forming a tyrosyl–cysteine (Tyr–Cys) dimeric amino acid (Ito *et al.*,1991). Spectroscopic and biochemical studies have demonstrated that this Tyr–Cys site is the redox-active site in the protein, forming a stable free radical upon mild oxidation (Whittaker and Whittaker, 1990; Babcock *et al.*, 1992; Gerfen *et al.*, 1996).

Galactose oxidase is unusual among metalloenzymes in appearing to catalyze two-electron redox chemistry at a mononuclear metal ion active site (Kosman, 1985; Ettinger and Kosman, 1982; Hamilton, 1982). The absence of additional cofactors or metal ions required that the second redox site be a protein group, most likely one of the potentially redox active amino acids tyrosine, cysteine, cystine, or tryptophan (Whittaker and Whittaker, 1988). The second redox active centre necessary for the reaction was found to be situated at a tyrosine residue (Whittaker and Whittaker, 1990).

The active site of galactose oxidase is a shallow, exposed copper complex, in which the metal is bound by four amino acid side chains: two tyrosines (Tyr272 and Tyr495) and two histidines (His496 and His581) (Ito *et al.*, 1991). As indicated above, one of the two tyrosines (Tyr272) has been found crystallographically to be cross-linked at the Cε carbon of the phenolic side chain to the Sγ sulfur of Cys228, forming tyrosyl–cysteine (Tyr–Cys). The thioether bond that links the two residues affects both the structure and reactivity of the protein. Structurally, the cross-link contributes to the rigidity of the active site, similar to the effect a disulfide bond would have on the protein. However, unlike a disulfide bond, the thioether bond is formed irreversibly and is not susceptible to reductive cleavage. The cross-link forms spontaneously in the protein in the presence of reduced copper (Cu^{1+}) and dioxygen (vide infra) (Whittaker and Whittaker, 2003).

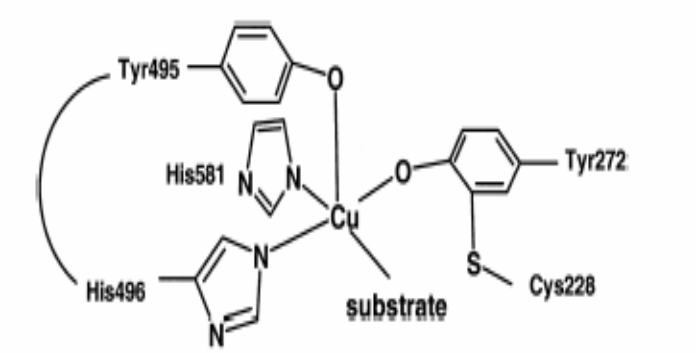


Figure 1.5 Schematic drawing of the active site of galactose oxidase

Galactose oxidase is the first representative of a new class of copper active sites, direct participation of the protein in active redox chemistry. Oxidation step is

required for catalytic activity, and is the basis of two-electron redox activity. If anaerobic addition of substrates takes place, this results in reduction of the two-electron redox unit (Whittaker and Whittaker, 1988). The enzyme functions as a 2-equiv oxidase with RCH_2OH primary alcohol/sugar substrates,

$$RCH_2OH + GOase_{ox} \rightarrow RCHO + GOase_{red}$$ (Whittaker, 1995)

followed by reaction with O_2

$$GOase_{red} + O_2 \rightarrow GOase_{ox} + H_2O_2$$ (Kain *et al*., 1996)

In GOase, the number of metal ions involved in the reaction does not match the number of electrons transferred. This paradox has been solved with the identification of a tyrosyl free radical incorporated into the redox unit during the catalytic cycle. The enzyme exists in three well-defined and stable oxidation levels. (Fontecave and Pierre, 1998)

$$\underset{(GOase_{ox})}{Cu^{2+}\text{-}Tyr^{\cdot}} \overset{+e^-}{\underset{-e^-}{\rightleftarrows}} \underset{(GOase_{semi})}{Cu^{2+}\text{-}Tyr} \overset{+e^-}{\underset{-e^-}{\rightleftarrows}} \underset{(GOase_{red})}{Cu^{+}}$$

After the substrate binds to the equatorial copper position the first step is a proton transfer from the alcohol to the axial tyrosinate (Tyr495). Next, a hydrogen atom is transferred from the substrate to the modified tyrosyl radical. The resulting substrate-derived ketyl radical is then oxidized through electron transfer to the copper center, yielding Cu(I) and aldehyde product. The two latter steps have been suggested to occur simultaneously (Wachter and Branchaud, 1996 and 1996; Wachter *et al.*, 1997). The Cu(I) and tyrosine are, finally reoxidized by molecular oxygen, regenerating Cu(II) and tyrosyl, and giving hydrogen peroxide as product (Himo and Siegbahn, 2003). The proposed catalytic mechanism for GOase is shown in Figure1.6.

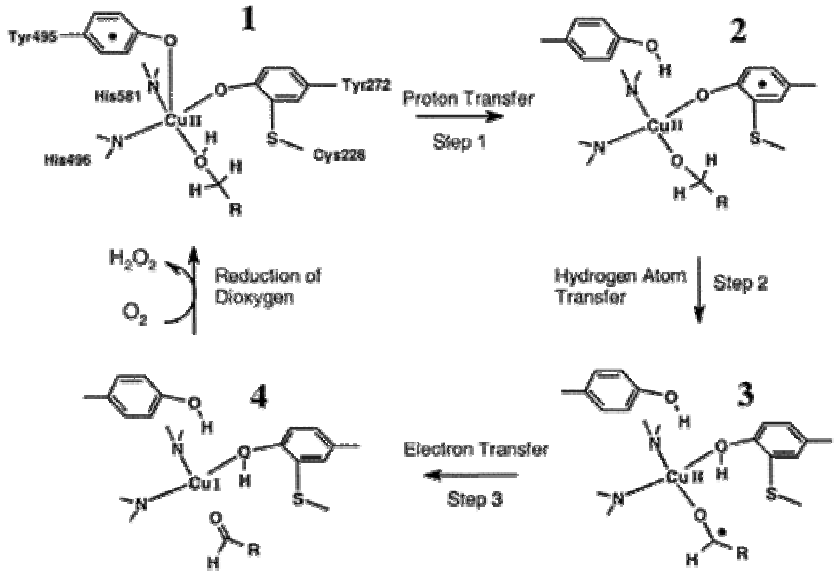


Figure 1.6 The mechanistic scheme for the catalytic cycle of galactose oxidase (Whittaker and Whittaker, 1988 and 1993; Branchaud *et al.*,1993)

1.1.3.3.2 Self-Catalytic Maturation of Galactose Oxidase

A growing number of enzymes have been reported that undergo posttranslational modifications of amino acids within their active sites to create a wide variety of structurally and functionally diverse cofactors. These modifications can be divided into two general classes. One involves proteins that undergo one-electron oxidations of amino acids to provide amino acid radicals on tyrosine, glycine, tryptophan and cysteine residues (Stubbe and van der Donk, 1998). The second class undergoes more extensive posttranslational modifications that involve new bond-forming reactions (Okeley and van der Donk, 2000). Amine oxidases and lysyl oxidase contain the quinone cofactors 2,4,5-trihydroxyphenylalanine quinone (TPQ) and lysyl tyrosylquinone (LTQ), respectively (Janes *et al.*, 1990; Wang *et al.*, 1996). A notable recent cryocrystallographic study of a copper amine oxidase has revealed the structures of some intermediates in the autocatalytic conversion of tyrosine into 2,4,5-trihydroxyphenylalanine quinone in the presence of copper and oxygen (Kim *et al.*, 2002). The most complex cofactor described to date 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 terminal electron transport protein cytochrome *c* oxidase (CcO) is posttranslationally modified through a crosslink between histidine and tyrosine in both bacteria (Ostermeier *et al.,*1997) and mammals (Yoshikawa *et al.,* 1998). His and Tyr residues are also crosslinked in catalase HPII from *Escherichia coli*, but the linkage in this protein involves a bond between the C_β of tyrosine and N_δ of histidine (Bravo *et al.,* 1997).

In galactose oxidase, as well as glyoxal oxidase, a tyrosine residue is crosslinked by a thioether bond between Cε of the aromatic ring and the sulfur atom of a cysteine (Ito *et al.,* 1991). This crosslinked tyrosine serves as a ligand to a catalytically essential copper and is oxidized to the tyrosyl radical form in the active state of the protein (Whittaker and Whittaker, 1988 and 1990).

The discovery of posttranslationally modified endogenous cofactors has led to great interest into the mechanisms of their formation (Xie and van der Donk, 2001). Some of these structures, such as tryptophan tryptophyl quinone (TTQ) in methylamine dehydrogenase (Graichen *et al.,* 1999) and formylglycine in sulfatases (Szameit *et al.,* 1999) are generated by accessory proteins. Others, on the other hand, including TPQ (Cai and Klinman, 1994; Matsuzaki *et al.,* 1994; Ruggiero *et al.,* 1997), the MIO structure in phenylalanine ammonia lyase (Schwede *et al.,* 1999), and the chromophore in green fluorescent protein (Heim *et al.,* 1994) are produced by autocatalytic processes.

Unique among this latter group is galactose oxidase because its self-catalytic maturation following the removal of its signal sequence, involves at least three processing events:

1. Cleavage of an 8 amino acid pre-peptide

2. Cleavage of a 17-aa N-terminal pro-peptide

3. Formation of the thioether bond (a two-electron process) and a further one electron oxidation to give the cation radical active form of the enzyme (McPherson *et al.,* 1992)

The latter two events have been shown to take place *in vitro* on the addition of copper and oxygen, even in the presence of protease inhibitors. It was shown that both reactions do not proceed when the protein is heterologously expressed (in *Aspergillus nidulans*) and purified under strictly metal-free conditions (Rogers *et al.,* 2000). In copper-limited conditions, heterologous expression of galactose oxidase results in three forms of the protein identifiable as distinct bands on SDS-PAGE (Figure 1.7). The lower band is mature galactose oxidase, which runs anomalously on SDS-PAGE (at 65.5 kDa) (McPherson *et al.,* 1993; Baron *et al.,*1994) owing to the presence of the thioether bond that introduces an intramolecular loop between residues Tyr228 and Cys272. The middle band (68.5 kDa) migrates with the expected apparent molecular mass of mature galactose oxidase, indicating that this is an incompletely processed form of the protein mirrored by the variant C228G, which is unable to generate a thioether bond (Baron et al, 1994).The upper band (70.2 kDa) is a precursor of galactose oxidase, which has been revealed by N-terminal sequencing to have the additional 17 amino acid at the N terminus (Rogers *et al.,* 2000).

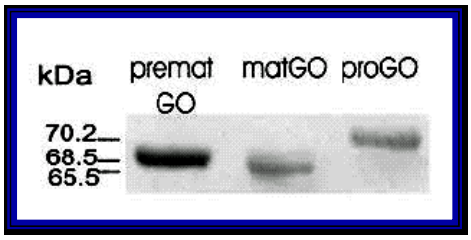


Figure 1.7 SDS-PAGE mobility assay showing three species of GOase. **premature GO**, (no pro-sequence, no thioether bond); **mature GO,**(no pro-sequence, thioether bond formed); **pro-GO,** pro-sequence present, no thioether bond) (Firbank *et al.,* 2003)

1.2 Proteolytic Processing of Extracellular Proteins

Proteolytic processing of fungal extracellular proteins as outlined below, was analyzed by Ögel, Z.B.(1993, unpublished).The targeting and translocation of newly synthesized proteins to various cellular compartments or to the extracellular medium

requires an efficient sorting system involving the precise recognition and transfer of proteins to their target locations. Although this process is not fully understood and may vary between different cell types and organelles it is, in general based on the presence or absence of peptide signals within the structure of individual proteins. In the case of extracellular proteins, this signal is most exclusively 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 (Ögel, unpublished study). In eukaryotes, following entry to the ER usually by co-translational translocation, the signal peptide is cleaved by signal peptidase and is therefore not associated with the mature form of the protein (Rapoport *et al.,* 1992). Those proteins destined for secretion from the cell are transferred from the ER to the golgi complex and subsequently, often following glycosylation, to the outer surface via secretory vesicles (Kelly, 1985).

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

Transient peptides, that undergo proteolytic processing, may be found at the N- or C terminus of proteins. At the N-terminus transient peptides may consist of only a signal peptide or may also contain one or more additional peptides, namely pro-peptide(s) between the signal peptide and mature protein. Such N-terminal transient peptides are often referred to as a leader peptide. In light of the widely accepted

passive bulk flow hypothesis, proteins translocated into the ER are secreted unless they possess additional signals for targeting to other cellular compartments (Wieland *et al.,* 1987). Since passage into the ER requires only an N-terminal signal peptide, this should normally be sufficient for export. However, many extracellular proteins appear to be processed in more than one stage. For example, pro-peptides have consistently been identified in proteases where they serve to maintain an intracellular inactive configuration or are important for proper folding (Silen and Agard, 1989; Zhu *et al.,* 1989; Simonen and Palva, 1993). In mammalian cells, examples exist where pro-regions are processed to produce alternative forms of secretory hormones (Benoit *et al.,* 1987).

In recent years several filamentous fungal genes encoding extracellular proteins have been cloned and sequenced allowing the analysis of their transient peptides. Such a comparative analysis should contribute to studies directed towards an understanding of protein secretion in filamentous fungi because the efficiency of secretion and access to the correct secretory route, as well as proper folding and stability could all be influenced by transient peptides and particularly by those at the N-terminus (Ögel, unpublished study).

The presence of a pro-peptide in filamentous fungal extracellular proteins has been mostly proposed due to the nature of residues at the leader peptide-mature protein junction, rather than the length of the leader peptide. This has been the case particularly in those proteins where a basic residue is found preceding the cleavage site at the mature N-terminus. In none of the 161 eukaryotic and 36 prokaryotic signal sequences analysed by von Heijne (1986a) is a basic residue found at position -1, relative to the cleavage site. Although unusual residues can be used in position -1 when no better cleavage site is available in the vicinity (von Heijne, 1986a), in almost all of the proteins with a basic residue preceding the cleavage site a favourable signal sequence cut site is found further upstream, supporting additional proteolytic processing (Ögel, unpublished study).

1.2.1 Proteolytic Processing of Propeptides

1.2.1.1 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 cleavage takes place following a 'Lys-Arg' pair whereas 'Lys-Lys' and 'Arg-Arg' pairs are less encountered. Dibasic processing of pro-peptides is likely to be a conserved feature rather than a similarity due to high homology and there is no correlation between protein function and the presence of a pro-peptide. Among the proteases, neutral protease II (NpII), a zinc metalloprotease of *A. oryzae*, has been shown to have a pro-peptide that is cleaved following two basic residues (Tatsumi *et al.,*1991).

1.2.1.2 Monobasic Processing

A considerable number of filamentous fungal extracellular proteins possess a monobasic cleavage site at their leader-mature protein junction. It is suggested that the presence of a pro-hexapeptide 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 (Bussink et *al.,* 1991b). 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 adjacent to a Leu or Ile. The fact that the pro-peptides contain both hydrophilic and hydrophobic residues and the absence of sequence homology 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 significant. In filamentous fungi there are no examples where proline is present immediately before or after the basic residue at the cleavage site. Nevertheless, since the role of proline is suggested to be one 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 where monobasic processing takes place.

It is clear from previous studies that the mature part of the protein also plays a role in the export process (Li *et al*., 1988). Minor factors such as the net charge or hydrophobicity at the N-terminus of filamentous fungal exported proteins could likewise influence translocation across the cell membrane or the cell wall. It has been pointed out that there is usually a net charge of zero or less in the region surrounding the signal peptide cleavage site and the first five residues at the N-terminus of the mature protein (von Heijne, 1986b; Li *et al*., 1988; Boyd and Beckwith, 1990). Hydrophobic regions are also found in the membrane spanning domains of integral membrane proteins and are also important for the export process due to the essential hydrophobic core of the signal peptide. It is not known whether hydrophobicity at the N-terminus of mature proteins is a factor that may influence the translocation of proteins into the extracellular medium (Ögel, 1993).

1.2.2 Autocatalytic Processing

As it is situated above some precursor proteins undergo proteolytic processing, whereas cleavage of pre-sequences of some precursors is not due to extraneous or intrinsic protease activity. As it is observed in GOase, cleavage of pro-sequence occurred despite the protease inhibitor cocktail just by using available reagents such as copper and dioxygen (Rogers *et al.*, 2000). At least one other protein generates its cofactor by posttranslational modification involving an autocatalytic cleavage of a peptide bond. Histidine decarboxylase is composed 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 derives from Ser-82 of the pro-enzyme (van Poelje and Snell, 1990). In GOase, however, cleavage of the pro-sequence and formation of the cofactor must be separate processes because an intermediate form lacking the N-terminal pro-peptide but without the crosslink has been identified (Rogers *et al.*, 2000). This observation suggests that the mechanisms of both modifications may be elucidated in future investigations.

Amyloid beta precursor protein (APP) which is the source of fibrillary peptides that are 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. The process in the two proteins could be related, their functions may overlap and their distributions coincide. It is proposed that PrP catalyses its own cleavage, the C-terminal fragment functions as an α secretase and the N-terminal segment chaperones the active site; the α secretase releases anticoagulant and neurotrophic ectodomains from APP. Nothing is known of the enzyme responsible for the endoproteolytic cut except that it is inhibited by metal chelators (Jimenez-Huete,1998). A radically new function of PrP is suggested: a proenzyme that catalyses its own cleavage and the release of neurotrophic APP from its membrane anchor. The prosegment, as in many pro-enzymes, may act as a chaperone (Abdulla, 2001).

1.2.3 Functional Roles of Putative Propeptides

In Ögel's unpublished study, transient peptides of filamentous fungal extracellular proteins and their potential roles were also analysed. Based on previous studies on bacterial proteases, long pro-peptides (50 to over 100 residues) are important for proper folding and activation of the enzyme; rather than the export process (Silen and Agard, 1989; Simonen and Palva, 1993). This could also be the case for filamentous fungal proteases. There is no clear evidence as to the role of short pro-peptides (6-9 residues). They could be involved in aspects of the secretion process, 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 studies on heterologous expression of active cutinase of *Fusarium solani* in E. *coli* (Soliday *et al.,* 1984; Martinez *et al.,* 1992). There is some evidence that the leader region of galactose oxidase may also be required for protein folding. These enzymes, and perhaps certain other filamentous fungal secreted proteins, may have evolved a mechanism for folding into their active conformation with the aid of transient pro-regions. In addition to allowing protein folding, pro-peptides could also be involved in retarding the folding process (Hardy

and Randall, 1992). From studies in prokaryotes it is known that an early function of the leader sequence is to slow down the folding process to allow interaction with the chaperone. This process appears to be essential for export because in order to be transferred across a membrane a soluble protein should not be in its final thermodynamically stable folded form (Hardy and Randall, 1992).

It has been suggested that the leader peptide protects producing strains from suicide by keeping the enzyme in an inactive state until processing takes place in the Golgi system (Lamy and Davies, 1991). Evidence for the function of such short pro-peptides in *Bacillus* species indicates that they are unlikely to play an active role in secretion but one possibility is that they might stabilize the secreted proteins by allowing folding into a protease-resistant conformation (Simonen and Palva, 1993). In higher eukaryotes, there are mainly two routes for the secretion of proteins; constitutive and the regulated pathways (Kelly, 1985). Although relatively short pro-peptides, often at the N-terminus, are mostly responsible for intracellular targeting, deletions resulting in the removal of pro-peptide regions of certain mammalian proteins have not affected targeting of the mature protein to the regulated pathway (Chidgey, 1993). There is some evidence that at least some of the secreted proteins may pass through the vacuole before being transported to the extracellular medium (Peberdy, 1994). This is supported by an early study on the regulation of galactose oxidase synthesis and secretion which states that the export of the enzyme to the extracellular medium is a regulated process influenced by pH and culture density (Shatzman and Kosman, 1977). It was suggested that at pH below 7.0 the enzyme is stored in the vacuole portion of the cell and secreted when the pH is raised to 7.0, accompanied by a decrease in vacuolization (Shatzman and Kosman, 1977). The vacuole in filamentous fungi may indeed have a role in the storage and/or processing of certain secreted proteins prior to export and certain pro-peptides could have a role in the sorting process (Ögel, unpublished study).

1.2.3.1 Possible Roles of GOase Pro-sequence

The N-terminal part of cloned *gaoA* gene includes a 16 amino acid signal peptide, an 8 amino acid putative pre-peptide and 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 self-processing reaction (Rogers *et al*., 2000). However, functional expression of low levels of mature GOase in *E. coli* was accomplished in the absence of the pro-sequence (Sun *et al.,* 2001).

Structural comparison of pro-GOase with mature GOase reveals overall structural similarity, but with some regions showing significant local differences in main chain position and some active-site-residue side chains differing significantly from their mature enzyme positions. These structural effects of the pro-peptide suggest that it may act as an intramolecular chaperone to provide an open active-site structure conducive to copper binding and chemistry associated with cofactor formation (Firbank *et al.,* 2003). An intriguing aspect of pro-GOase processing is the mechanism by which the pro-sequence is cleaved in a copper-dependent manner. Although structural data reveal the site of pro-sequence cleavage, it does not point to any obvious mechanism. Copper may bind transiently, near the cleavage point. It is interesting to note that site-specific cleavage events have been observed when 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 GOase is also near the disulphide formed between Cys-515 and Cys-518, and experiments are underway to probe any role for this site in pro-sequence cleavage (Firbank *et al.,* 2003).

1.3 GOase Processing in Different Heterologous Systems and Mutational Variants

1.3.1 Processing in Fungi

According to the predictive algorithm of von Heijne, if Arg(-1) can be a target for signal peptidase cleavage and 16-18 amino acid can act as a a signal, the remaining 23-25 residues of the leader sequence must act as a pro- or pre-pro-peptide, removed at a later stage during the secretion process. To provide evidence for the two-stage processing of galactose oxidase leader peptide, a most direct approach would be the isolation of pro-galactose oxidase and determination of its N-terminal amino acid

sequence. For the isolation of galactose oxidase precursor, mutagenesis at the leader-mature enzyme junction was carried out to prevent processing of the putative pro-peptide. At the monobasic cleavage site of galactose oxidase Arg-Ala was substituted into Phe-Phe. A second Arg-Ala pair is not found within the 639 amino acids of mature galactose oxidase. This mutation resulted in an enzyme that was correctly processed but which displayed a significantly low efficiency of export compared to wild type galactose oxidase heterologously expressed in *A. nidulans* (Ögel, 1993). The mutant enzyme was located to a region close to the outer surface, either the cell membrane or the cell wall. Since the only difference between the mutant and the wild-type enzyme appeared to be a single amino acid residue at the N-terminus (Ala1Phe), this has raised the question as to whether 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 caused by a single basic amino acid difference at the N-terminus. Phe at the beginning of mature galactose oxidase may influence export by altering the balance at the N-terminus, in this case by changing the degree of hydrophobicity (Ögel, 1993).

In later studies, this proposed two-step cleavage by Ögel (1993) and the presence of the pro-sequence was confirmed by copper limited and metal-free growth and the crystallographic studies, as explained before in section 1.1.4.1 and 1.1.4.3.2 (Rogers *et al*., 2000; Firbank *et al.*, 2001).

Heterologous expression (Baron *et al*., 1994) of the *Fusarium* protein in *Aspergillus nidulans* under copper-limited conditions resulted in the appearance of multiple protein forms (Figure1.7). These different species of GOase undergo similar posttranslational modifications as it does in its native host *Fusarium ssp*. As it is seen in Table 1.1 pro-sequence initiation site is the same with the wild type pro-GOase.

Table 1.1 Summary of galactose oxidase forms

	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, pre- and a pro-peptide. The pre-peptide cleavage site appears to be after the His-Lys at position -19 and -18 (Figure 1.8). This pre-peptide is likely to be cleaved, after the removal of the signal peptide, by a dibasic processing protease (Kex2-like) during secretion. Following cleavage of the pre-peptide, the pro-peptide is cleaved by autocatalytic cleavage.

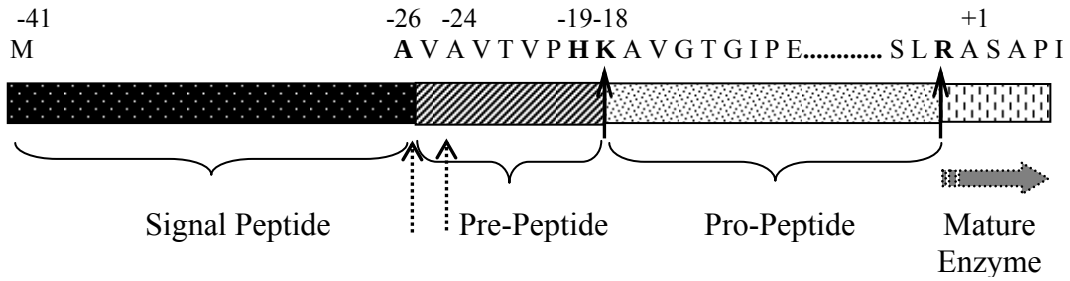


Figure 1.8 N-terminal part of GOase and cleavage sites for different processing events

1.3.2 Processing in Yeast

In order to improve the activity of GOase towards appropriate substrates there have been recent reports of directed evolution based on error-prone PCR (Delagrave *et al.*, 2001). Galactose oxidase from *Fusarium spp.* (NRRL 2903) was expressed in *Pichia*

pastoris X-33. The constructs used in this work contained the coding sequence for the 17 amino acid pro-form of the enzyme linked in-frame to the α-mating factor secretion signal of the vector, and are therefore referred to as Pro-GOase.

The results identified mutations at Cys383, Tyr436 and Val494 that were subsequently combined by subcloning and assayed for additive contributions to enhanced activity. Series of single and double mutations comprising V494A, C383S and Y436H were generated for comparison with the triple mutant clone, C383S/Y436H/V494A. These mutants were all tested with respect to catalytic efficiency and molecular masses.C383S/V494A double mutant and C383S/Y436H/V494A show enhancement in activity when compared with wild-type. For the majority of the samples there is good agreement between the expected and observed molecular masses whereas some of the variants differ from the expected value by 10 or more mass units, although the reason for this is unclear.

The pro-GOase protein produced in *P. pastoris* displayed a mixture of N-terminal extended species rather than the expected 17 amino acid N-terminal pro-sequence observed when produced in *Aspergillus nidulans* (Rogers *et al*., 2000). The majority of the protein has the 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 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 at Ala +1 (Baron, *et al*., 1994; McPherson *et al*., 1992). This observation is consistent with a previous report of mixed N-terminal sequences from constructs expressing GOase containing a pro-sequence in *P. pastoris* (Whittaker and 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. Expression of the full-length native sequence results in a mixture of partly processed and mature galactose oxidase. The reason for the differences in N-terminal processing by these two hosts is unclear; however, the recombinant enzymes prepared from *P. pastoris* and *A.*

nidulans exhibit similar properties including kinetics, UV/vis spectra and three dimensional structure (S. 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, indicating that the polypeptide has a ragged end, consistent with inefficient processing of the pro-peptide and partial digestion of the presequence linker. One of the possible signal peptidase cleavage sites consistent with these sequence results (AVA-VTV) is indicated in Table 1.2. The clean N-terminal sequence of the protein expressed as a gla fusion construct (Table 1.2) confirms that the gla pro-sequence is correctly processed to generate the authentic mature protein.

Table 1.2 N-terminal sequence analysis of recombinant galactose oxidase expressed in *P. pastoris* (Whittaker and 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 because the pro-sequence contains a suboptimal cleavage site. Inspection of the signal sequence indicates that correct processing would require cleavage at an isolated Arg residue rather than the preferred Lys-Arg dibasic recognition site for Kex2 cleavage, resulting in a protein product that is heterogeneous and exhibits low specific activity (Whittaker and Whittaker, 2000). The requirement for a Lys-Arg motif for Kex2-like maturation of recombinant secretory proteins in *Pichia* has previously been observed (Martı´nez-Ruiz *et al.*, 1998).

In this study we suggest that the above interpretations of Whittaker are incorrect or incomplete. This is because the signal sequence is likely to be cleaved following Ala(-26). This site is one of the sites giving a high score by the analysis of von Heijne (Ögel, 1993). The remaining peptide is cleaved as explained before in section 1.3.1 and 1.1.4.3.2.

1.3.3 Processing in Bacteria

According to the study of McPherson *et al*., (1993), in which *E. coli* expression system was investigated for GOase, it was found that *E. coli*-expressed GOase is a cytoplasmic protein even when the leader sequence is present. The majority of the GOase produced is inactive; GOase activity has only been observed when the leader sequence is present, suggesting it may be important for correct protein folding in *E. coli.* A similar observation has been reported for heterologous expression of active cutinase of *Fusarium solani* which has a leader sequence of 32 residues could be achieved in E. *coli* only in the presence of the 15 amino acids long putative pro-peptide (Soliday *et al.,* 1984; Martinez *et al.,* 1992). Without the pro-peptide cutinase activity was not detected. However, the pro-peptide was not efficiently processed in *E. coli* and, therefore, procutinase was used in X-ray crystallographic structural studies although no density was observed for the 16 amino acids at the N-terminus indicating a disordered structure (Martinez *et al.,* 1992). Alternatively, the leader sequence may direct the inefficient export of GOase to the periplasm of *E. coli*, resulting in low levels of active enzyme that arise due to the thioether-bond formation. It was thought that, bacterial expression does not represent a convenient system for the purification of GOase variants.

The application of GOase mentioned before would benefit from access to enzyme variants that are more stable and active towards non-natural substrates. A prerequisite to enzyme modification by powerful directed evolution methods (Arnold, 1998; Petrounia and Arnold, 2000) is functional expression in a host organism that permits creation and rapid screening of mutant libraries. *E. coli* is an excellent host for directed evolution, but does not support functional expression of many important eukaryotic enzymes. To date, all biochemical studies of GOase 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 and Whittaker, 2000) expression systems not suitable for directed evolution. Expression of GOase has been attempted as it was mentioned above (McPherson *et al*., 1993), but functional enzyme was obtained only as a *lacZ* fusion (Lis and Kuramitsu, 1997). Functional expression of GOase in *E. coli* was later achieved by directed evolution (Sun *et al*., 2001).

According to this study, to increase the total activity of GOase in *E. coli*, random mutagenesis was applied to the entire mature *GOase* gene also to just the region of the gene encoding domains II and III which are responsible for catalytic activity (McPherson *et al*., 1993). The highest activity mutant was identified which carries S10P, M70V, P136, G195E, V494A, and N535D mutations having advantageous effects on thermostability and expression. Also it was shown that the effects of mutations were not cumulative. The 30-fold increase in total activity for this best variant relative to wild-type reflects an 18-fold increase in GOase expression and a 1.7 fold increase in catalytic efficiency. The broad substrate specificity of wild type GOase is retained in the evolved enzymes (Sun *et al.,* 2001). Oxidized form of the enzyme produced in *E. coli* has redox potential and stability comparable to the GOase from *Fusarium* (Baron *et al.,* 1994; Reynolds *et al*., 1997). Enhanced thermostability and improved expression in *E. coli* facilitate protein purification and characterization.

In this thesis study, *pro-GOase* gene which was previously developed by directed evolution method (Sun *et al.,* 2001) was used. Site-directed mutagenesis was carried out on this gene and expression studies were performed in *E. coli* to understand the autocatalytic pro-peptide cleavage mechanism of GOase.

1.4 Aim of the Study

The discovery of posttranslationally modified endogenous cofactors has led to great interest into the mechanisms of their formation (Xie and van der Donk, 2001).The studies on galactose oxidase have demonstrated that no enzymes or accessory

proteins are required for the maturation of GOase except the addition of Cu^{+2} and dioxygen. This oxidative and copper-mediated proteolysis event is new to the literature (Rogers *et al*, 2000).

