"Molecular studies on Glycosyltransferase gene(s) from *Withania somnifera*"

A THESIS
SUBMITTED TO THE
UNIVERSITY OF PUNE

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

BY
R. J. Santosh Kumar

UNDER THE GUIDANCE OF
Dr. B. M. KHAN

PLANT TISSUE CULTURE DIVISION
NATIONAL CHEMICAL LABORATORY
PUNE- 411008
INDIA

August, 2011

Dedicated to my
Guru
&
Parents
&
Friends

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Molecular studies on Glycosyltransferase gene(s) from *Withania somnifera*"** submitted by R. J. Santosh Kumar for the degree of Doctor of Philosophy, was carried out under my supervision at the Plant Tissue Culture Division, National Chemical Laboratory, Pune. Materials obtained from other sources have been duly acknowledged in the thesis.

Dr. B. M. Khan
(Research guide)

Declaration

I hereby declare that the thesis entitled **"Molecular studies on Glycosyltransferase gene(s) from *Withania somnifera*"** has been carried out at Plant Tissue culture Division, National Chemical Laboratory, Pune, under the guidance of **Dr. Bashir M. Khan**. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other university. I further declare that the materials obtained from other sources have been duly acknowledged in the thesis.

(R. J. Santosh Kumar)

Date: August, 2011
Place: Plant Tissue Culture Division,
National Chemical Laboratory (NCL),
Pune – 411008,
India.

ACKNOWLEDGEMENTS

First of all I would like to express my deep sense of gratitude towards my research guide, **Dr. B. M. Khan** for his subtle guidance, unmatched humanity, constant support and encouragement, constructive criticism, valuable suggestions and for giving me freedom in finalizing the research and preparation of this manuscript without which the task would have been difficult. His enthusiastic attitude, innovative ideas and scientific knowledge have inspired me profoundly. I truly feel privileged to have joined his research group. I sincerely thank for the care and affection that I received from him and his family in the entire period.

I would be equally indebted to **Dr. S. K. Rawal**, Former Head, PTC Division, NCL, for his moral support, valuable comments and impeccable suggestions during the progress and completion of this work.

I am really greatful to Mrs. S. Kendurkar for her help in tissue culture work, Dr. Srinivas Hotha for providing me the LC-MS facility, Dr. M. Kulkarni for providing me the MALDI facility, Dr. Subhada Thengane for providing me the HPLC facility.

I would like to extend my thanks to Dr. Vidya S. Gupta, Dr. Sushma gaikwad, Dr. D.C. Agarwal, Dr. Ashok Giri, Dr. john, Dr. M. Phadka, Dr. D. K. Kulkarni, Mrs. M. V. Shirgurkar, Mrs. Varsha parushurami, Dr. Urmil J. Mehta, Mr. Surya Prasad, Mr. Jattar and all other members of the scientific and supporting staff of the Plant Tissue Culture Division, Biochemical Sciences Division and NCL Library.

My special and sincere thanks to my friends, seniors and advisors **Dr. Rajashekhar reddy, Dr. Bhushan, Dr. Ajay, Dr. Suhas** and **Dr. Abhay** the people behind my success.

My sincere thanks to Krunal & Disha for their help in modeling and docking studies, Asif shiekh for his help in conducting LC-MS experiments and Yashwant and Ms. Sweta for their help in MALDI analysis.

Many thanks to my senior friends Dr. Ramachandra , Dr. Sucheta, Dr. Rohini, Dr. Neelima, Dr. Nookaraju, Dr. Manish, Dr. Sushim, Dr. Noor, Dr. Sameer, Dr. Arun, Dr. Pallavi, Dr. Abhilash, Dr. Ruby, Dr. Bhuban, Dr. Sujata, Dr. Kiran for their encouragement and help during the course of this work.

I would like to thank my friends Jay kumar, Azfar, Ulfat, Rajesh kumar and Malini.

I would like to thank my labmates, colleagues and friends Rishi, Somesh, Sweta S, Sameena, Poonam, Neha, Parth, Krunal, Sumita, Santosh gupta, Kannan, suchi, Neeraja, kishoree, Manika, Shakeel, Uma, Sonali, Nazneen, Zohaib, Sneha, Pratik, Amar, Anuja, Sobia, Shruti, Richa, Srikanth, Sudarshan, Yasir, Nishi, Rajshree, Gayatri .I, Lumbini, Sulakshana, Asma S, Aditi, preeti, kiran, gugan, jai, Neha .K, Tripti, prashant, Amol for maintaining lively and cordial atmosphere in the lab.

I thank my students Kriti, Dhanajay and Ankita for their help during my work period.

It gives me great pleasure to thank my friends Ravi P, Anand J, Kailash D, Monali P, Manjae, Prabhas, Krishna, Shashi anna, Nagraj, Atul thakur, Atul Kumar, Sajid bhai, Mujju, Fazal, Swapna D, Trupti J, Raju D, Sweta S, Nitasha, Asad, Shadab, Srikanth, Ravi, Viku, Ashish, Chetan, Pushkar, Ejaz, Pooja, Tonima, Saurabh, Ansari, Avinash, Sridevi, Ambrish, Nishant, Ram, vittal, Abhisek M, Shameem, Upasan S, Sampa S.

Special thanks must for my college friends Chandoo, maruthi, G.V., Ravi kanth, pratap, Prasanth, shamsundar, jay kumar, Malathi, Basha, Pulla reddy.

I am highly indebted to my friends Trupti K, Prashant S and Gayatri S for constant support and making my stay more comfortable by their presence. I also cherish all those moments I shared with them.

Many thanks to Deepak, Vaibhav and Varsh.

It is difficult to express my gratitude towards my parents and family members, without whose encouragement and unstinted moral support, this endeavor would have never been possible. The main credit of my thesis goes to my dear parents (**Mr. Ramachandra jadhav and Hirabai jadhav**) whose unending love, encouragement and sacrifice has made me achieve this milestone today. Also to my dearest brother **Sunil** and sister **Nirmala** for their cheerful, loving nature and blessings was also instrumental in the completion of my thesis and obtaining the degree of Philosophy.

Last but not the least, it is difficult to express my gratitude towards my beloved wife **Vaishali**, who is indeed, the great force behind my perseverance; without whom, nothing would have been possible and nothing is going to be possible in future too.

I am grateful to Dr. Saurav Pal, Director, NCL for providing necessary facilities and permitting me to submit my findings in the form of thesis.

The fellowship awarded by the CSIR, India is duly acknowledged.

I thank **almighty God** for giving me the strength and courage at every step of life.

R. J. Santosh Kumar

CONTENTS

Abbreviations

AA	Amino acid
ABA	Abscisic acid
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
Ci/ mmol	Curie per milli mole
CIAP	Calf Intestinal Alkaline Phosphatase
Cps	Counts per second
Da	Dalton
DEPC	Diethylpyrocarbonate
DNA	Deoxyribose nucleic acid
DTT	Dithiothritol
EDTA	Ethylene Diamine Tetra Acetic acid
ELISA	Enzyme linked immuno sorbent assay
EtBr	Ethidium bromide
GT	Glycosyltransferase
g /L	Grams per litre
g	Gram
G	Guaiacyl
gDNA	Genomic DNA
GSP	Gene specific primers
Glc T	Glycosyltransferases
Gat T	Galactosyltransferases
h	Hour(s)
ha	Hectares
IPTG	Isopropyl ß-D-thiogalactoside
IEF	Isoelectric Focusing
Kb	Kilobase pairs

KDa	Kilo Daltons
Kg	Kilogram
L	Litre
LB	Luria-Bertani
M	Molar
MS	Murashige and Skoog
MCS	Multiple cloning sites
M	meter
mg	Milligram
min	Minute(s)
mL	Millilitre
µL	Micro liter
µg	Microgram
mM	Millimolar
mRNA	Messenger RNA
nM	Nano molar
nm	Nanometer
NUP	Nested Universal Primers
O/N	Overnight
OD	optical density
pI	Isoelectric point
PEG	Polyethylene glycol
pg	Picogram
pmol	Picomole
PMSF	Phenyl methyl sulphonyl fluoride
ppm	Parts per million
psi	pounds per square inch
PVPP	Poly vinyl pyro phosphate
pmol	Pico mole
PSPG	Plant secondary product glycosyltransferase
Q-PCR	Quantitative polymerase chain reaction

RACE	Rapid amplification of cDNA ends
RNase A/ (H)	Ribonuclease A/(H)
RNA	Ribose nucleic acid
rpm	Rotations per minute
RT	Room temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
s	Second(s)
SDS	Sodium dodecyl sulphate (sodium lauryl sulphate)
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMQ	Sterile Milli Q
Soln	Solution
sp.	Species
SSC	Saline sodium citrate
TAE	Tris acetic EDTA buffer
TE	Tris EDTA buffer
TEMED	Tetramethylethylenediamine
TLC	Thin layer chromatography
U	Units
UDP	Uridine diphospate glucose
UGT	UDP-glucose glycosyltransferases
UDP-GT	UDP-glycosyltransferase
UPM	Universal Primer Mix
UTR	Untranslated Region
UV	Ultraviolet
V	Volt
v/v	Volume / Volume
WsGT	*Withania somnifera* Glycosyltransferase
WsSAPGT	*Withania somnifera* saponin glycosyltransferase
w/v	Weight / Volume
X-gal	5-bromo-4-chloro-3-indolylß-D-galactoside
α	Alpha

β	Beta
λ	Lamda
%	Percentage
°C	Degree Celsius
µg	Microgram
µg/L	Micrograms per liter
µL	Microlitre
µm	Micrometer
µM	Micromolar

Abstract

Withania somnifera (Ashwagandha) belongs to the family Solanaceae. The roots are a constituent of over 200 formulations in Ayruvedha, Siddha and Unani medicine, which are used in the treatment of various physiological disorders. *Withania* is widely claimed to have potent aphrodisiac, sedative, rejuvenative and life prolonging properties. It is used as a general energy-enhancing tonic known as 'Medharasayana' which means 'to promote learning and a good memory' and in geriatric problems.

The chemistry of *Withania* species has been extensively studied and several groups of chemical constituents such as steroidal lactones, alkaloids, flavonoids, tannin and many more have been identified, extracted and isolated. At present, more than 12 alkaloids, 40 withanolides, and several sitoindosides (a withanolide containing a glucose molecule at carbon 27) have been isolated and reported from aerial parts, roots and berries of *Withania* species. The major chemical constituents of these plants, withanolides, are mainly localized in leaves and roots.

Glycosylation is the final step in the biosynthesis of secondary plant products resulting in the formation of an overwhelming number of natural glucosides with numerous applications. Plant-derived glycosides have attracted much attention due to their widespread applications. The isolation and purification of such glycosides from plant sources is, however, tedious and usually results in low yields. Thus, simple approaches for the generation of such glycosides would be highly beneficial. Since glycosylation confers amphipathic properties on the molecule and is normally critical for biological activity, this is clearly an important area in which an effort should be invested.

Glycosyltransferases (GTs)

Glycosylation reactions are of great biological importance to both prokaryotes and eukaryotes, and require the coordinated action of a large number of enzymes, the glycosyltransferases (GTs). These enzymes transfer the sugar moiety from an activated

nucleotide–sugar to an acceptor, which may be a growing oligosaccharide, a lipid or a protein. The enzymes can be classified into families on the basis of sequence similarity, catalytic specificity and the existence of consensus sequences. There are over 91 glycosyltransferase families. Despite the fact that many GTs recognize identical donor or acceptor substrates, few regions of sequence homology have been found among the different classes of eukaryotic glycosyltransferases and enzymes that are structurally related most often catalyze the same or a similar reaction. It seems these glycosyltransferases are "managers of metabolism", playing a role in cellular homeostasis and through their activity, regulating metabolic flux, levels of active hormones and the detoxification of xenobiotics.

Plants are exposed to a wide range of toxic and bioactive low-molecular-weight molecules from both exogenous and endogenous sources. Glycosylation is one of the primary sedative mechanisms that plants utilize in order to maintain metabolic homeostasis. Plants are capable of synthesizing several thousand different low molecular weight compounds, defined as secondary plant metabolites. Part of this diversity arises from decoration with glycosyl-, carboxyl-, methyl- and hydroxyl- groups by glycosyltransferases, acyltransferases, methyltransferases and cytochrome P450s, respectively. The conjugation of endogenous and exogenous organic molecules with sugar is one such important tool employed by all organisms. Secondary plant metabolites are glycosylated to O (OH- and COOH-), N, S and C atoms by glycosyltransferases (GTs) using nucleotide-activated sugars as donor substrates. Hydroxylated molecules are the most common acceptors, whilst UDP-glucose is the most common donor.

Glycosylation converts reactive and toxic aglycones into stable and non-reactive storage forms, thereby limiting their interaction with other cellular components. Generally speaking, the attachment of carbohydrate moieties to nucleophilic molecules will reduce the possibility of electron transfer from the unglycosylated molecule to other cellular components, thereby lowering the reactivity, and consequentially improving the stability of that molecule. Since the nucleophilic sites are, in many cases, the particular part of the molecules that interact damagingly with other cellular components, the addition of sugars will block the interactive site and consequently reduce toxicity. Sugars are highly polar, and the addition of

carbohydrate moieties to hydrophobic substrates will serve to increase the water solubility of the resultant glycoside.

However, biosynthetic pathway of withanolides, their functions in *W. somnifera* and metabolic steps leading to their glycol-transformation are unknown. One of the reasons for limited knowledge of their functions is non-availability of the relevant enzymes and genes. The present study was taken up with the objective to isolate, clone and characterize glycosyltransferase gene (s) (GT) from *W. somnifera*, which plays crucial role in solubilization of medicinally important secondary metabolites. The other objective was to study the expression of protein in heterologous system and see the expression pattern in different parts of the plant. The present work in this thesis divided into five chapters.

Chapter 1. General Introduction

This chapter gives general information on medicinal plants in India. Background of research done on *Withania somnifera*, a medicinal plant has been dealt in detail. A thorough literature survey of work done in the area of glycosyltransferase genes with regards to the current status of research in this area has been presented. A special emphasis on the techniques and strategies used in study of the glycosyltransferase gene has been dealt with. Role of the key enzyme in the formation of different secondary metabolites with special reference to health has been provided. Finally, the scope of the present study and objectives of the thesis work have been discussed.

Chapter 2. Materials and Methods

Overall materials and the protocols followed during the course of this work in detail have been included in this chapter.

Chapter 3. Isolation, Cloning and Characterization of *W. somnifera* glycosyltransferase gene(s).

This chapter deals with the PCR based approach for fishing out the full-length c-DNA clones of glycosyltransferase gene(s) from *W. somnifera*. Primers were designed on the basis of consensus regions of various reported nucleotide sequences of the GT genes from the NCBI GenBank database. A partial GT gene sequence was amplified and the sequence showed maximum homology with GT members of lamiaceae family. The partial GT gene sequence was used to design gene specific primers in order to get the full length genes. Rapid Amplification of cDNA ends (RACE) was performed to fish out the full length cDNA clones. The two different glycosyltransferase genes WsGT and WsSAPGT were isolated cloned and sequenced. Phlogenetic analysis of WsGT gene (Ac no FJ560880) shows maximum similarity with *Lycium barbarum* UDP-glucose glucosyltransferase and WsSAPGT (Ac no FJ560881) shows maximum similarity with *Solanum aculeatissimum* UDP-glucose glucosyltransferase. Characterization of the gene(s) encoding GT is embodied in this chapter.

Chapter 4

(4.A): Heterologous expression of *W. somnifera* GT gene(s), its purification and characterization

This chapter deals with the cloning of GT gene(s) (cDNA) in the expression vector pET30 b (+) and its expression in *E. coli* BL21 (DE3). The protein was purified from inclusion bodies using Ni-chelated affinity column. Polyclonal antibodies have been raised against purified WsGT protein in Newzeland white rabbit which has been used for western analysis. This section also covers purification of WsGT enzyme using FPLC system in active form and characterization of the protein. WsGT enzyme assay was performed with various substrates and most of the flavonoid substrates are glycosylated at 7^{th} position. Enzyme kinetics was done with these substrates and it is concluded that diadzein was the most specific substrate for WsGT. WsSAPGT enzyme was purified and enzyme assay was done with steroidal sapogenin substrates and it shows specificity with steroidal sapogenin digitoxigenin.

(4.B): Homology modeling and docking studies of the *W. somnifera* GT gene

This chapter deals with the molecular modeling studies of the WsGT gene using MODELLER 9v9 software and for the docking studies Autodock vina software was used. The WsGT structural validation was done with PROCHECK and ProSAII. Docking study was done with WsGT template and different ligand molecules along with UDP-glucose as a sugar donor. It also shows active residues for UDP-glucose and substrates.

Chapter 5

Tissue specific expression of *W. somnifera* GT gene – Real time PCR analysis

This chapter will describe the expression pattern of GT gene in different plant part. The tissue specific expression of the GT gene was studied using quantitative Real Time PCR. Primers were designed from the non conserved region of the gene so as to distinguish it from the other GT genes. The 5.8S RNA was used as the internal standard. Different plant parts *i.e*, leaf, stem and root were used for expression studies. This section also deals with the expression studies of WsGT gene under different stress conditions like salicylic acid, methyl jasmonate and heat. Tissues were incubated at different concentrations of stressess and time intervals to understand the effect of various stresses on WsGT expression.

Chapter: 1

Introduction

1. Introduction

1.1 *Withania somnifera*

Withania somnifera (Ashwagandha) belongs to the family Solanaceae. The roots are a constituent of over 200 formulations in Ayruvedha, Siddha and Unani medicine, which are used in the treatment of various physiological disorders (Asthana, 1989; Singh, 1998). *Withania* appears in WHO monographs on Selected Medicinal Plants and an American Herbal Pharmacopoeia monograph (Marderosion, 2001).

Withania is widely claimed to have potent aphrodisiac, sedative, rejuvenative and life prolonging properties. It is used as a general energy-enhancing tonic known as 'Medharasayana' which means 'to promote learning and a good memory' and in geriatric problems (Williamson, 2002). The plant was traditionally used to promote youthful vigor, endurance, strength, health and increasing the production of vital fluids, muscle fat, blood, lymph, semen and cells. The similarity between these restorative properties and those of ginseng roots has led to Ashwagandha roots being called Indian ginseng (Singh, 1998). It also helps counteract chronic fatigue, weakness, dehydration, bone weakness, loose teeth, thirst, impotency, premature ageing, emaciation, debility, and muscle tension. The leaves of the plant are bitter in taste and used as an antihelmanthic. Bruised leaves and fruits are locally applied to tumors and tubercular glands, carbuncles and ulcers (Kapoor, 2001). The roots are used as a nutrient and health restorative in pregnant women and old people. The decoction of the roots boiled with milk and ghee is recommended for curing sterility in women. The roots are also used in constipation, senile debility, rheumatism, general debility, nervous exhaustion, loss of memory, loss of muscular energy and spermatorrhoea (Watt, 1972; Mohammad Hossein Mirjalili, *et al.*, 2009).

1.1.2 Classification

Kingdom: Plantae
Division: Angiosperma
Class: Dicotyledoneae
Order: Tubiflorae

Family: Solanaceae

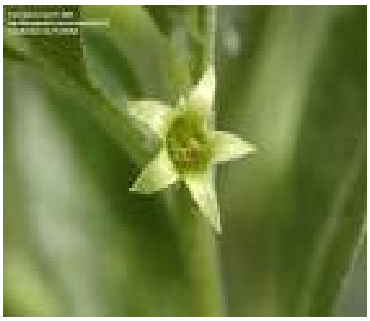
Genus: *Withania*

Species: *somnifera*

1.1.3 Botanical description

An erect branching under shrub reaching about 150 cm in height, usually clothed with minutely stellate tomentum; leaves ovate upto 10 cm long, densely hairy beneath and sparsely above, flowers greenish or lurid yellow in axillary fascicles, bisexual, pedicel long, fruits globose berries which are orange coloured when mature, enclosed in a persistent calyx (**Figure 1**). The fleshy roots when dry are cylindrical, gradually tapering down with a brownish white surface and pure white inside when broken.

Withania seed



Withania flower

Withania plant

Figure 1 ***Withania somnifera***

1.1.4 Geographical distribution

This plant grows wildly in all drier parts of subtropical India. It occurs in Madhya pradesh, Uttar pradesh, Punjab plains and northwestern parts of India like Gujarat and Rajasthan. It is cultivated in 5000 ha in Rajasthan, Madhya Pradesh, Andhra Pradesh and Uttar Pradesh (Arun Kumar, 2007). It is also found in Congo, South Africa, Egypt, Morocco, Jordan, Pakistan and Afganistan.

W. somnifera grows well in sandy loam or light red soil, having pH 7.5-8.0 with good drainage. It can be cultivated between 600-1200 m altitudes. The semi-tropical areas receiving 500-750 mm rainfall are suitable for cultivation of this rained crop. The crop requires dry season during its growing period. Temperature between 20 ºC to 35 ºC is most suitable for cultivation. Late winter rains are conducive for the proper development of the plant roots.

1.1.5 Chemical constituents

The chemistry of *Withania* species has been extensively studied and several groups of chemical constituents such as steroidal lactones, alkaloids, flavonoids, tannin and many more have been identified, extracted and isolated (Atta-ur-Rahman, 1991; 1993). At present, more than 12 alkaloids, 40 withanolides, and several sitoindosides (a withanolide containing a glucose molecule at carbon 27) have been isolated and reported from aerial parts, roots and berries of *Withania* species (Choudary, 1996). The major chemical constituents of these plants, withanolides, are mainly localized in leaves, and their concentration usually ranges from 0.001 to 0.5% dry weight (DW) (Bandyopadhyay, 2007). The withanolides are a group of naturally occurring C28-steroidal lactones built on an intact or rearranged ergostane framework, in which C-22 and C-26 are appropriately oxidized to form a six-membered lactone ring (Glotter, 1991). The basic structure is designated as the withanolide skeleton (**Figure 2**) (Alfonso, 1994).

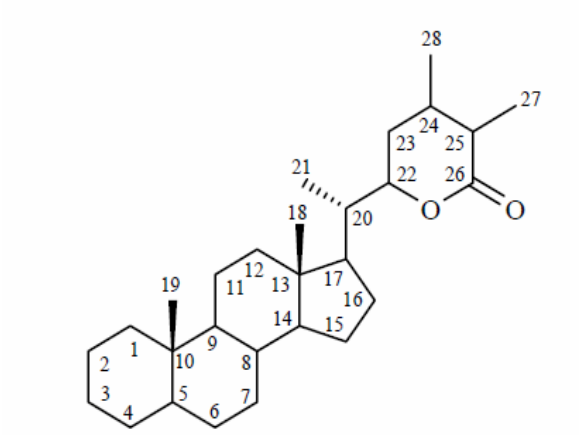


Figure 2 The basic structure of Withanolides

The biosynthetic pathways of withanolides and other chemical constituents of *W. somnifera* are not fully known and there is very little information about their biogenetic aspects (Kirson, 1977). It has been reported that with a very few exceptions, the plants that synthesize the 20-H withanolides are unable to produce the 20-OH counterparts and vice versa. Since withanolides are probably derived from cholesterol. Therefore, we can assume that this is a pertinent starting point for the biosynthesis of withanolides (Nittala, 1981).

Withaferin A (4β, 27-dihydroxy-1-oxo-5β,6β-epoxywitha-2-24-dienolide, **Figure 3**) was the first member of this group of compounds to be isolated from the well-known South-Asian medicinal plant, *W. somnifera* (Lavie, 1965). The structural novelty and interesting biological activities elicited by this compound led to a thorough chemical investigation of the plant and numerous compounds with similar structural features were isolated (Leet, 1982).

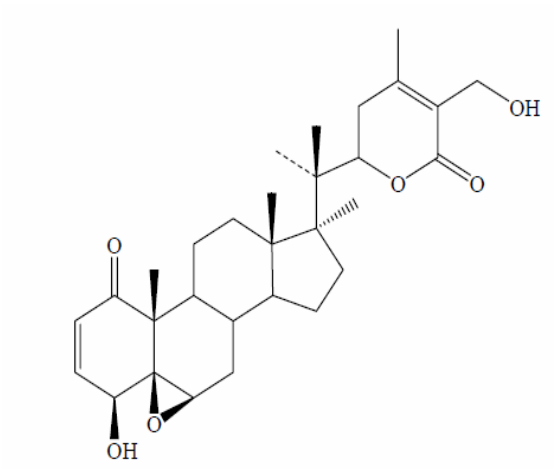


Figure 3 Structure of withaferin A

Withaferin A is mainly valued for its anti-cancerous properties. The yields of withaferin A from intact plants of *Withania* spp. (Israel chemotype) are 0.2-0.3% of DW of leaves (Abraham, 1968). Quantitative analysis of Indian chemotypes of *W. somnifera* performed and observed that withaferin A is totally absent in roots, stems, seeds and persistent calyx of fruits of intact plants but present in leaves (1.6%) (Gupta, 1996).

1.1.6 Other Chemical Compounds

Examination of *W. somnifera* roots has resulted in the isolation of a new dimeric thiowithanolide, named ashwagandhanolide (**Figure 4**) (Subaraju, 2006). A bio assay-guided purification of the methanolic extract of *W. somnifera* fruits yielded withanamides A-I. The structure of these compounds was determined by using serotonin, glucose and long-chain hydroxyl fatty acid moieties (Jayaprakasam, 2004). The general belief is that tropane alkaloids are restricted to the roots of *Withania* spp. Extraction with 45% alcohol yields the highest percentage of alkaloids. The isolation of nicotine, somniferine, somniferinine, withanine, withananine, pseudowithanine, tropine, pseudotropine, 3α-tigloyloxytropane, choline, cuscohygrine, *dl*-isopelletierine and new alkaloids anaferine and anhygrine has been described (Gupta, 2007). The reported total alkaloid content in the roots of Indian *W. somnifera* varies between 0.13 and 0.31%. However, much higher yields up to 4.3% have been recorded in plants of other regions/countries. In addition to the alkaloids, the roots are reported to contain starch, reducing sugars, hentriacontane, glycosides, dulcitol, withanicil, and four types of peroxidases. The leaves are reported to contain five unidentified alkaloids chlorogenic acid, calystegines (nitrogen-containing polyhydroxylated heterocyclic compounds) withanone, condensed tannin and flavonoids (Johri, 2005).

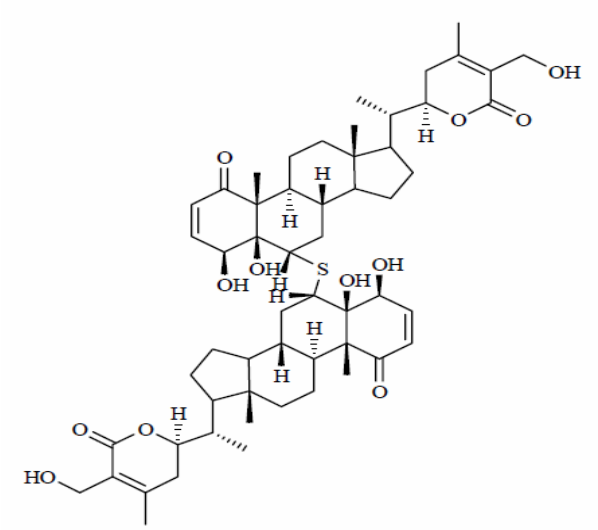


Figure 4 Ashwagandhanolide, a new compound isolated from *W. somnifera*

1.1.7 Therapeutic Use of *W. somnifera*

W. somnifera is one of the important medicinal cash crops in many states of india. It has antibiotic, antiviral, antiamoebic, antiarthritic and anti-inflammatory properties. Withaferin A, found in this plant, shows marked tumour-inhibitory activity. Its fruits and seeds are diuretic, hypnotic, maticatory and employed in curdling plant milk to prepare vegetarian cheese. They are also rich in saponins and can be used as substitutes of soap (Saritha, 2007). Saponins are important components of a number of herbal medicines and known for a wide range of bioactivities like allelopathic, antimicrobial, anticholesterolemic, anti-cancer, adjuvant, immunomodulatory, antioxidant and haemolytic activities (Bhaskara Reddy Madina , 2007).

The natural wild populations of *W. somnifera* contain a tongue paralysing constituent in their roots and hence cannot be used for ayurvedic treatments. The improved, cultivated varieties however do not have such an ingredient. The wild and cultivated plants contain the same alkaloids but exhibit different morphologies and therapeutic actions (Kaul, 1957). However, since the cultivated plants show many different chemotypes, production of a large number of plants of superior selected chemotypes (chemotypes are plants with similar morphological characters but with different chemical compositions), without seasonal constraints, is desirable (Kulkarni, 2000).

1.1.8 Anti-inflammatory Properties

The effectiveness of ashwagandha in a variety of rheumatologic conditions may be due in part to its anti-inflammatory properties, which have been studied by several authors. In a study, powdered root of *W. somnifera* (1 g/kg suspended in 2% gumacacia, 50 mg/mL) was given orally one hour before the induction of inflammation by injection of Freund's complete adjuvant in rats and continued daily for three days; phenylbutazone was given as a positive control. *W. somnifera* was found to cause considerable reduction in inflammation.

Few studies have been conducted on the mechanism of action for the anti-inflammatory properties of *W. somnifera*. Rats injected with 3.5-percent formaline in the hind leg footpad showed a decrease in absorption of 14_C-glucose in rat jejunum (Somasundaram, 1983). Glucose absorption was maintained at the normal level by both *W. somnifera* and the cyclooxygenase inhibitor oxyphenbutazone. Both drugs produced anti-inflammatory effects. Similar results were obtained in parallel experiments using 14_C-leucine absorption from the jejunum (Somasundaram, 1983). These studies suggest cyclooxygenase inhibition may be involved in the mechanism of action of *W. somnifera*.

Another study by Begum *et al.*, 1988 examined the effect of *W. somnifera* root powder on paw swelling and bony degenerative changes in Freund's adjuvant-induced arthritis in rats. *Withania* caused significant reduction in both paw swelling and degenerative changes as observed by radiological examination. The reductions were better than those produced by the reference drug, hydrocortisone.

1.1.9 Antitumor Properties

To investigate its use in treating various forms of cancer, the antitumor and radiosensitizing effects of *Withania* have been studied. *W. somnifera* was evaluated for its anti-tumor effect in urethane-induced lung adenomas in adult male albino mice. Simultaneous administration of *Withania* and urethane reduced tumor incidence significantly (tumor incidence: untreated control, 0/25; urethane treated, 19/19; *Withania* treated, 0/26, and *Withania* plus urethane treated, 6/24, $p<0.05$). The histological appearance of the lungs of animals protected by

Withania was similar to those observed in the lungs of control animals. No pathological evidence of any neoplastic change was observed in the brain, stomach, kidneys, heart, spleen, or testes of any treated or control animals. In addition to providing protection from carcinogenic effects, *Withania* treatment also reversed the adverse effects of urethane on total leukocyte count, lymphocyte count, body weight, and mortality (Devi, 1992).

The growth inhibitory effect of *Withania* was also observed in Sarcoma 180 (S-180), a transplantable mouse tumor (Devi, 1996). Ethanol extract of *Withania* root after intra-dermal inoculation of $5x10^5$ cells of S-180 in BALB/c mice produced complete regression of tumor after the initial growth. *Withania* was also found to act as a radio and heat sensitizer in mouse S-180 and in Ehrlich ascites carcinoma. Antitumor and radiosensitizing effects of withaferin (a steroidal lactone of *Withania*) were also seen in mouse Ehrlich ascites carcinoma *in vivo* (Sharad, 1996). Withaferin A from *Withania* gave a radiosensitizer ratio of 1:5 for *in vitro* cell killing of V79 Chinese hamster cell at a non-toxic concentration of about 2 mM/L. These studies are suggestive of antitumor activity as well as enhancement of the effects of radiation by *W. somnifera*.

1.1.10 Antistress Effect

To evaluate the antistress effect of *W. somnifera*, an alcohol extract from defatted seeds of *Withania* dissolved in normal saline was given to 20-25g mice in a swimming performance test in water at 28 ºC-30 ºC (Singh, 1982). Controls were given saline. The extracts approximately doubled the swimming time when compared to controls. In another study, *Withania* prevented both a weight increase of the adrenals and a reduction in ascorbic acid content of the adrenals normally caused by this swimming test. The authors suggested that *W. somnifera* induced a state of nonspecific increased resistance during stress.

1.1.11 Antioxidant Effect

The brain and nervous system are relatively more susceptible to free radical damage than other tissues because they are rich in lipids and iron, both known to be important in generating reactive oxygen species. The brain also uses nearly 20 percent of the total oxygen supply.

Free radical damage of nervous tissue may contribute to neuronal loss in cerebral ischemia and may be involved in normal aging and neurodegenerative diseases, e.g.,epilepsy, schizophrenia, Parkinson's, Alzheimer's and other diseases (Lakshmi-Chandra Mishra, 2000).

Since traditional Ayurvedic use of *W. somnifera* has included many diseases associated with free radical oxidative damage, it has been considered likely the effects may be due to a certain degree of antioxidant activity. The active principles of *Withania*, sitoindosides VII-X and withaferin A (glycowithanolides), have been tested for antioxidant activity using the major free-radical scavenging enzymes, superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPX) levels in the rat brain frontal cortex and striatum. Decreased activity of these enzymes leads to accumulation of toxic oxidative free radicals and resulting degenerative effects. An increase in these enzymes would represent increased antioxidant activity and a protective effect on neuronal tissue. Active glycowithanolides of *Withania* (10 or 20 mg/kg intraperitoneally) were given once daily for 21 days to groups of six rats. Dose-related increases in all enzymes were observed; the increases comparable to those seen with deprenyl (a known antioxidant) administration (2 g/kg/day intraperitoneally). This implies that *Withania* does have an antioxidant effect in the brain which may be responsible for its diverse pharmacological properties (Bhattacharya, 1997).

1.1.12 Immunomodulatory Properties

The use of *W. somnifera* as a general tonic to increase energy and prevent disease may be partially related to its effect on the immune system. Glycowithanolides and a mixture of sitoindosides IX and X isolated from *Withania* were evaluated for their immunomodulatory and central nervous system effects (antistress, memory, and learning) in Swiss mice (15-25 g, 5-6 months old) and Wistar strain albino rats (120-150 g and 250-300 g) (Ghosal, 1989). Both materials produced statistically significant mobilization and activation of peritoneal macrophages, phagocytosis, and increased activity of the lysosomal enzymes. Both compounds (50-200 mg/kg orally) also produced significant antistress activity in albino mice and rats, and augmented learning acquisition and memory retention in both young and old rats.

Root extract of *Withania* was tested for immunomodulatory effects in three myelosuppression models in mice: cyclophosphamide, azathioprin, or prednisolone. Significant increases ($p<0.05$) in hemoglobin concentration, red blood cell count, white blood cell count, platelet count, and body weight were observed in *Withania*-treated mice compared to untreated control mice. Also few reports indicate significant increases in hemolytic antibody responses toward human erythrocytes which indicated immunostimulatory activity (Ziauddin, 1996).

1.1.13 Hemopoetic Effect

Administration of *W. somnifera* extract was found to significantly reduce leukopenia induced by cyclophosphamide (CTX) treatment in Swiss albino mice. The major activity of *Withania* may be the stimulation of stem cell proliferation. These studies indicated that *Withania* reduced CTX-induced toxicity and may prove useful in cancer chemotherapy (Davis, 1998).

1.1.14 Rejuvenating Effect

In another clinical trial, *W. somnifera* purified powder was given 3 g/day for one year to 101 normal healthy male volunteers, age 50-59 years. All subjects showed significantly increased hemoglobin and RBC count, and improvement in hair melanin and seated stature. They also showed decreased SED rate, and 71.4 percent of the subjects reported improvement in sexual performance. In summary, these studies indicate that *Withania* may prove useful in younger as well as older populations as a general health tonic (Kuppurajan, 1996).

1.1.15 Nervous System Effects

Total alkaloid extract (ashwagandholine, AG) of *Withania* roots has been studied for its effects on the central nervous system (Malhotra, 1965). Effects of sitoindosides VII-X and withaferin isolated from aqueous methanol extract of roots of cultivated varieties of *Withania* were studied on brain cholinergic, glutamatergic and GABAergic receptors in male Wistar rats (Schliebs, 1997).

Ashwagandholine, total alkaloids extracted from extract of *Withania* roots, caused relaxant and antispasmodic effects against various agents that produce smooth muscle contractions in

intestinal, uterine, tracheal, and vascular muscles. The pattern of smooth muscle activity was similar to that of papaverine, but several-fold weaker, which indicated a direct musculotropic action. These results are consistent with the use of *W. somnifera* to produce relaxation (Malhotra, 1965).

1.1.16 Effects on the Endocrine System

Based on the observations that *Withania* provides protection from free radical damage in the mouse liver, studies were conducted to determine the efficacy of *Withania* in regulating thyroid function. Mice were given *Withania* root extract, which significantly reduced hepatic lipid peroxidation and increased the activity of superoxide dismutase and catalase. The results suggest *W. somnifera* stimulates thyroidal activity and also promotes hepatic antioxidant activity (Lakshmi-Chandra Mishra, 2000).

1.1.17 Effects on the Cardiopulmonary System

W. somnifera useful as a general tonic shows beneficial effects on the cardiopulmonary system. The effect of AG was studied on the cardiovascular and respiratory systems in dogs and frogs. The alkaloids had a prolonged hypotensive, bradycardiac, and respiratory-stimulant action in dogs. The alkaloids produced immediate predominant but short lived cardio-depressant effects and a weak but prolonged cardiotonic effect in isolated normal and hypodynamic frog hearts. The pharmacological actions of the total extract of *Withania* roots on the cardiovascular and respiratory systems appeared to be due to its alkaloid content. The total alkaloids were more than twice as active as the 70-percent alcohol extract of *Withania* root. These studies were found to be consistent with the use of *W. somnifera* as a tranquilizing agent (Panda, 1998).

