CLONING, CHARACTERIZATION AND EXPRESSION of *A. thaliana* G PROTEIN α-SUBUNIT GENE FOR STRUCTURAL STUDIES

by

E. SÜPHAN BAKKAL

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University
February 2003

CLONING, CHARACTERIZATION AND EXPRESSION of *A. thaliana* G PROTEIN α-SUBUNIT GENE FOR STRUCTURAL STUDIES

APPROVED BY:

Assoc. Prof. Zehra Sayers …….
(Dissertation Supervisor)

Prof. Beki Kan …….

Assistant Prof. Damla Bilgin …….

DATE OF APPROVAL: …….

ABSTRACT

Heterotrimeric G proteins are known to play a vital role in signal transduction cascades of animals and plants by transmitting signals from membrane bound receptors to downstream effectors. They consist of α, β, and γ subunits. Mammals possess 20 α, 6β and 12 γ subunits. Four main classes of G proteins have been detected: G_s, which activates adenyl cyclase; G_i, which inhibits adenyl cyclase; G_q, which activates phospholipase C; and G_{12} and G_{13} of unknown functions.

Activation of plasma membrane spanning heptahelical receptors by an extracellular ligand causes the activation of G-proteins by inducing the exchange of GDP for GTP at the binding site on G_α. Dissociated G_α and $G_{\beta\gamma}$ subunits bind and activate their downstream effectors including adenyl cyclase, phospholipase C-β, cGMP phosphodiestarase. According to the type of the subunits different pathways are activated or inhibited.

In plants, 13 α, 7 β and 2 γ subunits have been detected from monocots and dicots. *GPA1* is the first identified α-subunit from *Arabidopsis thaliana*. It is a single copy gene, consisting of 14 exons and 13 introns, and located on chromosome II. GPα1 is the protein containing 383 amino acids, encoded by *GPA1*. Immunolocalization and membrane fractionation have shown that GPα1 is associated with the plasma and ER membrane. According to the expression pattern *GPA1* is detected in all organs and cell types, it is most abundant in vegetative tissues including leaves and roots, less in floral stems and least in floral buds and floral meristems.

Among previously characterized G protein α-subunits, rat G_{i1-3} and bovine rat transducin are most similar to GPα1. 36% of its amino acids are identical and 73% are

similar to rat transducin.

Gain and loss of function and overexpression studies have shown that GPα1 is involved in a number of cellular processes including hormone signaling, seed germination, defense response to pathogens, growth, regulation of biosynthetic pathways as well as regulation of ion (K^+ and Ca^{+2}) channels and opening of stomatal guard cells.

In addition to functional studies, sequence-based homology modeling has shown that GTP binding and hydrolysis, and switch regions for effector and β-subunit binding are highly conserved while putative AC binding region shows little homology with mammalian counterparts. Despite several functional studies, mainly *in situ*, there is no experimental work on structure analysis of plant heterotrimeric G-proteins. Information on folding, effector- GPα1 interactions and mechanism of function is lacking.

Detailed functional characterization of GPα1 and comparison with mammalian counterparts would be greatly facilitated with X-ray crystal and solution scattering data. In this thesis characterization of *GPA1* and results of cloning and expression using a variety of systems, with the broader aim of structural studies on the purified recombinant protein are presented.

ÖZET

Heterotrimerik G-proteinleri bitkilerde ve hayvanlarda hücre zarına bağlı reseptörlerden hücrenin iç birimlerine sinyal iletmekle görevli olan önemli proteinler olarak bilinirler. α-, β- ve γ- alt birimlerinden oluşmuşlardır. Memelilerde 20 α, 6 β, ve 12 γ alt birimleri tespit edilmiştir. Dört ana G-proteini sınıfı vardır: G_s, adenyl siklazı uyarır; G_i, adenyl siklazı baskılar; G_q, phospholipaz C'yi uyarır; G_{12} ve G_{13}'nın fonksiyonları bilinmemektedir.

Hücre zarındaki reseptörün uyarılması, α-alt birimindeki GDP'nin GTP ile değiştirilmesine neden olur. Bunun sonucunda uyarılan heterotrimer G-proteininin α-alt birimi βγ-alt birimlerinden ayrılır ve farklı proteinlere bağlanırlar. Adenyl siklaz, phospholipaz C-β, cGMP phosphodiesteraz gibi hücre içi proteinlere bağlanan alt birimlerin familyasına göre farklı sistemler uyarılır yada baskılanır.

Bitkilerde tek-çenekli ve çift-çeneklilerden 13 α, 7 β ve 2 γ alt birimleri tespit edilmiştir. *GPA1, Arabidopsis thaliana'*dan belirlenen ilk ve tek α-alt birimidir. *GPA1* tek kopya bir gen olup, 14-egzondan ve 13-introndan oluşmaktadır ve kromozom II'nin üzerinde tespit edilmiştir. GPα1, *GPA1* geninin kodladığı protein olup, 383 amino asitten oluşur. Immunolokalizasyon ve hücre analiz çalışmaları bu proteinin hücre zarı ve ER zarında bulunduğunu tespit etmiştir. GPα1 bütün organ ve hücre tiplerinde, ama en çok yapraklarda ve kökte bulunmaktadır.

Daha önce karakterize edilmiş G-proteini α-alt birimleri arasında amino asit dizi karşılaştırması yöntemi kullanılarak yapılan çalışmalara göre GPα1 36%'lık bir oranla birebir ve 73%' lük bir oranla fare transducin proteinine benzer bulunmuştur.

Fonksiyon çalışmaları göstermiştir ki GPα1, hormon sinyallerinin iletimi, tohum gelişimi, zararlılara karşı bağışıklık sisteminin oluşturulması, biyosentez mekanizmalarının ve iyon (K^+ ve Ca^{+2}) kanallarının düzenlenmesi gibi birçok hücre içi mekanizmada rol almaktadır.

Fonksiyon çalışmalarına ek olarak yapılan dizi analizine dayalı homoloji modellemesi çalışmaları göstermiştir ki GTP bağlanması ve hidrolizi ile ilgili olan bölgeler ve efektör ve β-alt birimini bağlandıkları switch bölgeleri çok yüksek oranda korunmuştur. Buna karşılık AC bağlanan bölge memeli alt birimleriyle kıyaslandığında çok az bir homoloji göstermiştir. Hücre içi fonksiyonları anlamak üzere yapılan çalışmalara karşılık, bitki heterotrimer G-proteinlerinin yapısal analizini ortaya koymak üzere herhangi bir deneysel çalışma yapılmamıştır. Proteinin katlanması, effector-GPα1 ilişkileri ve fonksiyon mekanizmalarıyla ilgili bilgiler eksiktir.

GPα1'ın X-ışını kristallizasyonu ve solüsyon içindeki X-ışını saçılımlarının, memeli eşlenikleriyle karşılaştırılması sunucunda detaylı yapısal bilginin elde edilmesi beklenmektedir. Bu tezde *GPA1*'ın karakterizasyonu, farklı sistemler kullanılarak yapılan klonlama ve protein üretme çalışmaları sonucunda elde edilen bilgiler sunulacaktır.

To my family with all my hearth

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Assoc. Prof. Zehra Sayers, who gave the great responsibility and self-confidence to me throughout this study. She is a great mentor and taught me that the most important thing is the ability of discriminate between what you are looking and what you see and be patient in my work. I am grateful for her guidance and advice about every detail of this thesis. I hope that I would not lose her support and friendship throughout my life.

I would like to express my special thanks to Mert Şahin, who is my lab partner. He is a brilliant and succesfull scientist. Without his support it is impossible to finish this work. It is a great pleasure for me to have his friendship. I am also grateful to my graduate friends who are working at Sayer's lab: Ümit Öztürk, Kıvanç Bilecen, Çağdaş Seçkin, Erinç Şahin, and undergraduates Tolga Sütlü and Doğanay Duru. Their friendship and support are so important for me.

I would like to thank to Dr. Özgür Kütük, Melis Tiryakioglu, Pelin Akan, and Mazhar Adlı for their support and friendship. I am also grateful to all my roommates: Burcu Ünlü, Gamze Yurttutan and Burcu Dartan. They are so friendly and support me during hard days for two years. I would like to thank to Tuğba Baytekin for her support and friendship additional to her hard work.

Finally, I would like to thank faculty members: Prof. Hüveyda Başağa, Prof. İsmail Çakmak, Assist. Prof. Uğur Sezerman, Assist. Prof. Damla Bilgin, Assist. Prof. Metin Bilgin and all graduate and undergraduate students at the Biological Sciences and Bioengineering Program. It was a great pleasure and experience for me "working and creating together" with the all members of FENS at Sabancı University.

TABLE OF CONTENTS

ABBREVIATIONS

ABA: Abscisic acid

AC: Adenylyl cyclase

ATP:Adenosine triphosphate

BR: Brassinosteroids

C-terminus: Carboxyl terminus

GA: Gibberelic acid

GAP: GTPase activating protein

G_{α}: G-protein alpha subunit

G_{β}: G-protein beta subunit

GDP: Guanosine di-phosphate

G_{γ}: G-protein gamma subunit

GNEF: Guanine nucleotide exchange factor

GPα1: G_α protein from *A. thaliana*

GPA1: G_α gene from *A. thaliana*

GPCR: G-protein coupled receptor

GTP: Guanosine tri-phosphate

GTPase: enzyme converting GTP into GDP

G_α-GDP: G_α bound to GDP, in its inactive state

G_α-GTP: Gα bound to GTP, in its active state

DAG: Diacylglycerol

DRG: Developmentally regulated G proteins

FR: Far-red

IP3: Phosphatidylinotisol 3-kinase

JA: Jasmonic acid

NMR:Nuclear magnetic resonance

N-terminus: Amino terminus

PLC: PhospholipaseC

PhyA: Phytochrome A

PhyB: Phytochrome B

PKC: Protein kinase C

R: red

RGS: Regulators of G-protein signaling

SA: Salicylic acid

WD40 repeat: Tryptophane-Aspartate repeat consisting of 40 residues

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Guanine nucleotide binding proteins are mainly divided into three major classes: Small GTP binding proteins, heterotrimeric G proteins, and elongation factors. All these classes have important roles in signal transduction pathways. They interact with one another to regulate metabolic enzymes, ion channels, transporters, and other components of cellular machinery.

Small GTP binding proteins belongs to the Ras superfamily. This superfamily is divided into five families including Ras, Rho, Rab, Arf and Ran. Although they have guanine nucleotide binding domains and effector domains like α-subunit of heterotrimeric G proteins, they differ in the mechanisms by which they are regulated as well as those by which they activate the downstream targets. Upon activation by an upstream signal, a guanine nucleotide exchange factor (GNEF) exchanges GTP for GDP to activate the small GTPase protein. GTP-bound form interacts with the downstream effector proteins. Because active form exhibits very low intrinsic GTPase activity, a GTPase activating protein (GAP) is required for GTP hydrolysis. Plant homologues of these proteins are found including Ras, Raf, Arf, and ROP. *Arabidopsis* genome is predicted to encode 93 small GTPases, which regulate cellular processes ranging from vesicle trafficking to hormone signaling.

Eukaryotic and bacterial elongation factors are another class of GTP binding proteins. The GTP-bound form of EF-Tu brings the charged tRNA to the ribosome while GTP-bound EF-G catalyzes the translocation of amino acid from the A to P site on the ribosome.

Determined structures of EF-G, the ternary complex between EF-Tu, GTP and the aminoacyl-tRNA, and EF-Tu-EF-Ts provide information about the interaction of tRNA with the elongation factors as well as the mechanism of translocation.

Heterotrimeric G proteins are major signal transducers that integrate the signals from the extracellular signals including hormones, neurotransmitters, chemokines, autocrines and paracrine factors, by interacting the downstream effectors. They consists of α, β, and γ subunits. Upon activation of membrane-bound G-Protein coupled receptors (GPCR) by an external signal, GPCR acts as nucleotide exchange factors for G_{α} subunits. After exchange of GTP for GDP, the tightly associated G_{α} and $G_{\beta\gamma}$ subunits of the G protein separate from each other. Then GTP bound α-subunit and βγ-complex activates their effectors. Amplification of the signal through phosphorylation cascades leads to the activation of a variety of signal transduction pathways resulting the expression of specific genes, activation of ion channels, transporters and other cellular machinery. In plant system, 13 α, 7 β and 2 γ subunits were detected from monocots and dicots including *Arabidopsis thaliana*, tomato, rice, and soybean.

Three-dimensional structures of proteins can be determined both experimentally and computationally to support the results of functional studies. Small angle scattering of proteins in solution is the important tool for determining the protein-protein interactions and domain movements when compared with X-ray crystallography which gives detailed static information in the crystallized form of the protein or NMR which is more easily applicable to small molecules. In addition, X-ray scattering allows monitoring the dynamics of conformational changes in solution. For structural studies it is necessary to produce large quantities (more than tens of milligrams) of pure and soluble recombinant protein, which can be achieved using a variety of cloning and expression systems.

This thesis is organized as to give general information about the GTP binding proteins in animal and plant systems in the context of structural and functional studies. The characterization of *GPA1*, which is the α-subunit of heterotrimeric G protein from

Arabidopsis thaliana and the biochemical studies to determine its cellular localization and possible functions are the main focus of the overview. In addition, other plant α, β, and γ subunits and their functions in signal transduction processes are reviewed.

Results of experiments aiming to clone *GPA1* are presented and compared together with previous reports in the literature. Results of expression of the recombinant protein for structural studies are also presented and put into perspective.

2 OVERVIEW

2.1 Guanine Nucleotide Binding Proteins

There are three main groups of guanine nucleotide binding proteins: superfamily of Ras-related proteins, heterotrimeric G-proteins and elongation factors involved in protein synthesis including EF-Tu and EF-G (Reviewed by Hilgenfeld R., 1995).

2.1.1 Ras-related proteins

Ras is a small, monomeric GTP binding protein existing in all eukaryotic cells. Active GTP-bound form of Ras binds to growth factor-responsive tyrosine kinase-associated receptors resulting in activation of gene transcriptional factors through a cascade of phosphorylations. Cycling of Ras protein requires two proteins: Guanine nucleotide exchange factor (GNEF) and GTPase activating protein (GAP). GNEF facilitates exchange of GDP for GTP and after GTP is bound, it dissociates yielding active Ras-GTP form. Activated Ras binds to the N-terminal domain of Raf, a serine /threonine kinase. Hydrolysis of GTP regenerates the GDP bound inactive form of Ras; this process is accelerated by GAP. After Ras returns to the GDP bound form, it dissociates from Raf. MEK (a dual specificity protein kinase that phosphorylates both tyrosine and serine residues) and MAP kinase (serine/threonine kinase) lie downstream of Ras and Raf. Activated MAP kinase by MEK kinase translocate into the nucleus and phosphorylate many proteins including transcription factors that regulate expression

important cell cycle and differentiation specific proteins (Reviewed by Hilgenfeld, 1995).

2.1.2 Elongation Factors-EF-Tu and EF-G

Elongation factors are required for the initiation of translation. Bacterial elongation factors including EF-Tu and EF-G have been studied using structural techniques to understand the interaction between the ribosome and aminoacyl-tRNA. The GTP bound form of EF-Tu brings the charged tRNA to the ribosome while GTP-bound EF-G catalyzes the translocation of amino acid from the A to P site on ribosome during initiation of translation.

Three dimensional crystal structures of GTP-binding domains of EF-Tu and Ras p21 proteins, which is the mammalian Ras protein with the mass of 21 KDa, have provided information about GTP hydrolysis. GTP hydrolysis by both Ras and EF-Tu leads to inversion of configuration at the γ-phosphate as a consequence of a direct attack of a water molecule. According to the structural data, a conserved glutamine residue in Ras and a conserved histidine residue in EF-Tu have important functions in stabilization of transition state of $G_{t\alpha}$-GTP and effector- enhanced GTP hydrolysis, respectively. The crystal structure of ligand free and GDP-bound form of EF-G, which are similar, resemble to the complex of GTP.Mg^{+2}-bound EF-Tu with tRNA (Reviewed by Sprang, 1997; reviewed by Hilgenfeld, 1995).

2.1.3 Heterotrimeric G proteins

2.1.3.1 General information, types and evolution

Heterotrimeric G-proteins consist of three subunits: α, β, and γ where α- and γ-subunits are post-translationally modified for membrane binding. Although the β- (~35 KDa) and γ- subunits (~7-10 KDa) show little variation the α- subunit (~45 KDa) appears to be tissue specific. In the inactive form the heterotrimer contains a GDP bound to the α- subunit. The α- subunit interacts with a heptahelical transmembrane receptor upon activation of the receptor by binding of the ligand and GDP is exchanged with GTP. The scheme according to which the heterotrimeric G-protein activation leads to signal transduction is shown in figure 2.1. In the activated form the α- subunit dissociates from the βγ complex and interacts with target effectors including adenylate cyclase, phosphodiesterase, and ion channels and transporters, secretory machinery. Regulators of G-proteins (RGS) and phosducin regulate both G_{α} and $G_{\beta\gamma}$ complexes, respectively. RGS accelerates the rate of GTP hydrolysis to GDP.

Phosducin, a soluble phosphoprotein present in the retina, forms complexes with βγ-subunit of transducin, the retinal G protein. It binds tightly to $G_{\beta\gamma}$ dimer and regulate the number of βγ-subunits available to interact with G_{α} subunits. Phosducin inhibits the intrinsic GTPase activity of G_{α} subunits until phosphorylation by protein kinase A. Upon phosphorylation, phosducin dissociates from βγ-subunit due to the conformational change. Although phosducin is specific for retinal tissues, phosducin mRNA has been detected in brain, heart and liver and lung tissue (Danner *et al*., 1996).

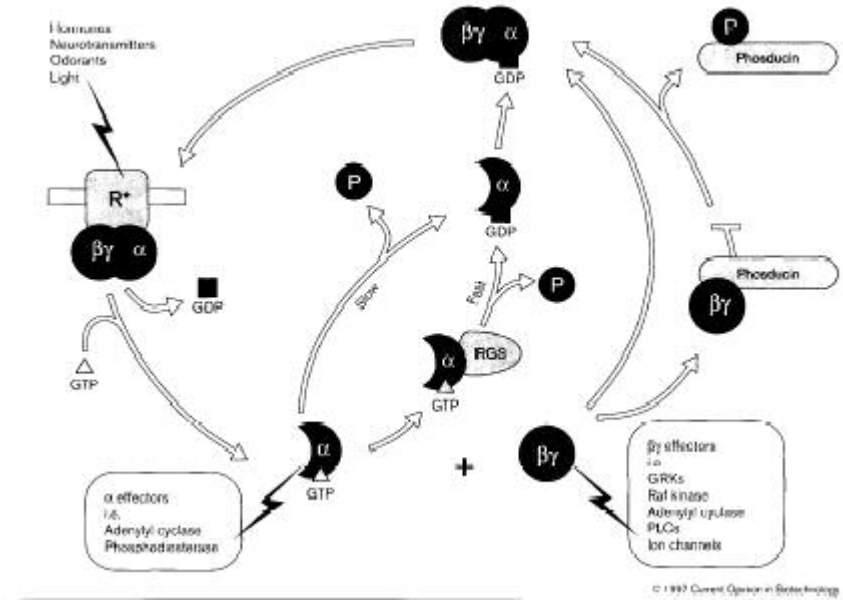


Figure 2.1 Schematic diagram of G-protein coupled signal transduction pathways (Bohm *et al*., 1997)

2.1.3.2 G protein α- subunit

There are three main groups of G_α subunits classified according to the sequence comparisons: GPA, Gα-I and Gα-II Groups. GPA contains genes from plants, fungi and slime mold. Phylogenetic tree based on the d_{12} (number of nucleotide substitutions per site) and K values (number of amino acid replacements) show the evolutionary the α-subunits within these groups. Gα-I Group includes G_o group, G_t group, G_x group and G_i group. The common ancestor of G_i and G_x diverged from the ancestor of G_t group. Gα-II Group contains G_q group, G_{12} group and Gs group. According to the sequence and intron comparisons of the most conserved regions, it suggests that G_α multigene family evolved in the common ancestor of fungi, plants and animals, slime molds and ciliates, after their divergence from simpler eukaryotes (Yokoyoma *et al.,* 1994). These four classes consist of more than 20 different proteins in mammalians. G_s (stimulatory) activates adenyl cyclase (AC), G_i (inhibitory) inhibits the AC. G_q activates phospholipase C (PLC) and G_{12} regulate small GTP binding proteins (Neves *et al*., 2002).

Structural data about $G_{\alpha t}$ and $G_{i\alpha 1}$ shows that there are two domains of α-subunit, which can be seen in figure 2.2. One is the guanine nucleotide binding domain (the G domain) and the other is the unique helical domain containing secondary structure elements keeping the GTP in the core of the protein. The extended N-terminal helix points away from the core of the protein (Wall *et al*., 1995). There are also 5 loops, G1 to G5, that correlate with conserved regions, which have been assigned functional roles (Reviewed by Sprang, 1997).

The interaction between G_{α} and $G_{\beta\gamma}$ complex occur via two distinct interfaces. SwitchII changes conformation between GDP-bound and GTP-bound state of the α-subunit. Conformation change occurs in the region that makes most of the contact with the β-subunit in the complex. The second important interface is the N-terminal region, which has 900 $Å^2$ interface area although switch II has 1800 $Å^2$ interface area. Switch I region corresponds to G-2 loop which is also called the effector loop because effector and Ras-GAP binding regions are there. In addition, G-2 loop forms part of Mg^{+2} binding loop.

Comparison of structure of Ras and its homologues α- and β-phosphate recognition loop (G-1) and guanine base recognition loop (G-4) shows that they are similar in structure whereas Mg^{+2} (G-2) and GTP-γ- phosphate binding sites (G-3) show variation in conformation. (Reviewed by Sprang, 1997)

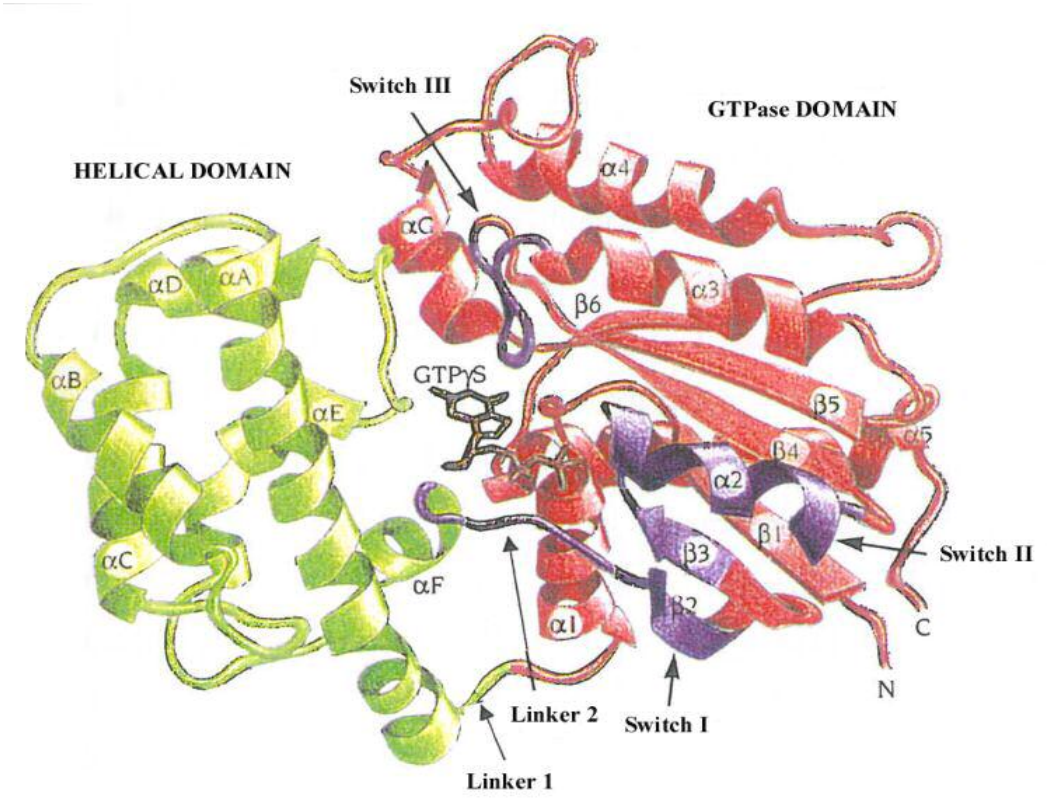


Figure 2.2 Structure of G-protein α-subunit (Rens-Domiano S *et al*., 1995)

2.1.3.3 G protein-β- and γ-subunits

β- and γ- subunits of the heterotrimer form a functional unit after dissociating from G_{α} –GTP complex. As shown in figure 2.3 the β-subunit has a long N-terminus helix followed by a short stretch of polypeptide chain that links it to a seven-bladed propeller structure containing a water-filled pore (Reviewed by Hamm, 1998).

β-subunit belongs to the β-propeller family of proteins. Propeller is composed of a sevenfold repeat of four-stranded β-sheet units and its structural symmetry arises from an internally repeated WD40 sequence motif, ~40 residue motif often ending with Trp (W) and Asp(D) amino acids. (Reviewed by Wittinghofer, 1996)

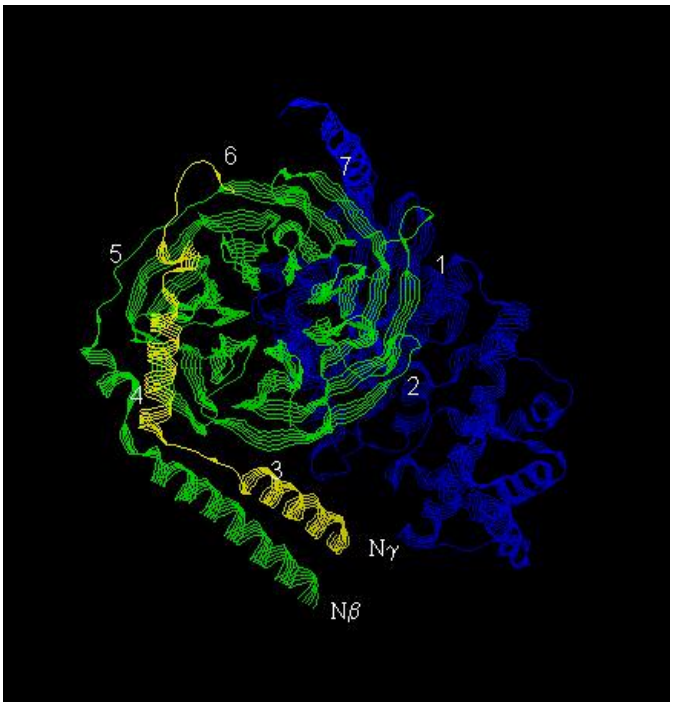


Figure 2.3 G protein heterotrimer $G_{i\alpha\beta1\gamma2}$ with GDP bound complex (extracted from PDB entry 1GG2). α-subunit is shown in blue, β-subunit is shown in green and γ-subunit is shown in yellow. Orientation of the molecule has been chosen to maximize visualization of the βγ-subunit.

G_{γ} forms a coiled coil with N-terminus of β-subunit. Approximately 80 residue of γ-subunit is highly extended and embedded on the surface of the torrodial β-subunit. Many of the $G_{\gamma2}$ residues at βγ interface make only hydrophobic contacts. Although all of the $G_{\beta1}$ residues that contact $G_{i\alpha}$ and G_{γ} are highly conserved, $G_{\beta1}$ is tolerant of sequence variation between different G_{γ} and G_{α} isoforms (Wall *et al*., 1998). βγ-subunit, as a signaling complex, keep the α-subunit is the non-signaling, GDP bound form, presents α-subunit to the cytoplasmic sides of the receptors (Iñiguez-Lluhi *et al.,* 1993).