The mechanism of maturation of the pro-GOase has not been definitely determined yet. This thesis aims finding out autoprocessing mechanism of galactose oxidase pro-peptide by applying site-directed mutagenesis to the amino acid residues that may play an essential role in self-cleavage. Multiple mutations within and around the cleavage site were carried out for this aim. Since pro-sequence may act as an intramolecular chaperone responsible for positioning of active site residues, an inhibitor of catalytic activity or may enhance the level of secretion (Eder and Fersht, 1995; Baardsnes, 1998), the idea of the possibility of clarifying the self-processing mechanism became stronger by the mutations put into practice here.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and Enzymes

The chemicals and the enzymes involved within the experiments are listed in Appendix A with their suppliers.

2.1.2 Growth Media, Buffers and Solutions

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

2.2 Methods

2.2.1 Bacterial strains

XL1-Blue and BL21 Star (DE3) *E .coli* strains were chosen for propagation of the plasmid carrying *gao* gene and for high-level heterologous expression, respectively.

Table 2.1 Bacterial strains and their genotype

STRAINS	**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 hsd*S_B ($r_B^- m_B^-$) *gal dcm rne131* (DE3)

2.2.1.1 BL21 Star (DE3) as a Host Strain

The most widely used host is BL21 Star (DE3), which has the advantage of being naturally deficient in both *lon* and *ompT* proteases. This results in a higher yield of intact recombinant proteins. The suffix "DE3" indicates that the host is a lysogen of *DE3, a lambda derivative that has the immunity region of phage *21*, carrying a DNA fragment containing the *lacI* gene and T7 RNA polymerase gene under the control of the IPTG-inducible *lacUV5* promoter. Such strains are used to induce high-level protein expression in T7 promoter-based systems.

2.2.2 Cultivation and Storage of the Strain

For plasmid isolation, XL1-Blue cells were cultivated in proper amount of antibiotic containing LB Medium (Appendix B) at 37 ˚C by shaking at 200 rpm overnight in a Falcon tube or flask having a volume of nearly 4 times larger than the culture volume.

For expression of *gao*, BL21 Star (DE3) cells were cultivated in 5 ml LB medium containing 5 µl ampicillin as a preculture. After 6 hours, 25 µl of the culture was transferred into 25 ml fresh medium and incubated overnight. Then 5 ml of this overnight culture was used for 250 ml medium inoculation. During the experiment, the conditions were 37˚C and 200 rpm.

Glycerol stocks were prepared both for XL1-Blue and BL21 Star (DE3) strains including all types of mutations and stored at -80 ˚C. For 1 ml glycerol stock; 800µl of the overnight culture was mixed with 200 µl previously sterilized 100% glycerol.

Strains were also kept on plates containing appropriate antibiotics (Appendix B), and were subcultured at suitable intervals and maintained at 4 ˚C.

2.2.3 Competent *E .coli* Preparation

2.2.3.1 Preparation of Heat Shock Competent *E. coli* Cells with $CaCl_2$

A single colony was picked from agar plate containing the desired strain and inoculated into 5 ml of LB-broth and was incubated overnight at 37˚C in orbital incubator at 200 rpm. 100 ml LB-broth was inoculated with 1 ml of overnight culture and grown at 37°C until the optical density at 550 nm reached 0.4-0.5. The culture was dispensed into two 50 ml falcon tubes and chilled on ice for 10 minutes. Then the tubes were centrifuged at 6000 rpm for 5 minutes at 4˚C. After discarding the supernatant, cells were resuspended in a total volume of 50 ml ice-cold solution I (25 ml for each tube), (Appendix B) kept on ice for 15 minutes and centrifuged at 6000 rpm for 5 minutes at 4°C. Then the pellet was resuspended in a total volume of 7 ml ice-cold solution I (3.5 ml for each tube) (Appendix B).Then, glycerol was added to a final concentration of 20%. 300 µl of cells were dispensed into sterile eppendorf tubes and stored at -80°C until use.

2.2.3.2 Preparation of Heat Shock Competent *E. coli* Cells with $RbCl_2$

Required strain was streaked from glycerol stock onto LB-agar plate (Appendix B) and was grown overnight at 37°C. A single colony was picked and inoculated into 5 ml of SOB (LB can be used instead) and was grown overnight at 37°C in an orbital shaker incubator at 200 rpm. 50 ml of pre-warmed SOB media (or LB) was inoculated with 1 ml overnight culture and was grown at 37°C in orbital incubator at 200 rpm until OD_{595} was 0.4 (2-3 hrs). Cells were then transferred into 50 ml Falcon Tube and chilled on ice for 5 minutes, followed by centrifugation (Hermle rotor) at 3000 rpm for 10 minutes. The supernatant was discarded and cells were resuspended

in 20 ml of ice-cold Tfb1 buffer and incubated on ice for 5 minutes. After centrifuging cells at 3000 rpm for 10 minutes and discarding the supernatant the tube was wiped dry with tissue. Pellet was resuspended in 2 ml of Tfb 2 buffer and incubated on ice for 15 minutes. Upon dispensing 250 □l aliquots into sterile eppendorf tubes, competent cells were frozen in liquid nitrogen and stored at –70°C.

2.2.4 Site-Directed Mutagenesis

In this study, all mutations on ProGON1 and ProGOMN1 constructs were done by using *QuikChange® Site-Directed Mutagenesis Kit* (Stratagene) according to the manufacturer's instructions with some modifications, as described below in section 2.2.4.2. This method is based on four major steps; (for schematic drawing see Figure 3.5)

1. Preparation of the plasmid
2. Thermal cycling
3. Digestion of parental DNA by *Dpn*I
4. Transformation and repair of the mutated DNA

2.2.4.1 Mutagenic Primer Design

Primer oligonucleotides were designed as follows: Both of the primers contained the desired mutations and annealed to the same sequence on opposite strands of the plasmid. Primers were between 25 and 45 bases in length with a melting temperature (*T*m) of 78°C or higher. The desired mutation (deletion or insertion) was placed in the middle of the primer with ~10–15 bases of correct sequence on both sides, GC content of the primers was 40% optimally and primers terminated in one or more C or G bases. Primers were purified in order to prevent unwanted mutations, and to increase mutation efficiency.

2.2.4.2 Mutant Strand Synthesis Reaction (Thermal Cycling)

During the mutant strand synthesis and transformation manufacturer's instructions were followed by slight modifications.

A 50 µl reaction mixture contains;

- 10 x reaction buffer for KOD Hot Start DNA Polymerase to give a final concentration of 1x reaction buffer
- 0.2mM dNTP mix
- 1mM $MgSO_4$
- ~20 ng of dsDNA template
- 125 ng of 5' oligonucleotide primer
- 125 ng of 3' oligonucleotide primer
- 1 unit KOD Hot Start DNA polymerase (Novagen) (1U/µl)
- Sterile ddH_2O was added to a final volume of 50 µl.

Mineral oil was not added to prevent the evaporation since the gradient 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® Site-Directed Mutagenesis Kit* (Stratagene)

Segment	Cycles	Temperature	Time
1	1	94˚C	30 seconds
2	24	94˚C	30 seconds
		55˚C	1 minute
		68˚C	8 minutes (1 minute/kb of plasmid length)

2.2.4.3 *Dpn* I Digestion of the Amplification Products

In order to digest the parental DNA, *Dpn*I restriction digestion was applied. *Dpn*I is a restriction enzyme digesting only the methylated DNA isolated from *E. coli*, however, it can not act on 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 by using a small, pointed pipet tip. Each reaction mixture was gently and thoroughly mixed by pipetting the solution up and down several times. Reaction mixtures were spun down in the microcentrifuge for 1 minute and immediately incubated at 37°C for 1 hour to digest the parental (i.e., the nonmutated) supercoiled dsDNA.

2.2.4.4 Transformation of XL1-Blue Supercompetent Cells

XL1-Blue supercompetent cells (from Stratagene) were gently thawed on ice. For each sample reaction to be transformed and (-) control, 25 µl of the supercompetent cells was aliquoted to a prechilled Falcon® 2059 polypropylene tube.1.42 M β-mercaptoethanol was added to each tube and kept on ice for 10 minutes by being swirled every 2 minutes. 1 µl of the *Dpn* I-treated DNA (20 ng/µl) from the sample reaction was transferred to separate aliquots of the supercompetent cells, except (-) control. The transformation reactions were swirled gently to mix and were incubated on ice for 30 minutes. They were heat-shocked for 45 seconds at 42°C and then placed on ice for 2 minutes. This heat pulse was optimized for transformation in Falcon 2059 polypropylene tubes. 250µl of SOC medium (Appendix B) preheated to 42°C was added to the transformation reactions and were incubated at 37°C for 1 hour with shaking at 225–250 rpm. Each transformation reaction was aliquoted onto two agar plates containing the appropriate antibiotic for the plasmid vector (Appendix B).The transformation plates were incubated at 37°C for >16 hours.

2.2.4.5 Determination of the Plasmids Carrying Desired Mutations

After transformation of mutated plasmids obtained by *QuikChange® Site-Directed Mutagenesis Kit* to the XL1-Blue cells was completed, four colonies from each mutation type were chosen and sequenced by LI-COR Dye Primer Labelled Sequencing technique at Automated DNA Sequencing Service, Biochemistry and Microbiology, Leeds University, UK.

2.2.5 DNA Isolation and Analysis

2.2.5.1 Plasmid DNA Isolation with QIAprep® Spin Miniprep Kit

In order to harvest the bacterial cells carrying plasmid of interest, 5ml overnight culture was centrifuged at 6500 rpm at 4°C for 5 minutes. Pelleted bacterial cells were resuspended in 250 µl buffer P1 (RNase was added to a final concentration of 100 µg/ml). 250 µl P2 buffer was added, inverted gently 4-6 times to mix until it became viscous and slightly clear and the mixture was incubated for a total of exactly 5 minutes. 350 µl N3 buffer was then added, and was mixed gently 4-6 times, avoiding localized precipitation and tubes were then centrifiuged at 13000 rpm for 10 minutes. After supernatants were collected, they were applied onto QIAprep Spin Column, centrifuged for 30-60 seconds, and flow through was discarded. Columns were washed optionally by adding 0.5 ml buffer PB and centrifuged for 30-60 seconds, flow through was discarded. Then column was washed by 0.75 ml PE Buffer (contains ethanol), and centrifuged for 30-60 seconds. After flow through was discarded, an additional 1-minute centrifugation took place to remove residual wash buffer. Unless the flow through was discarded, residual wash buffer was not completely removed resulting in the inhibition of subsequent enzymatic reactions due to the ethanol present in the buffer. Column was placed into a clean 1.5 ml centrifuge tube. To elute DNA 50 µl EB buffer (10mM Tris-Cl, pH=8,5) or water was added, left for 1 minute and then centrifuged for 1 minute at 13000 rpm.

2.2.5.2 Plasmid DNA Isolation with QIAfilter® Plasmid Maxi Kits

A single colony was picked from a freshly streaked selective plate and a starter culture of 2 ml LB medium containing the ampicillin (50µg/ml) was inoculated. Culture was incubated for nearly 8 hours at 37°C with vigorous shaking (~300 rpm). A flask with a volume of at least 4 times the volume of the culture was used for growth. The starter culture was diluted 1/500 to 1/1000 into selective LB medium. Cells were grown at 37°C at 300 rpm for 12-16 hours then harvested by centrifugation at 6000 rpm for 15 minutes at 4°C. All the traces of supernatant was removed by inverting the tube. Bacterial pellet was resuspended completely in RNase added 10 ml Buffer P1 (to a final concentration of 100 µg/ml).Next, 10 ml Buffer P2 was added and mixed thoroughly by inverting 4-6 times and incubated at room temperature for precisely 5 minutes. During incubation QIAfilter Cartridge was prepared.10 ml chilled Buffer P3 was added, mixed immediately but gently by inverting 4-6 times. The lysate was transferred into the QIAfilter Cartridge immediately in order to prevent later disruption of the precipitate layer. After pouring the lysate into the barrel of the cartridge it was incubated at room temperature for 10 minutes. Precipitate was floated and formed a layer on the top of the solution At the end of the 10 minutes QIAGEN-tip 500 was equilibrated by applying 10 ml Buffer QBT(gravity flow). The cap was removed from the QIAfilter Cartridge outlet nozzle and the plunger was gently inserted into the QIAfilter Cartridge and the cell lysate was filtered into the previously equilibrated QIAGEN-tip. The cleared lysate was allowed to enter the resin by gravity flow. QIAGEN-tip was washed with 2 x 30 ml Buffer QC and DNA was eluted with 15 ml Buffer QF. Glass tubes were recommended to collect the eluate, they were soaked into 0.1 M HCl, then washed with dH_2O and 70% ethanol before using. DNA was precipitated by adding 10.5 ml (0.7 volumes) room temperature isopropanol. The eluate was mixed and centrifuged immediately at 11000 rpm for 30 minutes at 4°C, the supernatant was discarded carefully. DNA was washed with 5 ml 70% ethanol (freshly prepared) and following centrifugation at 11000 rpm for 10 minutes, the supernatant was decanted. The pellet was air-dried for 20 minutes and redissolved in a suitable volume of TE buffer pH=8 or 10 mM Tris-Cl pH=8.5.

2.2.5.3 Plasmid DNA Isolation with Alkaline Extraction Procedure

5 ml overnight culture of bacterial cells carrying plasmid of interest was centrifuged at 6500 rpm at 4°C for 5 minutes. Cells were resuspended in 200 µl Solution I (Appendix B) and incubated at room temperature for 15 minutes. 200 µl Solution II was added and mixture was gently inverted 7-8 times (Appendix B) and incubated on ice for precisely 5 minutes. After addition of Solution III, it was gently mixed 7-8 times and incubated on ice for 15 minutes. Mixture was centrifuged at 13000 rpm at 4°C for 10 minutes. Supernatant was transferred into a new 1.5 ml microcentrifuge tube. 2 volume of cold ethanol (96%) was added and kept at -80°C for nearly 1 hour. Then it was centrifuged at 13000 rpm at 4°C for 10 minutes to discard the supernatant. Pellet was resuspended in 200 µl NE buffer and incubated on ice for 1 hour. 5 µl plasmid was loaded onto agarose gel (optional) in order to detect plasmid DNA at this step. Suspension was centrifuged at 13000 rpm at 4°C for 15 minutes. Supernatant was removed and put into a new 1.5 ml microcentrifuge tube. 2 volumes of cold ethanol (96%) was added and kept at -20°C for nearly 30 minutes. Mixture was centrifugated at 13000 rpm at 4°C for 10 minutes. Supernatant was discarded, pellet was air dried for 5-10 minutes and was resuspended in 20-30 µl ddH_2O. RNase treatment was done either after plasmid isolation (1/2-1 hour at 37 °C) or during the incubation with Solution I (Appendix B).

2.2.5.4 Agarose Gel Electrophoresis and Visualization of DNA

For the investigation of DNA samples, both after plasmid isolation and restriction enzyme digestion, appropriate amount of agarose (1 and 1,2 %, respectively) was dissolved in 1 x TAE buffer (Appendix B) by boling the mixture in a microwave oven for nearly 3 minutes. After cooling the gel to 50-60 °C ethidium bromide solution (Appendix B) was added to a final concentration of 0,5 µg/ml in order to stain the DNA and make it visible under UV light. The gel was then poured into a mould with a comb of proper size, and was affixed and allowed to solidify for approximately 15 minutes. As it became completely polymerized, it was placed into an electrophoresis tank filled with 1 x TAE buffer and combs were removed. The DNA samples were loaded into then wells of the agarose gel by mixing with

6 x loading dye (commercial). DNA size markers were also loaded. Separation of the DNA fragments by electrophoresis was carried out at 100 V for nearly 40 minutes. Then, the gel was visualized on UV transilluminator at 320 nm and photographed by Nikon digital camera.

Agarose concentration (%, w/v)	**DNA fragment range (kb)**
0.3%	5-60
0.4%	1-30
0.7%	0.8-12
1.0%	0.5-10
1.2%	0.4-7
1.5%	0.2-3
2.0%	0.1-2

DNA concentration by agarose gel electrophoresis was calculated according to the following formula:

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)

2.2.5.5 Restriction Enzyme Digestion

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

Sample DNA : 0.5 µl plasmid DNA (85 ng/µl)
Buffer : 2 µl 10 x Multi-core buffer
Restr.Enzyme : 0.5 µl (10U/µl)
ddH_2O : 17 µl
Total volume : 20 µl

In *Nco*I single enzyme digestion, 500-600 ng of the relevant purified plasmid DNA was added into reaction mixture.

Sample DNA : 13 µl plasmid DNA (41 ng/µl)
Buffer : 1.5µl (10 x Buffer Tango, MBI Fermentas)
Restr.Enzyme : 0.5 µl (10U/µl)
Total volume : 15 µl

In *Pst*I and *Eco*RI double digestion, nearly 600-900 ng of pure plasmid DNA was used. 10 x Buffer O^+ was added in order to achieve one time and one-fold concentration finally.

Sample DNA : 10 µl plasmid DNA (85 ng/µl)
Buffer : 1.5µl (10 x Buffer O^+, MBI Fermentas)
Restr.Enzyme : 1 µl (0.5 µl from each) (10U/µl)
Total volume : 12.5 µl

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

2.2.6 Heterologous Expression in *E .coli*

2.2.6.1 Transformation of BL21 Star (DE3) Cells

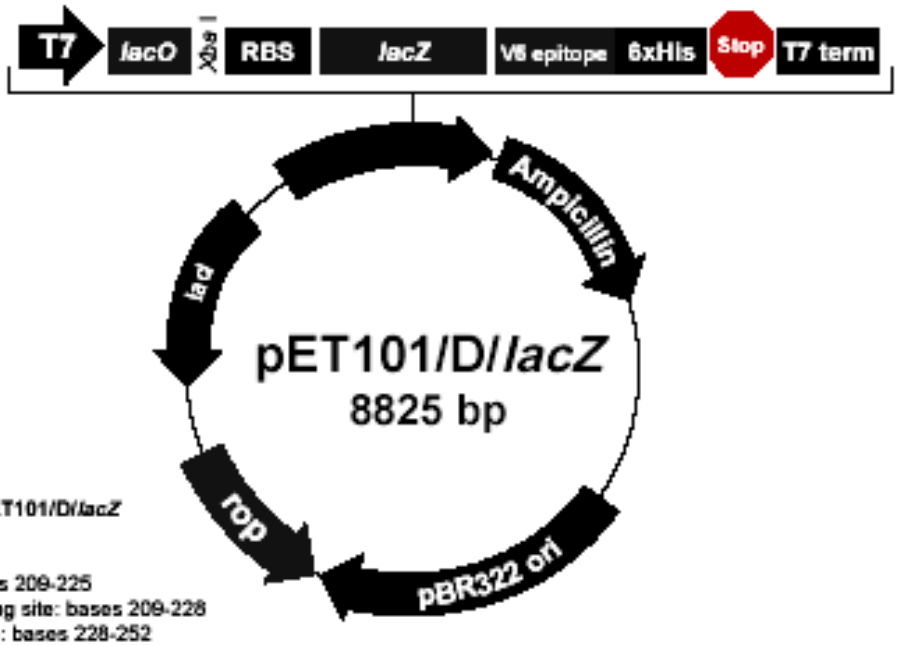
BL21 Star (DE3) competent cells were gently thawed on ice. 50 µl of the supercompetent cells were aliquoted to a prechilled Falcon® 2059 polypropylene tube. 1 µl of the plasmid DNA (20 ng/µl) was transferred to separate aliquots of the supercompetent cells, except the (-) control. The transformation reactions were mixed by tapping, not by pipetting and incubated on ice for 30 minutes. Then they were incubated for exactly 30 seconds at 42°C and quickly placed on ice and kept on ice for 2 minutes. 450-500 µl of SOC medium preheated to 42°C was added to the transformation reactions and they were incubated at 37°C for 1 hour with shaking at 225–250 rpm. Each transformation reaction was aliquoted onto two agar plates containing the appropriate antibiotic for the plasmid vector (Appendix B). The transformation plates were incubated at 37°C for >16 hours.

2.2.6.2 pET System

The pET System is one of the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli*. It is based on the T7 promoter-driven system.

Target genes are cloned in pET plasmids (Figure 2.1) under the control of the strong bacteriophage T7 promoter, transcription and translation signals; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that almost all of the cell's resources are converted to target gene expression. The desired product can comprise more than 50% of the total cell protein a few hours after induction.

After plasmids are established in a non-expression host, they are most often transformed into a host bearing the T7 RNA polymerase gene (λDE3 lysogen) under the control of the *lacUV5* promoter, and expression is induced by the addition of IPTG (Isopropyl-β-thiogalactopyronoside) for expression of target proteins. Figure 2.2 illustrates in schematic form the host and vector elements available for control of T7 RNA polymerase levels and the subsequent transcription of a target gene in a pET vector.



Comments for pET101/D/*lacZ*
8826 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
lacZ ORF: bases 306-3362
V5 epitope: bases 3405-3446
Polyhistidine (6xHis) region: bases 3456-3473
T7 reverse priming site: bases 3527-3546
T7 transcription termination region: bases 3488-3616
bla promoter: bases 3917-4015
Ampicillin (*bla*) resistance gene (ORF): bases 4016-4876
pBR322 origin: bases 5021-5694
ROP ORF: bases 6062-6253 (complementary strand)
lacI ORF: bases 7565-8656 (complementary strand)

Figure 2.1 The map of pET101/D/*lacZ* vector

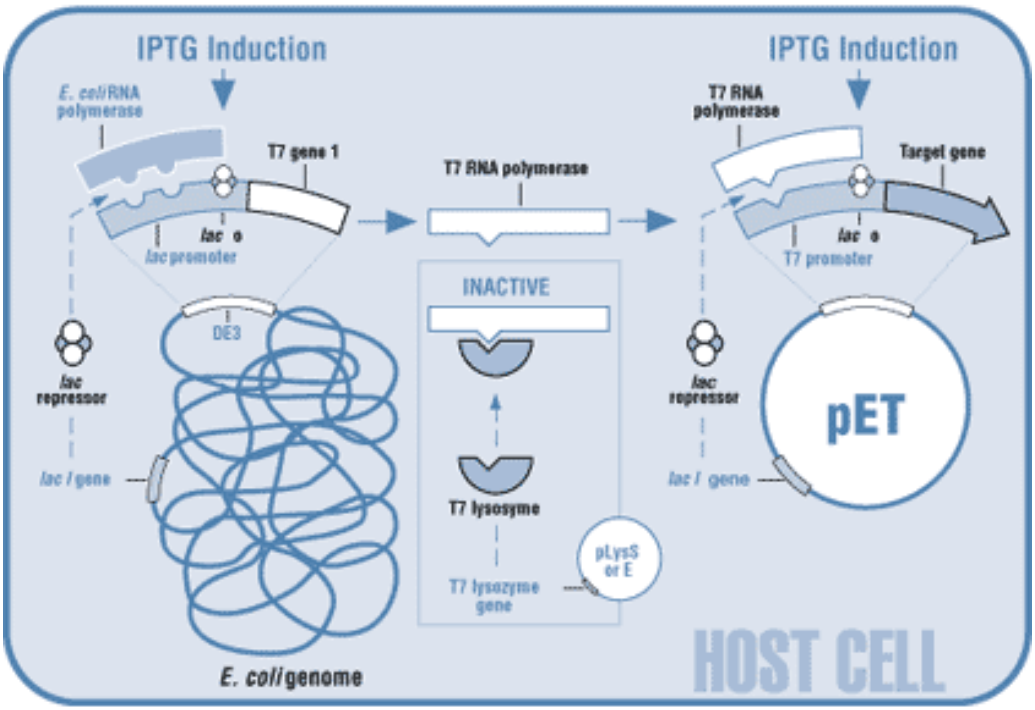


Figure 2.2 Control elements of the pET system

2.2.6.3 Heterologous Expression in BL21 Star (DE3)

Day 1

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

Day 2

A single colony was picked and inoculated into 50 µg/ml ampicillin containing 5 ml LB, in a 50 ml falcon tube. Cells were grown at 37˚C at 200 rpm. At the end of 6 hours 25 µl of the culture was transferred into fresh 25 ml LB in a 100 ml flask and was grown overnight at 37˚C at 200 rpm.

Day 3

With 5 ml of the overnight culture 250 ml LB containing the 50 μg/ml ampicillin in a 1L flask was inoculated and was incubated at 37˚C at 200 rpm to an OD_{595} of 0.4 for about 2 hours. When the OD_{595} value was captured, culture was transferred to 25˚C and monitoring was continued until OD_{595} reaches to 0.6-0.7. 1 ml pre-induction sample was taken in order to compare pre-induction and post-induction conditions. Then the culture was induced with 1M IPTG to a final concentration of 1 mM and was incubated at 25˚C at 200 rpm overnight. 1 ml sample at 1, 2, 4 and 6 hour time-points can also be taken and kept at -20˚C for analysis later if something goes wrong.

Day 4

After 16-21 hours, the culture was transferred into 250 ml Sorvall tubes and the cells were harvested by centrifugation in a Sorvall GSA rotor at 6000 rpm at 4˚C for 10 minutes. The pellets were resuspended in 5 ml 1x PBS (phosphate-buffered saline) (Appendix B) containing one complete protease inhibitor tablet (Complete Mini, ROCHE). After being pipetted and aliquoted into 50 ml Falcon tubes the cells were sonicated using bursts of 10 seconds on 40 seconds off for 7 cycles by placing the Falcon tubes in an ice-filled beaker to prevent heating. Sonicated cells were centrifuged in a Sorvall SH MT rotor at 11000 rpm at 4˚C for 5 minutes. Supernatant was collected and transferred into clean 15 ml falcon tube. Meanwhile, dialysis tubing (10k MWCO dialysis tubing, Sigma) was prepared by boiling in dH_2O twice for 3 minutes and waited for cooling down. The boiled dialysis tubing was kept in 20% ethanol at 4˚C. Supernatant was dialysed against 2 x 1L 1 x PBS buffer at 4˚C, after being dialysed once for 3 hours the buffer was refreshed and left overnight.

2.2.7 Enzyme Assays and Protein Analysis

2.2.7.1 Qualitative GOase Activity Assay

In order to observe the activity of the mutant GOases crude extracts were assayed. 100 mM $CuSO_4$ was added into the supernatants to a working concentration of

50 μM–also the pellets can be resuspended in 1 x PBS containing 50 μM $CuSO_4$ for further protein analysis-.

GOase assay solution including the substrate ABTS was prepared (Appendix B) to detect galactose oxidase activity. 30-40 μl of supernatant assumed to be containing the desired enzyme was added into 200 μl assay solution and colour change was monitored–if the enzyme is active, conversion of colourless transparent solution into green should be detected-.

This assay was performed with pure samples as well. They were all dialysed against 20mM PIPES+1mM $Cu(NO_3)_2$ pH=6,1 at 4˚C overnight to ensure full processing of the enzyme.

For a quantitative assay, absorbance can be measured at 414 nm.

2.2.7.2 *Strep*-tag Based Purification of the Mutant GOases

Purification is the next step after obtaining the cell extracts of the mutant samples that have been induced with 1mM IPTG as to express the mutant *gao*. In order to gain the purified GOase, *Strep*-Tactin® Sepharose® column was used which works on the principle of affinity basis. It was imported from Germany (IBA GmbH's, Göttingen) with financial support from TÜBITAK Scientific Research Group (TBAG).

One-step purification by creating a resin that is completely specific to the target protein is the ideal solution in order to get rid of the requirement of considerable amount of optimization and sample of contaminants.

In affinity chromatography a ligand that specifically interacts with the target protein is immobilized on a chromatography matrix; the target protein binds to the column, and unwanted proteins are eluted. In some cases, the affinity ligand is an antibody against the protein of interest; in others, the target protein is expressed from a plasmid that encodes for an "affinity tag" specific to a particular ligand.
Strep-tag II™ purification system (IBA GmbH's, Göttingen) also relies on the streptavidin-biotin interaction. N- or C-terminal fusion proteins of this tag can then

be purified by immobilization on a unique StrepTactin affinity column have been developed for this purpose. *Strep*-Tactin is a specially designed streptavidin derivative with a high binding affinity for *Strep*-tag II. The eight-residue *Strep*-tag II (WSHPQFEK) represents an improvement over its predecessor, the *Strep*-tag, which is nine amino acids long and is restricted to C-terminal fusions. Competitive elution of the fusion protein is achieved by adding small amounts of desthiobiotin, a biotin analog, to the washing buffer.

All operations were performed at a temperature amenable to the stability of the recombinant GOase; at 4 °C. To achieve optimal purification results, it was complied with the specified volumes and their ratios (column bed, washing volumes etc.). Throughout this thesis, purification experiments were performed with a column filled with 5 ml *Strep*-Tactin Sepharose® 4FF, 4% agarose. At low expression levels, applied cell extract volumes were increased to take advantage of the column capacity, without changing other volumes.

First the top then the bottom cap was removed from the column and the excess storage buffer was allowed to drain off. By adding 2 CV (CV = column volume) of Buffer W (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA); the column was equilibrated. Throughout the protocol, however, for recombinant GOase purification through *Strep*-Tactin column, buffers were prepared without EDTA, since it is a metalloprotein.

Then the cell extract having a volume between 0.5 and 10 CV was added to the column (concentrated cell extracts are preferable; if quantification is possible, cell extract containing between 50 and 100 nmol recombinant *Strep*-tag II fusion protein per 1 ml column bed volume can be applied). Frozen cell extracts were centrifuged (Eppendorf mini spinplus microcentrifuge) at 14,000 rpm at 4°C for 5 minutes before applying them to the column in order to remove any aggregates that might have formed. After the cell extract had completely entered the column, the column was washed 5 times with 1 CV of Buffer W. The eluate was collected in a 50 ml Falcon tube as to apply 20 µl of it to an analytical SDS gel.

In order to elute the recombinant GOase, 6 times 0.5 CV Buffer E (Appendix B) containing a reversibly binding specific competitor was added and the eluate was collected in 6 fractions. 20 µl of each fraction could be used for SDS-PAGE analysis. The purified *Strep*-tag II fusion protein was usually eluted in the 3rd to 5th fractions. Collected fractions were stored at -20 °C.

For regeneration, the column was washed 3 times with 5 CV Buffer R (Appendix B). The colour change from yellow to red indicates the regeneration process and the intensity of the red colour is an indicator of the column activity status. Buffer R was removed by adding 2 times 4 CV of Buffer W before the next purification run. The column was stored at 4°C overlaid with 2 ml of Buffer W or Buffer R.

If generation of authentic recombinant protein was desired desthiobiotin and EDTA (if it was used) can be removed by gel chromatography or dialysis after purification of the recombinant protein (with N-terminal *Strep*-tag II and subsequent factor Xa cleavage site). Biotinylated factor Xa can be applied according to the manufacturers instructions (Roche Diagnostics GmbH, Mannheim). After digestion, *Strep*-tag II, biotinylated factor Xa, and uncleaved recombinant protein can be separated from the authentic recombinant protein by another *Strep*-Tactin chromatography whereby the authentic recombinant protein is collected from flow through fractions. Because of its small size, *Strep*-tag generally does not interfere with the folding or bioactivity of the fusion partner. Thus, removal of the tag becomes superfluous. Therefore, this last application was not performed in this study.

Short protocol of the *Strep*-Tactin chromatography cycle

2.5 - 50 ml cell extract was loaded to the column.

1. After the protein extract was loaded onto the *Strep*-Tactin matrix, column was washed 5 times with 1 CV (column volume) Buffer W (5 x 5 ml).
2. Recombinant protein was eluted by the addition of 6 times 0.5 CV Buffer E (6 x 2.5 ml).
3. The column was regenerated by the addition of 3 times 5 CV Buffer R (3 x 25 ml).