1.2 Review of literature

W. somnifera (Solanaceae) is used extensively in Indian traditional system of medicine and is equated with ginseng in its therapeutic benefits. It is especially attractive for studying the enzymes involved in steroidal transformation. It is a rich source of a variety of steroidal

compounds. It contains a variety of glucosylated steroids called withanosides in roots and leaves.

Glucosylation of the sterols are catalyzed by sterol glucosyltransferases (SGTs), which are members of family 1 glycosyltransferases. One of the sterol glucosyltransferase gene (3β-hydroxy) is reported from *W. somnifera* leaves. Also cytosolic sterol glucosyltransferase was purified from *Withania* leaves and studied for its biochemical and kinetic properties.

Kulkarni *et al*., (1996) reported direct regeneration of *W. somnifera* with multiple shoots from leaf explants. The production of secondary metabolites from callus and shoot cultures has also been reported earlier (Roja *et al*., 1991). Regeneration of shoots from nodes, internodes, hypocotyls and embryos also reported (Kulkarni *et al*., 2000). *In vitro* flowering, *in vitro* fruiting and effective micropropagation protocol were studied in *Withania* (Saritha, 2007).

All parts of the plant contain chemical compounds called withanolides/steroidal lactones to which most of the pharmacological activity is attributed. Although occurrence of five distinct chemotypes of this species was reported (Abraham *et al*., 1968; Kirson *et al*., 1971). Five morphotypes exhibiting morphological variability have also been reported (Atal *et al*., 1975).

The chemical constituents of this plant have been the targets of many investigations and the structures of many withanolides have been characterized. Recently seven new withanolide glycosides termed withanosides I, II, III, IV, V, VI and VII have been reported from roots. This paper describes the structure elucidation of these withanosides based on chemical and physicochemical evidence (Hisashi Matsuda *et al*., 2001). Also five selected withanolides (withanone, withaferin A, withanolide B, withnolide E, withanolide A) were identified from in vitro cultures of *Withania* (Sharada *et al*., 2007).

Expressed sequence tag (EST) quickly gives the gene index of an organism and is used to identify the genes involved in specific plant metabolic pathways. The analysis of ESST offers a rapid and cost effective approach to elucidate the transcriptome of an organism. It is inexpensive, easy and less time consuming compared to tranditional gene cloning and whole genome sequencing. The first transcripomes, expressed sequence tags (ESTs) in a leaf and root from *W. somnifera* were reported (kalaiselvi senthil *et al*., 2010).

Glycowithanolides, consisting of equimolar concentrations of sitoindosides VII-X and withanferin A, isolated from the roots of *W. somnifera* have been reported to have an antioxidant effect in the rat brain frontal cortex and striatum (Bhattacharya *et al*., 2000).

However, biosynthetic pathway of withanolides, their functions in *W. somnifera* and metabolic steps leading to their glycol-transformation are unknown. One of the reasons for limited knowledge of their functions is non-availability of the relevant enzymes and genes.

1.3 Medicinal Plants

Humans are dependent on nature for their basic needs, for the production of food, shelter, clothing, transportation, fertilizers, flavours and fragrances, and medicines (Cragg and Newman, 2005). Plants have formed the basis of traditional medicine systems that have been in existence for thousands of years and continue to provide mankind with new remedies. Medicinal plant therapy is based on the empirical findings of hundreds and probably thousands of years of use. Until the advent of modern medicine, man depended on plants for treating human and livestock diseases. The interest in nature as a source of potential chemotherapeutic agents continues. Natural products and their derivatives represent more than 50% of all the drugs in clinical use in the world today and higher plants contribute 25% of the total (Farnsworh *et al*., 1985; Cragg and Newman, 2005).

Medicinal and aromatic plants (MAPs) play an important role in the healthcare of people around the world, especially in developing countries. Throughout the world human societies have a vast knowledge on medicinal uses of plants, and for related uses including as poison for fish and hunting, purifying water, and for controlling pests and diseases of crops and livestock. About 80% of the population of most developing countries still use traditional medicines derived from plants for treating human diseases (de Silva, 1997). China, Cuba, India, Sri Lanka, Thailand, and a few other countries have endorsed the official use of traditional systems of medicine in their healthcare programs. For example, the Indian systems of medicine ‘Ayurveda,’ ‘Sidha’ and ‘Unani’ entirely, and homeopathy to some extent, depend on plant materials or their derivatives for treating human ailments (Prajapati *et al*., 2003).

People in villages and remote areas primarily depend on traditional medicines, because the modern system is out of reach and expensive. Apart from health care, medicinal plants are mainly the alternate income-generating source of underprivileged communities (Chandra Prakash Kala *et al.*, 2006). Most of the Asian and Africans believe that traditional medicines are more effective than modern medicines and traditional medicines do not have side effects. Thus, in many societies, traditional and modern systems of medicines are used independently. Some important medicinal plants given in the table 1.

Table 1 List of few important medicinal plants and their uses

Botanical Name	Common name	Parts Used	Medicinal Use
Withania somnifera	Aswagandha	Root, Leafs	Restorative tonic, stress, nerves disorder, aphrodiasiac.
Bacopa monnieri	Brahmi	Whole plant	Nervous, memory enhancer, mental disorder.
Azardirchata indica	Neem	Leaf,fruit	Sedative, analgesic, epilepsy, hypertensive.
Ocimum sanclum	Tulsi	Leaves/Seed	Cough, Cold, bronchitis, expectorand
Vincea rosea	Sada Bahar	Whole Plant	Leaukamia, Hypotensiv, Antispasmodic, Atidot.
Cinnamomum zeylanicum	Dalchin	Bark, Oil	Bronchitis, Asthma, Cardiac, Disorder, Fever
Ranwolfia serpentina	Sarpa gandha	Root	Hyper tension, insomnia.
Aloe verra	Gritkumari	Leaves	Laxative, Wound healing, Skin burns & care, Ulcer.
Mesua ferrea	Nageswar	Bark, Leaf and Flower	Asthma, Skin, Burning, Vomiting, Dysentery, Piles.

1.3.1 Use and diversity of medicinal plants

In India, the total species of higher plants are 17,000 and 7500 are known for medicinal uses. This proportion of medicinal plants is the highest proportion of plants known for their medical purposes in any country of the world for the existing flora of that respective country (Shiva, 1996). Ayurveda, the oldest medical system in Indian sub-continent, has alone reported

approximately 2000 medicinal plant species, followed by Siddha and Unani. The Charak Samhita, an age-old written document on herbal therapy, reports on the production of 340 herbal drugs and their indigenous uses. The northern part of India has a great diversity of medicinal plants because of the majestic himalayan range. So far about 8000 species of angiosperms, 44 species of gymnosperms and 600 species of pteridophytes have been reported in the Indian Himalaya, of these 1748 species are known as medicinal plants (Prajapati, 2003).

4,22,000 plant species documented worldwide and out of this 12.5% are reported to have medicinal value. The proportion of medicinal plants varies in different countries from 4.4% to 20% (Schippmann *et al*., 2002). About 25% of drugs in modern pharmacopoeia are derived from plants (phytomedicines) and many others are synthetic analogues built on prototype compounds isolated from plants. Upto 60% of the drugs prescribed in Eastern Europe consists of unmodified or slightly altered higher plant products (Lancet, 1994). These drugs carry important therapeutic properties including contraceptives, steroids and muscle relaxants for anesthesia and abdominal surgery, quinine and artemisinin against malaria; digitalis derivatives for heart failure; and the anti-cancer drugs vinblastin, etoposide and taxol. These compounds cannot be synthesized cost effectively, which means that their production requires reliable supplies of plant material (van seters, 1997).

The global importance of medicinal and aromatic plant materials is evident from a huge volume of trade at national and international levels. The reported annual international import of medicinal and aromatic plants for pharmaceutical use amounted on an average 3,50,000 tonnes valued at over 1 billion USD (Table 2) (Rao *et al*., 2004).

Table 2 Leading countries of import and export of medicinal and aromatic plant material based on annual average.

Country of Import	Volume (Tonnes)	Value 1000USD	Country of export	Volume (Tonnes)	Value 1000USD
Hong kong	73 650	314 000	China	139 750	298 650

Japan	56 750	146 650	USA	11 950	114 450
USA	56 000	133 350	Germany	15 050	72 400
Germany	45 850	113 900	Singapore	11 250	59 850
Singapore	6 550	55 500	India	36 750	57 400
Rep.Korea	31 400	52 550	Chile	11 850	29 100
France	20 800	50 400	Egypt	11 350	13 700
China	12 450	41 750	Mexico	10 600	10 050
Italy	11 450	42 250	Bulgaria	10 150	14 850
Pakistan	11 350	11 850	Albania	7 350	14 050
Spain	8 600	27 450	Morocco	7 250	13 200
UK	7 600	25 500	Pakistan	8 100	5 300
Total	342 550	1 015 200	Total	281 550	643 200

A few countries dominate the international trade with over 80% of the global import and export allotted to 12 countries each. Japan and Korea are the main consumers of medicinal plants, whereas China and India are the world's leading producing nations. Hong Kong, United states and Germany stand out as important trade centers. It is estimated that the total number of medicinal and aromatic plants in international trade is around 2500 species worldwide (Schippmann *et al*., 2002) (Table 3).

Table 3 how many plants are used medicinally world-wide?

Country	Plant species	Medicinal plant species	% Usage
China	26 092	4 941	18.9
India	15 000	3 000	20.0
Indonesia	22 500	1 000	4.4
Malaysia	15 500	1 200	7.7
Nepal	6 973	700	10.0
Pakistan	4 950	300	6.1
Philippines	8 931	850	9.5
Sri Lanka	3 314	550	16.6
Thailand	11 625	1 800	15.5
USA	21 641	2 564	11.8
Vietnam	10 500	1 800	17.1
Average	13 366	1 700	12.5
World	**422 000**	**52 885**	

1.3.2 Drug discovery from medicinal plants

Drug discovery from medicinal plants has evolved to include numerous fields of inquiry and various methods of analysis. The process typically begins with a botanist, ethnobotanist, ethnopharmacologist, or plant ecologist who collects and identifies the plant(s) of interest. Collection may involve species with known biological activity for which active compound(s) have not been isolated (i.e traditionally used herbal remedies) or may involve taxa collected randomly for a large screening program. It is necessary to respect the intellectual property rights of a given country where plant(s) of interest are collected (Baker *et al*., 1995).

Phytochemists (natural product chemists) prepare extracts from the plant material, subject these extracts to biological screening in pharmacologically relevant assays, and commence the process of isolation and characterization of the active compound(s) through bioassay-guided fractionation. Molecular biology has become essential to medicinal plant drug discovery through the determination and implementation of appropriate screening assays directed towards physiologically relevant molecular targets. Pharmacognosy encapsulates all the above fields into a distinct interdisciplinary science.

1.4 Secondary metabolites

Plants play an important role of our everyday diet; their constituents and nutritional value have been intensively studied for decades. Higher plants synthesize essential primary metabolites (e.g. carbohydrates, lipids and amino acids), and also able to synthesize a wide variety of low molecular weight compounds – the secondary metabolites (Kirsi-Marja Oksman-Caldentey, 2004). Plant secondary metabolites have no recognized role in the maintenance of fundamental life processes in the plants, but they do have an important role in the interaction of the plant with its environment. The production of these compounds is very low (less than 1% dry weight) and it depends greatly on the physiological and developmental stage of the plant. The production of secondary metabolites in plants is enhanced by both biotic and abiotic stress conditions. Secondary metabolites have complex and unique structures, stored in specific cells and/or organs of the plant and often accumulate in vacuoles (Dixon, 2001).

Polysaccharides, sugars, proteins and fats are compounds derived from primary pathways of plants, which are the building blocks for plant growth. Alkaloids, terpenoids, phenolics, steroids and flavonoids are low concentration secondary pathway compounds and have a wide diversity in structure and size. These are found in very large numbers throughout the plant kingdom. There were approximately 100,000 different plant-derived compounds and large number of new ones added to the list every year (Verpoorte *et al.,* 1998). Numerous plant secondary metabolites such as alkaloids, anthocyanins, flavonoids, quinones, lignans, steroids, and terpenoids are used in commercial applications such as drugs, dyes, flavours, fragrance, insecticides, etc. Such fine chemicals are extracted and purified from plant materials (Verpoorte, 2000).

One of the functions of secondary compounds is that they form biochemical defense mechanism against pathogens and predators (Bannet *et al.,* 1994). The secondary compounds may be present at constitutively high levels, such as in the highly differentiated bark and heartwood of perennials, or as induced compounds resulting from a rapid synthesis of large amounts of low molecular weight phytoalexins followed by pathogen invasion or insect attack. The secondary compounds may be a convenient sink, into which excess carbon and nitrogen can be diverted away from an inactive part of primary metabolism. The secondary compounds are degraded and the stored carbon and nitrogen are recycled back into the primary metabolism whenever there is a demand. The balance between the activities of the primary and secondary metabolism is a dynamic one, which will be largely affected by growth, tissue differentiation and development of the plant body. The factors which determine the location and accumulation of secondary products in the intact plant are important as they also control the production of secondary products in plant cell cultures. For example, elicitation of a secondary pathway by a pathogen will lead to a localized production of a phytoalexin in plants and elicitors have been used to stimulate secondary product formation in tissue cultures (Collin, 2001).

Plants exploit secondary metabolites in defense responses against pathogens and in symbiotic relationships such as nitrogen fixation and pollinator attraction. Thus, the vast structural diversity in plant secondary metabolites is considered to be the consequence of chemical

adaptation by plants to specific ecological niches (Akio Noguchi *et al*., 2009). A particular plant lineage develops specialized metabolites by acquisition of novel functions of biosynthetic enzymes to increase fitness to its environment. Therefore, specialized metabolites are ascribed to structural changes associated with functional differentiation during the course of enzyme/gene evolution (**Figure 5**) (Pichersky and Gang, 2000).

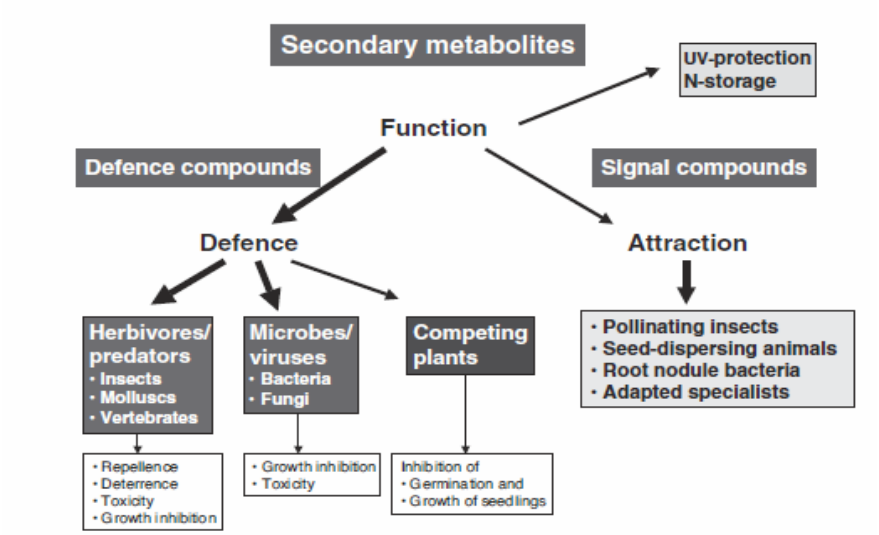


Figure 5 Ecological and Physiological functions of Plant Secondary metabolites

Originally, secondary products were seen as end points in metabolism, with a less specific role. Many of the compounds were shown to have an active turnover and now it is accepted that they have a much more defined function in plants. The relationship between primary compounds, intermediary metabolism and the groups of secondary compounds is described below (**Figure 6**) (Dörnenburg, 1996).

1.4.1 Role of plant secondary metabolites

In plants, as a result of metabolic processes, many different kinds of organic compounds or metabolites are produced. These metabolites are grouped into primary and secondary metabolites. The primary metabolites like chlorophyll, amino acids, nucleotides, simple carbohydrates or membrane lipids, play recognized roles in photosynthesis, respiration, solute transport, translocation, nutrient assimilation and differentiation. However, the secondary metabolites differ from primary metabolites in having a restricted distribution in the plant

kingdom. Few of secondary metabolites are often found in only one plant species or a taxonomically related group of species, whereas the basic primary metabolites are found throughout the plant kingdom (Taiz and Zeiger, 2006). Experimental evidence has made it clear that many secondary metabolites do have functions that are vital for the fitness of a plant.

The main roles are:

- Defense against herbivores (insects, vertebrates)
- Defense against fungi and bacteria
- Defense against viruses
- Defense against other plants competing for light, water and nutrients
- Signal compounds to attract pollinating and seed dispersing animals
- Signals for communication between plants and symbiotic micro-organisms (e.g. N-fixing Rhizobia or mycorrhizal fungi)
- Protection against UV-light or other physical stress (Wink, 1999)

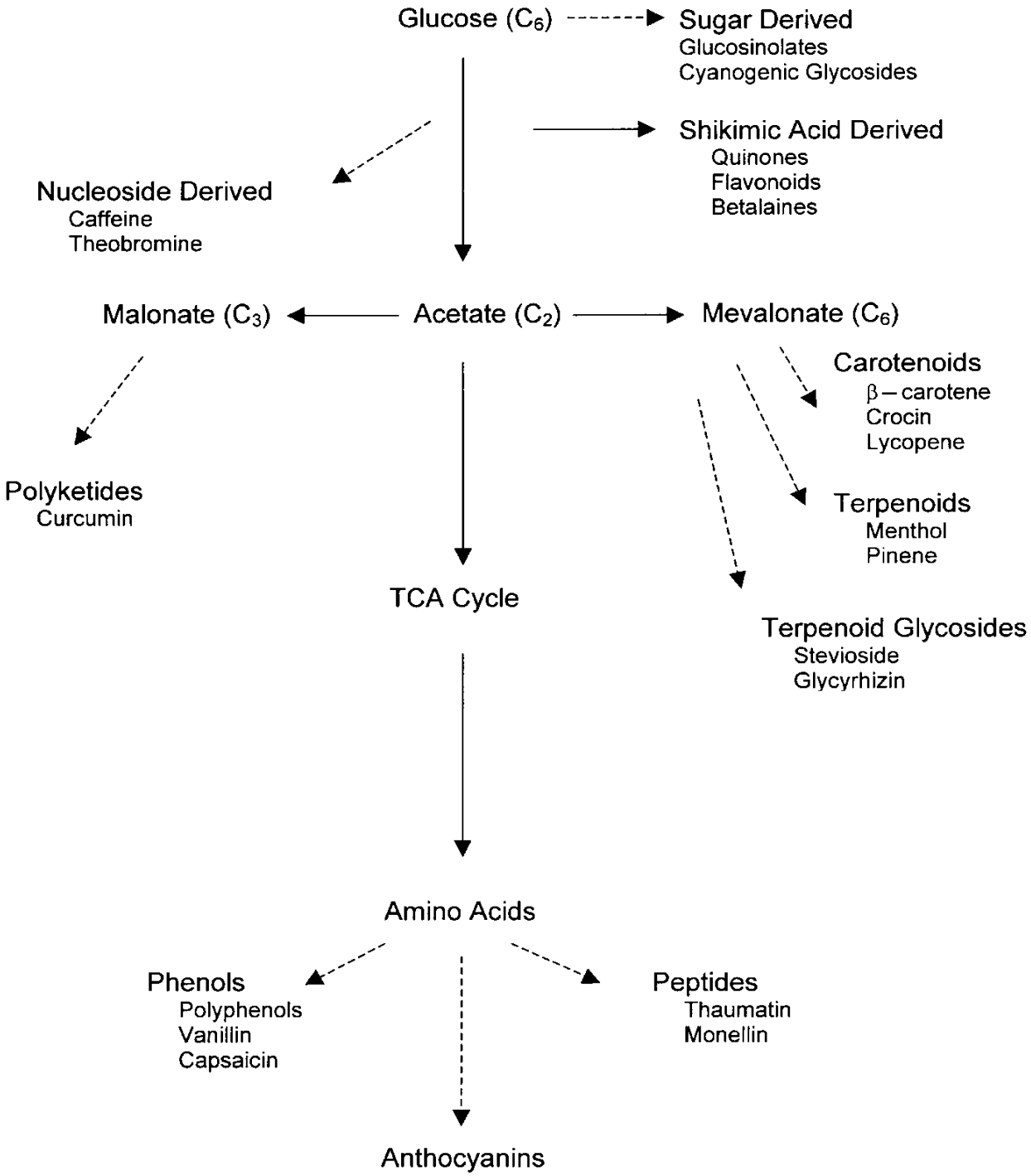


Figure 6 Relationship between primary and secondary pathways in plants

1.4.2 Molecular engineering for secondary products

Initially transgenic plants were used to increase the level of most secondary products. A more analytical approach is now used for the synthesis of secondary products. It includes identification of the enzymes in a secondary pathway and an attempt to modify these enzymes by enhancing gene expression. The problem with this approach is that the concentration of enzymes in secondary product pathways is too low for them to be isolated and purified in sufficient quantities for analysis, which is an essential step prior to identification of the corresponding gene(s). Recognition of the regulatory genes would be the ideal approach, since these genes are going to control a large part, or the whole, of the pathway. However, many secondary pathways are not linear; rather, there are so complex network of interactions that increasing total flux through a pathway requires the engineering of a number of genes. In addition to this, each gene codes a rate limiting step. These regulatory genes would also have to be modified for continuous expression, so that their products are not sensitive to the usual controls such as feedback inhibition (Collin, 2001).

Since the low level of production of most secondary product enzymes is a major problem in identifying enzymes and subsequently genes, the methods that are now being used to clone secondary product genes are transposon tagging and differential screening. Using a PCR approach, with differential screening, it is possible to isolate a number of genes relating to a pathway. The individual genes must then be heterologously expressed, the protein identified and its function judged by assay or by comparison with other proteins.

1.5 Glycosylation: a key modification of secondary metabolites

Plants are well known for producing a huge diversity of low molecular weight natural products through secondary metabolism. More than 1,00,000 such compounds have been described and many of them having commercial interest. Glycosylation (i.e. conjugation to a sugar moiety) is a mechanism which adds various modifications to the plant secondary metabolites and is catalyzed by a family of enzymes called glycosyltransferases (UGTs). It is involved in various functions, including the regulation of hormones homeostasis, the detoxification of xenobiotics and the biosynthesis and storage of secondary compounds

(Clarie, 2005). Glycosylation of small molecular weight, lipophilic acceptors has been proved to be a key mechanism in the metabolic homeostasis of plant cells. It plays a major role both in cellular "housekeeping" and in buffering the impact of biotic and abiotic challenges on the plant (Dianna Bowles, 2005).

Glycosylation is the final step in the biosynthesis of secondary plant products resulting in the formation of an overwhelming number of natural glucosides with numerous applications. The isolation and purification of glucosides from plant sources is tedious and usually results in low yields. Glycosylation plays important role in the synthesis of natural β-D-glucosides. During synthesis of glucosides, the major problem is production of α- and β-glucosides instead of the desired pure β-compounds (Joachim Arend, 2001).

1.5.1 Physiological roles of secondary metabolite Glycosylation

From a chemical point of view, sugar conjugation results in both increased stability (through the protection of reactive nucleophilic groups) and water solubility. Glycosylation enhances water solubility of lipophilic membrane sterols and can lead to a change in cellular mobility, fluidity, permeability, hydration and phase behavior. It is considered to be a biological flag controlling the compartmentalization of metabolites, for example, the accumulation of compounds in the vacuole (Claire, 2005). Glycosylation stabilizes the products, modulates their physiological activities and governs intracellular distribution (Bhaskara Redday Madina 2007). One of the studies from *Arabidopsis thaliana* suggests that the glycosylation plays important role in defense response (Glambitza, 2004).

Chemical stabilization is well illustrated by the anthocyanins, in which glcosylation at the 3-OH position is crucial for the stability of the aromatic ring. It also strongly affects the bioavailability of the dietary compounds, which display antioxidant, atheroprotective or anticancer activity (Kramer *et al.,* 2003). Glycosylation also plays a role in the regulation of the active levels of several hormones. e.g. IAA, abscisic acid, cytokinins, salicylic acid (Xu, *et al.,* 2002).

Overall glycosylation roles in plants are: the stabilization of pigments, enhancement of solubility, storage of secondary metabolites and regulation of plant growth regulators (Jae Hyung Ko, 2006).

1.6 Glycosyltransferases: managers of small molecules

Glycosylation is a mechanism to improve the hydrophilicity of lipophilic compounds, thereby increasing their pharmacokinetics. This reaction can be carried out through enzymatic or chemical methods. Enzymatic glycosylation usually employs glycosyltransferases, glycosidases or glycosynthases. Glycosidases catalyse the hydrolysis of glycosidic linkages and also reverse hydrolysis and transglycosylation for glycoconjugate synthesis. The engineered glycosidases no longer possess the hydrolysis ability and the resulting mutants are named as glycosynthases. These enzymes act as biocatalysts with higher catalytic activity to synthesise glycosidic linkages (Eng-Kiat Lim, 2005).

Unlike glycosidases and glycosynthases, glycosyltransferases (GTs) are enzymes that have evolved naturally for glycosylation reactions. GTs transfer sugar moieties from activated sugar donors to acceptor molecules forming a glycosidic bond with high efficiency and regiospecificity (**Figure 7**).

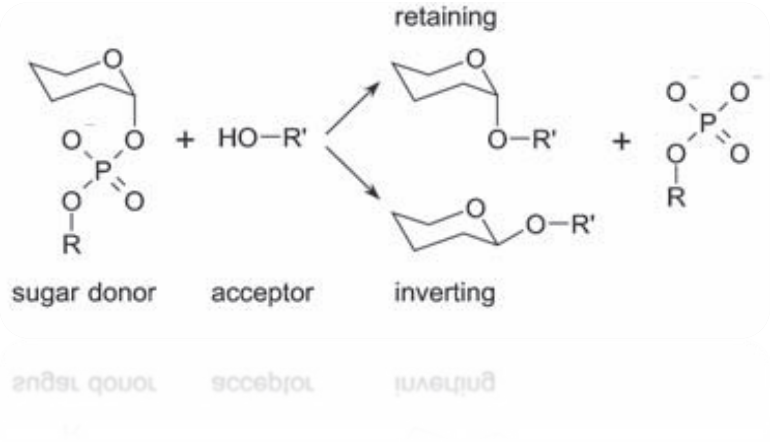


Figure 7 The retaining and inverting catalytic mechanism of GTs

Glycosyltransferases are found in all the living organisms. Glycosyl transfer reactions have been highlighted as the most important biotransformation on the earth, in quantitative terms they account for the assembly and degradation of the bulk of biomass (Yi Li, 2001).

1.6.1 Types of fold in glycosyltransferases

The modest degree of sequence homology within and sometimes among the various families has made the prediction of tertiary structures difficult; however, structural determinations in recent years have revealed that the catalytic domains of most glycosyltransferases display one of two fold types designated GT-A or GT-B (Coutinho *et al*., 2003; Bourne and Henrissat 2001).

GT-A and GT-B fold types consists of two closely associated domains at least one of which contains a Rossmann-fold responsible for donor nucleotide recognition. The Rossmann fold is a nucleotide binding domain (Dodson *et al*., 1966), and is a ubiquitous structural motif among enzymes with either GT-A or GT-B fold type, where it constitutes a dominant portion of the catalytic center in a cleft between the two domains. The Rossmann fold often contains much of the limited sequence homology that is observed across many glycosyltransferase families due to a finite repertoire of donor nucleotides utilized (Campbell *et al*., 1997).

The GT-A type fold (**Figure 8a**) is believed to be ancestral to enzymes with the GT-B type fold (Coutinho *et al*., 2003; Unligil *et al*., 2000). The donor usually binds the N-terminal domain's Rossmann fold. Typically one or two of this domain's β-sheets extend into the C-terminal domain, rendering definite separation of the folds difficult. The second domain is usually responsible for acceptor recognition and has greater sequence and structural variability among the different families than the nucleotide binding domain. This is presumably due to the limited number of different donor nucleotide sugars, but a vast number of different acceptor molecules. The two domains in the GT-A type fold cooperate to form the active- site cleft.

The GT-B type fold (**Figure 8b**) has two somewhat homologous Rossmann or Rossmann-like folds in distinct domains which are usually separated by a deep wide crevice while the enzyme is in an unliganded "open" conformation. Enzymes with the GT-B type fold have been suggested to have step-wise reaction mechanisms, where substrate binding in the "open" form can induce a conformational shift by a pair of main chain rotations to generate the "closed" form to align the nucleophile and substrates in the active site for catalysis (Coutinho *et al*., 2003; Unligil *et al*., 2000; Mulichak *et al*., 2003).

There have been two further fold types predicted using iterative BLAST searches and other methods (Rosen *et al*., 2004; Liu *et al*., 2003; Kikuchi *et al*., 2003). These so called GT-C and GT-D folds are thought to contain catalytic centers within transmembrane loops. As the catalytic domains themselves consist of integral-membrane proteins, they are intrinsically difficult to crystallize, and there is currently no structurally characterized example of either the GT-C or GT-D fold types.

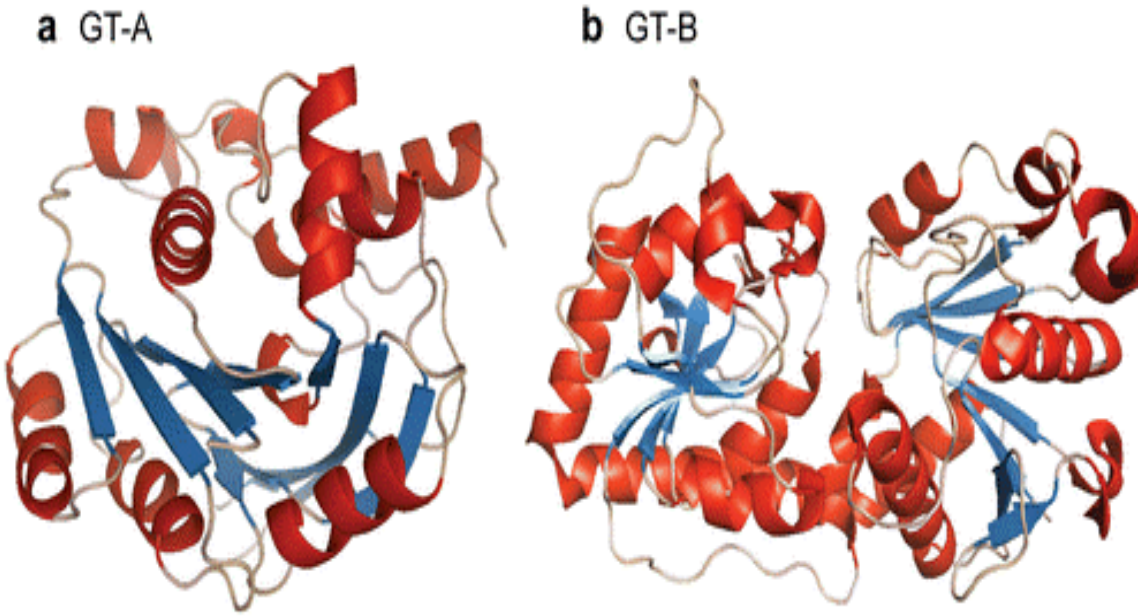
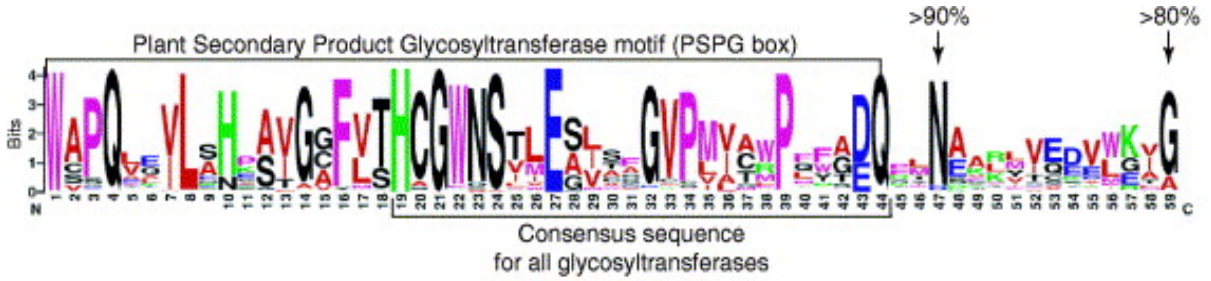


Figure 8: a) GT-A-fold type with a single Rossmann fold on the right, and b) GT-B fold type with two Rossmann folds.

1.6.2 Multigene families of plant glycosyltransferases

Glycosyltransferases are classified into families on the basis of sequence similarity, catalytic specificity and the existence of consensus sequences. According to carbohydrate-active enzyme database (CAZY.org) 12000 sequences from different organisms have been reported. These sequences were classified into 91 distinct families (Eng-Kiat Lim 2005). Among these, family 1 has the most number of GTs which have a close relationship with plant GTs. Most of the substrates for family 1 GTs are low-molecular weight lipophilic compounds. Glycosylation in these compounds occur at single or multiple sites and take place at the –OH, -COOH, -NH_2, -SH and C-C groups (Jun WANG, 2009).

Glycosyltransferases consists of Uridine-diphosphate glucose (UDP) glycosyltransferases (UGTs) and the glucose donors for GTs are UDP-glucose, UDP-glucuronic acid, UDP-xylose, UDP-galactose, and UDP-rhamnose (Bowles *et al.*, 2005). 50% of GTs contains a carboxy-terminal consensus sequence called the Plant Secondary Product Glycosyltransferase box (PSPG box). This box consists of 44 amino acids close to the C-terminal part of the protein, and is believed to be in binding of the activated donor sugars (**Figure 9**).



':' Highly conserved amino acids

Figure 9 Consensus sequence defining PSPG box

The amino-terminal regions of GTs are more variable and this domain is involved in the recognition and binding of diverse acceptors. A three dimensional model of glycosyltransferase from *Dorotheanthus bellidiformis* (Aizoaceae), indicates a direct

interaction between the uracil moiety of UDP-glucose and the highly conserved HCGWNS residues. Whereas, the last amino acid of the PSPG box contributes to the discrimination between UDP-glucose and UDP-galactose addition (Kubo, *et al*., 2004). The first plant UGT-encoding gene identified was an unexpected outcome of work by Nobel Prize winner Barbara McClintock on the genetic instability of transposons in maize. She studied the dark pigmentation of maize grains conferred by the mutation of Bronze1, which later turned out to encode a flavonoids UGT (Claire, 2005).

Sugar donor preference in plant UGTs is generally very specific. As it significantly impacts the vast structural diversity of natural phytochemicals, it is also of evolutionary and biochemical interest (Gachon *et al*., 2005). Elucidation of crystal structures of flavonoids glycosyltransferase provided information that was very useful for understanding the molecular basis of UGT catalysis. The PSPG box, a highly conserved motif among plant UFTs, is thought to be involved in sugar donor binding. Recent studies on UGT using homology modeling and biochemical analysis successfully identified a critical Arg residue for sugar donor specificity at the N terminus, which is far from the PSPG box (Akio Noguchi *et al*., 2009).

1.6.3 Identification of Glycosyltransferases

Several advanced approaches have been successfully used to clone, identify and analyze genes that encode plant GTs. The following approaches are: (Jun WANG, 2009)

- Biochemical methods
- Bioinformatics method
- Molecular biological methods
- Genetic methods

1.6.3.1 Biochemical methods

Isolation and purification of the GT enzymes directly from plants is the first step of the biochemical methods. Because of the complexity of protein components in plants, getting

highly purified enzyme is very difficult. Once the target proteins are purified, their enzyme activities for glycosylation of specific substrates can be investigated and the corresponding genes can be cloned by the derived nucleotide sequences from the amino acids.

1.6.3.2 Bioinformatics method

The development of genomics and bioinformatics greatly facilitates the identification of Plant GTs. The amino acid sequences encoded by the GT genes containing the consensus sequences vary in length from 435 to 507 amino acids. These amino acids have been found to possess nine conserved regions, including the UGT-defining consensus sequence. The level of similarity between these UGT amino acid sequences varies from over 95% to lower than 30% identity. Large number of GT amino acids contains a carboxy-terminal consensus sequence called the plant secondary product glycosyltransferases box (PSPG box) (Joe Ross, 2001). This consensus sequence provides a good starting point for searching new glycosyltransferases from a database. Once the related expressed sequence tags (ESTs), cDNAs or genes are identified by bioinformatics, a series of investigations such as cloning the full length of cDNAs, expressing and purifying recombinant proteins in vitro, analyzing the substrate specificity can be performed.

1.6.3.3 Molecular biological methods

Due to the rapid development of molecular biology, many methods are used in the identification of plant GTs. Based on the conserved amino acid sequences, degenerate primers can be designed and RT-PCR can be carried out using plant RNA material to clone putative GTs.

1.6.3.4 Genetic methods

The use of mutants is very important for gene identification and functional analysis. However, the lack of mutants of plant glycosyltransferases makes it difficult to use this method to identity GT genes currently.

1.6.4 Physiological roles of glycosyltransferases

Functional characterization of plant GTs indicates that glycosyltransferases play an important role in plant growth, development and interaction with the environment. The following physiological functions are:

- Hormone homeostasis
- Defense response
- Detoxification
- Biosynthesis and storage of secondary metabolites

1.6.4.1 Hormone homeostasis

The homeostasis of hormones is crucial for plant growth, development and adaptation responses to changes in the environment. The different mechanisms have been involved to control the level of different hormones in plant cells and tissues. Glycosylation is one of the mechanisms, because glycosides of all classical hormones except ethylene have been identified in plant extracts. After glycosylation the activities of hormones are reduced or lost. It was assumed that glycosylation could alter recognition between acceptors and hormones or change properties of hormones.

1.6.4.2 Defense response

Glycosyltransferases play an important role in glycosylating secondary metabolites and enhancing plant resistance to pathognes. TOGTs are *tobacco* glycosyltransferases with the highest *in vitro* enzyme activity towards hydroxycoumarin, scopoletin, and hydroxycinnamic acids. Down regulation of TOGTs in transgenic *tobacco* decreased the levels of scopoletin glucoside and simultaneously impaired the resistance to *Tobacco Mosaic Virus*. Over expression of TOGTs in transgenic tobacco caused enhanced resistance to *Potato Virus Y* (Jun WANG, 2009).