2.1.3.4 Regulators of G-protein signalling

Strength and duration of a signal for activation of heterotrimeric G-proteins depend on the receptor phosphorylation by protein kinases, effector enzymes, phosducin, arrestin and RGS. RGS proteins act as GAPs for G_{α} subunits. SST2 is the

first RGS identified in yeast. Then flbA, EGL-10 were identified from fungus and nematode, respectively with yeast two hybrid experiments. RGS1, RGS2, RGS3 and RGS4 are regulators of G-protein signaling in mammalian cells. All RGS proteins have GAP-like domains. RGS binds to the G_α in such a way that RGS-bound form of G_α-GTP complex structurally approximates Ras-GAP bound form of Ras-GTP. Hence, intrinsic GTPase activity of G_α subunits has increased (Reviewed by Dohlman and Thorner, 1997).

2.1.3.5 Lipid modifications

Membrane association of G-protein subunits is a complex process including multiple interactions. N-myristoylation (co-translational addition of the saturated 14-carbon fatty acid myristate to a glycine residue at the extreme N-terminus after removal of the initiating methionine), and S-palmitoylation (addition of palmitate, a 16 C saturated fatty acid) of G_α and prenylation (attachment of geranyl-geranyl or farnecyl moiety to a cysteine residue located in the C-terminal CAAX box of G_γ) provide the attachment of heterotrimer to the inner surface of plasma membrane. These lipid modifications also stabilize the interaction of the α-subunit with the βγ-dimer. Although mutant non-prenylated γ-subunit can form a stable dimmer with β-subunit, prenylation of γ is required for both the formation of stable βγ dimer and for the correct membrane targeting of the βγ dimer. In addition, prenylation is absolutely required for interactions of βγ with α-subunits, receptors and effectors (Wedegaertner *et al.*, 1995).

Most G-alpha subunits excluding $G_{t\alpha}$ are S-palmitoylated at a cysteine near the amino terminus and others G_α (o, i, z, t) are N-myristoylated at Gly-2. Lipid modifications help G_α subunits bind to the plasma membrane and juxtapose them to their cognate receptors and effector targets. Members of $G_{\alpha i}$ family can be ADP-ribosylated by pertussis toxin at a cysteine residue four residues removed from C-terminus, thereby inhibiting interaction with their receptors. $G_{\alpha s}$ and $G_{t\alpha}$ can be ADP-ribosylated at a conserved arginine residue in G-2 box, which permits GTP binding but abolishes GTPase activity (Reviewed by Sprang, 1997).

2.1.3.6 GTP binding

According to the structural data obtained from a variety of proteins ranging from Ras to heterotrimeric G-proteins, the 200 amino acid guanine nucleotide binding domain has 6 stranded β-sheets surrounded by 5 α-helices on both sides. Five polypeptide loops from G1 to G5 are highly conserved in this domain for EF-Tu, $p21^{ras}$, and $G_{\alpha.}$

βγ binding loop (P-loop or G1-box) presents GXXXXGK (S/T) consensus sequence and connects β1 strand to α1 helix resulting in contact between α- and β-phosphates of the guanine nucleotide. The G2 box provides connection between α1 helix and β2 sheets. The conserved threonine residue is involved in Mg^{+2} coordination. The G3 box has a consensus sequence DXXG at N terminus of α2 helix and links the sites for binding Mg^{+2} and γ phosphate of GTP. Guanine nucleotide base is recognized by a conserved N/TKXD sequence of the G4 box. The G5 box is located between β6 sheet and α5 helix and contains a consensus sequence providing guanine base recognition site. Comparison between guanine nucleotide binding domain of Ras with others shows the additions and substitutions among them (Reviewed by Sprang, 1997; Vetter *et al*., 2001).

Structural studies on G_α isoforms have shown that there are two tandem glycine residues (Gly203 and Gly226) and that are conserved. These provide flexibility to the N-terminus of switch II region and form part of the γ-phosphate binding site of GTP, complexed with Mg^{+2}. Particularly G_α subunits are unstable in the absence of Mg^{+2}. GTP induced conformational change within switch II region of G_α subunit results in releasing of G_α from $G_{\beta\gamma}$ dimer (Wall *et al*., 1998). Loss of interaction between switch II region and the γ-phosphate of GTP causes deactivation.

2.1.3.7 G protein coupled receptors (GPCR)

GPCRs are the largest gene family in most animal genomes. About 1% of total genes in Drosophila genome, >5% of C. elegans genome and >1% of human genome are coding the proteins containing heptahelical structures (Marinissen *et al*., 2001).

As schematically illustrated in figure2.4, different stimuli including neurotransmitters, hormones, lipids, photons, odorants, taste ligands, nucleotides and calcium ions act through the seven transmembrane spanning G-protein coupled receptors.

Upon activation by ligand binding, transmembrane helices 3 and 6 that mask G-protein binding sites on the second, third and the fourth cytoplasmic loops are uncovered (Ernst *et al*., 2000; Marin *et al*.; 2000). Association of GPCR with heterotrimeric G-protein promotes the exchange of GDP for GTP from G_{α} subunit resulting dissociation of G_{α} from $G_{\beta\gamma}$ dimer (Hur *et al*., 2002).

Sequences in the second and the third cytoplasmic loops of the receptors are important for selectivity and/or affinity of the receptor-G protein interactions. In addition to carboxyl terminus, the amino terminus of the α-subunit and the identity of the γ-subunit are also important for the receptor-G-protein interactions. Fluorescent labelled surface Cys210 located in switch II and Cys347 located in the C-terminus were used to report local conformational changes in $G_{t\alpha}$ upon induction. Results of these studies indicated that activation-dependent conformational change at the COOH terminus of $G_{t\alpha}$ might lead to a lowered receptor affinity and dissociation from the ternary complex (Yang *et al.,* 1999).

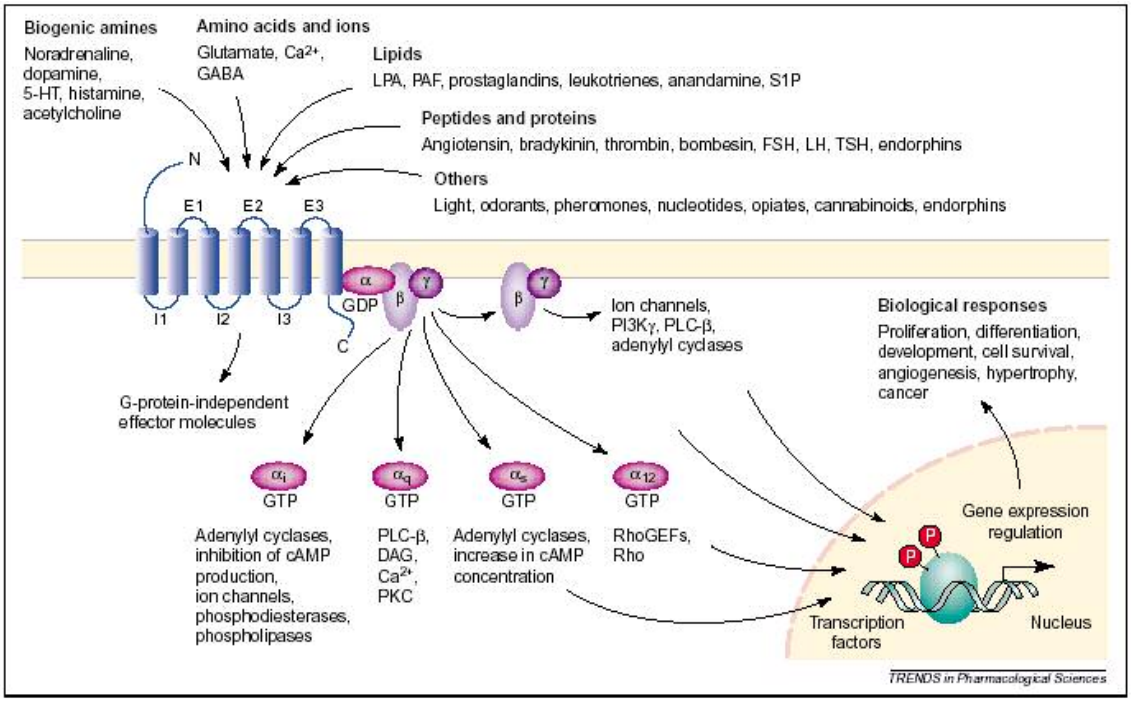


Figure 2.4 GPCRs and a variety of ligands (Marinissen *et al*., 2001)

High resolution structure of rhodopsin (visual light receptor) and crosslinking experiments on transducin have shown that N-terminal and C-terminal receptor binding regions of G_α are very close together in the receptor-bound to G_t (Hamm HE, 1998). In addition, GPCRs are in direct contact with $G_{\beta\gamma}$ subunits (Taylor *et al*., 1996; Yasuda *et al*., 1996).

Activation of one signaling pathway through the GPCR can either amplify or inhibit another signaling pathway depending on the cross talk among the downstream effectors' network. Dimerization of GPCRs results in distinct functional properties from the receptor components of the dimer (Hur *et al*., 2002).

2.1.3.8 Signalling pathways of G-Proteins

Many extracellular agents like hormones, neurotransmitters, chemokines, and local mediators activate four different G_α subunits through the activation of membrane spanning G-protein coupled receptors. Upon activation, different G_α subunits regulate different cellular machinery like ion channels, metabolic enzymes and transcriptional

regulators. Regulation of cellular machinery gives rise to the altered cellular functions including glucose metabolism, steroid production, chemotaxis, cardiac function, pituitary function, and pacemaker activity.

G_i signaling pathway, which mediates the light detection in the eye, is the best characterized G-protein signaling pathway. It inhibits the activity of adenyl cyclase upon activation of the receptor by many hormones and neurotransmitters, including epinephrine, acetylcholine, dopamine and serotonin. Phospholipase C-β (PLC-β), K^+ channels, adenyl cyclase, phosphatidylinotisol 3-kinase are the effectors regulated by $G_{\beta\gamma}$ of the $G_{\alpha i}$ pathway.

PLC-β hydrolyzes several inositol lipids to form secondary messengers like phosphatidylinotisol 3-kinase (IP_3) and diacylglycerol (DAG). These secondary messengers mediate the release of intracellular Ca^{+2} and activate protein kinase C (PKC), respectively. The activation of the PKC plays a key role in many aspects of cellular growth and metabolism (Hur *et al*., 2002).

G_s pathway activates the enzyme adenyl cyclase, which catalyze the production of cAMP from ATP, upon activation of a GPCR. β-adrenergic receptors are coupled with stimulatory G proteins. Activation of the receptor by a hormone ligand results in dissociation of $G_{\alpha s}$ from βγ-complex and $G_{\alpha s}$-GTP binds to adenyl cyclase. Activated enzyme catalyzes the synthesis of more cAMP as long as G_s-GTP is bound to it.

G_q pathway is activated by calcium-mobilizing hormones and stimulates PLC- β to produce secondary messengers inositol triphosphate (IP3) and diacyl glycerol (DAG). Increased Ca^{+2} concentration by IP3 activates Ca^{+2} channels at the cell surface to allow influx of Ca^{+2}.

There is not much information about $G_{\alpha 12}$ and $G_{\alpha 13}$. G_{12} is thought to stimulate phospholipase D, c-Src, and PKC; however, the endpoint physiological responses are not fully understood. Two receptors, lysophosphatidic acid receptor and thromboxane A2 receptor have been identified for G_{13}. It is suggested that $G_{\alpha 13}$ directly interacts with a guanine nucleotide exchange factor for the GTPase Rho, leading to the regulation of Na^+-H^+ exchanger (Neves *et al*., 2002).

2.2 Plant Guanine Nucleotide Binding Proteins

Guanine nucleotide binding proteins in plants include small GTPases, heterotrimeric G proteins and several guanine nucleotide binding proteins, which are not members of either classes (Reviewed by Bischoff F., 1999).

Like mammalian GTP binding proteins, they are involved in many signal transduction mechanisms including regulation of ion channels, plant hormone signaling, Nod factor signaling, cell division, development and light signaling.

2.2.1 Small G proteins

Small GTPases are monomeric guanine nucleotide binding proteins with their masses are in the range between 21 to 30 kD. *Arabidopsis thaliana* genome is predicted to encode 93 small GTPases (Arabidopsis Genome Initiative, 2000).

Plants have 4 distinct classes of small G proteins including Ran, Arf, Rab, and ROP. Unlike mammalian genome, there are no Ras-GTPase homologs in plants according to the *Arabidopsis* genome sequence. Functional analyses based on loss of function mutations, dominant negative mutants, constitutively active mutants and overexpression have shown the potential functions of these proteins. Rab is the largest family of small GTPases and has 57 distinct members. Proteins in this class have potential roles in vesicle trafficking. Arf is another class and members have the role of protein trafficking between plasma membrane and cytoplasm. The proteins in the Ran class may be nuclear matrix attachment proteins. ROP (Rho-related GTPases from plants) evolved from the ancestor Rho, Rac, or Cdc42 GTPases according to the phylogenetic analysis. Various ROPs has been placed into 4 distinct groups. Their functional roles are control of pollen tube growth, root hair development, regulation of H_2O_2 production and negative regulation of abscisic acid (ABA) signaling (Reviewed by Yang, 2002; Asmann, 2002; Bischoff *et al*., 1999).

2.2.2 Unconventional GTP binding proteins

. Plants also have GTP binding proteins other than small GTPases and heterotrimeric G proteins. *Arabidopsis thaliana* extra large GTP-binding protein was identified as a result of comparison of all previously identified G_{α} proteins with the *Arabidopsis* EST DNA sequences using the TBLASTN program. The encoded protein is 99 kD and is expressed in all tissues including roots, leaves, stems, flowers and fruits. Carboxy-terminal half of AtXLG1 has 32 % identity and 52 % similarity with GPα1 (protein encoded by *GPA1*, the first plant α-subunit detected from *Arabidopsis thaliana*). Although there are missing and different amino acids at different positions in functional domains when compared with mammalian α-subunits, recombinant AtXLG1 binds GTP with specificity. There is no ADP ribosylation site. The amino-terminus half contains a cysteine-rich region and the amino acid sequence-DSITVSPT exactly matches with TonB box, which interacts with TonB proteins in the bacterial innermembranes to mediate energy transfer to the outer membrane for using TonB proteins in macromolecular transport (Lee *et al*., 1999).

Developmentally regulated G proteins (DRGs) are another type of unclassified GTP binding proteins in plants. They are expressed in all tissues especially higher level of transcripts and proteins are found in growing organs like root apices, elongating stems and growing auxiliary buds. They are primarily found in microsomal and soluble cell fractions and detected less in the fractions containing nuclei, mitochondria and plastids (Devitt *et al*., 1999).

Two additional smaller G proteins were identified and isolated from *Arabidopsis* and tomato, ATGB1 and ORFX, respectively. Based on phylogenetic comparisons, both are related to $G_{\alpha s}$ although they have smaller size than $G_{\alpha s}$. The function of ATGB1 is unknown. ORFX may affect fruit size by regulating cell division (Reviewed by Assmann, SM., 2002).

2.2.3 Plant heterotrimeric G proteins

Several genes, listed in table 2.1, encoding α-, β- and γ- subunits of heterotrimeric G proteins have characterized by molecular cloning from monocots and dicots. Similar to the mammalian case, they have different roles in a variety of signal transduction pathways regulating growth and development (Reviewed by Millner, 2001). The sequence based analysis of the complete genome of *Arabidopsis thaliana* has indicated that there are possible genes coding for the α-, β- and γ- subunits of heterotrimeric G protein (Arabidopsis Genome Initiative, 2000).

GPA1 is the first defined α-subunit from *A. thaliana*. Degenerate primers designed for the highly conserved parts of the mammalian and yeast G-protein α-subunits have been used to identify the α-subunit from the plant genomic DNA. As a result of screening of cosmid clones and cDNA clones with the radioactively labeled PCR product, genomic and the overlapping cDNA clones were isolated. Comparisons of these sequences have shown that *GPA1* is 1149 nucleotide long including 14 exons and 13 introns. *GPA1* encodes a protein called GPα1, which has 383 amino acid residues corresponding to a molecular weight of 44,482 Da. It is 36 % identical and ~ 73 % similar to rat G_{i1-3} and bovine rat transducin based on the amino acid sequence analysis (Ma *et al.,* 1990; Şahin, 2002). There is single mRNA for *GPA1* and according to the expression pattern GPα1 is detected in all organs and cell types. It is most abundant in vegetative tissues including leaves and roots, less in floral stems and least in floral buds and floral meristems (Ma *et al*., 1994; Ma H, 2001).

Gene	Species	Classification	Reference
GPA1	Arabidopsis	Gα	Ma et al., 1990
TGA1	Tomato	Gα	Ma et al., 1991
LjGPA1	Lotus	Gα	Poulsen et al., 1994
RGA1/D1	Rice	Gα	Ishikawa et al., 1995; Seo et al., 1995
SGA1	Soybean	Gα	Kim et al., 1995
SGA2	Soybean	Gα	Gotor et al., 1996
NtGPα1	Tobacco	Gα	Saalbach et al., 1999
NtGA2	Tobacco	Gα	Ando et al., 2000
LGPα1	Lupin	Gα	Kusnetsov and Oelmueller, 1996b
AtGα1	Wild oat	Gα	Jones et al., 1998
PGA1, PGA2	Pea	Gα	Marsh and Kaufman, 1999
SOGA1	Spinach	Gα	Perroud et al., 2000
NPGPA1	*Nicotiana plumbaginifolia*	Gα	Kaydamov et al., 2000
AGB1	Arabidopsis	Gβ	Weiss et al., 1994
ZGB1	Maize	Gβ	Weiss et al., 1994
TGB1	Tobacco	Gβ	Kusnetsov and Oelmueller, 1996a
RGB1	Rice	Gβ	Ishikawa et al., 1996
AtGβ1	Wild oat	Gβ	Jones et al., 1998
AtGβ2	Wild oat	Possible Gβ	Jones et al., 1998
NPGPB1	*Nicotiana plumbaginifolia*	Gβ	Kaydamov et al., 2000
AGG1	Arabidopsis	Gγ	Mason and Botella, 2000
AGG2	Arabidopsis	Gγ	Mason and Botella, 2001
GCR1	Arabidopsis	Potential heterotrimeric G protein receptor	Josefsson and Rask, 1997; Plakidou-Dymock et al., 1998
MLO	Barley	Potential heterotrimeric G protein receptor	Devoto et al., 1999
AtXLG1	Arabidopsis	Extra large GTP binding protein	Lee and Assmann, 1999
PsDRG	Pea	Developmentally regulated G protein	Devitt et al., 1999
AtDRG	Arabidopsis	Developmentally regulated G protein	Etheridge et al., 1999; Devitt et al., 1999
RDH3	Arabidopsis	Root hair defective (putative GTP-binding protein)	Schiefelbein and Somerville, 1990
ATGB1	Arabidopsis	GTP-binding protein	Biermann et al., 1996
fw2.2/ORFX	Tomato	Fruit weight 2.2 (putative GTP-binding protein)	Frary et al., 2000

Table 2.1 Plant heterotrimeric G protein subunits

GPα1, recombinant protein encoded by *GPA1* binds the guanine nucleotide with an apparent K_d within the range around 10^{-9} M for heterotrimeric G-proteins. In addition, a functional form of the protein was purified with 4aaGTP affinity separation. In the presence of divalent cations, GTPγS binding occurred effectively; especially with $ZnCl_2$ binding was found to be stimulated 2-fold (Wise *et al*., 1997).

Immunofluorescence and subcellular fractionation methods, used to isolate different membranes, have shown that GPα1 is associated with the plasma membrane and ER membrane. The possibility of association of GPα1 with mitochondria and Golgi apparatus has also been suggested. However, the level of the protein in these organelles, appears to much lower than on plasma and ER membranes. Membrane bound location of GPα1 may indicate not only involvement in transmission of extracellular signals across the plasma membrane and in the cytoplasm, but also in regulation of protein

transport from ER to Golgi and in docking and fusion processes of membrane trafficking (Weiss *et al*., 1997).

Overexpression of GPα1 from *Arabidopsis thaliana* has shown that it is involved in PhyA (primary photoreceptor for far-red light mediated inhibition of hypocotyl elongation) signal transduction mechanism through FHY1, PhyA specific signaling components upon induction by FR (far-red) light and PhyB (primary photoreceptor for red light mediated inhibition of hypocotyl elongation) signal transduction mechanism upon induction by R light (red). These results support involvement of GPα1 in the light regulation of *Arabidopsis* seedling development (Okomoto *et al*., 2001).

In addition to overexpression and subcellular fractionation experiments, *GPA1* mutants were used to understand the physiological and biochemical roles of the α-subunit in a variety of signaling pathways. Nonfunctional *GPA1* mutants created by T-DNA insertion have short hypocotyls length. This appears due to a reduced number of elongating cells, indicating impaired cell division. It is well known that auxin regulates cell division in plants. Indeed, control tobacco cells without *GPA1* overexpression required auxin to reach the level of cell division seen in the *GPA1*-overexpressing cells.

However, it is not known whether *GPA1* is involved in an auxin-dependent pathway or parallel pathway of cell-cycle regulation. In addition, *GPA1* mutants exhibited reduced brassinolide responsiveness, which is consistent with the rotundifolia-like leaf shape (Ullah *et al*., 2001). (Rotundofolia encodes cytochrome P450 that might be involved in brassinolide synthesis)

Nonfunctional *GPA1* mutants were also used to test ion channel regulation for stomata aperture and ABA signaling in guard cells. Work on *GPA1*-1 and *GPA1*-2 mutants, that do not express full length *GPA1* transcript, has shown that ABA inhibition of inward K^+ channels and stomatal opening require the presence of functional *GPA1*. In addition to K^+ channels, *GPA1* also regulates plant anion channels independent of the pH (Wang *et al.,* 2001).

Another evidence for the role of heterotrimeric G protein α-subunit in regulation of plasma membrane ion channels came from experiments done on tomato G-protein α-subunit. In isolated patches of membrane from GTPase deficient and wild type TGα1, mean "open state" probability of the plasma membrane Ca^{+2}-channel was increased. But this result is not sufficient to prove a direct physical interaction between the α-subunit and plasma membrane Ca^{+2} –channel (Aharon *et al*., 1998).

Seeds integrate many intrinsic signals through the action of hormones including brassinosteroids (BR), ethylene, abscisic acid (ABA) and gibberelic acid (GA) to control the germination. According to the comparisons between *GPA1* mutants with wild type *Arabidopsis* have shown that *GPA1* mutants are altered in sugar, ABA, GA and BR signaling during seed germination. Experiments were carried out by isolation of one hormone pathway from another to understand different hormone sensitivities of seed germination because these pathways overlap during signal transduction processes. BR is coupled by heterotrimeric G-protein to activate GA induced seed germination because BR mutants lacked BR rescue of GA dependent seed germination and *GPA1* mutants were completely insensitive to BR (Ullah *et al*., 2002).

Other G protein α-subunits were characterized from rice (Iwasaki *et al.,* 1997), soybean (Kim *et al*., 1995; Gotor *et al*., 1996), tobacco (Saalbach *et al*., 1999; Ando *et al.,* 2000), tomato (Ma *et al.,* 1991), pea (March and Kaufman, 1999), spinach (Perroud et al., 2000), rice (Seo *et al.,* 1995) and *Nicotina plumbaginifolia* (Kaydamov *et al.,* 2000).

Rice α-subunit defective mutant d1 has shown decreased α-amylase activity induced by GA. Low levels of expression of α-amylase and GA-induced genes were observed in the comparison of aleurone (special cells in the peripheral layer of the seed) cells of wild type and d1 mutant. In addition, increased concentration of GA above physiological limits was induced in the wild type phenotype. This result was inconsistent with the possible role of α-subunit of G protein in GA signaling. However, epistatic expression of SLR gene, which is the negative regulator of GA signaling and D1, proved that there might be two GA signaling pathways, one of which is G_α -dependent and the other is G_α-independent (Ueguchi-Tanaka *et al*., 2000).

Time-course analysis of SOGA1, which is the cDNA clone of α-subunit of G protein of spinach, has shown that SOGA1 transcript accumulation in green tissues is in a rhythmic manner under short-day conditions versus continuous light conditions. SOGA1 might have a role in light signaling for spinach (Perroud *et al*., 2000).

Heterotrimeric G-protein β-subunit genes were isolated from maize and *Arabidopsis thaliana.* They are single-copy genes and are named as ZGB1 and AGB1 for maize and *Arabidopsis,* respectively. They are approximately 41% identical with animal G protein β-subunits and contain seven copies of WD40 motif, which is the common property of β-subunits. 76% similarity between *ZGB1* and *AGB1* suggests that they have the same function. According to the expression patterns tested by Northern hybridization, *AGB1* was detected in the root, leaf and the flower (Weiss *et al.,* 1994).

The two plant γ-subunits, *AGG1* and *AGG2* have been cloned from *Arabidopsis thaliana*. They have conserved characteristics of γ-subunits like small size, C-terminal CAAX box and N-terminal α-helix region capable of forming a coiled-coil interaction with β-subunit. These genes code for a 98 amino acid peptide with a molecular weight of 10.8 kDa. Expression patterns for *AGG1* and *AGG2* are similar to the *AGB1*, showing comparable expression levels in roots, leaves and flowers. A strong interaction of *AGB1* with *AGG1* and *AGG2,* respectively has been defined with yeast two-hybrid system. Coiled-coil domain of AGB1 was detected as an essential part for interaction with AGG1 (Mason and Botella, 2000; Mason and Botella, 2001).

2.2.4 Plant G protein coupled receptor

There are three sequences annotated as putative GPCRs in *Arabidopsis* but the only one; GCR1 shows significant homology with the Dictyostelium cAMP receptor CAR1 within the heptahelical transmembrane region. GCR1 is predicted to have an extracellular N-terminus and a cytosolic C-terminus, as expected for GPCRs. In addition, an extracellular N-linked glycosylation site is found in the sequence of GCR1. It is expressed at a low level in leaves, stems, and roots (Josefsson and Rask, 1997).

Barley MLO defines the "founder" of this plant-specific family of integral membrane proteins with seven membrane-spanning helices. MLO proteins are unique to plants and they are organized in small gene families. 15 members of MLO were detected in the genome of *Arabidopsis thaliana* from sequence alignments with the known MLO family members defined for barley (Devoto *et al*., 1999).

2.2.5 Plant defense system and possible role of heterotrimeric G proteins

Like mammalian systems, plants have evolved sophisticated mechanisms to recognize a wide range of pathogens including bacteria, viruses, nematodes, insects and even other plants (McDowell and Dangl, 2000).

When there is an invasion of pathogen, a local reaction, called the hypersensitive response (HR), including cell wall reinforcement, production of anti-microbial metabolites and expression of defense-associated genes, is established. After the first response, a secondary and long lasting defense is established throughout the plant. This secondary reaction is called the systemic acquired resistance (SAR). Concentration of salicylic acid (SA) increased in the vicinity of infection causes both the activation of HR and SAR.

Transgenic *Arabidopsis* containing *NahG* gene from *Pseudomonas putida,* which converts SA to biologically inactive form catechol, was unable to show SAR indicating the importance of SA for establishment of SAR (Delaney *et al*., 1995).

SA-dependent pathways seem to be effective to biotrophic pathogens (use the host cell for feeding without killing the cell) while jasmonic acid (JA)/ethylene dependent defense is more effective for necrotrophic pathogens (kill the host cell) (Thomma *et al*., 1999).