4. The column was equilibrated by the addition of 2 times 4 CV Buffer W prior to the next purification run (2 x 20 ml).

5. The column was stored at 4° C overlaid with 2 ml Buffer W or R.

Table 2.3 Recommended volumes for working with *Strep*-Tactin columns

Column Volume	Protein Extract Volume	Washing Buffer Volume	Elution Buffer Volume
0.2 ml	0.1-0.2 ml	5 x 0.2 ml	6 x 0.1 ml
1 ml	0.5-10 ml	5 x 1 ml	6 x 0.5 ml
5 ml	2.5-50 ml	5 x 5 ml	6 x 2.5 ml
10 ml	5-100 ml	5 x 10 ml	6 x 5 ml

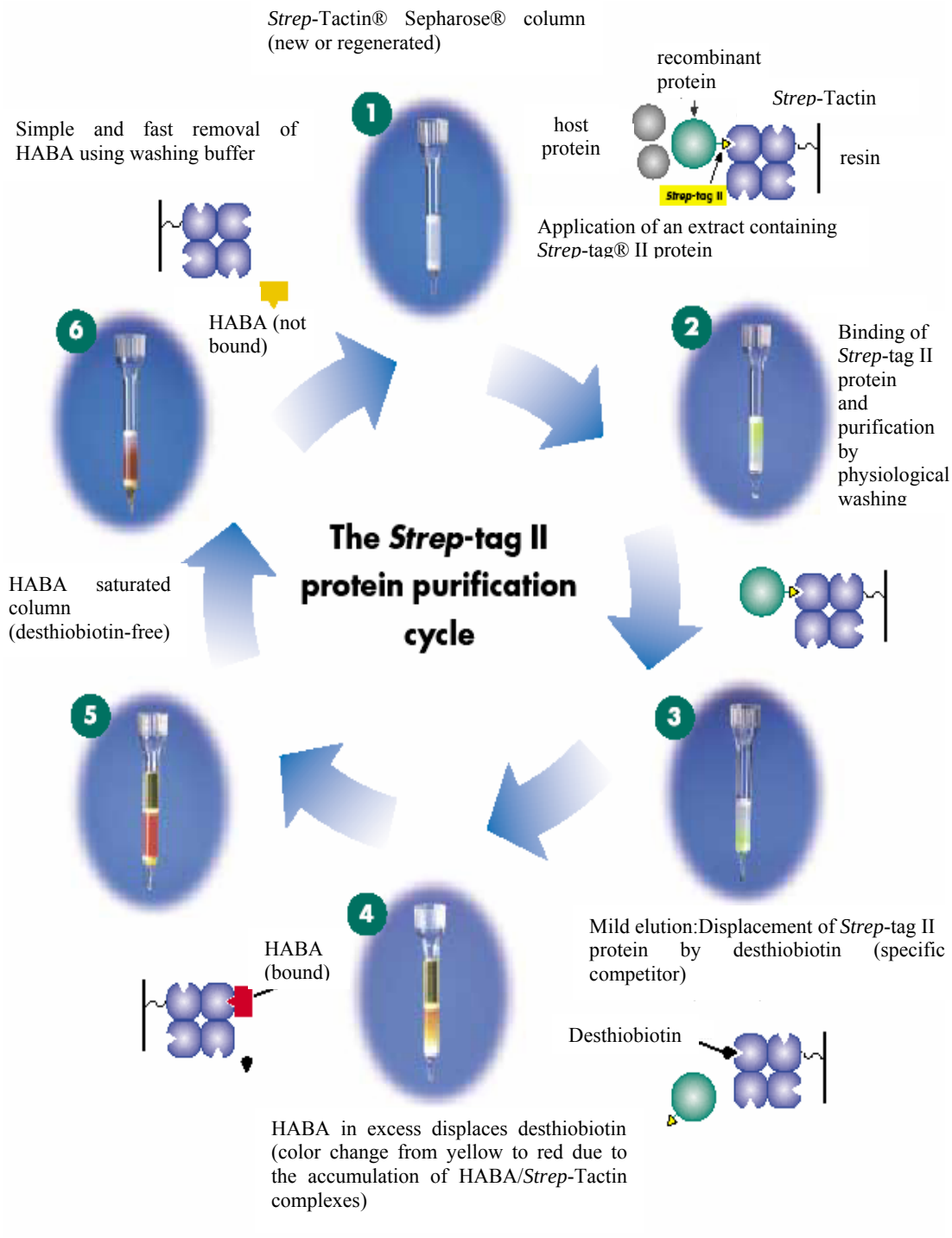


Figure 2.3 *Strep*-tag® II protein purification cycle

2.2.7.3 Laemlli SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was performed via Blue Vertical 102 and Blue Power 500 Electrophoresis System, (SERVA Electrophoresis GmbH). 2 combs and 4 spacers, 2 plain glass plates and 2 notched (the small one) glass plates were rinsed with dH_2O and also wiped with ethanol. They were dried and it was ensured that they were tissue-free. After placing spacers flush with the edges of notched plate, large gel was overlaid to create "sandwich". Sandwich was slid into the electrophoresis unit, with the small plate innermost. Glass plates were allowed to sit flush with the bottom of the running unit. By inserting the wedges and pressing them down the sandwich was affixed to the inner core running unit. When running only one gel, a dummy plate was required on the other side of the unit to retain the top of the buffer level. The inner core was placed onto the silicon pads of the gel casting module. Cams were loosely tightened to assemble the inner core running unit onto the silicon pads. Combs were checked forming tight fit in glass plates. 1 cm beneath well base was marked to indicate height of separating gel before removing combs.

A suitable percentage separating gel was prepared by combining the following components in a plastic universal or flask. TEMED and APS (freshly made) (Appendix B) was added just before the gel was ready to pour in order to avoid polymerization. Using plastic Pasteur pipette, before pouring the gel mixture ddH_2O was poured between gel plates up to the top to observe any leakage. Water was drained off and gel mixture was poured up to the level indicated. It was overlaid with ddH_2O and left to set for about 1 hour. Stacking gel was prepared by combining the following components in a plastic bijoux or flask (Preparation of both separating gel buffer and stacking gel buffer were given in Appendix B). Water overlay was poured off and the area above gel was dried by using 3MM filter paper. Stacking gel was poured above separating gel after addition of proper amounts of TEMED and APS. Combs were inserted by avoiding bubble formation and allowed to set for approximately 30 minutes.

Table 2.4 Required amounts of components of separating gel

Separating gel	Volume (ml)							
	12.5%			15%			10%	8%
	(x1)	(x2)	(x20)	(x1)	(x2)	(x20)	(x1)	(x1)
GB	1.2	2.4	24	1.2	2.4	24	1.2	1.2
30% acrylamide stock	2.1	4.2	42	2.6	5.2	52	1.7	1.3
H_2O	1.7	3.4	34	1.2	2.4	24	2.1	2.5
APS (25%)	0.03	0.06		0.03	0.06		0.03	0.03
TEMED	0.003	0.006		0.003	0.006		0.003	0.003
(Total)	(5)	(10)	(100)	(5)	(10)	(100)	(5)	(5)

Table 2.5 Required amounts of components of stacking gel

Stacking gel	Volume (ml)		
7.5%	(x1)	(x2)	(x20)
SGB	0.6	1.2	12
30% acrylamide stock	0.4	0.8	8
H_2O	1.4	2.8	28
APS (25%)	0.02	0.04	
TEMED	0.002	0.004	
(Total)	(2.4)	(4.8)	(48)

When the gels were polymerised, the gel casting adapter was removed by loosening the cams. Afterwards, the running unit was placed into the tank and the appropriate volume of 1 x running buffer (from 10x stock) (Appendix B) was added to the upper (700 ml) and lower chambers (200 ml). Lower buffer chamber was filled so almost to

top of gel sandwich to let the current throughout the gel. Combs were removed and wells were washed out with running buffer. Samples were prepared (for 2 gels, 20 maximum) including molecular weight markers. For denatured proteins, sample denaturing mix was made by adding 1:4 β-mercaptoethanol to sample buffer (Appendix B). 1/3 volume sample denaturing mix (x4 concentrated) was added to each sample (7 µl SDM, 21 µl sample), so SDM was at a final concentration of x1. It was tried to make all samples to same volume.

For 0.75 cm spacers and 10-well comb, maximum volume was ~35 µl. Samples were boiled for 5 min and allowed to cool on bench. Samples were centrifuged at 12500 rpm at room temperature for 5 minutes in Eppendorf mini spinplus microcentrifuge. By using a fine-pointed gilson tip sample was delivered as close as possible to base of well thus displacing running buffer bewaring of bubbles as this might have caused sample overflow. If there were less samples than 8 loaded/gel, they were filled with 1:1 H_2O:sample denaturing mix in order to prevent "smiling" of outer samples (equivalent volume of sample maintains the uniform electrical resistance across the gel). After placing the lid the leads were attached to the Blue Power 500 power supply (in each case red-red, black-black).Constant voltage of 150 V was applied and current at start was ~60 mA per gel. When 2 gels were required, the current chosen was twice the current of one but the same voltage. The gel was run for about normally 2 hours. Once blue dye was migrated at base of gel, power supply was disconnected, lid was removed and inner core running unit was pulled from lower buffer chamber. Buffer was poured off and wedges were removed. To release the gel from glass plates one spacer was pushed 1/2 way out of sandwich and gently twisted, instead a knife could be used for same purpose. Any corner of gel can be cut to record orientation; loading turn of the samples in accordance with the marker or ladder might have also help orientation after staining. Preceding staining with Coomassie Blue R 250, the gel was fixed with 20% trichloroacetic acid (Appendix B) for ½ hour at room temperature. Then it was rinsed with dH_2O 2-3 times for 3 minutes. The gel was transferred to 50-100 ml staining solution (Appendix B) and stained for overnight. After replacement of staining solution with 100 ml destaining solution (Appendix B) the gel was destained for 45 minutes refreshing the solution

every 15 minutes. The photograph of the gel was taken and the gel was kept in preserving solution (Appendix B). If dry gel is required, 2 sheets damp 3MM paper can be placed on the gel and it can be covered with cling film. After trimming, the gel can be dried at 80^{o}C for 45 minutes-1 hour, by using vacuum gel drier.

Size separation	**%age gel**
70-200 kDa	5.0%
40-150 kDa	7.5%
20-100 kDa	10.0%
10-70 kDa	12.5%
8-50 kDa	15.0%

Optimal protein loading	
Individual polypeptide	0.5-5 µg
Complex mixture	25-50 µg

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

To prepare a standard the amounts given in the table below were used.

Table 2.6 The amounts required for standard 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. Then BSA protein standard was dispensed to each tube at proper amounts. After addition of the Bradford's solution (it must be at room temperature before addition) the tubes were vortexed and mixed well. These tubes were kept in dark for 10 minutes and then the optical density was monitored at 595 nm within 1 hour. According to the results a standard curve was attained. The optical density of the sample protein was measured for different amounts until the

value read from the spectrophotometer fell into the range between the values obtained from standard curve. Then with the formulation given below the unknown protein concentration was calculated accordingly.

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

Protein	OD at 595 nm
0	0
0,01	0,043
0,02	0,079
0,03	0,108
0,04	0,145
0,05	0,171

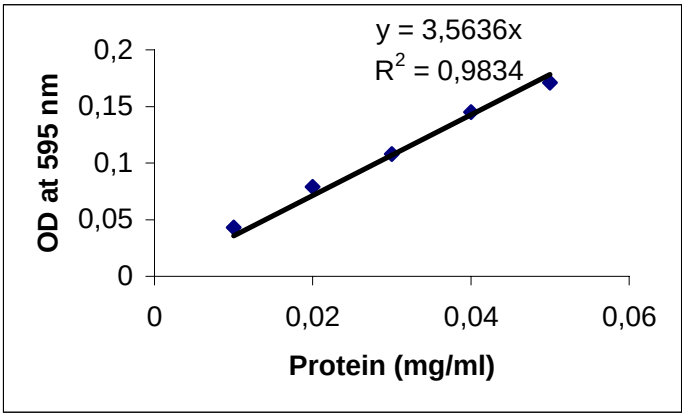


Figure 2.4 Standard curve for calculating ProGOMN1 concentration.

150 µl sample was taken and the absorbance value obtained was just within the range that change between 0.043-0.171 at OD_{595}.

Protein (mg/ml) : $\frac{0.089}{3.5636} \times \frac{500}{150} = 0{,}083\,\text{mg/ml} = 83\mu\text{g/ml}$

Similarly, the fractions 3 and 6 (Figure 3.31) were collected in the same tube. The concentration was calculated (23 µg/ml) after the absorbance was measured at OD_{595}. This dilute sample was used in electroblotting.

2.2.7.5 Protein Electroblotting for N-terminal Sequencing

As to get clear results of the N-terminal sequencing of the protein samples, electrophoretic transfer onto the PVDF membrane was required since this is the best way preventing proteins from degradation during transportation. PVDF membrane

has high mechanical strength, chemical stability and enhanced binding capacity in the presence of SDS.

In this experiment Mini Trans-Blot Cell Assembly, Bio-Rad was used. SDS-PAGE gel was kept at 4°C to provide exact polymerization. Gel loading should be estimated to provide sufficient protein in only a few tracks. Generally, a few 10s to a few 100s of picomoles of protein is appropriate. SDS-PAGE was run with 4 wells filled with 140 µl sample (105 µl sample in total and 35 µl sample denaturing mix) with a concentration of 23 µg/ml sample; providing approximately 3,5 fold of the amount desired for the N-terminal sequencing. After SDS-PAGE was run, this fresh gel was used in electroblotting. First of all, sequencer grade (low porosity) PVDF-type membrane was thoroughly pre-wetted with 100% HPLC-grade methanol for nearly 5 minutes, then equilibrated for 15 minutes in the blotting buffer. It was noted that the membrane should have turned translucent in methanol and remained so after equilibrating in blotting buffer (Appendix B); the reappearance of white (*i.e.* dry) membrane must have been avoided**.**

The gel was rinsed in blotting buffer for 15 minutes to wash away excess Tris/glycine/SDS electrophoresis buffer (TGS), which may leave a residue on the blot which interferes with sequencing, and the blotting papers were also prewetted in blotting buffer for 15 minutes.

Buffer Volumes

Recommended buffer volumes were determined as follows:

3 MM Filter Paper	0.2 ml/cm^2 x # of sheets
Membrane	1.0 ml/cm^2
Gel	1.5 ml/cm^2

The sizes of the membrane and the blotting papers were exactly the same as polyacrylamide gel to be blotted so as to minimise the surface area exposed to the electrode: this reduces the amount of current that needs to be passed in order to affect transfer (and reduces heating effects). Buffer type, pH and pI were very important

since this determines the direction of movement, thus accurate orientation of the membrane and the gel is the crucial step in blotting. Because the pI value of galactose oxidase is 8,07 it will be negatively charged in blotting buffer having a pH value of 10,5. This wil make the protein move towards anode. After placing fiber pad onto the gray side of the cassette 2-3 sheets of pre-wetted filter papers were overlaid. Then the equilibrated gel was placed, afterwards the membrane was placed onto the gel. Here, the orientation took place under basic conditions, so during set up membrane was resting on the anode side of the gel. Gel sandwich was completed by laying the equilibrated filter papers flat on to the membrane and the fiber pad lastly. The bubbles which might have formed were removed by rolling a glass tube gently on each layer in order to get good results. Finally the cassette was closed firmly and placed in the module. Frozen Bio-Ice cooling unit was added and following the placement of the module into the tank, the tank was completely filled with buffer.

The gel was electroblotted at 100 V, 350 mA for 1 hour. Following the detachment of the module, the gel was subsequently stained with Coomassie Blue to establish the effectiveness of the transfer. Also the membrane was stained with Coomassie Blue immediately after blotting while it was still wet for 2-3 minutes. It was destained for 15 minutes and kept in ddH_2O for 4 hours. The blot was dried between 2 sheets of filter paper and stored in a sealed bag at -20 ^{O}C ready for N-terminal sequence analysis.

In order to be sequenced, generally a few 10s to a few 100s of picomols of protein was appropriate which makes approximately 680-700 ng of pure GOase. Two fold of the beginning culture was used to obtain bulk amount of protein. Cultivation and purification was done as mentioned before in Section 2.2.6.3 and 2.2.7.2.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Strategy Followed to Explore the Mechanism of Pro-Peptide Self-Processing

Galactose oxidase is a unique enzyme which is maturated in a stepwise manner as explained in section 1.1.4.3.2.

In order to clarify the mechanism of autocatalytic cleavage of GOase pro-peptide, site-directed mutagenesis was carried out to obtain different mutant types. Following confirmation of the desired amino acid changes by DNA sequencing, expression studies took place in *E. coli.* Purification of GOase was performed by using affinity chromatography technique. Finally, these pure enzymes were analysed by SDS-PAGE as to see whether pro-sequence was removed or not to make a statement about the effectiveness of the mutations. The experimental strategy followed is demonstrated in Figure 3.1.

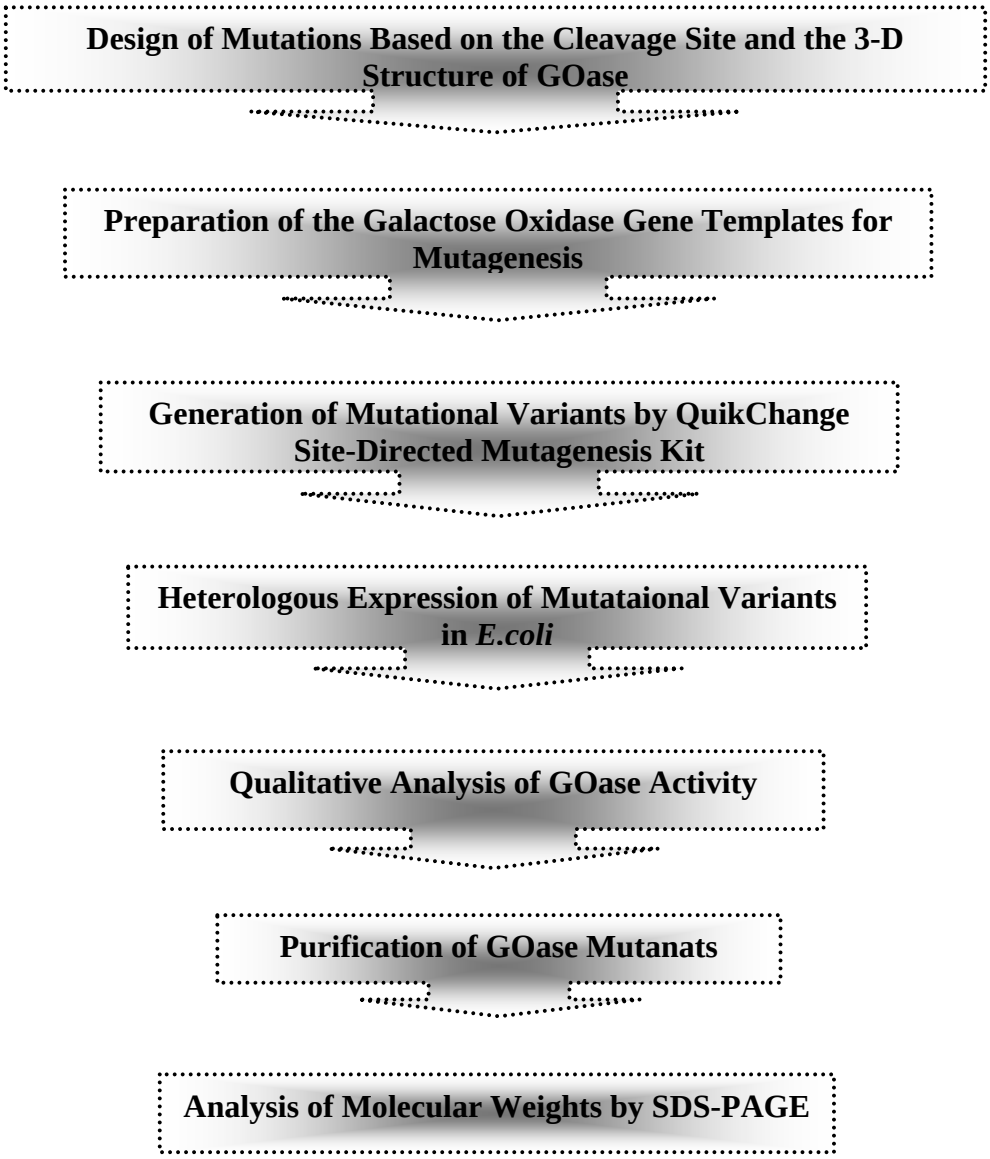


Figure 3.1 Experimental strategy followed in this study

N- terminus of GOase starts with the putative 8 amino acid pre-sequence preceding propeptide after removal of the signal peptide.

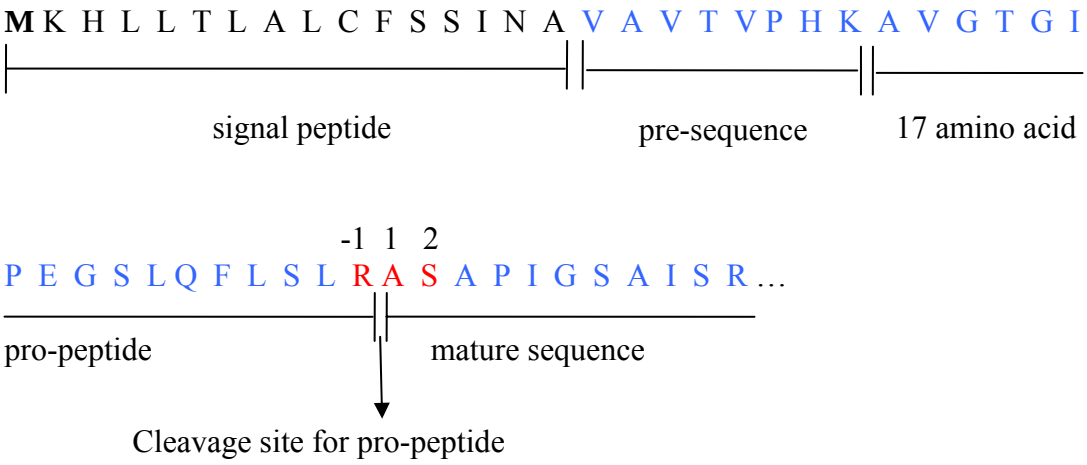


Figure 3.2 Amino acid sequence of the N-terminus of GOase

3.1.1 Constructs Carrying GOase Gene

In this thesis, the manipulation of galactose oxidase gene was carried out by two constructs which were previously developed by directed evolution (Frances H.Arnold , Leeds University, U.K.); ProGON1 and ProGOMN1. ProGON1 contains the wild type GOase carrying only silent mutations at the N-terminus, whereas ProGOMN1 carries 5 non-synonymus and 1 synonymus mutation in the mature protein coding region, in addition to silent mutations at N-terminus corresponding to the pro-peptide region. Both ProGON1 and ProGOMN1 constructs contain a *Strep*-tag at the C-terminus to allow purification by affinity chromatography and a pre-pro-peptide at the N-terminus.

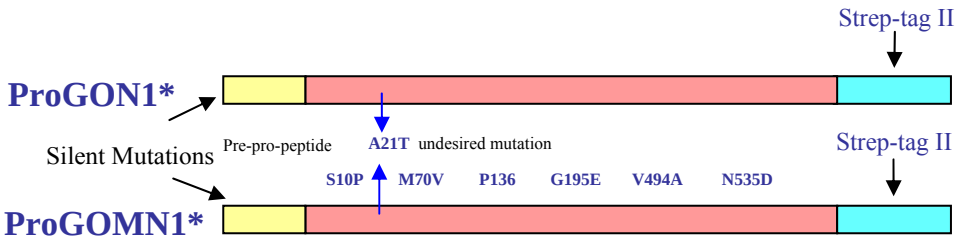


Figure 3.3 Shematic diagram of pro-galactose oxidase genes used in this study

The maps and the sequences of ProGON1 and ProGOMN1 constructs and native *gao* together with the amino acid alignments of these templates are also given in Appendix C.

3.1.2 Preparation of the ProGON1 and ProGOMN1 templates for Site-Directed Mutagenesis

In addition to those mutations, both ProGOMN1* and ProGON1* constructs contained an undesired mutation (Ala21Thr, demonstrated with an asterisk) which had to be reverted back by site-directed mutagenesis (Figure 3.3). It was shown that the one with the T21A reversion among mature constructs has slightly higher expression than the one with the mutation (Deacon, personal communication). Therefore, it was decided to first correct this mutation in the ProGO constructs. Mutations were carried out as described below in section 3.3.

3.2 Design of Mutational Primers

To study the mechanism of autocatalytic cleavage of GOase pro-sequence, four different primers were designed. Three of these primers were planned to constitute mutations at the cleavage site and one was close to the cleavage site in the three dimensional structure of the enzyme.

In this study, the mutations introduced to the appropriate points that are likely to possess a crucial role in processing were;

1. **R-1P/A1P**
2. **R-1X/A1X**
3. **S2A**
4. **H522A**

3.3 Construction of ProGON1 and ProGOMN1 by Reversion of A21T Mutation on the Templates

XL1-Blue Supercompetent cells (Stratagene) were transformed with the plasmids carrying undesired A21T mutation according to the procedure given in section 2.2.4.4. Plasmid isolation was done in order to get starting material for *QuikChange® Site-Directed Mutagenesis Kit* (Stratagene). The band with an expected size of 7805bp was visualised by agarose gel electrophoresis as seen in Figure 3.4**.**

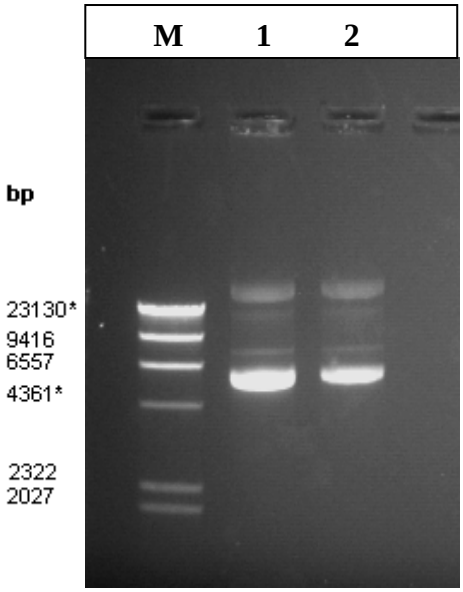


Figure 3.4 Isolation of plasmid carrying the undesired mutation A21T. **M,** λ DNA/*Hin*d III (50μg/ml); **1,** PROGON1*; **2,** ProGOMN1*.

3.3.1 General Strategy of the *QuikChange*® Mutagenesis System

The *QuikChange® Site-Directed Mutagenesis Kit* is used to make point mutations, switch amino acids, and delete or insert single or multiple amino acids.A high fidelity DNA polimerase (KOD Hot Start / *Pfx* DNA Polymerase, two synthetic mutant oligonucleotide primers (each complementary to opposite strands of the vector) and a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest were used. The procedure is simply based on degradation of parental DNA

by *Dpn*I endonuclease which is specific for methylated and hemimethylated DNA after temperature cycling. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue cells and the nick is repaired. The basic procedure is demonstrated in Figure 3.5.

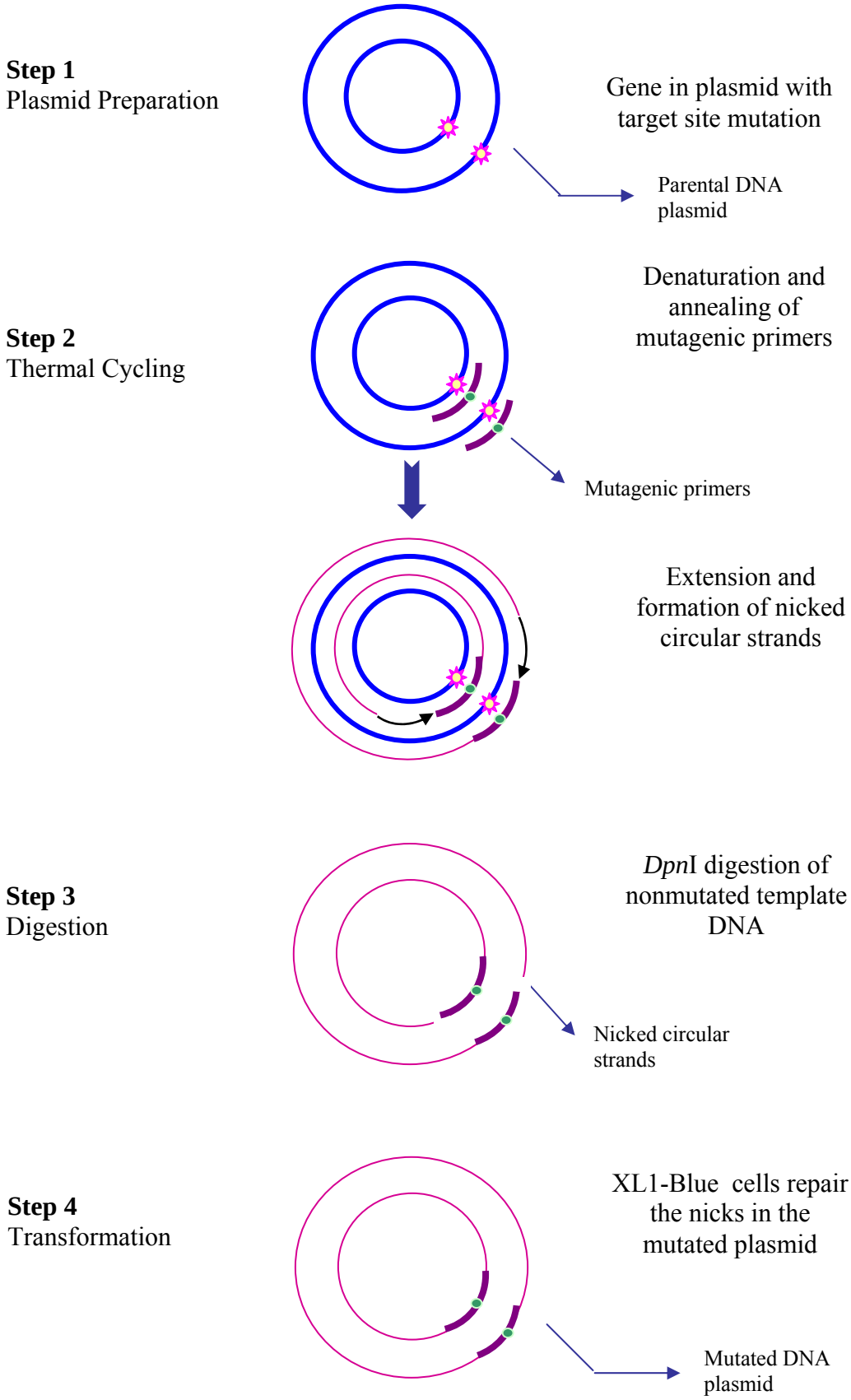


Figure 3.5 Overwiew of the *QuikChange® Site-Directed Mutagenesis Method*

3.3.2 *QuikChange*® Application and Optimization of Annealing Temperature

In order to find the optimum annealing temperature, five different reaction mixtures were prepared to run at temperatures from 53ºC to 61 ºC by using a gradient thermal cycler.ProGON1* and ProGOMN1* were used as templates with primers FGGO_T21A_F and FGGO_T21A_R which were given in Appendix D.

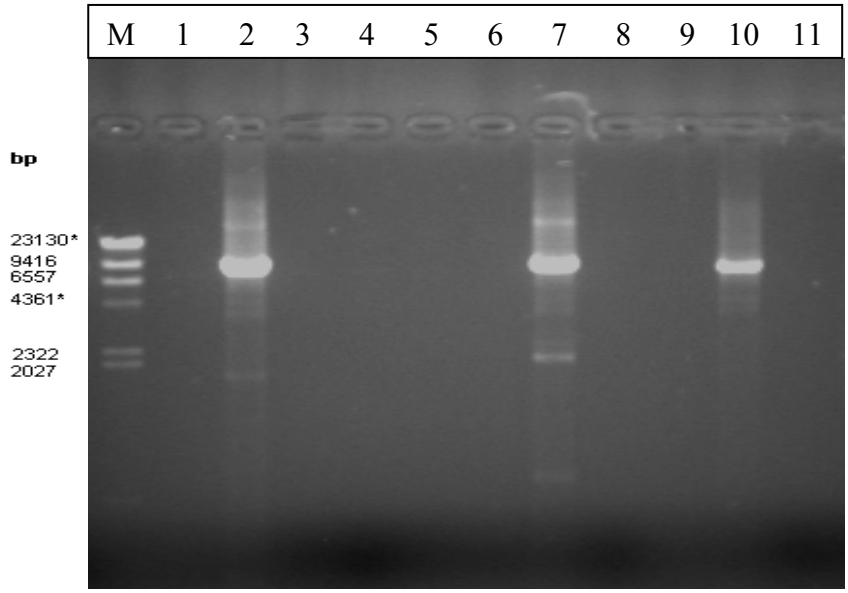


Figure 3.6 Agarose gel electrophoresis showing the results of gradient temperature cycling for establishing the optimum annealing temperature for the *QuikChange*® *Site-Directed Mutagenesis System.* **M,** λ DNA/*Hin*d III (50µg/ml); **1-5,** Putative ProGON1 obtained at temperatures 53 ºC, 55 ºC, 57 ºC, 59 ºC, 61 ºC, respectively; **6-10,** Putative ProGOMN1 obtained at temperatures 53 ºC, 55 ºC, 57 ºC, 59 ºC, 61 ºC, respectively; **11,** Negative control.