1.6.4.3 Detoxification

Fungal pathogens detoxify plant glycoside molecules by hydrolyzing the glycosidic bonds. In the battle between plants and pathogens, plants form glycosidic bonds to detoxify the toxicity of pathogens. The detoxification of trichothecene deoxynivalenol (DON) is produced by the fungus *Fusarium*, one of the most familiar fungi of cereal species such as wheat, barley and maize. DON is not only harmful to plant growth but also to the human health, and it is considered as a virulence factor in fungal pathogenesis. The recombinant protein of a putative glycosyltransferases from *Arabidopsis* catalyzes DON to form DON-3-O-glucoside. Glycosylated DON could lose toxicity, and over expression of putative GT in transgenic *Arabidopsis* could enhance the resistance of transgenics to DON (Poppenbergerger et al., 2003).

1.6.4.4 Biosynthesis and storage of secondary metabolites

Plants innately have developed many metabolic pathways to synthesize a vast number of secondary metabolites for environmental adaptation. Many studies have shown that glycosyltransferases and glycosylation reactions could be involved in the biosynthesis, modification, transportation and storage of secondary metabolites.

In the lignin biosynthesis pathway, lignin monomers (coumaryl, coniferyl and sinapyl alcohols) need to be translocated from the cytosol to the cell wall, where they are polymerized into lignin. The glucosides of lignin monomers have been considered as the transport forms (Lim *et al.,* 2005b).

1.7 Flavonoids and their glycosyltransferases

Flavonoids, which share a common 15-carbon polyphenolic skeleton, are secondary metabolites found mainly in plants. Thousands of flavonoids have been discovered, and they differ in the position of their hydroxyl and methoxyl groups, and in the number and identity of saccharide moieties.

Flavonoids, a large class of phenylpropanoid-derived secondary metabolites, are mostly glycosylated by UGTs with one or more sugar groups. The increasing number of biochemically characterized flavonoids UGTs allowed for the examination of the phylogenetic relationship between UGT structure and function (Tohge *et al*., 2005). Each flavonoid, 3-0-glycosyltransferse, 5-0-glycosyltransferase, and 7-0-glycosyltransferase forms a unique phylogenic cluster that is not limited by species. The regiospecificity of flavonoids UGTs for the sugar acceptor (i.e., the specificity of the glycosylation site on the sugar acceptor) was established prior to speciation. The recent identification of new cluster UGTs, which are specifically involved in glycosylation of the sugar moieties of glycosides, showed that regiospecificity for the sugar acceptor, but not the sugar donor specifically, is the basis for flavonoids UGT clusters (Akio Noguchi *et al*., 2009).

One of the most widely studied classes of plant glycosides is the large and heterogenic group of polyphenols. Till date, an overwhelming number of polyphenolic glycosides including flavonoids glycosides have been identified. Flavonoids are an important group of polyphenolic natural products and exhibit a wide range of biological natural products and exhibit a wide range of biological activities including antioxidant and estrogenic properties (Jae Hyung Ko *et al*., 2006).

Attachment of sugar to flavonoids occurs at the last step of biosynthesis pathway. Various flavonoids UGT genes have been cloned and characterized. Some of UGTs have been used for biocatalytic synthesis of flavonoids glycosides due to the complex and labor-intensive chemical synthesis. In addition, the nature of sugars and the glycosylation positions in flavonoids affect their absorption and utilization in humans (Byoung seok Hong *et al.*, 2007).

1.8 Rationale of thesis

W. somnifera is a medicinal plant known for several pharmacological properties attributed to its characteristic steroidal compounds, called withanolides and glycowithanolides. The biosynthetic pathway of withanolides, their functions in *withania* and metabolic steps leading to their glycol-transformations are unknown. One of the reasons for limited knowledge of

their functions is non-availability of the relavant enzymes and genes. For a long period of time glycosyltransferase activities and glycosylated products were known from a variety of plants, the enzymes and genes involved in the glycosylation failed to be isolated. In recent years, dozens of glycosyltransferases genes have been identified, and among these quite a few have been characterized functionally.

Plant-derived glucosides have attracted much attention due to their widespread applications. This class of products is difficult to isolate or to synthesize in pure form because of the resulting low yields. The isolation and purification of glucosides from plant sources is tedious and usually results in low yields. An enzyme-catalyzed transfer of glucose to aglycones would be an alternative approach to chemical synthesis, but requires the appropriate glycosyltransferases. These enzymes occur in only minute amounts in plant material, are mostly unstable and generally exhibit very limited substrate specificities. Substrate specificites have been poorly investigated and clearly none of the enzymes could be used to generate glucosides of a broader interest.

The present thesis deals with glycosyltransferase from *W. somnifera* glycosylates different flavonoid substrates. This flavonoid glycosyltransferase is expressed in active form and checked for substrate specificity. Also the work deals with another glycosyltransferase, which glycosylates steroidal sapogenins.

Chapter: 2

Materials and Methods

2. Materials and methods

This chapter includes general laboratory techniques routinely followed in the studies. Other important specific methodologies followed are discussed separately in the respective chapters.

2.1 Plant material

2.1.1 *Withania somnifera*

Withania somnifera plants were collected from the nursery of Pune University. These plants were grown in green house. They were watered at regular intervals. Seeds were collected from these plants and stored for further *in vitro* inoculation. Seeds were surface sterilized and grown in MS1/2 medium. Seedlings were transfered into liquid MS1/2 medium for rooting.
For genomic DNA isolation green house plant material was used. For RNA isolation and cDNA preparation (Gene isolation and Quantitative Real time PCR) *in vitro* cultures were used.

2.1.2 Media used

All media were based on Murashige and Skoog basal medium (Murashige & Skoog, 1962). The pH of the media was adjusted to 5.8 prior to autoclaving (121 $^{\circ}$C and 1.05 kg/cm^2 pressure for 20 minutes). Seeds were cultured on MS1/2 solid media with gelrite and seedlings for rooting inoculated on MS1/2 liquid media.

2.2 Glassware

Glassware used in all the experiments was procured from “Borosil”, India. Test tubes (25 mm x 150 mm), glass bottles (70 mm x 125 mm), petridishes (85 mm x 15 mm; 55 mm x 15 mm), conical flasks (100, 250 & 500 mL; 1, 2 & 5 L capacity) and pipettes (1, 2, 5, 10 and 25 mL capacity) were used during the course of study.

2.2.1 Preparation of glassware

Glassware used for all the experiments were cleaned by boiling in a saturated solution of Sodium bicarbonate for 1 h followed by repeated washing in tap water. Thereafter, it was immersed in 30% HNO_3 solution for 30 min followed by repeated washing in tap water and rinsed with distilled water. Washed glassware was thereafter dried at room temperature. Test tubes and flasks were plugged with absorbent cotton (Mamta Surgical Cotton Industries Ltd., Rajasthan, India). Autoclaving of the glassware and above items was done at 121 ºC and 1.05 kg/cm^2 pressure for 1 h.

2.3 Plasticware

Sterile disposable filter sterilization units (0.22 μM) and petridishes (55 mm and 85 mm diameter) were procured from "Laxbro", India. Microfuge tubes (1.5 mL and 2 mL capacity), microtips (10, 200 and 1000 μL capacity) and PCR tubes (0.2 mL and 0.5 mL capacity) were obtained from "Tarsons" and "Axygen", India.

2.4 Chemicals

Tris, IPTG, X-gal, SDS, BSA, EDTA, PMSF, PIPES, DTT, PVPP, APS, Imidazole, Urea, Ethidium bromide, Ampicillin, Kanamycin, TEMED, dNTPs, Agarose, Acrylamide, Bis-Acrylamide, RNase A were purchased from Sigma-Aldrich (USA) and Bioworld (USA). Restriction enzymes, T4 DNA ligase, Taq DNA polymerase and lysozyme were obtained from NEB (USA), Promega (USA), Bioenzymes (USA), Amersham (UK) and Banglore Genei (India). Different kits were purchased from SMART RACE KIT (BD CLONETECH, JAPAN), Gene Race Kit (Invitrogen, USA). Plasmid vectors, pGEM-T Easy Vector and pET30b (+) were purchased from Invitrogen (USA), Promega (USA) and Novagen (USA) respectively. Megaprime labeling kit and Hybond-N+ membrane were obtained from Amersham (UK). [α- ^{32}P]-dATP and [α- ^{32}P]-dCTP were obtained from Bhabha Atomic Research Centre (BARC), India. All other chemicals and solvents of analytical grade were purchased from HiMedia, Qualigens fine chemicals and E-Merck Laboratories, India. All chemicals used in the tissue culture study were of analytical grade (AR) and were obtained

from Qualigens, S.D. Fine Chemicals and HiMedia, India. The Sucrose, glucose and agar-agar were obtained from Hi- Media. Bacto-Agar for microbial work was obtained from DIFCO laboratories, USA. Substrates and standards for enzyme assays i.e. naringenin, diadzein, genistein, apigenin, catechin, myricetin, isorhamnetin, luteolin, hesperetin, curcumin, kaempferol, diosgenin, capsaicin, vanilic acid, salicylic acid and all the respective glycosides were also obtained from Sigma-Aldrich (USA) and Chromadex (USA). Immobiline Drystrip gel (3-10 pH) 13cm, Destreak rehydration buffer, IPG buffer, Mineral oil, Iodoacetamide from GE life sciences (Sweden).

2.5 Equipments

Table 2.1 List of equipments used in present study

S. No	Equipment	Make
1	Balances	Contech/Sartorious
2	Water bath	Fisher Scientific/Julabo
3	Dry Bath	Eppendorf/Banglore Genei
4	Incubator/shaker	New Brunswick
5	Centrifuge	Sorvall/Haereus/Eppendorf/Sigma
6	Gel Documentation system	Bio-Rad
7	Thermo Cycler PCR machine/ Real Time PCR	BioRad/Stratagene
8	Spectrophotometer	Perkin Elmer Lambda 650
9	Power pack	Bio-Rad
10	Agarose Gel Electrophoresis Units	Bangalore Genei/Bio-Rad
11	Protein Gel Electrophoresis Units	GE life science

12	Speed Vac concentrator	Savant/Eppendorf
13	pH-Meter	Microset
14	Water purification system	Millipore Unit (Milli RO/Milli Q)
15	Microwave oven	Electrolux
16	Fridge/ Deep freezer	Vestfrost/Leonard/Godrej
17	Magnetic rotator	REMI
18	Laminar Air Flow	Microfilt India
20	Typhoon Trio + Scanner	GE life science (USA)
21	HPLC	Perkin Elmer
22	ELISA Plate Reader	Bio-Rad(USA)
23	iBlot Gel Transfer System	Invitrogen
24	LC-MS	Waters
25	MALDI-MS	Waters
26	AKTA Explorer	GE life science
27	Sonicator	Misonix-XL-2000
28	NanoVue	GE life science
29	Hybridization Incubator	SciGene
30	2D Gel system	Ettan ^{TM} IPGphorIII (GE)

2.6 Buffers and solutions

Table 2.2 Buffers and Solutions for DNA Electrophoresis

Name	Components	Preparation and Storage
50x TAE	2 M Tris Acetic acid 0.05 M EDTA	pH was adjusted to 8.0 using glacial acetic acid and stored at room temperature
TBE buffer	90 mM Tris 90 mM Boric acid 2 mM EDTA	Room temperature
DNA loading buffer	0.25 g Xylencyanol 0.25 g Bromophenol blue 0.25 g Ficoll 400 1.46 g EDTA Make up the volume to 100 mL with H_2O	The solutions were filter sterilized using 0.22 micron (µ) filter and stored at room temperature

2.6.1 Buffers and Solutions for g-DNA isolation

Table 2.3 Buffers and Solutions for g-DNA isolation

Name	Components	Preparation and Storage
Extraction buffer	100 mM Tris-HCl (pH 8.0) 20 mM Na EDTA (pH 8.0) 1.4 M NaCl 2.0% (w/v) CTAB Add β-mercaptoethanol 0.2 % before use.	Room temperature.

2.6.2 Buffers and solutions for Southern Hybridization

Table 2.4 Buffers and solutions for Southern Hybridization

Name	Components	Preparation and Storage
Depurination buffer	0.25 N HCL	Freshly prepared.
Denaturation buffer	1.5 M NaCl 0.5 M NaOH	Room temperature.
Neutralization buffer	1.5 M NaCl 1.0 M Tris HCl (pH 7.4)	Room temperature.
20 X SSC	3 M NaCl 0.3 M Sodium citrate (pH 7.0)	Room temperature.
Hybridization buffer	1% BSA 1.0 mM EDTA pH 8.0 0.5 M Sodium phosphate pH 7.2 7% SDS	Room temperature.
Low stringency Wash buffer	6 X SSC 0.1% SDS	Room temperature.
Moderate stringency Wash buffer	2 X SSC 0.1% SDS	Room temperature.
High stringency Wash buffer	0.2 X SSC 1% SDS	Room temperature.

2.6.3 Solutions for the transformation and selection of *E. coli*

E. coli was grown at 37 ºC with shaking at 200 rpm in Luria Bertani (LB) broth/ SOB broth And maintained on LB/SOB plates with 1.5% agar (Sambrook *et al*., 1989). For plasmid DNA preparation recombinant *E. coli* was grown in LB media supplemented with appropriate antibiotics. See Table for Stock solutions for *E. coli* transformation and selection.

Table 2.5 Stock solutions for *E. coli* transformation and selection

Name	Components	Preparation and Storage
IPTG solution	200 mg/mL in SMQ	Sterile filtration and storage at -20 ºC
X-Gal (5-bromo-4chloro -3-indolyl-β-D-galactoside)	20 mg/mL in N,N'-Dimethyl formamide (DMF)	Light sensitive, covered & store at -20 ºC
Ampicillin	100 mg/mL in SMQ	Sterile filtration and storage at -20 ºC
Tetracycline	12.5 mg/mL in 70% ethanol	Sterile filtration and storage at -20 ºC
Kanamycin	50 mg/mL in SMQ	Sterile filtration and storage at -20 ºC

2.6.4 Buffers and Solutions for plasmid isolation (Alkaline lysis method)

Table 2.6 Buffers and Solutions for plasmid isolation (Alkaline lysis method)

Name	Components	Preparation and Storage
Solution I (GTE) or Resuspension buffer	50 mM Glucose, 25 mM Tris-HCl (pH 8.0),10 mM EDTA (pH 8.0)	Store at 4 ºC
Solution II or Lysis buffer	0.2 N NaOH and 1% SDS	Freshly prepared

Solution III or Neutralisation buffer	3M Potassium acetate (pH-5.3)	Store at 4 ºC
RNase A	10 mg/mL	Store at -20 ºC
Other solutions or Reagents	Chloroform, Absolute ethanol, 3.0 M Sodium acetate, 70% ethanol & Deionized sterile water	Room temperature

2.6.5 Buffers and Solutions for Gel Electrophoresis (native/ SDS-PAGE)

Table 2.7 Buffers and Solutions for Gel Electrophoresis (native/ SDS-PAGE)

Name	Components	Preparation and Storage
Monomer solution	29.2% acrylamide 0.8% bis-acrylamide in water	Store at 4 ºC (in darkness)
Stacking gel	Distilled water 2.76 mL 1 M Tris-HCl (pH 6.8) 0.5 mL Acrylamide/bis 30% 0.65 mL 10% SDS (w/v) 0.04 mL 10% APS 0.04 mL TEMED 4 µL	Freshly prepared
Separating gel (10%)	Distilled water 3.162 mL 1.5 M Tris-HCl (pH 8.8) 2.0 mL Acrylamide/Bis 30% 2.67 mL 10% SDS (w/v) 0.08 mL 10% APS 0.08 mL TEMED 8 µL	Freshly prepared
2x Protein loading	Distilled water 2.7 mL 0.5 M Tris-HCl (pH 6.8)	Store at 4 ºC

buffer	1.0 mL Glycerol 2.0 mL 10% SDS 3.3 mL β-Mercaptoethanol 0.5% Bromophenol blue	
10x SDS-electrode buffer	Tris base 15.1 g Glycine 94.0 g SDS 0.5 g Adjust pH-8.3 and make-up the volume up to 500 mL.	Store at 4 ºC Dilute 1:10 before use
Staining solution	Coomassie-blue R 250, 0.25 g (in Methanol 40 mL, Acetic acid 10 mL) Make-up to 100 mL	Store at RT
Destaining solution	Methanol 40 mL Acetic acid 10 mL Make- up to 100 mL	Store at RT
Silver staining Fixer solution	40% Methanol, (150 mL) 10% acetic acid, (50 mL) Make-up to 100 mL	Store at RT
Sensitizing solution	0.2% Na_2S_2O3	Store at RT
Silver solution	0.2% silver nitrate (0.6 g) 0.01 % formaldehyde (225 µL) Make-up to 300 mL	Prepare fresh in darkness
Developing solution	6% Na_2CO_3 (18 g) 0.02% formaldehyde (150 µL) Make-up to 300 mL	Prepare fresh
Stop solution	1.5% Na_2EDTA (4.5 g) Make-up to 300 mL	Store at RT

2.6.6 Buffers and solutions for protein purification (based on imidazole conc.)

Table 2.8 Buffers and solutions for protein purification (based on imidazole conc.)

Name	Components	Preparation and Storage
Binding buffer	50 mM Tris 150 mM NaCl 20 mM imidazole 5% glycerol (pH 8.0)	Adjust pH by adding concentrated HCl and store at 4 ºC
Wash buffer	50 mM Tris 300 mM NaCl 30 mM imidazole 10% glycerol (pH 8.0)	Adjust pH by adding concentrated HCl and store at 4 ºC
Elution buffer	50 mM Tris 300 mM NaCl 250 mM imidazole 10% glycerol (pH 8.0)	Adjust pH by adding concentrated HCl and store at 4 ºC

2.6.7 Buffers and solutions for protein extraction under denaturing conditions

Table 2.9 Buffers and solutions for protein extraction under denaturing conditions

Name	Components	Preparation and Storage
Lysis buffer	50 mM Tris-HCl (pH 8.0) 5 mM EDTA 100 mM NaCl 0.5% TritonX-100 0.7 mM DTT 0.1 mM PMSF (Freshly added) 10 mM $MgSO_4$ Lysozyme 100 µg/mL (Added freshly)	Store at 4 ºC
Sonication buffer	100 mM Tris HCl (pH8.0)	Store at 4 ºC

	50 mM Glycine	
Dispersion buffer	100 mM Tris-HCl (pH8.0) 50 mM Glycine & 8 M Urea /6 M GuHCl	Store at 4 ºC

2.6.8 substrates used for the study

Table 2.10 substrates used for the study

Substrate/Products	Manufacturer
Apigenin	Sigma-Aldrich (USA)
Daidzein	Sigma-Aldrich (USA)
Naringenin	Sigma-Aldrich (USA)
Catechin	Sigma-Aldrich (USA)
Myricetin	Sigma-Aldrich (USA)
Isorhamnetin	Sigma-Aldrich (USA)
Genistein	Sigma-Aldrich (USA)
Luteolin	Sigma-Aldrich (USA)
Hesperetin	Sigma-Aldrich (USA)
Curcumin	Sigma-Aldrich (USA)
Kaempferol	Sigma-Aldrich (USA)
Diosgenin	Sigma-Aldrich (USA)
Capsaicin	Sigma-Aldrich (USA)
Vannilic acid	Sigma-Aldrich (USA)

Salicylic acid	Sigma-Aldrich (USA)
Resorcinol	Sigma-Aldrich (USA)
Thymol	Sigma-Aldrich (USA)
Euginol	Sigma-Aldrich (USA)
Benzoic acid	Sigma-Aldrich (USA)
4-nitro phenol	Sigma-Aldrich (USA)
Methyl vanilate	Sigma-Aldrich (USA)
3,4dimethoxyphenol	Sigma-Aldrich (USA)
Quercetin	Sigma-Aldrich (USA)
Biochanin A	Sigma-Aldrich (USA)
Syringic acid	Sigma-Aldrich (USA)
Sarsasapogenin	Sigma-Aldrich (USA)
Digitoxigenin	Sigma-Aldrich (USA)

2.6.9 Standards used for the study

Table 2.11 Standards used for the study

Genistein 4`-glucoside	Chromadex (USA)
Genistein 7-glucoside	Chromadex (USA)
Naringenin 7-O-glucoside	Chromadex (USA)
Kaempferol 7-glucoside	Chromadex (USA)

Apigenin-7-O-glucoside	Chromadex (USA)
Luteolin-4-glucoside	Chromadex (USA)
Luteolin-7-O-glucoside	Chromadex (USA)
Diadzin	Chromadex (USA)

2.6.10 Sugar donors for glycosyltransferase assay

Table 2.12 sugar donors for glycosyltransferases assay

UDP-Glucose	Sigma-Aldrich (USA)
UDP-Galactose	Sigma-Aldrich (USA)
UDP-Xylose	Sigma-Aldrich (USA)
UDP-Glucuronic acid	Sigma-Aldrich (USA)

2.6.11 Different inducing media and hormones

Table 2.13 Component of Murashige & Skoog media/Different inducing media and hormones

Name	Components	Preperation and storage
Major component	20.61 mM NH_4NO_3 18.75 mM KNO_3 2.99 mM $CaCl_2.2H_2O$ 1.5 mM $MgSO_4.7H_2O$ 1.24 mM KH_2PO_4	Store at 4 ºC
Minor components	0.147 mM $MnSO_4$ 5.3×10^{-2} mM $ZnSO_4$	Store at 4 ºC

	1.56 x 10^{-4} mM $CuSO_4$ 1.05 x 10^{-4} mM $CoCl_2.6H_2O$ 4.99 x 10^{-3} mM KI 0.1mM H_3BO_4 1.03 x 10^{-3} mM $Na_2MO_4.2H_2O$	
Vitamins	5.55 x10^{-2} mM Myoionsitol 4.06 x 10^{-3} mM Nicotinic acid 2.43 x 10^{-3}mM Pyridoxine HCl 2.96 x 10^{-4} mM Thymine HCl 2.66 x 10^{-2} mM Glycine	Store at 4 ºC
Iron	0.1 mM $FeSO_4.7H_20$ 0.1 mM Na_2EDTA	Store at 4 ºC
BAP	1.776 mM BAP (dissolve in NaOH and make up the volume by adding ethanol)	Store at 4 ºC

2.6.12 Different media used for bacterial growth

Table 2.14 Different media used for bacterial growth

Name	Components	Preparation and Storage
Luria Bertani Broth (LB)	1% Bactotryptone 0.5% Yeast extract 1% NaCl	pH adjusted to 7.0 with NaOH, store at room temperature or at 4 ºC
SOB media	2% Bactotryptone 0.5% Yeast extract 10 mM NaCl 10 mM $MgCl_2.6H_2O$	pH adjusted to 6.8 with NaOH, store at room temperature or at 4 ºC

	2 mM KCl	
TB buffer	10 mM PIPES 15 mM CaCl2 250 mM KCl	pH was adjusted 6.8 with KOH. $MnCl_2$ was added to final concentration of 55 mM and filter sterlized.

2.7 Host cells

E.coli	**Genotype**
DH 5α	F' _80_lacZ_M15 end A1 hsdR17 (rk-mk+) supE44 thi-1_-gyrA96 relA1 _(lacZYA-argFV169) deoR
JM 109	e14-(McrA-) recA1 endA gyrA96 th-1 hsdR17(rk-mk+) supE44 relA1 _(lac-proAB) [F' traD36 proAB lacqZ _M15
XL1 Blue	RecA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)
TOPO 10	F- mcrA Δ(mrr-hsdRMS-mcrBC) f80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG
BL 21	F-, ompT hsdSB (rB – mB -) gal dcm (DE3) pLysS (CamR)

2.8 Methods

2.8.1 Withania seed surface sterilization

The seeds of *Withania somnifera* were first rinsed twice with sterile distilled water in sterile Jam bottles in laminar air flow. Then treated with Teepol (1 ml) added in Savlon (90%) for 3 min and rinsed by sterile distilled water. Seeds were further treated with 70% ethanol for 5min and rinsed by distilled water. Finally, seeds were treated with 0.05% $HgCl_2$ for 5min and rinsed thoroughly by distilled water and were taken out on whattman filter paper for further inoculation.

2.8.1.1 Inoculation and Incubation

The surface sterilized seeds were inoculated on MS1/2 solid medium under aseptic conditions. Cultures were incubated at temperature 26±1°C under 16 hour photoperiod at 11.7 µmol/m^2/sec light intensity/8 h dark cycles for about 15 days. The sprouting was observed after 15 days and the seedlings were grown up to 3-5 cm within 4 weeks after inoculation. For rooting, seedlings were transferred on half strength MS without supplementing gelling agent. The seedlings were supported by 2 cm^2 whattman filter paper no.1 in culture tubes. The seedlings were grown up to 9-10 cm within 15 days after shifting. Well grown cultures were used for further experiments.

2.8.1.2 Stress treatment for Quantitative expression

In vitro seedlings were used for study of different stress treatments like methyl jasmonate, salicylic acid, cold and heat. Leaf, stem and root tissues were collected and incubated in MS1/2 liquid medium for different time intervals (2 h, 4 h, 6 h, 8 h and 10 h) at different concentrations of methyl jasmonate and salicylic acid (50µM and 100 µM). Approximately 0.2-0.4 gm of tissue was inoculated on the medium containing methyl jasmonate, salicylic acid. Tissues were also incubated for heat (42 ºC) treatment for 0 min to 90 min time period.

After incubation period RNA was isolated and cDNA was prepared for Quantitative PCR analysis.

2.8.2 Bacterial growth & transformation

2.8.2.1 Bacterial culture conditions

E. coli was grown at 37 ºC with shaking at 200 rpm in Luria Bertani (LB) broth/ SOB broth and maintained on LB/SOB plates with 1.5% agar (Sambrook *et al*., 1989). For plasmid DNA preparation recombinant *E. coli* was grown in LB medium supplemented with appropriate antibiotic(s).

2.8.2.2 Preparation of competent cells using TB buffer

A single colony of *E. coli* XL1 Blue was inoculated in 5 mL of LB medium containing tetracycline (12.5 mg/L) and grown overnight at 37 ºC in incubator shaker at 200 rpm. Overnight grown culture (500 µL) was added to 50 mL of SOB medium and grown for 2-3 hours at 37 ºC in incubator shaker at 200 rpm. Cells were kept on ice for 10 min and harvested by centrifugation at 5,000 rpm for 10 min at 4 ºC. The cell pellet was suspended in 5 mL ice-cold TB buffer containing 7% DMSO. This was then dispensed into aliquots of 100 µl in 1.5 mL eppendorf tubes, frozen in liquid nitrogen and stored at -80ºC.

2.8.2.3 Preparation of competent cells using $CaCl_2$

A single colony of *E. coli* XL1 Blue was inoculated in 5 mL of LB medium containing tetracycline (12.5 mg / L). LB medium (50 mL) was inoculated with 1% of the overnight grown *E. coli* culture and allowed to grow till 0.5 O.D. at 600 nm. The cells were harvested by centrifugation at 5,000 g for 10 min at 4 ºC, suspended in ice-cold 100 mM $CaCl_2$ and kept on ice for 30 min. Cells were centrifuged, the pellet suspended in 1 mL of ice-cold 100 mM $CaCl_2$ and stored as aliquots of 200 µL at 4ºC or in -70 ºC after adding 15% glycerol or 7% DMSO.

2.8.2.4 *E. coli* transformation

The competent *E. coli* cells were transformed according to Sambrook *et al.,* (1989). Gently, DNA (~50 ng in 10 µL or less) was added to the (200 µL) competent *E. coli* cells, mixed and kept on ice for 30 min. The cells were then incubated at 42 ºC for 2 min (heat shock) and immediately kept back on ice. To each tube 800 µL of LB broth was added and further incubated at 37 ºC for 1 h. Cells were pelleted by centrifugation at 5,000 rpm for 5 min. The supernatant media was discarded and pellet was dissolved in 100 µL of LB broth. The resuspended cells were plated on LB agar medium with appropriate antibiotics, IPTG and X-gal as per need (Sambrook *et al*., 1989).

Stock solutions and Final concentration

Solutions	Stocks	Final concentration
IPTG	200 mg/mL in sterile distilled water.	40 µg/ mL
X-gal	20 mg/mL in dimethylformamide.	40 µg/ mL

2.8.2.5 Colony PCR for screening recombinant colonies after bacterial transformation

This method bypasses DNA purification, and relies on the selectivity of PCR amplification to determine whether a bacterial colony of interest does indeed contain the desired DNA. Simply adding a small portion of a bacterial colony to a PCR master mix will introduce enough template DNA for amplification. A single bacterial colony was picked up from the agar plate containing transformants with the help of microtip and added to 1.5 mL eppendorf PCR tube containing 20 µL of sterile miliQ water. The microtip was agitated in the water to suspend the bacterial cells. Subsequently, 2 –3 µL of this suspension was used in 15 µL of PCR reaction. The remaining components were added to the PCR reaction and subjected to normal cycling parameters for the particular primers. If insert orientation, as well as presence, needs to be determined, utilization of a forward vector-specific primer and a reverse insert-specific primer, or vice versa, allows such determination. If only the presence of the insert needs to be determined, then gene specific primers can be used. An additional 5 min denaturation step at 95 ºC before the amplification cycles will aid the bacterial lysis to enhance PCR product

amplification success. The resulting PCR products were checked on an agarose gel for the presence of the cloned gene of interest.

2.8.3 Isolation of nucleic acids and Polymerase Chain Reaction (PCR)

2.8.3.1 Isolation of plasmid DNA from *E. coli* cells

The alkaline lysis method of Sambrook *et al*., (1989) was improvised upon so that 12-24 samples could be processed conveniently for plasmid DNA extraction within 3 h, with yields of 5-30 μg per 1.5 mL culture depending on the host strain and the plasmid vector. An important feature of this protocol was the use of PEG for purification, which resulted in precipitation of high quality super-coiled plasmid DNA free of contamination. The bacterial cultures were grown overnight with shaking (200 rpm) at 37 ºC in LB broth, with appropriate antibiotic(s). About 1.5 to 3 mL culture was centrifuged for 1 min at 7,000 g to pellet the bacterial cells. The pellet was resuspended in 100 μL of GTE buffer (Table 2.6) by vigorous pipetting, 200 μL of Solution. II (Table 2.6) was added, mixed by inversion till the solution becomes clear, normally for 2-3 min. The cell lysate was neutralized by addition of 150 μL of Solution. III (Table 2.6) mixed well and incubated on ice for 5 min. The cell debris was removed by centrifugation for 10 min at 12,000 g at 4 ºC. The supernatant was transferred to another microfuge tube, RNase A was added to a final concentration of 20 μg/mL (Sambrook *et al*., 1989) and incubated at 37 ºC for 20 min. To the above solution equal volume of chloroform:isoamyl alcohol (24:1) was added, vortexed for 2 min and centrifuged for 10 min at 12, 000 g at room temperature. The upper aqueous layer was transferred to a clean tube, 1/10th volume sodium acetate and one volume absolute ethanol was added with gentle mixing and kept at -20 ºC for 1-2 h. The sample was centrifuged at 12,000 g for 10 min at 4 ºC. The pellet was washed thrice with 70% ethanol and dried under vacuum. The dried pellet was dissolved in 40 μL of deionized water and 40 μL of PEG/NaCl solution (20% PEG 8000 in 2.5 M NaCl) was added. The mixture was incubated on ice for 20 min and the plasmid DNA pelleted out by centrifugation at 12,000 g for 15 min at 4 ºC. The supernatant was aspirated carefully, the pellet washed with 70% ethanol and air-dried. The dried pellet was resuspended in 20 μL deionized water and stored at -20 ºC.

2.8.3.2 Isolation of plant genomic DNA

Extraction buffer: 100 mM Tris-HCl and 20 mM sodium EDTA adjust pH to 8.0 with HCl; add NaCl to 1.4 M and 2.0 % (w/v) CTAB (cetyltrimethylammonium bromide). Dissolve CTAB by heating to 60 °C and store at room temperature. Add β-mercaptoethanol to 0.2% before use. Chloroform: isoamyl alcohol 24:1 (v/v), 5 M NaCl, RNAase A (10 mg/mL), 95% ethanol, 70% ethanol, TE buffer: 10 mM Tris-HCl and 1 mM EDTA, adjust pH to 8.0 and autoclave. Genomic DNA was isolated by using the protocol of Lodhi *et al*., (1994). Fresh young leaves were collected, frozen in liquid nitrogen and crushed to a fine powder. About 1.0 g of ground tissue was extracted with 10 mL extraction buffer. The slurry was poured into a clean, autoclaved 50 mL centrifuge tube and 100 mg insoluble polyvinylpolypyrrolidone (PVPP) as well as 20-40 µL of β-mercaptoethanol was also added. The tube was inverted several times but very gently to thoroughly mix the slurry, incubated at 65 ºC for 30 min and then allowed to cool down to room temperature. 12 mL of chloroform: isoamyl alcohol mixture was added and the contents mixed by inverting the tube gently till an emulsion formed. The mixture was then centrifuged at 6,000 g for 15 min at room temperature. Supernatant was carefully collected in a fresh tube and chloroform: isoamyl alcohol (24:1) extraction step repeated. To the clear supernatant 0.5 volume of 5 M NaCl was added and mixed gently and two volumes of cold (-20 ºC) 95 % ethanol was added and the sample kept at 4 ºC until DNA strands appeared. The tube was centrifuged at 3,000 g for 3 min and then at 5,000 g for next 3 min. The supernatant was poured off, and DNA pellet washed with cold (4 ºC) 70 % ethanol and air-dried. DNA was dissolved in 400 µL of TE buffer. The DNA solution was treated with 1 µL RNase A (10 mg/mL) per 100 µL DNA solution and incubated at 37 ºC for 30 min. The sample was extracted with chloroform: isoamyl alcohol to remove RNAase A. DNA was re-precipitated and dissolved in 40-100 µL TE buffer. Purity of DNA was checked spectrophotometrically by measuring the absorbance ratio (A_{260}/A_{280}) and also by visualization on 0.8 % agarose gel with 1X TAE. Genomic DNA was stored at 4 ºC.

2.8.3.3 Restriction digestion of DNA

Plasmid and genomic DNA restriction digestion was set up as per restriction enzymes manufacturer's recommendations.

2.8.4 Nucleic acids blotting/hybridization

2.8.4.1 Southern blotting

For Southern hybridization (Southern, 1975) the DNA samples were electrophoresed on an agarose gel in 1X TAE buffer containing 0.5 μg/mL ethidium bromide. The gel was rinsed with deionized water (D/W) and placed in depurination solution for 15 min. The gel was rinsed with deionized water and immersed in denaturation solution for 30 min with gentle shaking. The gel was again rinsed with deionized water and transferred to neutralization solution for 45 min. The gel was next set up for capillary transfer of DNA to solid membrane support. A tray was filled with the transfer buffer (20X SSC). A platform was made and covered with a wick, made from 2 sheets of Whatman 3 MM filter paper saturated with transfer buffer and the gel was placed on it. It was surrounded with saran wrap to prevent the transfer buffer from being absorbed directly by the paper towels stacked above the membrane. A sheet of Hybond N+ membrane (Amersham, UK) of the exact gel size was wetted with deionized water followed by transfer buffer (20X SSC) and then placed on top of the gel. A glass rod was rolled over the membrane to remove any trapped air bubbles. One piece of Whatman 3MM paper wetted with 20X SSC was placed on the membrane followed by Whatman 3MM paper pre-wetted in 2X SSC. On this paper another dry Whatman 3MM paper was placed followed by a stack of absorbent paper towels. A glass plate and a ~0.5 kg weight were placed on the top of the paper towels. Transfer of DNA was allowed to proceed for 18 h. The membrane was marked for orientation, removed carefully and washed with 6X SSC. The membrane was air dried and baked for 2 h at 80 ºC to immobilize DNA onto the nylon membrane. Hybridization and autoradiography were carried out as is described in the following section.

2.8.4.2 Random primer labeling

Random primer labeling of the DNA probes (Feinbeng and Vogelstein, 1983) was done using the Megaprime DNA labeling kit (Amersham, UK). Reaction (50 μL) was set up as follows:

Component	Volume
25 ng DNA (used as probe)	5.0 µL
Primer solution (Random hexamers (3.5 A260 U)	5.0 µL
Final Volume	10.0 µL

Above mixture was heated in a boiling water bath for 10 min and cooled to room temperature facilitating primer annealing to the DNA. This reaction mixture was added to the following reaction mixture.

Components	**Volume**
10 X reaction buffer (500 mM Tris-HCl, pH 8.0; 100 mM $MgCl_2$; 10 mM DTT; 0.5 mg/mL acetylated BSA)	5.0 µL
dATP (0.5 mM)	4.0 µL
dGTP (0.5 mM)	4.0 µL
dTTP (0.5 mM)	4.0 µL
(α-3^2P)-dCTP (Sp. activity 3000 Ci/mmol)	5.0 µL
Sterile deionized water	16.0 µL
Exonuclease free Klenow fragment (2 U/ µL)	2.0 µL
Total volume	50.0 µL

The reaction was carried out at 37 ºC for 1 h and stopped by incubation at 95 ºC in a dry bath for 10 min and snap chilled on ice.