Although antagonism between these pathways has been shown, a large number of genes that are coordinately up or down regulated by both SA and JA, have been identified by microarray analysis (Schenk *et al.,* 2000).

A number of signal transduction pathways are involved in the early response of the plant resistance system and it appears that G-proteins have important roles in transmitting the elicitor-receptor recognition that provides the activation of defense-related genes. Although the ligands and potential roles of putative GPCRs are not yet known, it appears that binding of elicitors to plant plasma membrane receptors triggers to the activation of heterotrimeric G-proteins. Activation leads to an increase in cytosolic Ca^{+2}-concentration by stimulating plasma membrane H^{+} -ATPase and inhibition of Ca^{+2} –ATPase. These events trigger downstream signaling cascades, including superoxide and H_2O_2 formation, which lead the cell death at the site of infection (Blumwald *et al*., 1998).

Elicitor induction of oxidative burst in soybean, parsley and *Arabidopsis* is inhibited by diphenylene iodinium, an inhibitor of mammalian NADPH oxidase, might prove the presence of plant homologue of NADPH oxidase-like complex (Legendre *et al.*, 1992).

In mammalian system, NADPH oxidase is the main source of reactive oxygen species, causing the oxidative burst, which leads to programmed cell death. In plant, two NADPH oxidase components ($gp91^{phox}$ and Rac) have been cloned. Ca^{+2}-binding motifs were found in $gp91^{phox}$, providing the regulatory connection to pathogen induced ion fluxes (Torres *et al*., 2002).

Tomato cells containing *Cfr5* resistance gene were also used to investigate involvement of redox processes in the production of active oxygen species associated with plant response to different fungal races containing and lacking *avr5*. NADPH oxidase and Cyt c oxireductase transfer electrons from ascorbate and ferricyanide and lead to production of superoxide anions. Addition of guanidine nucleotide analogues and mastarporan (stimulator of GDP/GTP exchange) result in increasing plasma membrane redox reactions might support the role of G-proteins for the activation of the

tomato defense mechanisms against fungal pathogen *C. fulvum* (Vera-Estrella *et al*., 1994).

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 Chemicals

All chemicals were supplied by Fluka (Switzerland), Merck (Germany), Riedel de Häen (Germany), and SIGMA (USA).

3.1.2 Primers

Primers were designed according to the coding sequence of *GPA1* (NCBI accession number: AC004484) reported by Ma H (Ma H *et al*., 1990). Integrated DNA Technologies, Inc., USA, synthesized primers without restriction enzyme sites. Primers with restriction enzyme sites were purchased from SEQLAB (Germany) and SIGMA (USA).

3.1.3 Enzymes

3.1.3.1 Restriction enzymes

*Eco*RI, *Xho* I, *Not* I, *Xma* I, *Xba* I, *Bam* HI, *Spe* I, *Sac* I (Promega and Fermentas)

3.1.3.2 Ligase

T4 DNA Ligase (Promega and Fermentas)

LigaFast™ Rapid Ligation System (Promega)

3.1.3.3 Taq polymerase

Taq DNA Polymerase in Storage Buffer A (Promega)

Pfu DNA Polymerase (Promega)

Herculase™ Enhanced DNA Polymerase (Stratagene)

3.1.3.4 Fill-in

DNA Polymerase I Large (Klenow) Fragment (Promega)

3.1.3.5 5'-dephosphorylation

Calf Intestine Alkaline Phosphatase (Fermentas)

3.1.4 Commercial Kits

PCR Core System II (Promega)

pGEM® -T and pGEM® -T Easy Vector Systems (Promega)

Qiaquick® PCR Purification Kit (250) (QIAGEN**)**

Qiaquick® Gel extraction Kit (250) (QIAGEN)

Qiaprep® Spin Miniprep Kit (250) (QIAGEN)

QIAGEN® Plasmid Midi Kit (100) (QIAGEN)

TOPO® XL PCR Cloning Kit (Invitrogen)

TOPO® TA Cloning Kit (Invitrogen)

TOPO® T7 Expression System (Invitrogen)

TOPO® pTrcHis Expression System (Invitrogen)

3.1.5 Vectors

Maps of all vectors can be found in Appendix A

pCIT 857 was kindly donated by Dr. Hong Ma (PennState University,USA)

pGEM® -T Easy (Promega)

pGEX-4T2 (Amersham Pharmacia)

pGFPuv (Clonetech)

pETM-11 (EMBL, Heidelberg)

pETM-30 (EMBL, Heidelberg)

pCR® II- TOPO® (Invitrogen)

pCR® -XL-TOPO (Invitrogen)

TOPO-T7 (Invitrogen)

TOPO-pTrc (Invitrogen)

3.1.6 Cells

Different *E. coli* strains containing TOP10, XL1 Blue, BL21 (DE3), BL21(DE3)pLysE, Rosetta(DE3), Rosetta(DE3)pLysS were kindly provided by EMBL, Hamburg

BL21-CodonPlus® (DE3)-RIL Competent Cells (Strategene)

TOP10 F' (Invitrogen)

BL21 (DE3) pLysS (Invitrogen)

3.1.7 Buffers and solutions

Standard buffers and solutions used in cloning and molecular manipulations were prepared according to Sambrook J and Russell DW, 2001

3.1.7.1 Culture medium

3.1.7.1.1 Liquid medium

LB Broth (Luria Bertani) from Sigma was used for liquid culture of bacteria. This mixture contains the tryptone, yeast extract, and sodium chloride, which are mixed in appropriate amounts. 20 g of LB Broth was used for preparation of 1 L liquid medium. The liquid medium was autoclaved at 121^0C for 20 min. before using.

3.1.7.1.2 Solid medium

LB Agar (Luria-Bertani) from Sigma was used for preparation of solid medium for the growth of bacteria. This mixture contains the tryptone, yeast extract, sodium chloride, and agar, which are mixed in appropriate amounts. 40 g of LB Agar was used for preparation of 1 L solid medium. The appropriate amount of LB Agar is dissolved in correspondent amount deionized water for autoclaving at 121^0C for 20 min. Autoclaved medium was poured to petri plates (~20 ml/plate) after cooling down to the room temperature.

3.1.7.2 Buffers for denaturing polyacrylamide gel electrophoresis

All buffers for gel electrophorase were prepared according to the protocols from Maniatis *et al.,* 1989

1X TAE (Tris -acetate EDTA)

1XTBE (Tris-borate EDTA)

1X Tris Glycine-SDS

3.1.8 Sequencing

Sequencing service was commercially provided by SEQLAB (Germany) or MWG-The Genomic Company (Germany).

3.1.9 Equipment

All equipments used in this study can be found at Appendix D

3.2 Methods

3.2.1 Culture Growth

Cells were grown overnight (12-16h) in LB Broth (Luria Bertani) medium prior to any application. LB Agar (Miller's LB agar) solid medium was used as selective and unselective solid medium for the growth of bacteria.

Protocols for liquid and solid culture growth and the other applications including competent cell preparation, glycerol stocks were done according to Maniatis *et al*.,1989 and Sambrook J and Russell DW, 2001.

3.2.2 PCR

Recommended reaction volumes and final concentrations of the PCR Core System II components were used for PCR reaction mixture. Annealing temperatures of primers with and without restriction enzyme sites were estimated according to the formula in the PCR Core System II (Promega). Either *Eco* RI digested *GPA1* or pCIT 857 was used as template. Following PCR conditions was used for amplification process.

1. 94^{o} C…1 min.
2. 94^{o} C…1 min.
3. 53^{o} C…1 min. **30 cycles**
4. 72^{o} C…1 min.
5. 72^{o} C…1 min.
6. 4^{o} C…HOLD

3.2.3 Purification of PCR product

PCR product was purified either from the 1% Agarose gel with Qiaquick® Gel extraction Kit (250) (QIAGEN) or directly with Quiaquick® PCR Purification Kit (250) (QIAGEN).

3.2.4 Cloning

Basic procedures were carried out according to Manniatis *et al*., 1989.

3.2.4.1 Subcloning

Amplified *GPA1* was subcloned into pGEM®T-Easy vector (Promega), pCR® II-TOPO vector (Invitrogen) and pCR® -XL-TOPO (Invitrogen) according to the optimized protocols of each kit.

3.2.4.2 Ligation

PCR amplified and purified *GPA1* containing different restriction enzyme sites was ligated with pGEM ®T-Easy vector (Promega) in such a way that 3:1 insert: vector ratio was used. The reaction mixture was incubated overnight at 4^{o} C.

PCR amplified and gel purified *GPA1* containing different restriction enzyme sites was ligated into pCR® II- TOPO® (Invitrogen), pCR® -XL-TOPO (Invitrogen) vectors, individually. Reaction mixture was incubated at least 30 min. at room temperature (~25^{0} C) and 1µl 6X TOPO® Cloning Stop Solution was added to stop the ligation reaction.

3.2.4.3 Transformation

Ligation mixtures were transformed into different endonuclease deficient strains of *E. coli*- XL1 Blue, TOP10. Transformed cells and controls were plated on appropriate antibiotic selective LB plates prepared according to the ligation vectors.

3.2.4.4 Colony Selection

Positive colonies were selected and grown on liquid LB culture containing appropriate antibiotic for both preparing glycerol stocks and plasmid isolation.

3.2.4.5 Plasmid isolation

Plasmid isolation was done either with Qiaprep® Spin Miniprep Kit (250) (QIAGEN) or following to the alkaline lysis protocol from Maniatis *et al.,* 1989.

3.2.4.6 Restriction enzyme digestion

Purified plasmids containing *GPA1* were digested with appropriate restriction enzymes according suppliers instructions to verify the presence of *GPA1*. (Enzyme/reaction mix) v/v ratio was kept at 1/10 or smaller in all digestions.

3.2.4.7 DNA analysis

Purified plasmids and digested plasmids were analyzed by agarose gel electrophoresis. Appropriate DNA markers were used for size and concentration

determination. In addition, concentration and $OD_{260/280}$ ratio were monitored by absorption measurements.

3.2.4.8 Frozen stocks of cells

Frozen stocks of *E. coli* containing different plasmids with *GPA1* were prepared in 15% glycerol in LB with antibiotics and kept at -80^{o} C according to the protocol from Maniatis *et al*, 1989.

3.2.4.9 Sequence verification

QIAGEN® Plasmid Midi Kit (100) (QIAGEN) purified plasmids containing *GPA1* were sent for sequence analysis. Plasmids were checked by restriction and electrophoretic analysis before sequencing.

3.2.5 Directional cloning using expression vectors

GPA1 was digested from subcloning vectors with appropriate restriction enzymes and ligated into different expression vectors containing pGEX-4T2, pGFPuv, pETM-11, pETM-30 digested with the same restriction enzymes. Different fragment: vector ligation ratios (1:1, 3:1, 5:1, 10:1, 20:1, 50:1, 1:3, and 1:5) were used. Ligation mixtures were transformed into TOP10 or XL1 Blue cells and transformants were checked with restriction enzyme digestion for the presence of *GPA1*. Glycerol stocks of *GPA1* for positive colonies were done. Purified plasmids were sent for sequencing.

Meanwhile, purified plasmids were transformed into different expression competent cells containing BL21 (DE3), BL21 (DE3) pLysS, Rosetta(DE3), Rosetta(DE3)pLysS, BL21-CodonPlus® (DE3)-RIL to express *GPA1*.

3.2.6 Cloning directly into expression vectors

PCR amplified *GPA1* without restriction enzyme sites was directly ligated into TOPO®T7/NT vector (Invitrogen) and TOPO® pTrcHis vector (Invitrogen), separately.

Ligation and transformation reactions were done according to the protocol supplied in the kit. Transformation was done with TOP10 F' competent cells. After verification of the *GPA1* in the expression vector with restriction enzyme digestion, construct was transformed into BL21(DE3)pLysS expression cells. Construct was also verified by sequencing the insert.

3.2.7 Expression

GPA1 within different expression vectors including pGEX-4T2 (Amershiam), pGFPuv (Clonetech), TOPO® T7/NT vector (Invitrogen) and TOPO® pTricHis vector (Invitrogen) was expressed according to the protocol from Maniatis et al.,(1989) and Sambrook J and Russell DW,2001. Different IPTG final concentrations for inducing recombinant protein expression were used ranging from 0.5mM to 1mM. Aliquots corresponding a total OD_{600} of 1.4 were taken from induced cells.

Cells were pelleted and stored at -20^0 C. Pellets were prepared for SDS-PAGE gel according to the protocol from Maniatis et al., 1989 and Sambrook et al., 2001. 10 µl of the samples were loaded on 5%-12% SDS-PAGE gels. The gels were run at 30mA constant current for 1 h and stained with Coomasie blue for protein bands were observed after destaining solution overnight.

Protein molecular weight markers (Fermentas) were used to identify the molecular weights of expressed proteins observed on the gel.

4 RESULTS

4.1 Localization of *GPA1* on pCIT 857

Plasmid pCIT857 had been constructed by inserting the longest *GPA1* cDNA into the *Eco* RI site of Promega vector pGEM 7Zf(+) (Ma *et al.*,1990). pCIT 857 a gift by Ma, H. (PennState University, USA), was digested with *Eco* RI to verify the presence of *GPA1* fragment. Figure 4.1 shows a preparative agarose gel electrophoresis analysis of pCIT857 digested with *Eco* RI. A fragment of approximately 1300 bp was indicated on the gel. It is expected that 1149 bp region of this fragment would corresponds to *GPA1* whereas the remaining 227 bp is the noncoding part which comes from 5' and 3' flanking regions of the chromosome II of *A. thaliana.*

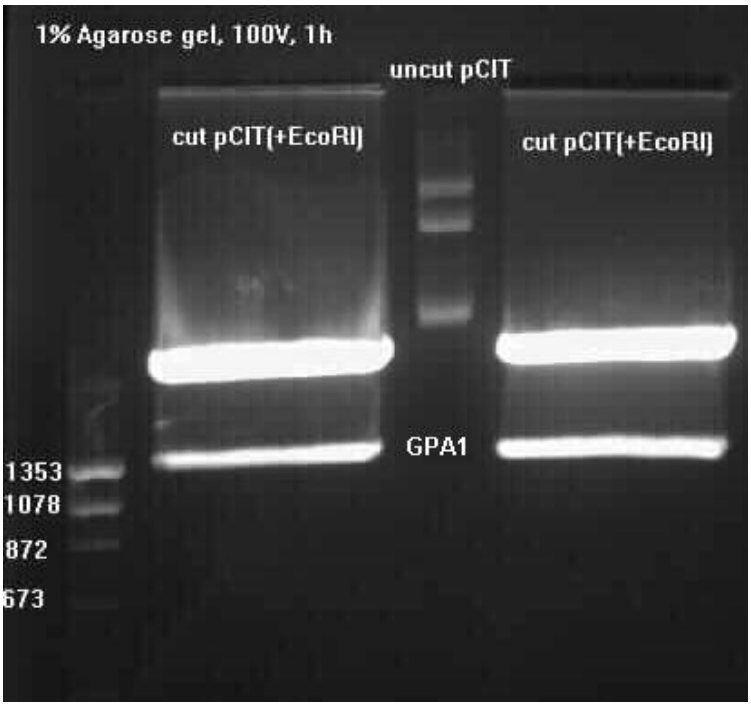


Figure 4.1 1% Agarose gel showing *GPA1* insert in pCIT 857

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

The presence of *GPA1* in pCIT 857 was also checked by sequencing (Apendix C.1). Sequence analysis for *GPA1* fragment and restriction enzyme sites were carried out using the Biology Workbench (http://workbench.sdsc.edu) providing integrated DNA and protein database searching.

Results of restriction mapping approximately of the 1800bp region, which was sequenced, is shown in table 4.1

AatII	1	BslI	2	HincII	1	PspGI	1
AccI	1	BsmAI	4	HindIII	3	PspOMI	1
Acc65I	1	BsmFI	1	HinfI	6	PsrI	1
AclI	2	Bsp1286I	3	HpaII	5	PsrI	1
AflIII	1	BspCNI	2	HphI	2	PstI	1
AloI	1	BspEI	1	Hpy8I	4	PvuII	1
AloI	1	BsrI	5	Hpy99I	3	RsaI	5
AluI	9	BssSI	1	Hpy188I	9	SacI	1
AlwI	4	BstBI	1	Hpy188III	9	SalI	1
AlwNI	1	BstF5I	4	HpyCH4III	2	Sau96I	4
ApaI	1	BstKTI	2	HpyCH4IV	6	ScrFI	3
ApoI	3	BstNI	1	HpyCH4V	7	SfaNI	1
AvaI	2	BstUI	1	HpyF10VI	4	SfcI	4
AvaII	2	BstXI	1	KasI	1	SfoI	1
BamHI	2	BstYI	2	KpnI	1	SmaI	1
BanI	2	Cac8I	3	MaeIII	2	SmlI	1
BanII	2	ClaI	1	MboI	2	SphI	2
BbeI	1	Csp6I	5	MboII	10	StyI	2
BbvI	2	CviJI	19	MfeI	1	StyD4I	3
BbvCI	1	DdeI	5	MluI	1	TaiI	6
BccI	5	DpnI	2	MlyI	2	TaqI	9
BceAI	2	EarI	1	MmeI	2	TaqII	1
BfaI	2	Eco57I	3	MnlI	10	TatI	1
BfrBI	2	EcoICRI	1	MseI	1	TfiI	4
Bme1580I	1	Eco57MI	5	MslI	1	TseI	2
BmgBI	2	EcoO109I	1	MspA1I	1	Tsp45I	1
BmrI	1	EcoRI	2	MwoI	4	Tsp509I	11
BpmI	2	EcoRV	1	NarI	1	TspDTI	5
Bpu10I	1	FatI	6	NciI	2	TspGWI	2
BsaI	1	Fnu4HI	2	NlaIII	6	Tth111I	1
BsaHI	2	FokI	4	NlaIV	8	XbaI	1
BsaJI	4	HaeII	1	NsiI	2	XhoI	1
BsaWI	2	HaeIII	2	NspI	2	XmaI	1
BseMII	2	HhaI	1	PfoI	1	XmnI	2
BseRI	3	Hin4I	4	PleI	2	ZraI	1
BseYI	1	Hin4I	4	PpuMI	1		
BsiHKAI	2	HinP1I	1	PsiI	1		

Table 4.1 Restriction enzyme mapping of approximately 1800 bp region of pCIT857

4.2 PCR amplification of *GPA1*

4.2.1 PCR amplification for subcloning

GPA1 was amplified using primers designed without restriction enzyme sites

Forward primer (G1): 5'- ATG GGC TTA CTC TGC AGT-3' and

Reverse primer (G3): 5'-TCA TAA AAG GCC AGC CTC-3'

G1 and G3 were used for subcloning *GPA1* using pGEM®T-Easy, pCR® II TOPO and pCR®XL TOPO vectors. These primers were also used for cloning of *GPA1* directly into expression vectors T7/NT TOPO and pTrcHis TOPO vectors.

Both *Eco* RI digested and gel purified *GPA1* from pCIT 857 and the intact pCIT 857 were used as template for PCR amplification. PCR conditions were basically as described in section 3.2.2 with 1 μM final primer concentration.

In later experiments clone #44, which is pGEM®T-Easy vector containing *GPA1* was also used as template for amplification. *Taq* DNA Polymerase was used for A-tailing ligation and *pfu* DNA polymerase, Herculase™ Enhanced DNA Polymerase were used for blunt end ligation.

Results of amplification of *GPA1* using different templates were monitored by agarose gel electrophoresis. Some typical examples are shown in Figure 4.2. For all templates, a fragment, approximately 1149 bp long, corresponding to the size of the coding region of *GPA1* was seen on the gel. The size of the amplified DNA (~1150bp) is less than the size of digested *GPA1* (~1300bp) from pCIT857. This result also verified that only the coding part of *GPA1* was amplified (see also subcloning results).

4.2.2 PCR amplification for insertion into expression vectors

Primers containing different restriction enzyme sites were designed for in-frame insertion of *GPA1* into expression vectors including pGEX-4T2, pGFPuv , pETM-11 and pETM-30. Primer sequences, restriction sites and the corresponding expression vectors are shown in table 4.2. PCR carried out using these primers also resulted in amplification of a fragment of about 1150 bp long similar to that shown in Figure 4.2 (data not shown).

vector name	Forward Primer	Reverse Primer
pGEX-4T2	5'- GCG TC G AAT TCC CAT GGG CTT ACT CTG-3'	5'-AAA CCC CTC GAG TCA TAA AAG GCC A-3'
pGFPuv(5'MCS)	5'- AAA CCC TCT AGA GAT GGG CTT ACT CTG-3'	5'- AAA CCC CCC GGG GTA AAA GGC CAG-3'
pGFPuv(3'MCS)	5'-AAA CCC GAG CTC ATG GGC TTA CTC T-3'	5'-AAA CCC ACT AGT TCA TAA AAG GCC A-3'
pETM11	5'- GCG TCG AAT TCG ATG GGC TTA CTC TG-3'	5'-AAA CCC CTC GAG TCA TAA AAG GCC A-3'
pETM30	5'- GCG TCG AAT TCG ATG GGC TTA CTC TG-3'	5'-AAA CCC CTC GAG TCA TAA AAG GCC A-3'

Table 4.2 Primer sequences, restriction sites and the corresponding expression vectors

Restriction sites are underlined

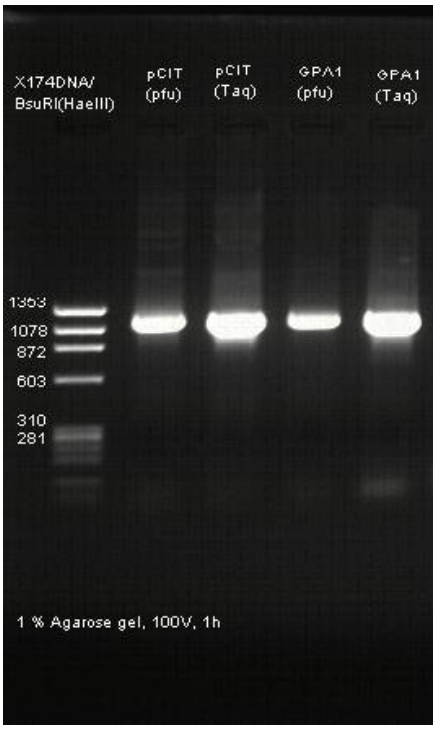


Figure 4.2 1% Agarose gel showing results of PCR carried out using different templates and different DNA polymerase enzymes

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph

4.2.3 PCR amplification at different annealing temperatures

In order to check nonspecific amplification annealing temperature was varied in PCR. Additional bands other than the DNA band for *GPA1* have shown that, unspecific primer binding to pGEX-4T2 could occur at the PCR conditions used (Figure 4.3)

Optimal Global Alignment results between pGEX-4T2 and forward and reverse primers, respectively by using the global aligment tools from Biology Workbench (http://workbench.sdsc.edu) have been shown that additional DNA bands at the positions of approximately at the 3000 bp, 2000bp, 1200 bp, 800 bp and 700 bp bands might be obtained under non-stringent PCR conditions (When problems with insertion into expression vectors were encountered possible roles of these nonspecific fragments as contaminants were investigated, see Chapter 5 Discussion).

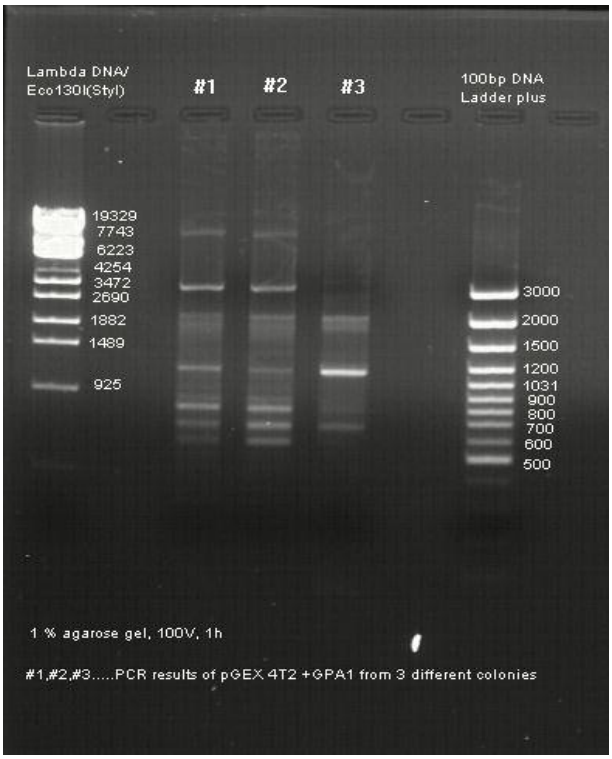


Figure 4.3 Colony PCR results with pGEX-4T2+*GPA1* construct

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

4.3 Subcloning and Sequence Verification of GPA1 in pGEM-Teasy, pCR-XLTOPO and pCR-II TOPO vectors

For sequence verification all amplified *GPA1* fragments were first inserted into pGEM®T-Easy vector and JM109 cells were transformed with the construct. *GPA1* positive colonies were selected by blue-white colony screening. Two colonies selected for sequencing were labelled as # 33 and #44 (construct map of #44 is in Appendix B1)and plasmids isolated from liquid cultures of these cells are shown in Figure 4.4. pGEM®T-Easy vector facilitates A-tailing ligation of the fragment between two flanking *Eco* RI sites and presence of the insert can be verified by digestion of the isolated plasmids with *Eco* RI. Results of digestion revealing approximately 1100 bp

fragments are shown in Figure 4.4. Subsequently clones # 33 and #44 were sequenced verifying the presence of *GPA1* in the pGEM®T-Easy vector. Alignment results of # 33 and #44 with *GPA1* are given in Appendix C.2.