As seen in Figure 3.6, optimum annealing temperature for both of the constructs was 55 ºC.

QuikChange was performed in order to revert the undesired mutation (A21T) according to the mutant strand synthesis reaction given in section 2.2.4.2.The templates and the primers were the same which were used in annealing temperature

optimization above. Thermal cycling parameters were: 94 ºC x 30 sec., 24 x (94 ºC x 30 sec., 55 ºC x 1 min., 68 ºC x 8 min) as given in Table 2.2 in details.

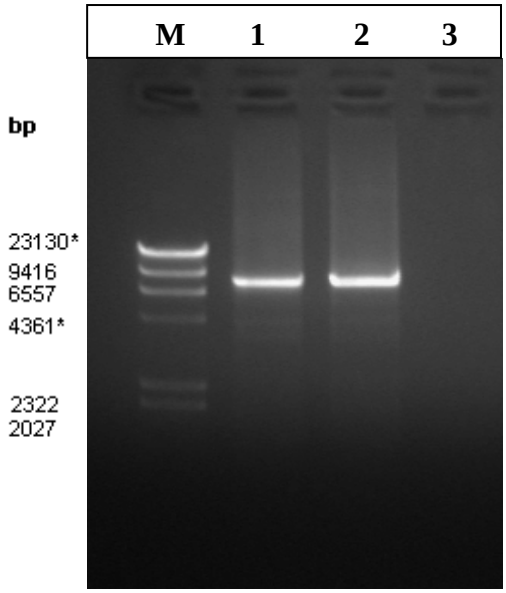


Figure 3.7 *QuikChange® Site-Directed Mutagenesis System* results visualized by agarose gel electrophoresis. **M**, λ DNA/*Hin*d III (50µg/ml); **1,** Putative ProGON1; **2,** Putative ProGOMN1; **3,** (-) control

Two bright bands were obtained with an expected size (7805bp) as shown in Figure 3.7.

3.3.3 Transformation of XL1-Blue Cells byProGOMN1 and ProGON1

The amplification products which were thought to be carrying T21A conversion were transformed into XL1-Blue cells according to the procedure explained in section 2.2.4.4.

3.3.4 Plasmid Isolation and Detecting Reversion of Undesired Mutation by Sequence Analysis

Several colonies were selected for plasmid isolation and further sequence determination to detect T21A reversion, the right sequence.

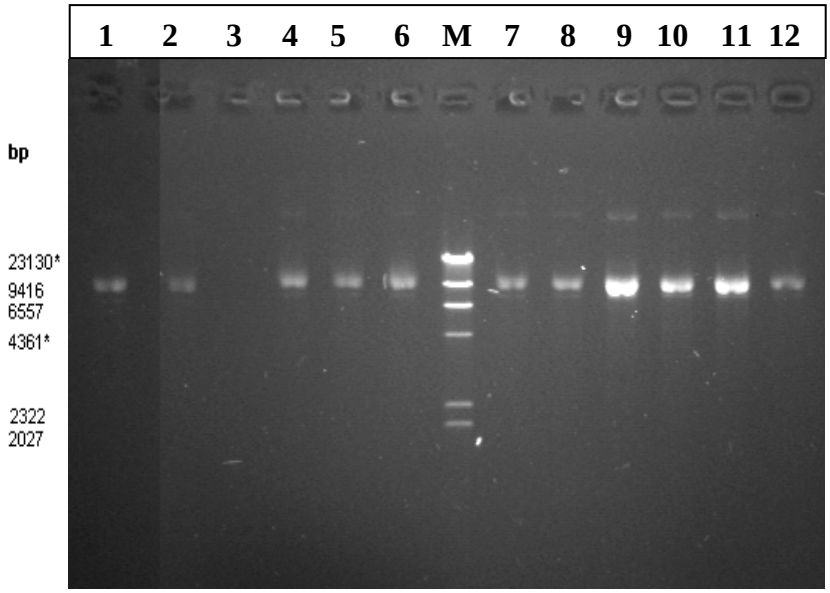


Figure 3.8 Isolation of plasmid from E.coli transformants following *QuikChange*® mutagenesis for T21A reversion. **M**, λ DNA/*Hin*d III (50µg/ml); **1-6,** Putative ProGON1 clones; **7-12,** Putative ProGOMN1 clones.

After estimation of plasmid DNA concentration, 250 ng of each sample was subjected to sequence analysis as described in section 2.2.4.5. According to those data, among the ProGON1 and ProGOMN1 samples $ProGON1_6$ and $ProGOMN1_4$ had the correct conversion (T21A).

To obtain a sufficiently high amount of plasmid for further use in mutagenesis, large scale plasmid isolation was done for ProGON1 and ProGOMN1 with QIAfilter® Plasmid Maxi Kits as described in section 2.2.5.2.

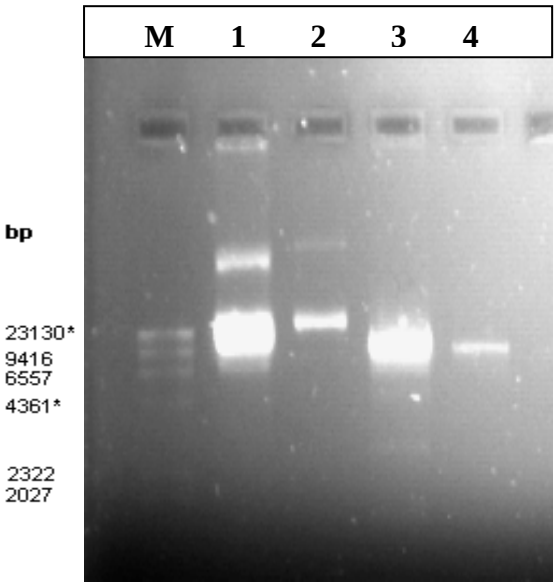


Figure 3.9 Large scale plasmid isolation of ProGON1 and ProGOMN1 and restriction enzyme digestion results with *Cla*I. **M**, λ DNA/*Hin*d III (50µg/ml); **1**, ProGON1; **2,** ProGOMN1; **3,** ProGON1 digested with *Cla*I; **4,** ProGOMN1 digested with *Cla*I.

3.4 Generation of Mutational Variants by Site-Directed Mutagenesis

The mutations were designed in order to prevent the possible roles of important conformational determinants that might participate in self-processing. In a previous study by Ögel (1993), Arg-Ala was substituted into Phe-Phe as explained in section 1.3.1. It was shown that this mutation did neither alter the cleavage process nor the site of cleavage (Ögel, 1993). However, for a thorough investigation further mutations are necessary at or surrounding the cleavage site.

Here, codons were altered as little as possible modifications (see Appendix C and D). Arg-1Pro, Ser2Ala and His522Ala substitutions were carried out supposing the inhibition of the cleavage by changing polarity and putative reactivity of side chains. Similarly, the cleavage site mutation Arg-1X/Ala1X was designed in order to generate a library of cleavage site mutations. Here, four types of this library were analyzed; this can be continued in a separate study.

3.4.1 Results of Thermal Cycling and *Dpn*I Digestion

In order to achieve mutations explained in Section 3.2 *QuikChange® Site-Directed Mutagenesis Kit* was used by making use of suitable primer oligonucleotides as described in section 2.2.4.1 and shown in Table D.1. For all mutations the protocol explained in section 2.2.4.2 was followed. All these mutations were performed on ProGON1 and ProGOMN1.

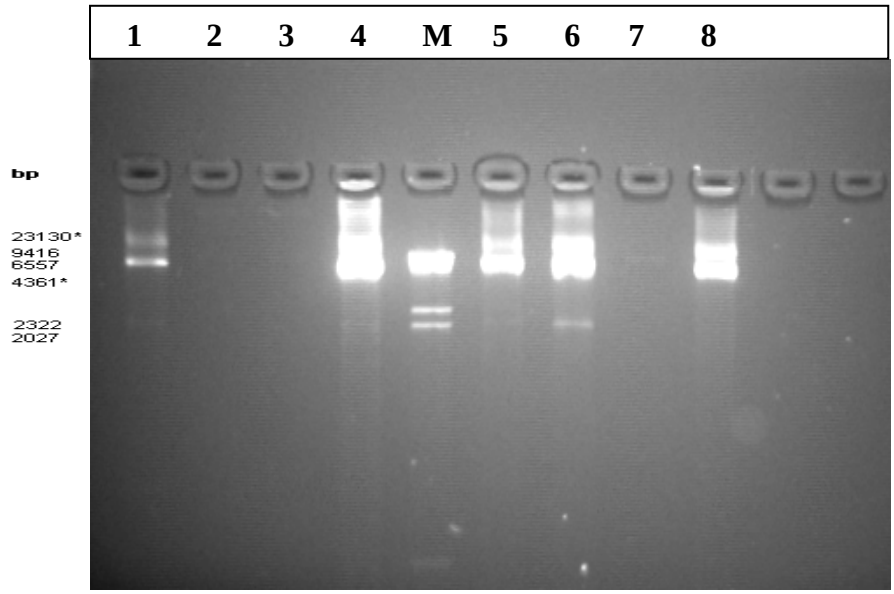


Figure 3.10 Results of site-directed mutagenesis application demonstrating all mutant types attained on each construct. **1-4,** ProGON1**; 1,** N-RPAP; **2,** N-RXAX**; 3**, N-S2A; **4,** N-H522A; **M,** λ DNA/*Hin*d III (50µg/ml); **5-8,** ProGOMN1; **5**, MN-RPAP; **6,** MN-RXAX**; 7,** MN-S2A; **8,** MN-H522A (see for the abbreviations)

Below are given the *Dpn*I digestion (see section 2.2.4.3) results to determine the size matching the bands expected.

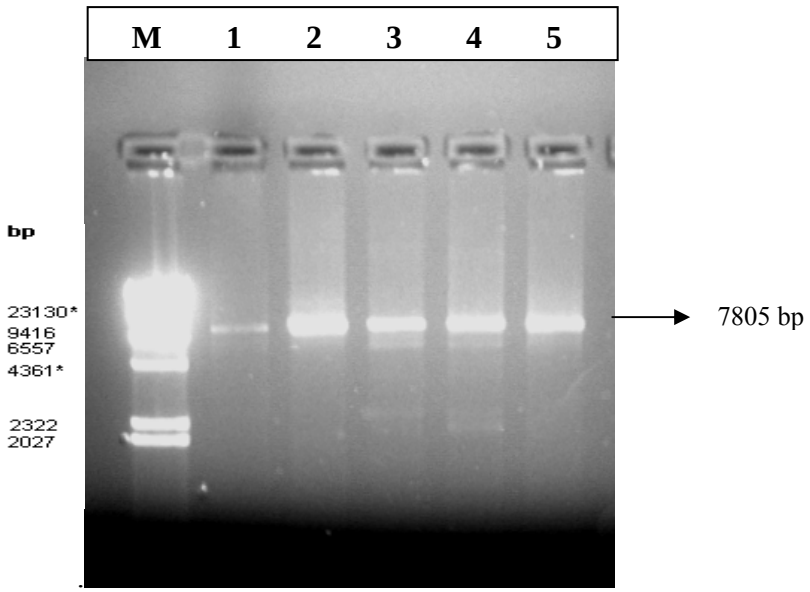


Figure 3.11 The bands representing mutant *gao* genes with 20ng/µl template DNA after *Dpn*I digestion. **1,** N-RPAP; **2,** N-HisA; **3,** MN-RPAP; **4,** MN-RXAX; **5,** MN-H522A.

As it is seen in Figure 3.10; N-RXAX**,** N-S2A and MN-S2A mutant forms could not be obtained at this stage of the experiments.

However, several trials of thermal cycling were performed in order to achieve all mutants by altering the amount of the template plasmid DNA without changing the annealing temperature. 5ng/µl template DNA was used instead of 20ng/µl. Some mutant fragments obtained after thermal cycling were strangely getting stuck within the wells, the reason for this was thought to be concatamer formation. It was shown that increasing the annealing temperature gradually could solve this problem.

N-S2A (see Figure 3.12) mutation was generated at a later stage of this study, although ProGON1 template later shown was not very suitable for mutations as explained below.

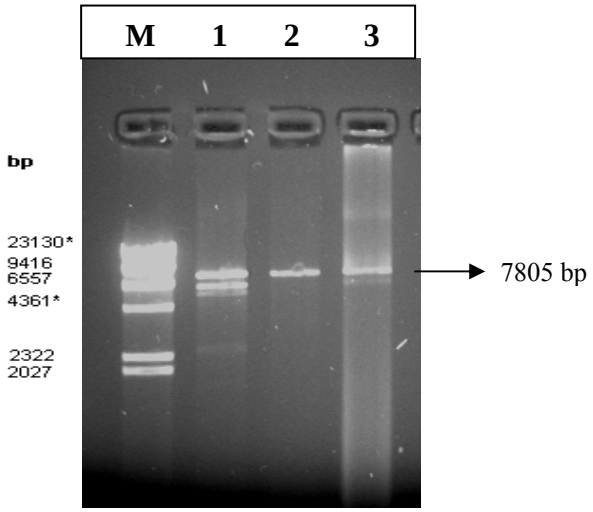


Figure 3.12 *Dpn*I digestion results of the mutants amplified over the templates having a concentration of 5ng/µl. **M,** λ DNA/*Hin*d III (50µg/ml); **1,** N-RPAP; **2,** N-S2A; **3,** MN-RPAP.

3.4.2 Transformation of the Mutants into *E. coli* and Sequence Analysis

As soon as the mutant *gao* genes were attained, these constructs were transformed into XL1-Blue cells as described in section 2.2.4.4. This was followed by plasmid isolation and sequence analysis.

For DNA sequencing, four representative colonies were chosen per mutation and were cultivated as it explained in section 2.2.4.5. After transformation into XL1-Blue cells, plasmid mini-preparations were done using QIAprep® Spin Miniprep Kit as explained before. Isolated plasmid DNAs were digested using the restriction enzyme *Cla*I (Section 2.2.5.5). Both the plasmids before digestion (Figure 3.13. and Figure 3.15.) and the digested products (Figure 3.14 and Figure 3.16) were then loaded on an agarose gel to analyze the isolation efficiency and the accuracy. Yield is important for providing the appropriate amount of DNA for sequencing and it also indicates efficient isolation. Digestion using *Cla*I which has a unique restriction site on the vector would confirm the isolate of interest by giving a single fragment.

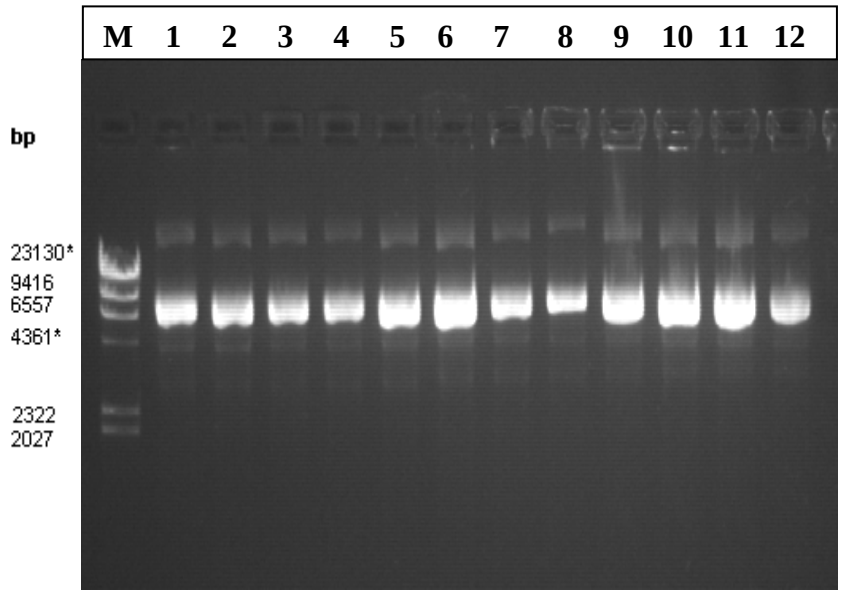


Figure 3.13 Plasmid isolation results of ProGON1 mutants. **M,** λ DNA/*Hin*d III (50µg/ml); **1-4,** N-RPAP; **5-8,** N-S2A; **9-12**, N-H522A.

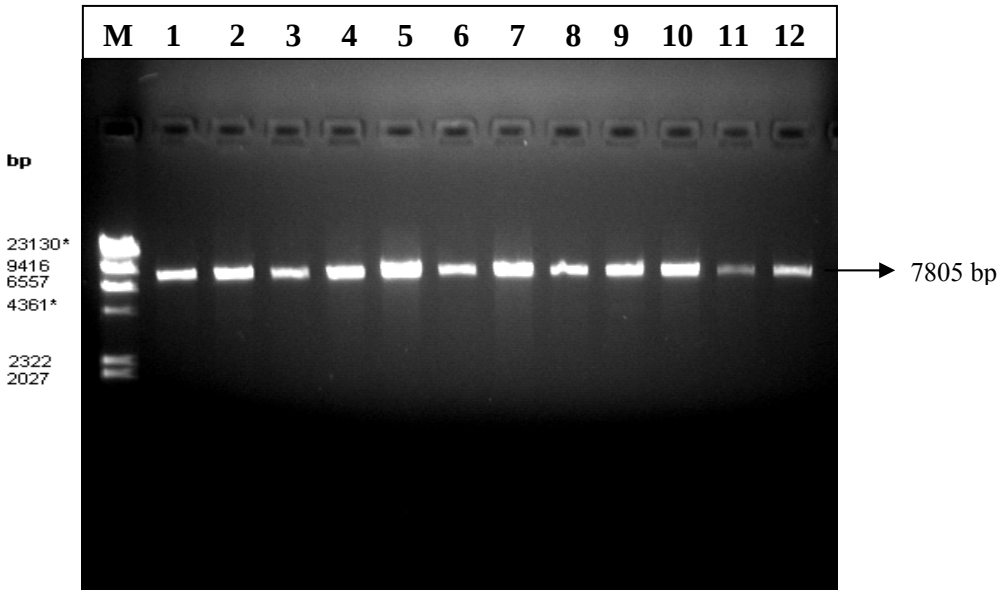


Figure 3.14 Mutated ProGON1 plasmid after *Cla*I digestion. **1-4,** N-RPAP; **5-8,** N-H522A**; 9-12**, N-S2A.

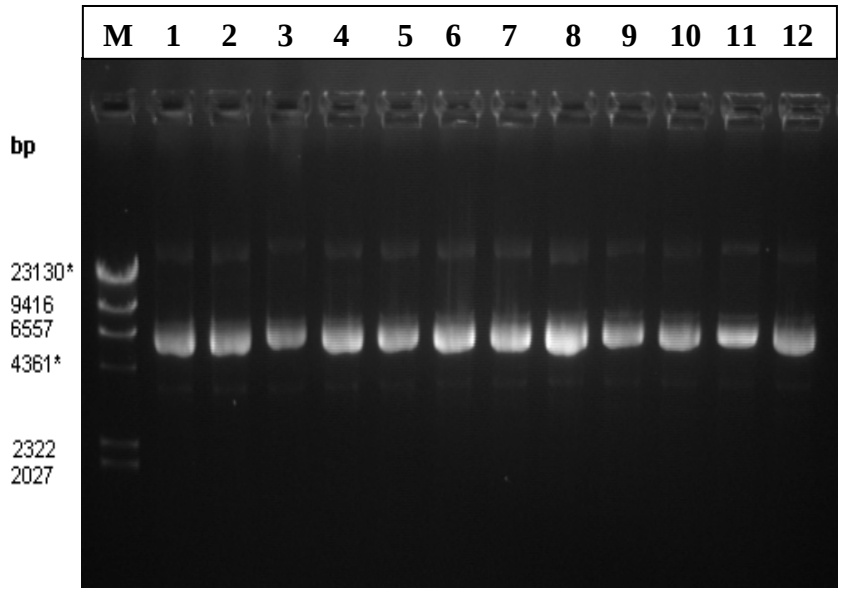


Figure 3.15 Plasmid isolation results of ProGOMN1 mutants **M,** λ DNA/*Hin*d III (50µg/ml); **1-4,** MN-RPAP; **5-8,** MN-RXAX; **9-12**, MN-H522A**.**

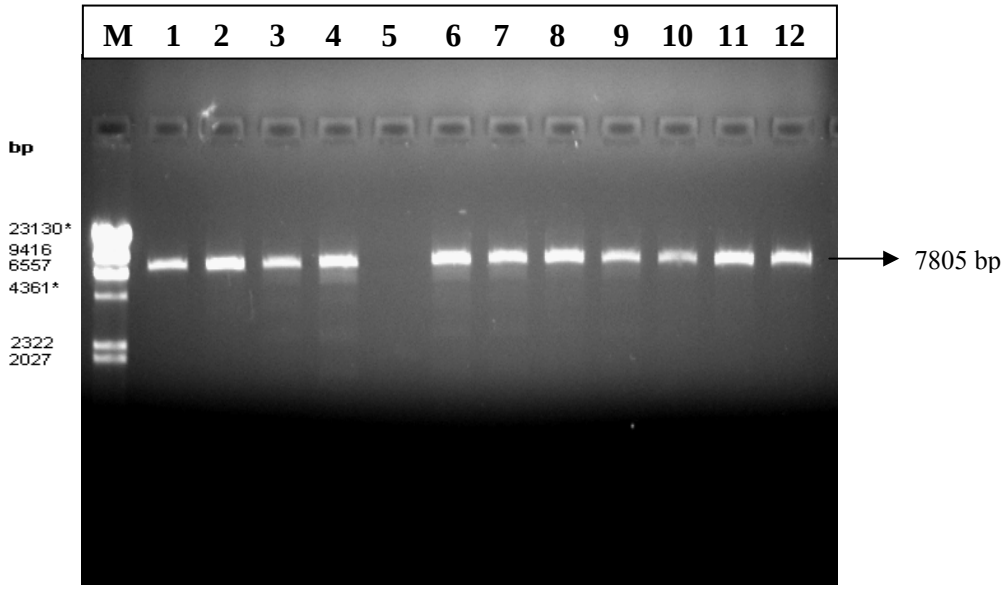


Figure 3.16 Restriction enzyme digestion with *Cla*I for mutated ProGOMN1 plasmid. **M,** λ DNA/*Hin*d III (50µg/ml); **1-4,** MN-RPAP; **5-8,** MN-RXAX; (lane 5 was not loaded) **9-12**, MN-H522A.

As it was mentioned before and seen in Figure 3.13 and Figure 3.15. N-RXAX mutant of ProGON1 and MN-S2A mutant of ProGOMN1 could not be obtained. These mutations were later repeated and successfully obtained by using same conditions given in section 3.4.1.

As explained above, four representative mutant clones were obtained for each mutation. These mutants were then analyzed by DNA sequencing. Alignments of both the DNA sequence and amino acid sequence were done by Clustal W programme (Appendix E). As it is seen in the data given, a 36 bp triplication (72 bp extra bit of sequence) appeared within all samples of ProGON1 constructs except one of the S2A clones, N-S2A 1. On the other hand, ProGOMN1 constructs did not seem to have such an undesirable triplication.

Taking into account the possibility that the problem resulted from sequencing, restriction enzyme digestion was applied both to the unmutated templates and the mutants to determine whether there is indeed a triplication or not (Appendix C). According to the plasmid restriction digests of *Pst* I & *Eco*RI, the expected band sizes were;

4550 bp
1776 bp
860 bp
559 bp
60 bp

OR → 631 bp if there was 72 bp extra sequence

N-S2A 1 which did not seem to have having the triplication was taken as a reference point for comparison with other samples and especially with the ProGON1 and ProGOMN1 templates.

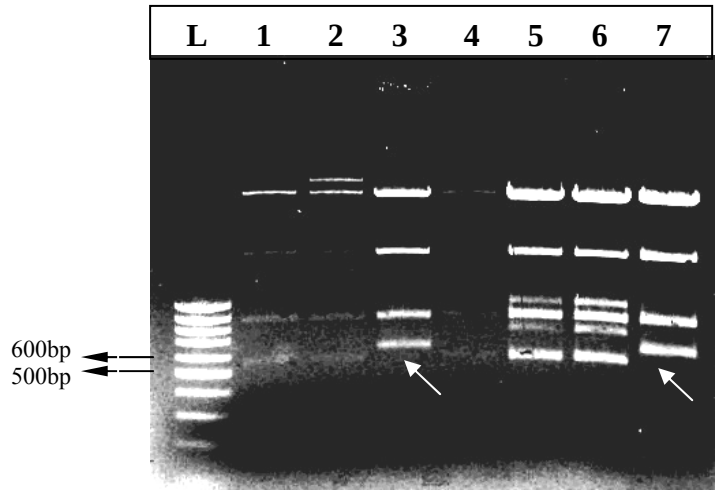


Figure 3.17 Double restriction enzyme digestion with *Pst* I & *Eco*RI for the detection of the 72 bp extra sequence **L**, 100bp ladder; **1,** ProGON1*; **2,** ProGOMN1*; **3,** ProGON1; **4,** ProGOMN1; **5,** MN-RPAP 1; **6,** N-S2A 1**;7,** N-S2A 2.

According to the results ProGON1* and ProGOMN1* showed the expected band sizes. On the other hand, it was obvious that the template ProGON1 and the mutants (here only the N-S2A 2 was shown as a representative) amplified over ProGON1, all carry a 631 bp band indicating triplication within the *gao* gene.

3.4.3 Preparation of ProGON1 without Triplication

QuikChange® Site-Directed Mutagenesis Kit was used for a second time with primers FGGO_T21A_F and FGGO_T21A_R which were given in Table D.1 to obtain original ProGON1 (this time using P*fx* DNA polymerase as described in section 2.2.4.2).

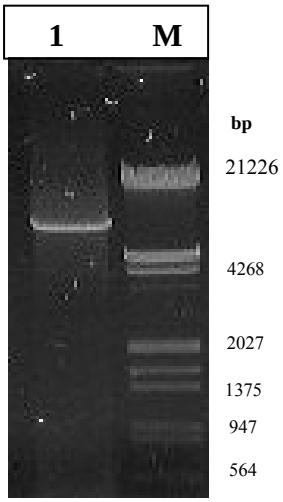


Figure 3.18 Agarose gel electrophoresis showing a single band of amplified putative ProGON1 **M,** λ DNA/*Eco*RI+*Hin*d III (0.5 µg/µl); **1,** putative ProGON1.

Restriction digestions were also made using *Nco*I. The expected band sizes for *Nco*I digestion were;

7031 bp
774 bp ⟶ 846 bp if there was 72 bp extra sequence

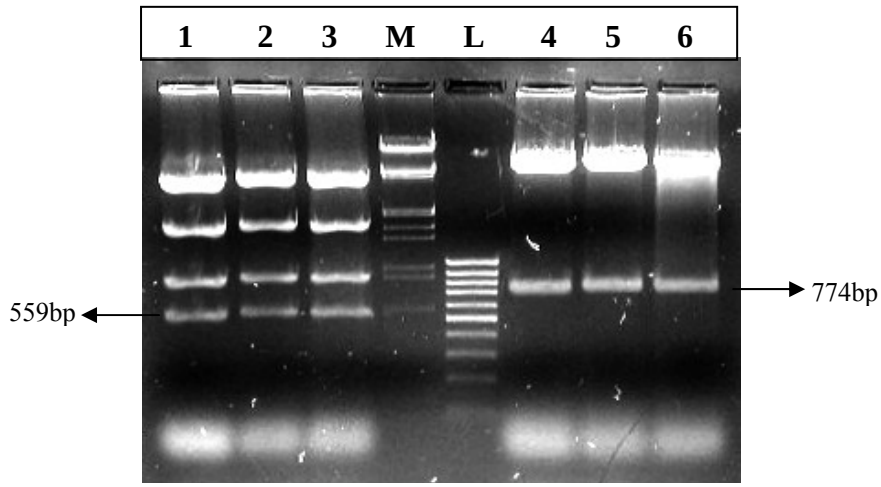


Figure 3.19 Detection of expected fragments as a product of multiple digestions. Lane 1-3 denote *Pst* I & *Eco*RI double digestion, lane 4-6 denote for *Nco*I single digestion. **M,** λ DNA/*Eco*RI+*Hin*d III (0.5 µg/µl); **1,** ProGON1**; 2,** MN-S2A 1 ; **3,** MN-S2A 2 ; **4,** ProGON1**; 5,** MN-S2A 1 ; **6,** MN-S2A 2.

These results confirmed that the ProGON1 constructs had an extra sequence whereas the ProGOMN1 constructs did not (Appendix E). Therefore, further studies were continued with the ProGOMN1 mutants. As an advantage, those clones are based on the template generated by directed evolution and were previously shown to have much higher efficiency of expression in *E. coli* (Deacon, personal communication). In Table 3.2, besides the desired mutations, some extra mutations were also present on some of the clones (Appendix E).

3.5 Heterologous Expression and Purification of Mutant GOase Proteins

3.5.1 Heterologous Expression of Mutant GOase Enzymes

Following the verification of mutants by DNA sequencing, the plasmids were transformed into BL21 Star (DE3) (Invitrogen) cells by heat shock transformation method (section 2.2.6.1). BL21 Star (DE3) is an optimised strain for recombinant protein expression allowing high-level expression of T7-regulated genes with IPTG induction. Its protease and RNase deficient nature enhance the capability of recombinant protein expression.

After transformation, single colonies were picked from plates as a representative of each type of mutation. Cells were then grown for the expression of mutant GOase. The experimental procedure is indicated in Section 2.2.6.3. When the cells were harvested, cell extracts were obtained without using protease inhibitor in order to compare its effect on cleavage. However later, experiments in the presence of protease inhibitor (Complete Mini, ROCHE) were also performed to hinder possible protease activity and qualitative GOase assay was performed as described in section 2.2.7.1.Accordingly, the results shown in Table 3.1 were obtained (also see Figure 3.20).

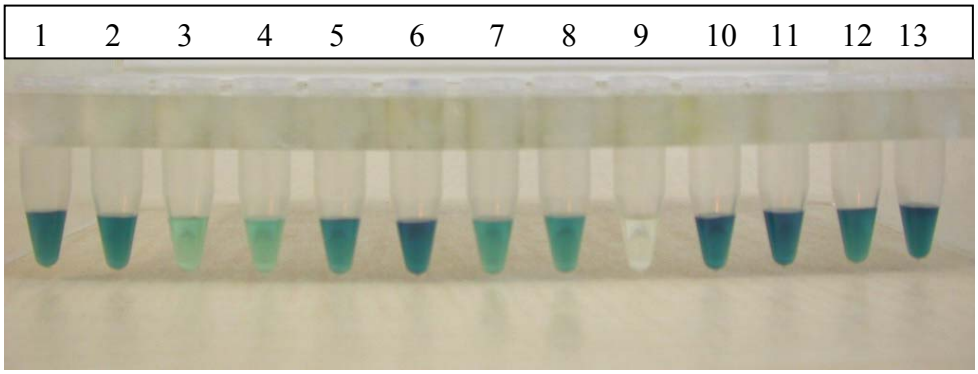


Figure 3.20 A set of qualitative assay results of the ProGON1, ProGOMN1 and the mutants. **1,** ProGOMN1; **2,** ProGON1; **3**, MN-RPAP 1; **4,** MN-RPAP 3; **5,** MN-RPAP 4; **6,** MN-RXAX 3; **7,** MN-RXAX 4; **8**, MN-H522A 1; **9,** MN-H522A 2; **10,** MN-S2A 1; **11,** MN-S2A 2; **12,** MN-S2A 3; **13,** MN-S2A 4.