2.8.4.3 Pre-hybridization and hybridization Solutions

20X SSC: 3 M NaCl; 0.3 M Sodium citrate (pH 7.0)

Hybridization buffer: 1% BSA; 1.0 mM EDTA, pH 8.0; 0.5 M Sodium phosphate, pH 8.0; 7% SDS

Low stringency wash buffer: 2 X SSC, 0.1% SDS

High stringency wash buffer: 0.2 X SSC, 1% SDS

The blots made as in section 2.8.4 mentioned were pre-hybridized at 62 ºC in 30 mL of hybridization buffer for 6-8 h in a hybridization incubator (Robin Scientific, USA). The buffer was decanted and fresh buffer added with the denatured radiolabelled probe. Hybridization was carried out at 60 ºC for 14-18 h. The solution was discarded and the membrane washed with low stringency buffer at 55-65 ºC for 5 min followed by a high stringency wash at 55-65 ºC for 5 min. The moist blot was wrapped in saran wrap and placed in the developing / intensifying screen provided with the Typhoon TRIO+ scanner. After 6-8 h of exposure, remove the membrane and scan the screen with Typhoon trio scanner system. One can increase the exposure time if the band intensity is weak. The membrane scanning is done usually in the resolution mode instead of sensitivity mode of the scanner. Good quality of band pictures can be obtained by scanning at 500 micron resolution with the system. This instrument will provide many more options and functions to get exquisite band pictures of the exposed membrane. After scanning the intensifying screen, the signals can easily be erased from it, upon exposure to a white light provided with the system for 10 min and the same screen can be used to develop the other membrane.

2.8.5 Extraction and purification of DNA from agarose gels

The restriction digested DNA or PCR amplified products were run on an agarose gel in 1X TAE buffer. The gel was stained with ethidium bromide (0.5 µg/mL) and viewed using a hand held long wavelength UV illuminator. The fragment of interest was excised from the gel and weighed. A 100 µg gel slice was transferred to a 1.5 mL micro centrifuge tube and 300 µL buffer DE-A (AxygenTM GEL elution kit, Biosciences, USA) added. The tube was incubated at 70 ºC for 5 to 10 min with intermittent mixing until the gel slice was completely dissolved. The gel mixture was cooled down to room temperature and 150 µL of buffer DE-B was added. The above melted agarose was put into Axyprep column and placed into 2 mL microfuge collection tube. The assembly was centrifuge at 12,000 g for 1 min and filtrate was discarded. 500 µL of wash buffer 1 (provided by Axygen) was added and centrifuged at

12,000 g for 30 s, filtrate was discarded. 700 µL of wash buffer 2 was added and spin at 12,000 g for 30 s, filtrate was discarded. It was repeated again with wash buffer 2. One min empty spin was given to ensure the complete removal of salt as well as ethanol. Axyprep column was transferred into a fresh 1.5 mL microfuge tube and 25-30 µL of elution buffer was added to the centre of the membrane and kept it as such for 1 min at room temperature. Centrifuge at 12000 g for 1 min. The eluted DNA was stored at 4 ºC. This eluted PCR product or any DNA fragments are of good quality and can be visualized on 1% agarose gel by taking an aliquots of 3-4 µL. The eluted DNA /PCR product was stored at -20 ºC and was used for subsequent reactions.

2.8.6 Total RNA Isolation

RNase free environment was created and maintained as described by Blumberg (1987). All glass and plasticware were DEPC (0.1% in water) treated overnight and autoclaved. The pestle and mortar were also DEPC treated and then baked at 300 ºC for 6 h. All materials were dried in a vacuum oven. Total RNA from different plant tissues was isolated using TRIzol reagent. The plant tissue was collected, washed with DEPC treated water, frozen in liquid nitrogen and crushed to a fine powder. To 100 mg of the fine powder, 1 mL TRIzol reagent was added and mixed thoroughly using a vortix. Chloroform: isoamyl alcohol (300 µL) was added and mixed thoroughly using vortix. The tubes were centrifuged at 13,000 g for 15 min at 4 ºC. The supernatant was transferred to 1.5 mL tubes and the chloroform: isoamyl alcohol step was repeated. The aqueous phase was transferred to 1.5 mL tubes and 0.6 volume isopropanol added. It was mixed thoroughly and kept for RNA precipitation for 1 h at room temperature. Total RNA was pelleted out by centrifugation at 13,000 g for 15 min at 4 ºC. The RNA pellet was washed with 70% ethanol twice and dried in a Speedvac centrifugal concentrator. RNA pellet was dissolved in 40 µL of DEPC treated water and stored at -80 ºC in aliquots. Purity of RNA was confirmed by measuring OD at 260/280 nm and also by visualization on 1.0% TAE agarose gel.

2.8.6.1 mRNA purification

Total RNA was quantified spectrophotometrically. The amount of RNA was in the range of 1to 3 mg. Appropriate amount of OBB buffer and Oligotex suspension were added according to manufacturer's instruction. The sample was incubated for 3 min at 70 ºC in a heating block. Sample was removed from the heating block, and placed at 20 – 30 ºC for 10 min. Oligotex: mRNA complex was pelleted down by centrifugation for 2 min at maximum speed (14,000–18,000 g) and carefully removed the supernatant by pipetting. Oligotex: mRNA pellet was resuspended in the appropriate amount of OW2 buffer by vortexing and pipetted onto a small spin column placed in a 1.5 mL microcentrifuge tube and centrifuged for 1 min at maximum speed. Spin column was transferred to a new RNase-free 1.5 mL micro-centrifuge tube, appropriate amount of OW2 buffer added to the column and again centrifuged for 1 min at 12,000 g, flow-through was discarded. Spin column was transferred to a new RNase free 1.5 mL micro-centrifuge tube. Appropriate amount of hot (70 ºC) OEB buffer was pipetted onto the column and resuspended by pipetting up and down; spin it down for 1 min at 12,000 g. The eluted mRNA sample was quantified spectrophotometrically and used for further downstream processes.

2.8.6.2 Spectrophotometric determination of nucleic acids concentration

DNA concentration was determined by measurement of the absorption at 260 nm. Lambda 650 Perkin Elmer Photometer and Nanovue (GE) was used to determine the concentration of 1:50 diluted RNA or DNA samples in a volume of 1mL in a 1 cm light path quartz cuvette. Absorbance value (A_{260}) should fall between 0.1-1.0 to be accurate. Sample dilution was adjusted accordingly. An absorbance of 1.0 at 260 nm corresponds to 50 µg DNA/mL.

2.8.6.3 cDNA first strand synthesis by reverse transcription

Complementary DNA (cDNA) was synthesized from a mature mRNA template in a reaction catalyzed by the enzyme reverse transcriptase. The resulting molecule is a DNARNA hybrid and the process is called as cDNA 1st strand synthesis. For DNA double strand synthesis this hybrid molecule is digested with RNase H (specific for degrading RNA strand in a DNA-RNA hybrid), DNA second strand is synthesized using DNA polymerase I (Kimmel and

Berger, 1987). In the present study cDNA first strand was synthesized using ImProm-II™ Reverse Transcription System (Promega, USA). The reactions were set up as per the manufacturer's guidelines. Briefly, reverse transcription reactions of up to 1 µg of total RNA performed in 20 µL reactions comprised of components of the ImProm-II™ Reverse Transcription System. Experimental RNA was combined with the oligo $(dT)_{15}$ primer. The primer/template mixture was isothermally denatured at 70 ºC for 5 min and snap chilled on ice. A reverse transcription reaction mix was assembled on ice to contain nuclease-free water, reaction buffer, reverse transcriptase, magnesium chloride, dNTPs and ribonuclease inhibitor, RNasin®. As a final step, the template-primer combination was added to the reaction mix on ice. Following an initial annealing at 25 ^{o}C for 5 min, the reaction was incubated at 42 ºC for up to 1-1.5 h. The cDNA synthesized was directly added to amplification reactions. The first strand reaction was set up as follows:

Reagent and concentration	**Volume**
Experimental RNA (1 µg)	1.0 µL
Primer (Oligo$(dT)_{15}$ or Random (10 pmol))	1.0 µL
DEPC treated Water	3.0 µL
Final volume	5.0 µL

The tubes were incubated at 70 ºC for 5 min and then chilled in ice-water for 5 min. Tubes were briefly spun in a micro centrifuge to collect the condensate and maintain the original volume. The tubes were kept closed on ice until addition of the reverse transcription reaction mix. The reverse transcription reaction mix was prepared by adding the following components of the ImProm-II™ Reverse Transcription System in a sterile 1.5 mL micro centrifuge tube on ice.

Reagent and concentration	**Volume**
ImProm-II. 5X Reaction Buffer	4.0 µL

$MgCl_2$ (15 mM)	2.0 µL
dNTP Mix (7.5 mM)	1.0 µL
RNasin® Ribonuclease Inhibitor (40 U/ µL)	0.5 µL
ImProm-II Reverse Transcriptase	1.0 µL
Nuclease-free water	6.5 µL
Final volume	15.0 µL

An aliquot of 1.0 µg total RNA and oligo $(dT)_{15}$ or Random hexamer primer (10 pmol) mix was added to the above reaction for a final reaction volume of 20 µL per tube. The reaction mixure was incubated at 25 ºC for 5 min for primer annealing and then at 42 ºC for 1 h for cDNA first strand synthesis. Reverse transcriptase was thermally inactivated by incubation at 70 ºC for 15 min prior to proceeding with PCR amplification (Chumakov, 1994).

2.8.7 Polymerase Chain Reaction (PCR)

PCR is a powerful technique to amplify a desired nucleotide sequence using sequence specific primers. This amplification may be either of and from a single template or of a template from a mixture of templates (Saiki *et al*., 1985; Mullis and Faloona, 1987; Saiki *et al*., 1988; Mullis, 1990; Arnheim and Erlich, 1992). This technique has been successfully used for various purposes like fishing out of gene(s) from genomic DNA or from cDNA population (Todd *et al*., 1987), introducing restriction sites of interest in the amplified product for directional cloning (Scharf *et al*., 1986), creating sequence mismatch/ deletion/ addition resulting in mutant version of a gene or nucleotide sequence (Goodenow *et al*., 1989), differentiating between two alleles (Kwok *et al*., 1990) etc. In the present study applications of PCR were exploited for a few of the above specified applications. The PCR reaction mixture and cycling conditions used were as follows:

Reaction mixture

Reagent and concentration	Volume
Sterile deionized water	10.7 µL
Template (100 ng/ µL)	1.0 µL
Forward primer (10 pmol)	1.0 µL
Reverse primer (10 pmol)	1.0 µL
dNTPs (0.2 mM)	4.0 µL
10 x Buffer (Mg^{2+} 1.5 mM)	2.0 µL
Taq DNA Polymerase (1 U/µL)	0.3 µL
Total volume	20.0 µL

2.8.7.1 PCR cycle conditions

95 ºC - 5 min

95 ºC - 1 min

45-65 ºC - 45-90 s } 35 cycle

72 ºC - 1.5 min

Annealing temperature (Ta) was dependent on melting temperature of the primer (Tm); Ta=Tm -5 º C

1 cycle 72 ºC 10 min

1 cycle 4 ºC hold.

2.9 Rapid amplification of cDNA ends (RACE, Invitrogen)

Generally, using reverse transcription PCR, either partial cDNA fragments (both 5' and 3' ends missing) or cDNA with full 5'end missing are amplified from total cDNA. If a partial cDNA sequence is known, unknown sequences to the 5' and 3' of the known sequence can be reverse transcribed from RNA, amplified by PCR using RACE (Frohmman *et al*., 1988).

RACE ready cDNA Kit (Invitrogen, USA) was used to perform RACE. The reactions were set up as per the manufacturer's guidelines with minor modification. Briefly, the RACE technique is based on oligo-capping and RNA ligase mediated (RLM) RACE method (Maruyama and Sugano, 1994; Volloch *et al*., 1994). The GeneRacer method involves selectively ligating an RNA oligo nucleotide (GeneRacer RNA Oligo) to the full-length 5' ends of decapped mRNA using T4 RNA ligase. Application of GeneRacer race ready cDNA Kit is as follows:

- Identifying the 5' and 3' untranslated regions of genes.
- Studying heterogeneous transcriptional start sites.
- Characterizing promoter regions.
- Obtaining the complete cDNA sequence of a gene.
- Amplifying the full length cDNA for downstream cloning and expression.

GeneRacer RACE Ready cDNA Synthesis

The pre-requisite to begin the RACE was to have the good quality RNA or mRNA. 1-5 µg total RNA or 50-250 ng purified mRNA in DEPC water was used for the RACE reaction**.** Dephosphorylation reaction of 10 µL was set in 1.5 mL sterile micro centrifuge tube. The reaction mix is as follows:

Reagent and concentration	**Volume**
RNA	7.0 µL
10X CIAP buffer	1.0 µL
RNaseOut™ (40 U/µL)	1.0 µL
CIP (10 U/µL)	1.0 µL
Total Volume	10.0 µL

The reaction mixture was mixed gently by pipetting, followed by brief spin to collect the liquid at the bottom and incubated at 50 ºC for 1 h. The reaction volume was made up to 100 µL by adding DEPC water and precipitated by adding 2 µL of 10 mg/mL glycogen, 10 µL of 3 M sodium acetate (pH 5.2) and 220 µL of 95% ethanol. The precipitated sample was washed with 70% ethanol and pellet was air-dried. Air-dried pellet was dissolved in 7 µL of DEPC water. 5' mRNA cap structure from full- length mRNA was removed by following reaction.

Reagent and concentration	**Volume**
Dephosphorylated RNA	7.0 µL
10X TAP buffer	1.0 µL
RNaseOut™ (40 U/µL)	1.0 µL
TAP (0.5 U/µL)	1.0 µL
Total Volume	10.0 µL

The reaction mixture was mixed gently by pipetting, followed by brief spin to collect the liquid at the bottom and incubated at 37 ºC for 1 h. The reaction volume was made up to 100 µL by adding DEPC water and precipitated by adding 2 µL of 10 mg/mL glycogen, 10 µL of 3 M sodium acetate (pH 5.2) and 220 µL 95% ethanol. The precipitated sample was washed with 70% ethanol and pellet was air-dried. Air dried pellet was dissolved in 7µL of DEPC water. After decapping of full length mRNA, the RNA Oligo was ligated to the 5' end of the mRNA. The reaction of ligating the RNA Oligo to decapped mRNA is as follows:

Reagent and concentration	**Volume**
Decapped RNA	7.0 µL
10 X Ligase buffer	1.0 µL

10 mM ATP	1.0 µL
RNaseOut™ (40 U/ µL)	1.0 µL
T4 RNA ligase (5 U/µL)	1.0 µL
Total Volume	11.0 µL

7 µL of dephosphorylated, decapped RNA was added to the tube containing the prealiquoted, lyophilized GeneRacer™ RNA Oligo (0.25 µg). RNA Oligo was resuspended by pipetting up and down several times. Mixture was incubated at 65 ºC for 5 minutes and kept on ice, reaction mixture was mixed gently by pipetting, followed by brief spin to collect the liquid at the bottom and incubated at 37 ºC for 1 h. The reaction volume was made up to 100 µL by adding DEPC water and precipitated by adding 2 µL of 10 mg /mL glycogen, 10 µL of 3 M sodium acetate (pH 5.2) and 220 µL 95% ethanol. The precipitated sample was washed with 70% ethanol and pellet was air dried. Air dried pellet was dissolved in 10 µL of DEPC water. The decapped full length mRNA ligated with GeneRacer™ RNA - Oligo was used to prepare cDNA by reverse transcription. 1 µL of the desired primer and 1 µL of dNTP mix (25 mM each) were added to the ligated RNA and incubated at 65 ºC for 5 min. Following reagents were added to the 11 µL ligated RNA and primer.

Reagent and concentration	**Volume**
5 X RT buffer	4.0 µL
SuperScript™ III RT (200 U/µL)	1.0 µL
0.1 M DTT	1.0 µL
RNaseOut™ (40 U/ µL)	1.0 µL
Sterile water	2.0 µL
Total Volume	20.0 µL

The reaction mixture was mixed gently by pipetting, followed by brief spin to collect the liquid at the bottom and incubated at 45 ºC for 1 h. The reaction was inactivated by incubating the reaction mixture at 70 ºC for 15 min. 1 µl of RNase H (2 U) was added to reaction mixture to chew up the DNA / RNA hybrid by incubation at 37 ºC for 20 min. Centrifuged the sample and kept at -20 ºC for further use in PCR.

PCR reaction setup

Reactions were setup to amplify either the 5′ end or the 3′ end of gene of interest. The reaction was set up as given below:

Reagent and concentration	**5′ RACE**	**3′ RACE**
GeneRacer 5′ Primer 10 µM	3.0 µL	-
Reverse GSP 10 µM	1.0 µL	-
GeneRacer™ 3′ Primer 10 µM	-	3.0 µL
Forward GSP 10 µM	-	1.0 µL
RT Template	1.0 µL	1.0 µL
10X High Fidelity PCR buffer	5.0 µL	5.0 µL
dNTP Solution (10 mM each)	1.0 µL	1.0 µL
Platinum® Taq DNA Polymerase or High Fidelity, 5 U/ µL	0.5 µL	0.5 µL
$MgSO_4$, 50 mM	2.0 µL	2.0 µL
Sterile Water	36.5 µL	36.5 µL
Total Volume	50.0 µL	50.0 µL

Cycling conditions

Temperature	Time	Cycles
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
50 ºC	0.45 min	
72 ºC / 1 kb	1.30 min	
72 ºC / 1 kb	5 min	1 cycle

Nested PCR

Nested PCR was done to increase the specificity and sensitivity of RACE products for the 5′ or 3′ ends of gene. 0.5 µL of the original amplification reaction was used as template for nested PCR. Reactions were set up as described below:

Reagent	5′ RACE	3′ RACE
Gene Racer 5′Nested Primer10 µM	1.0 µL	-
Reverse Nested GSP 10 µM	1.0 µL	-
Gene Racer 3′ Nested Primer 10 µM	-	1.0 µL
Forward Nested GSP 10 µM	-	1.0 µL
Primary PCR Product	0.5 µL	0.5 µL
10 X High Fidelity PCR Buffer	2.0 µL	2.0 µL
dNTP Solution (10 mM each)	1.0 µL	1.0 µL

Platinum® Taq DNA Polymerase High Fidelity, 5 U/µL	0.5 µL	0.5 µL
$MgSO_4$, 50 mM	0.5 µL	0.5 µL
Sterile Water	13.5 µL	13.5 µL
Total Volume	20.0 µL	20.0 µL

Following cycling condition was used for the nested PCR reactions.

Temperature	**Time**	**Cycles**
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
55-58 ºC	0.45 min	
72 ºC / 1 kb	1.30 min	
72 ºC / 1 kb	5 min	1 cycle

Nested PCR product was analyzed on a 1% agarose/ethidium bromide gel. Nested PCR products were shorter by the number of bases between the original primers and the nested primers. The band was excised, cloned and sequenced.

2.10 Rapid Amplification of cDNA Ends (B.D.Clontech SMART RACE KIT)

In the present study, to isolate full length gene SMART RACE cDNA Amplification Kit (BD Biosciences, Clontech, USA) was used. The reactions were set up as per the manufacturer's guidelines.

Briefly, SMART technology provides a mechanism for generating full length cDNA's in reverse transcription reactions. This is done by the joint action of the SMART II™ A Oligonucleotide and the PowerScript™ Reverse Transcriptase (RT). PowerScript RT is a variant of MMLV RT, which upon reaching the end of a RNA template exhibits terminal transferase activity by adding 3–5 residues (predominantly dC) to the 3' end of the first strand cDNA. The SMART oligo contains a terminal stretch of G residues that anneal to the dC-rich cDNA tail and serves as an extended template for RT. PowerScript RT switches templates from the mRNA molecule to the SMART oligo, generating a complete cDNA copy of the original RNA with the additional SMART sequence at the end. Following reverse transcription, the first strand cDNA is used directly in 5'- and 3'-RACE PCR reactions. The only requirement for SMART RACE cDNA amplification is 22–28 nucleotides of sequence information in order to design gene specific primers (GSPs) for the 5'- and 3'-RACE reactions.

Using SMART RACE Kit two separate cDNA populations, 5'-RACE cDNA and 3'-RACE cDNA are synthesized. The cDNA for 5'-RACE is synthesized using a modified lock-docking oligo (dT) primer and the SMART II A oligo as described above. The modified oligo (dT) primer termed the 5'-RACE CDS Primer A (5'-CDS), has two degenerate nucleotide positions at the 3' end. These nucleotides position the primer at the start of the A+ tail and thus eliminate the 3' heterogeneity inherent with conventional oligo (dT) priming. Once, RACE cDNAs are prepared, 5'- and 3'-RACE can be performed using gene specific primers. All PCR reactions in the SMART RACE protocol are carried out using the Advantage® 2 Polymerase Mix. The Polymerase Mix is comprised of TITANIUM™ Taq DNA Polymerase - a nuclease-deficient N-terminal deletion of Taq DNA polymerase plus TaqStart® Antibody to provide automatic hot-start PCR and a minor amount of a proofreading polymerase.

The reactions were set up for 5’ and 3’ RACE cDNA as below:

RACE cDNA preparation

For preparation of 5' RACE cDNA

RNA sample (1 µg/µL)	5 µL
5' CDS primer A (12 µM)	1 µL

SMART II A oligo (12 µM)	1 µL
SMQ	3 µL

For preparation of 3' RACE cDNA

RNA sample (1 µg/µL)	5 µL
3'-CDS primer A (12 µM)	1 µL
SMQ	4 µL

Sterile H_2O was added to a final volume of 10 µL for each of the above reaction. Contents were mixed and the tubes centrifuged briefly. The tubes were incubated at 70 °C for 2 min and cooled on ice for 2 min. The tubes were briefly centrifuged and to each reaction tubes following reagents were added:

5X First-Strand Buffer	4 µL
DTT (20 mM)	1 µL
dNTP Mix (10 mM)	1 µL
$MgCl_2$ (25 mM)	3 µL
PowerScript Reverse Transcriptase (5U/ µL)	1 µL
Total volume	10 µL

The contents of the tube were mixed by gentle pipetting and were briefly centrifuged. The tubes were incubated at 42 ºC for 1.5 h in a hot-lid thermal cycler. The first strand reaction mixture was diluted to 100 µL with Tricine- EDTA buffer (provided with the kit) and heated at 70 ºC for 7 min. The diluted first strands were used for 5’ and 3’ RACE. The following master mix was prepared according to the number of PCR reactions to be set up:

Primary and Nested PCR

Master Mix for RACE PCR reaction

PCR-grade Water	34.5 µL
10X Advantage 2 PCR Buffer	5.0 µL
dNTP Mix (10 mM)	1.0 µL
50X Advantage 2 Polymerase Mix	1.0 µL
Total volume	41.5 µL

The above master mix was used for 5' and 3' RACE PCR.

The reaction was set up as follows, for 5'RACE

5' RACE cDNA	2.5 µL
UPM (10X)	5.0 µL
GSP R (10 µM)	1.0 µL
Master mix	41.5 µL
Total volume	50.0 µL

The reaction was set up as follows, for 3'RACE

3' RACE cDNA	2.5 µL
UPM (10X)	5.0 µL
GSP F (10 µM)	1.0 µL
Master mix	41.5 µL

Total volume	50.0 µL

Following PCR cycling conditions were used for amplifying 5’ and 3’ RACE products

No. of Cycles	Temperature	Time
5 cycles	94 °C 72 °C	30 s 3 min
5 cycles	94 °C 70 °C 72 °C	30 s 30 s 1.5 min
30 cycles	94 °C 65 °C 72 °C	30 s 30 s 1.5 min

Nested PCR reaction mix

Nested 5’ RACE and 3’ RACE were performed as separate reactions. PCR-Grade Water	41 µL
10X Advantage 2 PCR Buffer	5.0 µL
dNTP Mix (10 mM)	1.0 µL
50X Advantage 2 Polymerase Mix	1.0 µL
Forward primer (for 3’ RACE)	1 µL
Reverse primer (for 5’ RACE)	1 µL
NUP Primer	1.0 µL
5’ and 3’ Primary reaction	2.0 µL
Total volume	50 µL

Following is the list of primers (refer Table), which were used for the RACE provided with kit.

List of Primers used for the RACE.

Name	Sequence 5'-3'
SMART II™ A Oligonucleotide	AAGCAGTGGTATCAACGCAGAGTACGCGGG
3'-RACE CDS Primer A	AAGCAGTGGTATCAACGCAGAGTAC $(T)_{30}$V N
5'-RACE CDS Primer A (5'-CDS)	$(T)_{25}$V N
10X universal Primer A Mix (UPM)	Long: CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT Short: CTAATACGACTCACTATAGGGC
Nested universal Primer (NUP)	AAGCAGTGGTATCAACGCAGAGT

2.11 Quantitative real time PCR (QRT PCR)

Quantitative Real Time PCR is a powerful tool for gene expression analysis and was first demonstrated by Higuchi *et al*., (1992, 1993). QPCR quantitates the initial amount of the template most specifically, sensitively and reproducibly, and is a preferred alternative to other forms of quantitative reverse transcriptase PCR that detect the amount of final amplified product at the end-point (Freeman *et al*., 1999; Raeymaekers, 2000). QPCR monitors the fluorescence emitted during the reaction as an indicator of amplicon production during each PCR cycle (i.e., in real time) as opposed to the end point detection (Higuchi *et al*., 1992, 1993). The QPCR system is based on the detection and quantitation of a fluorescent reporter (Lee *et al*., 1993; Livak *et al*., 1995). There are three main fluorescence-monitoring systems for DNA amplification (Wittwer *et al*., 1997a): (1) hydrolysis probes (2) hybridizing probes and (3) DNA-binding agents (Wittwer *et al*., 1997b; Vander Velden *et al*., 2003). Most commonly used are the hydrolysis probes, which include TaqMan probes (Heid *et al*., 1996), molecular beacons (Tan *et al*., 2004; Vet and Marras, 2005) and scorpions (Saha *et al*., 2001; Terry *et al*., 2002).

They use the 5' exonuclease activity of *Taq* polymerase to measure the amount of target sequences in cDNA samples. The use of fluorescent probe technologies reduces the risk of sample contamination while maintaining convenience, speed and high throughput screening capabilities (Reischl *et al*., 2002). In the present study the Brilliant® II QPCR Master Mix (Stratagene, USA) was used. This kit supports quantitative amplification and detection with multiplex capability, and shows consistent high performance with various fluorescent detection systems, including molecular beacons and TaqMan® probes. The kit supports PCR amplifications and detection of a variety of DNA targets, including genomic DNA, plasmid DNA and cDNA. The Brilliant II QPCR master mix includes SureStart® *Taq* DNA polymerase, a modified version of Taq2000™ DNA polymerase with hot start capability. A passive reference dye (an optional reaction component) is provided so as to control the final dye concentration, increasing the flexibility of the reagents for use with different platforms.

2.11.1 QRT-PCR considerations

The optimal concentration of the upstream and downstream PCR primers is determined empirically. The optimal concentration is the lowest concentration that results in the lowest Ct and an adequate fluorescence for a given target concentration. The best concentrations of the upstream and downstream primers are not always of equal molarity. In this study, 100 nM was considered optimum. Reaction was standardized in such a way that there was no primer dimer formation. Acquisition of real-time data generated by SYBR Green was done as recommended by the instrument manufacturer. Data collection was either at the annealing step (3- step cycling protocol) or extension step of each cycle.

Magnesium chloride concentration in the PCR reaction mix affects the specificity of the PCR primers and probe hybridization. The SYBR Green Brilliant® II QPCR Master Mix kit contains $MgCl_2$ at a concentration of 5.5 mM (in the 1X solution), which is suitable for most targets. A passive reference dye may be added to compensate for non-PCR related variations in fluorescence. Fluorescence from the passive reference dye does not change during the course of the PCR reaction but provides a stable baseline to which samples are normalized. In this way, the reference dye compensates for changes in fluorescence between wells caused by slight volume differences in reaction tubes. The excitation and emission wavelengths of the

reference dye are 584 nm and 612 nm respectively. The SYBR Green Brilliant® II QPCR Master Mix kit contains reference dye which is suitable for most targets.

2.11.2 Preparing the reactions

Real time PCR model Max 3000P (Stratagene, USA) was used in the present study. The experimental reactions were prepared by adding the following components in order.

Reaction Mixture

Nuclease-free PCR-grade H_2O to adjust the final volume to 25 µL (including experimental DNA)

Reagent	Volume
2X master mix	12.5 µL
Upstream primer (optimized concentration)	1.0 µL
Downstream primer (optimized concentration)	1.0 µL
Total	25.0 µL

The reactions were mixed without creating bubbles and 1 µL of experimental genomic DNA, cDNA or plasmid DNA was added to each experimental reaction. Reaction was mixed gently and briefly spun.

2.11.3 PCR cycling programs

The reactions were placed in the instrument and a 2 step program was run (recommended for TaqMan reactions).

Two step cycling Protocol

Temperature	Time	No. of cycles
95 ºC	10 min	1
95 ºC	30 s	40
55- 57 ºC	30 s	
72 ºC	1 min	

Initial 10 min incubation was to activate the DNA polymerase completely. The temperature cycler was set to detect and report fluorescence during the annealing/extension step of each cycle.

2.12 Expression and purification of recombinant Protein

Heterologous expression and activity of glycosyltransferase (*GT*) gene was attempted in *E. coli* (BL21) host cell. The transformants were grown at 37 ºC in Luria-Bertani medium containing kanamycin (50 µg/mL). A single isolated bacterial colony from freshly streaked plates (grown on LB agar medium containing 50 µg/mL kanamycin) was used to inoculate 5 mL liquid LB medium containing the same concentrations of the antibiotic. Culture was grown overnight with shaking at 200 rpm at 37 °C. One mL aliquot of each culture was used to inoculate 100 mL liquid LB containing 50 µg/mL kanamycin. Once the cultures reached A_{600} 0.6 - 0.8, recombinant protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG), and the culture was grown for 4 to 6 h at 37 °C with shaking at 150 rpm. Cells were harvested by centrifugation at 6,000 rpm for 10 min at 4 °C. Pellets were resuspended in 6.25 mL lysis buffer. Cells were disrupted by sonication for 5 mins at 70 amplitude on an ultrasonic liquid processor, XL 2000 model (MESONIX). $MgSO_4$, final concentration of 10 mM and lysozyme final concentration 100 µg/mL was added to the disrupted cells and kept at 37 ºC for ½ h. It was centrifuged at 10,000 rpm for 10 minute and supernatant was saved as lysate and pellet was resuspended in 2 mL sonication buffer.

Suspension was again sonicated for 1 min at 70 amplitude to disrupt the inclusion bodies and the disrupted inclusion bodies were dissolved in 3 mL of dispersion buffer and an aliquot of 20 μL checked on SDS PAGE (Chapter 2: section 2.6.5 and Table: 2.7) to check heterologous expression of *GT* genes.

2.12.1 Protein isolation from inclusion body

Cells were isolated by centrifugation and resuspended in 100 mM Tris-HCl buffer, pH 8, containing 2 mM EDTA, 20% glycerol, 1.5 mM DTT, 1 mM PMSF and 0.4% Triton. Lysozyme (0.5 mg/mL) was added, and the suspension was incubated at 37 ºC. Bacterial cells were further lysed by sonication on ice with an ultrasonic liquid processor, XL 2000 model (MESONIX) for 5 min (5 seconds on and 5 sec off) at 70% amplitude. Cell debris was removed by centrifugation and the pellet was resuspended in sonication buffer which contain 100 mM Tris (pH 8.0) and 50 mM glycine. Cells were sonicated once again for 5-6 pulses 10 s on and 20 second cooling.

2.12.2 Affinity purification of recombinant protein Using Ni+ NTA beads

The recombinant protein, among several other bacterial proteins is loaded on affinity matrix column such as Ni-agarose. This affinity matrix contains bound metal ion nickel, to which the polyhistidine-tag binds with micromolar affinity. The matrix is then washed with buffer to remove unbound proteins. This can be achieved either by increasing the imidazole concentration in buffer or by lowering the pH of the washing and elution buffer. The column(s) and buffers were equilibrated to room temperature. The Ni^{+} beads column was equilibrated with binding buffer for 30 min and then three bed volume of binding buffer was passed through the column followed by addition of soluble protein in dispersion buffer. The column was kept at 4 ºC for 1 h for binding of recombinant protein to Ni^{+} beads. Flow through was collected in different tubes after 1 h and column was washed with two bed volume of washing buffer (chapter 2: section 2.6.7). The washing efficiency may be improved by the addition of 20 mM imidazole and histidine-tagged proteins are then usually eluted with 250 mM imidazole. (Chapter 2: section 2.6.7 and Tab 2.8). The 6x His-tagged protein was eluted in 4 aliquots of elution buffer, 1 mL each. Protein elution was monitored by measuring

the absorbance at 280 nm of collected fractions. The eluted protein was separated and analyzed by SDS-PAGE.

2.12.3 Polyacrylamide gel electrophoresis (PAGE)

PAGE system is the widely used electrophoresis system for protein separations (Laemmeli, 1970). The resolution in a Laemmeli gel is excellent because the treated peptides are stacked in a stacking gel before entering the separating gel.

2.12.4 Preparation of the separating gel

A vertical slab gel (Hoeffer Scientific, U.S.A.) was assembled using 1.0 mm spacers. In a side armed vacuum flask, 10% separating gel solution was made according to Table 2.7, except for the addition of ammonium persulfate and TEMED. The solution was degassed by applying vacuum while stirring on a magnetic stirrer. TEMED and ammonium persulfate were added, and mixed gently without generating bubbles. The solution was pipetted into the gel cassette leaving 1.5 cm from the top unfilled. The gel solution was overlaid with water saturated n-butanol to remove trapped air bubbles and to layer evenly across the entire surface. When a sharp liquid-gel inter-surface was observed after the gel polymerization, the slab was tilted to pour off the overlay.

2.12.5 Preparation of the stacking gel

Stacking gel solution was prepared according to Table 2.7, excluding ammonium per sulfate and TEMED. As in the separating gel, this solution was degassed. TEMED and ammonium per sulfate were added, mixed and overlaid on the separating gel. A comb was inserted taking care not to trap air bubbles beneath the comb teeth. The gel was left to polymerize.

2.12.6 Preparation of the sample

Equal parts of the protein sample and the loading buffer were mixed in a microcentrifuge tube and kept in a boiling water bath for 3-5 min for SDS-PAGE. Then the samples were centrifuged at 10,000 g for 10 min, supernatant collected and used for electrophoresis. Gel was run at room temperature at constant ampere 20 mA or 80 V. The protein samples for

native PAGE were mixed with native dye and gel was run at 80 V at low temperature ranging from 4-10 ºC.

2.12.7 Loading and running the polyacrylamide gel

Once wells have been formed in the gel, flushed it with distilled water and drained off completely. The wells were filled with tank buffer and the samples under laid using a syringe. Lower and upper buffer chambers were filled with tank buffer. Voltage was set between 70-80 units. The run was stopped when the dye reached the bottom of the polyacrylamide gel.

2.12.8 Coomassie Blue staining of the gel

Coomassie blue staining solution: 45 mL Methanol, 10 mL acetic acid, 45 mL de-ionized water and 0.25% Coomassie blue (R 250). After running the gel it was transferred directly to a tray containing Coomassie blue staining solution and was kept for three to four hours at room temperature on rocker. Staining solution was poured off and de-staining solution (Table 2.7) was poured in. De-staining step was repeated two-three times till clear bands appeared.

2.12.9 Silver staining of the gel

Protein separated on SDS or native PAGE, was transferred to the fixer solution for 1 h. This was followed by 2 x 20 min washes in 50% ethanol. The gel was transferred to sodium thiosulfate solution for 1 min and rinsed thrice with deionized H_2O (20 s each). The gel was silver stained (HCHO was added to the solution just before use) for 20 min with intermittent shaking in dark. The gel was then rinsed thrice with deionized H_2O (20s each) and transferred to the developer till the bands developed. The gel was washed with deionized H_2O and stored in fixer.

2.13: 2D gel electrophoresis

2-D electrophoresis is a powerful and widely used method for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples. This technique separates proteins according to two independent properties in two discrete steps.

The first-dimension step, isoelectric focusing (IEF), separates proteins according to their isoelectric points (pI); the second-dimension step, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their molecular weights (M, relative molecular mass). Each spot on the resulting two-dimensional gel potentially corresponds to a single protein species in the sample. Thousands of different proteins can thus be separated, and information such as the protein pI, the apparent molecular weight, and the amount of each protein can be obtained.

2.13.1 Rehydration of Immobiline Dry Strips

1. Pipette the appropriate volume of prepared DeStreak Rehydration Solution into the Reswelling Tray or into the regular Strip Holder. Distribute the solution evenly over the channel length.
2. Carefully remove the cover foil from the Immobiline DryStrip, starting from the anodic end (+ end).
3. Carefully place the Immobiline DryStrip into the tray/holder channel, gel-side down. Take care to distribute the rehydration solution evenly under the strip. To help coat the entire gel, gently lift and lower the strip and slide it back and forth along the surface of the solution. Be careful not to trap bubbles under the Immobiline DryStrip gel.
4. Load the sample in a rehydration solution evenly on the reswelling tray.
5. Overlay the strip with Immobiline DryStrip Cover Fluid.
6. Rehydrate for 10–15 h.

2.13.2 Isoelectric Focusing

The IPGphor is connected to an external computer via the serial port to control and monitor the electrical conditions. This allows to judge from the shape of the graphs, whether the separation will give good or bad 2-D results.

- Level the IPGphor chamber, must be horizontal on the bench.
- Be sure, that the manifold is carefully cleaned and dried.
- Never use new strip holders without cleaning them before the first run.
- Place the Manifold on the cooled electrode contact areas of the power supply.
- Starting at the basic side, place the IPG strip – gel side facing up – into the strip holder with the acidic end towards the anode side. Be sure that the protruding film at the basic end touches the end of the groove.
- *Check:* Now the number printed on the strip must be mirror-converted, the gel surface has to be turned upside.
- Align protrusions along the grooves inside the manifold, align the rehydrated IPG strips, keeping them straight and centered when placed inside the manifold.
- Soak electrode pads with Milli-Q water.
- Blot them on filter paper and place them on top of the ends of the strip.
- The pads should sit completely on the gel surface. If longer pads are required for removal of salt, there must be an overlapping of at least 5 mm. The pads must be damp, not wet.
- The electrode assembly has electrode teeth on one side and hold-down teeth (for paperbridge-loading) on the other side. It is important to choose the correct orientation, to get contact with the electrode pads.
- Place the electrode assemblies on the pads. Secure them on place with the cams.
- Apply the loading cups at correct side of the strip. Press them down with the "sample cup insertion tool" to prevent leakage. Mostly anodal sample application is employed.