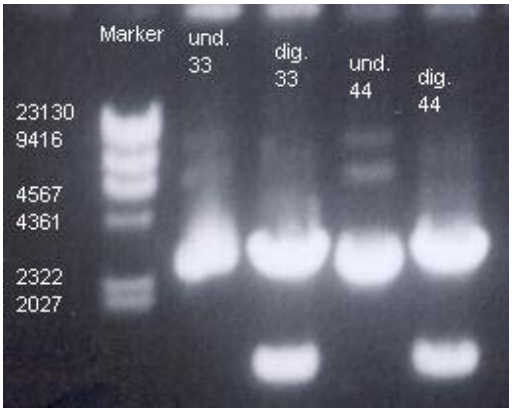


Figure 4.4 Restriction enzyme digestion analyses of constructs # 33 and #44 with *Eco* RI

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

GPA1 amplified with primers containing restriction enzyme sites designed for different expression vectors were also subcloned in XL1 Blue cells using pCR®XL TOPO and pCR®-II TOPO vectors. Constructs were isolated and digested with respective enzymes to show the intact presence of the of the restriction sites.
Figure 4.5 shows the electrophoretic analysis of TOPO1 to TOPO8 clones. These constructs were pCR®-XL TOPO +*GPA1* containing *Eco*RI and *Xho* I restriction enzyme sites for insertion into pGEX-4T2 vector. As expected when the constructs were digested only with *Xho* I, the linearized vector at 4.5kb was obtained. *Xho* I and *Eco* RI double digestion on the other hand yielded the 1100 bp *GPA1* fragment. Further sequence analyses confirmed the uncorrupted amplification of *GPA1* before insertion into pGEX-4T2. Sequencing results and construct map for TOPO8 are given in Appendix C.3 and Appendix B.2, respectively)

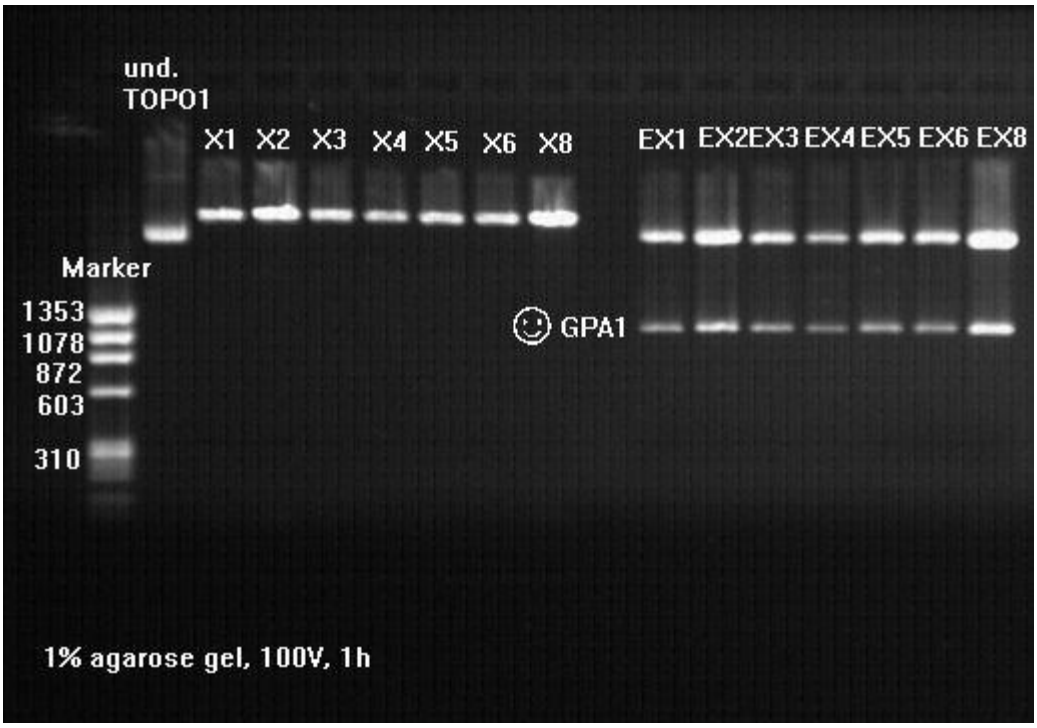


Figure 4.5 1% Agarose gel analysis of digestion of TOPO1 to 8 with *Xho* I, and *Eco* RI and *Xho* I

X1 to X8 correspond to digested TOPO1 to 8 with *Xho* I. EX1 to EX8 correspond to digested TOPO1 to TOPO8 with *Eco* RI and *Xho* I. DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

pCR®-II TOPO vector was used for subcloning *GPA1* amplified with the primers containing *Spe* I (3') and *Sac* I (5') restriction enzyme sites and the ones containing *Xma* I(3') and *Xba* I (5'). These fragments were designed for insertion into 3' and 5' MCS respectively of the pGFPuv expression vector. The clones containing *GPA1* with *Spe* I and *Sac* I restriction enzyme sites were labeled as P1 and P20 and XX1 was used as a label for the clones containing *GPA1* with *Xma* I and *Xba* I restriction enzyme sites.

Presence of *GPA1* in the pCR®-II TOPO vector was controlled by restriction enzyme analysis with the corresponding enzymes as shown in Figure 4.6.

A **B**

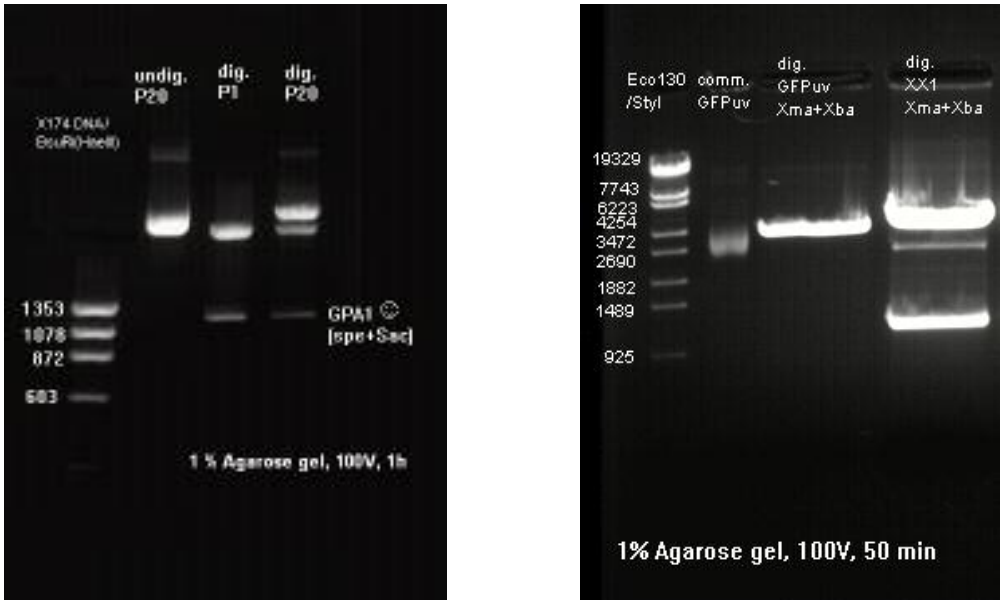


Figure 4.6 A and B Restriction enzyme analyses showing the presence of *GPA1* in pCR®-II TOPO vector

P1 and P20 are the constructs for insertion into 3' MC site of pGFPuv whereas XX1 is the construct for the 5' MC site insertion into the pGFPuv vector. DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

Sequence analysis results for P1 and P20 are in given in Appendix C.4. Construct map for P1 is given in Appendix B.3.

4.4 Cloning of *GPA1* using expression vectors pGEX-4T2, pGFPuv,pT7/NT TOPO and pTrcHis TOPO

General strategy followed for cloning using expression vectors is was digestion of the *GPA1* fragment out of the subcloning vector using the appropriate restriction enzymes followed by ligation with the expression vector prepared by digestion with the same enzymes. Various stages of this procedure for cloning with pGEX-4T2 vector are illustrated in figures 4.7 A, B and C. GPA1 digested from TOPO8 construct, shown in Figure 4.7 B was ligated with the vector pGEX-4T2/Eco RI-Xho I, shown in figure 4.7

A. Relative concentrations of vector and insert were determined by gel electrophoresis as shown in Figure 4.7 C.

A **B**

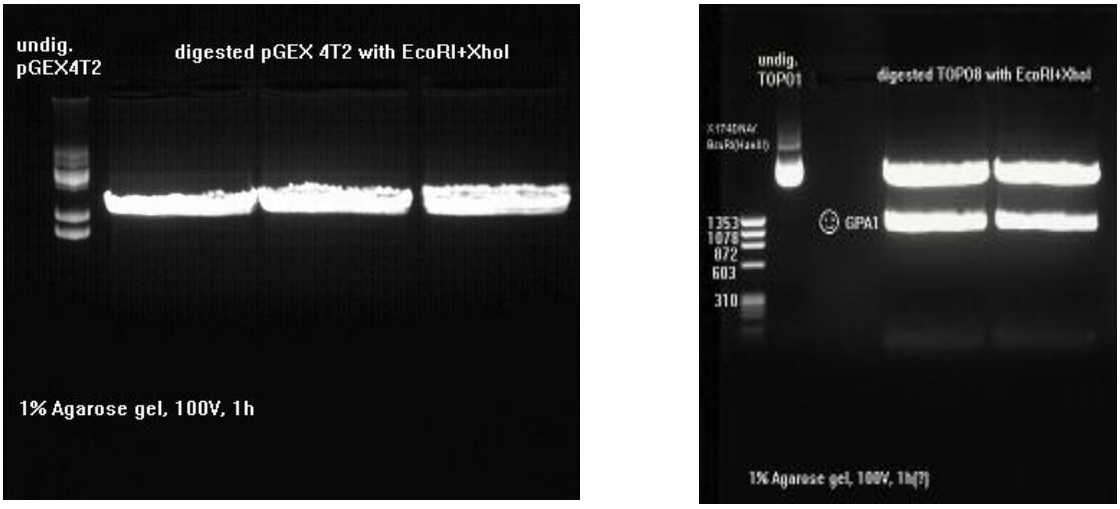
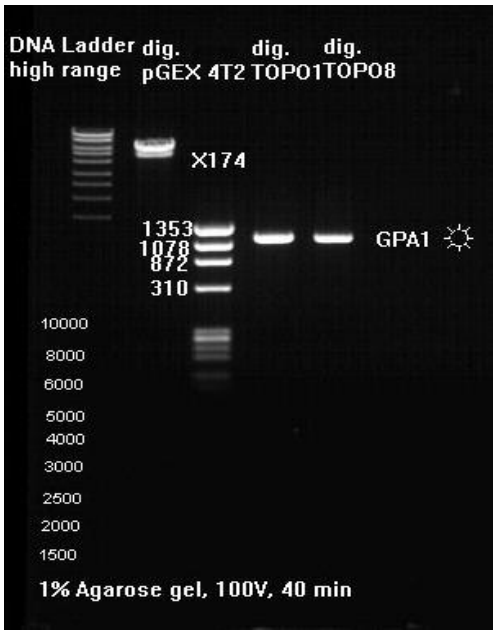


Figure 4.7 Double digestions of pGEX-4T2 and TOPO vectors

A-Double digested pGEX-4T2 with *Eco* RI and *Xho* I
B-Double digested TOPO1 and TOPO8 with *Eco* RI and *Xho* I



C-Eco RI and Xho I digested and gel purified pGEX-4T2 and TOPO1, TOPO8 to determine molar ratios of fragment to vector before ligation. DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

Possible positive clones (T1/2 to T1/5 and T8/1 to T8/5) were selected and analyzed for the presence of *GPA1* with *Eco* RI and *Xho* I digestion.

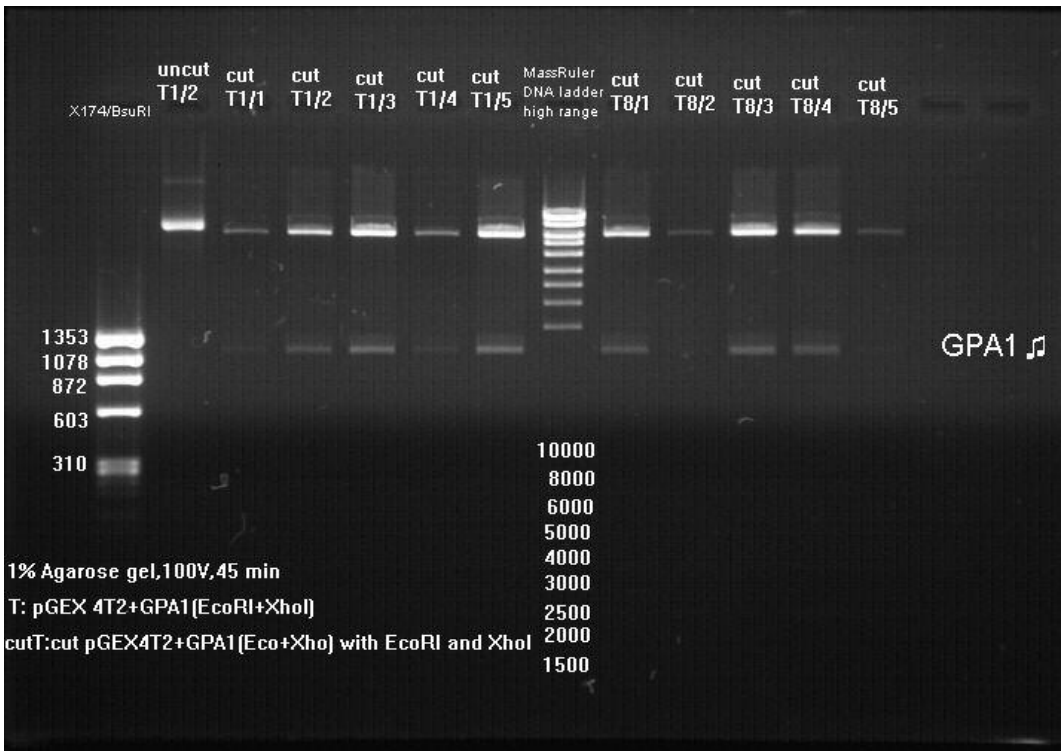


Figure 4.8 Double digest of T clones (pGEX-4T2+*GPA1*) with *Eco* RI and *Xho* I to verify the presence of *GPA.1*

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

Presence of *GPA1* was detected at the expected size except the clones T8/2 and T8/5. *GPA1* positive constructs were transformed into BL21 (DE3) cells for expression. It was however noticed that forward primer for this clone was wrongly designed and that the gene was not inserted into pGEX-4T2 in frame with the start codon of the GST fusion protein. New primers were designed and cloning and subcloning were repeated. However, although PCR amplification and subcloning using pCRII TOPO vector gave positive resuts it was not possible to insert *GPA1* fragment in frame into the pGEX-4T2 between the *Eco* RI and *Xho* I restriction enzyme sites (see Chapter 5 Discussion).

For cloning using pGFPuv vector *GPA1* digested from the P1 construct as indicated in Figure 4.9 A was ligated into 3' MCS. Ratio of insert to vector was determined from the agarose gel shown in figure 4.9 B. This construct was labeled as Clone #4 and transformed into XL1 Blue cells. Restriction analysis of isolated plasmids shown in Figure 4.10 verified the insert.

A **B**

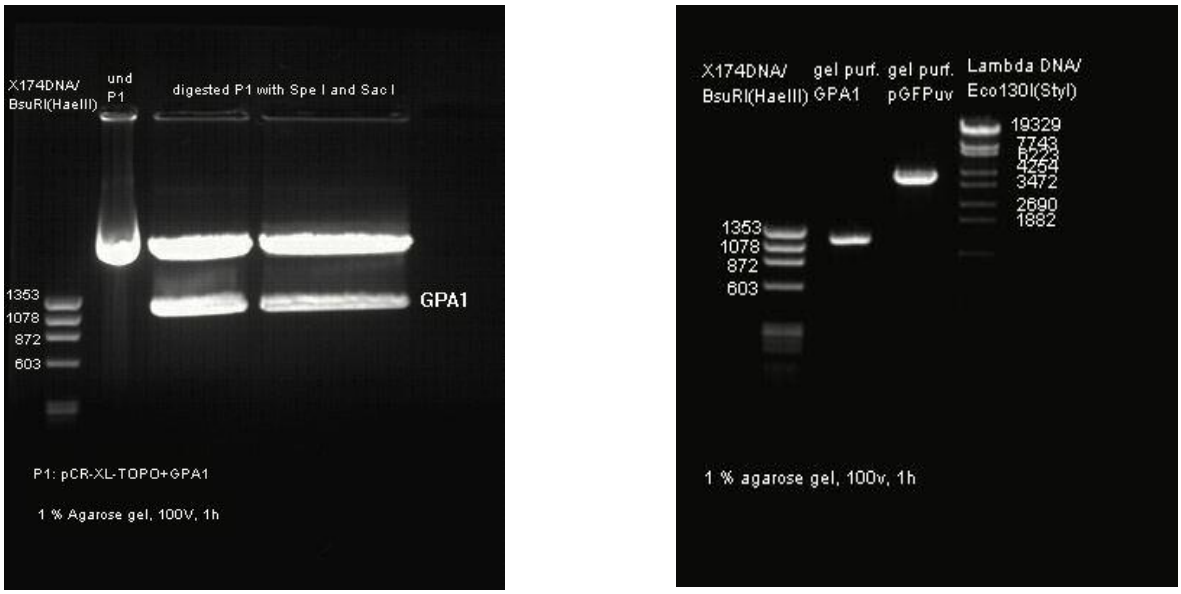


Figure 4.9 Double digestion of P1 and pGFPuv with *Spe* I and *Sac* I

A- *Spe* I and *Sac* I digested P1

B- *Spe* I and *Sac* I digested and gel purified pGFPuv and *GPA1* to determine fragment to vector ratio for ligation

Figure 4.10 *Spe* I and *Sac* I digested construct #4 to check the presence of *GPA1*

Construct #4 was further transformed into BL21(DE3) expression cells and the clone labeled 4/4_3'MCS was sequenced. Sequencing results and construct map are given in Appendix C.4 and Appendix B.4, respectively. Analyses and alignments of 4/4_3'MCS has shown that the sequence of the 3' -end agrees completely with *GPA1* 3'-end sequence for about 180 , the *Spe I* restriciton site is intact and the pGFPuv sequences flanking the *GPA1* 3'-end are also present. This result is in agreement and complementary with the results of restriction enzyme analysis shown figure 4.10. The 5'-end of the insert could not be sequenced unambiguously, however a previous sequencing of the same construct had verified the GFPuv region of the construct (data not shown) (see Chapter 5 Discussion).

GPA1 was also inserted into direct cloning vectors pT7/NT TOPO and pTrcHis TOPO to avoid extensive handling of PCR fragments as well as ligation problems. PCR was conducted with low concentrations of the template (to avoid false positives during cloning) and results were controlled by agarose gel electrophoresis as can be seen in Figure 4.11.

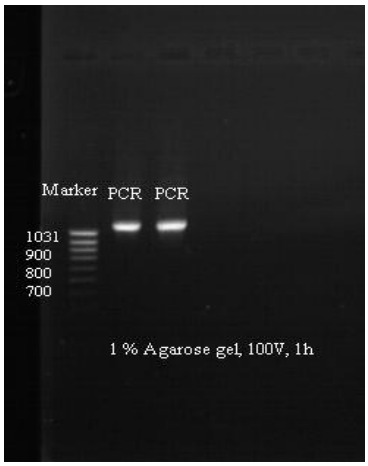


Figure 4.11 Analysis of results of PCR for direct cloning into expression vectors

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

As can be seen in figure 4.11, no template bands were visible and the size of the PCR fragment was as expected. PCR products were directly ligated into the expression vector and TOP10F' cells were transformed with the construct. Presence and direction of the inserts were verified by restriction enzyme analysis of isolated plasmid with *Bam* HI.

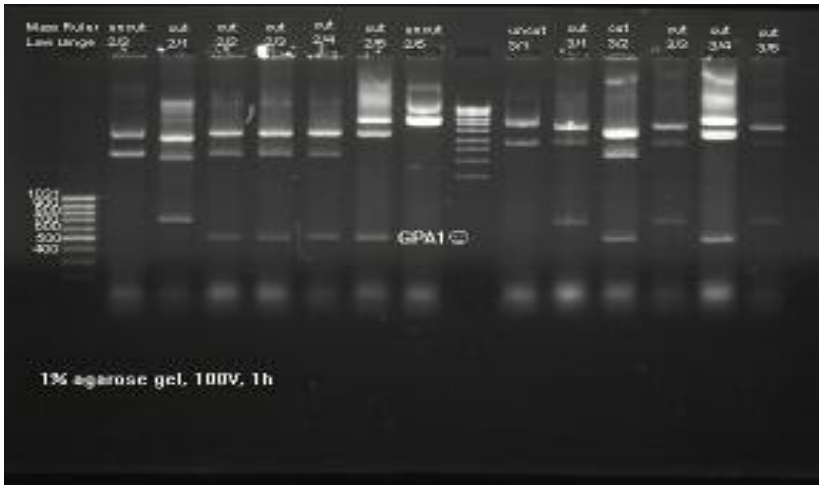


Figure 4.12 Restriction enzyme analysis of pT7/NT TOPO+ *GPA1* and pTrcHis TOPO+ *GPA1* constructs

DNA molecular weight markers, different samples and electrophoresis conditions are indicated on the photograph.

As illustrated in figure 4.12 directional inserts yielded about 450 bp fragments. *GPA1* positive clones were labeled as constructs 2/3 (pT7/NT clone) and 3 /4 (pTrcHis clone) were subjected to sequence analysis. Results are given in Appendices C.6 and C.7. Additionally construct maps for 2/3 and 3 /4 are given in Appendix B.5 and B.6, respectively. The usual sequencing was encountered with the construct 2/3. A region more 1000 bp long could be sequenced with construct 3 /4 and this showed almost perfect agreement with *GPA1* at the 5' end. 3' end was again problematic (see Chapter 5 Discussion). Although restriction enzyme analysis has verified the directional presence of *GPA1* in these expression vectors false positives due to contamination with pCIT 857 carried from PCR would also have given a result similar to that expected from pTrc_*GPA1* and pT7_*GPA1* (Table 4.1 and vector maps in Appendix B) by using same restriction enzymes. However the correct construct was verified by sequencing data.

4.5 Expression of *GPA1*

4.5.1 Expression using pGEX-4T2

pGEX-4T2 was used to express GPA1 as a fusion protein with GST and figure 4.13 shows SDS PAGE analysis of the cell extracts of the clone T1/3 in BL21 (DE3) with 1 mM final IPTG concentration at 2 hours after induction. Lanes 1 and 2 are extracts from cells without the construct and +IPTG and –IPTG, respectively. Lanes 5 and 6 are extracts from cells with the construct (T1/3) and +IPTG and –IPTG, respectively. Lanes 7 and 8 are the samples from the pellet fraction of +IPTG and -IPTG T1/3. Size of GST is about 27 KDa which would make the expected size of the fusion protein to be about 70 KDa . The strong bands at about 30 KDa in lanes 5 and 6, are the induced recombinant GST protein. The additional bands above the GST band in lanes 7 and 8 cuold be the recombinant protein with a premature termination codon due to the wrong forward primer.

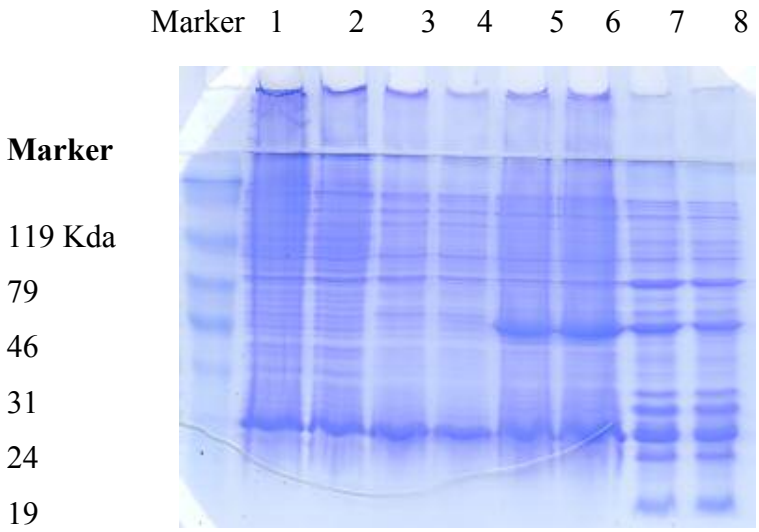


Figure 4.13 SDS PAGE analysis of the induction of T1/3 clones

As mentioned above it has not been possible to obtain expression with correctly inserted *GPA1*

4.5.2 Expression using pGFPuv

Expression with pGFPuv vector would produce a GFP-GPA1 fusion protein and would faciliate localization of *GPA1*. Results of expression experiments are shown in figure 4.14. Lanes 1 and 2 are +IPTG and -IPTG induced of BL21 (DE3) cells without the construct, respectively. Lanes 3 and 4 are +IPTG and -IPTG induced of BL21 (DE3) cells with pGFPuv vector without *GPA1*, respectively. Lanes 5 and 6 are +IPTG and -IPTG induced of BL21 (DE3) cells with 4/4_3'MCS, respectively. Lanes 7 and 8 are the samples from the pellet fraction of +IPTG and -IPTG 4/4_3'MCS.

Expression of GFPuv, which is a 26 KDa protein, was detected in lanes 3 and 4. But there was no recombinant protein (approximately 70 KDa) detected in the lanes from 5 to 8 corresponding the samples from supernatant and pellet fractions of the transformed cell lysates.

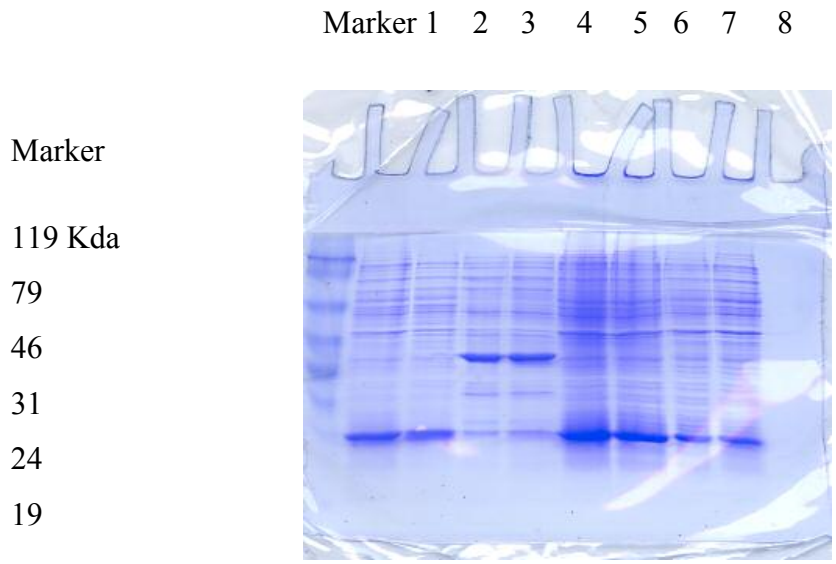


Figure 4.14 Induction analysis of 4/4_3'MCS (pGFPuv+*GPA1*) in BL21(DE3) 4 hours after induction

OD_{600} values regularly increased during expression proving the viability of the cells. Although we could not observe green colour in cells containing the construct green fluorescence was observed with the control cells expressing GFPuv alone.

4.5.3 Expression using pT7/NT TOPO and pTrcHisTOPO vectors

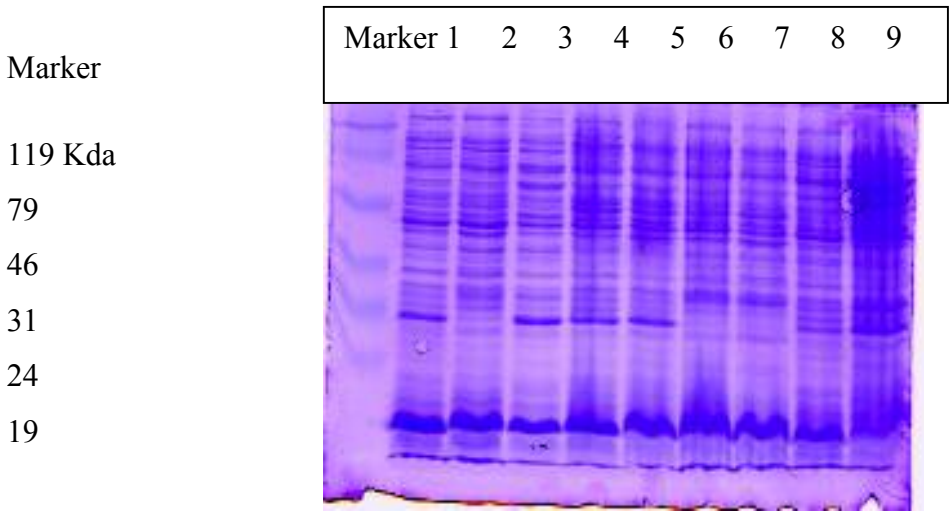


Figure 4.15 Induction analysis of 2/3 (pT7/NT TOPO +*GPA1*) in BL21 (DE3)pLysS

Lane1 is BL21(DE3)pLysS, lane 2 is 2/3 and lane3 is positive control before induction. Lanes 5 and 6 are +IPTG and –IPTG BL21(DE3) at 2 hours after induction, respectively. Lanes 7 and 8 are +IPTG and –IPTG 2/3 at 2 hours after induction, respectively. Lanes 9 and 10 are +IPTG and –IPTG positive controls at 2 hours after induction, respectively. The expected size of the fused protein is about 50 KDa but there was no specific fusion protein observed in lanes 7 and 8 as indicated figure 4.15.