As it is seen in Figure 3.20 above MN-H522A 2 has a light green colour which may indicate low expression level or degradation.

According to the result in Table 3.1 no colour change among mutant samples may denote either lack of expression of the recombinant protein or lack of processing. This was later clarified by SDS-PAGE after purification.

Table 3.1 Activity results of the mutants and mutants with extra unexpected mutations

CONTROLS	TYPE	ACTIVITY	APPROXIMATE MOLECULAR MASS
GOase (commercial)	**(+)**	+	65.5 kDa
PROGON1	**(+)**	+	above 70 kDa
PROGOMN1	**(+)**	+	above 70 kDa
PROGOMN1	**(-)**	-	
MUTATIONAL VARIANTS	**EXTRA MUTATIONS**	**ACTIVITY**	**APPROXIMATE MOLECULAR MASS**
N-S2A 1	S20C,Q22L	+	-
MN-R-1P/A1P 1	-	+	above 70 kDa
MN-R-1P/A1P 2	H522A	-	-
MN-R-1P/A1P 3	-	+	above 70 kDa
MN-R-1P/A1P 4	S7N	+	above 70 kDa
MN-R-1F(L)/A1V 1	deletion	-	-
MN-R-1N/A1P 2	deletion	-	-
MN-R-1C/A1S 3	-	+	above 70 kDa
MN-R-1G/A1G 4	L-2F	+	-
MN-S2A 1	-	+	above 70 kDa
MN-S2A 2	-	+	above 70 kDa
MN-S2A 3	-	+	above 70 kDa
MN-S2A 4	-	+	above 70 kDa
MN-H522A 1	-	+	above 70 kDa
MN-H522A 2	N521K	+ (very low)*	above 70 kDa
MN-H522A 3	-	+	-
MN-H522A 4	T519P, T520P,N521Y	-	-

* Hardly detectable activity by qualitative GOase assay

3.5.2 Purification and Analysis of GOase Mutants

Each dialysed cell extract carrying a mutational variant of GOase was loaded to the equilibrated column. After the adsorption of the Strep-tag II fused GOase to the column, the column was washed. Then the recombinant mutant GOase enzyme was eluted in 6 fractions. Subsequently the column was regenerated and equilibrated to be ready for the next run (section 2.2.7.2).

3.5.2.1 Effect of Copper Treatment on ProGOMN1-Generated Unmutated GOase

SDS-PAGE analysis was made first with ProGOMN1 as a positive control. Each fraction of ProGOMN1 was loaded to the wells of the gel. In these trials, ProGOMN1 was not treated with Cu^{2+}.

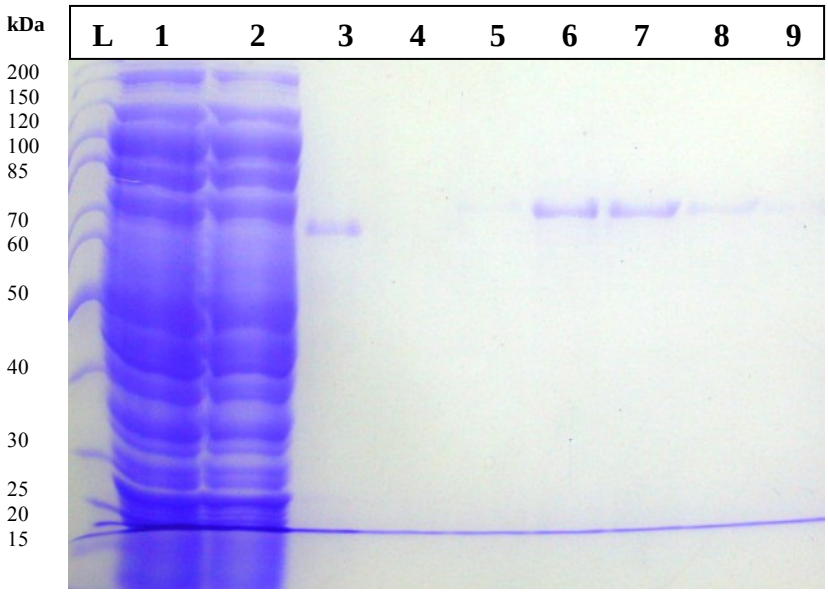


Figure 3.21 Purification of unmutated ProGOMN1 from *Strep*-Tactin® Sepharose® column in 6 fractions. **L,** Protein ladder; **1**, Crude Extract (Pre-load); **2,** Flow through; **3,**GOase (commercial, Sigma); **4-9,** Fractions 1-6.

The bands belonging to ProGOMN1 in Figure 3.21 migrated slightly above the 70 kDa band. Even though there was a trace amount of copper in the cells, which

was realized from the colour change in qualitative assay, $Cu(SO_4)$ was added to a final concentration of 50 μM and cell extracts were incubated for several hours at room temperature in order to assure cleavage of the pro-sequence. Results of copper-treated ProGOMN1-generated unmutated GOase were shown in Figure 3.22.

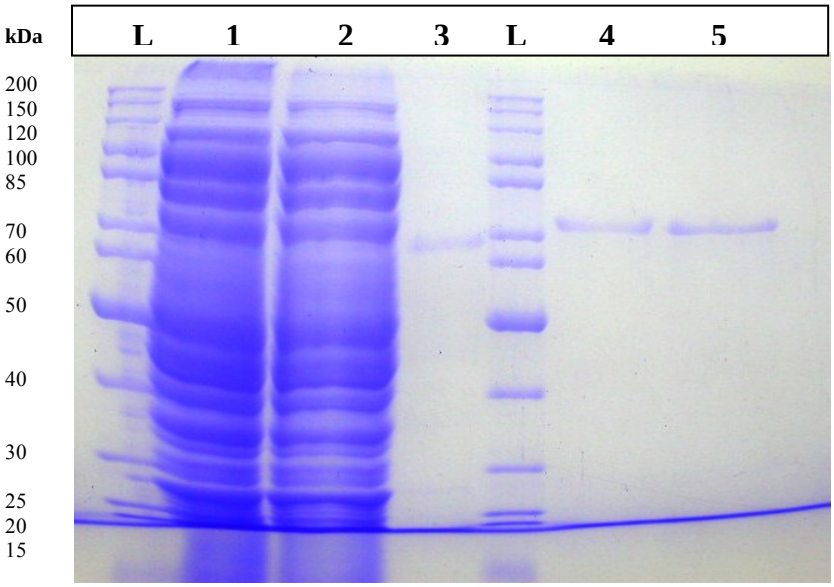


Figure 3.22 SDS-PAGE to realize ProGOMN1 processing after Cu^{2+} treatment. **L,** Protein ladder; **1**, Crude Extract (Pre-load); **2,** Flow through; **3,** GOase (commercial, Sigma); **4,** ProGOMN1 (-) copper; **5,** ProGOMN1 (+) copper.

As it is seen in Figure 3.22, there was no difference between the sizes of bands after copper treatment. After aerobic addition of copper into the pure GOase, self-catalytic removal of pro-sequence is expected yielding a 65.5 kDa processed enzyme. An intriguing point here is the migration of the protein above 70 kDa. The difference between the observed and expected band sizes was first supposed to be resulting from the *Strep*-tag II at the C-terminus allowing purification. Mutant GOases were purified and analyzed in order to find an explanation to this ambiguity.

3.5.2.2 Effect of Different Protease Inhibitors

In expression applications for the original construct and the mutational variants, the same type of protease inhibitor was used (Complete Mini,ROCHE). The effect of the

protease inhibitor was also investigated by using two different kinds of protease inhibitor, one was EDTA-free and the other may have contained EDTA. Since EDTA is a chelating agent and may act on metalloproteins, it was possible that it effects self-processing by GOase. This experiment was carried out on unmutated GOase from ProGOMN1 (Figure 3.23).

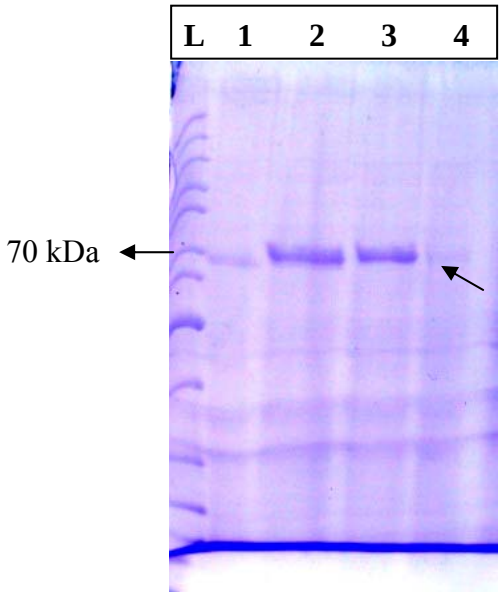


Figure 3.23 Protease inhibitor type effect on cleavage. **L,** Protein ladder; **1**, GAO (commercial); **2,** ProGOMN1 (- protease inhibitor); **3,** ProGOMN1 (+ protease inhibitor); **4,** ProGOMN1 (+ EDTA-free protease inhibitor)

It was seen that protease inhibitor with or without EDTA has no crucial role on self-processing, as all proteins were the same size.

3.5.2.3 SDS-PAGE Analysis of GOase Mutants

As mentioned before, the mutations were designed so that they can possibly prevent the cleavage of the pro-peptide. So, if cleavage did not take place this should yield a protein of size larger than ProGOMN1-generated unmutated GOase. Although the size of the unmutated GOase was larger than expected, as mentioned above, this was likely to be due to the existence of the *Strep*-tag. Since cleavage of the pro-peptide was autocatalytic, absence of cleavage in the unmutated GOase was something

unexpected. Thus in any case, mutations should yield a larger protein if they were able to prevent cleavage. Except MN-S2A mutations (they are all shown together on the same gel, Figure 3.27); only one single gel was shown per mutation as a representative (see Figures 3.24-3.26).

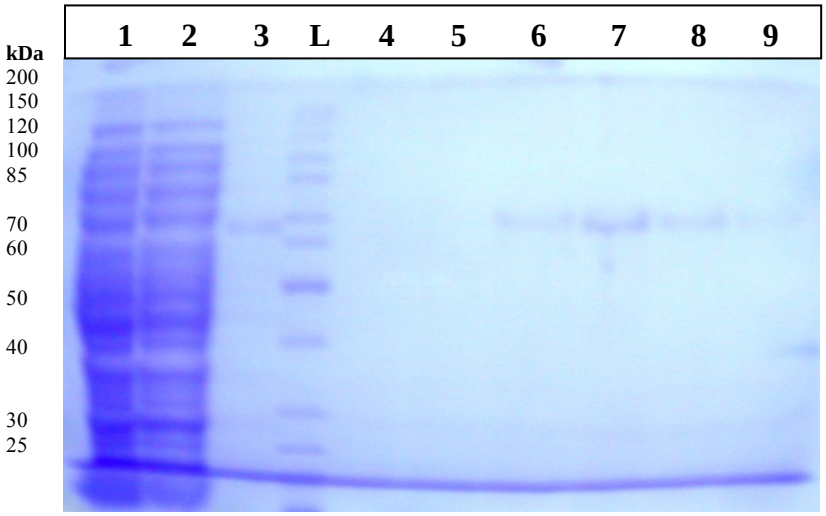


Figure 3.24 MN-RPAP 1 fractions collected from *Strep*-Tactin® Sepharose® column **L,** Protein ladder ; **1**, Crude Extract (Pre-load); **2,** Flow through; **3,**GOase (commercial, Sigma); **4-9,** Fractions 1-6.

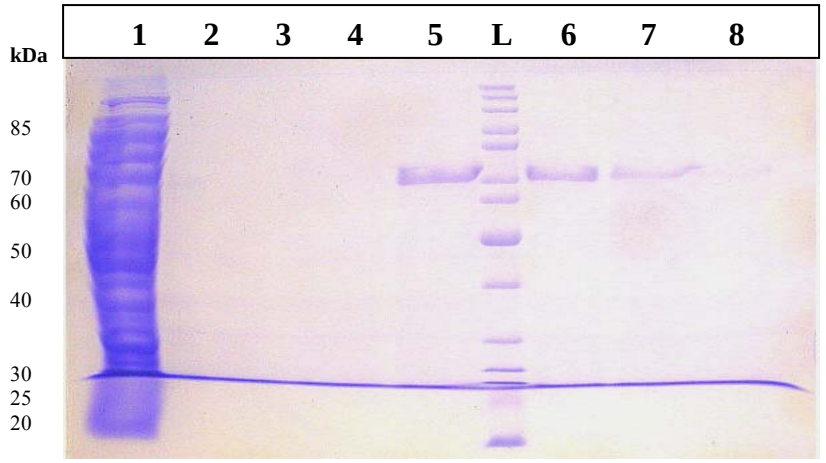


Figure 3.25 MN-RXAX 3 fractions gathered from *Strep*-Tactin® Sepharose® column **L,** Protein ladder ; **1**, Flow through; **2,** GOase (commercial, Sigma); **3-8,** Fractions 1-6.

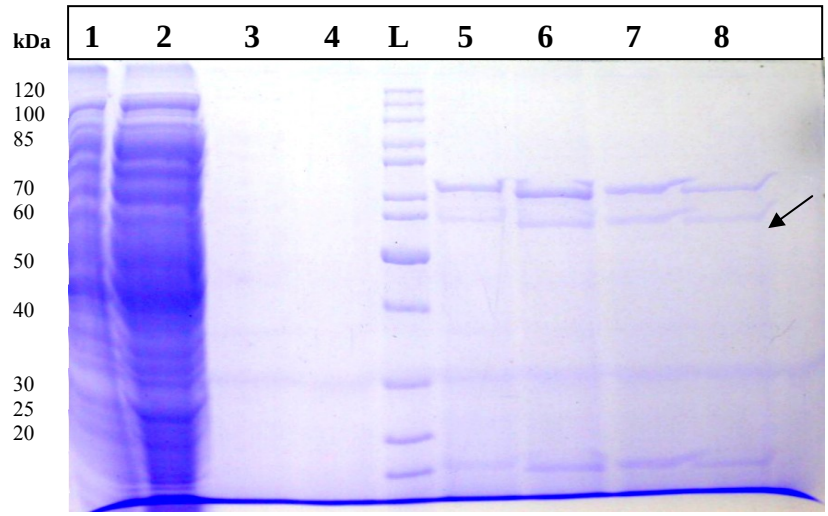


Figure 3.26 SDS-PAGE for MN-H522A 2 fractions **L,** Protein ladder ; **1**, Crude Extract (Pre-load); **2,** Flow through; **3-8**;Fractions 1-6.

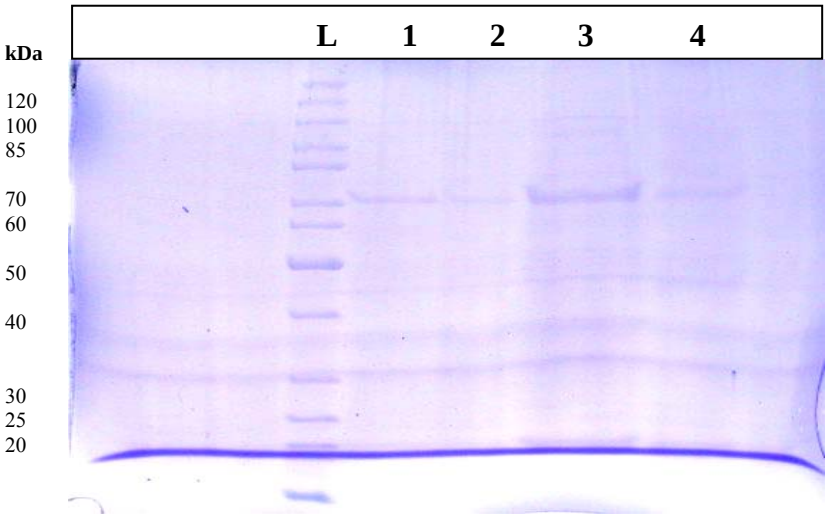


Figure 3.27 SDS-PAGE analysis for MN-S2A samples altogether **L,** Protein ladder ; **1**, MN-S2A 1; **2,** MN-S2A 2; **3,** MN-S2A 3; **4,** MN-S2A 4.

In MN-H522A 2 sample (indicated with arrows in Figure 3.26) there was a lower-size band pointing out a possible degradation. MN-H522A 2 contains an extra mutation (N521K) which may generate a putative processing site (Lys-Ala) for a specific protease (see Table 3.1 and Appendix E). Low activity by GOase-HRP

coupled assay, previously shown in Figure 3.20, may be due to these two point mutations located against the cleavage site on three dimensional configuration.

When the Table 3.1 was examined, it can be seen that in some samples there is activity in spite of no band was detected such as N-S2A 1, MN-RXAX 4, and MN-H522A 3 possibly due to the low expression level.

According to the Table 3.1, in some mutant types neither the bands nor the activity was observed. MN-RXAX 1 and MN-RXAX 2 have deletions leading a frame shift which results in a totally different protein. On the other hand, MN-RPAP 2 and MN-H522A 4 mutants may have conformational changes on their main structure caused by extra mutations.

As can be seen in Figures 3.24-3.27, the size of bands of all mutants indicates the same molecular weight. It was expected that especially on unmutated GOase and also the mutants which yielded green colour formation on ABTS assays indicating activity, pro-sequence should have been removed. However, the size of the protein bands were all the same and larger than expected.

As it was mentioned before, all GOase constructs including the mutants were carrying an eight amino acid encoding *Strep*-tag at C- terminus. This would result in a 1 kDa increase in molecular mass. If the pro-sequence had been removed the expected band size would be 66.5 kDa (mature peptide +1 kDa *Strep*-tag).

All the bands migrated slightly above the 70 kDa band. This size is unlikely to have resulted only from the *Strep*-tag itself. The peptide mass was estimated by using Peptide Mass Prediction program. According to this data, if the pro-sequence had not been removed the expected size of the protein is approximately 69.5 kDa (65.5 kDa mature peptide + ~1 kDa for the 8 amino acid preceding sequence before pro-sequence + 1.7 kDa pro-sequence + 1 kDa *Strep*-tag).

In order to explain the situation there were two possible approaches. Since it was not evident that processing had occurred, it was decided to evaluate *Strep*-tag carrying mature GOase (without the pre-pro-peptide) and the others together in an

SDS-PAGE. The second approach was to submit ProGOMN1-generated GOase for N-terminal amino acid sequencing.

3.5.3 Analysis Based on Comparison with MatGOMN6

MatGOMN6 is the best construct among the developed mature GOases which shows the highest expression in *E.coli* (Deacon, personal communication). The schematic diagram of this construct can be seen in Figure 3.28.

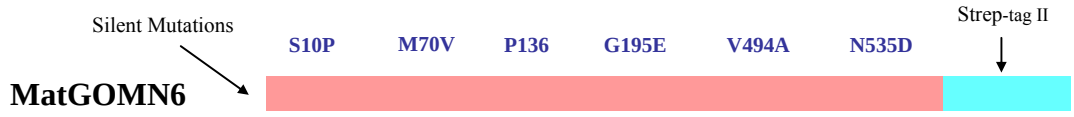


Figure 3.28 Schematic diagram of MatGOMN6 construct

In addition to MatGOMN6, the protein product of ProGON1 was also obtained following its recovery without the unwanted triplication. The mutants were loaded to the same SDS-PAGE gel together with the control samples (Figure 3.29 and Figure 3.30).

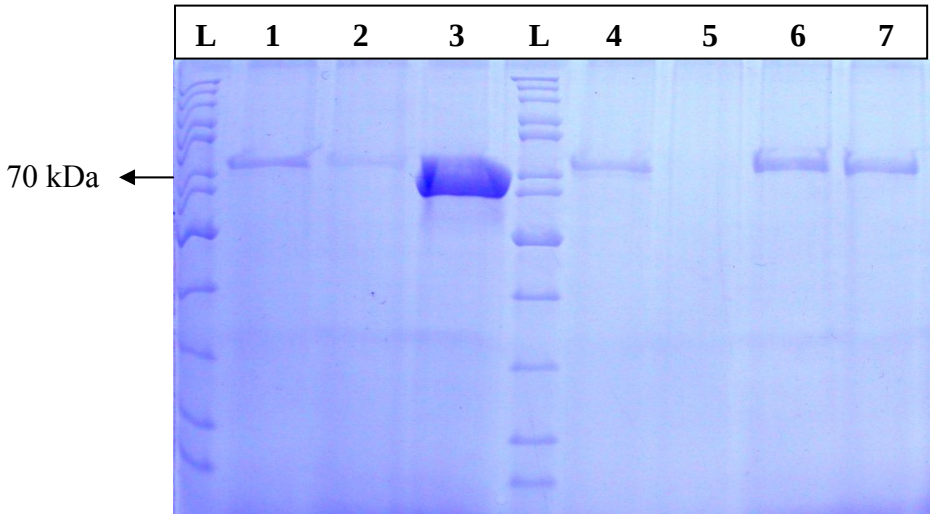


Figure 3.29 Comparison of the mutant GOases with mature enzyme MatGOMN6. **L,** Protein ladder; **1**, ProGOMN1; **2,** ProGON1; **3,** MatGOMN6; **4,** MN-RPAP 1; **5,** MN-RPAP 3 (very dilute); **6,** MN-RPAP 4; **7,** MN-RXAX 3.

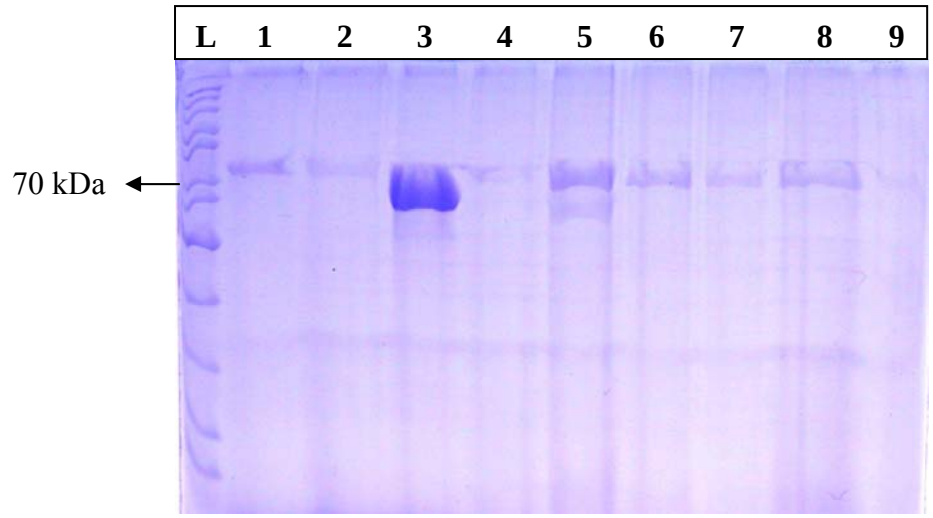


Figure 3.30 SDS-PAGE for comparison of the mutant GOases with mature enzyme MatGOMN6. **L,** Protein ladder; **1**, ProGOMN1; **2,** ProGON1; **3,** MatGOMN6; **4,** MN-H522A 1; **5,** MN-H522A 2; **6,** MN-S2A 1; **7,** MN-S2A 2; **8,** MN-S2A 3 ; **9,** MN-S2A 4.

There is a distinct difference between the sizes of the MatGOMN6 and the other samples. Since the significant difference between the mature construct and the pro-constructs is the existence of pro-sequence, this experiment strongly indicates that all the mutational variants and the original constructs ProGOMN1 and ProGON1 still carry the pro-sequence.

3.6 Analysis of the N-terminal Amino Acid Sequence of ProGOMN1-generated GOase

Samples to be sent for N-terminal sequence analysis were first loaded on SDS-PAGE gel (Figure 3.31) for the determination of concentration.

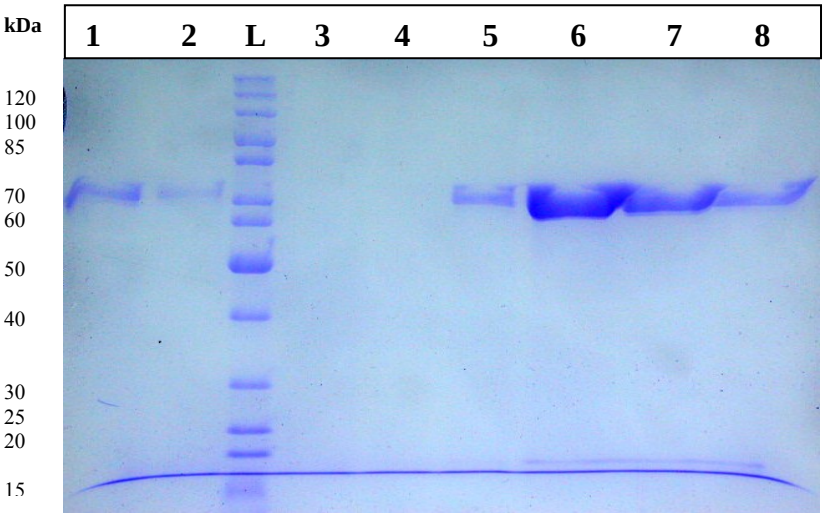


Figure 3.31 SDS-PAGE for detecting the amount of protein from two fold cell extract. **L,** Protein ladder; **1**, ProGOMN1 (- protease inhibitor); **2,** ProGOMN1 (+ EDTA-free protease inhibitor); **3-8;** Fractions 1-6 of ProGOMN1(+ protease inhibitor).

3.6.1 Electroblotting

After concentration was determined (section 2.2.7.4), 4 wells of gel were loaded with adequate amount of sample in order to provide the desired quantity for N-terminal sequencing. After the SDS-PAGE was run, the procedure explained in section 2.2.7.5 was followed. The membrane on which the samples were electroblotted was kept at -20 ^{o}C (Figure 3.32).

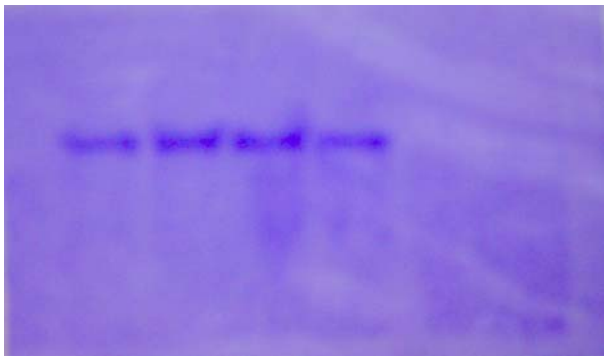


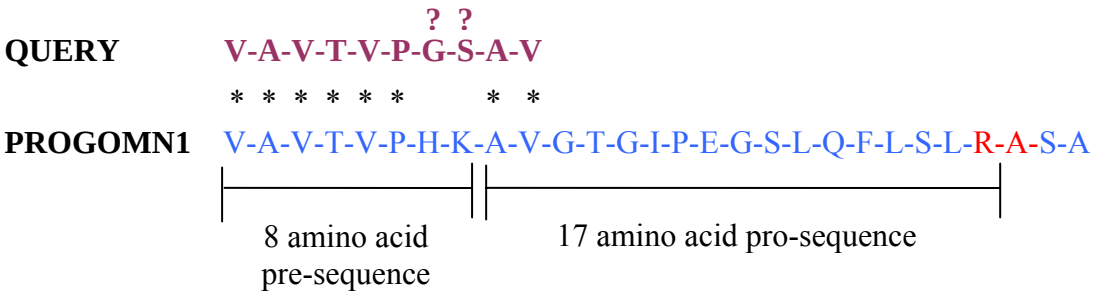
Figure 3.32 ProGOMN1 electroblotted PVDF membrane after staining.

3.6.2 N-terminal Sequencing

Sequence analysis was performed in the Institute of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, UK. According to the report the following sequence had been obtained:

Val-Ala-Val-Thr-Val-Pro-Gly?-Ser?-Ala-Val (initial yield c. 4pmol).

When these residues were aligned with ProGOMN1 deduced amino acid sequence of GOase; it is seen that the pro-sequence is still there together with the 8 amino acid preceding sequence (below). So, even the unmutated GOase was not processed when heterologously expressed in *E. coli.*



Taking into account other studies, especially the ones based on eukaryotic expression hosts, the pre-peptide cleavage may require a proteolytic process. So, the lack of specific proteases for this site or any different posttranslational modifications in *E. coli* might hinder this cleavage. Furthermore, the presence of pre-peptide may inhibit the autocatalytic cleavage of the pro-peptide. The pre-peptide may also have a role in vacuole targeting and sorting prior to secretion as explained in section 1.2.3. Following intracellular targeting in filamentous fungi, the pre-peptide might be cleaved.

Until now, mature, pre-mature and pro- forms of the enzyme have been detected. The intermediate pre-mature form lacking both the Cys-Tyr cross-link and the

pro-sequence indicated that they were separate reactions, in vivo (Rogers *et al.*,2000). Since a form which carries both the pro-sequence and the thioether bond was not detected so far, it was suggested that pro-sequence cleavage precedes thioether bond formation. However, in this study this fourth form of the enzyme (PRO-MATURE enzyme) is explored for the first time. It is asserted that thioether bond formation is independent of pro-peptide removal and may take place before self-processing of pro-peptide, refuting the idea that gives priority to self-cleavage.

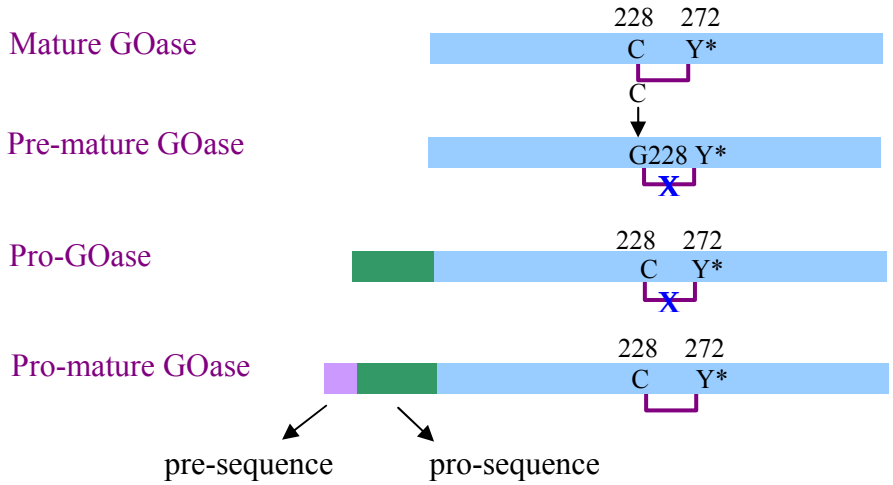


Figure 3.33 Forms of GOase

This also brings out a further strong statement .It is certain that GOase can be active in spite of the pro-sequence presence at its N-terminus.

3.7 Future Prospects

Recently, developments particularly in protein engineering and directed evolution studies may progress with the use of different enzymes in areas which have not been used before.

Galactose oxidase is a member of a growing class of proteins with novel posttranslationally modified redox-active amino acids. A greater understanding of these modifications might allow incorporation of similar centers into designed enzymes (Firbank *et al*, 2001).

The current work provides a contribution to the study of the mechanism of autocatalytic cofactor generation in galactose oxidase and may also provide insights into the biogenesis of cross-linked cofactors found in other proteins. Self-processing capability, using available reagents such as metal ions and dioxygen to generate new types of reactivity might represent a key step in the evolution of enzymes (Rogers *et al*, 2000).

Galactose oxidase may prove a prototype for understanding copper-dependent proteolytic events that can occur in certain disease states, such as Alzheimer precursor protein cleavage.

Which folding properties of the cupredoxins are necessary for the function is the subject of site-directed mutagenesis studies. It is obvious that the efforts to clarify the activation mechanism in galactose oxidase by mutations in pro-sequence will offer a solution for modelling and theoretical studies of several new enzymes.