Note: The cup can straddle on the alignment protrusions, if necessary pour 100 mL Drystrip cover fluid (paraffin oil) over the strips, around the cups. Any leakage would be detected, because oil would flow into a cup.

- Pipette samples into the cups.
- Pipette 20 μL paraffin oil on each sample.
- Close the safety lid.

2.13.3 Set up running conditions

- Enter the running conditions in the computer.
- Start Ettan™ IPGphor3 program.
- Select the instrument connected (usually instrument 1 of four)
- Select pI range, strip length and number of strips. A programmed voltage running curve will show up.

Enter the running parameters as shown in the table below.

Rehydration time			0 hour
Temperature			20º C
Current per strip			75 µA
Strip length			13 cm
pH gradient			3-10
Step 1	Step & hold	300V	3 h
Step 2	gradient	500V	2 h
Step 3	gradient	1,000V	2 h
Step 4	gradient	8,000V	3 h
Step 5	Step & hold	8,000V	0:45 min
Total time (h)			10:45 h

- Save method under a new name.
- Transfer to IPGphor instrument.

2.13.4 Equilibration of the IPG strips

- Remove the electrodes, the loading cups, and the electrode pads from the Manifold.
- Pour the dry strip cover fluid out from the Manifold.
- Add 1 g DTT to 100 mL equilibration buffer, mix thoroughly and pour into the manifold.
- Place the manifold on an orbital shaker for 15 minutes, which is set to 30 rpm.
- After 15 min, pour out the first equilibration buffer.
- Add 2.5 g iodoacetamide to 100 mL equilibration buffer, mix thoroughly and pour into the Manifold.

- Place the Manifold on an orbital shaker for another 15 minutes, which is set to 30rpm.
- After 15 min, pour out the second equilibration buffer.
- Do not leave the strips longer in equilibration buffer, because this would elute a part of the proteins from the strip.

2.13.5 Second-dimension SDS-PAGE using SE 600 Ruby system

1. Prepare separating and stacking gel solutions according to Laemmeli method.
2. Dip the Immobiline DryStrip gel in SDS buffer.
3. While the SDS gels still are in the gel caster, apply the Immobiline DryStrip gels on top of them. Push the strips gently down to the gel surface.
4. Seal the Immobiline DryStrip gel in place with melted agarose.
5. Place the gel in electrophoresis chamber according to manual instructions and run the gel at 25 ºC, 30 mA current.
6. Stop the run when dye reaches the bottom of the polyacrylamide gel.
7. Transfer the gel directly either into Coomassie staining solution or into fixer for silver staining.

2.14 Bradford protein assay

Principle

The assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions stabilize the anionic form of the dye, causing a visible color change. The assay is useful since the extinction coefficient of a dye-albumin complex solution is constant over a 10-fold concentration range.

Reagent required

1. Bradford reagent: Dissolve 10 mg Coomassie Brilliant Blue G-250 in 5 mL 95% ethanol, add 10 mL 85% (w/v) phosphoric acid. Dilute to 100 mL when the dye has completely dissolved, and filter through Whatman #1 paper just before use.

2. (Optional) 1 M NaOH (to be used if samples are not readily soluble in the color reagent).

The Bradford reagent should be a light brown in color. Filtration may have to be repeated to rid the reagent of blue components.

ASSAY

1. Warm up the spectrophotometer 15 min before use.

2. Dilute samples with buffer to an estimated concentration of 20 to 200 micrograms/mL

3. Prepare standards containing a range of 20 to 200 micrograms protein (albumin or gamma globulin are recommended) to a standard volume (generally 1 mL or less).

4. Prepare unknowns to estimated amounts of 20 to 200 micrograms protein per tube, same volume as the unknowns.

5. Add 0.25 mL 1 M NaOH (Optional) to each sample and vortex.

6. Add 5 mL dye reagent and incubate 5 min.

7. Measure the absorbance at 590 nm

ANALYSIS

Prepare a standard curve of absorbance versus micrograms protein (or *vice versa*), and determine amounts from the curve. Determine concentrations of original samples from the amount protein, volume/sample, and dilution factor, if any.

2.15 Raising polyclonal antibody against *GT* in rabbit

The purified recombinant glycosyltransferase protein was used for raising polyclonal antibodies in rabbit (New Zealand White).

2.15.1 Pre-treatment of serum

Rabbit immune serum (stored at -70 ºC), containing polyclonal antibodies against the antigen, was thawed overnight at 4 ºC. The Serum was kept at 55 °C for one hour and the immune serum was centrifuged at 12,000 g for 15 min at 4 ºC. Aliquots according to the requirements were made and stored at -70 ºC. Prior to use again antiserum was thawed overnight at 4 ºC.

2.15.2 Determination of titer of antibodies and ELISA Buffers

1X Phosphate buffer saline (PBS): 8 gm NaCl; 1.44 g Na_2HPO4; 0.24 g KCl; 0.2 g K_2HPO4 in 1 L de-ionized water. Wash Buffer (PBST): 1X PBS, 0.05% Tween 20 and 0.1% BSA. Antibody diluting buffer: 1X PBS and 0.25% BSA. Blocking reagent: 1% BSA in PBS. Substrate diluting buffer: 200 mM Tris-HCl, pH 9.5 and 10 mM $MgCl_2$. All reagents were prepared in sterile milliQ water. ELISA was performed to determine the titre of first, second and third bleed of rabbit serum. Equal quantity of antigen *i.e* 100 ng was coated in triplicates on ELISA plates and kept overnight at 4 °C. Next morning, the plate was washed with 250 µL of PBST, three times for 5 min and 300 µL of blocking reagent was added. The plate was wrapped in aluminium foil and kept at 37 °C for 2 h. ELISA plate was washed again as described earlier and challenged to different dilution of serum such as, 1:10000, 1:20000, 1:30000, 1:40000, 1:60000, 1:80000 and 1:100000. All dilutions were in triplicates and plate was kept at 37 °C for 2 h again. Plate was washed as described earlier with PBST buffer and secondary antibodies (anti goat IgG against rabbit IgG) tagged with alkaline phosphatase was added to a dilution of 1:20000. Plate was incubated at 37 °C for two more hours. After 2 hours of incubation, plate was washed with 250 µL of PBST, three times for 5 min and 100 µL of 1 mg/mL substrate (*p*- Nitro phenyl phosphate) was added and incubated for 45 min. Reaction was stopped by adding 10 mM EDTA. Once the antibody titer was determined then a fixed dilution of antibodies was used for rest of the experiments.

2.16 MALDI MS/MS

For MALDI analysis 10% SDS-PAGE was run for both recombinant glycosyltransferase proteins. Gel was then stained with coomassie Blue. Expected stained protein bands were excised from the gel. The gel pieces were then destained by destaining solution (50% acetonitrile / 50% 50 mM NH_4HCO_3) till colour was gone. Gel was then dehydrated by treating with 100% acetonitrile. After dehydration acetonitrile was completely removed by evaporating briefly in speedvac till noticeably shrunken and white. Then gel pieces were dissolved in 10 mM DTT in 100 mM NH_4HCO_3 and proteins were reduced for 45-60 min. at 56 ^oC. Cooled to room temperature, DTT solution removed and 55 mM iodoacetate in 100 mM NH_4HCO_3 was added. This mixture was vortexed, spun briefly and incubated for 45 min in dark place at room temperature. This was followed by iodoacetamide removal and gel pieces were washed with 100 mM NH_4HCO_3 for 5 min. Again gel pieces were washed twice with 50% acetonitrile / 50% 50 mM NH_4HCO_3) and dehydrate with 100% acetonitrile as mentioned above. Then enough trypsin solution was added to cover the gel pieces (usually around 20 µL) and the gel pieces were rehydrated at 4 ^oC for 30 min in buffer containing 50 mM NH_4HCO_3 and trypsin. Spun briefly and more NH_4HCO_3 was added to cover gel pieces (typically another 25 µL). This was followed by overnight digestion at 37 ^oC. The digested solution (supernatant) was transferred into clean 1.5 mL eppendorf tube. Add 50 % acetonitrile / 2% formic acid solution to the gel pieces which was incubated and vortexed for 20 min. This was spun and sonnicated for 5 mins in a water bath with no heat. Supernatant was then removed and combined with initial digestion solution (supernatant). Extracted digests was then vortexed, evaporated to reduce to 5-10 µL. The remaining 5-10 µL was spun at 14k rpm for at least 10 min to remove any microparticulates. The supernatant was carefully transferred to a fresh 1.5 mL eppendorf tube. The sample was then ready for loading onto MALDI MS/MS. Simultaneously MALDI plate was also washed inorder to remove the particulate matter deposited if any. Then sample and the MALDI matrix (α cyano 4 hydroxy cinnamic acid) in the proportion of 2:1 ratio were loaded on to the MALDI plate. This was followed by reading of MALDI plate with laser energy of about 280 Volts using MALDSNYPT equipment (Waters).

2.17 Glycosyltransferase enzyme assay

The reaction mixture (500 µL) consisted of 5 mM glucose donor (UDP-Glucose), 200 µM substrate, 20 mM Tris-HCl buffer, pH 8.0, and purified enzyme. The reaction was incubated at 30 ^{o}C for 3 hours. After incubation the reaction was extracted thrice by the addition of equal volume of ethyl acetate. The extracted reaction was then completely evaporated to dryness and dissolved in HPLC grade methanol. The substrates and glycosylated products were analyzed by LC-MS and HPLC.

2.18 LC-MS

LC-MS analysis was performed using a Q-TOF Premier mass spectrometer (Waters) outfitted with an electrospray ion source operated in the V-Optics negative mode. A Develosil C18 column was used for LC on a Shimadzu LC-20AD HPLC. For LC-MS analysis solvent system used was 20% water and 80% (v/v) methanol for 10 min.

2.19 HPLC

Substrates and glycosylated products were analyzed by high performance liquid chromatography (HPLC) (Perkin Elmer series 200) on a reversed-phase supelco C18 column (5µm, 25cm x 4.6 mm) and a photodiode array detector (DAD). The mobile phase and UV detection for every glycosylated product was different. The flow rate used in separation was 1mL/min.

2.19.1 Luteolin method

- Run time : 20 min
- Injection volume : 20 µL
- Wave length : Channel A 270 nm & Channel B 350nm
- Mobile phase channel B-Acetonitrile and channel D-Sterile milliQ water.

Step	Time	Flow	A	B	C	D
0	0.5 min	1.00 mL	0.0	10%	0.0	90%
1	5 min	1.00 mL	0.0	10%	0.0	90%
2	5 min	1.00 mL	0.0	30%	0.0	70%
3	5 min	1.00 mL	0.0	70%	0.0	30%
4	5 min	1.00 mL	0.0	95%	0.0	5%

2.19.2 Diadzein method

- Run time : 20 min
- Injection volume : 20 µL
- Wave length : Channel A 250 nm & Channel B 270 nm
- Mobile phase channel B-Acetonitrile and channel D-Sterile milliQ water.

Step	Time	Flow	A	B	C	D
0	0.5 min	1.00 mL	0.0	10%	0.0	90%
1	5 min	1.00 mL	0.0	10%	0.0	90%
2	5 min	1.00 mL	0.0	30%	0.0	70%
3	5 min	1.00 mL	0.0	60%	0.0	40%
4	5 min	1.00 mL	0.0	90%	0.0	10%

2.19.3Naringenin method

- Run time : 20 min
- Injection volume : 20 µL
- Wave length : Channel A 250 nm & Channel B 270 nm
- Mobile phase channel B-Acetonitrile and channel D-Sterile milliQ water

Step	Time	Flow	A	B	C	D
0	0.5 min	1.00 mL	0.0	20%	0.0	80%
1	5 min	1.00 mL	0.0	40%	0.0	60%
2	5 min	1.00 mL	0.0	70%	0.0	30%
3	5 min	1.00 mL	0.0	90%	0.0	10%
4	5 min	1.00 mL	0.0	20%	0.0	80%

2.19.4 Apigenin Method

- Run time : 20 min
- Injection volume : 20 µL
- Wave length : Channel A 320 nm & Channel B 340 nm
- Mobile phase channel B-Acetonitrile and channel D-Sterile milliQ water.

Step	Time	Flow	A	B	C	D
0	0.5min	1.00 mL	0.0	10%	0.0	90%
1	5 min	1.00 mL	0.0	10%	0.0	90%

2	5 min	1.00 mL	0.0	40%	0.0	60%
3	5 min	1.00 mL	0.0	70%	0.0	30%
4	5 min	1.00 mL	0.0	90%	0.0	10%

2.19.5 Genistein Method

- Run time : 15 min
- Injection volume : 20 µL
- Wave length : Channel A 270 nm & Channel B 340 nm
- Mobile phase channel B-Acetonitrile and channel C-50 mM Tris-Hcl.

Step	Time	Flow	A	B	C	D
0	0.5 min	1.00 mL	0.0	10%	90%	0.0
1	10 min	1.00 mL	0.0	40%	60%	0.0
2	5 min	1.00 mL	0.0	60%	40%	0.0

Chapter: 3

Isolation, Cloning and Characterization of glycosyltransferase (genes) from Withania somnifera

3.1 Introduction

Plant derived glucosides have attracted great attention due to their widespread applications. This class of products is difficult to isolate/synthesize in pure form because of the resulting low yields. The isolation and purification of such glucosides from plant sources is tedious and results in low yields. Thus, simple approaches for generation of such glucosides would be highly beneficial. Functional properties, such as solubility, physicochemical stability, bioactivity, pharmacokinetics and cellular localization of natural products, such as flavonoids, are greatly affected by glycosylation.

Glycosylation of plant secondary products, such as flavonoids, coumarins, terpenoids, and cyanohydrins is generally catalyzed by plant secondary product glycosyltransferases, which are family-1 glycosyltransferases.

Glycosylation and glycosyltransferases have been already discussed in detail in Chapter 1 of this thesis. The present Chapter deals about the family 1 member of flavonoid glycosyltransferase and the second glycosyltransferase, which glycosylate sterol compounds called UDP-Glycosyltransferase.

Isolation of a gene is the pre-requisite for its thorough study or characterization. To isolate any gene of interest generally there are two major approaches as follows:-

1. Genomic/cDNA library screening: - It is the classical method of gene isolation in which genomic/cDNA libraries are screened with homologous or heterologous probe. After 2-3 rounds of screenings, the plaques showing strong positive signals are excised and the DNA fragment sequenced.
2. PCR based approach: - It is one of the most popular approaches of gene isolation because of its simplicity and rapidity. Forward and reverse primers are designed on the basis of available sequences in the database. PCR is performed using these primers and genomic/cDNA as template. The amplicon is sequenced to confirm its identity.

In the present study, PCR based approach was followed to fish out the glycosyltransferase gene(s) from *W. somnifera.*

3.2 Materials and methods

3.2.1 Genomic DNA extraction

Genomic DNA was extracted from *W. somnifera* using the method given by Lodhi *et al*., (1994) (Chapter 2, Section 2.8.3.2).

3.2.2 Restriction digestion of plasmid DNA

Plasmid DNA restriction digestion was set up as per manufacturers (Promega, USA; NEB, UK; Amersham, USA) recommendations (Chapter 2, Section 2.8.3.3).

3.2.3 Bacterial strains and plasmids used in the study

Escherichia coli XL-1 blue (Stratagene, USA)

Escherichia coli XL-10 (Novagen, USA)

pGEM-T Easy Cloning vector (Promega, USA)

3.2.4 RNA isolation and cDNA first strand synthesis

Total RNA was isolated from *W. somnifera* leaves (Chapter 2, Section 2.8.6); mRNA was purified from total RNA using oligotex dT resins (Chapter 2, Section 2.8.6.1) and cDNA was synthesized using Reverse Transcription (RT) system (Chapter 2, Section 2.8.6.3).

3.2.5 Polymerase Chain Reaction

PCR amplification was done using cDNA or genomic DNA as template (Chapter 2, Section 2.8.7). Amplified PCR products were eluted from Agarose gel (Chapter 2, Section 2.8.5). The eluted PCR products were used for ligation into PGEM-T easy cloning vector.

3.2.6 Transformation and selection

The ligation mixture was used for transformation of *E. coli* XL1-Blue, host cell line. Putative transformants were selected using Blue-white screening by adding X-gal and IPTG. Colony PCR (Chapter 2, Section 2.8.2.4) was done to screen the putative transformants. Plasmid was isolated from clones showing positive colony PCR results. The plasmid was restriction digested to check for the insert size and the insert was further sequenced to confirm its identity.

3.2.7 Bioinformatic analysis

The glycosyltransferase gene/nucleotide sequences available in the NCBI GeneBank database were aligned and multiple sets of primers were designed from the conserved regions (Primer3 software). Nucleotide and amino acid sequence analysis was done using software pDRAW 32, ClustalX 1.8 and online bioinformatics analysis facility available at www.justbio.com, www.expasy.org and www.ncbi.nlm.nih.gov. Multiple alignments of the amino acid sequences were carried out with the Clustal W1.8 program (http://www.ebi.ac.uk/clustalw/). The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA 4 (Tamura, 2007).

3.2.8 Rapid Amplification of cDNA Ends (RACE)

RACE (Chapter 2, Section 2.9 & 2.10) was done to isolate the full length gene with its 5' and 3' UTRs.

3.2.9 Sequencing

DNA sequencing was performed with the ABI Prism Big-Dye Terminator Cycle Sequencing Kit on the ABI Prism 3730 DNA analyzer (Applied Biosystems) at Genomebio Biotech Pvt. Ltd., Pune, Maharastra, India.

3.2.10 Estimation of glycosyltransferase gene copy number

Genomic DNA was isolated from *W. somnifera* using Lodhi *et al.,* (1994) method. Gene copy number was estimated by Southern hybridization. DNA was digested with three restriction enzymes, *XhoI, EcoRI and HindII*. One of these enzymes, *XhoI* does not cut inside the gene while *EcoRI* has two sites in the gene and *HindIII* has one site in the gene. Southern hybridization was done using an approximately ~400bp fragment (from 5' region of *GT* coding region) as probe at 60 ºC hybridization temperature (Sambrook *et al.*, 1989).

3.3 Results and discussion

3.3.1 PCR based approach for the isolation of glycosyltransferase gene(s).

3.3.1.1 Multiple sequence alignment of glycosyltransferase gene sequences reported in NCBI data base

GT gene nucleotide sequences available at NCBI GenBank database were aligned using Clustal W (1.8) multiple sequence alignment software. Forward and reverse primers were designed from regions showing highest homology. The conserved regions selected for primer designing have been shown high-lighted (yellow) (Figure 3.1).

```
AJ889012        ATGACTACTCACAAAGCTCATTGCTTGATCTTGCCATATCCAGTCCAAGGTCATATCAAC 60
AF190634        ATGACTACTCAAAAAGCTCATTGCTTGATCTTACCATATCCAGCTCAGGGTCATATCAAC 60
                ***********.******************.**********  **.************

AJ889012        CCAATGCTTCAATTCTCCAAACGTTTACGATCCAAACGCGTTAAAATCACTATAGCACTC 120
AF190634        CCTATGCTCCAATTCTCCAAACGTTTGCAATCCAAAGGTGTCAAAATCACTATAGCAGCC 120
                **:***** *****************.*.******* * ** ***************  *

AJ889012        ACAAAATCCTTTTTGAAAAACATGAAGGAATTGCCAACTTCTATGTCAATTGAGGCCATA 180
AF190634        ACCAAATCATTCTTGAAAACCATGCAAGAATTGTCAACTTCTGTGTCAGTCGAGGCTATC 180
                **.*****.** *******.****.*.****** ********.*****.* ***** **.

AJ889012        TCTGATGGCTATGATGATGGTGGTCGCGATCAAGCAGGAACTTTCGTGGCCTATATTACA 240
AF190634        TCCGATGGCTATGATGATGGCGGACGCGAGCAAGCTGGAACCTTTGTGGCCTATATTACA 240
                ** ***************** **:***** *****:***** ** ***************
```

```
AJ889012        CGATTCAAAGAAATTGGTTCGGATACTCTGTCTCAACTTATTCAAAAATTGGCAATTAGT 300
AF190634        AGATTCAAAGAAGTTGGCTCGGATACTTTGTCTCAGCTTATTGGAAAGTTAACAAATTGT 300
                .***********.**** ******** *******.****** .***.**..***:*:**

AJ889012        GGATGTCCTGTAAATTGCATAGTATATGATCCATTCCTCCCTTGGGCTGTTGAAGTTGCA 360
AF190634        GGTTGTCCTGTGAGTTGCATAGTTTACGATCCATTTCTTCCTTGGGCTGTTGAAGTGGGA 360
                **:********.*.*********:** ******** ** ***************** * *

AJ889012        AAACAATTTGGATTAATTAGTGCTGCATTTTTCACACAAAATTGTGTAGTGGATAATCTT 420
AF190634        AATAATTTTGGAGTAGCTACTGCTGCTTTTTTCACTCAATCTTGTGCAGTGGATAACATT 420
                **:.*:****** **. ** ******:********:***:.***** ********* .**

AJ889012        TATTACCATGTACATAAAGGGGTGATAAAACTTCCACCTACTCAAAATGACGAAGAAATA 480
AF190634        TATTACCATGTACATAAAGGGGTTCTAAAACTTCCTCCAACTGACGTTGATAAAGAAATC 480
                ********************** .**********:**:*** *..:*** .*******.

AJ889012        TTAATTCCTGGATTTCCAAATTCGATCGATGCATCAGATGTACCTTCTTTTGTTATTAGT 540
AF190634        TCAATTCCTGGATT---ATTAACAATTGAGGCATCAGATGTACCTAGTTTTGTTTCTAAT 537
                * ************   *::::*.** ** ***************: *******: **.*

AJ889012        CCTGAAGCAGAAAGGATAGTTGAAATGTTAGCAAATCAATTCTCAAATCTTGACAAAGTT 600
AF190634        CCTGAATCTTCAAGAATACTTGAAATGTTGGTGAATCAGTTCTCGAATCTTGAGAACACA 597
                ****** *: .***.*** **********.* .*****.*****.******** **.. :

AJ889012        GATTGTGTTCTAATCAATAGCTTCTATGAGTTGGAGAAAGAGGTAATTGATTGGATGTCG 660
AF190634        GATTGGGTCCTAATCAACAGTTTCTATGAATTGGAGAAAGAAGTAATTGATTGGATGGCC 657
                ***** ** ******** ** ********.***********.*************** *

AJ889012        AAGATATATCCAATAAAGACAATTGGACCAACAATACCATCAATGTACTTAGACAAGAGA 720
AF190634        AAGATCTATCCAATCAAGACAATTGGACCAACTATACCATCAATGTACCTAGACAAGAGG 717
                *****.********.****************:*************** **********.

AJ889012        CTACATGATGATAAAGAGTATGGTCTTAGTATGTTCAAGCCAATGACAAATGAATGCCTA 780
AF190634        CTACCAGATGACAAAGAATATGGCCTTAGTGTCTTCAAGCCAATGACAAATGCATGCCTA 777
                ****.:***** *****.***** ******.* ******************.*******

AJ889012        AATTGGTTAAACCATCAACCAATTAGCTCAGTGTTGTATGTATCATTTGGAAGTTTAGCC 840
AF190634        AACTGGTTAAACCATCAACCAGTTAGCTCAGTAGTATATGTATCATTTGGAAGTTTAGCC 837
                ** ******************.**********. *.************************

AJ889012        AAACTAGGAAGTGAGCAAATGGAAGAATTGGCATGGGGTTTGAAGAATAGCAACAAGAGC 900
AF190634        AAATTAGAAGCAGAGCAAATGGAAGAATTAGCATGGGGTTTGAGTAATAGCAACAAGAAC 897
                *** ***.*. :*****************.*************. *************.*

AJ889012        TTCTTGTGGGTTGTTAGGTCTACTGAAGAGCCCAAACTCCCCAACAACTTTATTGAGGAA 960
AF190634        TTCTTGTGGGTAGTTAGATCCACTGAAGAATCCAAACTTCCCAACAACTTTTTAGAGGAA 957
                ***********:*****.** ********. ******* ************:*:******

AJ889012        TTAACAAGTGAAAAAGGCTTAGTGGTGTCATGGTGTCCACAATTACAAGTGTTGGAACAT 1020
AF190634        TTAGCAAGTGAAAAAGGATTAGTCGTGTCATGGTGTCCACAATTACAAGTCTTGGAACAT 1017
                ***.*************.***** ************************** *********

AJ889012        GAGTCGATAGGGTGTTTTCTGACGCACTGTGGATGGAATTCAACTCTGGAAGCGATTAGT 1080
AF190634        AAATCAATAGGGTGTTTTCTCACGCACTGTGGCTGGAATTCAACTTTGGAAGCAATTAGT 1077
                .*.**.************** ***********.************ *******.******

AJ889012        TTGGGAGTGCCAATGGTGGCAATGCCACAATGGTCTGATCAACCAACAAATGCAAAGCTT 1140
AF190634        TTGGGAGTACCAATGATTGCAATGCCACATTGGTCAGACCAGCCAACAAATGCGAAGCTT 1137
                ********.******.* ***********:*****:** **.***********.******

AJ889012        GTGAAAGATGTTTGGGAAATAGGTGTTAGAGCCAAACAAGATGAAAAAGGGGTAGTTAGA 1200
AF190634        GTGGAAGATGTTTGGGAGATGGGAATTAGACCAAAACAAGATGAAAAAGGATTAGTTAGA 1197
                ***.*************.**.**:.***** *.****************. ********

AJ889012        AGAGAAGTTATAGAAGAATGTATAAAGCTAGTGATGGAAGAAGATAAAGGAAAACTAATT 1260
AF190634        AGAGAAGTTATTGAAGAATGTATTAAGATAGTGATGGAGGAAAAGAAAGGAAAAAAGATT 1257
```

```
                ***********:***********:***.**********.***.* *********.:.***

AJ889012        AGAGAAAATGCAAAGAAATGGAAGGAAATAGCTAGAAATGTTGTGAATGAAGGAGGAAGT 1320
AF190634        AGGGAAAATGCAAAGAAATGGAAGGAATTGGCTAGGAAAGCTGTGGATGAAGGAGGAAGT 1317
                **.************************:*.*****.**:* ****.**************

AJ889012        TCAGATAAAAACATTGAAGAATTTGTTTCCAAGTTGGTTACTATTTCCTAA--------- 1371
AF190634        TCAGATAGAAATATTGAAGAATTTGTTTCCAAGTTGGTGACTATTGCCTCAGTGGAAAGC 1377
                *******.*** ************************* ****** ***.*

AJ889012        ---
AF190634        TAA 1380
```

Figure 3.1 Multiple sequence alignment of glycosyltransferase (GT) nucleotide sequences of Solanaceae member *Lycopersicon esculentum* (AJ889012) and *Nicotiana tabacum* (AF190634). Highlighted are conserved regions considered for primer synthesis.

3.3.1.2 Isolation of total RNA from *W. somnifera* and its cDNA synthesis

RNase free environment was created and maintained as described, the glassware and plasticware were treated overnight with DEPC (0.1% in water) and autoclaved. The pestle and mortar were also DEPC treated and then baked at 300 ºC for 6 h.

Total RNA (Figure 3.2) was isolated from leaf tissue of mature *W. somnifera*. The quantity and integrity of the RNA was checked by Perkin Elmer Lambda 650 system or Nanovue (GE) and also by agarose gel electrophoresis (Section 2.8.6.2). First strand cDNA was synthesized using 2 µg of total RNA by AMV reverse transcriptase based ImProm-II™ Reverse Transcription System (Promega, USA) (as mentioned in Chapter 2, Section 2.8.6.3).

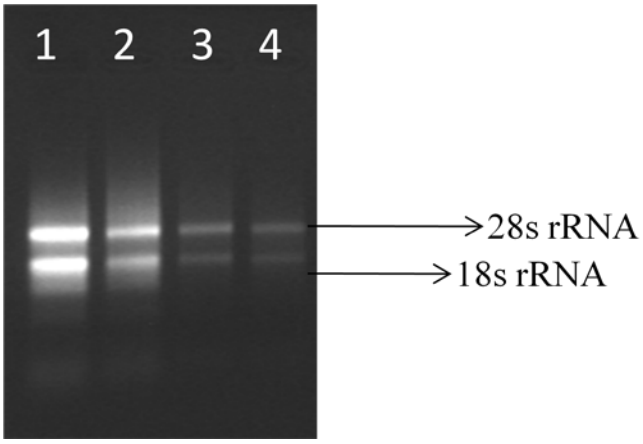


Figure 3.2: Total RNA isolated from leaf tissue (lanes 1, 2, 3 & 4) of *W. somnifera*, resolved on 1% Agarose gel. Arrows indicate both 28S rRNA and 18S rRNA bands

3.3.1.3 PCR amplification of partial cDNA fragment of *GT* genes from *W. somnifera*

Among the several set of primers one primer set, i.e. WsGTF1 – WsGTR1 (as mentioned in section 3.3.1.1) was used to amplify ~500 bp fragment of *GT* gene using cDNA 1st strand as a template (Figure 3.3). The fragment is cloned in pGEM-T easy vector (Figure 3.4) and confirmed by restriction analysis (Figure 3.5) and sequencing (Figure 3.6). This partial clone will be referred to as WsGT in the subsequent discussion. The partial clone of WsGT showed maximum homology (84%) at amino acid level with the Solanaceae member, *Lycopersicon esculentum* (Accession no AJ889012).

3.3.1.4 Primers used for PCR amplification

WsGTF1 5` AGGAAGTGAGCAAATGGAAGAA 3`

WsGTR1 5` ATCTGAACTTCCTCCTTCATTCAC 3`

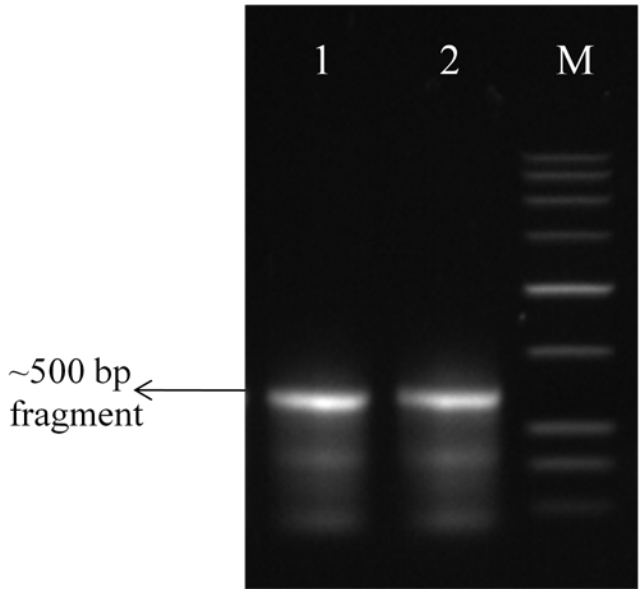


Figure 3.3: PCR amplification product of ~500 bp fragment of WsGT gene separated on 1% agarose gel. Lane 1-2 amplified product and lane M- 100 bp marker.

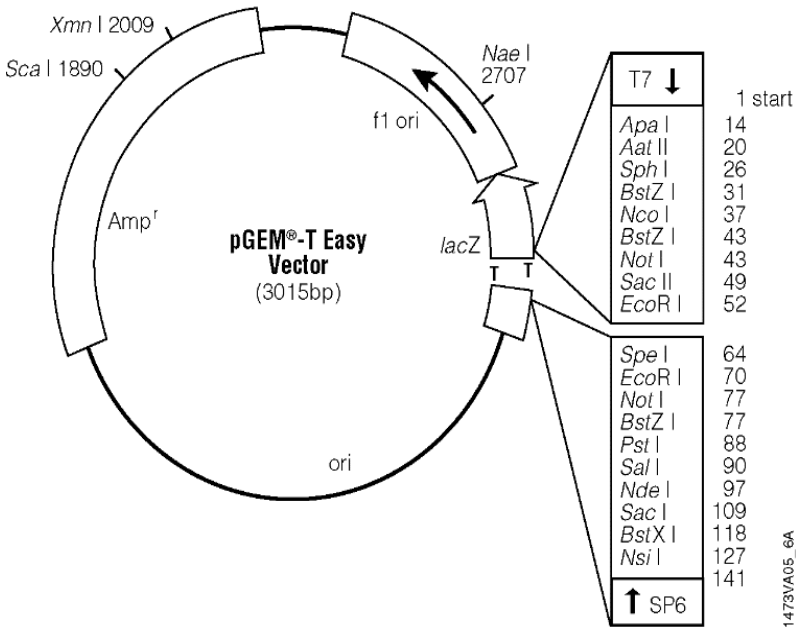


Figure 3.4 Map of pGEM-T Easy vector.

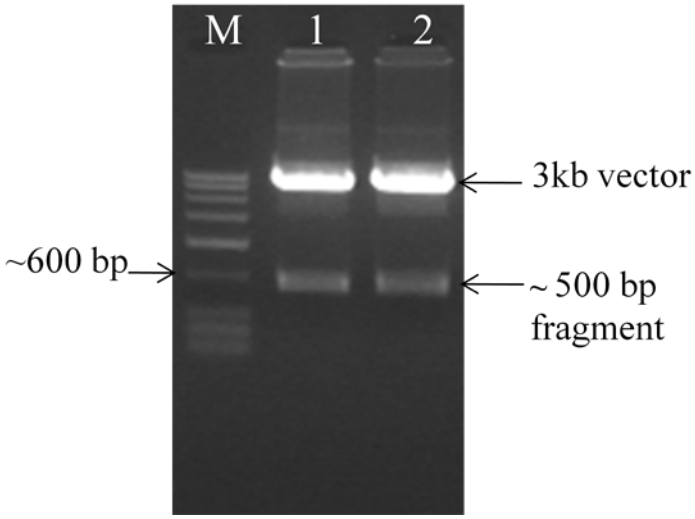


Figure 3.5 pGEM-T clone of WsGT fragment digested with *Eco* RI, Lane M: marker, lane 1-2: cloned ~500 bp insert.

WsGT partial nucleotide sequence

```
   1 AGGAAGTGAG CAAATGGAAG AACTGGCATG GGGTTTGAAG AATAGCAACA
  51 AGAACTTCTT GTGGGTAGTT AGGTCCGCTG AAGAACCCAA ACTTCCGAAG
 101 AACTTCATAG AGGAATTACC AAGTGAAAAA GGCCTAGTGG TATCATGGTG
 151 TCCACAATTA CAGGTGTTGG AACATGAATC AATAGGGTGT TTTATGACGC
 201 ACTGTGGGTG GAATTCGACT TTGGAAGCAA TTAGTTTGGG AGTGCCAATG
 251 GTGACATTGC CACAATGGTC AGATCAACCC ACAAATACAA AGCTTGTGAA
 301 GGATGTTTGG GAGATGGGAG TTAGAGCCAA ACAAGATGAC AAAGGGCTAG
 351 TTAGACGAGA AGTTATCGAA GAGTGTGTAA AACTAGTGAT GGAAGAAGAG
 401 AAAGGAAAAG TGATTAGGGA AAATGCTAAG AAATGGAAGG AATTGGCTAG
 451 AAATGCTGTG AATGAAGGAG GAAGTTCAGA T
```

Figure 3.6: ~500 bp partial WsGT nucleotide sequence. Forward and reverse primers are highlighted in red colour and under lined.

WsGT partial gene sequence shows maximum homology with *Solanacaea* family. Therefore primers were designed from extreme forward and extreme reverse sequence of *Lycopersicon esculentum* to get the entire gene sequence.