Additionally figure 4.15 shows that approximately the same size (50 KDa) of fusion protein was not detected in pTrcHisTOPO vector system.

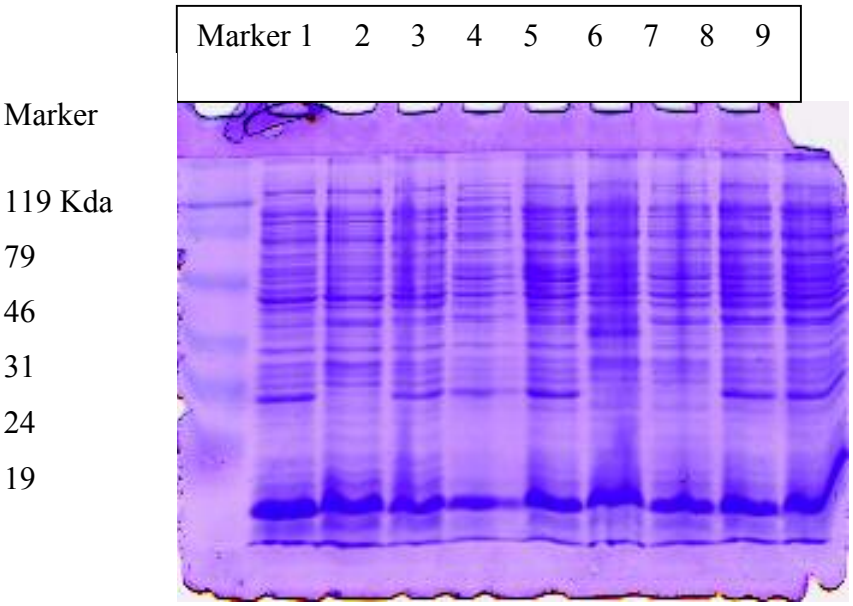


Figure 4.16 Induction analysis of 3 /4 (pTrcHis TOPO +*GPA1*) in BL21 (DE3) pLysS

Lane1 is BL21(DE3)pLysS, lane 2 is 3 /4 and lane3 is positive control before induction. Lanes 5 and 6 are +IPTG and –IPTG BL21(DE3) at 2 hours after induction, respectively. Lanes 7 and 8 are +IPTG and –IPTG 3 /4 at 2 hours after induction, respectively Lanes 9 and 10 are +IPTG and –IPTG positive controls at 2 hours after induction, respectively.

5 DISCUSSION

Dynamics of the conformational changes in protein structure, can be followed through measurements of structural parameters during interactions with other proteins and cofactors. These studies help to understand and interpret the biochemical data on possible functions of proteins. Development of techniques for recombinant protein production and progress in structural techniques permitting dynamic as well as static experiments help to broaden the scope of investigations of structure-function relationships.

There are a number of biochemical studies for the *GPA1,* which is the α-subunit of the heterotrimeric G protein from *Arabidopsis thaliana,* to determine its sequence-based characterization, subcellular localization and functions in a variety of metabolic processes including regulation of ion channels and stomatal opening, seed germination, growth, hormone regulation and pathogen response (Reviewed by Ma H, 2001)

Most of the functional studies on GPα1 were carried out with expression of *GPA1* mutants in transgenic *Arabidopsis thaliana.* The only study on the characterization of recombinant GPα1 came from Wise *et al.*, (1997) who expressed GPα1 in *E. coli* BL21 (DE3) strain by co-transformation with the expression construct and *dnaY* gene which encodes $tRNA^{Arg}_{AGA/AGG}$.

Although *GPA1* is a eukaryotic gene, *E. coli* was the choice for expression system in our study because it is the most common organism for convenient and high level production of recombinant proteins. Production of plant G_{α} using similar expression systems had also been reported in the literature previously. (Wise et al., 1997; Aharon et al., 1998)

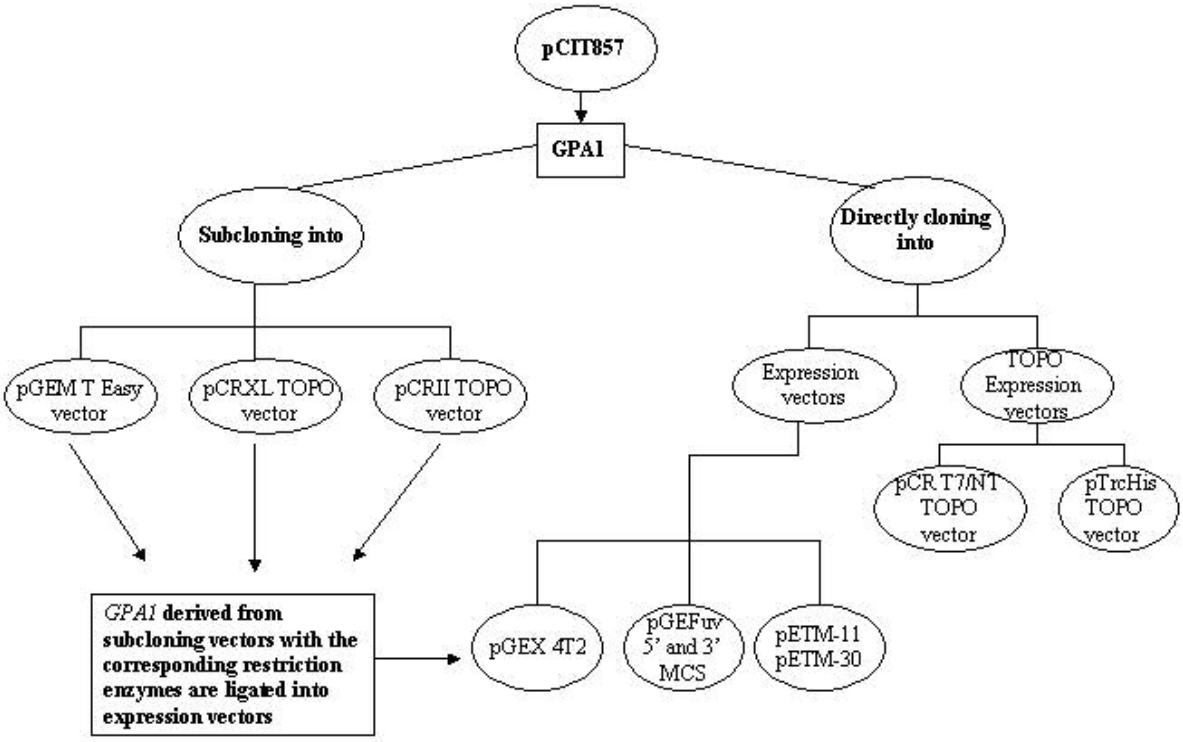


Figure 5.1 Cloning strategies of GPA1 into different cloning vectors

Three different approaches, shown schematically in Figure 5.1, were used to clone *GPA1* into expression vectors. For this purpose, different strains of *E. coli* including XL1Blue, TOP10 and TOP10 F' were used as hosts for subcloning and BL21(DE3), BL21(DE3)pLysS, Rosetta(DE3), Rosetta(DE3)pLysS and BL21-CodonPlus® (DE3)-RIL were used for expression. Subcloning vectors included pGEM® -T Easy (Promega), pCR® II- TOPO (Invitrogen), pCR® -XL-TOPO (Invitrogen) vectors, and a variety of expression vectors including pGEX-4T2 (Amersham Pharmacia), pGFPuv (Clonetech), pETM-11 and pETM-30(EMBL, Heidelberg), pCR® T7/NT TOPO (Invitrogen) and pTrcHis® TOPO (Invitrogen) vectors, were used.

Expression of GPα1 as a fusion partner of Glutathione-S-transferase (GST) in pGEX-4T2 and of GFPuv in pGFPuv vector as well as of His-tag in TOPO and pETM vectors offers several advantages. Fusion proteins may provide protection from

proteolytic degradation may act as reporter proteins and facilitate assays for rapid detection of the recombinants. On the other hand one step purification procedures using affinity columns are possible with his-tags or GST fusion proteins.

In this study prior to cloning experiments, the *GPA1* insert in the construct pCIT857, provided kindly by Dr. Hong Ma (PennState University, USA), was verified both by restriction enzyme digestion with *Eco* RI and sequence analysis of pCIT857. pCIT 857 had been constructed by inserting the longest *GPA1* cDNA fragment into the *Eco* RI site of the pGEM7Zf (+) vector (Promega) (Ma et al., 1990). After *Eco* RI digestion a fragment with the approximate size of 1400 bp fragment was observed on agarose gels (Figure 4.1). The *GPA1* gene is 1149 bp corresponding 383 amino acids (Ma *et al*., 1990) and sequence analysis of pCIT857 has shown that the larger size of the fragment observed on the gel, is due the 5'- and 3'- flanking regions of chromosome II of *Arabidopsis thaliana*. (Appendix C1, figure1).

After sequence verification, a restriction enzyme digestion map was prepared by using the tools on the website of SDSC Biology Workbench. (Table3.1) and different primers containing the appropriate restriction enzyme digestion sites were designed to clone *GPA1* into a variety of subcloning and expression vectors. (Table 3.2). In addition, primers without restriction enzyme digestion sites were designed.

When the size of the PCR product was compared with the pCIT 857/ *Eco* RI fragment (Figure 4.2), it was concluded that the *GPA1* coding region was properly amplified since the size of PCR product was between 1363bp and 1078bp bands of the size marker while the size of the fragment from *Eco* RI digestion of pCIT857 was above 1363. *GPA1* was also verified by sequencing the PCR product in subcloning vectors (Appendices C2, C3).

In the PCR reactions appropriate amounts of pCIT857, according to the manufacturer's instructions, was used as templates and PCR products were purified from agarose gels before ligation with vectors. However, restriction enzyme digestion and sequence analysis results have shown that *GPA1* positive clones were actually pCIT857 for many subcloning and cloning trials (data not shown). Different templates

consisting of digested and gel purified *GPA1* from pCIT857, construct #44 (pGEM T-Easy+*GPA1*), and very low amounts of pCIT857, were used as templates for PCR reactions in order to avoid template contamination after PCR. (Figure 4.2) In addition, PCR products were gel purified before ligation.

In summary as a result of exhaustive attempts the following constructs could be established as stable plasmids harbored in different types of non-expressing *E. coli* hosts

#44: pGEM T-Easy+*GPA1*;

TOPO8: pCRXL-TOPO+*GPA1* with *Eco* RI and *Xho* I sites;

P1 and P20: pCRII TOPO+*GPA1* with *Sac* I and *Spe* I sites,

4/4_ 3'MCS: pGFPuv+*GPA1* with *Sac* I and *Spe* I sites

T1/3: pGEX-4T2+*GPA1* with *Eco* RI and *Xho* I,

2/3: pT7/NT TOPO+ *GPA1*

3 /4: pTrcHis TOPO+*GPA1*

4/4_3'MCS, T1/3, 2/3 and 3 /4 were introduced also into various types of expression cells.

After sequence verification (Appendix C.3), *GPA1* from TOPO*8* could be inserted into pGEX-4T2 without difficulties. However, an error in the design of the forward primer containing *Eco* RI site caused the premature termination of the transcript and no expression was observed (Figure 4.14). Subsequently a new forward primer was designed for in frame insertion of *GPA1* into pGEX-4T2. Interestingly *GPA1* amplified with the correct forward primer yielded no positive results for insertion into pGEX-4T2.

In parallel experiments, *GPA1* with restriction enzyme sites for insertion into 5' and 3' MCS of pGFPuv, were subcloned using pCRII TOPO vector. Constructs pCR® II- TOPO+ *GPA1* with *Sac* I and *Spe* I at 5'- and 3'- ends of the insert respectively, were labeled as P1 and P20 (Appendix C.4) and these were used for insertion of *GPA1* into the 3' MCS of the pGFPuv vector (Appendix C.5). However, when it came to expression of the recombinant protein, similar to the pGEX-4T2 case, no fused protein could be detected either on SDS-PAGE or under UV light at different temperatures (37 ^{0}C and 30 ^{0}C) or at different IPTG concentrations (0.5 mM and 1mM) (Figure 4.14). Sequencing problems frequently encountered with expression constructs may be due to secondary structure of the inserted gene.

In addition to following a subcloning-expression cloning strategy alternative procedures, which would circumvent problems that may arise from alterations in the gene during extensive manipulatons, were also tried. Here the gene was amplified with primers containing appropriate restriction enzyme sites for the corresponding expression vectors and after digestion and purification of the PCR products and attempts were made to ligate these directly with the expression vectors, which were digested with the same enzymes. No *GPA1* positive clones could be obtained also with these alternative procedures (Data not shown).

To overcome the problems, which might due to the multiple steps invoved in subcloning and cloning, attempts were made to introduce *GPA1* directly into the TOPO expression vectors including pCR® T7/NT TOPO (Invitrogen) and pTrcHis® TOPO (Invitrogen). These are designed for cloning the PCR product immediately after the reaction using the inherent topoisomerase activity of the vector. The direction of the *GPA1* insert in these vectors was determined by restriction enzyme analysis with appropriate enzymes internal to *GPA1* and the vectors. (Figure 4.12) In addition, one of the sequencing primers supplied by the kit and a complement insert specific primer were used for PCR amplification to verify the insertion and direction of *GPA1*. These results were also confirmed by sequencing

In summary although *GPA1* could be PCR amplified subcloned readily for sequence verification and characterization, trials for insertion *GPA1* into pGEX-4T2,

pETM-11, pETM-30, and 5'MCS of pGFPuv vectors, which facilitate expression, failed. Constructs with expression vectors were obtained only with pT7/NT TOPO direct expression vector, when *GPA1* was inserted into the 3'-MCS of pGFPuv and when *GPA1* was inserted out of frame with GST. These constructs, however, failed to give significant expression of recombinant GPα1.

As mentioned earlier *GPA1* cloning (Ma et al., 1990) and expression in *E.coli* (Weiss et al., 1997) have been reported previously. However, not every gene can be expressed in this organism. This may be due to the unique structural features of the gene sequence, stability and translational efficiency of mRNA, protein misfolding, degradation by the host cell proteases, major differences in codon usage between the foreign gene and native *E. coli*, potential toxicity of the protein for the host cell.

5.1 Nucleotide Sequence of *GPA1*

Nucleotide sequence analysis shows that AT content (59%) of *GPA1* is higher than GC content (41 %). Several complementary nucleotides containing TTT, AAA, CC, GG, GC, CG are detected in the sequence. (Figure 4.2) These complementary repeats might create secondary structures and prevent efficient PCR, ligation, and translation by blocking the accession of cellular machinery during these reactions. This might explain the unsuccessful subcloning and cloning trials.

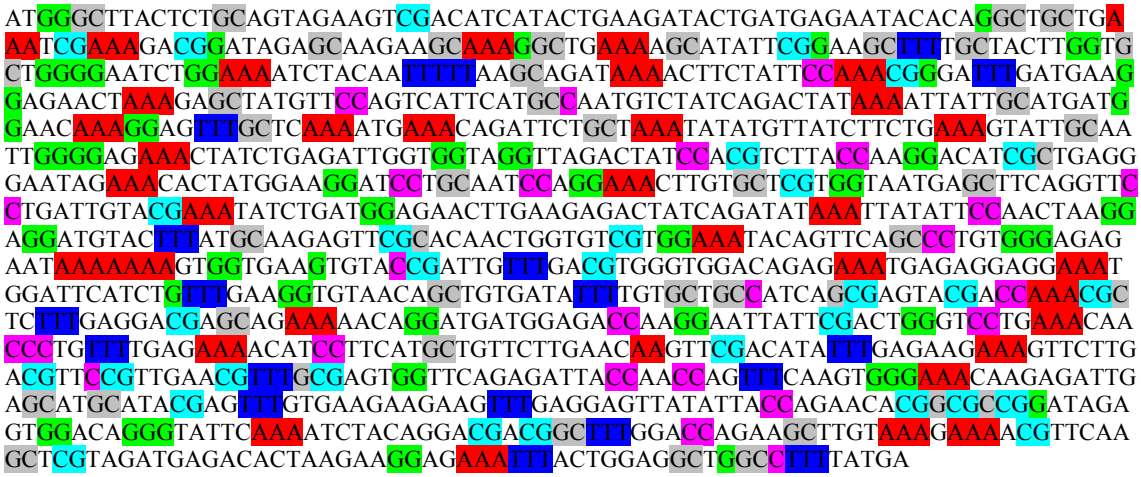


Figure 5.2 Several nucleotide repeats in the sequence of *GPA1*

5.2 Nonrandom Usage of Codons

Nonrandom usage of synonymous codons in prokaryotes and eukaryotes might be one possible explanation of the failure to see the recombinant protein from the construct 4/4_3'MCS in BL21(DE3) cells. Although expression was detected on SDS PAGE, there is a possibility that production of undetectable amount of recombinant protein is due to the abundance of the low usage codons in GPα1.

Codon(s)	Amino acid
AGA, AGG, CGA, CGG	Arg
UGU, UGC	Cys
GGA, GGG	Gly
AUA	Ile
CUA, CUC	Leu
CCC, CCU, CCA	Pro
UCA, AGU,UCG, UCC	Ser
ACA	Thr

Table 5.1 Low usage codons in *E. coli* (Makrides, SC., 1996)

Glu(41), Leu(35), Lys(35), Thr(24), Val(23) are the most frequent amino acids in the sequence of GPα1. Among them, Leu and Thr are the low usage codons in *E. coli*. Weiss et al.(1997), reported increased expression of GPα1 in *E. coli* BL21 (DE3) strain by co-transformation with the expression vector and *dnaY* gene which encodes $tRNA^{Arg}_{AGA/AGG}$ because minor $tRNA^{Arg}_{AGA/AGG}$ has been shown to be a limiting factor in bacterial expression of several mammalian genes.

5.3 Potential toxicity of GPα1

Potential toxicity of the protein by the host cell is another possibility for the unexpressed protein although it is cloned at 3'MCS of pGFPuv. To observe the propagation and the maintenance TOP10F' competent cells, which do not contain T7 RNA polymerase, were transformed with the construct. There was no significant

difference in the number of the colonies obtained with TOP10F' and those with any of the expression cells. In addition, there was no extreme OD_{600} changes observed during expression procedure between the positive control and 4/4_3'MCS.

Although it is not usually essential for expression, posttranslational modifications are not carried out in *E. coli* unless these are introduced by other means. There is no guarantee that *E. coli* has the correct chaperons for all foreign proteins and that correct folding will take place and a thermodynamically stable end product as that observed *in vivo*, will be produced.

The usage of *E. coli* as an expression host might cause folding problems because of the requirement for specific chaperons and the absence of posttranslational modifications.

5.4 mRNA Instability

Our inability to detect the fusion protein may be due to mRNA instability, improper folding of GFPuv due to size orvery low level of expression due to the abundant presence of rare codons in the amino acid sequence of *GPA1*

Results of cloning and expression studies presented in this thesis show that despite the previous report by Weiss et al (1997), expression of *GPA1* in a prokaryotic expression system was not possible under standard conditions. Wise et al. have reported a yield of 1-2 mg recombinant protein from one liter of culture which is very low and and was obtained only after introduction of *dna Y* gene encoding $tRNA^{Arg}_{AGA/AGG}$. It appears that expression and detection requires special precautions for secondary structure formation and limited codon availability. In addition, the effect of posttranslational modifications and the chaperon specificity of *GPA1* for proper folding, raise doubts about a prokaryotic system for expression.

In *in vitro* translation experiments the yeast two-hybrid system was used to show the interaction of β-subunit from *Arabidopsis thaliana* with a possible γ-subunit

(Mason and Botella, 2000). However, there is no *in vivo* evidence for the interaction of *GPA1* with βγ-dimer other than the detection of possible binding regions by sequence-based homology modeling. In other words there is no conclusive experimental evidence showing the presence of a heterotrimeric G-protein in plants and no experimental data suggesting the mechanism by which the heterotrimer may function.

6 CONCLUSION

Functional studies on isolated GTP binding proteins from plants are lacking in the literature. Similarly structural investigations, which would contribute to a better understanding of these proteins in the signaling pathways of plants are also lacking and much needed. The broader perspective of this thesis involves structural measurements on the components of the heterotrimeric plant G-proteins. In particular small angle solution X-ray scattering would provide dynamic information on the interaction of the α-, β-, and γ-subunits and conformational changes during the signal transduction reactions. This type of structural work however requires production of significant quantities of well-characterized recombinant proteins. The work presented in this thesis provides a basis for the expression and characterization of recombinant plant G_{α} (*GPA1*).

In conclusion, we have been able to amplify and characterize *GPA1* gene and subclone it in several vectors for further manipulation. These could be sequenced and sequence fidelity to *GPA1* cDNA is shown. Additionally *GPA1* has been inserted into vectors where the recombinant GPα1 could be expressed either on its own or as a fusion protein with GFPuv.

In contrast to an earlier report by Wise et al, (1997) the expression of the recombinant protein could not be detected under standard prokaryotic expression conditions. An improved prokaryotic or a eukaryotic system would facilitate expression of the recombinant GPα1. Our studies also indicate that secondary structure features of the *GPA1* gene need to be taken into account when designing these improved systems. Furthermore the recombinant protein needs to be fully characterized for posttranslational modifications, GTP-binding and βγ-subunit interactions before functional studies.

Determination of the biochemical parameters of the recombinant protein and can provide a reference for *in situ* studies and pure and soluble recombinant protein can be used for structural experiments. Solving the structure of *GPA1* by crystallography and solution scattering experiments would answer questions about not only the functional similarities and differences of the plant G-protein α-subunit with its animal counterparts but also the presence of the heterotrimer in plants, interaction of *GPA1* with GTP, interactions with βγ-subunits and with other possible downstream effectors.

In the aspects mentioned above site-directed mutagenesis experiments would also provide valuable information however our cloning work suggests extereme caution about the artifacts which may arise from the secondary structure features of the *GPA1* gene.

7 REFERENCES

1. Aharon GS, Gelli A, Snedden WA, Blumwald E, "Activation of a plant plasma membrane Ca^{+2} channel by TGα1, a heterotrimeric G protein α-subunit homologue" *FEBS Letters* 424 (1998) 17-21.

2. Ando S, Takumi S, Ueda Y, Ueda T, Mori N, Nakamura C, "Nicotina Tobaccum cDNAs encoding α- and β-subunits of a heterotrimeric GTP binding protein isolated from hairy root tissues" *Genes Genet. Syst.* 75 (2000) 211-221.

3. Arabidopsis Genome Initiative, " Analysis of the geome sequenceof the flowering plant Arabidopsis thaliana" *Nature* 408 (2000) 796-815.

4. Assmann SM, "Heterotrimeric and Unconventional GTP Binding Proteins in Plant Cell Signalling" *The Plant Cell* 14 (2002) S355-S373.

5. Bischoff F, Molendijk A, Rajendrakumar CS, Palme K, "GTP-binding proteins in plants" *Cell Mol. Life Sci.* 55 (1999) 233-256.

6. Blumwald E, Aharon GS, Lam BC-H, "Early Signal Transduction Pathways in Plant-Pathogen Interactions" *Trends in Plant Science* 3:9 (1998) 342-346.

7. Bohm A, Gaudet R, Sigler PB, "Structural aspects of heterotrimeric G-protein signaling" *Current Opinion in Biotechnology* 8 (1997) 480-487.

8. Danner S, Lohse M, "Phosducin is a ubiquitous G-protein regulator" *PNAS* 93 (1996) 10145-10150.

9. Devitt ML, Maas KJ, Stafstrom JP, "Characterization of DRGs, developmentally regulated GTP-binding proteins, from pea and *Arabidopsis*" *Plant Molecular Biology* 39 (1999) 75-82.

10. Devoto A, Piffanelli P, Nillsson I, Wallin E, Panstruga R, Heijne G, Leferd P, "Topology, Subcellular Localization, and sequence Diversity of the Mlo Family in Plants" *The Journal of Biol. Chem.* 274:49 (1999) 34993-350004

11. Dohlman HG, Thorner J, "RGS Proteins and Signalling by Heterotrimeric G Proteins" *The Journal of Biol. Chem.* 272:7 (1997) 3871-3874.

12. Earnst OP, Meyer CK, Marin EP, Henklein P, Fue W, Sakmar TP, Hofman KP, "Mutation of Fourth Cytoplasmic Loop of Rhodopsin Affects Binding of Transducin and Peptides Derived from the Carboxyl-terminal Sequences Transducin α- and γ-subunits" *The Journal of Biol. Chem.* 275:3 (2000) 1937-1943.

13. Fujisawa Y, Kato H, Iwasaki Y, "Structure and functions of heterotrimeric G proteins in plants" *Plant Cell. Physiol.* 42:8 (2001) 789-794.

14. Gotor C, Lam E, Cejudo FJ, Romero LC, "Isolation and analysis of the soybean SGA2 gene (cDNA), encoding a new member of the plant G-protein family of signal transducers" *Plant Mol. Biol.* 32 (1996) 1227-1234.

15. Hamm HE, "The many faces of G protein signaling" *J.Biol. Chem.* 273:2 (1998) 669-672.

16. Hilgenfeld R, "Regulatory GTPases" *Current Opinion in Structural Biology* 5 (1995) 810-817.

17. Hur EM, Kim KT, "G protein-coupled receptor signaling and cross-talk Achieving rapidity and specificity" *Cellular Signalling* 14:4 (2002) 397-405.

18. Iñiguz-Lluhi J, Kleuss C, Gilman AG, “The importance of G-protein β/γ subunits” *Trends in Cell Biology* 3 (1993) 230-235.

19. Iwasaki Y, Kato T, Kaidoh T, Ishikawa A, Asadi T, “Characterization of the putative α-subunit of a heterotrimeric G protein in rice” *Plant Molecular Biology* 34 (1997) 563-572.

20. Josefsson LG, Rask L, “Cloning of a putative G-protein-coupled receptor from *Arabidopsis thaliana*” *Eur. J. Biochem.* 249:2 (1997) 415-420.

21. Kaydamov C, Tewes A, Adler K, Manteuffel R, “Molecular characterization of cDNAs encoding G protein alpha and beta subunits and study of their temporal and spatial expression patterns in Nicotiana plumbaginifolia Viv” *Biochim. Biophys. Acta* 149 (2000) 143-160.

22. Kim WY, Cheong NE, Lee DC, Je DY, Bahk JD, Cho MJ, Lee SY, “Cloning and sequencing analysis of a full-length cDNA encoding a G protein alpha subunit, SGA1, from soybean” *Plant Physiol.* 108 (1995) 1315-1316.

23. Lee YR, Assmann SM, “Arabidopsis thaliana 'extra-large GTP-binding protein' (AtXLG1): a new class of G-protein” *Plant Mol. Biol.* 40:1 (1999) 55-64.

24. Legendre L, Heinstein PF, Low PS, “Evidence for participation of GTP-binding proteins in elicitation of oxidative burst in cultured soybean cells” *J.Biol.Chem.* 267: 20140-20147.

25. Ma H, Yanofsky MF, Meyerowitz EM, “Molecular cloning an characterization of *GPA1*, a G protein alpha subunit gene from *Arabidopsis thaliana*” *Proc. Natl. Acad. Sci. USA* 87:10 (1990) 3821-3825.

26. Ma H, Yanofsky MF, Huang H, “Isolation and sequence analysis of TGA1 cDNAs encoding a tomato G protein α subunit” *Gene* 107 (1991) 189-195.