On the basis of this study the encoding sequence of pro-GOase for 8 amino acid residues, preceding the 17-amino acid pro-sequence, can also be deleted by site-directed mutagenesis in order to analyze whether pro-sequence could be removed without that short oligopeptide. If it is detected that the pro-sequence is being cleaved, this time any desired mutations can also be carried out on this newly generated construct to explore autocatalytic processing.

CHAPTER 4

CONCLUSION

The aim of this study was to analyze the autocatalytic cleavage mechanism of galactose oxidase pro-sequence by site-directed mutagenesis in a heterologous system, *E. coli*. It was clear that this finding would bring light to such ambiguities observed in other posttranslationally modified enzymes.

In this thesis, four point mutations were carried out on previously developed constructs carrying manipulated *gao* by directed evolution. The successfully generated mutations were R-1P/A1P, R-1X/A1X, S2A, and H522A. Following expression in BL21Star (DE3) the mutant galactose oxidases were purified in order to detect the mutation effect on cleavage. The SDS-PAGE results indicate that all the mutants including the unmutated templates, still have the pre-pro-sequence at their N-termini, which was later confirmed by N-terminal amino acid sequencing. On the other hand, it was signified by qualitative galactose oxidase assays that all mutational variants together with the unmutated constructs show activity inspite of the pre-pro-sequnce.

Apart from the effect of mutations, as an unexpected result, pro-sequence does not affect thioether bond formation at the active site. As a matter of great importance, this pro-sequence carrying active pro-GOase represents the fourth elusive form of the enzyme (carries both the pro-sequnce and thioether bond) which has been able to be observed for the first time. Furthermore, this finding also contradicts with the supposition that pro-sequence cleavage may take place before the formation of the thioether bond, since it was shown here that these events are totally irrelevant, seperate events either of which can occur formerly.

Of further importance, is the observation of lack of cleavage in the pre-pro-GOase construct. Since pro-GOase processing is autocatalytic, in the presence of copper and dioxygen, absence of processing is likely to be prevented by the pre-peptide. This suggestion should be further studied by deleting the sequence corresponding to the pre-peptide, on ProGON1 and ProGOMN1 constructs.

REFERENCES

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

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

Andreana P.R., Xie W., Cheng H.N., Qiao L., Murphy D.J., Gu Q.M., Wang P.G. (2002) *Org. Lett.* **4,** 863-1866.

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

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

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.

Baron, A. J., Stevens, C., Wilmot, C.M., Knowles, P. F., Phillips, S. E. V., & McPherson, M. J. (1993) *Biochem. Soc.Trans.* **21,** 319S.

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.

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

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

Branchaud, B.P., Montague-Smith, M.P., Kosman, D.J., McLaren, F.R.J. (1993) *Am. Chem. Soc.* **115**, 798.

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

Bullock, W.O., Fernndez, J.M. and Short, j.M. (1987) *Biotechniques* **5(4),** 376-378.

Bussink, H.J.D., Brouwer, K.B., De Graaff, L.H., Kester, H.C.M. & Visser, J. (1991a).*Current Genetics **20,*** 301-307.

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

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

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

Delagrave, S., Murphy, D. J., Pruss, J. L. R., Maffia, A. M., Marrs, B. L., Bylina, E. J., Coleman, W. J., Grek, C. L., Dilworth, M. R., Yang, M. M., and Youvan, D. C. (2001) *Protein Engineering* **14,** 261-267.

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

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

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

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.

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

Fontecave, M. and Pierre, J.L. (1998) *Coordination Chemistry Rewievs* **170**, 125-140.

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

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

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.

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

Hardy, S.J.S. and Randall, L. L. (1992) *Antonie van Leeuwenhoek: International Journal of General and Molecular Microbiology* **61(2)**, 101-103

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

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

Himo, F.and Siegbahn, Per E.M. (2003) *Chem. Rev.* **103.** 2421-2456.

Ito, N., Phillips, S.E.V., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D, 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,** 000-000.

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

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., and Heineman, W. R. (1982) *Anal. Chem.* **54,** 1394-1399

Kain, W., Rall, J.(1996) *Angew. Chem. Int. Ed. Engl*. **35**, 43.

Kelleher, F.M., and Bhavanandan, V.P.(1986) *J. Biol. Chem*. **261**, 11045-11048.

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

Kim, M., Okajima, T., Kishishita, S., Yoshimura, M., Kawamori, A.,Tanizawa, K. and Yamaguchi, H. (2002) *Nat. Struct. Biol*. **9**, 591–596.

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

Kosman, D.J., (CRC Press, Boca Raton publishers, Boca Raton 1984) *Copper Proteins and Copper Enzymes,* **1** (ed Lontie, R.), 1–26.

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

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

Li, P., Beckwith, J. and Inouye, H. (1988) *Proceedings of the National Academy of Sciences* USA **85,** 7685-7689

Lis, M. and Kuramitsu, H. K. (1997) *Antimicrob. Agents Chemother*. **41**, 999-1003
Liu, X. C. And Dordick, J. S.(1999) J. Am. Chem. Soc., **121**, 466-467

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

Maradufu, A., Cree, G. M., and Perlin, A. S. (1971) *Canadian Journal of Chemistry* **49,** 3429-3437.

Maradufu, A., and Perlin, A. S. (1974). *Carbohydrate Research* **32,** 93-99.Whittaker, M.M and Whittaker, J.W (1988) *J.Biol.Chem*. **263,** 6074-6080**.**

Martı´nez-Ruiz, A., del Pozo, A. M., Lacadena, J., Manchen˜o, J. M.,On˜aderra, M., Lopez-Otı´n, C., and Gavilanes, J. G. (1998) *Pichia pastoris, Protein Expr. Purif.* **12,** 315–322.

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

Matsuzaki, R. , Fukui, T. , Sato, H. , Ozaki, Y. & Tanizawa, K. (1994) *FEBS Lett.* **351**, 360-364.

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

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

McPherson, M.J., Ogel, Z.B., Stevens, C., Yadav, K.D.S., Keen, J.N. and Knowles, P.F., (1992) *J.Biol.Chem* .**267**, 8146-8152.

McPherson, M.J., Stevens, C., Baron, A.J., Ogel, Z.B., Seneviratne, K.,Wilmot, C., Ito, N., Brocklebank, I., Phillips, S.E.V. and Knowles, P.F.(1993) *Biochem. Soc. Trans*. **21**, 752–756

Mendonca, M. H., and Zancan, G. T. (1987). *Archives of Biochemistry and Biophysics* **252,** 507-514.

Multhaup, G., Ruppert, T., Schlicksupp, A., Hesse, L., Bill, E., Pipkorn, R.,Masters, C.L. and Beyreuther, K. (1998) *Biochemistry* **37**, 7224–7230.

Ögel, Z.B. (1993).*Ph.D. thesis, Univ. of Leeds, Leeds, U.K.*

Ögel, Z.B., Brayford, D. and McPherson, M.J. (1994) *Mycol. Res*. **98,** 474–480

Ögel, Z. B. and Özilgen, M. (1995) *Enzyme and Microbial Technology* **17**, 870-876.

Okeley, N.M. and van der Donk, W.A. (2000) *Chem. Biol* **.7,** R159-R171.

Ostermeier, C. , Harrenga, A. , Ermler, U. and Michel, H. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10547-10553

Peberdy, J. F. (1994) *TIBTECH* **12**, 50-57

Perlman, D. and Halvorson, H.O. (1983).*Journal of Molecular Biology* **167**, 391-409.

Petrounia, I. P. and Arnold F.H.(2000) *Curr. Opin. Biotechnol*. **11,** 325-330

Pristle, J.P. (1988) *J.appl.Crystallogr.* **21,** 572-576.

Rapoport, T.A., Gorlich, D., Musch, A., Hartman, E., Prehn, S., Wiedman, M., Otto, A., Kostka, S. and Kraft, R. (1992) *Antonie van Leeuwenhoek: International Journal of General and Molecular Microbiology* **61(2)**, 119-122

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

Rogers, M. S., Baron, A. J., McPherson, M. J., Knowles, P. F., and 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.

Root, R.L.,Durrwachter.J.R. and Wong.C.H. (1985) *J. Am. Chem. Soc*., **107,** 2997-2999

Ruggiero, C. E. , Smith, J. A. , Tanizawa, K. & Dooley, D. M. (1997) *Biochemistry* **36**, 1953-1959.

Said, I.T., Shamsuddin,A.M., Sherief,M.A., Taleb,S.G., Aref,W.F. and Kumar,D. (1999) *Histol. Histopathol.,* **14,** 351-357.

Sarah E. Deacon, Khaled Mahmoud, R. Kate Spooner, Susan J. Firbank, Peter F. Knowles, Simon E. V. Phillips, and Michael J. McPherson (2004) *ChemBioChem* **5**, 972 - 979

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

Schwartz, T. W. (1986) *FEBS Letters* **200**, 1-10

Schwede, T. F. , Rétey, J. and Schulz, G. E. (1999) *Biochemistry* **38**, 5355-5361.

Shatzman, A.R. and Kosman, D. J. (1997) *Journal of Bacteriology* **130**, 455-463

Silen, J.L. and Agard, D.A. (1989) *Nature* **341**, 462-464.

Simonen, M.and Palva, I. (1993) *Microbiological Reviews* **57**(1), 109-137

Soliday, C. L., Flurkey, W. H., Okita, T.W. and Kolattukudy, P.E. (1984) *Proceedings of the National Academy of Sciences* USA **81**, 3939-3943

Springer, G. F. (1997). *J. Mol. Med.***75,** 595–602.

Stubbe, J.and van der Donk, W.A.(1998) *Chem. Rev.* **98,** 705-762

Sun, L., Petrounia, I.P., Yagasaki, M., Bandara, G.,Arnold, F.H., (2001) *Protein Engineering* **14,** 699-704.

Sun, L. H, Bulter, T., Alcalde, M., Petrounia, I.P., Arnold, F.H. (2002) *ChemBioChem* **3,** 781-783.

Szabo, E.E. Adanyi, N. and Varadi,M. (1996) *Biosens. Bioelectron.* **11,**1051-1058.

Szameit, C. , Miech, C. , Balleininger, M. , Schmidt, B. , von Figura, K. & Dierks, T. (1999) *J. Biol. Chem.* **274**, 15375-15381.

Tatsumi, H., Murakami, S.,Tsuji, R.F., Ishida,Y., Murakami, K., Masaki, A., Kawabe, H., Arimura, H., Nakano, E. and Motai, H. (1991) *Molecular and General Genetics* **228**, 97-103

Tkac, J., Gemeiner,P. and Sturdik,E. (1999) *Biotechnol. Tech*, **13,**931-936.

Tressel, P.S., and Kosman, D.J.(1982) *Methods Enzymol*. **89**, 163-171

van Poelje, P. D. & Snell, E. E. (1990) *Annu. Rev. Biochem.* **59**, 29-59.

Vega, F.A., Numez,C.G., Weigel,B., Hitzmann,B. and 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. and Marek,M. (1992) *Collect. Czech. Chem. Commun.*, **57,** 2287-2294.

Wachter, R.M.; Branchaud, B.P.(1996) *Biochemistry* **35,** 14425.

Wachter, R.M.; Branchaud, B.P.(1996) *J. Am.Chem.Soc.* **118,** 2782.

Wachter, R.M.; Montague-Smith,M.P.; Branchaud, B.P.(1997) *J.of Am.Chem.Soc*.**119,** 7743.

Wang, S. X. , Mure, M. , Medzihradszky, K. F. , Burlingame, A. L. , Brown, D. E.,Dooley, D. M., Smith, A. J. , Kagan, H. M. and Klinman, J. P. (1996) *Science* **273**, 1078-1084.

Wieland, F. T., Gleason, M. L., Serafini, T. A. and Rothman, J. E. (1987) *Cell* **50,** 289-300

Whittaker, J.W., (1995) *Meth.Enzym., Acad.Press,* **258**, 262.

Whittaker, J.W. (2003) *Chem. Rev* **103**,2347-2363.

Whittaker, J.W. (2005) *Archives of Biochemistry and Biophysics* **433** , 227–239.

Whittaker, M.M., Whittaker, J.W. (1988) *J. Biol. Chem*. **263**, 6074–6080.

Whittaker, M.M., Whittaker, J.W. (1990) *J. Biol. Chem*. **265**, 9610–9613.

Whittaker, M.M, Whittaker, J.W. (1993) *Biophys. J.* **64,** 762–772.

Whittaker, M.M., and Whittaker, J. W. (2000). *Protein Expression and Purification* **20,** 105-111.

Whittaker, M.M, Whittaker,J.W. (2003) *J. Biol. Chem*. **278,** 22090–22101

Wilkinson, D., Akumanyi, N., Hurtado-Guerrero, R., Dawkes, H., Knowles, P.F., Phillips, S.E.V. and McPherson, M.J. (2004) *Protein Engineering, Design and Selection* **17(2),** 141-148.

Xie, L. and 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. and Klotz, A. (2000) *Appl. Biochem. Biotechnol.* **88**, 23-32

Yang, G.Y. and Shamsuddin,A.M. (1996) *Histol. Histopathol.*, **11,** 801-806.

Yoshikawa, S. , Shinzawaitoh, 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.

Zhu, X., Ohta, Y., Jordan, F. and Inouye, M. (1989) *Nature* **339,** 483-484.

APPENDIX A

CHEMICALS, ENZYMES AND THEIR SUPPLIERS

Chemicals

Agar	Difco
Agarose	Sigma
ABTS	Applichem
Acrylamide	Applichem
Amonium peroxidisulfate	Fluka
β-Mercaptoethanol	Merck
Bisacrylamide	Applichem
Bovine Serum Albumin	Sigma
Bromophenol Blau	Merck
$CaCl_2$	Sigma
Coomassie Brillant Blue G 250	Merck
Coomassie Brillant Blue R 250	Merck
$Cu(NO_3)_2$	Merck
Cupric sulfate	Sigma
d-Desthiobiotin	Sigma
D-Galactose	Sigma
D(+)-Glucose monohydrate	Merck
EDTA	Merck
Ethidium Bromide	Sigma
Glacial Acetic Acid	Merck
Glucose	Sigma
Glycerol	Merck

HABA	Sigma
HCl	Merck
IPTG	MBI Fermentas
Isopropanol	Merck
KCl	Fluka
KH_2PO_4	Merck
Methanol	Merck
MOPS	Calbiochem
NaCl	Merck
Na_2CO_3	Merck
$NaHCO_3$	Merck
$Na_2HPO_4.7H_2O$	Merck
NaH_2PO_4	Merck
NaOH	Merck
Phosphoric acid	Merck
PIPES	Sigma
SDS	Merck
Sodium Acetate	Merck
TEMED	Fluka
Trichloroacetic acid	Riedel-de Haen
Tris Base	Merck
Trypton	Merck
Yeast Extract	Merck

Enzymes

Galactose oxidase	Sigma
Horseradish peroxidase	Sigma
KOD Hot Start DNA Polymerase	Novagen
Pfx DNA polymerase	Invitrogen
RNase A	Sigma
*Cla*I	Promega

*Dpn*I	MBI Fermentas
*Eco*RI	MBI Fermentas
*Nco*I	MBI Fermentas
*Pst*I	MBI Fermentas

Protease Inhibitors

Complete, Mini	Roche
Complete, Mini EDTA-free	Roche

DNA and Protein size markers

GeneRuler® 100 bp DNA Ladder Plus	MBI Fermentas
GeneRuler® Protein Ladder	MBI Fermentas
λ DNA/*Hin*d III DNA Marker	MBI Fermentas
λ DNA/*Eco*RI+*Hin*d III Marker	MBI Fermentas

Kits

QIAprep® Spin Miniprep Kit	Qiagen
QIAfilter® Plasmid Maxi Kits	Qiagen
Dialysis Membrane	Sigma
pET101/D/lacZ Vector	Invitrogen
Strep-Tactin® Sepharose®	IBA GmbH's

APPENDIX B

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.

Ampicillin (50 mg.ml^{-1} 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^{o}C. It is used at a dilution of 1/1000 (400 µl in 400 ml) to give 50 µg.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.

Blotting Buffer (pH=10,5) (500 ml)

$NaHCO_3$	0.42 g	10 mM
$NaCO_3$	0.159 g	3 mM
Methanol	100 ml	20%

After mixing all, ddH_2O is added to 500ml

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 0.85% NaCl and 0.1% sodium azide. 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 P3 (Neutralization Buffer)

3.0 M potassium acetate, pH = 5.5

Buffer QBT (Equilibration Buffer)

750 mM NaCl
50 mM MOPS, pH = 7.0
15% isopropanol
0.15% Triton X-100

Buffer QC (Wash Buffer)

1.0 M NaCl
50 mM MOPS, pH = 7.0
15% isopropanol

Buffer QF (Elution Buffer)

1.25 M NaCl
50 mM Tris·Cl, pH = 8.5;
15% isopropanol

Buffer W (Washing buffer)

100 mM Tris-Cl pH 8.0
150 mM NaCl

Buffer E (Elution buffer)

100 mM Tris-Cl pH 8.0
150 mM NaCl
2.5 mM desthiobiotin

Buffer R (Regeneration buffer)

100 mM Tris-Cl pH 8.0
150 mM NaCl
1 mM HABA (hydroxy-azophenyl- benzoic acid)

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 H_2O are added to 1 L. It is sotered in dark bottle at room temperature, recycled if necessary.

$Cu(NO_3)_2$ (1 mM)

0.188 mg is dissolved in 1 ml dH_2O.

$CuSO_4$ (100 mM)

15.95 g is dissolved in 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

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

Fixing Solution (20%)(100 ml)

20 g trichloroacetic acid

dH_2O is added to 100 ml and trichloroacetic acid is dissolved in it.It is kept at room temperature.

Galactose Oxidase Assay Solution

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.

LB (Luria-Burtani 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-Burtani) 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.

LB Tetracycline Agar (per Liter)

10 g NaCl
10 g tryptone
5 g yeast ex tract
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

$MgCl_2$ (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.

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)

PBS (x1)

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

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^oC.

Sample buffer (x4) (4ml)

20 % SDS	2.0 ml
1 M Tris-HCl, pH7	1.0 ml
Glycerol	1.0 ml
Bromophenol blau	few grains

Everything is mixed, stored at room temperature.

SDM

Sample buffer (x4)	4.0 ml	(80%)
Mercaptoethanol	1.0 ml	(20%)

It is prapered just before use (freshly made).

SDS (10%)

10g 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 5 N 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˚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.

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.

TAE Buffer (50x, per Liter)

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.

TE Buffer

10 mM Tris-HCl (pH 8.0)
1.0 mM EDTA (pH 8.0)

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 RbCl2	3.025g	121
10 mM CaCl2	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 CaCl2	1.1g	147.02
10 mM RuCl2	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.

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.

APPENDIX C

MAPS AND SEQUENCE DATA OF THE PROGON1 AND PROGOMN1

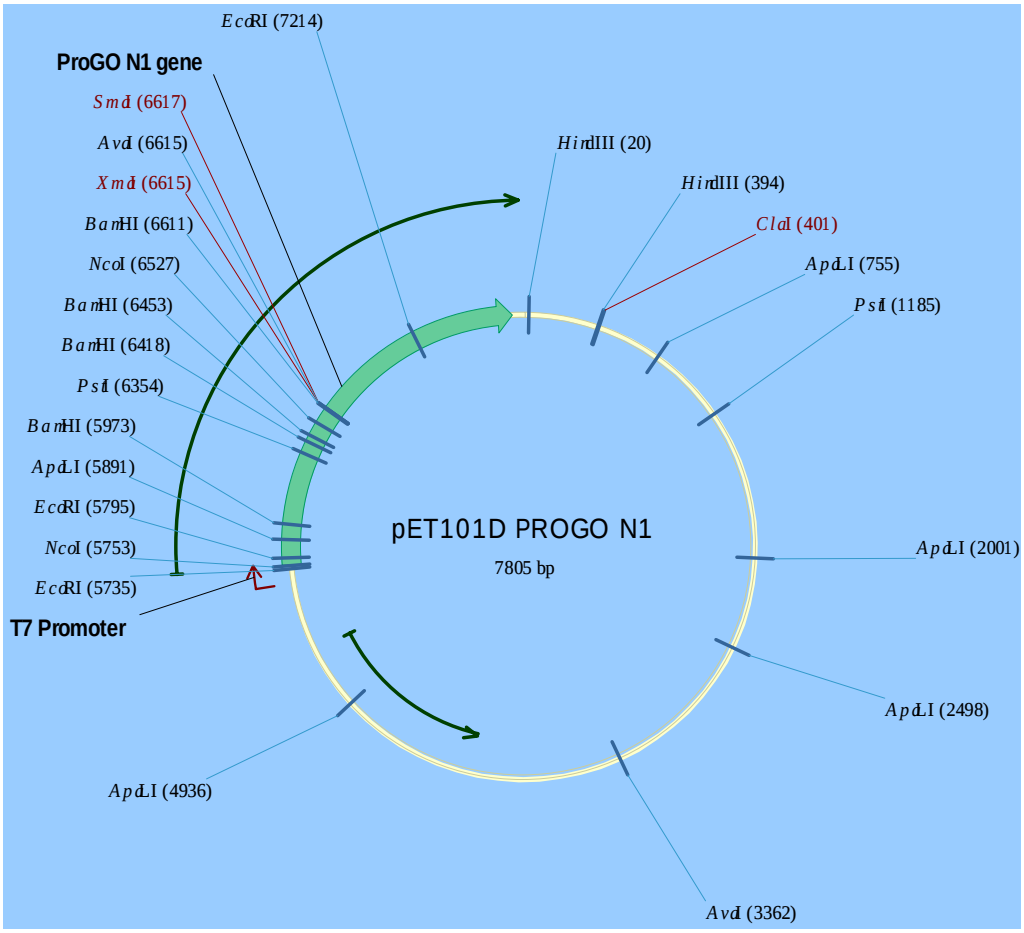


Figure C.1. The map of pET101D ProGON1 vector

ProGON1

atggttgcagttaccgttcctcacaaggccgtaggaactggaattcctgaagggagtcttcagttcctgagccttcgagcct
cagcacctatcggaagcgccatttctcgcaacaactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaa
caaggccattgatggcaacaaggatacctttggcacacattctatggcgccaacggggatccaaagccccctcacacat
acacgattgacatgaagacaactcagaacgtcaacggcttgtctatgctgcctcgacaggatggtaaccaaaacggctgg
atcggtcgccatgaggtttatctaagctcagatggcacaaactggggcagccctgttgcgtcaggtagttggttcgccgac
tctactacaaaatactccaactttgaaactcgccctgctcgctatgttcgtcttgtcgctatcactgaagcgaatggccagcct
tggactagcattgcagagatcaacgtcttccaagctagttcttacacagccccccagcctggtcttggacgctggggtccg
actattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcgggacgagtccttatgtggtcttcatatcgcaat
gatgcatttggaggatcccctggtggtatcactttgacgtcttcctgggatccatccactggtattgtttccgaccgcactgtg
acagtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgtagtcacaggtggcaacgat
gccaagaagaccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgcaagtggctcgtgggtatcag
tcatcagctaccatgtcagacggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaag
tctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagcaaggat
tgtaccgttcagacaaccacgcgtggctctttggatggaagaagggttcggtgttccaagcgggacctagcacagccatg
aactggtactataccagtggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgatgc
catgtgcggaaacgctgtcatgtacgacgccgttaaaggaaagatcctgacctttggcggctccccagattatcaagactc
tgacgccacaaccaacgcccacatcatcaccctcggtgaacccggaacatctcccaacactgtctttgctagcaatgggtt
gtactttgcccgaacgtttcacacctctgttgttcttccagacggaagcacgtttattacaggaggccaacgacgtggaattc
cgttcgaggattcaaccccggtatttacacctgagatctacgtccctgaacaagacactttctacaagcagaaccccaactc
cattgttcgcgtctaccatagcatttccctttgttacctgatggcagggtatttaacggtggtggtggtctttgtggcgattgta
ccacgaatcatttcgacgcgcaaatctttacgccaaactatctttacaatagcaacggcaatctcgcgacacgtcccaagat
taccagaacctctacacagagcgtcaaggtcggtggcagaattacaatctcgacggattcttcgattagcaaggcgtcgtt
gattcgctatggtacagcgacacacacggttaatactgaccagcgccgcattcccctgactctgacaaacaatggaggaa
atagctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatgaactcggccggtgtt
cctagtgtggcttcgacgattcgcgttactcagggcggtggcggttcttggagccatccgcagtttgagaaatgatgagcg
gccgccagc

The *gao* gene has a 4 x glycine + 1 x serine linker and an 8 amino acid *Strep*-tag II at its C-terminal.

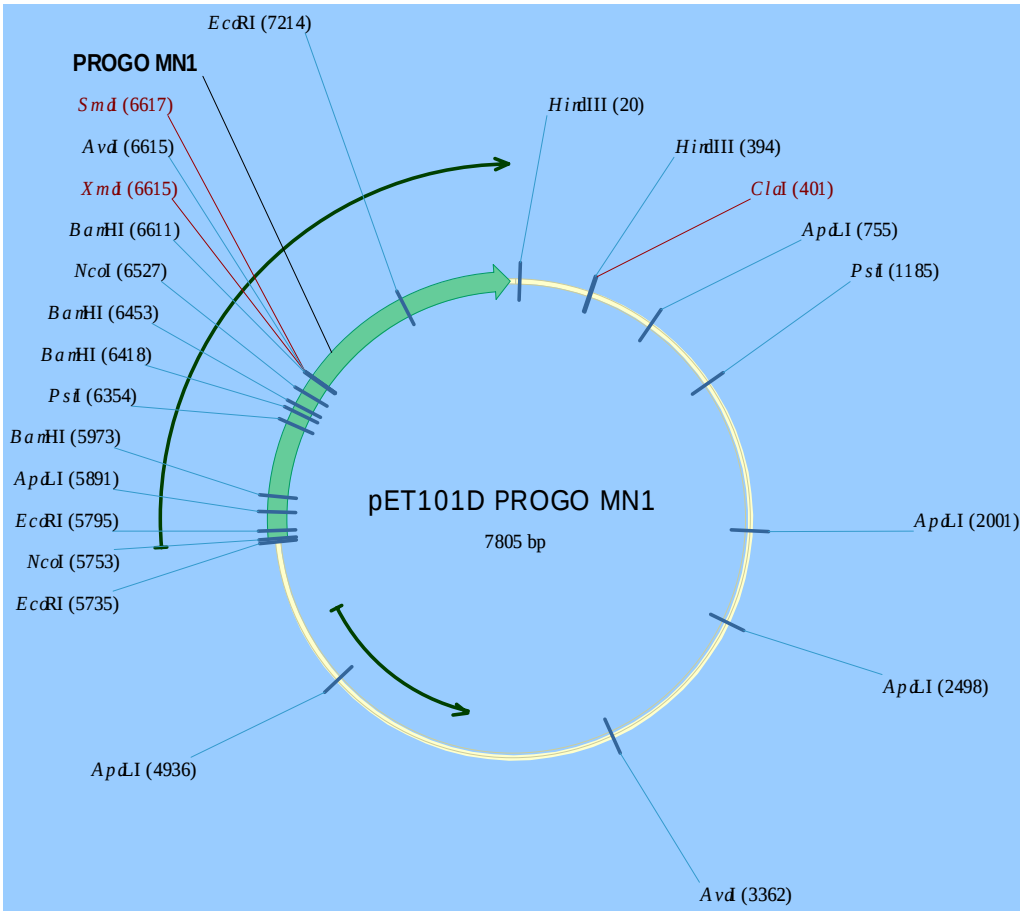


Figure C.2 The map of pET101D ProGOMN1 vector

ProGOMN1

atggttgcagttaccgttcctcacaaggccgtaggaactggaattcctgaagggagtcttcagttcctgagccttcgagcct
cagcacctatcggaagcgccattcctcgcaacaactgggccgtcacttgcgacagtgcacagtcgggaaatgaatgcaa
caaggccattgatggcaacaaggatacctttggcacacattctatggcgccaacggggatccaaagccccctcacacat
acacgattgacatgaagacaactcagaacgtcaacggcttgtctgtgctgcctcgacaggatggtaaccaaaacggctgg
atcggtcgccatgaggtttatctaagctcagatggcacaaactggggcagccctgttgcgtcaggtagttggttcgccgac
tctactacaaaatactccaactttgaaactcgccctgctcgctatgttcgtcttgtcgctatcactgaagcgaatggccagcc
ctggactagcattgcagagatcaacgtcttccaagctagttcttacacagccccccagcctggtcttggacgctggggtcc
gactattgacttaccgattgttcctgcggctgcagcaattgaaccgacatcgggacgagtccttatgtggtcttcatatcgca
atgatgcatttgaaggatcccctggtggtatcactttgacgtcttcctgggatccatccactggtattgtttccgaccgcactgt
gacagtcaccaagcatgatatgttctgccctggtatctccatggatggtaacggtcagatcgtagtcacaggtggcaacga
tgccaagaagaccagtttgtatgattcatctagcgatagctggatcccgggacctgacatgcaagtggctcgtgggtatca
gtcatcagctaccatgtcagacggtcgtgtttttaccattggaggctcctggagcggtggcgtatttgagaagaatggcgaa
gtctatagcccatcttcaaagacatggacgtccctacccaatgccaaggtcaacccaatgttgacggctgacaagcaagg
attgtaccgttcagacaaccacgcgtggctctttggatggaagaagggttcggtgttccaagcgggacctagcacagcca
tgaactggtactataccagtggaagtggtgatgtgaagtcagccggaaaacgccagtctaaccgtggtgtagcccctgat
gccatgtgcggaaacgctgtcatgtacgacgccgttaaaggaaagatcctgacctttggcggctccccagattatcaaga
ctctgacgccacaaccaacgcccacatcatcaccctcggtgaacccggaacatctcccaacactgtctttgctagcaatgg
gttgtactttgcccgaacgtttcacacctctgttgttcttccagacggaagcacgtttattacaggaggccaacgacgtggaa
ttccgttcgaggattcaaccccggtatttacacctgagatctacgtccctgaacaagacactttctacaagcagaaccccaa
ctccattgttcgcgcttaccatagcatttccctttgttacctgatggcagggtatttaacggtggtggtggtctttgtggcgatt
gtaccacgaatcatttcgacgcgcaaatctttacgccaaactatctttacgatagcaacggcaatctcgcgacacgtcccaa
gattaccagaacctctacacagagcgtcaaggtcggtggcagaattacaatctcgacggattcttcgattagcaaggcgtc
gttgattcgctatggtacagcgacacacacggttaatactgaccagcgccgcattcccctgactctgacaaacaatggagg
aaatagctattctttccaagttcctagcgactctggtgttgctttgcctggctactggatgttgttcgtgatgaactcggccggt
gttcctagtgtggcttcgacgattcgcgttactcagggcggtggcggttcttggagccatccgcagtttgagaaatgatgag
cggccgccagc

The *gao* gene has a 4 x glycine + 1 x serine linker and an 8 amino acid *Strep*-tag II at its C-terminal.