3.4 Multiple sequence alignment of glycosyltransferase gene sequences reported in NCBI data base

```
AJ889012        ATGACTACTCACAAAGCTCATTGCTTGATCTTGCCATATCCAGTCCAAGGTCATATCAAC 60
AF190634        ATGACTACTCAAAAAGCTCATTGCTTGATCTTACCATATCCAGCTCAGGGTCATATCAAC 60
                ***********.*******************.**********  **.************

AJ889012        CCAATGCTTCAATTCTCCAAACGTTTACGATCCAAACGCGTTAAAATCACTATAGCACTC 120
AF190634        CCTATGCTCCAATTCTCCAAACGTTTGCAATCCAAAGGTGTCAAAATCACTATAGCAGCC 120
                **:***** ****************.*.******* * ** ***************  *

AJ889012        ACAAAATCCTTTTTGAAAAACATGAAGGAATTGCCAACTTCTATGTCAATTGAGGCCATA 180
AF190634        ACCAAATCATTCTTGAAAACCATGCAAGAATTGTCAACTTCTGTGTCAGTCGAGGCTATC 180
                **.*****.** *******.****.*.****** ********.*****.* ***** **.

AJ889012        TCTGATGGCTATGATGATGGTGGTCGCGATCAAGCAGGAACTTTCGTGGCCTATATTACA 240
AF190634        TCCGATGGCTATGATGATGGCGGACGCGAGCAAGCTGGAACCTTTGTGGCCTATATTACA 240
                ** ***************** **:***** *****:***** ** ***************

AJ889012        CGATTCAAAGAAATTGGTTCGGATACTCTGTCTCAACTTATTCAAAAATTGGCAATTAGT 300
AF190634        AGATTCAAAGAAGTTGGCTCGGATACTTTGTCTCAGCTTATTGGAAAGTTAACAAATTGT 300
                .***********.**** ********* *******.****** .***.**..***:*:**

AJ889012        GGATGTCCTGTAAATTGCATAGTATATGATCCATTCCTCCCTTGGGCTGTTGAAGTTGCA 360
AF190634        GGTTGTCCTGTGAGTTGCATAGTTTACGATCCATTTCTTCCTTGGGCTGTTGAAGTGGGA 360
                **:********.*.*********:** ******** ** ***************** * *

AJ889012        AAACAATTTGGATTAATTAGTGCTGCATTTTTCACACAAAATTGTGTAGTGGATAATCTT 420
AF190634        AATAATTTTGGAGTAGCTACTGCTGCTTTTTTCACTCAATCTTGTGCAGTGGATAACATT 420
                **:.*:****** **. ** ******:********:***:.***** ********* .**
```

```
AJ889012        TATTACCATGTACATAAAGGGGTGATAAAACTTCCACCTACTCAAAATGACGAAGAAATA 480
AF190634        TATTACCATGTACATAAAGGGGTTCTAAAACTTCCTCCAACTGACGTTGATAAAGAAATC 480
                ********************** .**********:**:*** *..:*** .*******.

AJ889012        TTAATTCCTGGATTTCCAAATTCGATCGATGCATCAGATGTACCTTCTTTTGTTATTAGT 540
AF190634        TCAATTCCTGGATT---ATTAACAATTGAGGCATCAGATGTACCTAGTTTTGTTTCTAAT 537
                * ************   *::::*.** ** ***************: *******: **.*

AJ889012        CCTGAAGCAGAAAGGATAGTTGAAATGTTAGCAAATCAATTCTCAAATCTTGACAAAGTT 600
AF190634        CCTGAATCTTCAAGAATACTTGAAATGTTGGTGAATCAGTTCTCGAATCTTGAGAACACA 597
                ****** *: .***.*** **********.* .*****.*****.******** **.. :

AJ889012        GATTGTGTTCTAATCAATAGCTTCTATGAGTTGGAGAAAGAGGTAATTGATTGGATGTCG 660
AF190634        GATTGGGTCCTAATCAACAGTTTCTATGAATTGGAGAAAGAAGTAATTGATTGGATGGCC 657
                ***** ** ******** ** ********.***********.*************** *

AJ889012        AAGATATATCCAATAAAGACAATTGGACCAACAATACCATCAATGTACTTAGACAAGAGA 720
AF190634        AAGATCTATCCAATCAAGACAATTGGACCAACTATACCATCAATGTACCTAGACAAGAGG 717
                *****.********.****************:*************** **********.

AJ889012        CTACATGATGATAAAGAGTATGGTCTTAGTATGTTCAAGCCAATGACAAATGAATGCCTA 780
AF190634        CTACCAGATGACAAAGAATATGGCCTTAGTGTCTTCAAGCCAATGACAAATGCATGCCTA 777
                ****.:***** *****.***** ******.* ******************.*******

AJ889012        AATTGGTTAAACCATCAACCAATTAGCTCAGTGTTGTATGTATCATTTGGAAGTTTAGCC 840
AF190634        AACTGGTTAAACCATCAACCAGTTAGCTCAGTAGTATATGTATCATTTGGAAGTTTAGCC 837
                ** *****************.**********. *.************************

AJ889012        AAACTAGGAAGTGAGCAAATGGAAGAATTGGCATGGGGTTTGAAGAATAGCAACAAGAGC 900
AF190634        AAATTAGAAGCAGAGCAAATGGAAGAATTAGCATGGGGTTTGAGTAATAGCAACAAGAAC 897
                *** ***.*. :*****************.*************. *************.*

AJ889012        TTCTTGTGGGTTGTTAGGTCTACTGAAGAGCCCAAACTCCCCAACAACTTTATTGAGGAA 960
AF190634        TTCTTGTGGGTAGTTAGATCCACTGAAGAATCCAAACTTCCCAACAACTTTTTAGAGGAA 957
                ***********:*****.** ********. ******* ************:*:******

AJ889012        TTAACAAGTGAAAAAGGCTTAGTGGTGTCATGGTGTCCACAATTACAAGTGTTGGAACAT 1020
AF190634        TTAGCAAGTGAAAAAGGATTAGTCGTGTCATGGTGTCCACAATTACAAGTCTTGGAACAT 1017
                ***.*************.***** ************************** *********

AJ889012        GAGTCGATAGGGTGTTTTCTGACGCACTGTGGATGGAATTCAACTCTGGAAGCGATTAGT 1080
AF190634        AAATCAATAGGGTGTTTTCTCACGCACTGTGGCTGGAATTCAACTTTGGAAGCAATTAGT 1077
                .*.**.************** ***********.************ *******.******

AJ889012        TTGGGAGTGCCAATGGTGGCAATGCCACAATGGTCTGATCAACCAACAAATGCAAAGCTT 1140
AF190634        TTGGGAGTACCAATGATTGCAATGCCACATTGGTCAGACCAGCCAACAAATGCGAAGCTT 1137
                ********.******.* ***********:*****:** **.***********.******

AJ889012        GTGAAAGATGTTTGGGAAATAGGTGTTAGAGCCAAACAAGATGAAAAAGGGGTAGTTAGA 1200
AF190634        GTGGAAGATGTTTGGGAGATGGGAATTAGACCAAAACAAGATGAAAAAGGATTAGTTAGA 1197
                ***.*************.**.**:.***** *.**************. ********

AJ889012        AGAGAAGTTATAGAAGAATGTATAAAGCTAGTGATGGAAGAAGATAAAGGAAAACTAATT 1260
AF190634        AGAGAAGTTATTGAAGAATGTATTAAGATAGTGATGGAGGAAAAGAAAGGAAAAAAGATT 1257
                ***********:***********:***.**********.***.* *********.:.***

AJ889012        AGAGAAAATGCAAAGAAATGGAAGGAAATAGCTAGAAATGTTGTGAATGAAGGAGGAAGT 1320
AF190634        AGGGAAAATGCAAAGAAATGGAAGGAATTGGCTAGGAAAGCTGTGGATGAAGGAGGAAGT 1317
                **.************************:*.*****.**:* ****.**************

AJ889012        TCAGATAAAAACATTGAAGAATTTGTTTCCAAGTTGGTTACTATTTCCTAA--------- 1371
AF190634        TCAGATAGAAATATTGAAGAATTTGTTTCCAAGTTGGTGACTATTGCCTCAGTGGAAAGC 1377
                *******.*** ************************* ****** ***.*

AJ889012        ---
AF190634        TAA 1380
```

Figure 3.7 Sequence alignment of reported *GT* genes for primer synthesis

3.4.1 Primers used in PCR amplification

WsGTF2 5` ATGACTACTCACAAAGCTCA 3`

WsGTR2 5` GGAAATAGTAACCAACTTGG 3`

3.4.2 PCR amplification

RNA was isolated from *Withania* leaf tissue and cDNA is synthesized according to manual instructions. Extreme forward and reverse primers (see above section 3.4.1) were used in PCR amplification. ~1.3kb fragment was amplified (Figure 3.8), it was cloned into pGEM-T Easy vector and analyzed by restriction analysis (Figure 3.9) and sequenced (Figure 3.10).

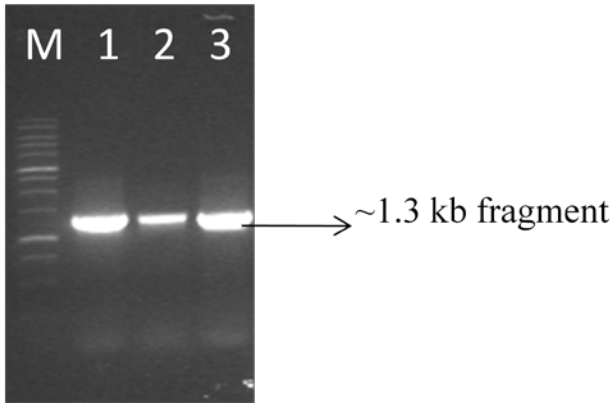


Figure 3.8 PCR amplicon of ~1.3 kb fragment of WsGT gene separated on 1% agarose gel. Lane 1, 2 & 3 amplified product and lane M- 1kb marker.

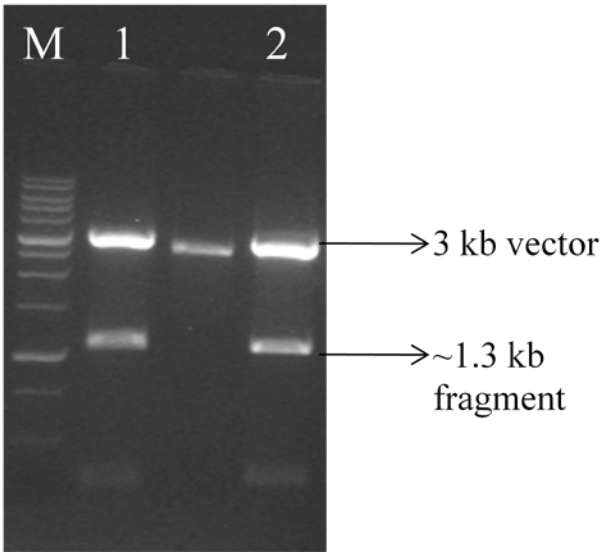


Figure 3.9 WsGT ~1.3 kb insert was released when digested with *Eco* RI. Lane M: 1 kb marker, Lane 1 & 2 digested pGEM-T vector and insert.

```
   1 ATGACTACTC ACAAAGCTCA CTGCTTAATC TTGCCATATC CAGGCCAAGG
  51 TCATGTCAAC CCAATGCTCC AATTCTCCAA ACGTTTGCAA TCCAAAAGTG
 101 TCAAAATCAC AATAGCAACC ACAAAATCCT TCTTGAAAAA AATGCAAAAA
 151 TTACCAACTT CTATTTCAAT CGAAGCCATC TCTGATGGCT ACGATGATGA
 201 TGGCCTCGAC CAAGCAAGAT CTTATGCAGC CTATTTAACA AGATTCAAAG
 251 AAGTTGGCTC AGATACTCTG TCTCAACTTA TTGAAAAGTT AGCAAATTCT
 301 GGGAGCCCTG TGAATTGCAT AGTATATGAT CCATTCCTTC CTTGGGTTGT
 351 TGAAGTGGCA AAGAACTTTG GATTAGCTAT TGCTGCATTT TTCACACAAT
 401 CTTGTGCAGT GGACAACATT TATTACCATG TACATAAAGG GGTACTAAAA
 451 CTTCCTCCTA CTCAAGTTGA TGAAGAAATA TTAATTCCTG GATTATCATA
 501 TGCAATTGAG AGTTCAGACG TACCTAGTTT TGAGTCTACT TCTGAACCAG
 551 ATTTACTTGT TGAACTGTTG GCGAATCAGT TTTCAAATCT TGAGAAAACT
 601 GATTGGGTCC TAATCAACAG CTTCTATGAG TTGGAAAAAC ATGTAATTGA
 651 TTGGATGTCC AAGATTTATC CAATCAAGGC AATTGGACCA ACAATACCAT
 701 CCATGTACCT AGACAAGAGG CTACCAGATG ACAAAGAATA CGGCCTTAGT
 751 ATGTTCAAGC CAATAACAGA TGCATGCATA AATTGGCTAA ACCACCAACC
 801 AATTAACTCA GTGTTATATG TATCATTTGG AAGCTTAGCC AAACTAGAAG
 851 CTGAGCAAAT GGAAGAACTG GCATGGGGTT TGAAGAATAG CAACAAGAAC
 901 TTCTTGTGGG TAGTTAGGTC CGCTGAAGAA CCCAAACTTC CGAAGAACTT
 951 CATAGAGGAA TTACCAAGTG AAAAAGGCCT AGTGGTATCA TGGTGTCCAC
1001 AATTACAGGT GTTGGAACAT GAATCAATAG GGTGTTTTAT GACGCACTGT
1051 GGGTGGAATT CGACTTTGGA AGCAATTAGT TTGGGAGTGC CAATGGTGAC
1101 ATTGCCACAA TGGTCAGATC AACCCACAAA TACAAAGCTT GTGAAGGATG
1151 TTTGGGAGAT GGGAGTTAGA GCCAAACAAG ATGACAAAGG GCTAGTTAGA
1201 CGAGAAGTTA TCGAAGAGTG TATAAAACTA GTGATGGAAG AAGAGAAAGG
1251 AAAAGTGATT AGGGAAAATG CTAAGAAATG GAAGGAATTG GCTAGAAATG
1301 CTGTGGATGA AGGTGGAAGT TCAGATAAAA ACATTGAAGA ATTTGTTTCC
1351 AAGTTGGTTA CTATTTCA
```

Figure 3.10 ~1.3kb WsGT partial nucleotide sequence. Extreme forward and reverse primers are highlighted in red colour and under lined.

3.5 Gene copy number by Southern Hybridization

To understand the distribution of the *WsGT* gene in the *W. somnifera* genome, Southern hybridization was performed (Chapter 2, Section 2.8.4). A 20 µg aliquot of *W. somnifera* genomic DNA was restriction digested individually with three restriction enzymes, *Xho1, EcoRI* and *HindIII* (Figure 3.11). One enzyme, *XhoI* does not cut inside the gene while other enzyme, *Eco RI* has two sites within the gene and HindIII has one site within the gene. As shown in Figure 3.12, two bands were detected in the *Xho1* digests, while the other one *EcoRI* shows two distinct bands and the *HindIII* doesn't show any distinct bands, may be due to the reason that bands may be very small and of low intensity, which is not seen in blot. Southern hybridization was done using a part of coding region. Post hybridization membrane was exposed on intensifying screen. The screen was scanned for signal development using a very robust and much sensitive Typhoon TRIO+ scanner. An approximately 400 bp fragment was used as a probe for hybridization at 60 °C. Banding pattern in Southern hybridization suggested that at least 2 copy of *WsGT* gene is present in *W. somnifera*, as two bands are observed in well digested with *Eco* RI, and subjected to Southern hybridization (Figure 3.12).

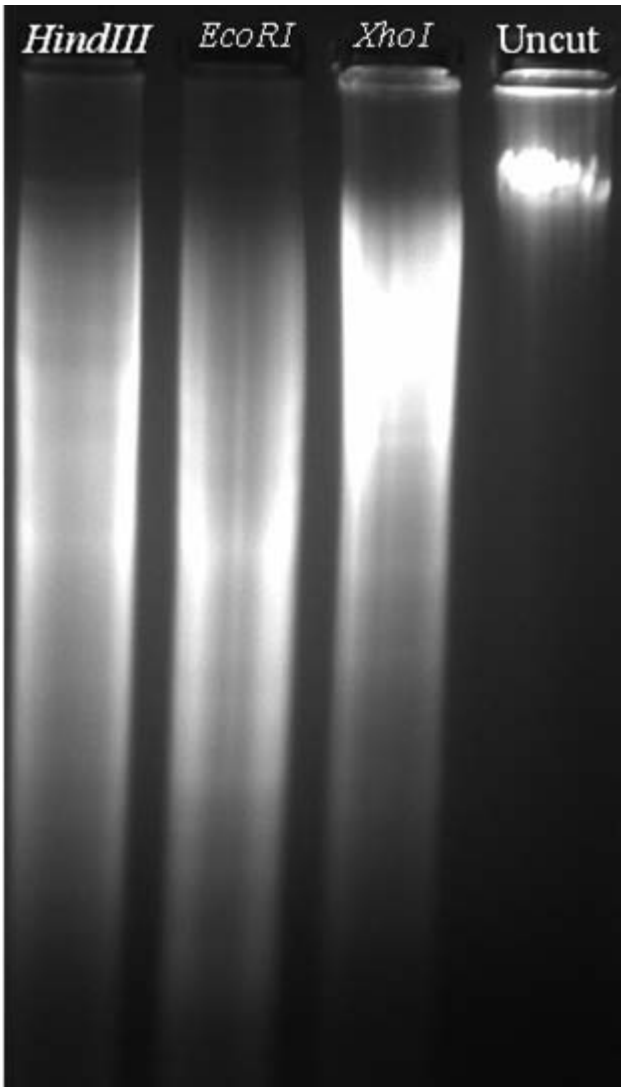


Figure 3.11 0.8% agarose gel showing *W. somnifera* gDNA digested with *XhoI*, *Eco* RI and *HindIII*. In one Lane gDNA undigested.

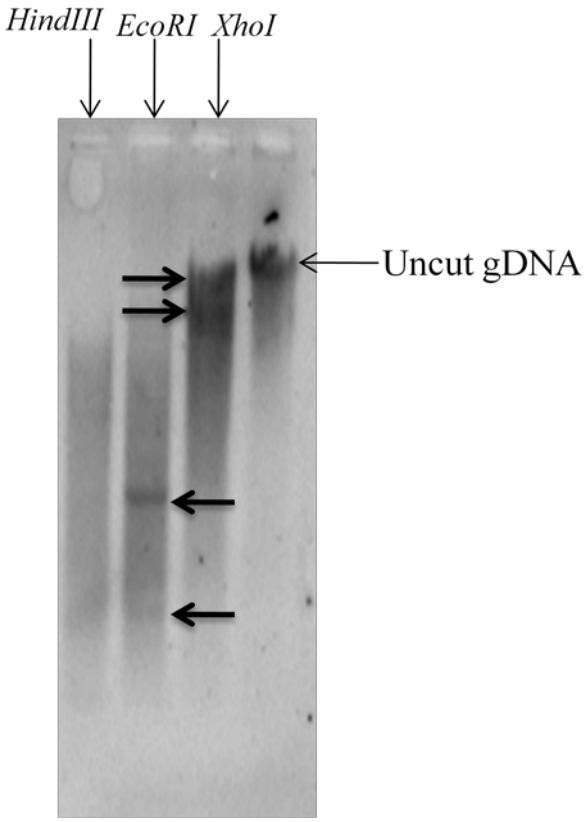


Figure 3.12 Southern hybridized signals of the above mentioned gel scanned using Typhoon TRIO+ scanner. Arrows indicate hybridization signals with respective enzymes.

3.6 Rapid Amplification of cDNA Ends (RACE) using BD Clontech kit

RACE is an important tool to obtain the UTRs (Un-translated regions) of a particular gene and is also useful to obtain full length gene of unknown partial fragment.

3.6.1 Primer designing for RACE

Based on the 1368 bp partial WsGT gene, one 5'Gene specific primer (GSP) and one 5'Nested GSP were designed for 5'RACE as shown below highlighted and underlined with green colour. For 3'RACE one 3` Gene specific primer and one 3'Nested GSP was designed as shown below highlighted with red colour and underlined (Figure 3.13).

```
   1 ATGACTACTC ACAAAGCTCA CTGCTTAATC TTGCCATATC CAGGCCAAGG
  51 TCATGTCAAC CCAATGCTCC AATTCTCCAA ACGTTTGCAA TCCAAAAGTG
 101 TCAAAATCAC AATAGCAACC ACAAAATCCT TCTTGAAAAA AATGCAAAAA
 151 TTACCAACTT CTATTTCAAT CGAAGCCATC TCTGATGGCT ACGATGATGA
 201 TGGCCTCGAC CAAGCAAGAT CTTATGCAGC CTATTTAACA AGATTCAAAG
 251 AAGTTGGCTC AGATACTCTG TCTCAACTTA TTGAAAAGTT AGCAAATTCT

                  5` Nested Gene specific primer
 301 GGGAGCCCTG TGAATTGCAT AGTATATGAT CCATTCCTTC CTTGGGTTGT

            5` Gene specific primer
 351 TGAAGTGGCA AAGAACTTTG GATTAGCTAT TGCTGCATTT TTCACACAAT
 401 CTTGTGCAGT GGACAACATT TATTACCATG TACATAAAGG GGTACTAAAA
 451 CTTCCTCCTA CTCAAGTTGA TGAAGAAATA TTAATTCCTG GATTATCATA
 501 TGCAATTGAG AGTTCAGACG TACCTAGTTT TGAGTCTACT TCTGAACCAG
 551 ATTTACTTGT TGAACTGTTG GCGAATCAGT TTTCAAATCT TGAGAAAACT
 601 GATTGGGTCC TAATCAACAG CTTCTATGAG TTGGAAAAAC ATGTAATTGA
 651 TTGGATGTCC AAGATTTATC CAATCAAGGC AATTGGACCA ACAATACCAT
 701 CCATGTACCT AGACAAGAGG CTACCAGATG ACAAAGAATA CGGCCTTAGT
 751 ATGTTCAAGC CAATAACAGA TGCATGCATA AATTGGCTAA ACCACCAACC


                                        3` Gene specific primer
 801 AATTAACTCA GTGTTATATG TATCATTTGG AAGCTTAGCC AAACTAGAAG

                                  3` Nested Gene specific primer
 851 CTGAGCAAAT GGAAGAACTG GCATGGGGTT TGAAGAATAG CAACAAGAAC
 901 TTCTTGTGGG TAGTTAGGTC CGCTGAAGAA CCCAAACTTC CGAAGAACTT
 951 CATAGAGGAA TTACCAAGTG AAAAAGGCCT AGTGGTATCA TGGTGTCCAC
1001 AATTACAGGT GTTGGAACAT GAATCAATAG GGTGTTTTAT GACGCACTGT
1051 GGGTGGAATT CGACTTTGGA AGCAATTAGT TTGGGAGTGC CAATGGTGAC
1101 ATTGCCACAA TGGTCAGATC AACCCACAAA TACAAAGCTT GTGAAGGATG
1151 TTTGGGAGAT GGGAGTTAGA GCCAAACAAG ATGACAAAGG GCTAGTTAGA
1201 CGAGAAGTTA TCGAAGAGTG TATAAAACTA GTGATGGAAG AAGAGAAAGG
1251 AAAAGTGATT AGGGAAAATG CTAAGAAATG GAAGGAATTG GCTAGAAATG
1301 CTGTGGATGA AGGTGGAAGT TCAGATAAAA ACATTGAAGA ATTTGTTTCC
1351 AAGTTGGTTA CTATTTCA
```

Figure 3.13 Gene specific primer designing for RACE

3.6.2: 3' RACE reaction

To isolate the 3' end sequence of the WsGT gene, 3' RACE reaction was performed (as described in Chapter 2, Section 2.10). For WsGT gene, 3'RACE was performed using forward Gene specific primer WsGTRACE F1 (5` TGGAAGCTTAGCCAAACTAGAA 3`) and reverse 10X Universal Primer Mix (10XUPM) provided with the kit. 3' RACE ready cDNA was used as the template. The RACE reaction resulted in an amplified product of

approximately 700 bp; which was further re-amplified using 3' Nested WsGTRACE F2 (5` GGCATGGGGTTTGAAGAATAGCAACAAG 3') and Nested Universal Primer (NUP) provided with kit. The resultant PCR product (Figure 3.14A) was cloned into pGEMT-Easy vector (Promega, USA) and the recombinant construct was confirmed by *Eco* RI restriction analysis (Figure 3.14B) and by sequencing (Figure 3.15).

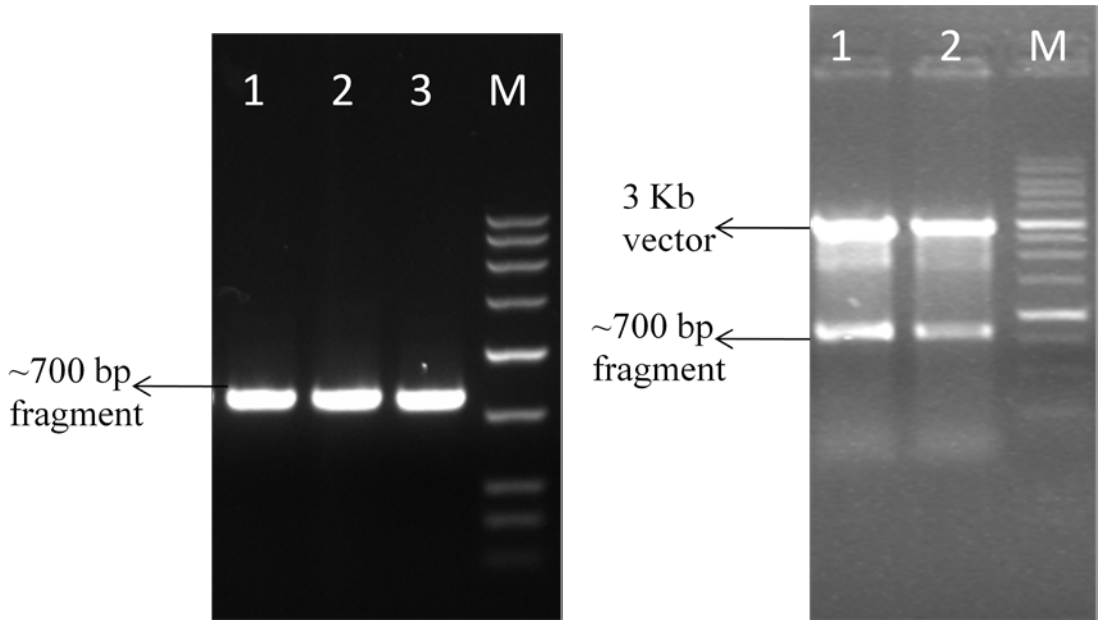


Figure A **Figure B**

Figure 3.14 Figure A 1% agarose gel showing 3'RACE product of WsGT gene. Lane M- marker, lane 1, 2 & 3 is ~700 bp 3'RACE product. **Figure B** pGEM-T digestion with *EcoR1* enzyme; Lane M- marker, Lane 1 & 2 *Eco RI* digested recombinant pGEMT-Easy vector releasing ~ 700bp fragment.

```
  1 GGCATGGGGT TTGAAGAATA GCAACAAGAA CTTCTTGTGG GTAGTTAGGT
 51 CCGCTGAAGA ACCCAAACTT CCGAAGAACT TCATAGAGGA ATTACCAAGT
101 GAAAAAGGCC TAGTGGTATC ATGGTGTCCA CAATTACAGG TGTTGGAACA
151 TGAATCAATA GGGTGTTTTA TGACGCACTG TGGGTGGAAT TCGACTTTGG
201 AAGCAATTAG TTTGGGAGTG CCAATGGTGA CATTGCCACA ATGGTCAGAT
251 CAACCCACAA ATACAAAGCT TGTGAAGGAT GTTTGGGAGA TGGGAGTTAG
301 AGCCAAACAA GATGACAAAG GGCTAGTTAG ACGAGAAGTT ATCGAAGAGT
351 GTATAAAACT AGTGATGGAA GAAGAGAAAG GAAAAGTGAT TAGGGAAAAT
401 GCTAAGAAAT GGAAGGAATT GGCTAGAAAT GCTGTGGATG AAGGTGGAAG
451 TTCAGATAAA AACATTGAAG AATTTGTTTC CAAGTTGGTG ACTATTTCCT
501 AAGTAAGAAA TAAACAGCAA CCTAATCAGT GTGTTGCAAG TTATATTTTT
551 CAGGTTGAAA GTATCAAATT TTGACGTGAA TCTTCCTTTG TTTTCTCCTT
601 GGCTCATATT GTAACAAGCT ATAATTGTAA CATTGAATAT TAATTATATG
651 GATTGATGAT AAAAAAAAAA AAAAAAAAAA AAAAAAAAGT ACTCTGCGTT
701 GATACCACTG CTT
```

Figure 3.15 Sequence analysis of 3' RACE product of WsGT gene:

The 3'UTR is represented by 158 bp and it extends from stop codon till end of poly-A tail. Stop codon (TAA) is highlighted in red, poly-A tail is highlighted in red and Nested Universal Primer is represented by pink letters and underlined. 3` forward Gene specific primer is shown in green colour and underlined.

3.6.3: 5' RACE reaction

In order to get the 5'region of the WsGT genes, 5' RACE was performed as described in Chapter 2, Section 2.10. For WsGT gene, 5'RACE was performed using gene specific reverse primer 5'GSP (5' GAAGTGGCAAAGAACTTTGGATTA 3') and 10X Universal primer mix provided with the kit. In this reaction 5' RACE ready cDNA was used as the template. The above RACE product was diluted (1:20) and secondary PCR was done using 5'nested GSP (5' CCTG TGAATTGCATAGTATATGATCCA 3') and Nested universal primer provided with the kit. Agarose gel electrophoresis analysis revealed an amplification product of approximately 250 bp as depicted in figure 3.16A. This amplicon was cloned in pGEMT-Easy vector (Promega, USA) and the resulting construct was confirmed by *Eco*RI restriction analysis (Figure 3.16B) and by sequencing (Figure 3.17).

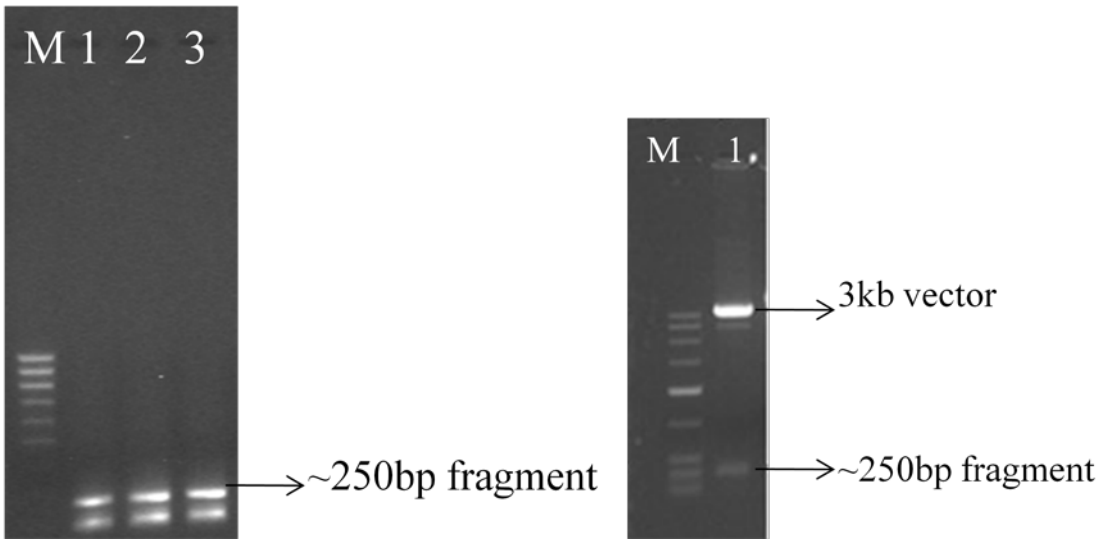


Figure A **Figure B**

Figure 3.16 Figure A: 1% agarose gel showing ~250bp 5'RACE product of WsGT gene. Lane M- 100 bp marker; Lane 1, 2 & 3 ~250 bp fragment. **Figure B:** pGEM-T digestion with *EcoR1* enzyme; Lane M- marker, Lane 1 *Eco RI* digested recombinant pGEMT-Easy vector releasing ~ 250 bp fragment.

```
  1 AAGCAGTGGT ATCAACGCAG AGTACGCGGG GGTTGCAAAA TATGACTACT
 51 CACAAAGCTC ACTGCTTGAT CTTGCCATAT CCAACCCAAG GTCATGTCCC
101 GATTCTCCAA AGGTTTACAA TTCAAAAAGT GTCAACTTCC ATATCAATCG
151 AGGCCATCTC TGATGGTGGC AGGACCTTAT GCGGCCTATT TAACAAGATT
201 CAAAAATTGA CATATTGTGG ATGCCCTGTG AATTGCATAG TATATGATCC
251 A
```

Figure 3.17 Sequence analysis of 5' RACE product of WsGT gene:

Nested Universal Primer shown in red colour and underlined, 5` UTR shown in blue colour; start codon (ATG) shown in green colour and 5` Nested Gene specific primer shown in pink colour and underlined.

3.7 Isolation, Cloning and Characterization of full length WsGT gene

To isolate the full length sequence of WsGT gene primers were designed from start codon (extreme forward) and from stop codon (extreme reverse). PCR was performed using cDNA

1st strand as a template. High fidelity *Taq pol* (*Pfx* Invitrogen) was used to amplify the WsGT gene. Approximately 1.4 kb fragment was amplified (Figure 3.18), it was cloned into pGEM-T Easy vector and analyzed by restriction analysis (Figure 3.19) and sequenced (Figure 3.20).

3.7.1 Primers used in PCR amplification

WsGT F 5` ATGACTACTCAC AAAGCTCACT 3`

WsGT R 5` TTAGGAAATAGTAACCAACTTGGA 3`

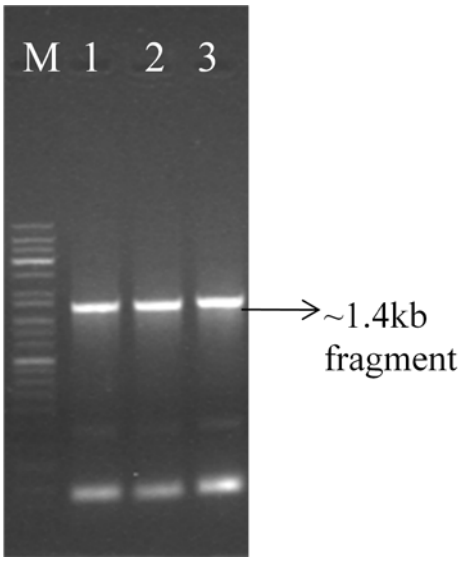


Figure 3.18: PCR amplicon of ~1.4 kb fragment of WsGT gene separated on 1% agarose gel. Lane M-Marker and Lane 1, 2 & 3 ~1.4 kb amplicon.

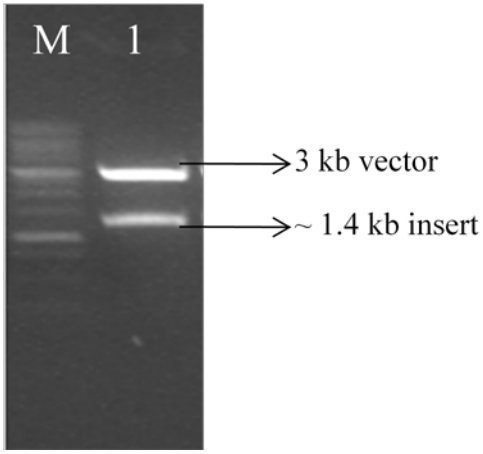


Figure 3.19: ~1.4 kb WsGT insert was released when digested with *Eco* RI. Lane M-Marker and Lane 1 digested pGEM-T Easy vector and insert.

```
   1 ATGACTACTC ACAAAGCTCA CTGCTTAATC TTGCCATATC CAGGCCAAGG
  51 TCATGTCAAC CCAATGCTCC AATTCTCCAA ACGTTTGCAA TCCAAAAGTG
 101 TCAAAATCAC AATAGCAACC ACAAAATCCT TCTTGAAAAA AATGCAAAAA
 151 TTACCAACTT CTATTTCAAT CGAAGCCATC TCTGATGGCT ACGATGATGA
 201 TGGCCTCGAC CAAGCAAGAT CTTATGCAGC CTATTTAACA AGATTCAAAG
 251 AAGTTGGCTC AGATACTCTG TCTCAACTTA TTGAAAAGTT AGCAAATTCT
 301 GGGAGCCCTG TGAATTGCAT AGTATATGAT CCATTCCTTC CTTGGGTTGT
 351 TGAAGTGGCA AAGAACTTTG GATTAGCTAT TGCTGCATTT TTCACACAAT
 401 CTTGTGCAGT GGACAACATT TATTACCATG TACATAAAGG GGTACTAAAA
 451 CTTCCTCCTA CTCAAGTTGA TGAAGAAATA TTAATTCCTG GATTATCATA
 501 TGCAATTGAG AGTTCAGACG TACCTAGTTT TGAGTCTACT TCTGAACCAG
 551 ATTTACTTGT TGAACTGTTG GCGAATCAGT TTTCAAATCT TGAGAAAACT
 601 GATTGGGTCC TAATCAACAG CTTCTATGAG TTGGAAAAAC ATGTAATTGA
 651 TTGGATGTCC AAGATTTATC CAATCAAGGC AATTGGACCA ACAATACCAT
 701 CCATGTACCT AGACAAGAGG CTACCAGATG ACAAAGAATA CGGCCTTAGT
 751 ATGTTCAAGC CAATAACAGA TGCATGCATA AATTGGCTAA ACCACCAACC
 801 AATTAACTCA GTGTTATATG TATCATTTGG AAGCTTAGCC AAACTAGAAG
 851 CTGAGCAAAT GGAAGAACTG GCATGGGGTT TGAAGAATAG CAACAAGAAC
 901 TTCTTGTGGG TAGTTAGGTC CGCTGAAGAA CCCAAACTTC CGAAGAACTT
 951 CATAGAGGAA TTACCAAGTG AAAAAGGCCT AGTGGTATCA TGGTGTCCAC
1001 AATTACAGGT GTTGGAACAT GAATCAATAG GGTGTTTTAT GACGCACTGT
1051 GGGTGGAATT CGACTTTGGA AGCAATTAGT TTGGGAGTGC CAATGGTGAC
1101 ATTGCCACAA TGGTCAGATC AACCCACAAA TACAAAGCTT GTGAAGGATG
1151 TTTGGGAGAT GGGAGTTAGA GCCAAACAAG ATGACAAAGG GCTAGTTAGA
1201 CGAGAAGTTA TCGAAGAGTG TATAAAACTA GTGATGGAAG AAGAGAAAGG
1251 AAAAGTGATT AGGGAAAATG CTAAGAAATG GAAGGAATTG GCTAGAAATG
1301 CTGTGGATGA AGGTGGAAGT TCAGATAAAA ACATTGAAGA ATTTGTTTCC
1351 AAGTTGGTTA CTATTTCCTA A
```

Figure 3.20: Full length sequence of WsGT gene. Start codon (ATG) shown in green colour and stop codon (TAA) shown in red colour.

The full length cDNA sequence of WsGT has been submitted to NCBI GenBank database under accession no.FJ560880. The WsGT sequence contains a 1371 bp open-reading frame, which codes for a protein of 456 amino acids with a molecular weight 52 kDa (Figure 3.19). The sequence analysis shows that RACE amplification has been well extended downstream to the stop codon (TAA) and has yielded a 158 bp 3'UTR and a poly-A tail. Also it shows 18 bp 5` UTR region. Upstream to poly-A tail an ubiquitous polyadenylation (poly-A) signal sequence '**AATAA**' was present in the 3'UTR (Figure 3.21) which is present in most

eukaryotes and forms a complex with U4 SnRNP for primary cleavage site selection in pre-mRNA (Joshi, 1987).