27. Ma H, “GTP-binding proteins in plants: new members of an old family” *Plant Mol. Biol.* 26 (1994) 1611-1636.

28. Ma H, “Plant G proteins: The different faces of *GPA1*” *Current Biology* 11 (2001) R869-R871.

29. Makrides SC, “Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*” *Microbiological Reviews* 60:3 (1996) 512-538.

30. Marin EP, Krishna AG, Zvyaga TA, Isele J, Siebert F, Sakmar TP, “ The Amino-terminus of the Fourth Cytoplasmic Loop of Rhodopsin Modulates Rhodopsin-Transducin Interaction” *J.Biol. Chem.* 275:3 (2000) 1930-1936.

31. Marinissen MJ, Gutkind JS, “G-protein-coupled receptors and signaling networks: emerging paradigms” *Trends in Pharmachological Sciences* 22:7 (2001) 368-376.

32. Marsh JF, Kaufman LS, “Cloning and characterization of *PGA1* and *PGA2*: Two G-protein α-subunits from pea that promote growth in the yeast *Sacchharomycyes cerevisiae*” *Plant Journal* 19 (1999) 237-47.

33. Mason MG, Botella JR, “Completing the heterotrimer: isolation and characterization of an *Arabidopsis thaliana* G protein γ-subunit cDNA” *PNAS* 97:26 (2000) 14784-14788.

34. Mason MG, Botella JR, “Isolation of a novel G-protein γ-subunit from *Arabidopsis thaliana* and its interaction with Gβ” *Biochim. Biophys. Acta.* 1520 (2001) 147-153.

35. McDowell JM, Dangl JL, “Signal transduction in plant immune response” *TIBS* 25 (2000) 79-82.

36. Millner PA, "Heterotrimeric G-proteins in plant cell signalling" *New Phytol.* 151 (2001) 165-174.

37. Neves SR, Ram PT, Iyengar R, "G protein pathways" *Science* 296 (2002) 1636-1639.

38. Okamoto H, Matsui M, Deng XW, "Overexpression of the Heterotrimeric G-Protein α-Subunit Enhances Phytochrome-Mediated Inhibition of Hypocotyl Elongation in *Arabidopsis*" *The Plant Cell* 13 (2001) 1639-1651.

39. Perroud PF, Diogon T, Crevecoeur M, Greppin H, "Molecular cloning, spatial and temporal characterization of spinach SOGA1 cDNA, encoding an α subunit of G protein"; *Gene* 248 (2000) 191-201.

40. Rens-Domiano S, Hamm HE, "Structural and functional relationships of heterotrimeric G-proteins" *FASEB J.* 9 (1995) 1059-1066.

41. Saalbach G, Natura B, Lein W, Buschmann P, Dashe I, Rohrbeck M, Nagy F, "The α-subunit from heterotrimeric G-protein from tobacco, NtGPα1, functions in K^+ channel regulation in mesophyll cells" *J.Exp.Bot*. 50 (1999) 53-61.

42. Sambrook J, Manniatis T, Fritsch EF, "Molecular Cloning: A Laboratory Manual" Cold Spring Harbor Laboratory Press, 2nd Edition, (1989).

43. Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerwille S, Manners JM, "Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis" *PNAS* 97:21 (2000) 11655-11660.

44. Seo HS, Kim HY, Jeong JY, Lee SY, Cho MJ, Bahk JD, "Molecular cloning and characterization of RGA1 encoding a G protein α-subunit from rice (Oryza sativa L. IR-36)" *Plant Mol. Biol.* 27 (1995) 1119-1131.

45. Sprang SR, "G protein mechanisms: Insights from structural analysis" *Annu. Rev. Biochem.* 66 (1997) 639-678.

46. Taylor JM, Jacob-Mosier GG, Lawton RG, VanDort M, Neubig RR, "Receptor and Membrane Interaction sites on G_{β} "*The Journal of Biological Chem.* 271:7 (1996) 3336-3339.

47. Thomma BPHJ, Eggermonth K, Penninckx IAMA, Mauch-Mani B, Vogelsang R, Cammue BPA, Brokeard WF, "Separate jasmonate-dependent and salicylate-dependent defense response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens" *PNAS* 95 (1998) 15107-15111.

48. Torres MA, Dangl JL, Jones JDC, "*Arabidopsis* gp91^{phox} homologues *AtrbohD* and *AtrbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response" *PNAS* 99:1 (2002) 517-522.

49. Ueguchi-Tanaka M, Fujisawa Y, Kobayashi M, Ashikari M, Iwasaki Y, "Rice dwarf mutant *d1*, which is defective in the α-subunit of heterotrimeric G protein, affects gibberellin signal transduction" *PNAS* 97:21 (2000) 11638-11643.

50. Ullah H, Chen J, Young JC, Im K, Sussman MR, Jones AM, "Modulation of Cell Proliferation by Heterotrimeric G Protein in *Arabidopsis*" *Science* 292 (2001) 2066-2069.

51. Ullah H, Chen J, Wang S, Jones AM, "Role of a Heterotrimeric G Protein in Regulation of *Arabidopsis* Seed Germination" *Plant Physiology* 129 (2002) 897-907.

52. Vera-Estrella R, Higgins VJ, Blumwald E, "Plant Defense Response to Fungal Pathogens" *Plant Physiol*. 106 (1994) 97-102.

53. Vetter IR, Wittinghofer A, "The Guanine Nucleotide-Binding Switch in Three Dimentions" *Science* 294 (2001) 1299-1304.

54. Wall MA, Coleman DE, Lee E, Iñiguz-Lluhi JA, Posner BA, Gilman AG, Sprang SR, "The structure of the G protein heterotrimer $G_{i\alpha 1}\beta_1\gamma_2$" *Cell* 83 (1995) 1047-1058.

55. Wall MA, Posner BA, Sprang SR, "Structural basis of activity and subunit recognition in G protein heterotrimers" *Structure* 6 (1998) 1169-1183.

56. Wang X-Q, Ullah H, Jones AM, Assmann SM, "G protein regulation of ion channels and abscisic acid signaling in *Arabidopsis* guard cells" *Science* 292 (2001) 2070-2072.

57. Wedegaertner PB, Wilson PT, Bourne PB, "Lipid Modifications of Trimeric G Proteins" *J.Biol. Chem.* 270:2 (1995) 503-506.

58. Weiss CA, Garnaat CW, Mukai K, Hu Y, Ma H, "Isolation of cDNAs encoding guanine nucleotide-binding protein β-subunit homologues from maize (ZGB1) and *Arabidopsis* (AGB1)" *PNAS* 91 (1994) 9554-9558.

59. Weiss CA, White E, Huang H, Ma H, "The G protein α-subunit (GPα1) is associated with the ER and the plasma membrane in meristematic cells of *Arabidopsis* and cauliflower" *FEBS Letters* 407 (1997) 361-367.

60. Yokoyoma S, Wilkie TM, "Evolution of the G Protein Alpha Subunit Multigene Family" *Molecular Evolution of Physiological Processes* (1994) 249-270.

61. Wise A, Thomas PG, Carr TH, Murphy GA, Millner PA, "Expression of the *Arabidopsis* G-protein GPα1:purification and characterization of the recombinant protein" *Plant Molecular Biology* 33 (1997) 723-728.

62. Wittinghofer A, “Deciphering the alphabet of G proteins: the structure of the α, β, γ heterotrimer” *Structure* 4 (1996) 357-361.

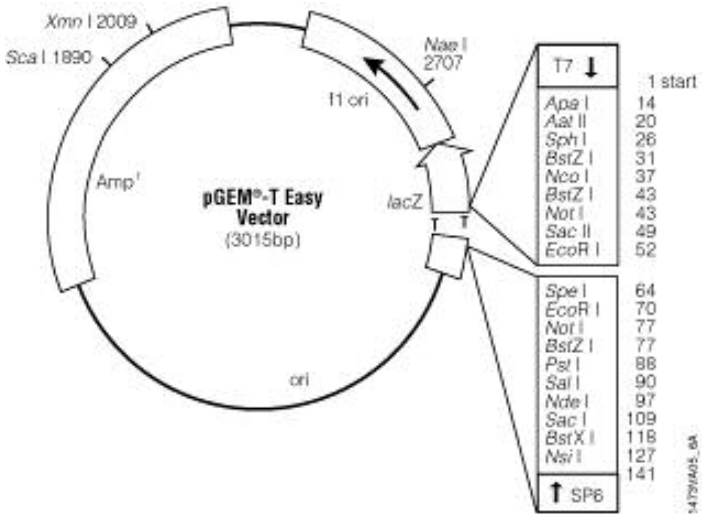
63. Yang C, Skibas NP, Mazzoni MR, Hamm HE, “Conformational changes at The Carboxyl Terminus of G_{α} Occur during G Protein Activation” *The Journal of Biological Chem.* 274:4 (1999) 2379-2385.

64. Yang Z, “Small GTPases: Versatile Signalling Switches in Plants” *The Plant Cell* (2002) S375-S388.

65. Yasuda H, Lindorfer MA, Woodfork KA, Fletcher JE, Garrison JC, “ Role of the Geranyl Group on the G-protein γ Subunit in Coupling Trimeric G Proteins to A1 Adenosine Receptors” *The Journal of Biological Chem.* 271:31 (1996) 18588-18595.

APPENDIX A

A.1 pGEM-T Easy vector circle map and sequence reference points



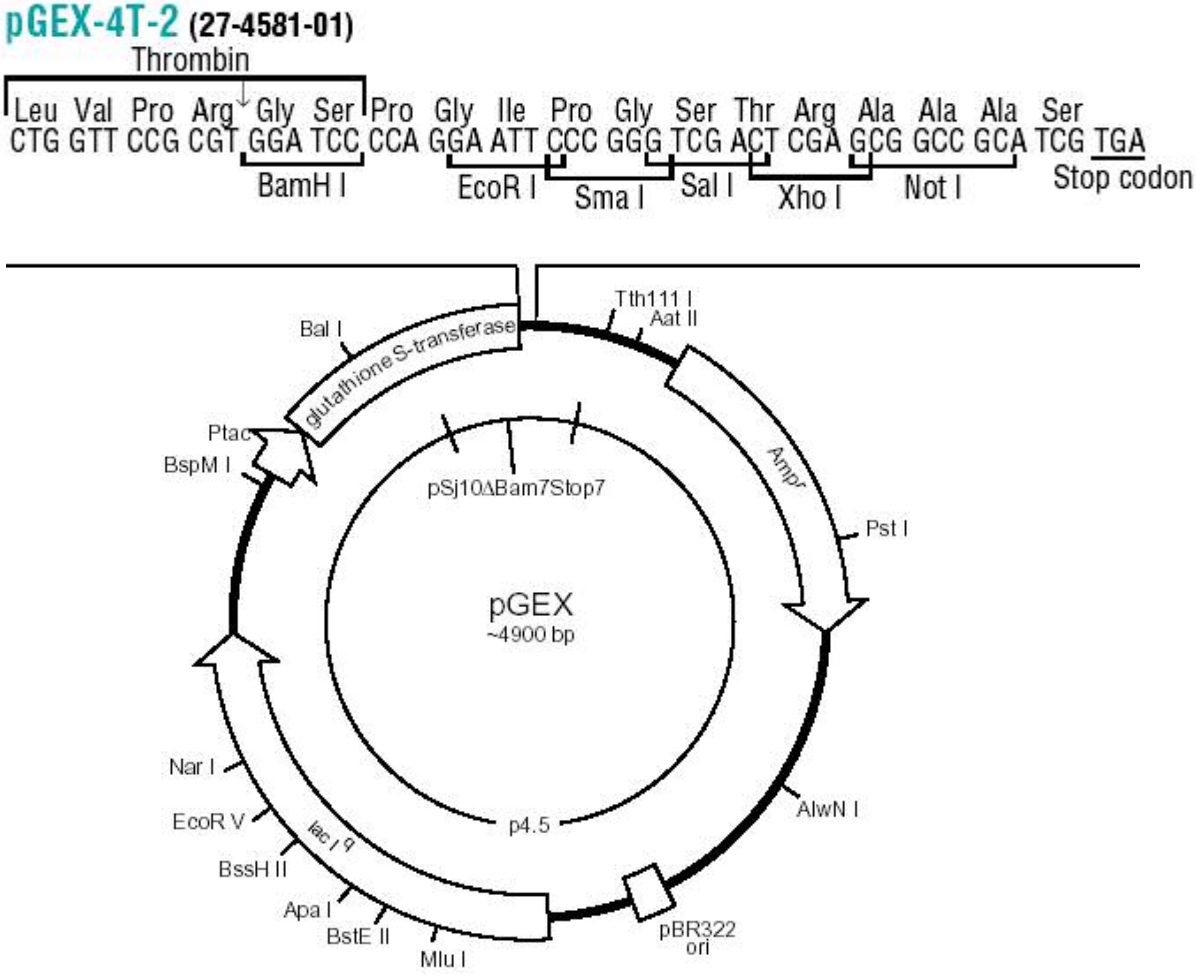
pGEM®-T Easy Vector Sequence reference points:	
T7 RNA Polymerase transcription initiation site	1
SP6 RNA Polymerase transcription initiation site	141
T7 RNA Polymerase promoter (–17 to +3)	2999–3
SP6 RNA Polymerase promoter (–17 to +3)	139–158
multiple cloning region	10–128
lacZ start codon	180
lac operon sequences	2836–2996, 166–395
lac operator	200–216
β-lactamase coding region	1337–2197
phage f1 region	2380–2835
binding site of pUC/M13 Forward Sequencing Primer	2956–2972
binding site of pUC/M13 Reverse Sequencing Primer	176–192

Enzyme	# of Sites	Location
Aat II	1	20
Acc I	1	76
Acy I	2	17, 1932
Afl III	2	99, 502
*Alw*26 I	2	1456, 2232
*Alw*44 I	2	816, 2062
*Alw*N I	1	918
Apa I	1	14
*Asp*H I	4	94, 820, 1981, 2066
Ava II	2	1533, 1755
Ban I	3	246, 1343, 2626
Ban II	3	14, 94, 2664
Bbu I	1	26
Bgl I	3	39, 1515, 2833
Bsa I	1	1456
*Bsa*A I	1	2589
*Bsa*H I	2	17, 1932
*Bsa*J I	5	37, 43, 241, 662, 2936
*Bsp*120 I	1	10
*Bsp*H I	2	1222, 2230
*Bsp*M I	1	62
*Bss*S I	2	675, 2059
*Bst*O I	5	242, 530, 651, 664, 2937
*Bst*X I	1	103

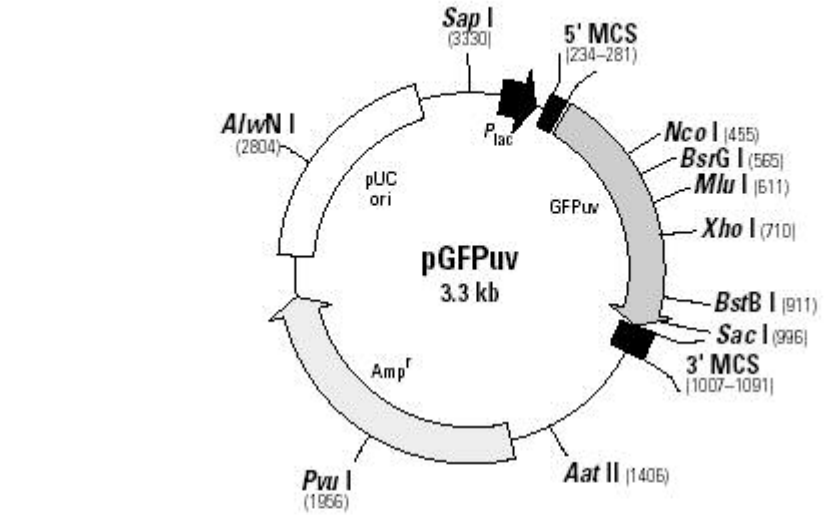
Enzyme	# of Sites	Location
Bst Z I	2	31, 62
*Cfr*10 I	2	1475, 2690
Dde I	4	777, 1186, 1352, 1892
Dra I	3	1261, 1280, 1972
Dra III	1	2589
Drd I	2	610, 2544
Dsa I	2	37, 43
Eag I	2	31, 62
Ear I	3	386, 2190, 2878
*Ecl*HK I	1	1395
*Eco*52 I	2	31, 62
*Eco*ICR I	1	92
*Eco*R V	1	51 (see above)
Fok I	5	119, 1361, 1542, 1829, 2919
Fsp I	2	1617, 2840
Hae II	4	380, 750, 2740, 2748
Hga I	4	613, 1191, 1921, 2806
Hinc II	1	77
Hind II	1	77
*Hsp*92 I	2	17, 1932
Mae I	5	56, 997, 1250, 1585, 2740
Mlu I	1	99

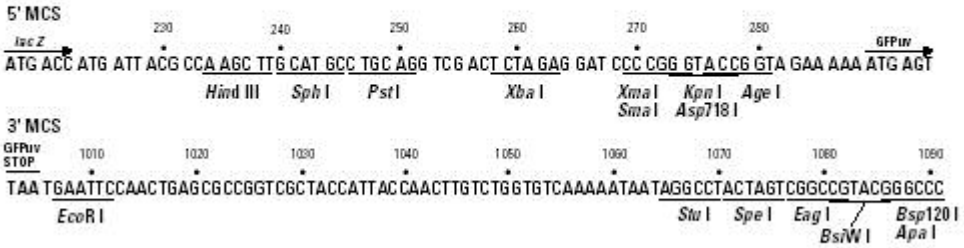
Total number of hits per restriction enzyme for pGEM- T Easy vector

A.2 Vector map of pGEX-4T2

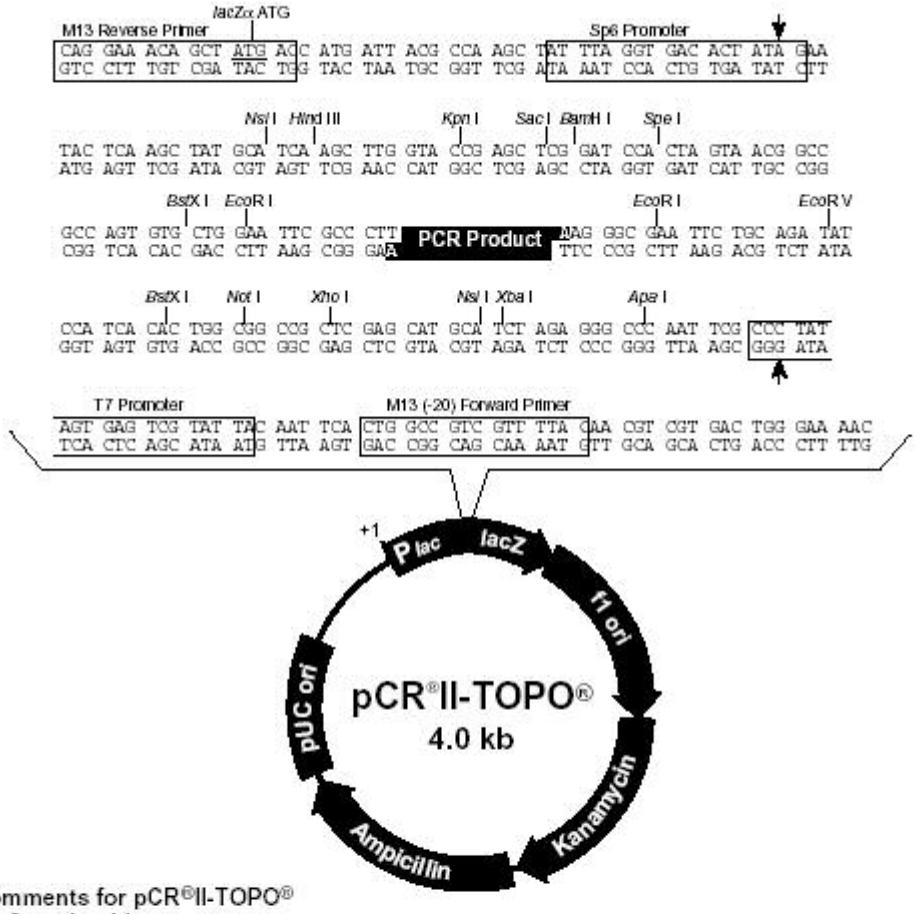


A.3 pGFPuv vector map showing 5’ and 3’ multiple cloning sites





A.4 pCRII-TOPO vector map and sequence reference points



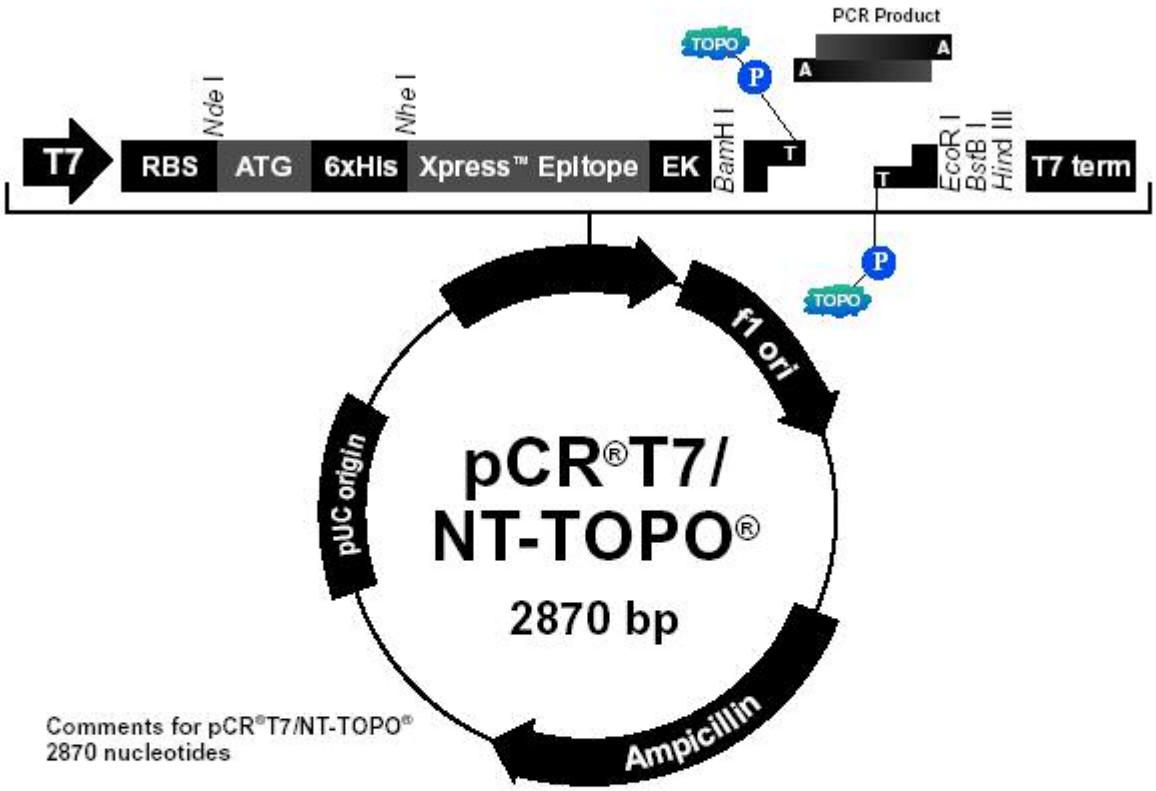
Comments for pCR®II-TOPO®
3973 nucleotides

LacZα gene: bases 1-589
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
Multiple Cloning Site: bases 269-383
T7 promoter: bases 406-425
M13 (-20) Forward priming site: bases 433-448
f1 origin: bases 590-1027
Kanamycin resistance ORF: bases 1361-2155
Ampicillin resistance ORF: bases 2173-3033
pUC origin: bases 3178-3851

Zeocin resistance ORF: bases 2238-2612
pUC origin: bases 2680-3393

pCRII-TOPO vector map including multiple cloning site and sequence reference points

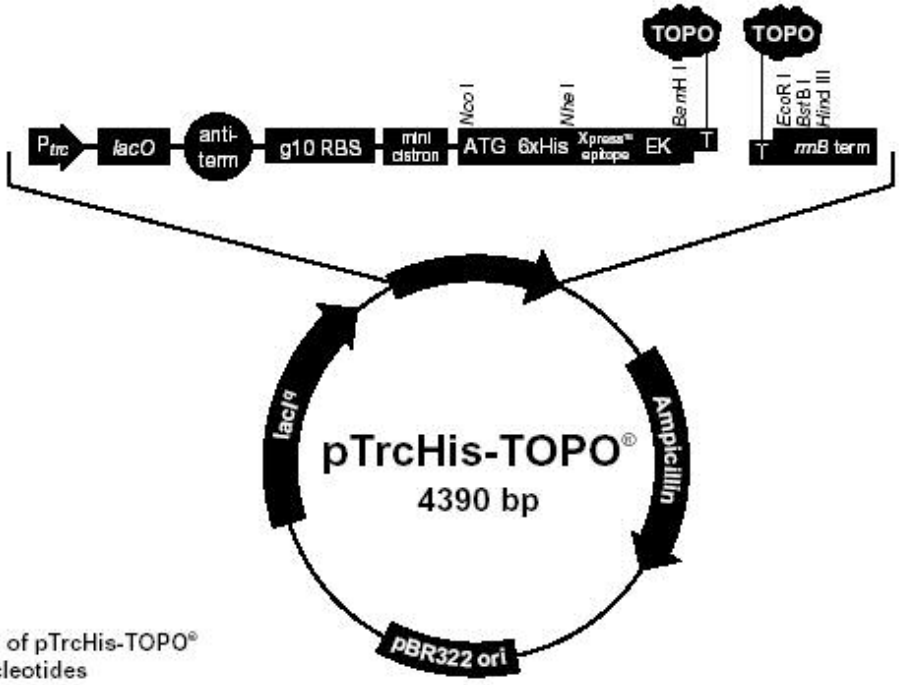
A.5 pCR T7/NT-TOPO vector including multiple cloning site and sequence



Comments for pCR®T7/NT-TOPO®
2870 nucleotides

T7 promoter: bases 20-36
T7 promoter priming site: bases 20-39
Ribosome binding site: bases 87-90
Initiation ATG: bases 100-102
Polyhistidine (6xHis) region: bases 112-129
Xpress™ epitope: bases 169-192
EK recognition site: bases 178-192
TOPO® Cloning site: bases 204-205
T7 reverse priming site: bases 270-289
T7 transcription termination region: bases 231-360
f1 origin: 431-886
Ampicillin resistance gene (ORF): bases 1017-1877
pUC origin: 2022-2695

A.6 pTrcHis-TOPO vector including multiple cloning site and sequence referemce points