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  -
```

GAO and PROGON1 (a.a seguence alignment)

CLUSTAL W (1.82) multiple sequence alignment

```
GAO             MKHLLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 60
PROGON1         --------------MVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 45
                              ****************************************

GAO             SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 120
PROGON1         SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 105
                ************************************************************

GAO             GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 180
PROGON1         GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 165
                ************************************************************

GAO             IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 240
PROGON1         IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 225
                ************************************************************

GAO             GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 300
PROGON1         GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 285
                ************************************************************

GAO             WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 360
PROGON1         WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 345
                ************************************************************

GAO             MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 420
PROGON1         MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 405
                ************************************************************

GAO             DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 480
PROGON1         DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 465
                ************************************************************

GAO             TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 540
PROGON1         TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 525
                ************************************************************

GAO             LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 600
PROGON1         LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 585
                ************************************************************

GAO             ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 660
PROGON1         ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 645
                ************************************************************

GAO             LFVMNSAGVPSVASTIRVTQ------------------ 680
PROGON1         LFVMNSAGVPSVASTIRVTQGGGGSWSHPQFEK--AAAS 682
                ********************
```

GAO and PROGOMN1 (a.a seguence alignment)

CLUSTAL W (1.82) multiple sequence alignment

```
GAO             MKHLLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCD 60
PROGOMN1        ---------------MVAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAIPRNNWAVTCD 45
                               **********************************.*********

GAO             SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQN 120
PROGOMN1        SAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGNQN 105
                ************************************************:*********

GAO             GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 180
PROGOMN1        GWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTS 165
                ************************************************************

GAO             IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFGGSPG 240
PROGOMN1        IAEINVFQASSYTAPQPGLGRWGPTIDLPIVPAAAAIEPTSGRVLMWSSYRNDAFEGSPG 225
                ************************************************** ****

GAO             GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 300
PROGOMN1        GITLTSSWDPSTGIVSDRTVTVTKHDMFCPGISMDGNGQIVVTGGNDAKKTSLYDSSSDS 285
                ************************************************************

GAO             WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 360
PROGOMN1        WIPGPDMQVARGYQSSATMSDGRVFTIGGSWSGGVFEKNGEVYSPSSKTWTSLPNAKVNP 345
                ************************************************************

GAO             MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 420
PROGOMN1        MLTADKQGLYRSDNHAWLFGWKKGSVFQAGPSTAMNWYYTSGSGDVKSAGKRQSNRGVAP 405
                ************************************************************

GAO             DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 480
PROGOMN1        DAMCGNAVMYDAVKGKILTFGGSPDYQDSDATTNAHIITLGEPGTSPNTVFASNGLYFAR 465
                ************************************************************

GAO             TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRVYHSIS 540
PROGOMN1        TFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVRAYHSIS 525
                ************************************************.*****

GAO             LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYNSNGNLATRPKITRTSTQSVKVGGR 600
PROGOMN1        LLLPDGRVFNGGGGLCGDCTTNHFDAQIFTPNYLYDSNGNLATRPKITRTSTQSVKVGGR 585
                **********************************:*************************

GAO             ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 660
PROGOMN1        ITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWM 645
                ************************************************************

GAO             LFVMNSAGVPSVASTIRVTQ------------------ 680
PROGOMN1        LFVMNSAGVPSVASTIRVTQGGGGSWSHPQFEK--AAAS 682
                ********************
```

APPENDIX D

MUTAGENIC PRIMERS

Table D.1. Mutagenic Primers Used in *QuikChange® Site-Directed Mutagenesis Kit*

MUTAGENIC PRIMERS	SEQUENCE (5'-3')	T_m	LENGTH
FGGO_T21A_F	CGTCACTTGCGACAGT**GCA**CAGTCGGGAAATGAATGC	85.6	37 bases
FGGO_T21A_R	GCATTCATTTCCCGACTG**TGC**ACTGTCGCAAGTGACG	85.6	37 bases
FGGO_ R-1P/A1P_F	GAGTCTTCAGTTCCTGAGCCTT**CGAGCC**TCAGCACCTATCGGAAGCG GAGTCTTCAGTTCCTGAGCCTT**CCGCCG**TCAGCACCTATCGGAAGCG	89.3	47 bases
FGGO_R-1P/A1P_R	CGCTTCCGATAGGTGCTGA**CGGCGG**AAGGCTCAGGAACTGAAGACTC	89.3	47 bases
FGGO_R-1X/A1X_F	GAGTCTTCAGTTCCTGAGCCTT**CGAGCC**TCAGCACCTATCGGAAGCG GAGTCTTCAGTTCCTGAGCCTT**NNSNNS**TCAGCACCTATCGGAAGCG	83.3	47 bases
FGGO_R-1X/A1X_R	CGCTTCCGATAGGTGCTGA**SNNSNN**AAGGCTCAGGAACTGAAGACTC	82.4	47 bases
FGGO_S2A_F	CTGAGCCTTCGAGCCGCC**TCA**CCTATCGGAAGCGCC CTGAGCCTTCGAGCCGCC**GCC**CCTATCGGAAGCGCC	72.4	36 bases
FGGO_S2A_R	GGCGCTTCCGATAGGTGCTGC**GGC**TCGAAGGCTCAG	72.4	36 bases
FGGO_H522A_F	GGCGATTGTACCACGAAT**CAT**TTCGACGCGCAAATCTTTACG GGCGATTGTACCACGAAT**GCC**TTCGACGCGCAAATCTTTACG	86.7	42 bases
FGGO_H522A_R	CGTAAAGATTTGCGCGTCGAA**GGC**ATTCGTGGTACAATCGCC	86.7	42 bases

N=A+C+G+T S=G+C

The each first line in the sequence column belongs to the unmutated original *gao.*

APPENDIX E

NUCLEOTIDE AND AMINO ACID SEQUENCE ALIGNMENTS OF MUTATIONAL VARIANTS

ProGON1 MUTANTS

N-RPAP 1 / T7P Primer

```
GAO             ------ATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGTTG
N-RPAP 1        NTNTAGANATAATTTTGTNTAAC-TTTAAGAAGGAATTCAGGAGCCCTTCACCATGGTTG
                      *   **     * ****  *          ***** ***   ***     ****

GAO             CTGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC
N-RPAP 1        CAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC
                * ** ***** **************************************************

GAO             TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA
N-RPAP 1        TGAGCCTTCCGCCGTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA
                *********   * ******************************************

GAO             CTTGCGACAGTGCACAGTCGGGAAATGAATGC---------------------------
N-RPAP 1        CTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAA
                ******************************

GAO             -------------------------------------------AACAAGGCCATTGATG
N-RPAP 1        ATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATG
                                                           ****************

GAO             GCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTC
N-RPAP 1        GCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTC
                ************************************************************
```

N-RPAP 1 and GAO Amino Acid Sequence Alignment

```
GAO               17 VAVTVPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAVGTGIPEGSLQFLSL  SAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAVGTGIPEGSLQFLSL**PP**SAPIGSAISRNNWAVTCDSA
N-RPAP 1          56 gggagccagggagacggacctcaccctgcagagatcaatggatgag
                     tctctcaactgcgtcaggtattgtcccctggctcgaagctcgagc
                     tatcttcgcaatattagttgcgctggaatcaccttcccgcctccta


GAO               63                            -QSGNECNKAIDGNKDTFWHTFY
                                                 QSGNECNKAIDGNKDTFWHTFY
                                 -:H[cac]       AQSGNECNKAIDGNKDTFWHTFY
N-RPAP 1         194 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                       <2-----[196  :  261]-2> cacgaagaactagaaactgacta
                                               aggataccgcttccgtctgcact
```

```
GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-RPAP 1        332 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-RPAP 2 / T7P Primer

```
GAO             --AT-GAAACACCTTTTAACA--CTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGTTGC
N-RPAP 2        CTNTAGAAATAATTTTGTNTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGGTTGC
                   * **** *  ***    *   *          ***** ***   ***     *****

GAO             TGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCT
N-RPAP 2        AGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCT
                 ** ***** **************************************************

GAO             GAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCAC
N-RPAP 2        GAGCCTTCCGCCGTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCAC
                ********   * ********************************************

GAO             TTGCGACAGTGCACAGTCGGGAAATGAATGC----------------------------
N-RPAP 2        TTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAAA
                ****************************

GAO             ----------------------------------------AACAAGGCCATTGATGG
N-RPAP 2        TGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGG
                                                        *****************

GAO             CAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCA
N-RPAP 2        CAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCA
                ************************************************************
```

N-RPAP 2 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSL  SAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSLPPSAPIGSAISRNNWAVTCDSA
N-RPAP 2         56 gggagccagggagacggacctcaccctgcagagatcaatggatgag
                    tctctcaactgcgtcaggtattgtccccctggctcgaagctcgagc
                    tatcttcgcaatattagttgcgctggaatcaccttcccgcctccta


GAO              63                            -QSGNECNKAIDGNKDTFWHTFY
                                                QSGNECNKAIDGNKDTFWHTFY
                                 -:H[cac]      AQSGNECNKAIDGNKDTFWHTFY
N-RPAP 2        194 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[196  :  261]-2> cacgaagaactagaaactgacta
                                            aggataccgcttccgtctgcact


GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-RPAP 2        332 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-RPAP 3 / T7P Primer

```
GAO             -------ATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGTT
N-RPAP 3        CNNNTAGAAANAATNNNGTNTAAC-TTTAAGAAGGAATTCAGGAGCCCCNCACCANGGNT
                       *   **     * ****  *          ***** ***    **     * *

GAO             GCTGTCAC-CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT
N-RPAP 3        GCAGTNACGCGTNCCNCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT
                ** ** ** *** ** ****************************************

GAO             CCTGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT
N-RPAP 3        CCTGAGCCTTCCGCCGTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT
                ***********   * ****************************************

GAO             CACTTGCGACAGTGCACAGTCGGGAAATGAATGC-------------------------
N-RPAP 3        CACTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGG
                *********************************

GAO             ---------------------------------------------AACAAGGCCATTGA
N-RPAP 3        AAATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGA
                                                             **************

GAO             TGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCC
N-RPAP 3        TGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCC
                ************************************************************
```

N-RPAP 3 and GAO Amino Acid Sequence Alignment

```
GAO               21 VPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                     ++HKAVGTGIPEGSLQFLSL  SAPIGSAISRNNWAVTCDSA
                     xxHKAVGTGIPEGSLQFLSLPPSAPIGSAISRNNWAVTCDSA
N-RPAP 3          70 gccagggagacggacctcaccctgcagagatcaatggatgag
                     tcaactgcgtcaggtattgtcccccctggctcgaagctcgagc
                     NNcgcaatattagttgcgctggaatcaccttcccgcctccta


GAO               63                             -QSGNECNKAIDGNKDTFWHTFY
                                                  QSGNECNKAIDGNKDTFWHTFY
                                -:H[cac]         AQSGNECNKAIDGNKDTFWHTFY
N-RPAP 3         196 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                       <2-----[198  :  263]-2> cacgaagaactagaaactgacta
                                               aggataccgcttccgtctgcact


GAO               85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-RPAP 3         334 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                     gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                     cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-RPAP 4 / T7P Primer

```
GAO             ---ATGAAACACCTTT--TAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGTTGC
N-RPAP 4        NTCTNGNCNTNCTTTTGTTTAACTTTAAGAAGGAATTCAGGAGCCCTTCACCATGGTTGC
                     *     * ***  * *   *          ***** ***   ***     *****

GAO             TGTCAC-CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC
N-RPAP 4        AGTTACGCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCC
                 ** ** *** ************************************************
```

```
GAO             TGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA
N-RPAP 4        TGAGCCTTCCGCCGTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCA
                *********   * ******************************************

GAO             CTTGCGACAGTGCACAGTCGGGAAATGAATGC--------------------------
N-RPAP 4        CTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAA
                *******************************

GAO             ------------------------------------------AACAAGGCCATTGATG
N-RPAP 4        ATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATG
                                                           ***************

GAO             GCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTC
N-RPAP 4        GCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTC
                ************************************************************
```

N-RPAP 4 and GAO Amino Acid Sequence Alignment

```
GAO              21 VPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VPHKAVGTGIPEGSLQFLSL  SAPIGSAISRNNWAVTCDSA
                    VPHKAVGTGIPEGSLQFLSLPPSAPIGSAISRNNWAVTCDSA
N-RPAP 4         69 gccagggagacggacctcaccctgcagagatcaatggatgag
                    tcaactgcgtcaggtattgtccccctggctcgaagctcgagc
                    ttcgcaatattagttgcgctggaatcaccttcccgcctccta


GAO              63                            -QSGNECNKAIDGNKDTFWHTFY
                                                QSGNECNKAIDGNKDTFWHTFY
                                 -:H[cac]      AQSGNECNKAIDGNKDTFWHTFY
N-RPAP 4        195 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[197  :  262]-2> cacgaagaactagaaactgacta
                                              aggataccgcttccgtctgcact


GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-RPAP 4        333 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-S2A1 / T7P Primer

```
GAO          ATGTCAACATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGT 58
N-S2A 1      ----------ATAATTTTGTNTAAC-TTTAAGAAGGAATTCAGGAGCCCTTCACCATGGT 49
                       **     * ****  *          ***** ***   ***     **

GAO           GCTGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT 118
N-S2A 1      TGCAGTTACCGTTCCTCACAAGGCCGTAG-AACTGGAATTCCTGAAGGGAGTCTTCAGTT 108
             *** ** ***** *************** ****************************

GAO          CCTGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT 178
N-S2A 1      CCTGAGCCTTCGAGCCGCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT 168
                 *************** ****************************************

GAO          CACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGA 238
N-S2A 1      CACTTGCGACTGTGCACTGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGA 228
             ********** ****** ****************************************

GAO          TACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACAC 298
N-S2A 1      TACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACAC 288
             ************************************************************
```

```
GAO          GATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGG 358
N-S2A 1      GATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGG 348
             ************************************************************
```

N-S2A 1 and GAO Amino Acid Sequence Alignment

```
GAO               17 VAVTVPHKAVGTGIPEGSLQFLSLRA**S**APIGSAISRNNWAVTCDSAQSG
                     VAVTVPHKAV TGIPEGSLQFLSLRA+APIGSAISRNNWAVTCD A SG
                     VAVTVPHKAV!TGIPEGSLQFLSLRA**A**APIGSAISRNNWAVTCDCALSG
N-S2A 1           48 gggagccagg2agacggacctcaccgggcagagatcaatggatgtgctg
                     tctctcaact cgtcaggtattgtgcccctggctcgaagctcgagctcg
                     tatcttcgca tattagttgcgctacaatcaccttcccgcctcctagga


GAO               66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                     NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                     NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
N-S2A 1          194 agtaagaggaagattcattggaggcacccataagaaaacagagttaccc
                     aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                     taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta


GAO              115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
N-S2A 1          341 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                     aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                     gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

N-S2A 2 /T7P Primer

```
GAO              ATGTCAACATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGT 58
N-S2A 2          ---------AATAATTTTGTNTAAC-TTTAAGAAGGAATTCAGGAGCCCTTCACCATGGT 50
                           **     * ****  *          ***** ***   ***     **

GAO              TGCTGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT 118
N-S2A 2          TGCAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT 110
                 *** ** ***** ********************************************

GAO              CCTGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT 178
N-S2A 2          CCTGAGCCTTCGAGCCGCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGT 170
                 *************** ****************************************

GAO              CACTTGCGACAGTGCACAGTCGGGAAATGAATGC------------------------- 212
N-S2A 2          CACTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGG 230
                 *********************************

GAO              ---------------------------------------------AACAAGGCCATTGA 226
N-S2A 2          AAATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGA 290
                                                               **************

GAO              TGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCC 286
N-S2A 2          TGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCC 350
                 ************************************************************
```

N-S2A 2 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSLRA+APIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSLRAAAPIGSAISRNNWAVTCDSA
N- S2A 2         49 gggagccagggagacggacctcaccgggcagagatcaatggatgag
                    tctctcaactgcgtcaggtattgtgcccctggctcgaagctcgagc
                    tatcttcgcaatattagttgcgctacaatcaccttcccgcctccta


GAO              63                            -QSGNECNKAIDGNKDTFWHTFY
                                                QSGNECNKAIDGNKDTFWHTFY
                                 -:H[cac]      AQSGNECNKAIDGNKDTFWHTFY
N-S2A 2         187 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[189  :  254]-2> cacgaagaactagaaactgacta
                                              aggataccgcttccgtctgcact


GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-S2A 2         325 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-S2A 3 /T7P Primer

```
GAO              ----------------------------------------------------ATGTCAA 7
N-S2A 3          TCCTCACAAGGCCGTAGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCGGAGCCTTCGA 120
                                                                      ** *

GAO              CATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCATCAATGCTGTTGCTGTCAC 67
N-S2A 3          GCCGCAGCACCTATCGGAAGCGCCCCTGAGCCTTCG-AGCCGCAGCACC--TATCGGAAG 177
                    * * ***** *    *    * **  ** *  * ***  **   *   *   *  *

GAO              CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAG-------TCTT-CAGTTC 119
N-S2A 3          CGCCCCTGA----GCCTTCGGAGCCGCAGCACCTATCGGAAGCGCCATTTCTCGCAACAA 233
                 ** **** *    *** * *** * * *   ***   ** **       ***  **

GAO              CTGAGCCTTCGA--GCCTCAGCACCTA-TCGGAA-------GCGCCATTTCTCGCA---- 165
N-S2A 3          CTGGGCCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGC 293
                 *** *** **    **  ***  *  * **** *       *** ** **    **

GAO              ACAACTGGGC--------CGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAA 217
N-S2A 3          ACAGTCGGGAAATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAA 353
                 ***   ***         *************** ************************

GAO              GGCCATTGATGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCC 277
N-S2A 3          GGCCATTGATGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCC 413
                 ************************************************************
```

N-S2A 3 and GAO Amino Acid Sequence Alignment

```
GAO              42 ASAPIGSAISRNNWAVTCDSA                          -Q
                    A+APIGSAISRNNWAVTCDSA                           Q
                    AAAPIGSAISRNNWAVTCDSA           -:H[cac]       AQ
N-S2A 3         196 gggcagagatcaatggatgagCAGTCGGGA  Intron 1   CAGCgc
                    cccctggctcgaagctcgagc  <2-----[261  :  326]-2> ca
                    caatcaccttcccgcctccta                          ag
```

```
GAO              64 SGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSML
                    SGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSML
                    SGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSML
N-S2A 3         334 tgagtaagaggaagattcattggaggcacccataagaaaacagagttac
                    cgaagaactagaaactgactagcagacaccacactataccaatagtctt
                    gataccgcttccgtctgcactcccgtagctcacgtcggatgccccgtgg


GAO             113 PRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRP
                    PRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWF DSTTKYSNFETRP
                    PRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWF!DSTTKYSNFETRP
N-S2A 3         481 cccggacagtagccggtcatggaatgacggtgatt4gtaaattatgacc
                    cgaagaaaggtggaatatgcagcagggctccgggt acccaacatacgc
                    tagttcaccgctctgttacatcacgccttgattgc cttaaccctatct
```

N-S2A 4 /T7P Primer

```
GAO             ATGTCAACATGAAACACCTTTTAACACTCGCTCTTTGCTTCAGCAGCA--TCAATGCTGT 58
N-S2A 4         --------AAATAATTTTGTNTAAC-TTTAAGAAGGAATTCAGGAGCCCTTCACCATGGT 51
                        *   **     * ****  *          ***** ***   ***     **

GAO             TGCTGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT 118
N-S2A 4         TGCAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTT 111
                *** ** ***** **********************************************

GAO             CCTGAGCCTT-CGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCG 177
N-S2A 4         CCTGAGCCTTTCGAGCCGCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCG 171
                ********** ****** ****************************************

GAO             TCACTTGCGACAGTGCACAGTCGGGAAATGAATGC------------------------ 212
N-S2A 4         TCACTTGCGACAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGG 231
                **********************************

GAO             ----------------------------------------------AACAAGGCCATTG 225
N-S2A 4         GAAATGAATGCGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTG 291
                                                              *************

GAO             ATGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCC 285
N-S2A 4         ATGGCAACAAGGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCC 351
                ************************************************************
```

N-S2A 4 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLS RA+APIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLS!RAAAPIGSAISRNNWAVTCDSA
N-S2A 4          50 gggagccagggagacggacctca4cgggcagagatcaatggatgag
                    tctctcaactgcgtcaggtattg gcccctggctcgaagctcgagc
                    tatcttcgcaatattagttgcgc acaatcaccttcccgcctccta


GAO              63                            -QSGNECNKAIDGNKDTFWHTFY
                                                QSGNECNKAIDGNKDTFWHTFY
                                 -:H[cac]       AQSGNECNKAIDGNKDTFWHTFY
N-S2A 4         189 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[191  :  256]-2> cacgaagaactagaaactgacta
                                             aggataccgcttccgtctgcact
```

```
GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-S2A 4         327 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-H522A 1/T7P Primer

```
GAO             CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 74
N-H522A 1       CGTTCCTCACAAGGCCGTAG-AACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 119
                *** *************** ************************************

GAO             TCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGA 134
N-H522A 1       TCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGA 179
                ************************************************************

GAO             CAGTGCACAGTCGGGAAATGAATGC---------------------------------- 159
N-H522A 1       CAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATG 239
                *************************

GAO             -----------------------------------AACAAGGCCATTGATGGCAACAA 182
N-H522A 1       CGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAA 299
                                                   **********************

GAO             GGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATA 242
N-H522A 1       GGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATA 359
                ************************************************************

GAO             CACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGA 302
N-H522A 1       CACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGA 419
                ************************************************************
```

N-H522A 1 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAV TGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAV!TGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
N-H522A 1        50 gggagccagg2agacggacctcaccgtgcagagatcaatggatgag
                    tctctcaact cgtcaggtattgtgcccctggctcgaagctcgagc
                    tatcttcgca tattagttgcgctacaatcaccttcccgcctccta


GAO              63                              -QSGNECNKAIDGNKDTFWHTFY
                                                  QSGNECNKAIDGNKDTFWHTFY
                                -:H[cac]         AQSGNECNKAIDGNKDTFWHTFY
N-H522A 1       187 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[189  :  254]-2> cacgaagaactagaaactgacta
                                              aggataccgcttccgtctgcact


GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-H522A 1       325 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-H522A 1/GOFORB Primer

```
GAO             TGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACA
N-H522A 1       TGGAATTCCGTTCGAGGATTCAACCCCGGTATTTACACCTGAGATCTACGTCCCTGAACA
                ************************************************************

GAO             AGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTCTACCATAGCATTTCCCT
N-H522A 1       AGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTCTACCATAGCATTTCCCT
                ************************************************************

GAO             TTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCAC
N-H522A 1       TTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACCAC
                ************************************************************

GAO             GAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAATCT
N-H522A 1       GAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAATCT
                ****   ****************************************************

GAO             CGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAAT
N-H522A 1       CGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGAAT
                ************************************************************

GAO             TACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGAC
N-H522A 1       TACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCGAC
                ************************************************************
```

N-H522A 1 and GAO Amino Acid Sequence Alignment

```
GAO               477 YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                      YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                      YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
N-H522A 1         148 ttgcatcatggccggaataaggcccgactggtacgtacgatgcgcgatt
                      atcgctacctttcaggcttcggagggtctaacccttccatatcaaacta
                      ctcagtccttttacacgttaacaatatgcgtacgatatgccctaactcc

GAO               526 KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYL
                      KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYL
                      KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYL
N-H522A 1         295 acacatagcgtcaatcttcggagtaggggctggtaaagtggcatacatc
                      aaacacttgtaagtctttcaggttaggggtggagccactacattccaat
                      ggccccttccctctctgattcgatctttttcttcgtcccgactgactt

GAO               575 YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                      YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                      YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
N-H522A 1         442 taaagacgaccaaaaatacagagggaaaatagttaaagttactgagaca
                      aagagatccgcatcgcccagtatgggtctccacctgaccttgagcccac
                      ctccctcgatcgtcactagccgctcatacggttgtcggggtcttagacg
```

N-H522A 2/T7P Primer

```
GAO             CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 74
N-H522A 2       CGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 120
                *** ********************************************************

GAO             TCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGA 134
N-H522A 2       TCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGA 180
                ************************************************************
```

```
GAO             CAGTGCACAGTCGGGAAATGAATGC---------------------------------- 159
N-H522A 2       CAGTGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATG 240
                ************************

GAO             -----------------------------------AACAAGGCCATTGATGGCAACAA 182
N-H522A 2       CGTCACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAA 300
                                                   **********************

GAO             GGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATA 242
N-H522A 2       GGATACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATA 360
                ************************************************************

GAO             CACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGA 302
N-H522A 2       CACGATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGA 420
                ************************************************************
```

N-H522A 2 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                    VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
N-H5222A 2       50 gggagccagggagacggacctcaccgtgcagagatcaatggatgag
                    tctctcaactgcgtcaggtattgtgcccctggctcgaagctcgagc
                    tatcttcgcaatattagttgcgctacaatcaccttcccgcctccta


GAO              63                              -QSGNECNKAIDGNKDTFWHTFY
                                                  QSGNECNKAIDGNKDTFWHTFY
                                  -:H[cac]       AQSGNECNKAIDGNKDTFWHTFY
N-H522A 2       188 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                      <2-----[190  :  255]-2> cacgaagaactagaaactgacta
                                             aggataccgcttccgtctgcact


GAO              85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                    GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-H522A 2       326 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                    gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                    cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-H522A 2/GOFORB Primer

N-H522A 2 and GAO Amino Acid Sequence Alignment

```
GAO             477 YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                    YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                    YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
N-H522A2        148 ttgcatcatggccggaataaggcccgactggtacgtacgatgcgcgatt
                    atcgctacctttcaggcttcggagggtctaacccttccatatcaaacta
                    ctcagtccttttacacgttaacaatatgcgtacgatatgccctaactcc


GAO             526 KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYL
                    KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYL
                    KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYL
N-H522A2        295 acacatagcgtcaatcttcggagtaggggctggtaaagtggcatacatc
                    aaacacttgtaagtctttcaggttaggggtggagccactacattccaat
                    ggccccttccctctctgattcgatctttttcttcgtcccgactgactt
```

```
GAO              575 YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                     YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                     YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
N-H522A2         442 taaagacgaccaaaaatacagagggaaaatagttaaagttactgagaca
                     aagagatccgcatcgcccagtatgggtctccacctgaccttgagcccac
                     ctccctcgatcgtcactagccgctcatacggttgtcggggtcttagacg
```

N-H522A 3 / T7P Primer

```
GAO              CCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCTTCG 77
N-H522A 3        TCCTCACAAGGCCGTAGANACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCTTCG 120
                  ***************  ****************************************

GAO              AGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGACAG 137
N-H522A 3        AGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGACAG 180
                 ************************************************************

GAO              TGCACAGTCGGGAAATGAATGC------------------------------------ 159
N-H522A 3        TGCACAGTCGGGAAATGAATGCGTCACTTGCGACAGTGCACAGTCGGGAAATGAATGCGT 240
                 **********************

GAO              ---------------------------------AACAAGGCCATTGATGGCAACAAGGA 185
N-H522A 3        CACTTGCGACAGCGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGA 300
                                                   *************************

GAO              TACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACAC 245
N-H522A 3        TACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACAC 360
                 ************************************************************

GAO              GATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGG 305
N-H522A 3        GATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGG 420
                 ************************************************************
```

N-H522A 3 and GAO Amino Acid Sequence Alignment

```
GAO               17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAV TGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAVxTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
N-H522A 3         47 gggagccagggagacggacctcaccgtgcagagatcaatggatgag
                     tctctcaactacgtcaggtattgtgcccctggctcgaagctcgagc
                     tatcttcgcaNtattagttgcgctacaatcaccttcccgcctccta


GAO               63                              -QSGNECNKAIDGNKDTFWHTFY
                                                   QSGNECNKAIDGNKDTFWHTFY
                                -:H[cac]          AQSGNECNKAIDGNKDTFWHTFY
N-H522A 3        185 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                       <2-----[187  :  252]-2> cacgaagaactagaaactgacta
                                               aggataccgcttccgtctgcact


GAO               85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIG HEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIG!HEVYLSSD
N-H522A 3        323 ggaggcacccataagaaaacagagttaccccggacagtag4cggtcatg
                     gcagacaccacactataccaatagtcttcgaagaaaggtg aatatgca
                     cccgtagctcacgtcggatgccccgtggtagttcaccgct tgttacat
```

N-H522A 3/GOFORB Primer

```
GAO             ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 1560
N-H522A 3       ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 322
                ************************************************************

GAO             TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 1620
N-H522A 3       TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 382
                ************************************************************

GAO             TGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 1680
N-H522A 3       TGTGGCGATTGTACCACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 442
                ********************   **********************************

GAO             AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 1740
N-H522A 3       AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 502
                ************************************************************

GAO             AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 1800
N-H522A 3       AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 562
                ************************************************************

GAO             CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 1860
N-H522A 3       CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 622
                ************************************************************
```

N-H522A 3 and GAO Amino Acid Sequence Alignment

```
GAO              485 SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
                     SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
                     SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
N-H522A 3        170 tggccggaataaggcccgactggtacgtacgatgcgcgattacacatag
                     ctttcaggcttcggagggtctaacccttccatatcaaactaaaacactt
                     ttttacacgttaacaatatgcgtacgatatgccctaactccggcccctt

GAO              534 RVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYLYNSNGNLA
                     RVYHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYLYNSNGNLA
                     RVYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYLYNSNGNLA
N-H522A 3        317 cgtcaatcttcggagtaggggctggtaaagtggcatacatctaaagacg
                     gtaagtctttcaggttaggggtggagccactacattccaataagagatc
                     ccctctctgattcgatctttttcttcgtcccgactgacttctccctcg

GAO              583 TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
                     TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
                     TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
N-H522A 3        464 accaaaaatacagagggaaaatagttaaagttactgagacagaagccca
                     cgcatcgcccagtatgggtctccacctgaccttgagcccactacaaggt
                     atcgtcactagccgctcatacggttgtcggggtcttagacgtttcgcct
```

N-H522A 4 / T7P Primer

```
GAO             ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 1560
N-H522A 3       ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 322
                ************************************************************

GAO             TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 1620
N-H522A 3       TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 382
                ************************************************************
```

```
GAO             TGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 1680
N-H522A 3       TGTGGCGATTGTACCACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 442
                ********************   **********************************

GAO             AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 1740
N-H522A 3       AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 502
                ************************************************************

GAO             AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 1800
N-H522A 3       AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 562
                ************************************************************

GAO             CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 1860
N-H522A 3       CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 622
                ************************************************************
```

N-H522A 4 and GAO Amino Acid Sequence Alignment

```
GAO               17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAV TGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
                     VAVTVPHKAV!TGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSA
N-H522A 4         50 gggagccagg2agacggacctcaccgtgcagagatcaatggatgag
                     tctctcaact cgtcaggtattgtgcccctggctcgaagctcgagc
                     tatcttcgca tattagttgcgctacaatcaccttcccgcctccta


GAO               63                              -QSGNECNKAIDGNKDTFWHTFY
                                                   QSGNECNKAIDGNKDTFWHTFY
                                -:H[cac]          AQSGNECNKAIDGNKDTFWHTFY
N-H522A 4        187 CAGTCGGGA  Intron 1   CAGCgctgagtaagaggaagattcatt
                       <2-----[189  :  254]-2> cacgaagaactagaaactgacta
                                               aggataccgcttccgtctgcact


GAO               85 GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
                     GANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSD
N-H522A 4        325 ggaggcacccataagaaaacagagttaccccggacagtagccggtcatg
                     gcagacaccacactataccaatagtcttcgaagaaaggtggaatatgca
                     cccgtagctcacgtcggatgccccgtggtagttcaccgctctgttacat
```

N-H522A 4/GOFORB Primer

```
GAO             ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 1560
N-H522A 4       ATCTACGTCCCTGAACAAGACACTTTCTACAAGCAGAACCCCAACTCCATTGTTCGCGTC 325
                ************************************************************

GAO             TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 1620
N-H522A 4       TACCATAGCATTTCCCTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTT 385
                ************************************************************

GAO             TGTGGCGATTGTACCACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 1680
N-H522A 4       TGTGGCGATTGTACTACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTAC 445
                ************** ******   **********************************

GAO             AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 1740
N-H522A 4       AATAGCAACGGCAATCTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTC 505
                ************************************************************

GAO             AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 1800
N-H522A 4       AAGGTCGGTGGCAGAATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATT 565
                ************************************************************
```

```
GAO             CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 1860
N-H522A 4       CGCTATGGTACAGCGACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTG 625
                ************************************************************
```

N-H522A 4 and GAO Amino Acid Sequence Alignment

```
GAO              477 YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                     YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                     YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
N-H522A 4        149 ttgcatcatggccggaataaggcccgactggtacgtacgatgcgcgatt
                     atcgctacctttcaggcttcggagggtctaacccttccatatcaaacta
                     ctcagtccttttacacgttaacaatatgcgtacgatatgccctaactcc

GAO              526 KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYL
                     KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYL
                     KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYL
N-H522A 4        296 acacatagcgtcaatcttcggagtaggggctggtaaagtggcatacatc
                     aaacacttgtaagtctttcaggttaggggtggagccactacattccaat
                     ggccccttccctctctgattcgatctttttctttgtcccgactgactt