Nucleotide sequence of WsGT (FJ560880)

```
   1 ACGCGGGGGT TGCAAAATAT GACTACTCAC AAAGCTCACT GCTTAATCTT
  51 GCCATATCCA GGCCAAGGTC ATGTCAACCC AATGCTCCAA TTCTCCAAAC
 101 GTTTGCAATC CAAAAGTGTC AAAATCACAA TAGCAACCAC AAAATCCTTC
 151 TTGAAAAAAA TGCAAAAATT ACCAACTTCT ATTTCAATCG AAGCCATCTC
 201 TGATGGCTAC GATGATGATG GCCTCGACCA AGCAAGATCT TATGCAGCCT
 251 ATTTAACAAG ATTCAAAGAA GTTGGCTCAG ATACTCTGTC TCAACTTATT
 301 GAAAAGTTAG CAAATTCTGG GAGCCCTGTG AATTGCATAG TATATGATCC
 351 ATTCCTTCCT TGGGTTGTTG AAGTGGCAAA GAACTTTGGA TTAGCTATTG
 401 CTGCATTTTT CACACAATCT TGTGCAGTGG ACAACATTTA TTACCATGTA
 451 CATAAAGGGG TACTAAAACT TCCTCCTACT CAAGTTGATG AAGAAATATT
 501 AATTCCTGGA TTATCATATG CAATTGAGAG TTCAGACGTA CCTAGTTTTG
 551 AGTCTACTTC TGAACCAGAT TTACTTGTTG AACTGTTGGC GAATCAGTTT
 601 TCAAATCTTG AGAAAACTGA TTGGGTCCTA ATCAACAGCT TCTATGAGTT
 651 GGAAAAACAT GTAATTGATT GGATGTCCAA GATTTATCCA ATCAAGGCAA
 701 TTGGACCAAC AATACCATCC ATGTACCTAG ACAAGAGGCT ACCAGATGAC
 751 AAAGAATACG GCCTTAGTAT GTTCAAGCCA ATAACAGATG CATGCATAAA
 801 TTGGCTAAAC CACCAACCAA TTAACTCAGT GTTATATGTA TCATTTGGAA
 851 GCTTAGCCAA ACTAGAAGCT GAGCAAATGG AAGAACTGGC ATGGGGTTTG
 901 AAGAATAGCA ACAAGAACTT CTTGTGGGTA GTTAGGTCCG CTGAAGAACC
 951 CAAACTTCCG AAGAACTTCA TAGAGGAATT ACCAAGTGAA AAAGGCCTAG
1001 TGGTATCATG GTGTCCACAA TTACAGGTGT TGGAACATGA ATCAATAGGG
1051 TGTTTTATGA CGCACTGTGG GTGGAATTCG ACTTTGGAAG CAATTAGTTT
1101 GGGAGTGCCA ATGGTGACAT TGCCACAATG GTCAGATCAA CCCACAAATA
1151 CAAAGCTTGT GAAGGATGTT TGGGAGATGG GAGTTAGAGC CAAACAAGAT
1201 GACAAAGGGC TAGTTAGACG AGAAGTTATC GAAGAGTGTA TAAAACTAGT
1251 GATGGAAGAA GAGAAAGGAA AAGTGATTAG GGAAAATGCT AAGAAATGGA
1301 AGGAATTGGC TAGAAATGCT GTGGATGAAG GTGGAAGTTC AGATAAAAAC
1351 ATTGAAGAAT TTGTTTCCAA GTTGGTTACT ATTTCCTAAG TAAGAATAAA
1401 ACAGCAACCT AATCAGTGTG TTGCAAGTTA TATTTTTCAG GTTGAAAGTA
1451 TCAAATTTTG ACGTGAATCT TCCTTTGTTT TCTCCTTGGC TCATATTGTA
1501 ACAAGCTATA ATTGTAACAT TGAATATTAA TTATATGGAT TGATGATAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAA
```

Fig 3.21 Full length nucleotide sequence with 5` & 3` UTR region

Deduced Amino acid sequence of WsGT (FJ560880)

M T T H K A H C L I L P Y P G Q G H V N P M L Q F S K R L Q S K S V K I T I A T T K
S F L K K M Q K L P T S I S I E A I S D G Y D D D G L D Q A R S Y A A Y L T R F K E
V G S D T L S Q L I E K L A N S G S P V N C I V Y D P F L P W V V E V A K N F G L A
I A A F F T Q S C A V D N I Y Y H V H K G V L K L P P T Q V D E E I L I P G L S Y A I

E S S D V P S F E S T S E P D L L V E L L A N Q F S N L E K T D W V L I N S F Y E L E
K H V I D W M S K I Y P I K A I G P T I P S M Y L D K R L P D D K E Y G L S M F K P
I T D A C I N W L N H Q P I N S V L Y V S F G S **L A K L E A E Q M E E L A W G L K**
N S N K N F L W V V R S A E E P K L P K N F I E E L P S E K G L V V S **W C P Q L**
Q V L E H E S I G C F M T H C G W N S T L E A I S L G V P M V T L P Q W S D Q P
T N T K L V K D V W E M G V R A K Q D D K G L V R R E V I E E C I K L V M E E E
K G K V I R E N A K K W K E L A R N A V D E G G S S D K N I E E F V S K L V T I S

Figure 3.22 Deduced amino acid sequence of ORF region

Analysis of the deduced amino acid sequence showed it contained the donor binding site (green coloured bold and underlined) and the 44 bp PSPG box which is conserved in all plant glycosyltransferase genes (coloured in red). Most conserved region of PSPG box i.e, **HCGWNS** (highlighted in turquoise) is also present in the sequence. It also contained the tryptophan residue in the most conserved region of PSPG box (**HCGWNS**) which is found to be responsible for glucose specificity.

3.7.2 CLUSTAL W (1.8) multiple sequence alignment of amino acid sequences of WsGT with *Lycium barbarum*

```
WsGT            MTTHKAHCLILPYPGQGHVNPMLQFSKRLQSKSVKITIATTKSFLKKMQKLPTSISIEAI 60
Lycium          MTTHRAHCLILPYPSQGHINPMLQFSKRLQSKGVKITIATTKSFLKTMQELTTSVSIEAI 60
                ****:*********.***:*************.*************.**:*.**:*****

WsGT            SDGYDDDGLDQARSYAAYLTRFKEVGSDTLSQLIEKLANSGSPVNCIVYDPFLPWVVEVA 120
Lycium          SDGYDDGGRDQAGSFVAYITRFKEVGSDTLAQLIKKLANSGCPVNCIVYDPFLPWAVEVA 120
                ******.* *** *:.**:***********:***:******.*************.****

WsGT            KNFGLAIAAFFTQSCAVDNIYYHVHKGVLKLPPTQVDEEILIPGLSYAIESSDVPSFEST 180
Lycium          KDFGLVSAAFFTQNCAVDNIYYHVHKGVLKLPPTQDDEEILIPGFSCPIESSDVPSFVIS 180
                *:***. ******.******************** ********:* .*********  :

WsGT            SEPDLLVELLANQFSNLEKTDWVLINSFYELEKHVIDWMSKIYPIKAIGPTIPSMYLDKR 240
Lycium          PEAARILDMLVNQFSNLDKVDWVLINSFYELEKEVIDWMSKIYPIKTIGPTIPSMYLDNR 240
                .*.  ::::*.******:*.*************.************:***********:*

WsGT            LPDDKEYGLSMFKPITDACINWLNHQPINSVLYVSFGSLAKLEAEQMEELAWGLKNSNKN 300
Lycium          LPDDKEYGLSVFKPMTNECLNWLNHQLISSVVYVSFGSLAKVEVEQMEELAWGLKNSNKN 300
                **********:***:*: *:****** *.**:*********:*.***************

WsGT            FLWVVRSAEEPKLPKNFIEELP--SE-KGLVVSWCPQLQVLEHESIGCFMTHCGWNSTLE 357
Lycium          FLWVVRSTEESKLPKNFLEELKLVSENKGLVVSWCPQLQVLEHKSTGCFLTHCGWNSTLE 360
                *******:**.******:***   ** ****************:* ***:**********

WsGT            AISLGVPMVTLPQWSDQPTNTKLVKDVWEMGVRAKQDDKGLVRREVIEECIKLVMEEEKG 417
Lycium          AISLGVPMLTMPQWTDQPTNAKLVKDVWEMGVRAKQDEKGIVRREVIEECIKLVMEEEKG 420
                ********:*:***:*****:****************:**:******************
```

```
WsGT            KVIRENAKKWKELARNAVDEGGSSDKNIEEFVSKLVTIS---- 456
Lycium          KMIKENAQKWKELARKAVDEGGSSDKNIEEFVSKLVTISSLGS 463
                *:*:***:*******:**********************
```

Figure 3.23 WsGT amino acid sequence alignment with *Lycium barbarum*.

The amino acid sequence of WsGT sequence shows 84% homology with UDP-glucose:glucosyltransferase from *Lycium barbarum* (AB360617) (Figure 3.23).

3.7.3 CLUSTAL W (1.8) multiple sequence alignment of nucleotide sequences of WsGT with *Lycium barbarum*

```
WsGT            ATGACTACTCACAAAGCTCACTGCTTAATCTTGCCATATCCAGGCCAAGGTCATGTCAAC 60
Lycium          ATGACTACTCACAGAGCTCATTGCTTAATCTTGCCATATCCAAGCCAAGGTCACATAAAC 60
                ************.****** ********************.********** .*.***

WsGT            CCAATGCTCCAATTCTCCAAACGTTTGCAATCCAAAAGTGTCAAAATCACAATAGCAACC 120
Lycium          CCGATGCTTCAATTCTCCAAACGTTTACAATCCAAAGGTGTCAAAATCACAATAGCAACC 120
                **.***** *****************.*********.**********************

WsGT            ACAAAATCCTTCTTGAAAAAAATGCAAAAATTACCAACTTCTATTTCAATCGAAGCCATC 180
Lycium          ACAAAATCCTTCTTGAAAACAATGCAAGAATTGACAACTTCGGTGTCAATTGAGGCAATC 180
                ******************.*******.****..******* .* ***** **.**.***

WsGT            TCTGATGGCTACGATGATGATGGCCTCGACCAAGCAAGATCTTATGCAGCCTATTTAACA 240
Lycium          TCCGATGGCTATGATGATGGTGGCCGCGATCAAGCAGGATCTTTCGTGGCCTACATAACA 240
                ** ******** *******.***** *** ******.******: * .***** :*****

WsGT            AGATTCAAAGAAGTTGGCTCAGATACTCTGTCTCAACTTATTGAAAAGTTAGCAAATTCT 300
Lycium          AGATTCAAAGAAGTTGGCTCGGATACTCTAGCTCAACTTATTAAAAAATTAGCAAATAGT 300
                ********************.********. ***********.****.*********: *

WsGT            GGGAGCCCTGTGAATTGCATAGTATATGATCCATTCCTTCCTTGGGTTGTTGAAGTGGCA 360
Lycium          GGGTGCCCAGTAAATTGCATAGTTTATGATCCATTCCTTCCTTGGGCAGTTGAAGTTGCA 360
                ***:****:**.***********:********************** :******** ***

WsGT            AAGAACTTTGGATTAGCTATTGCTGCATTTTTCACACAATCTTGTGCAGTGGACAACATT 420
Lycium          AAGGATTTTGGATTAGTTAGTGCTGCTTTTTTCACACAAAATTGTGCAGTGGATAACATT 420
                ***.* ********** ** ******:************:.************ ******

WsGT            TATTACCATGTACATAAAGGGGTACTAAAACTTCCTCCTACTCAAGTTGATGAAGAAATA 480
Lycium          TATTACCATGTACATAAAGGGGTTCTAAAGCTTCCACCTACTCAAGATGATGAAGAAATA 480
                ************************:*****.*****:**********:*************

WsGT            TTAATTCCTGGATTATCATATGCAATTGAGAGTTCAGACGTACCTAGTTTTGAGTCTACT 540
Lycium          TTAATTCCTGGATTTTCATGTCCAATTGAGAGTTCAGATGTACCTAGTTTTGTTATTAGC 540
                **************:****.* *************** *************: : **

WsGT            TCTGAACCAGATTTACTTGTTGAACTGTTGGCGAATCAGTTTTCAAATCTTGAGAAAACT 600
Lycium          CCTGAAGCAGCAAGAATACTTGATATGCTGGTGAATCAATTCTCAAATCTTGACAAAGTG 600
                 ***** ***.:: *.*: ****:.** *** ******.** *********** ***.

WsGT            GATTGGGTCCTAATCAACAGCTTCTATGAGTTGGAAAAACATGTAATTGATTGGATGTCC 660
Lycium          GATTGGGTACTAATCAACAGCTTCTATGAGTTGGAGAAAGAGGTAATTGATTGGATGTCC 660
                ********.*************************.*** * ******************

WsGT            AAGATTTATCCAATCAAGGCAATTGGACCAACAATACCATCCATGTACCTAGACAAGAGG 720
Lycium          AAGATTTATCCAATCAAGACAATTGGACCAACAATACCATCCATGTACCTAGACAATAGG 720
                ******************.*********************************** ***
```

```
WsGT            CTACCAGATGACAAAGAATACGGCCTTAGTATGTTCAAGCCAATAACAGATGCATGCATA 780
Lycium          CTACCGGATGACAAAGAGTATGGCCTTAGTGTCTTCAAGCCAATGACAAATGAATGCCTA 780
                *****.**********.** *********.* **********.***.***.****.**

WsGT            AATTGGCTAAACCACCAACCAATTAACTCAGTGTTATATGTATCATTTGGAAGCTTAGCC 840
Lycium          AATTGGTTAAATCATCAACTAATTAGCTCAGTGGTGTATGTATCATTTGGAAGTTTAGCC 840
                ****** **** ** **** *****.******* *.**************** ******

WsGT            AAACTAGAAGCTGAGCAAATGGAAGAACTGGCATGGGGTTTGAAGAATAGCAACAAGAAC 900
Lycium          AAAGTAGAAGTTGAACAAATGGAAGAATTGGCATGGGGTTTGAAAAATAGCAACAAGAAC 900
                *** ****** ***.*********** ****************.***************

WsGT            TTCTTGTGGGTAGTTAGGTCCGCTGAAGAACCCAAACTTCCGAAGAACTTCATAGAGGAA 960
Lycium          TTCTTGTGGGTTGTTAGATCCACTGAAGAATCCAAACTTCCTAAGAATTTTTTAGAGGAA 960
                **********:*****.***.******** ********** ***** ** :********

WsGT            TTACCAAGTG---------AAAAAGGCCTAGTGGTATCATGGTGTCCACAATTACAGGTG 1011
Lycium          TTAAAATTAGTAAGTGAGAATAAAGGCCTAGTTGTGTCATGGTGTCCGCAATTACAAGTT 1020
                ***..*: :*         *:*********** **.**********.********.**

WsGT            TTGGAACATGAATCAATAGGGTGTTTTATGACGCACTGTGGGTGGAATTCGACTTTGGAA 1071
Lycium          TTGGAACATAAATCAACAGGGTGTTTTCTGACTCACTGTGGATGGAATTCGACTTTGGAA 1080
                *********.****** **********.**** ********.******************

WsGT            GCAATTAGTTTGGGAGTGCCAATGGTGACATTGCCACAATGGTCAGATCAACCCACAAAT 1131
Lycium          GCGATTAGTTTGGGAGTACCAATGTTGACAATGCCACAATGGACAGATCAACCAACAAAT 1140
                **.**************.****** *****:**********:**********.******

WsGT            ACAAAGCTTGTGAAGGATGTTTGGGAGATGGGAGTTAGAGCCAAACAAGATGACAAAGGG 1191
Lycium          GCAAAGCTTGTAAAGGATGTTTGGGAGATGGGAGTTAGAGCCAAACAAGATGAAAAAGGG 1200
                .**********.****************************************.******

WsGT            CTAGTTAGACGAGAAGTTATCGAAGAGTGTATAAAACTAGTGATGGAAGAAGAGAAAGGA 1251
Lycium          ATAGTTAGAAGAGAAGTTATTGAAGAATGTATAAAGTTAGTGATGGAAGAAGAGAAAGGA 1260
                .********.********** *****.********. **********************

WsGT            AAAGTGATTAGGGAAAATGCTAAGAAATGGAAGGAATTGGCTAGAAATGCTGTGGATGAA 1311
Lycium          AAAATGATTAAGGAAAATGCACAGAAATGGAAGGAATTGGCTAGGAAAGCTGTGGATGAA 1320
                ***.******.*********:.**********************.**:************

WsGT            GGTGGAAGTTCAGATAAAAACATTGAAGAATTTGTTTCCAAGTTGGTTACTATTTCCTAA 1371
Lycium          GGAGGAAGTTCAGATAAAAATATAGAAGAATTTGTTTCCAAGTTGGTGACTATTTCCTCA 1380
                **:***************** **:********************** ********.*.*

WsGT            ------------
Lycium          TTAGGAAGCTGA 1392
```

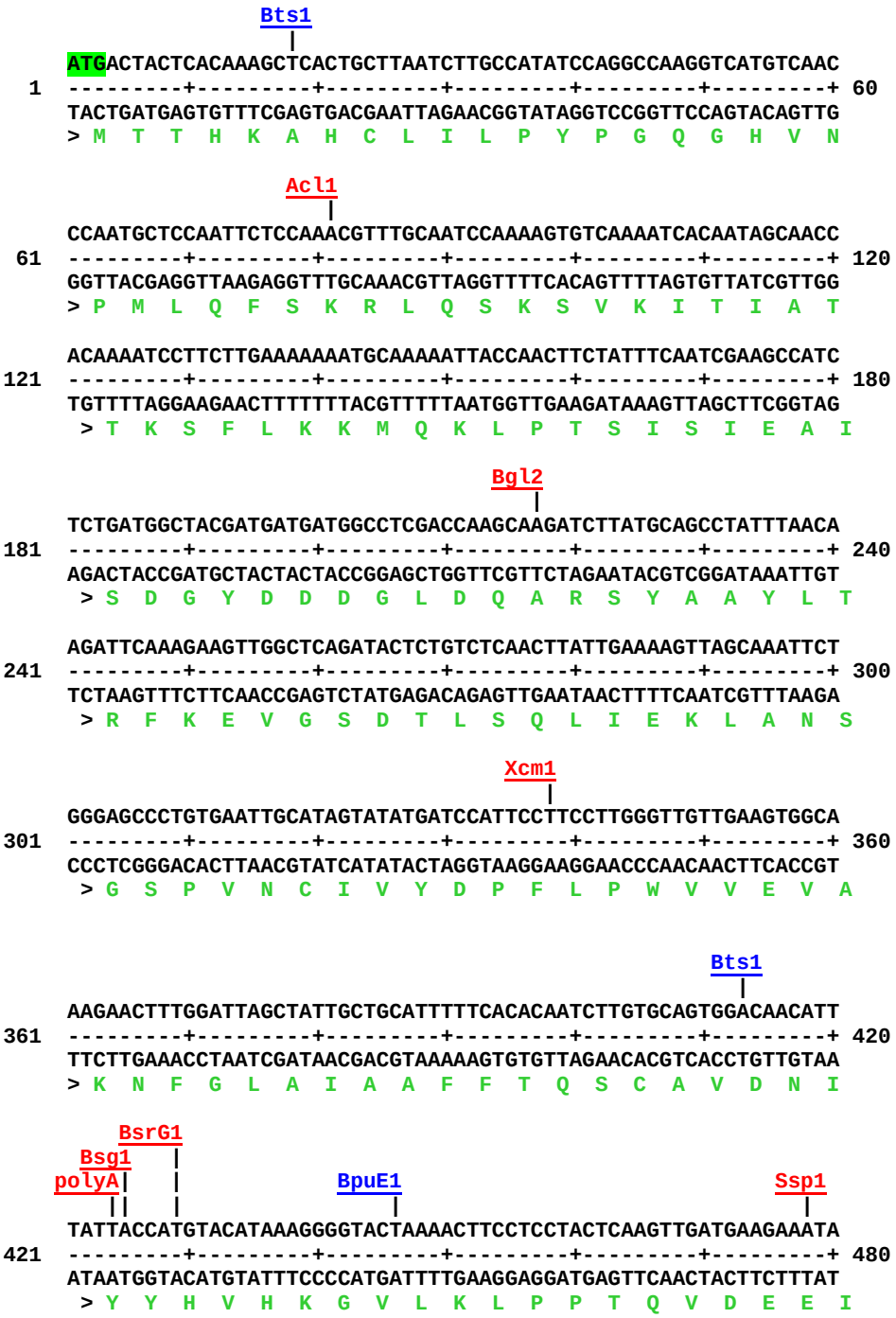
Figure 3.24 WsGT nucleotide sequence alignment with *Lycium barbarum*.

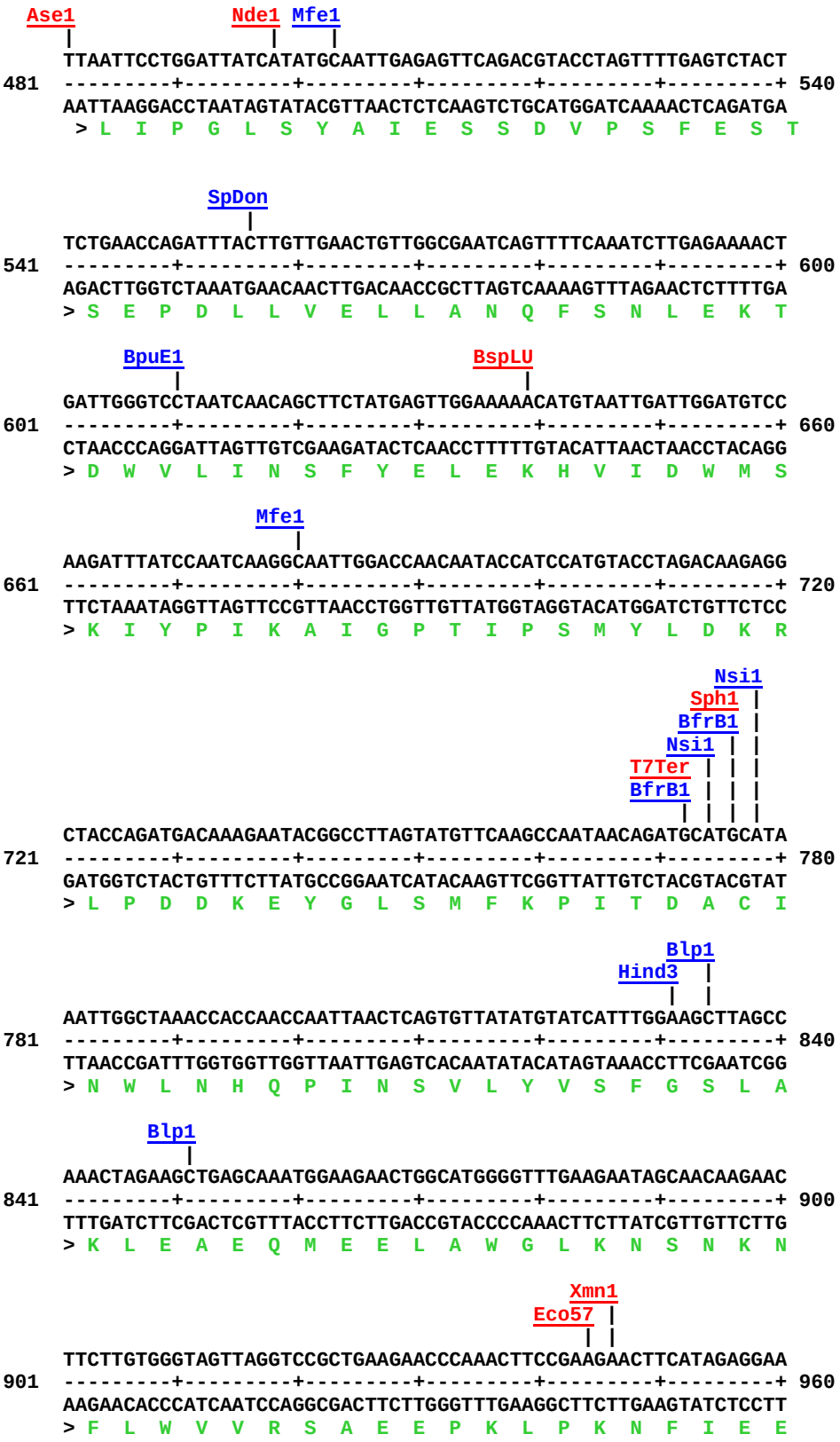
The nucleotide sequence of WsGT gene shows 87% homology with UDP glucose:glucosyltransferase from *Lycium barbarum* (AB360617) (Figure 3.24).

3.7.4 Restriction map of WsGT gene (FJ560880)

Restriction map of WsGT gene was generated by using an online bioinformatic tool, Webmap DNA. Map created at http://pga.mgh.harvard.edu/web_apps/web_map/start. Editor at

http://pga.mgh.harvard.edu/web_apps/dna_utilities.html Sites protected by *E. coli* methylation are not shown in the map.




Ase1
Nde1
Mfe1
TTAATTCCTGGATTATCATATGCAATTGAGAGTTCAGACGTACCTAGTTTTGAGTCTACT
481 ---------+---------+---------+---------+---------+---------+ 540
AATTAAGGACCTAATAGTATACGTTAACTCTCAAGTCTGCATGGATCAAAACTCAGATGA
> L I P G L S Y A I E S S D V P S F E S T
SpDon
TCTGAACCAGATTTACTTGTTGAACTGTTGGCGAATCAGTTTTCAAATCTTGAGAAAACT
541 ---------+---------+---------+---------+---------+---------+ 600
AGACTTGGTCTAAATGAACAACTTGACAACCGCTTAGTCAAAAGTTTAGAACTCTTTTGA
> S E P D L L V E L L A N Q F S N L E K T
BpuE1
BspLU
GATTGGGTCCTAATCAACAGCTTCTATGAGTTGGAAAAACATGTAATTGATTGGATGTCC
601 ---------+---------+---------+---------+---------+---------+ 660
CTAACCCAGGATTAGTTGTCGAAGATACTCAACCTTTTTGTACATTAACTAACCTACAGG
> D W V L I N S F Y E L E K H V I D W M S
Mfe1
AAGATTTATCCAATCAAGGCAATTGGACCAACAATACCATCCATGTACCTAGACAAGAGG
661 ---------+---------+---------+---------+---------+---------+ 720
TTCTAAATAGGTTAGTTCCGTTAACCTGGTTGTTATGGTAGGTACATGGATCTGTTCTCC
> K I Y P I K A I G P T I P S M Y L D K R
Nsi1
Sph1
BfrB1
Nsi1
T7Ter
BfrB1
CTACCAGATGACAAAGAATACGGCCTTAGTATGTTCAAGCCAATAACAGATGCATGCATA
721 ---------+---------+---------+---------+---------+---------+ 780
GATGGTCTACTGTTTCTTATGCCGGAATCATACAAGTTCGGTTATTGTCTACGTACGTAT
> L P D D K E Y G L S M F K P I T D A C I
Blp1
Hind3
AATTGGCTAAACCACCAACCAATTAACTCAGTGTTATATGTATCATTTGGAAGCTTAGCC
781 ---------+---------+---------+---------+---------+---------+ 840
TTAACCGATTTGGTGGTTGGTTAATTGAGTCACAATATACATAGTAAACCTTCGAATCGG
> N W L N H Q P I N S V L Y V S F G S L A
Blp1
AAACTAGAAGCTGAGCAAATGGAAGAACTGGCATGGGGTTTGAAGAATAGCAACAAGAAC
841 ---------+---------+---------+---------+---------+---------+ 900
TTTGATCTTCGACTCGTTTACCTTCTTGACCGTACCCCAAACTTCTTATCGTTGTTCTTG
> K L E A E Q M E E L A W G L K N S N K N
Xmn1
Eco57
TTCTTGTGGGTAGTTAGGTCCGCTGAAGAACCCAAACTTCCGAAGAACTTCATAGAGGAA
901 ---------+---------+---------+---------+---------+---------+ 960
AAGAACACCCATCAATCCAGGCGACTTCTTGGGTTTGAAGGCTTCTTGAAGTATCTCCTT
> F L W V V R S A E E P K L P K N F I E E

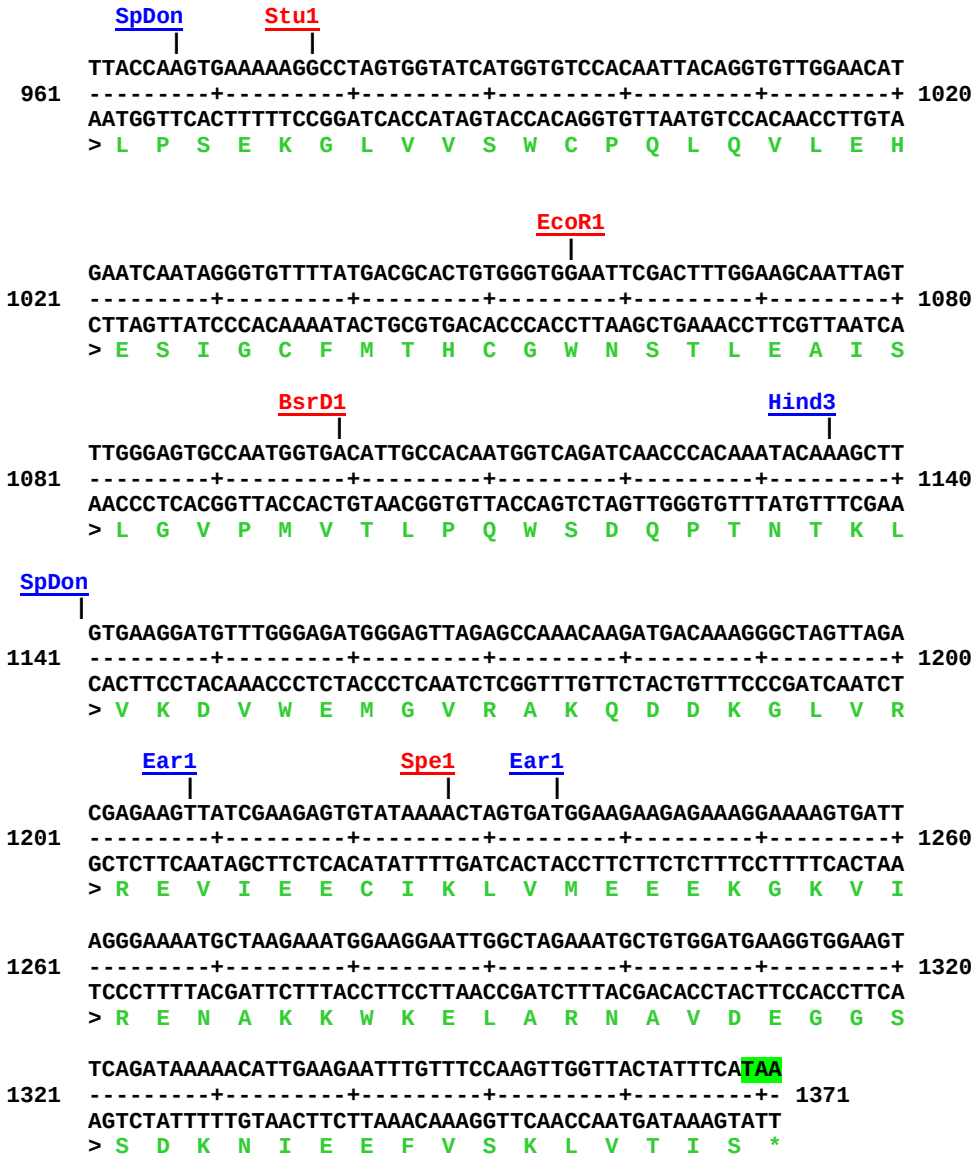


Figure 3.25: Restriction enzyme analysis of WsGT gene.

3.7.5 Amino acid composition, theoretical pI and molecular weight of WsGT gene (FJ560880)

The deduced amino acid sequences of WsGT gene show the coding region consists of 456 amino acids, with theoretical molecular weight of 51424.15 Daltons. The theoretical pI value of WsGT gene is 5.32 (http://www.expasy.org/cgi-bin/pi_tool). The amino acid composition of WsGT gene is given in table 3.1.1. In WsGT gene the total number of negatively charged

residues (Asp + Glu) equals 61 and positively charged residues (Arg + Lys) equal 48. The empirical formula of WsGT gene is $C_{2323}H_{3643}N_{589}O_{687}S_{19}$.

Table 3.1.1 Amino acid composition of WsGT (FJ560880)

Amino acid	Total number	Percentage
Ala (A)	25	5.5%
Arg (R)	10	2.2%
Asn (N)	20	4.4%
Asp (D)	23	5.0%
Cys (C)	8	1.8%
Gln (Q)	16	3.5%
Glu (E)	38	8.3%
Gly (G)	22	4.8%
His (H)	9	2.0%
Ile (I)	30	6.6%
Leu (L)	47	10.3%
Lys (K)	38	8.3%
Met (M)	11	2.4%
Phe (F)	16	3.5%
Pro (P)	25	5.5%
Ser (S)	39	8.6%
Thr (T)	20	4.4%
Trp (W)	11	2.4%
Tyr (Y)	13	2.9%
Val (V)	35	7.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

3.7.6 Hydropathy index of the WsGT amino acids (FJ560880)

The hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain (Jack Kyte and Russell Doolittle, 1982). The larger the number is, the more hydrophobic the amino acid. The most hydrophobic amino acids are isoleucine (4.5) and valine (4.2). The most hydrophilic ones are arginine (-4.5) and lysine (-3.9). This is very important in protein structure; hydrophobic amino acids tend to be internal (with regard to the protein's native 3-dimensional structure) while hydrophilic amino acids are more commonly found towards the protein surface. The amino acid sequence of WsGT

(FJ560880) is analyzed using Kyte-Doolittle Hydropathy plot at http://gcat.davidson.edu/rakarnik/kd.cgi (Figure 3.26) window size 9. When the window size is 9, strong negative peaks indicate possible surface regions of globular proteins.

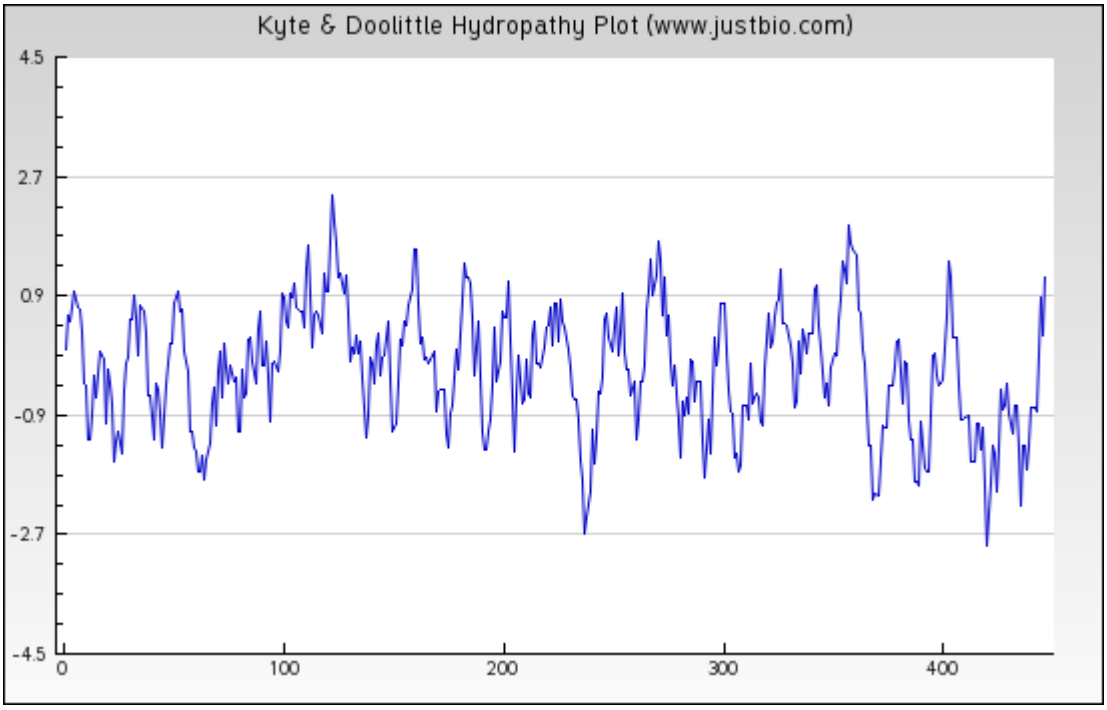


Figure 3.26 Hydropathic plot. A Kyte-Doolittle Hydropathy for WsGT (FJ560880). Window size of 9 suggested a good value for finding putative surface-exposed regions. When the window size is **9**, strong negative peaks indicate possible surface regions of globular proteins.

3.7.7 Codon usage of the WsGT gene (FJ560880)

The GC% and codon usage of the cDNA clone WsGT (FJ560880) was calculated using online software (www.justbio.com). The GC content of the cDNA clone WsGT (FJ560880) is 39.13%. The codon usage for WsGT cDNA clone (FJ560880) is given in Table 3.1.2. It is expressed as % of total codons. Standard genetic codes were used for determining codon usage.