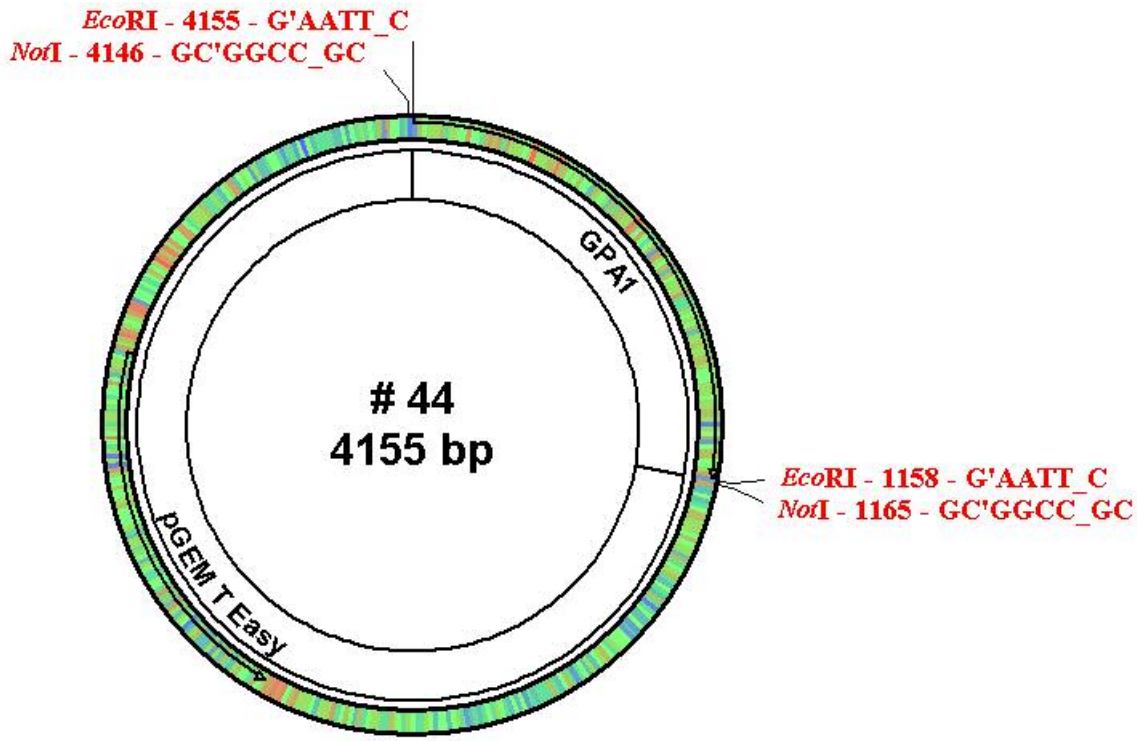


Features of pTrcHis-TOPO®
4390 nucleotides

trc promoter and 5' UTR: bases 190-382
-35 region: bases 193-198
-10 region: bases 216-221
lac operator site: bases 228-248
rrnB anti-termination sequence: bases 264-333
T7 gene 10 translational enhancer: bases 346-354
Ribosome binding site: 369-373
pTrcHis forward priming site: bases 370-390
Minicistron: bases 383-409
Reinitiation RBS: bases 398-403
Initiation ATG: bases 413-415
6xHis tag: bases 425-442
Xpress™ epitope: bases 482-505
Xpress™ forward priming site: bases 445-463
Enterokinase cleavage site: bases 491-505
TOPO® Cloning site: bases 517-518
pTrcHis reverse priming site: bases 574-591
rrnB T_1 and T_2 transcription termination sequence: bases 624-781
bla promoter: bases 1002-1059
Ampicillin resistance gene (*bla*): bases 1060-1920
pBR322-derived origin: bases 2065-2738
Lac Repressor (*lacI*^{q}) : bases 3267-4351

APPENDIX B

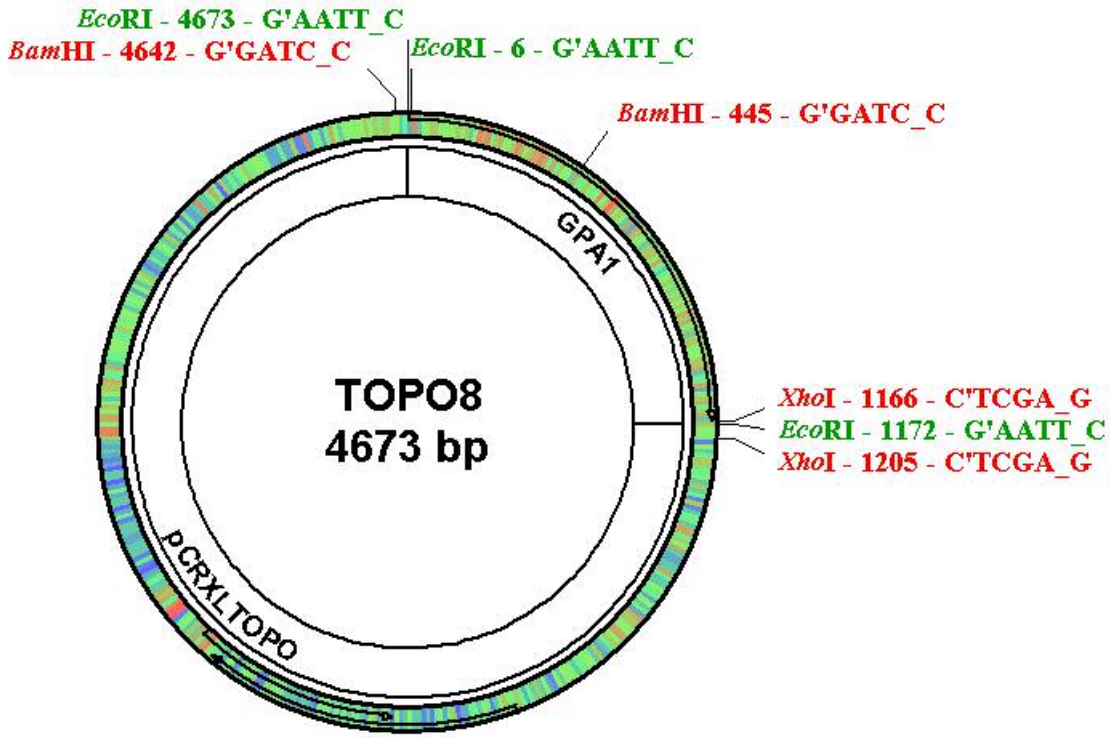
B.1 The construct named as # 44 which is the pGEM T Easy (Promega) containing *GPA1*



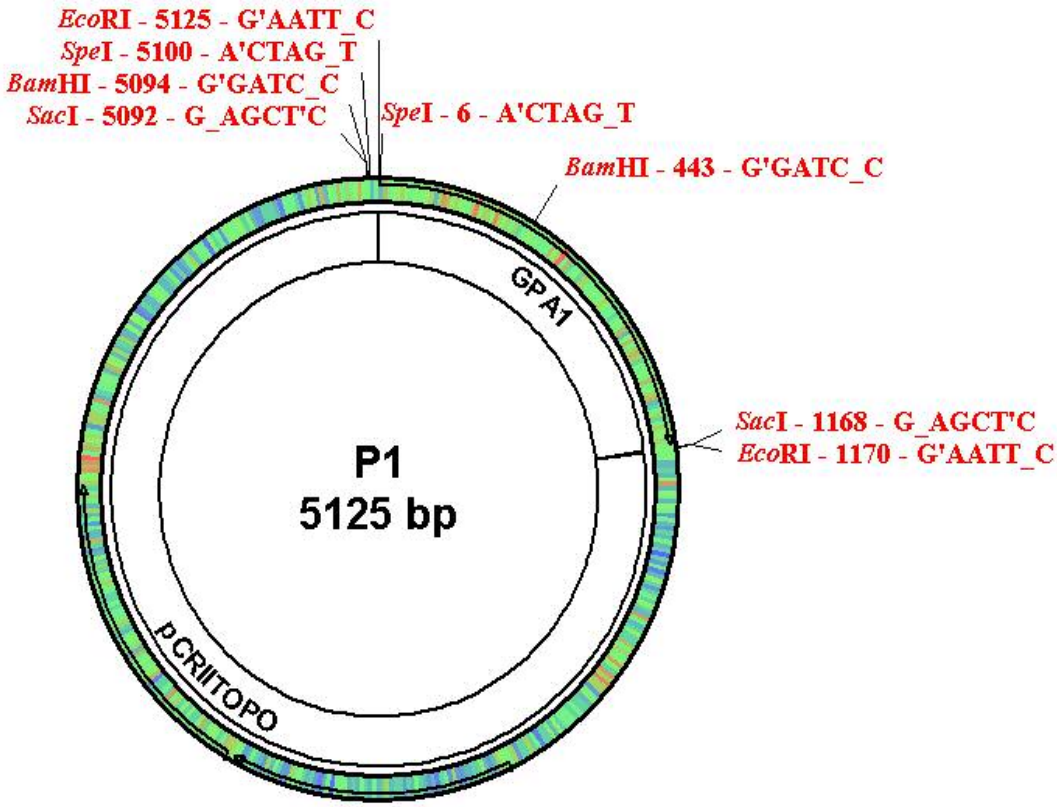
GPA1, cDNA sequence encoding α-subunit of heterotrimeric G protein from Arabidopsis thaliana was inserted between *Eco* RI restriction enzyme sites of pGEM T-

Easy vector. pGEM T-Easy is constructed in such a way that additional 3' terminal thymidine at both ends provides a compatible overhangs for PCR products at which the additional 5' terminal adenine is added by Taq Polymerase enzyme.

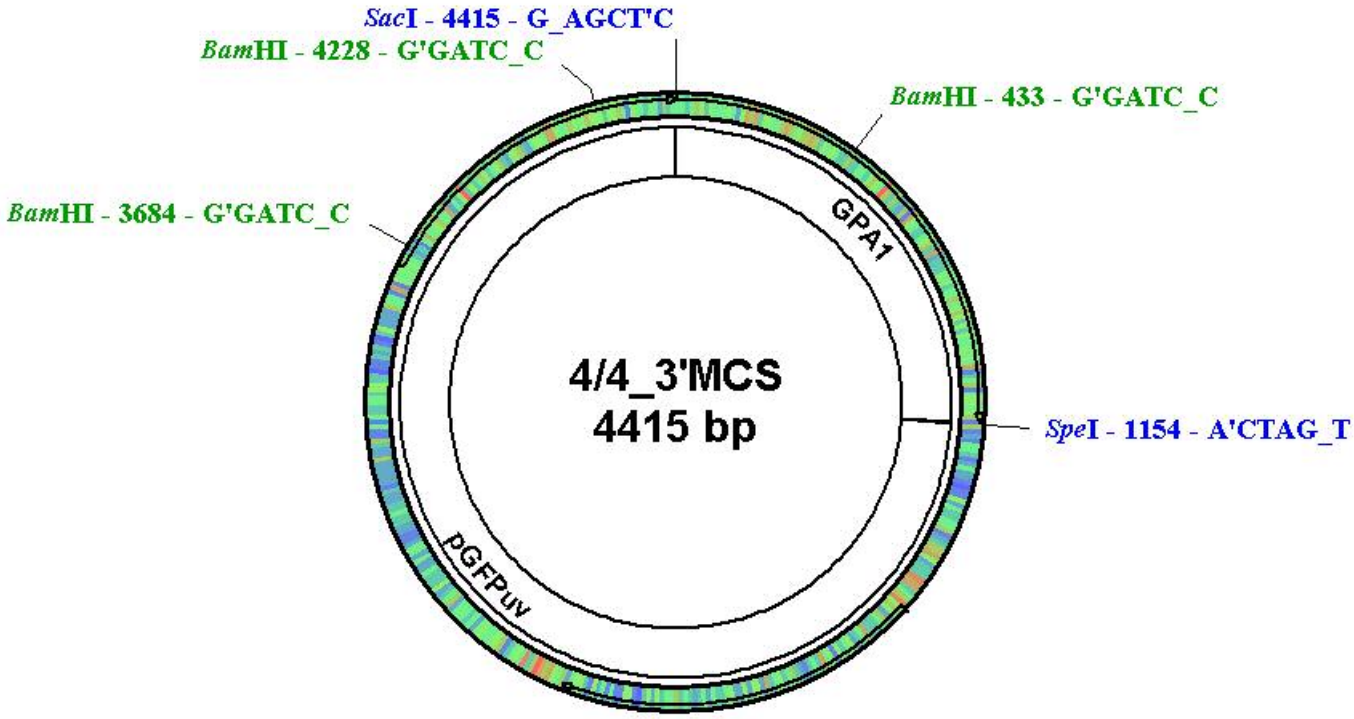
B.2 The construct named as TOPO8 which is pCRXLTOPO vector containing *GPA1* with *Eco* RI and *Xho* I restriction enzyme sites at 5' and 3' ends of *GPA1*, respectively



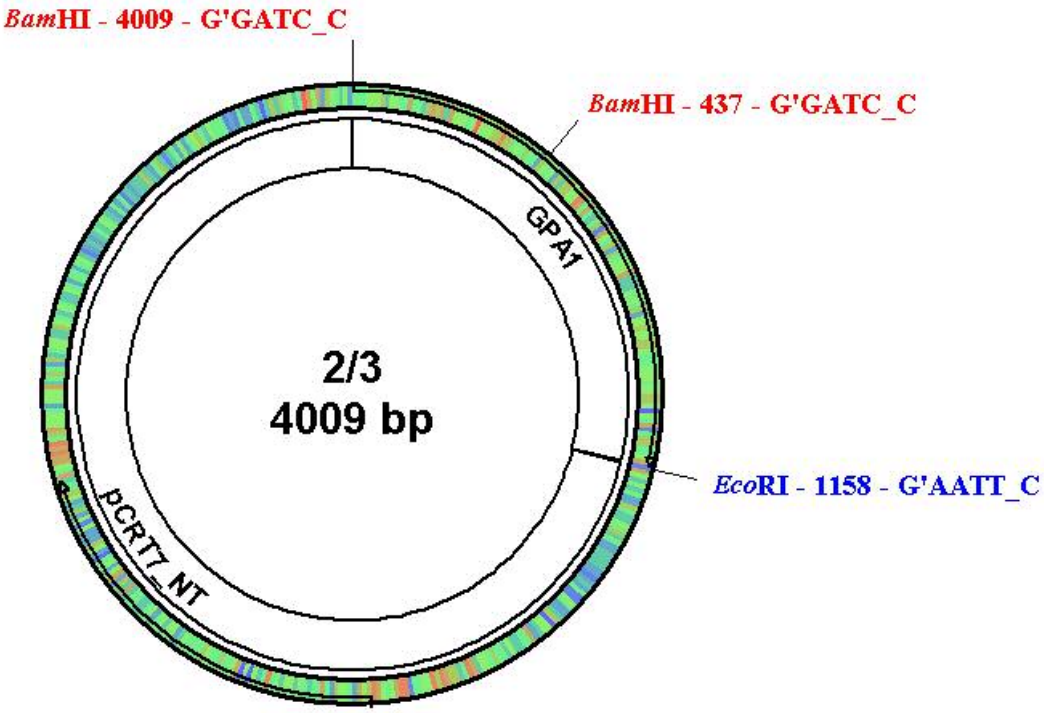
B.3 The construct named as P1 which is pCRIITOPO vector containing *GPA1* with *Sac* I and *Spe* I restriction enzyme sites at 5' and 3' ends of *GPA1*, respectively



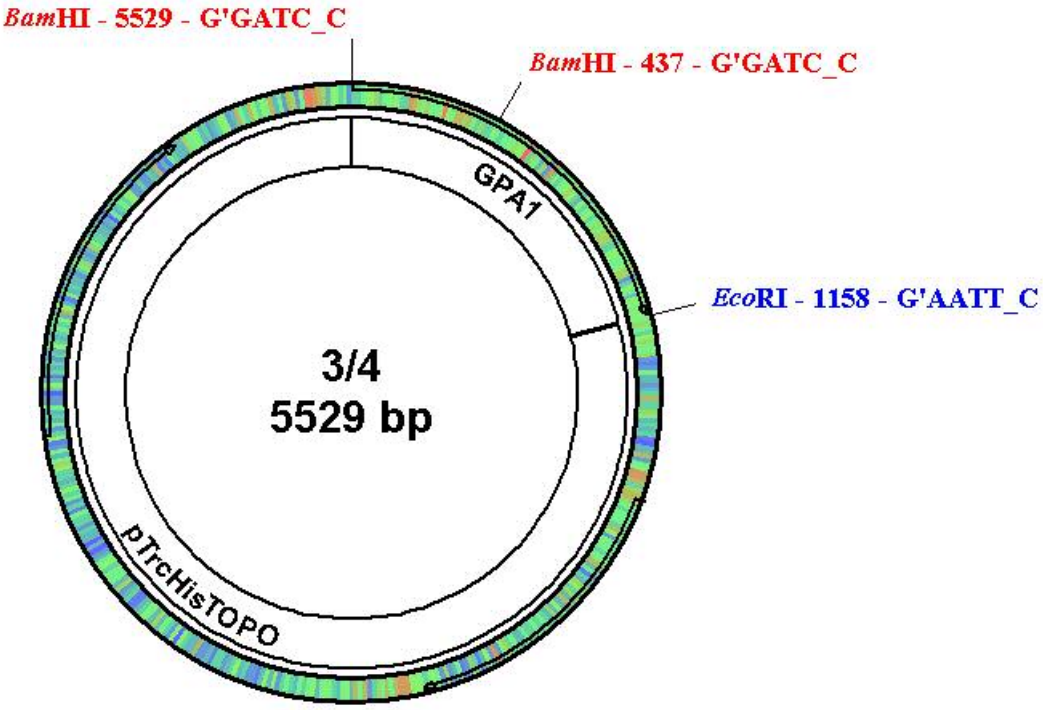
B.4 The construct named as 4/4_3'MCS is the pGFPuv vector containing *GPA1* at 3'MCS between the *Sac* I (5') and *Spe* I (3') restriction enzyme sites.



B.5 The construct named 2/3 is the pT7/NT TOPO vector containing *GPA1*



B.6 The construct named as 3 /4 is pTrcHis TOPO vector containing *GPA1*



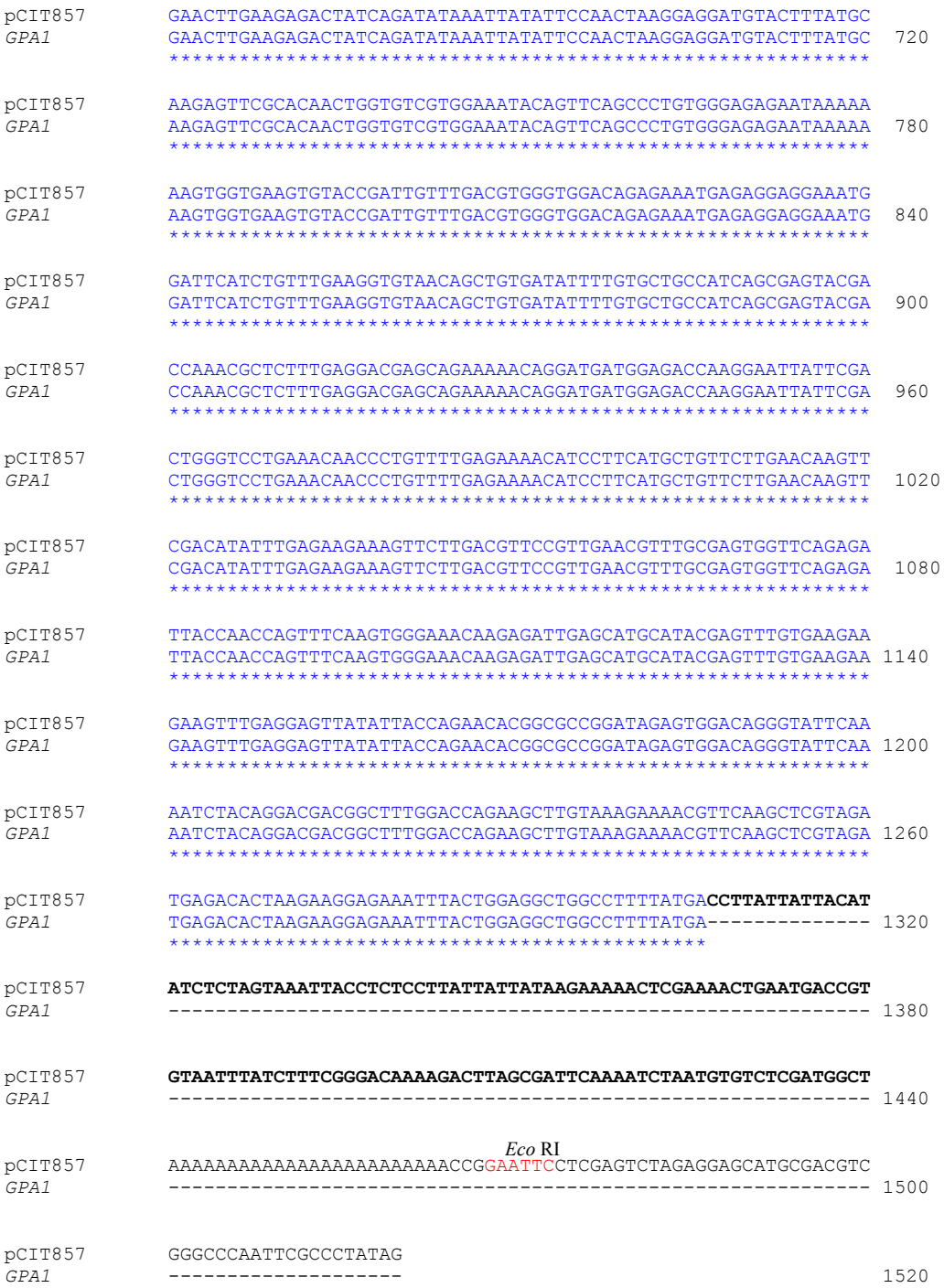
APPENDIX C

All the sequence alignments in the Appendix C were done on the SDSC Biology Workbench

C.1 Sequence alignment of pCIT857 with *GPA1* to verify the sequence of *GPA1*

Bam HI

```
pCIT857     GGTGACACTATAGAATACTCCAGCTATGCATCCAACGCGTTGGGAGCTCTCCGGATCCAA
GPA1        ------------------------------------------------------------  60

                                        Eco RI
pCIT857     GCTTATCGATTTCGAACCCGGGGTACCGAATTCCGGCTCCGATATCTTCTTCACTACCTT
GPA1        ------------------------------------------------------------  120

pCIT857     TGACTCCATTTCTTTTTCTTCTTCAGAAACAATCATGGGCTTACTCTGCAGTAGAAGTCG
GPA1        ----------------------------------ATGGGCTTACTCTGCAGTAGAAGTCG  180
                                               **************************

pCIT857     ACATCATACTGAAGATACTGATGAGAATACACAGGCTGCTGAAATCGAAAGACGGATAGA
GPA1        ACATCATACTGAAGATACTGATGAGAATACACAGGCTGCTGAAATCGAAAGACGGATAGA  240
            ************************************************************

pCIT857     GCAAGAAGCAAAGGCTGAAAAGCATATTCGGAAGCTTTTGCTACTTGGTGCTGGGGAATC
GPA1        GCAAGAAGCAAAGGCTGAAAAGCATATTCGGAAGCTTTTGCTACTTGGTGCTGGGGAATC  300
            ************************************************************

pCIT857     TGGAAAATCTACAATTTTTAAGCAGATAAAACTTCTATTCCAAACGGGATTTGATGAAGG
GPA1        TGGAAAATCTACAATTTTTAAGCAGATAAAACTTCTATTCCAAACGGGATTTGATGAAGG  360
            ************************************************************

pCIT857     AGAACTAAAGAGCTATGTTCCAGTCATTCATGCCAATGTCTATCAGACTATAAAATTATT
GPA1        AGAACTAAAGAGCTATGTTCCAGTCATTCATGCCAATGTCTATCAGACTATAAAATTATT  420
            ************************************************************

pCIT857     GCATGATGGAACAAAGGAGTTTGCTCAAAATGAAACAGATTCTGCTAAATATATGTTATC
GPA1        GCATGATGGAACAAAGGAGTTTGCTCAAAATGAAACAGATTCTGCTAAATATATGTTATC  480
            ************************************************************

pCIT857     TTCTGAAAGTATTGCAATTGGGGAGAAACTATCTGAGATTGGTGGTAGGTTAGACTATCC
GPA1        TTCTGAAAGTATTGCAATTGGGGAGAAACTATCTGAGATTGGTGGTAGGTTAGACTATCC  540
            ************************************************************

pCIT857     ACGTCTTACCAAGGACATCGCTGAGGGAATAGAAACACTATGGAAGGATCCTGCAATCCA
GPA1        ACGTCTTACCAAGGACATCGCTGAGGGAATAGAAACACTATGGAAGGATCCTGCAATCCA  600
            ************************************************************

pCIT857     GGAAACTTGTGCTCGTGGTAATGAGCTTCAGGTTCCTGATTGTACGAAATATCTGATGGA
GPA1        GGAAACTTGTGCTCGTGGTAATGAGCTTCAGGTTCCTGATTGTACGAAATATCTGATGGA  660
            ************************************************************
```



Identical nucleotides are shown in blue. Bold written nucleotides are 5' and 3' flanking regions of chromosome II of *Arabidopsis thaliana.*

C.2 Sequencing result of #44 (pGEM T-Easy+*GPA1*)

```
GPA1        TGGTGTCGTGGAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTA
#44         TGGTGTCGTGGAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTA  720
            ************************************************************

GPA1        CCGATTGTTTGACGTGGGTGGACAGAGAAATGAGAGGAGGAAATGGATTCATCTGTTTGA
#44         CCGATTGTTTGACGTGGGTGGACAGAGAAATGAGAGGAGGAAATGGATTCATCTGTTTGA  780
            ************************************************************

GPA1        AGGTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGA
#44         AGGTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGA  840
            ************************************************************

GPA1        GGACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACA
#44         GGACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACA  900
            ************************************************************

GPA1        ACCCTGTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTCGACATATTTGAGAA
#44         ACCCTGTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTCGACATATTTGAGAA  960
            ************************************************************

GPA1        GAAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTC
#44         GAAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTC  1020
            ************************************************************

GPA1        AAGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTT
#44         AAGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTT  1080
            ************************************************************

GPA1        ATATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGAC
#44         ATATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGAC  1140
            ************************************************************

GPA1        GGCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAG
#44         GGCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAG  1200
            ************************************************************

GPA1        GAGAAATTTACTGGAGGCTGGCCTTTTATGA----------------------------
#44         GAGAAATTTACTGGAGGCTGGCCTTTTATGAAATCACTAGTGAATTCGCGGCCGCCTGCA  1260
            ********************************          Eco RI  Not I

GPA1        ------------------------------------------------------------  1340
#44         GGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTG
                                                                SP6 Promoter

GPA1        ----------------------                                         1362
#44         TCACCTAAATATCTTGGCGTAA
```