GAO              575 YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                     YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                     YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
N-H522A 4        443 taaagacgaccaaaaatacagagggaaaatagttaaagttactgagaca
                     aagagatccgcatcgcccagtatgggtctccacctgaccttgagcccac
                     ctccctcgatcgtcactagccgctcatacggttgtcggggtcttagacg
```

ProGOMN1 MUTANTS

MN-RPAP 1/T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RPAP 1       CCGTTCCTCACAAGGCCGTAG-AACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 116
                **** ********************************************************

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RPAP 1       TTCCGCCGTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 176
                ***   * ************************ ************************** 

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RPAP 1       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 236
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RPAP 1       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 296
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RPAP 1       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 356
                ******************************** ************************** 

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RPAP 1       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 416
                ************************************************************
```

MN-RPAP 1 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSAQSG
                    VAVTVPHKAV TGIPEGSLQFLSL  SAPIGSAI RNNWAVTCDSAQSG
                    VAVTVPHKAV!TGIPEGSLQFLSL**PP**SAPIGSAIPRNNWAVTCDSAQSG
MN-RPAP 1        48 gggagccagg2agacggacctcaccctgcagagaccaatggatgagctg
                    tctctcaact cgtcaggtattgtccccctggctcgaagctcgagcacg
                    tatcttcgca tattagttgcgctggaatcaccttcccgcctcctagga


GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-RPAP 1       194 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                    aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                    taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta


GAO             115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
MN-RPAP 1       341 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                    aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                    gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

MN-RPAP 2 /T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RPAP 2       CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 116
                **** ********************************************************
```

```
GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RPAP 2       TTCCGCCGTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 176
                ***   * ********************** **************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RPAP 2       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 236
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RPAP 2       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 296
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RPAP 2       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 356
                ****************************** **************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RPAP 2       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 416
                ************************************************************
```

MN-RPAP 2 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL  SAPIGSAI RNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL**PP**SAPIGSAIPRNNWAVTCDSAQSG
MN-RPAP 2        47 gggagccagggagacggacctcaccctgcagagaccaatggatgagctg
                    tctctcaactgcgtcaggtattgtccccctggctcgaagctcgagcacg
                    tatcttcgcaatattagttgcgctggaatcaccttcccgcctcctagga

GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-RPAP 2       194 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                    aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                    taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta

GAO             115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
MN-RPAP 2       341 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                    aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                    gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

MN-RPAP 2 / GOFORB Primer

```
GAO             CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 1680
MN-RPAP 2       CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 400
                ************************************************************

GAO             ACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAAT 1740
MN-RPAP 2       ACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAAT 460
                ******   ********************************** **************

GAO             CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 1800
MN-RPAP 2       CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 520
                ************************************************************

GAO             ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 1860
MN-RPAP 2       ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 580
                ************************************************************
```

```
GAO             ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 1920
MN-RPAP 2       ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 640
                ************************************************************

GAO             AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 1980
MN-RPAP 2       AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 700
                ************************************************************
```

MN-RPAP 2 and GAO Amino Acid Sequence Alignment

```
GAO               477 YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                      YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
                      YFARTFHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFY
MN-RPAP 2         149 ttgcatcatggccggaataaggcccgactggtacgtacgatgcgcgatt
                      atcgctacctttcaggcttcggagggtctaacccttccatatcaaacta
                      ctcagtccttttacacgttaacaatatgcgtacgatatgccctaactcc


GAO               526 KQNPNSIVRVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYL
                      KQNPNSIVR YHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYL
                      KQNPNSIVRAYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYL
MN-RPAP 2         296 acacatagcgtcaatcttcggagtaggggctggtaaagtggcatacatc
                      aaacacttgcaagtctttcaggttaggggtggagccactacattccaat
                      ggccccttctctctctgattcgatctttttcttcgtcccgactgactt


GAO               575 YNSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                      Y+SNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
                      YDSNGNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHT
MN-RPAP 2         443 tgaagacgaccaaaaatacagagggaaaatagttaaagttactgagaca
                      aagagatccgcatcgcccagtatgggtctccacctgaccttgagcccac
                      ctccctcgatcgtcactagccgctcatacggttgtcggggtcttagacg
```

MN-RPAP 3 / T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RPAP 3       C-GTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 115
                * ** ************************************************************

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RPAP 3       TTCCGCCGTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 175
                ***   * ********************** ***************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RPAP 3       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 235
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RPAP 3       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 295
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RPAP 3       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 355
                ******************************* ***************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RPAP 3       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 415
                ************************************************************
```

MN-RPAP 3 and GAO Amino Acid Sequence Alignment

```
GAO              21 VPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSAQSGNECN
                    VPHKAVGTGIPEGSLQFLSL  SAPIGSAI RNNWAVTCDSAQSGNECN
                    VPHKAVGTGIPEGSLQFLSL**PP**SAPIGSAIPRNNWAVTCDSAQSGNECN
MN-RPAP 3        58 gccagggagacggacctcaccctgcagagaccaatggatgagctgagta
                    tcaactgcgtcaggtattgtccccctggctcgaagctcgagcacgaaga
                    ttcgcaatattagttgcgctggaatcaccttcccgcctcctaggatacc

GAO              70 KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGN
MN-RPAP 3       205 agaggaagattcattggaggcacccataagaaaacagagttgccccgga
                    actagaaactgactagcagacaccacactataccaatagtcttcgaaga
                    gcttccgtctgcactcccgtagctcacgtcggatgccccgtggtagttc

GAO             119 QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
MN-RPAP 3       352 cagtagccggtcatggaatgacggtgattggtaaattatgaccgctgcc
                    aaggtggaatatgcagcagggctccgggtcacccaacatacgccgatgt
                    accgctctgttacatcacgccttgattgcccttaaccctatcttctttt
```

MN-RPAP 4 / T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RPAP 4       CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 116
                **** ************************************************

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RPAP 4       TTCCGCCGTCAGCACCTATCGGAAACGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 176
                ***   * *************** ******* ************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RPAP 4       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 236
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RPAP 4       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 296
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RPAP 4       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 356
                ******************************** ************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RPAP 4       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 416
                ************************************************************
```

MN-RPAP 4 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSL**RA**SAPIG**S**AISRNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL  SAPIG+AI RNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL**PP**SAPIG**N**AIPRNNWAVTCDSAQSG
MN-RPAP 4        47 gggagccagggagacggacctcaccctgcagagaccaatggatgagctg
                    tctctcaactgcgtcaggtattgtcccctgactcgaagctcgagcacg
                    tatcttcgcaatattagttgcgctggaatcaccttcccgcctcctagga
```

```
GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-RPAP 4       194 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                    aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                    taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta


GAO             115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
MN-RPAP 4       341 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                    aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                    gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

MN-RXAX 1/ T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RXAX 1       CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 116
                **** ************************************************************

GAO             TT**CGA**GCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RXAX 1       TT**TT-**GTGTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 175
                **   *  ********************** *************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RXAX 1       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 235
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RXAX 1       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 295
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RXAX 1       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 355
                ******************************** ***************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RXAX 1       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 415
                ************************************************************
```

MN-RXAX 1 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL  SAPIGSAI RNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSL!VSAPIGSAIPRNNWAVTCDSAQSG
MN-RXAX 1        47 gggagccagggagacggacctcac2gtgcagagaccaatggatgagctg
                    tctctcaactgcgtcaggtattgt tccctggctcgaagctcgagcacg
                    tatcttcgcaatattagttgcgct gaatcaccttcccgcctcctagga


GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-RXAX 1       193 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                    aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                    taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta
```

```
GAO              115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
MN-RXAX 1        340 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                     aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                     gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

MN-RXAX 2/ T7P Primer

```
GAO              CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RXAX 2        CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 116
                 **** ************************************************** 

GAO              TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RXAX 2        TT-AACCGCCAGCAC-TATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 174
                 **  * *  ****** *************** ************************** 

GAO              ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RXAX 2        ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 234
                 ************************************************************ 

GAO              GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RXAX 2        GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 294
                 ************************************************************ 

GAO              TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RXAX 2        TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 354
                 ******************************** ************************** 

GAO              ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RXAX 2        ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 414
                 ************************************************************ 
```

MN-RXAX 2 and GAO Amino Acid Sequence Alignment

```
GAO               17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSAQSG
                     VAVTVPHKAVGTGIPEGSLQFLSL  ++ IGSAI RNNWAVTCDSAQSG
                     VAVTVPHKAVGTGIPEGSLQFLSLN!ASTIGSAIPRNNWAVTCDSAQSG
MN-RXAX 2         47 gggagccagggagacggacctcaca1gaaagagaccaatggatgagctg
                     tctctcaactgcgtcaggtattgta cgctggctcgaagctcgagcacg
                     tatcttcgcaatattagttgcgctc cctcaccttcccgcctcctagga


GAO               66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                     NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                     NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-RXAX 2        192 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                     aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                     taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta


GAO              115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
                     QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPAR
MN-RXAX 2        339 cggacagtagccggtcatggaatgacggtgattggtaaattatgaccgc
                     aagaaaggtggaatatgcagcagggctccgggtcacccaacatacgccg
                     gttcaccgctctgttacatcacgccttgattgcccttaaccctatcttc
```

MN-RXAX 3/ T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RXAX 3       C-GTTCCTCACAAGGCCGTAGNCACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 115
                * ** ****************  ************************************

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RXAX 3       TTTGCAGCTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 175
                ** *   ************************ *************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RXAX 3       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 235
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RXAX 3       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 295
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RXAX 3       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 355
                ******************************* *************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RXAX 3       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 415
                ************************************************************
```

MN-RXAX 3 and GAO Amino Acid Sequence Alignment

```
GAO              21 VPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSAQSGNECN
                    VPHKAV+TGIPEGSLQFLSL +SAPIGSAI RNNWAVTCDSAQSGNECN
                    VPHKAVxTGIPEGSLQFLSL**CS**SAPIGSAIPRNNWAVTCDSAQSGNECN
MN-RXAX 3        58 gccagggagacggacctcactatgcagagaccaatggatgagctgagta
                    tcaactNcgtcaggtattgtggccctggctcgaagctcgagcacgaaga
                    ttcgcactattagttgcgctccaatcaccttcccgcctcctaggatacc

GAO              70 KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGN
MN-RXAX 3       205 agaggaagattcattggaggcacccataagaaaacagagttgccccgga
                    actagaaactgactagcagacaccacactataccaatagtcttcgaaga
                    gcttccgtctgcactcccgtagctcacgtcggatgccccgtggtagttc

GAO             119 QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
MN-RXAX 3       352 cagtagccggtcatggaatgacggtgattggtaaattatgaccgctgcc
                    aaggtggaatatgcagcagggctccgggtcacccaacatacgccgatgt
                    accgctctgttacatcacgccttgattgcccttaaccctatcttctttt
```

MN-RXAX 4/ T7P Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-RXAX 4       C-GTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCT 112
                * ** ************************************************** *

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-RXAX 4       TTGGGGGCTCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 172
                ** * * ************************ *************************
```

```
GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-RXAX 4       ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 232
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 298
MN-RXAX 4       GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 292
                ************************************************************

GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 358
MN-RXAX 4       TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 352
                ******************************** **************************

GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-RXAX 4       ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 412
                ************************************************************
```

MN-RXAX 4 and GAO Amino Acid Sequence Alignment

```
GAO              21 VPHKAVGTGIPEGSLQFLSL**RA**SAPIGSAISRNNWAVTCDSAQSGNECN
                    VPHKAVGTGIPEGSLQFLS   SAPIGSAI RNNWAVTCDSAQSGNECN
                    VPHKAVGTGIPEGSLQFLSF**GG**SAPIGSAIPRNNWAVTCDSAQSGNECN
MN-RXAX 4        55 gccagggagacggacctcatggtgcagagaccaatggatgagctgagta
                    tcaactgcgtcaggtattgtggccctggctcgaagctcgagcacgaaga
                    ttcgcaatattagttgcgctgcaatcaccttcccgcctcctaggatacc

GAO              70 KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPRQDGN
                    KAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQDGN
MN-RXAX 4       202 agaggaagattcattggaggcacccataagaaaacagagttgccccgga
                    actagaaactgactagcagacaccacactataccaatagtcttcgaaga
                    gcttccgtctgcactcccgtagctcacgtcggatgccccgtggtagttc

GAO             119 QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
                    QNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRL
MN-RXAX 4       349 cagtagccggtcatggaatgacggtgattggtaaattatgaccgctgcc
                    aaggtggaatatgcagcagggctccgggtcacccaacatacgccgatgt
                    accgctctgttacatcacgccttgattgcccttaaccctatcttctttt
```

MN-S2A 1/GAOstr Primer

```
GAO             CGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 119
MN-S2A 1        CGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCCT 119
                *** ********************************************************

GAO             TCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCGA 179
MN-S2A 1        TCGAGCCGCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCGA 179
                ******* ********************** ****************************

GAO             CAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTTG 239
MN-S2A 1        CAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTTG 239
                ************************************************************

GAO             GCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACAT 299
MN-S2A 1        GCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACAT 299
                ************************************************************

GAO             GAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAAA 359
MN-S2A 1        GAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAAA 359
                ******************************** ***************************
```

```
GAO             CGGCTGGATCGGTCGCC-ATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 418
MN-S2A 1        CGGCTGGATCGGTCGCCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 419
                *************** ****************************************
```

MN-S2A 1 and GAO Amino Acid Sequence Alignment

```
GAO               4 LLTLALCFSSINAVAVTVPHKAVGTGIPEGSLQFLSLRA**S**APIGSAISR
                    L TL      +  VAVTVPHKAVGTGIPEGSLQFLSLRA+APIGSAI R
                    LFTLRRIxEPFTMVAVTVPHKAVGTGIPEGSLQFLSLRA**A**APIGSAIPR
MN-S2A 1         10 ttataaaNgctaagggagccagggagacggacctcaccgggcagagacc
                    ttctggtcactcttctctcaactgcgtcaggtattgtgcccctggctcg
                    gttaagtcgcccgtatcttcgcaatattagttgcgctacaatcaccttc

GAO              53 NNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
                    NNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
                    NNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
MN-S2A 1        157 aatggatgagctgagtaagaggaagattcattggaggcacccataagaa
                    aagctcgagcacgaagaactagaaactgactagcagacaccacactata
                    ccgcctcctaggataccgcttccgtctgcactcccgtagctcacgtcgg

GAO             102 TTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADS
                    TTQNVNGLS+LPRQDGNQNGWIG HEVYLSSDGTNWGSPVASGSWFADS
                    TTQNVNGLSVLPRQDGNQNGWIG!HEVYLSSDGTNWGSPVASGSWFADS
MN-S2A 1        304 aacagagttgccccggacagtag4cggtcatggaatgacggtgattggt
                    ccaatagtcttcgaagaaaggtg aatatgcagcagggctccgggtcac
                    atgccccgtggtagttcaccgct tgttacatcacgccttgattgccct
```

MN-S2A 2/GAOstr Primer

```
GAO             GCTGTCACCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTC 111
MN-S2A 2        GCAGTTACCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTC 120
                ** ** ***** ********************************************

GAO             CTGAGCCTTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTC 171
MN-S2A 2        CTGAGCCTTCGAGCCGCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTC 180
                ************** ********************** ********************

GAO             ACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGAT 231
MN-S2A 2        ACTTGCGACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGAT 240
                ************************************************************

GAO             ACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACG 291
MN-S2A 2        ACCTTTTGGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACG 300
                ************************************************************

GAO             ATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGT 351
MN-S2A 2        ATTGACATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGT 360
                ********************************** ********************

GAO             AACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGG 411
MN-S2A 2        AACCAAAACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGG 420
                ************************************************************
```

MN-S2A 2 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSLRA+APIGSAI RNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSLRAAAPIGSAIPRNNWAVTCDSAQSG
MN-S2A 2         58 gggagccagggagacggacctcaccgggcagagaccaatggatgagctg
                    tctctcaactgcgtcaggtattgtgcccctggctcgaagctcgagcacg
                    tatcttcgcaatattagttgcgctacaatcaccttcccgcctcctagga

GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPR
MN-S2A 2        205 agtaagaggaagattcattggaggcacccataagaaaacagagttgccc
                    aagaactagaaactgactagcagacaccacactataccaatagtcttcg
                    taccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggta

GAO             115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
MN-S2A 2        352 cggacagtagccggtcatggaatgacggtgattggtaa
                    aagaaaggtggaatatgcagcagggctccgggtcaccc
                    gttcaccgctctgttacatcacgccttgattgccctta
```

MN-S2A 3/GAOstr Primer

```
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 118
MN-S2A 3        CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 112
                **** **************************************************

GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 178
MN-S2A 3        TTCGAGCCGCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 172
                ******** ********************** **************************

GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 238
MN-S2A 3        ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 232
                ************************************************************

GAO             GGCACACATTCTATGGCGCCAACGGGG-ATCCAAAGCCCCCTCACACATACACGATTGAC 297
MN-S2A 3        GGCACACATTCTATGGCGCCAACGGGGGATCCAAAGCCCCCTCACACATACACGATTGAC 292
                *************************** ******************************

GAO             ATGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAA 357
MN-S2A 3        ATGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAA 352
                ******************************** *************************

GAO             AACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGC 417
MN-S2A 3        AACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGC 412
                ************************************************************
```

MN-S2A 3 and GAO Amino Acid Sequence Alignment

```
GAO              17 VAVTVPHKAVGTGIPEGSLQFLSLRASAPIGSAISRNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSLRA+APIGSAI RNNWAVTCDSAQSG
                    VAVTVPHKAVGTGIPEGSLQFLSLRAAAPIGSAIPRNNWAVTCDSAQSG
MN-S2A 3         43 gggagccagggagacggacctcaccgggcagagaccaatggatgagctg
                    tctctcaactgcgtcaggtattgtgcccctggctcgaagctcgagcacg
                    tatcttcgcaatattagttgcgctacaatcaccttcccgcctcctagga
```

```
GAO              66 NECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPR
                    NECNKAIDGNKDTFWHTFYGAN DPKPPHTYTIDMKTTQNVNGLS+LPR
                    NECNKAIDGNKDTFWHTFYGAN!DPKPPHTYTIDMKTTQNVNGLSVLPR
MN-S2A 3        190 agtaagaggaagattcattgga4gcacccataagaaaacagagttgccc
                    aagaactagaaactgactagca acaccacactataccaatagtcttcg
                    taccgcttccgtctgcactccc tagctcacgtcggatgccccgtggta


GAO             115 QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
                    QDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTT
MN-S2A 3        338 cggacagtagccggtcatggaatgacggtgattggtaa
                    aagaaaggtggaatatgcagcagggctccgggtcaccc
                    gttcaccgctctgttacatcacgccttgattgccctta
```

MN-S2A 4 /GAOstr Primer

```
MN-S2A 4        CCGTTCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 120
GAO             CCGTCCCTCACAAGGCCGTAGGAACTGGAATTCCTGAAGGGAGTCTTCAGTTCCTGAGCC 356
                **** ************************************************** 

MN-S2A 4        TTCGAGCCGCAGCACCTATCGGAAGCGCCATTCCTCGCAACAACTGGGCCGTCACTTGCG 180
GAO             TTCGAGCCTCAGCACCTATCGGAAGCGCCATTTCTCGCAACAACTGGGCCGTCACTTGCG 416
                ******** ********************** ************************** 

MN-S2A 4        ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 240
GAO             ACAGTGCACAGTCGGGAAATGAATGCAACAAGGCCATTGATGGCAACAAGGATACCTTTT 476
                ************************************************************ 

MN-S2A 4        GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 300
GAO             GGCACACATTCTATGGCGCCAACGGGGATCCAAAGCCCCCTCACACATACACGATTGACA 536
                ************************************************************ 

MN-S2A 4        TGAAGACAACTCAGAACGTCAACGGCTTGTCTGTGCTGCCTCGACAGGATGGTAACCAAA 360
GAO             TGAAGACAACTCAGAACGTCAACGGCTTGTCTATGCTGCCTCGACAGGATGGTAACCAAA 596
                ******************************** ************************** 

MN-S2A 4        ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 420
GAO             ACGGCTGGATCGGTCGCCATGAGGTTTATCTAAGCTCAGATGGCACAAACTGGGGCAGCC 656
                ************************************************************ 
```

MN-S2A 4 and GAO Amino Acid Sequence Alignment

```
GAO              19 VTVPHKAVGTGIPEGSLQFLSLRA**S**APIGSAISRNNWAVTCDSAQSGNE
                    +TVPHKAVGTGIPEGSLQFLSLRA+APIGSAI RNNWAVTCDSAQSGNE
                    xTVPHKAVGTGIPEGSLQFLSLRA**A**APIGSAIPRNNWAVTCDSAQSGNE
MN-S2A 4         57 gagccagggagacggacctcaccgggcagagaccaatggatgagctgag
                    Nctcaactgcgtcaggtattgtgcccctggctcgaagctcgagcacgaa
                    tcttcgcaatattagttgcgctacaatcaccttcccgcctcctaggata


GAO              68 CNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQD
                    CNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLS+LPRQD
                    CNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSVLPRQD
MN-S2A 4        204 taagaggaagattcattggaggcacccataagaaaacagagttgccccg
                    gaactagaaactgactagcagacaccacactataccaatagtcttcgaa
                    ccgcttccgtctgcactcccgtagctcacgtcggatgccccgtggtagt
```

```
GAO              117 GNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYV
                     GNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYV
                     GNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYV
MN-S2A 4         351 gacagtagccggtcatggaatgacggtgattggtaaattatgaccgctg
                     gaaaggtggaatatgcagcagggctccgggtcacccaacatacgccgat
                     tcaccgctctgttacatcacgccttgattgcccttaaccctatcttctt
```

MN-H522A 1/GOFORB Primer

```
GAO              CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 1680
MN-H522A 1       CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 389
                 ************************************************************

GAO              ACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAAT 1740
MN-H522A 1       ACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAAT 449
                 ******   ********************************** **************

GAO              CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 1800
MN-H522A 1       CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 509
                 ************************************************************

GAO              ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 1860
MN-H522A 1       ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 569
                 ************************************************************

GAO              ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 1920
MN-H522A 1       ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 629
                 ************************************************************

GAO              AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 1980
MN-H522A 1       AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 689
                 ************************************************************
```

MN-H522A 1 and GAO Amino Acid Sequence Alignment

```
GAO              485 SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
                     SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
                     SVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIV
MN-H522A 1       162 tggccggaataaggcccgactggtacgtacgatgcgcgattacacatag
                     ctttcaggcttcggagggtctaacccttccatatcaaactaaaacactt
                     ttttacacgttaacaatatgcgtacgatatgccctaactccggcccctt

GAO              534 RVYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYLYNSNGNLA
                     R YHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYLY+SNGNLA
                     RAYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYLYDSNGNLA
MN-H522A 1       309 cgtcaatcttcggagtaggggctggtaaagtggcatacatctgaagacg
                     gcaagtctttcaggttaggggtggagccactacattccaataagagatc
                     ctctctctgattcgatctttttcttcgtcccgactgacttctccctcg

GAO              583 TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
                     TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
                     TRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRI
MN-H522A 1       456 accaaaaatacagagggaaaatagttaaagttactgagacagaagccca
                     cgcatcgcccagtatgggtctccacctgaccttgagcccactacaaggt
                     atcgtcactagccgctcatacggttgtcggggtcttagacgtttcgcct
```

MN-H522A 2/GOFORB Primer

```
GAO             CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 1680
MN-H522A 2      CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 385
                ************************************************************

GAO             ACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAAT 1740
MN-H522A 2      ACGAAAGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAAT 445
                *****    ********************************** **************

GAO             CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 1800
MN-H522A 2      CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 505
                ************************************************************

GAO             ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 1860
MN-H522A 2      ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 565
                ************************************************************

GAO             ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 1920
MN-H522A 2      ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 625
                ************************************************************

GAO             AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 1980
MN-H522A 2      AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 685
                ************************************************************
```

MN-H522A 2 and GAO Amino Acid Sequence Alignment

```
GAO              482 FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
                     FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
                     FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
MN-H522A 2       149 tcatggccggaataaggcccgactggtacgtacgatgcgcgattacaca
                     tacctttcaggcttcggagggtctaacccttccatatcaaactaaaaca
                     tcctttacacgttaacaatatgcgtacgatatgccctaactccggccc

GAO              531 SIVRVYHSISLLLPDGRVFNGGGGLCGDCTT**NH**FDAQIFTPNYLYNSNG
                     SIVR YHSISLLLPDGRVFNGGGGLCGDCTT  FDAQIFTPNYLY+SNG
                     SIVRAYHSISLLLPDGRVFNGGGGLCGDCTT**KA**FDAQIFTPNYLYDSNG
MN-H522A 2       296 tagcgtcaatcttcggagtaggggctggtaaagtggcatacatctgaag
                     cttgcaagtctttcaggttaggggtggagccactacattccaataagag
                     cttctctctctgattcgatctttttcttcgacccgactgacttctccc

GAO              580 NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
                     NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
                     NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
MN-H522A 2       443 acgaccaaaaatacagagggaaaatagttaaagttactgagacagaagc
                     atccgcatcgcccagtatgggtctccacctgaccttgagcccactacaa
                     tcgatcgtcactagccgctcatacggttgtcggggtcttagacgtttcg
```

MN-H522A 3/GOFORB Primer

```
GAO             CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 1680
MN-H522A 3      CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 374
                ************************************************************

GAO             ACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAAT 1740
MN-H522A 3      ACGAATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAAT 434
                ******   ********************************** **************
```

```
GAO             CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 1800
MN-H522A 3      CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 494
                ************************************************************

GAO             ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 1860
MN-H522A 3      ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 554
                ************************************************************

GAO             ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 1920
MN-H522A 3      ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 614
                ************************************************************

GAO             AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 1980
MN-H522A 3      AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 674
                ************************************************************
```

MN-H522A 3 and GAO Amino Acid Sequence Alignment

```
GAO              486 VVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVR
                     VVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVR
                     VVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPNSIVR
MN-H522A 3       150 ggccggaataaggcccgactggtacgtacgatgcgcgattacacatagc
                     tttcaggcttcggagggtctaacccttccatatcaaactaaaacacttg
                     tttacacgttaacaatatgcgtacgatatgccctaactccggccccttc

GAO              535 VYHSISLLLPDGRVFNGGGGLCGDCTTN**H**FDAQIFTPNYLYNSNGNLAT
                      YHSISLLLPDGRVFNGGGGLCGDCTTN FDAQIFTPNYLY+SNGNLAT
                     AYHSISLLLPDGRVFNGGGGLCGDCTTN**A**FDAQIFTPNYLYDSNGNLAT
MN-H522A 3       297 gtcaatcttcggagtaggggctggtaaagtggcatacatctgaagacga
                     caagtctttcaggttaggggtggagccactacattccaataagagatcc
                     tctctctgattcgatctttttcttcgtcccgactgacttctccctcga

GAO              584 RPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRIP
                     RPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRIP
                     RPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRIP
MN-H522A 3       444 ccaaaaatacagagggaaaatagttaaagttactgagacagaagcccac
                     gcatcgcccagtatgggtctccacctgaccttgagcccactacaaggtc
                     tcgtcactagccgctcatacggttgtcggggtcttagacgtttcgcctc
```

MN-H522A 4/GOFORB Primer

```
GAO             CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTACC 1680
MN-H522A 4      CTTTTGTTACCTGATGGCAGGGTATTTAACGGTGGTGGTGGTCTTTGTGGCGATTGTCCC 385
                ************************************************************ **

GAO             ACGAATCATTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACAATAGCAACGGCAAT 1740
MN-H522A 4      CCGTATGCCTTCGACGCGCAAATCTTTACGCCAAACTATCTTTACGATAGCAACGGCAAT 445
                 ** **   ********************************** **************

GAO             CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 1800
MN-H522A 4      CTCGCGACACGTCCCAAGATTACCAGAACCTCTACACAGAGCGTCAAGGTCGGTGGCAGA 505
                ************************************************************

GAO             ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 1860
MN-H522A 4      ATTACAATCTCGACGGATTCTTCGATTAGCAAGGCGTCGTTGATTCGCTATGGTACAGCG 565
                ************************************************************

GAO             ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 1920
MN-H522A 4      ACACACACGGTTAATACTGACCAGCGCCGCATTCCCCTGACTCTGACAAACAATGGAGGA 625
                ************************************************************
```

```
GAO             AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 1980
MN-H522A 4      AATAGCTATTCTTTCCAAGTTCCTAGCGACTCTGGTGTTGCTTTGCCTGGCTACTGGATG 685
                ************************************************************
```

MN-H522A 4 and GAO Amino Acid Sequence Alignment

```
GAO              482 FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
                     FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
                     FHTSVVLPDGSTFITGGQRRGIPFEDSTPVFTPEIYVPEQDTFYKQNPN
MN-H522A 4       149 tcatggccggaataaggcccgactggtacgtacgatgcgcgattacaca
                     taccttcaggcttcggagggtctaacccttccatatcaaactaaaaca
                     tccttttacacgttaacaatatgcgtacgatatgccctaactccggccc


GAO              531 SIVRVYHSISLLLPDGRVFNGGGGLCGDC**TTNH**FDAQIFTPNYLYNSNG
                     SIVR YHSISLLLPDGRVFNGGGGLCGDC    FDAQIFTPNYLY+SNG
                     SIVRAYHSISLLLPDGRVFNGGGGLCGDC**PPYA**FDAQIFTPNYLYDSNG
MN-H522A 4       296 tagcgtcaatcttcggagtaggggctggtcctgtggcatacatctgaag
                     cttgcaagtctttcaggttaggggtggagccactacattccaataagag
                     cttctctctctgattcgatctttttttcttcgtcccgactgacttctccc


GAO              580 NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
                     NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
                     NLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQ
MN-H522A 4       443 acgaccaaaaatacagagggaaaatagttaaagttactgagacagaagc
                     atccgcatcgcccagtatgggtctccacctgaccttgagcccactacaa
                     tcgatcgtcactagccgctcatacggttgtcggggtcttagacgtttcg
```

Although all the sequence of the mutants was analyzed, here only the sequence analysis of the site of interest is given. The other sites are compatible with the original sequence.

APPENDIX F

THE GENETIC CODE AND SINGLE-LETTER AMINO ACID DESIGNATONS

		Second Position of Codon					
		T	**C**	**A**	**G**		
First Position	**T**	TTT Phe [F] TTC Phe [F] TTA Leu [L] TTG Leu [L]	TCT Ser [S] TCC Ser [S] TCA Ser [S] TCG Ser [S]	TAT Tyr [Y] TAC Tyr [Y] TAA *Ter* [end] TAG *Ter* [end]	TGT Cys [C] TGC Cys [C] TGA *Ter* [end] TGG Trp [W]	**T** **C** **A** **G**	**Third Position**
	C	CTT Leu [L] CTC Leu [L] CTA Leu [L] CTG Leu [L]	CCT Pro [P] CCC Pro [P] CCA Pro [P] CCG Pro [P]	CAT His [H] CAC His [H] CAA Gln [Q] CAG Gln [Q]	CGT Arg [R] CGC Arg [R] CGA Arg [R] CGG Arg [R]	**T** **C** **A** **G**	
	A	ATT Ile [I] ATC Ile [I] ATA Ile [I] ATG Met [M]	ACT Thr [T] ACC Thr [T] ACA Thr [T] ACG Thr [T]	AAT Asn [N] AAC Asn [N] AAA Lys [K] AAG Lys [K]	AGT Ser [S] AGC Ser [S] AGA Arg [R] AGG Arg [R]	**T** **C** **A** **G**	
	G	GTT Val [V] GTC Val [V] GTA Val [V] GTG Val [V]	GCT Ala [A] GCC Ala [A] GCA Ala [A] GCG Ala [A]	GAT Asp [D] GAC Asp [D] GAA Glu [E] GAG Glu [E]	GGT Gly [G] GGC Gly [G] GGA Gly [G] GGG Gly [G]	**T** **C** **A** **G**	

Figure F.1. The genetic code

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 F.2 Single-letter and three letter amino acid designations with their mass and pI values