C.3 sequencing result of TOPO8, wrong primer (pCR-XL + *GPA1* with *Eco* RI and *Xho* I restriction enzyme sites)

```
TOPO8          TATCGCCAGCTATTTGGTGACGCGTTAGAATACTCAAGCTATGCATCAAGCTTGGTACCG
GPA1           ------------------------------------------------------------  60
                                                *Eco* RI           *Eco* RI

TOPO8          AGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGCGTCGAATTCA
GPA1           -------------------------------------------------------------A 120
                                                                             *

TOPO8          TGGGCTTTCTCTGCAGTAGAAGTCGACATCATACTGAAGATACTGATGAGAATACACAGG
GPA1           TGGGCTTACTCTGCAGTAGAAGTCGACATCATACTGAAGATACTGATGAGAATACACAGG 180
               ******* **************************************************

TOPO8          CTGCTGAAATCGAAAGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATATTCGGAAGC
GPA1           CTGCTGAAATCGAAAGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATATTCGGAAGC 240
               ************************************************************

TOPO8          TTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGATAAAACTTC
GPA1           TTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGATAAAACTTC 300
               ************************************************************

TOPO8          TATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCATTCATGCCA
GPA1           TATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCATTCATGCCA 360
               ************************************************************

TOPO8          ATGTCTATCAGACTATAAAATTATTGCATGATGGAACAAAGGAGTTTGCTCAAAATGAAA
GPA1           ATGTCTATCAGACTATAAAATTATTGCATGATGGAACAAAGGAGTTTGCTCAAAATGAAA 420
               ************************************************************

TOPO8          CAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAAACTATCTG
GPA1           CAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAAACTATCTG 480
               ************************************************************

TOPO8          AGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGGAATAGAAA
GPA1           AGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGGAATAGAAA 540
               ************************************************************

TOPO8          CACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCTTCAGGTTC
GPA1           CACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCTTCAGGTTC 600
               ************************************************************

TOPO8          CTGATTGTACGAAATATCTGATGGAGAACTTGAAGAGACTATCAGATATAAATTATATTC
GPA1           CTGATTGTACGAAATATCTGATGGAGAACTTGAAGAGACTATCAGATATAAATTATATTC 660
               ************************************************************

TOPO8          CAACTAAGGAGGATGTACTTTATGCAAGAGTTCGCACAACTGGTGTCGTGGAAATACAGT
GPA1           CAACTAAGGAGGATGTACTTTATGCAAGAGTTCGCACAACTGGTGTCGTGGAAATACAGT 720
               ************************************************************

TOPO8          TCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTACCGATTGTTTGACGTGGGTG
GPA1           TCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTACCGATTGTTTGACGTGGGTG 780
               ************************************************************

TOPO8          GACAGAGAAATGAGAGGAGGAAATGGATTCATCTGTTTGAAGGTGTAACAGCTGTGATAT
GPA1           GACAGAGAAATGAGAGGAGGAAATGGATTCATCTGTTTGAAGGTGTAACAGCTGTGATAT 840
               ************************************************************

TOPO8          TTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGAGGACGAGCAGAAAAACAGGA
GPA1           TTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGAGGACGAGCAGAAAAACAGGA 900
               ************************************************************

TOPO8          TGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACAACCCTGTTTTGAGAAAACAT
GPA1           TGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACAACCCTGTTTTGAGAAAACAT 960
               ************************************************************

TOPO8          CCTTCATGCTGTTCTTGAACAAGTTCGACATATTTGAGAAGAAAGTTCTTGACGTTCCGT
GPA1           CCTTCATGCTGTTCTTGAACAAGTTCGACATATTTGAGAAGAAAGTTCTTGACGTTCCGT 1020
               ************************************************************
```

```
TOPO8          TGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTCAAGTGGGAAACAAGAGATTG
GPA1           TGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTCAAGTGGGAAACAAGAGATTG 1080
               ************************************************************


TOPO8          AGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTATATTACCAGAACACGGCGC
GPA1           AGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTATATTACCAGAACACGGCGC 1140
               ************************************************************


TOPO8          CGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACGGCTTTGGACCAGAAGCTTG
GPA1           CGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACGGCTTTGGACCAGAAGCTTG 1200
               ************************************************************

TOPO8          TAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGGAGAAATTTACTGGAGGCTG
GPA1           TAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGGAGAAATTTACTGGAGGCTG 1260
               ************************************************************

TOPO8          GCCTTTTATGACTCGAGGGGTTTAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGC
GPA1           GCCTTTTATGA-------------------------------------------------- 1310
               ********** Xho I             Eco RI

TOPO8          CGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTAAGTGAGTGATTCA           1360
GPA1           ---------------------------------------------
                   Xho I
```

Identical nucleotides are shown in blue

C.4 Sequencing results of P1 and P20 (pCR II-TOPO+*GPA1* with *Sac* I and *Spe* I restriction enzyme sites at 5' and 3' sites, respectively)

```
P20          GAATCACTCCTATAGGGCGAATTGGGCCCGCCGTCGCATGCTCCCGGCCGCCATGGCGGC
GPA1         ------------------------------------------------------------ 60
P1           ----ACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGC

                  Eco RI           Sac I
P20          CGCGGGAATTCGATTAAACCCGAGCTCATGGGCTTACTCTGCAGTAGAAGTCGACATCAT
GPA1         --------------------------ATGGGCTTACTCTGCAGTAGAAGTCGACATCAT 120
P1           CGCGGGAATTCGATTAA-CCCGAGCTCATGGGCTTACTCTGCAGTAGAAGTCGACATCAT
                                       *********************************

P20          ACTGAAGATACTGATGAGAATACACAGGCTGCTGAAATCGAAAGACGGATAGAGCAAGAA
GPA1         ACTGAAGATACTGATGAGAATACACAGGCTGCTGAAATCGAAAGACGGATAGAGCAAGAA 180
P1           ACTGAAGATACTGATGAGAATACACAGGCTGCTGAAATCGAAAGACGGATAGAGCAAGAA
             ************************************************************

P20          GCAAAGGCTGAAAAGCATATTCGGAAGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAA
GPA1         GCAAAGGCTGAAAAGCATATTCGGAAGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAA 240
P1           GCAAAGGCTGAAAAGCATATTCGGAAGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAA
             ************************************************************

P20          TCTACAATTTTTAAGCAGATAAAACTTCTATTCCAAACGGGATTTGATGAAGGAGAACTA
GPA1         TCTACAATTTTTAAGCAGATAAAACTTCTATTCCAAACGGGATTTGATGAAGGAGAACTA 300
P1           TCTACAATTTTTAAGCAGATAAAACTTCTATTCCAAACGGGATTTGATGAAGGAGAACTA
             ************************************************************

P20          AAGAGCTATGTTCCAGTCATTCATGCCAATGTCTATCAGACTATAAAATTATTGCATGAT
GPA1         AAGAGCTATGTTCCAGTCATTCATGCCAATGTCTATCAGACTATAAAATTATTGCATGAT 360
P1           AAGAGCTATGTTCCAGTCATTCATGCCAATGTCTATCAGACTATAAAATTATTGCATGAT
             ************************************************************

P20          GGAACAAAGGAGTTTGCTCAAAATGAAACAGATTCTGCTAAATATATGTTATCTTCTGAA
GPA1         GGAACAAAGGAGTTTGCTCAAAATGAAACAGATTCTGCTAAATATATGTTATCTTCTGAA 420
P1           GGAACAAAGGAGTTTGCTCAAAATGAAACAGATTCTGCTAAATATATGTTATCTTCTGAA
             ************************************************************


P20          AGTATTGCAATTGGGGAGAAACTATCTGAGATTGGTGGTAGGTTAGACTATCCACGTCTT
GPA1         AGTATTGCAATTGGGGAGAAACTATCTGAGATTGGTGGTAGGTTAGACTATCCACGTCTT 480
P1           AGTATTGCAATTGGGGAGAAACTATCTGAGATTGGTGGTAGGTTAGACTATCCACGTCTT
             ************************************************************

P20          ACCAAGGACATCGCTGAGGGAATAGAAACACTATGGAAGGATCCTGCAATCCAGGAAACT
GPA1         ACCAAGGACATCGCTGAGGGAATAGAAACACTATGGAAGGATCCTGCAATCCAGGAAACT 540
P1           ACCAAGGACATCGCTGAGGGAATAGAAACACTATGGAAGGATCCTGCAATCCAGGAAACT
             ************************************************************

P20          TGTGCTCGTGGTAATGAGCTTCAGGTTCCTGATTGTACGAAATATCTGATGGAGAACTTG
GPA1         TGTGCTCGTGGTAATGAGCTTCAGGTTCCTGATTGTACGAAATATCTGATGGAGAACTTG 600
P1           TGTGCTCGTGGTAATGAGCTTCAGGTTCCTGATTGTACGAAATATCTGATGGAGAACTTG
             ************************************************************

P20          AAGAGACTATCAGATATAAATTATATTCCAACTAAGGAGGATGTACTTTATGCAAGAGTT
GPA1         AAGAGACTATCAGATATAAATTATATTCCAACTAAGGAGGATGTACTTTATGCAAGAGTT 660
P1           AAGAGACTATCAGATATAAATTATATTCCAACTAAGGAGGATGTACTTTATGCAAGAGTT
             ************************************************************

P20          CGCACAACTGGTGTCGTGGAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGT
GPA1         CGCACAACTGGTGTCGTGGAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGT 720
P1           CGCACAACTGGTGTCGTGGAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGT
             ************************************************************

P20          GAAGTGTACCGATTGTTTGACGTGGGTGGACAGAGAAATGAGAGGAGGAAATGGATTCAT
GPA1         GAAGTGTACCGATTGTTTGACGTGGGTGGACAGAGAAATGAGAGGAGGAAATGGATTCAT 780
P1           GAAGTGTACCGATTGTTTGACGTGGGTGGACAGAGAAATGAGAGGAGGAAATGGATTTAT
             ************************************************************ **
```

```
P20          CTGTTTGAAGGTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACG
GPA1         CTGTTTGAAGGTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACG 840
P1           CTGTTTGAAGGTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACG
             ************************************************************

P20          CTCTTTGAGGACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTC
GPA1         CTCTTTGAGGACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTC 900
P1           CTCTTTGAGGACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTC
             ************************************************************


P20          CTGAAACAACCCTGTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTCGACATA
GPA1         CTGAAACAACCCTGTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTCGACATA 960
P1           CTGAAACAACCCTGTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTCGACATA
             ************************************************************

P20          TTTGAGAAGAAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAA
GPA1         TTTGAGAAGAAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAA 1020
P1           TTTGAGAAGAAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAA
             ************************************************************

P20          CCAGTTTCAAGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTT
GPA1         CCAGTTTCAAGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTT 1080
P1           CCAGTTTCAAGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTT
             ************************************************************

P20          GAGGAGTTATATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTAC
GPA1         GAGGAGTTATATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTAC 1140
P1           GAGGAGTTATATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTAC
             ************************************************************

P20          AGGACGACGGCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACA
GPA1         AGGACGACGGCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACA 1200
P1           AGGACGACGGCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACA
             ************************************************************
                                                  Spe I

P20          CTAAGAAGGAGAAATTTACTGGAGGCTGGCCTTTTATGAACTAGTGGGTTTAATCACTAG
GPA1         CTAAGAAGGAGAAATTTACTGGAGGCTGGCCTTTTATGA---------------------- 1260
P1           CTAAGAAGGAGAAATTTACTGGAGGCTGGCCTTTTATGAACTAGTGGGTTTAATCACTAG
             ************************************
              Eco RI
P20          TGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCAT
GPA1         ------------------------------------------------------------ 1320
P1           TGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCAT


P20          AGCTTGAGTATTCTATAGTGTCACCTAAAGCTTGCGATA
GPA1         ---------------------------------------                      1359
P1           AGCTTGAGTATTCTATAGTGTCCCCTAAATAGCTTGGCG
```

Identical nucleotides are shown in blue

C.5 Sequencing results of construct 4/4_3'MCS (*GPA1* inserted in 3'MCS of pGFPuv)

```
4/4_3'MCS      NAAGCCNTTGTNNTGTTGNTGGTTTAAACCGGACAAAACGGTTAGGATTCNAGNTCAGCN
GPA1           ------ATGGGCTTACTCTGCAGTAGAAGTCGACATCATACTGAAGATACT-GATGAGAA 60
                     * *   *  *       *  **   ****  *   * * *** *  * * **

4/4_3'MCS      CCCTCACGNNGTTTNTNACNTTTAAAAAA-ACGTCCCNCTTCAAACCATGTCAAANAAAG
GPA1           TACACAGGCTGCTGAAATCGAAAGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATAT 120
                 * ** *  * *     *     *   * *   *      ****   **  **  * *

4/4_3'MCS      TCATGNCCANTTTCTANCNGCCTTCGAGCATTCTCCCGAGNACCCCNCCCCGTCATTG-C
GPA1           TCGGAAGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGAT 180
               **     *  ** ***   *     * * * ***    *     *           *

4/4_3'MCS      AGTGATGCNGCTTTAAA----GTNCCATGGATAAATCNGTNAAAAGTGGGTTGGNNCTGG
GPA1           AAAACTTCTATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCAT 240
               *    * *   *  ***     *   *** *  *       * *  *  ***

4/4_3'MCS      TTAAGTCGA-----GTAAGA--GGAAAAGTACAATATGCAGACCTAGGAGCGGAAACCTG
GPA1           TCATGCCAATGTCTATCAGACTATAAAATTATTGCATGATGGAACAAAGGAGTTTGCTCA 300
               * * * * *      * ***    **** **    ***  *    *   * *    *

4/4_3'MCS      ----GAAACCTAT-CAGCAACTTACATGATATGTCATCTAGTCACT-CAAAGACTTTTAG
GPA1           AAATGAAACAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAA 360
                   *****  ** * ** *  ** *** *** *  *  *   * * ***        *

4/4_3'MCS      GCTATAAGAGGCTGACTAAAAGCAATCCAATC----TCATGCCAA--ATTCTATTGTATA
GPA1           ACTATCTGAGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGG 420
                ****  ***  **     * *  *  * ***    ** * ****  *      **

4/4_3'MCS      AATCT---CGCTCTAAAGG----------CCGAAAAAATCAGCGCTCGAC---ACGGACT
GPA1           AATAGAAACACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCT 480
               ***     * ** *  * *           * *  ***   * *****     * *  **

4/4_3'MCS      CATTGTCACCACCCGTCACCTAAAATCTACTCAGCGTCGGCAAAGGAGCCATGGATTCTA
GPA1           TCAGGTTCCTGATTGT-ACGAAATATCTGAT-GGAGAACTTGAAGAGACTATCAGATATA 540
                   **  *     ** **  ** ****  *  * *      ***   * **    * **

4/4_3'MCS      GCAACTAACTTACCTGTTGAAATTCGAACACCCAAACAACTTGTTAATATCTATTCGAAG
GPA1           AATTATATTCCAACTAAGGAGGATGTACTTTATGCAAGAGTTCGCACAACTGGTGTCGTG 600
                    **    * **   **   *  *        *  * **   *  *    *     *

4/4_3'MCS      CGAATGCAGATTGAAGAAACCTTCCGAGACTTGAAAAGTCCTGCCTACGGACTAGGCCTA
GPA1           GAAATACAGTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAA-GTGTACCGATTGTT 660
                 *** *** *                 *    ******  **     * **      *

4/4_3'MCS      CGCCATAGCCGAACGAGCAGCTCAGAGCGTTTTGATATCATGCTGCTAATCGCCCTGATG
GPA1           TGACGTGGGTGGACAGAGAAATGAGAG------GAGGAAATGG-ATTCATCTGTTTGAAG 720
                * * * *  * **    *  * ****      **    ***    * ***    *** *

4/4_3'MCS      CTTCAACTAACATGTTGGCTTGCG--GGCGTTCATGCTCAGAAACAAGGTTGGGACAAGC
GPA1           GTGTAACAGCTGTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGAGG 780
                *  ***     ** *   *** *  * * *    *   *  * ***          **

4/4_3'MCS      ACTTCCAGGCTAACACAGTCAGAAATCGAAACGTACT-CTCAACAGTTCGCTTAGGCATG
GPA1           ACGAGCAGAAAAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACAAC 840
               **   ***   ****   * *   *    ** * * *  ** ** *    ** *  **

4/4_3'MCS      GAAGTTTTGCGGCA------TTCTGGCTACACAATAACAAGGGAAGACTTACTCGTGGCT
GPA1           CCTGTTTTTGAGAAAACATCCTTCATGCTGTTCTTGAACAAGTTC-GACATATTTGAGAAG 900
                  ****** *  *      ***  ***   *   ******    *** ** * * *

4/4_3'MCS      GCAACCCTACTAGCTCAAA-ATTTATTCACACATGGTT---ACGCTTTGGGGAAATTATG
GPA1           AAAGTTCTTGACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTCA 960
                 *   **    * **          **  *   *****   *   *       * **

4/4_3'MCS      AGGGGATCTCTCAGGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTA
GPA1           AGTGGGAAACAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTA 1020
               ** **    *    ******************************************
```

```
4/4_3'MCS       TATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACG
GPA1            TATTACCAGAACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACG 1080
                ************************************************************

4/4_3'MCS       GCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGG
GPA1            GCTTTGGACCAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGG 1140
                ************************************************************

4/4_3'MCS       AGAAATTTACTGGAGGCTGGCCTTTTATGAACTAGTCGGCCGTACGGGCCCTTTCGTCTC 1200
GPA1            AGAAATTTACTGGAGGCTGGCCTTTTATGA------------------------------
                **************************** Spe I

4/4_3'MCS       GCG                                                          1203
GPA1            ---
```

Identical residues are shown in blue.

C.6 Sequencing result of 2/3 (pT7/NT TOPO expression+*GPA1*)

```
2/3          ------------------------------------------------------------ 60
GPA1         ATGGGCTTACTCTGCAGTAGAAGTCGACATCATACTGAAGATACTGATGAGAATACACAG

2/3          ------------------------------------------------------------ 120
GPA1         GCTGCTGAAATCGAAAGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATATTCGGAAG

2/3          ------------------------------------------------------------ 180
GPA1         CTTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGATAAAACTT

2/3          ------------------------------------------------------------ 240
GPA1         CTATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCATTCATGCC

2/3          ------------------------------------------------------------ 300
GPA1         AATGTCTATCAGACTATAAAATTATTGCATGATGGAACAAAGGAGTTTGCTCAAAATGAA

2/3          ------------------------------------------------------------ 360
GPA1         ACAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAAACTATCT

2/3          ------------------------------------------------------------ 420
GPA1         GAGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGGAATAGAA

2/3          ------------------------------------------------------------ 480
GPA1         ACACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCTTCAGGTT

2/3          ------------------------------------------------------------ 540
GPA1         CCTGATTGTACGAAATATCTGATGGAGAACTTGAAGAGACTATCAGATATAAATTATATT

2/3          -------------------------ANAGTTTTANGNTTAANNGTTTTCNTTTCCAAAAAA
GPA1         CCAACTAAGGAGGATGTACTTTATGCAAGAGTTCGCACAACTGGTGTCGTGGA-AATACA 600
                                       *    *  *   **  * * ** *    ** * *

2/3          ANCCCGCCCNCAAAGCAANATNNCGGAAAAANATAATGNANGCTACNCNT---ACNTNNN
GPA1         GTTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTACCGATTGTTTGACGTGGG 660
               * ****     *  * *      ***    * * *    *      *   ** *

2/3          NNTACNNNCAAAANAANGNNCTNNAGNNTNTTTTTTTTTNNGNAGGTCTNNCGGNNGAAN
GPA1         TGGACAGAGAAATGAGAGGAG-GAAATGGATTCATCTGTTTGAAGGTGTAACAGCTGTGA 720
               **    ***  *  *      *     **  * * *  * **** *  * *  *

2/3          NANAGCTNGANAGCATACNNAGGAGANAN-------------------------------- 780
GPA1         TATTTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGAGGACGAGCAGAAAAACA
              *      *    ***      *    *

2/3          ------------------------------------------------------------ 840
GPA1         GGATGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACAACCCTGTTTTGAGAAAA

2/3          ------------------------------------------------------------ 900
GPA1         CATCCTTCATGCTGTTCTTGAACAAGTTCGACATATTTGAGAAGAAAGTTCTTGACGTTC

2/3          ------------------------------------------------------------ 960
GPA1         CGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTCAAGTGGGAAACAAGAGA

2/3          ------------------------------------------------------------ 1020
GPA1         TTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTATATTACCAGAACACGG

2/3          ------------------------------------------------------------ 1080
GPA1         CGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACGGCTTTGGACCAGAAGC
```

```
2/3          ------------------------------------------------------------ 1140
GPA1         TTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGGAGAAATTTACTGGAGG


2/3          --------------                                               1154
GPA1         CTGGCCTTTTATGA
```

Identical residues are shown in blue

C.7 Sequencing result of 3 /4 (pTrcHisTOPO +*GPA1*)

```
3/4          TNTGTGTNCGTCTGAAAAACTCNNCTTTTTNTTTAAATTTCTTTCCCTTNGACNTCCCCC
GPA1         ------------------------------------------------------------ 60


3/4          TTATGGGCTTACTCTGCAGTAGANGTCGACATCATACTGAAGATACTGATGAGAATACAC
GPA1         --ATGGGCTTACTCTGCAGTAGAAGTCGACATCATACTGAAGATACTGATGAGAATACAC 120
               ******************** ************************************


3/4          AGGCTGNTTTTTTTGAANGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATATTCGGA
GPA1         AGGCTGCTGAAATCGAAAGACGGATAGAGCAAGAAGCAAAGGCTGAAAAGCATATTCGGA 180
             ****** *    * *** ****************************************


3/4          AGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGATAAAAC
GPA1         AGCTTTTGCTACTTGGTGCTGGGGAATCTGGAAAATCTACAATTTTTAAGCAGATAAAAC 240
             ************************************************************


3/4          TTCTATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCATTCATG
GPA1         TTCTATTCCAAACGGGATTTGATGAAGGAGAACTAAAGAGCTATGTTCCAGTCATTCATG 300
             ************************************************************


3/4          CCAATGTCTATCAGACTATAAAATTATTGCATGATGGNACAAAGGAGTTTGCTCAAAATG
GPA1         CCAATGTCTATCAGACTATAAAATTATTGCATGATGGAACAAAGGAGTTTGCTCAAAATG 360
             ************************************ ************************


3/4          AAACAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAAACTAT
GPA1         AAACAGATTCTGCTAAATATATGTTATCTTCTGAAAGTATTGCAATTGGGGAGAAACTAT 420
             ************************************************************


3/4          CTGAGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGGAATAG
GPA1         CTGAGATTGGTGGTAGGTTAGACTATCCACGTCTTACCAAGGACATCGCTGAGGGAATAG 480
             ************************************************************


3/4          AAACACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCTTCAGG
GPA1         AAACACTATGGAAGGATCCTGCAATCCAGGAAACTTGTGCTCGTGGTAATGAGCTTCAGG 540
             ************************************************************


3/4          TTCCTGATTGTACGAAATNTCTGATGGATAACTTTAAGAGACTATCAGATATAAATTATA
GPA1         TTCCTGATTGTACGAAATATCTGATGGAGAACTTGAAGAGACTATCAGATATAAATTATA 600
             ****************** ********* ***** ************************


3/4          TTCCAACTAAGGAGGNTGTACTTTATNCAAGAGTTCGTACAACTGGTGTNGTGGAAATAC
GPA1         TTCCAACTAAGGAGGATGTACTTTATGCAAGAGTTCGCACAACTGGTGTCGTGGAAATAC 660
             *************** ********** ********** *********** **********


3/4          ANGTTCAGCCCTGTGGGAGAGAATANAAAAAGCGGTGAAGTGTACCTGATTGNTTGACGT
GPA1         AG-TTCAGCCCTGTGGGAGAGAATAAAAAAAGTGGTGAAGTGTACC-GATTGTTTGACGT 720
             *  ********************** ****** ************* ***** *******
```

```
3/4          GGGTGGNCANAGAAATCGNGAGGAGGAAATGGATTCATTTGNTTTGTAGGTTTACACCCT
GPA1         GGGTGGACAGAGAAAT-GAGAGGAGGAAATGGATTCATCTGTTT-GAAGGTGTAACAGCT 760
             ****** ** ****** * ****************** ** ** * **** **    **

3/4          GNGATATTTTGTGCTGCCATTAGCTAGTACTACCAAACGCTCTTTTATGACGAGCCATAN
GPA1         GTGATATTTTGTGCTGCCATCAGCGAGTACGACCAAACGCTCTTTGAGGACGAGC-AGAA 820
             * ****************** *** ***** ************** * ******* * *

3/4          TAACACGATGATTGTGACCAANGGAATTTTCNACTGGGTCCNGAAACAACCCTGCTTTGA
GPA1         AAACAGGATGATGGAGACCAAGGAATTATTCGACTGGGTCCTGAAACAACCCTGTTTTGA 880
              **** ****** * ****** * * * *** ********* ************ *****

3/4          GAAAACAGNCTTTATGCGTGTTCTTGNACAANGTTCGTTNTATTTGNGA-GNAAGTTNTC
GPA1         GAAAACATCCTTCATGC-TGTTCTTGAACAAG-TTCGACATATTTGAGAAGAAAGTTCTT 940
             *******  *** **** ******** ****  ****   ****** ** * ***** *

3/4          GACGT-CCGNTGANCGTT-GCGAGTGGGTCNGAGNNTNCCA-CCNGTCCAAANNGGGAAA
GPA1         GACGTTCCGTTGAACGTTTGCGAGTGGTTCAGAGATTACCAACCAGTTTCAAGTGGGAAA 1000
             ***** *** *** **** ******** ** ***  * *** ** **   **  ******

3/4          CAANAGANTGNNCNTTNTTN----------------------------------------
GPA1         CAAGAGATTGAGCATGCATACGAGTTTGTGAAGAAGAAGTTTGAGGAGTTATATTACCAG 1060
             *** *** **  * *   *

3/4          ------------------------------------------------------------ 1120
GPA1         AACACGGCGCCGGATAGAGTGGACAGGGTATTCAAAATCTACAGGACGACGGCTTTGGAC

3/4          ------------------------------------------------------------ 1180
GPA1         CAGAAGCTTGTAAAGAAAACGTTCAAGCTCGTAGATGAGACACTAAGAAGGAGAAATTTA

3/4          ---------------------                                        1201
GPA1         CTGGAGGCTGGCCTTTTATGA
```

Identical residues are shown in blue

APPENDIX D

Autoclave:	Hirayama, Hiclave HV-110, JAPAN
	Certoclav, Table Top Autoclave CV-EL-12L, AUSTRIA
Balance:	Sartorius, BP211D, GERMANY
	Sartorius, BP221S, GERMANY
	Sartorius, BP610, GERMANY
	Schimadzu, Libror EB-3200 HU, JAPAN
Centrifuge:	Eppendorf, 5415C, GERMANY
	Eppendorf, 5415D, GERMANY
	Eppendorf, 5415R, GERMANY
	Kendro Lab. Prod., Heraeus Multifuge 3L, GERMANY
	Hitachi, Sorvall RC5C Plus, USA
	Hitachi, Sorvall Discovery 100 SE, USA

Deepfreeze:	-70^{o} C, Kendro Lab. Prod., Heraeus Hfu486 Basic, GERMANY
	-20^{o} C, Bosch, TÜRKİYE
Distilled Water:	Millipore, Elix-S, FRANCE
	Millipore, MilliQ Academic, FRANCE
Electrophoresis:	Biogen Inc., USA
	Biorad Inc., USA
Gel Documentation:	UVITEC, UVIdoc Gel Documentation System, UK
	Biorad, UV-Transilluminator 2000, USA
Ice Machine:	Scotsman Inc., AF20, USA
Incubator:	Memmert, Modell 300, GERMANY
	Memmert, Modell 600, GERMANY
Laminar Flow:	Kendro Lab. Prod., Heraeus, HeraSafe HS12, GERMANY
Magnetic Stirrer:	VELP Scientifica, ARE Heating Magnetic Stirrer, ITALY
	VELP Scientifica, Microstirrer, ITALY
Microliter Pipette:	Gilson, Pipetman, FRANCE
	Mettler Toledo, Volumate, USA

Microwave Oven:	Bosch, TÜRKİYE
pH meter:	WTW, pH540 GLP MultiCal®, GERMANY
Power Supply:	Biorad, PowerPac 300, USA
	Wealtec, Elite 300, USA
Refrigerator:	+4^{o} C, Bosch, TÜRKİYE
Shaker:	Forma Scientific, Orbital Shaker 4520, USA
	GFL, Shaker 3011, USA
	New Brunswick Sci., Innova™ 4330, USA
Spectrophotometer:	Schimadzu, UV-1208, JAPAN
	Schimadzu, UV-3150, JAPAN
	Secoman, Anthelie Advanced, ITALY
Speed Vacuum:	Savant, Speed Vac® Plus Sc100A, USA
	Savant, Refrigerated Vapor Trap RVT 400, USA
Thermocycler:	Eppendorf, Mastercycler Gradient, GERMANY
Vacuum:	Heto, MasterJet Sue 300Q, DENMARK
Water bath:	Huber, Polystat cc1, GERMANY