Table 3.1.2 Codon usage of WsGT (FJ560880)

codon	mean	codon	mean	codon	mean	codon	Mean
UUU ( )	2.02	UCU ( )	0.45	UAU ( )	1.79	UGU ( )	2.69
UUC ( )	2.24	UCC ( )	2.02	UAC ( )	1.12	UGC ( )	1.79
UUA ( )	0.90	UCA ( )	1.12	UAA ( )	0.90	UGA ( )	5.38
UUG ( )	1.57	UCG ( )	0.00	UAG ( )	2.02	UGG ( )	2.24
CUU ( )	1.79	CCU ( )	1.35	CAU ( )	1.57	CGU ( )	0.22
CUC ( )	0.90	CCC ( )	0.45	CAC ( )	0.90	CGC ( )	0.22
CUA ( )	0.67	CCA ( )	1.35	CAA ( )	4.26	CGA ( )	0.90
CUG ( )	0.22	CCG ( )	0.00	CAG ( )	0.22	CGG ( )	0.22
AUU ( )	3.14	ACU ( )	2.02	AAU ( )	4.26	AGU ( )	3.14
AUC ( )	2.02	ACC ( )	2.24	AAC ( )	1.35	AGC ( )	2.24
AUA ( )	1.35	ACA ( )	1.79	AAA ( )	4.26	AGA ( )	4.71
AUG ( )	1.12	ACG ( )	0.45	AAG ( )	1.35	AGG ( )	1.79
GUU ( )	1.79	GCU ( )	1.12	GAU ( )	1.12	GGU ( )	1.79
GUC ( )	1.12	GCC ( )	0.90	GAC ( )	0.67	GGC ( )	1.12
GUA ( )	0.22	GCA ( )	0.90	GAA ( )	3.36	GGA ( )	2.69
GUG ( )	1.12	GCG ( )	0.00	GAG ( )	0.67	GGG ( )	0.67

3.7.8 Analysis of amino acid sequence of the WsGT genes (FJ560880) for their conserved domains

The amino acid sequences of WsGT (FJ560880) analyzed for their sequence similarity. It matched with glycosyltransferases which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. The acceptor molecule can be a lipid, a protein, a heterocyclic compound, or another carbohydrate residue. The structures of the formed glycoconjugates are extremely diverse, reflecting a wide range of biological functions. The members of this family share a common GTB topology, one of the two protein topologies observed for nucleotide-sugar-dependent glycosyltransferases. GTB proteins have distinct N- and C- terminal domains each containing a typical Rossmann fold. The two domains have high structural homology despite minimal sequence homology. The large cleft that separates the two domains includes the catalytic center and permits a high degree of flexibility. The conserved domains of WsGT (Figure 3.27) are shown below.

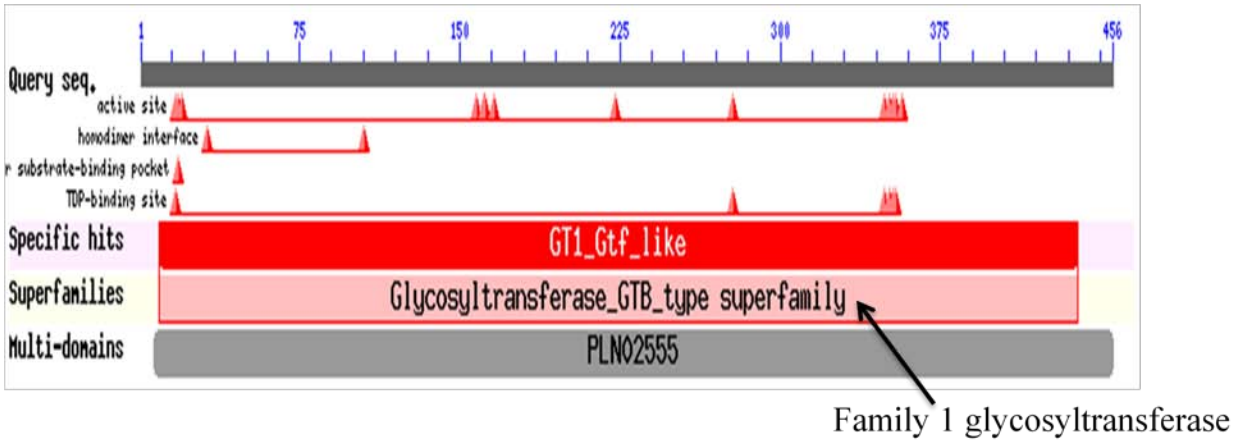


Figure 3.27 Conserved domains of WsGT genes.

3.7.9 Phylogenetic analysis of WsGT (FJ560880) gene with other reported glycosyltransferases

Phylogenetic analysis was done for WsGT gene using various reported protein sequences of GT, which were retrieved from the GenBank database. Multiple alignments of the predicted protein coding sequences were performed using Clustal X. The phylogenetic tree was constructed by neighbor joining methods using Mega 4.0 software program (Figure 3.28). It is clear from the phylogenetic tree that, WsGT is evolutionarily most similar with *Lycium barbarum* (AB360617) (UDP-glucose: glucosyltransferase), *Lycopersicon esculentum* (AJ889012) (UDP-xylose phenolic Glycosyltransferase) and *Nicotiana tabacum* (AAF61647) (UDP-glucose: salicylic acid glucosyltransferase).

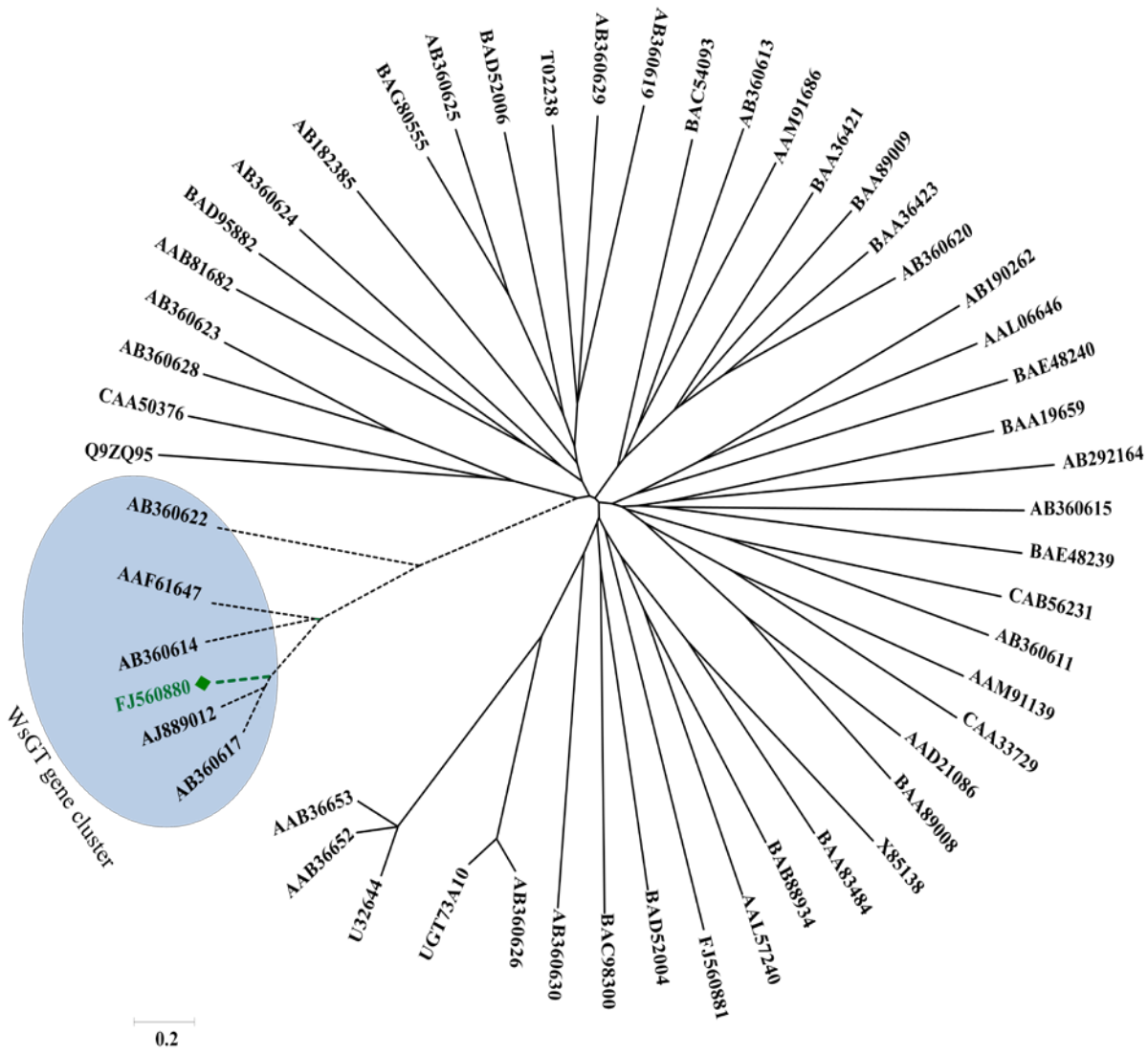


Figure 3.28 Phylogenetic tree of WsGT gene. WsGT (FJ560880) is marked with colour dot. The sequences used for the alignment were primarly those of glycosyltransferases that have been functionally characterized. The phylogenetic tree was constructed by neighbor joining methods using Mega 4.0 software program. The names and accession numbers of glycosyltransferases used for the alignment are as follows: *Lycopersicon esculentum* (AJ889012), *Lycium barbarum* (BAG80555), *L. barbarum* (AB360625), *L. barbarum* (AB360624), *Nicotiana tabacum* (TO2238), *L. barbarum* (AB360629), *L. barbarum* (AB360619), *Dianthus caryophyllus* (BAD52004), *Scutellaria baicalensis* (BAC98300), *Lycopersicum esculentum* (X85138), *S. baicalenis* (BAA88934), *N. tabacum* (BAB88934), *Dorotheanthus bellidiformis* (AAL57240), *D. bellidiformis* (BAD52006), *L. barbarum*

(AB360626), *L. barbarum* (UGT73A10), *N. tabacum* (U32644), *N. tabacum* (AAB36652), *N. tabacum* (AAB36653), *L. barbarum* (AB360613), *L. barbarum* (AB360620), *Petunia x hybrid* (BAA89009), *Torenia hybrid* (BAC54093), *Arabidopsis thaliana* (AAM91686), *A. thaliana* (Q92Q95), *Petunia x hybrid* (CAA50376), *L. barbarum* (ABB360628), *L. barbarum* (AB360623), *Perilla frutescins* (BAA36421), *Verbena x hybrid* (BAA36423), *Antirrhinum majus* (BAE48239), *A. thaliana* (AAM91139), *Hordum vulgare* (CAA33729), *Forsythia x intermedia* (AAD21086), *Petunia x hybrid* (BAA89008), *D. bellidiformis* (CAB56231), *L. barbarum* (AB360611), *Perilla frutescins* (BAA19659), *Glycine max* (AB292164), *L. barbarum* (AB360615), *Bellis peremnis* (AB190262), *Citrus maxima* (AAL06646), *Vitis vinefera* (AAB81682), *Ipomoea purpurea* (BAD95882), *L. barbarum* (AB360630), *Linaria vulgaris* (BAE48240), *L. barbarum* (AB360622), *N. tabacum* (AAF61647), *L. barbarum* (AB360614), *L. barbarum* (AB360617), *Scutellaria baicalenis* (BAA83484).

3.8 Isolation, Cloning and Characterization of Saponin related GT gene

Glycosyltransferase gene nucleotide sequences (Saponin related) available at NCBI GenBank database were aligned using Clustal W (1.8) multiple sequence alignment software. Forward and reverse primers were designed from regions showing highest homology. The conserved regions selected for primer designing has been shown high-lighted in yellow colour (Figure 3.29).

```
AB182386        ------------------------------------------------------------
AB182385        ATGGATAACGGCAGCAATCAACTACATGTCCTCTTCCTTCCTTACTTCGCCACTGGTCAT 60
DQ218277        ATGGATAACGGGAGCAAGCAACTACACGTCCTCTTCCTTCCTTACTTCGCCACTGGTCAT 60


AB182386        ------------------------------------------------------------
AB182385        ATCATCCCATTAGTTAACGCCGCCAGGCTATTCGTCTTTCACGCCGGCGTCAAAGTCACA 120
DQ218277        ATCATTCCATTAGTTAACGCTGCCAGGCTATTCGCCTCCCGTGGCGGTGTCAAAGTTACC 120


AB182386        ------------------------------------------------------------
AB182385        ATCCTCACTACCCACCACAACGCCTCCCTATTCCGATCTACTATTGACAACGACGTTGAA 180
DQ218277        ATTCTCACTACCCACCACAATGCTTCCCTCTTCCGATCTTCTATT-------------- 165


AB182386        --ATGGAATTCAATACTCG-AATCCATACTCTTAGGTTCCCCTCCACTGAAGTTGGGTTG 57
AB182385        GACGGCCATTCCGTAATCTCTATCCATACTCTTAGGTTCCCCTCCACTGAAGTTGGGTTG 240
DQ218277        ---GACAATTCCCTAATCTCTATCGCTACTCTTAAGTTCCCTTCCACTGAAGTTGGGTTG 222
                    . .****. **.**  :*** .********.****** ******************

AB182386        CCTGAAGGGATCGAGAACTTCAGCTCCGCCTCTTCACCTGAACTCGCGGGCAAAGTATTT 117
AB182385        CCTGAAGGGATCGAGAACTTCAGCTCTGCCTCTTCACCTGAACTCGCGGGCAAAGTATTT 300
DQ218277        CCTGAAGGGATCGAAAATTTCAGCTCCGCCTCTTCAACTGAAATCGCGAGCAAATTATTT 282
                **************.** ******** *********.*****.*****.***** *****

AB182386        TACGCCACTTATCTTCTACAGAAACCAATGGAAGATAAAATTCGTGAAATCCATCCTGAT 177
AB182385        TACGCCATTTATCTTCTACAGAAACCAATGGAAGATAAAATTCGTGAAATCCATCCGGAT 360
DQ218277        GGCGGCATTTATCTTCTGCAGAAACCAATGGAAGATAAAATTCGTGAAATCCATCCTGAT 342
                 .** ** *********.************************************ ***

AB182386        TGTATTTTCTCTGATATGTACTTTCCATGGACCGTCGATATTGCACTGGAGCTCAACATC 237
AB182385        TGTATTTTCTCTGATATGTACCTCCCATGGACTGTCAATATTGCCTTGGAGCTCAAAATC 420
DQ218277        TGTATCTTCTCTGATATGTATTTCCCATGGACTGTCGATATTGCACTGGAGCTCAAAATC 402
                ***** **************  * ******** ***.*******. **********.***

AB182386        CCCAGGCTATTGTTCAACCAGTCCAGCTACATGTACAATTCCATCCTGCACAAGCTTAGG 297
AB182385        CCCAGGCTATTGTTCAACCAGTCCAGCTACATGTACAATTCCATTCTGTACAATCTGAGG 480
DQ218277        CCCAGGCTATTGTTCAACCAATCTAGCTACATGTACAATTCCATTCTGTACAATCTTAGG 462
                ********************.** ******************* *** **** ** ***

AB182386        TTTTACAAACCTCACAAATCCAAAACTACTACTAGTAATGATGATATCTCAGTTCCTGGT 357
AB182385        CTTTACAAACCTCACAAATCCAAAACTATTACTAGTACTGATAGTATCTCAGTTCCTGGT 540
DQ218277        CTTTACAAACCTCACGAAT---------ATTCCAAAAGTAGTAATTTCTCGGTTCCGGGT 513
                 **************.***          *:* *.:* *..*..*:****.***** ***


AB182386        TTACCAGATAAGATCGAGTTCAAGCTAACGCAACTTACAGATGATCTGATAAAGCCCGAG 417
AB182385        TTACCCGATAAGATCGAGTTCAAGCTATCGCAACTTACAGATGATCTGATAAAGCCTGAG 600
DQ218277        TTACCTGATAAGATCGAGTTCAATCTATCGCAACTTACAGACGATCTGATAAAGCCTGCA 573
                ***** ***************** ***:************* ************* *..
```

```
AB182386        GATGAGAAGAATGCTTTTGACGAGTTGCTCGATCGAACCAGAGAATCCGAGGATCGAAGC 477
AB182385        GATGAGAAGAATGCTTTTGACGAATTGCTCGATCGAACTAGAGAATCCGAGGATCGAAGC 660
DQ218277        GATGAGAGGAATGGTTTTGATGAATTGCTCGATCGAACCAGAGAATCTGAGGATCAAAGC 633
                *******.***** ****** **.************** ******** *******.****

AB182386        TACGGCATTGTTCACGACACTTTTTACGAGCTAGAACCTGCCTACGCTGACTACTACCAG 537
AB182385        TACGGCATCGTTCACGATACTTTTTACGAGCTAGAACCTGCCTACGCTGATTATTATCAG 720
DQ218277        TACGGTATCGTTCATGATACTTTTTACGAACTAGAACCTGCCTACGCTGACTACTATCAG 693
                ***** ** ***** ** **********.******************** ** ** ***

AB182386        AAGGTGAAGAAAACCAAATGTTGGCAAATTGGTCCCATCTCCCATTTTTCTTCCACGTTA 597
AB182385        AAGGTGAAGAAAACTAAATGTTGGCAAATTGGTCCCATCTCCCATTTTTCTTCCAAGTTA 780
DQ218277        AAGATGAAGAAAACCAAATGTTGGCAAATTGGTCCCATTTCCTATTTTTCTTCCAAATTA 753
                ***.********** ********************** *** ************..***

AB182386        CTCCGAAGAAGAAAAGAGCTGGTTAATGCTGTCGATGAAAGTAACTCATGTGCC---ATT 654
AB182385        TTCCG---TAGAAAAGAGCTGATTAATGCTGTCGATGAAAGTAACTCATGTGCC---ATT 834
DQ218277        TTCCG---AAGAAAAGATCTGATTAATTCTTTTGATGAAAGTAACTCATCTGCCGCTGTT 810
                 ****   :******** ***.***** ** * *************** ****   .**

AB182386        TCAGAGTGGTTGAATGAGCAGAAGCATAAATCGGTCCTTTACATCTCTTTCGGGAGCGTA 714
AB182385        GTAGAGTGGTTGAACGAGCAGGAGCATAAATCAGTCCTCTATGTCTCTTTCGGGAGCGTA 894
DQ218277        GTAGAGTGGTTGAATAAACAGAAGCACAAATCGGTCCTCTACGTCTCTTTCGGGAGCACA 870
                  ************ .*.***.**** *****.***** ** .**************. *

AB182386        GTTAAATTCCCAGACGCACAACTTACTGAAATCGCAAAAGCTCTAGAAGCTTCAAGCATC 774
AB182385        GTTAGATTCCCAGAAGCCCAACTCACTGAAATCGCAAAAGCTCTAGAAGCTTCGAGCATC 954
DQ218277        GTTAAATTCCCAGAGGAGCAACTCGCTGAAATCGCAAAAGCTCTAGAAGCTTCTACCGTC 930
                ****.********* *. ***** .*************************** * *.**

AB182386        CCTTTCATTTGGGTAGTGAGGAAGGACCAATCAGCAGAAACCACGTGGTTGCCGAAGGAG 834
AB182385        CCTTTCATTTGGGTAGTGAAGAAGGACCAATCGGCAGAAACCACGTGTTTGCTGGAGGAG 1014
DQ218277        CCTTTCATTTGGGTAGTGAAGGAGGACCAATCAGCAAAAACCACCTGGTTACCGGAGAGT 990
                ******************.*.**********.***.******* ** **.* *.**..

AB182386        AACAAATTGAAGAAAAAGGGTCTGATTATTAGAGGGTGGGCGCCGCAAGTAACCATCTTG 894
AB182385        GAAAAATTGAAGAACAAGGGTCTGATTATTAGAGGGTGGGCGCCGCAACTAACCATCTTG 1074
DQ218277        TTGTTCGATGAGAAAAAAGGTCTGATTATTAAAGGGTGGGCTCCGCAACTAACCATCTTA 1050
                 : ::. : .****.**.*************.********* ****** **********.

AB182386        GACCATTCAGCCGTGGGAGGATTCATGACACACTGCGGATGGAATTCAATACTCGAATCC 954
AB182385        GACCATTCAGCCGTGGGAGGATTCATGACACACTGCGGTTGGAATTCCATACTCGAAGCC 1134
DQ218277        GATCATTCAGCAGTAGGAGGATTCATGACACACTGTGGATGGAATTCGGTGCTTGAAGCT 1110
                ** ********.**.******************** **:******** .*.** *** *

AB182386        ATCACGGCTGGCGTCCCGTTGGTGACGTGGCCAGTGTTCGCGGAGCAATTCTACAATGAA 1014
AB182385        ATCATCGCCGGCGTCCCGTTGGTGACGTGGCCGGTGTTCGCGGAGCAATTCTACAATGAA 1194
DQ218277        ATCATCGCTGGGGTGCCGTTGGTGACGTGGCCAGTGTTCGCTGAACAATTCTACAATGAA 1170
                ****  ** ** ** ****************.******** **.***************

AB182386        AAGCTAGTGGAGGTTATGGGATTAGGAGTGAAAGTGGGAGCAGAAGTACATATCTCGGAT 1074
AB182385        AAGCTTGTGGAGGTAATGGGGTTAGGAGTGAAGGTGGGGGCAGAAGTACATGAGTCGAAT 1254
DQ218277        AAACTTGTGGAGGTTATGGAGCTAGGAGTGAAAGTAGGGGCAGAAGTACATAACTCCGAC 1230
                **.**:********:****.. **********.**.**.************.: ** .*

AB182386        G---GCCTTGAGTTCTCGAGCCCTGTGATAGAGAGCGAAAAGATCAAAGAAGCAATAGAG 1131
AB182385        GGAGGTGTAGAGATCTCGAGCCTTGTGATAGAGAGCGAAAAGATCAAAGAAGCAATAGAG 1314
DQ218277        GGATGTGTTGAGATATCGAGCCCTGTGTTAAGGAGCGAAAAGATAAAAGAAGCAATTGAG 1290
                *   *  *:***:*.******* ****:**..************.***********:***

AB182386        AAATTAATGGATGACTCAAACGAAAGCCAGAAAATCAGGGAGAAAGCAATGGCTACGAGT 1191
AB182385        AAATTAATGGATGACTCAAAGGAAAGCCAGAAAATTAGAGAGAAAGTAATAGGTATGAGC 1374
DQ218277        AGGTTAATG-----------GAAAGTCAGAAAATAAGAGAGAAAGCAGTGAGTATGAGT 1338
                *..******           ***** ******** **.******* *.*.. ** ***
```

```
AB182386        GAGATGGCGAAAAGTGCAGTAGGAGAAGGCGGATCTTCATGGAACAATCTCACAGGTCTG 1251
AB182385        GAAATGGCTAAAAATGCAGTCGAAGAAGGTGGATCTTCATGGAACAATCTCACTGCACTT 1434
DQ218277        AAGATGGCTAAAAATGCAGTGGAAGAAGGTGGATCTTCATGGAACAATCTTACCGCACTT 1398
                .*.***** ****.****** *.****** ******************** ** * :**

AB182386        ATCAATGATATTAAGAATTTTACTTCTTCCTAG----------------- 1284
AB182385        ATTGATGATATCAAGAATTTTACTTCGACGACGAA---------CGTTTAA 1476
DQ218277        ATAGATGATATCAAGAATTTTACTTCTTCTTCATTGAAGATCATGGATTAA 1449
                ** .******* ************** :* :..
```

Figure 3.29 Multiple sequence alignment of Glycosyltransferase (*GT*).

Nucleotide sequences of *Solanaceae* member *Solanum tuberosum* (DQ218277), *Solanum aculeatissimum* (AB182386, AB18277). Highlighted are conserved regions considered for primer synthesis.

3.8.1 PCR amplification of partial cDNA fragment of *GT* gene from *W. somnifera*

Among the several set of primers one primer set i.e. WsSAPF1 – WsSAPR1 (as mentioned in Figure 3.29) is used to amplify ~600 bp fragment of *GT* gene using cDNA 1st strand as a template (Figure 3.30A). The fragment is cloned in pGEM-T easy vector and confirmed by restriction analysis (Figure 3.30B) and sequencing (Figure 3.31). This partial clone will be referred to as WsSAPGT in the subsequent discussion. The partial clone of WsSAPGT showed maximum homology (73%) at amino acid level with *Solanum aculeatissimum* of *Solanaceae* family. The reported *S. aculeatissimum* glycosyltransferase glycosylates steroidal sapogenins and steroidal alkaloids.

3.8.2 Primers used for PCR amplification

WsSAPF1 5` ATCTCAGTTCCTGGTTTACCC 3`

WsSAPR1 5` TTCGAGTATGGAATTCCAACCGCA 3`

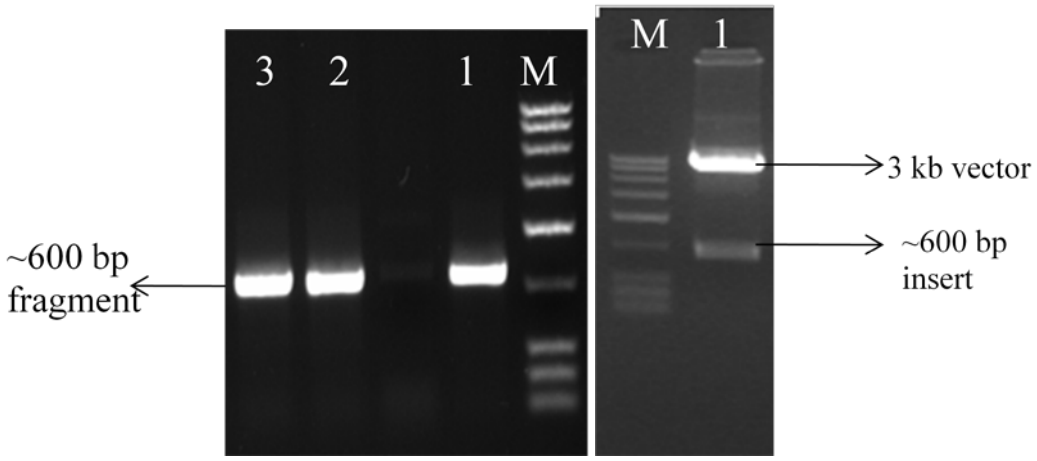


Figure A Figure B

Figure 3.30 Figure A: PCR amplicon of ~600 bp of WsSAPGT gene on 1% agarose gel. Lane 1, 2 & 3 amplified product and lane M- 1kb marker. **Figure B:** pGEM-T digestion with *EcoR1* enzyme; Lane M- marker, Lane 1 & 2 *Eco RI* digested recombinant pGEMT-Easy vector releasing ~600bp insert.

```
  1 ATCTCAGTTC CTGGTTTACC CAATAAGATC CACTTCAAGC GTTCCCAACT
 51 TACAGAGGAT CTTATAAAGT CTGCCGATGA GAGGACTGCT CATGACCAAC
101 TGCTCGATCA AATAAGAGAT TCCGAGGATC GAAGCTATGG CATCGTTCAC
151 GACACATTTT ATGAGCTGGA ACCTGCCTAT GCTGACTACT ATCAAAAGAT
201 GAAGAAAACA AAGTGTTGGC AAATTGGTCC TATCTCTCAT TTCTCTTCCA
251 AATTAATCCG AAGAAAAGAG CTAATTGATG CCTCCGATGA CGTTAACTCG
301 TGTGAGATTG ATAAATGGTT GAACAAGCAG GGGCAACGGT CGGTCCTGTA
351 CATCTCTTTC GGGAGCTTTG TCAGATTCCC AGAGGACCAA CTCACGGAAA
401 TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC CTTTTGTTTG GGTAATGAGG
451 AAGGACCAAT CAGCACAAAC CACGTGGTTA CCCGATGGTT TCAAGGAAAA
501 AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG GTGGGCACCG CAACAGACCA
551 TCTTGGACCA TTCAGCGGTT GGAGGATTCC TCACTCACTG CGGTTGGAAT
601 TCCATACTCG AA
```

Figure 3.31 ~600 bp partial WsSAPGT nucleotide sequence. Forward and reverse primers are highlighted in blue colour.

3.8.3 Rapid Amplification of cDNA Ends (RACE) (Invitrogen)

RACE is an important tool to obtain the UTRs (Un-Translated Regions) of a particular gene and is also useful to obtain full length gene of unknown partial fragment.

3.8.4 Primer designing for 3`RACE

Based on the 600 bp partial WsSAPGT sequence, one 3` Gene specific primer and one 3'Nested GSP was designed for 3` RACE reaction as shown below highlighted in colour and underlined.

3` Gene specific primer  3` Nested Gene specific primer

```
  1 ATCTCAGTTC CTGGTTTACC CAATAAGATC CACTTCAAGC GTTCCCAACT
 51 TACAGAGGAT CTTATAAAGT CTGCCGATGA GAGGACTGCT CATGACCAAC
101 TGCTCGATCA AATAAGAGAT TCCGAGGATC GAAGCTATGG CATCGTTCAC
151 GACACATTTT ATGAGCTGGA ACCTGCCTAT GCTGACTACT ATCAAAAGAT
201 GAAGAAAACA AAGTGTTGGC AAATTGGTCC TATCTCTCAT TTCTCTTCCA
251 AATTAATCCG AAGAAAAGAG CTAATTGATG CCTCCGATGA CGTTAACTCG
301 TGTGAGATTG ATAAATGGTT GAACAAGCAG GGGCAACGGT CGGTCCTGTA
351 CATCTCTTTC GGGAGCTTTG TCAGATTCCC AGAGGACCAA CTCACGGAAA
401 TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC CTTTTGTTTG GGTAATGAGG
451 AAGGACCAAT CAGCACAAAC CACGTGGTTA CCCGATGGTT TCAAGGAAAA
501 AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG GTGGGCACCG CAACAGACCA
551 TCTTGGACCA TTCAGCGGTT GGAGGATTCC TCACTCACTG CGGTTGGAAT
601 TCCATACTCG AA
```

Figure 3.32: Gene specific primer designing for 3` RACE.

3.8.5 Gene specific primers for 3` RACE

SAPRACE F1 5` CAGTTCCTGGTTTACCCAATAAGA 3`

SAPRACE F2 5` CACTTCAAGCGTTCCCAACT TACAGA 3`

3.8.6: 3' RACE reaction

To isolate the 3' end sequence of the WsSAPGT gene, 3' RACE reaction was performed (as described in Chapter 2, Section 2.9). For WsSAPGT gene, 3'RACE was performed using forward Gene specific primer SAPRACE F1 (5` CAGTTCCTGGTTTACCCAATAAGA 3`) and a reverse 3'GeneRacer primer provided with the kit. 3' RACE ready cDNA was used as the template. The RACE reaction resulted in an amplified product of approximately 1250 bp; which was further re-amplified using 3' Nested SAPRACE F2 (5` CACTTCAAGCGTTCCCAACTTACAGA 3') and 3'Nested GeneRacer primer (3'NGRP) provided with kit. The resultant PCR product (Figure 3.33A) was cloned into pGEMT-Easy

vector (Promega, USA) and the recombinant construct was confirmed by *Eco* RI restriction analysis (Figure 3.33B) and by sequencing (Figure 3.34).

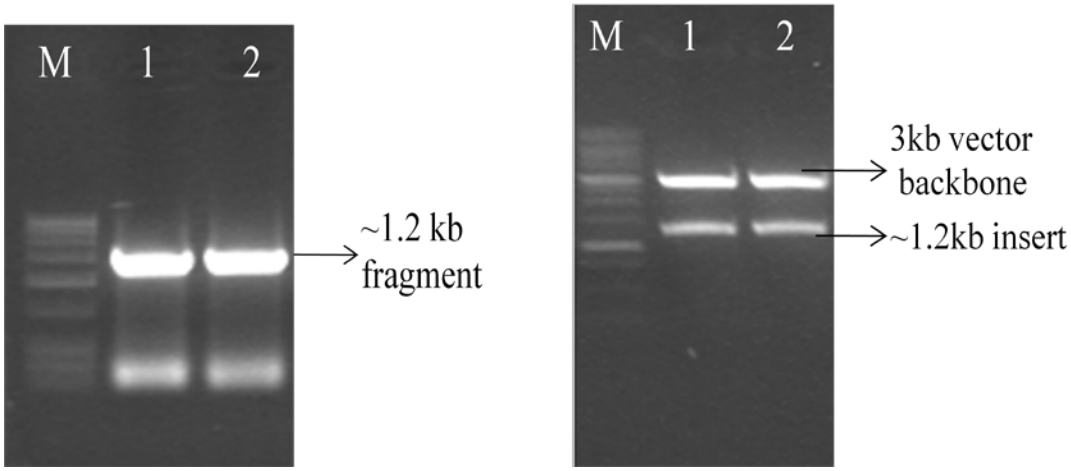


Figure A **Figure B**

Figure 3.33 Figure A: 1% agarose gel showing 3'RACE product of WsSAPGT gene. Lane M- marker, lane 1 & 2 is ~1.2 kb 3'RACE product. **Figure B** Lane M- marker, Lane 1 & 2 *Eco RI* digested recombinant pGEMT-Easy vector and ~ 1.2 kb insert.

```
   1 CACTTCAAGC GTTCCCAACT TACAGAGGAT CTCATAAAGT CTGCCGATGA
  51 GAGGACTGCT TATGACCAAC TGCTCGATCA AATAAGAGAT TCCGAGGATC
 101 GAAGCTATGG CATCGTTCAC GACACATTTT ATGAGCTGGA ACCTGCCTAT
 151 GCTGACTACT ATCAGAAGAT GAAGAAAACA AAGTGTTGGC AAATTGGTCC
 201 TATCTCTCAT TTCTCTTCCA AATTAATCCG AAGAAAAGAG CTAATTGATG
 251 CCTCCGATGA CGTTAACTCG TGTGAGATTG ATAAATGGTT GAACAAGCAG
 301 GGGCAACGGT CGGTCCTCTA CATCTCTTTC GGGAGCCTTG TCAGATTCCC
 351 AGAGGACCAA CTCACGGAAA TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC
 401 CTTTTGTTTG GGTAATGAGG AAGGACCAAT CAGCACAAAC CACGTGGTTA
 451 CCCGATGGTT TCAAGGAAAA AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG
 501 GTGGGCACCG CAACAGACCA TCTTGGACCA TTCAGCGGTT GGAGGATTCA
 551 TCACTCACTG TGGTTGGAAT TCCGTACTCG AAGCTATCGT TGCTGGCGTG
 601 CCGATGCTGA CATGGCCACT ATTCGCAGAT CAATTCTACG ACGAGAAGCT
 651 CGTAGAGGTT TTGGGTTTGG GAGTGAAAGT GGGGTCAGAG GTATGTAGTT
 701 TGGTAGGTGT TGACATAATG GGTCCTATAA TAGGGAGTGA AAAGATTAAA
 751 GAAGCAATAC ACCAATTGAT GAGTGGTGGT TCCAAGGAAA GGGAAAATAT
 801 TAGGGAGAAA TCAATGGTTA TGAGTAAGAT GGCGAAAAAG GCAACCGAAG
 851 GAAATGGATT TTCATGCAAC AGTCTCACAG CGCTCATTGA TGATATCAAG
 901 AATTTTACTT TTATCGAAAA AGAAAGTGGT TAAGGGATGA CAACCAGGAG
 951 GATGGGATTG GAAGCAACCA TCGGTTAGTC ATCAACAATA TAGCTTAAAT
1001 GATCGTTGGT TTGGTTTAGT TTTCCTCCTT TTCGTTATTT TTTTTTGTCA
1051 TGTTACCATG CACTAATGCT GTGTTTGTGC ACAACTACTC CATAAACAAG
1101 CATGTTATTG TATGTTGTTA AAATCTATGT ATATTGTGGG AACTAGCTTC
```

```
1151 CCTCCAGTTG CAATAAATCA AAGACATGAA AAATCAATCC TCATAGAAAT
1201 AAATAGATTA AATTTTGGAA AAAGAAAAAA AAAAAAAAAA AAAAAAAAAC
1251 ACTGTCATGC CGTTACGTAG CG
```

Figure 3.34: WsSAPGT partial 1272 bp sequence.

Partial 1271 bp nucleotide sequence of WsSAPGT gene with 3` UTR 287 nucleotides highlighted in blue colour after stop codon TAA, 3` Nested Gene Specific primer shown in Pink colour, 3` Nested Gene Racer primer shown in green colour provided in Kit.

3.8.7 Primer designing for 5`RACE

Based on the 612 bp partial WsSAPGT gene, one 5'Gene specific primer (GSP) and one 5'Nested GSP primer were designed for 5'RACE as shown below highlighted and underlined with colour.

```
   1 CACTTCAAGC GTTCCCAACT TACAGAGGAT CTCATAAAGT CTGCCGATGA

                                          5` Nested Gene specific primer
  51 GAGGACTGCT TATGACCAAC TGCTCGATCA AATAAGAGAT TCCGAGGATC

                    5` Gene specific primer
 101 GAAGCTATGG CATCGTTCAC GACACATTTT ATGAGCTGGA ACCTGCCTAT
 151 GCTGACTACT ATCAGAAGAT GAAGAAAACA AAGTGTTGGC AAATTGGTCC
 201 TATCTCTCAT TTCTCTTCCA AATTAATCCG AAGAAAAGAG CTAATTGATG
 251 CCTCCGATGA CGTTAACTCG TGTGAGATTG ATAAATGGTT GAACAAGCAG
 301 GGGCAACGGT CGGTCCTCTA CATCTCTTTC GGGAGCCTTG TCAGATTCCC
 351 AGAGGACCAA CTCACGGAAA TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC
 401 CTTTTGTTTG GGTAATGAGG AAGGACCAAT CAGCACAAAC CACGTGGTTA
 451 CCCGATGGTT TCAAGGAAAA AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG
 501 GTGGGCACCG CAACAGACCA TCTTGGACCA TTCAGCGGTT GGAGGATTCA
 551 TCACTCACTG TGGTTGGAAT TCCGTACTCG AA
```

Figure 3.35 Gene specific primers designing for 5` RACE

3.8.8 Primers designed for 5` RACE

5` Gene specific primer 5' GCTCATAAAATGTGTCGTGAACGATG 3'

5` Nested Gene specific primer 5' AGCTTCGATCCTCGGAATCTCT 3'

3.8.9 5’ RACE reaction

In order to get the 5’region of the WsSAPGT genes, 5’ RACE was performed as described in Chapter 2, Section 2.9. For WsSAPGT gene, 5’RACE was performed using gene specific reverse primer 5’GSP (5’ GCTCATAAAATGTGTCGTGAACGATG 3’) and 5` Gene Racer primer provided with the kit. In this reaction 5’ RACE ready cDNA was used as the template. The above RACE product was diluted (1:20) and secondary PCR was done using 5’ Nested GSP (5’ AGCTTCGATCCTCGGAATCTCT 3’) and 5` Nested Gene Racer primer provided with the kit. Agarose gel electrophoresis analysis revealed an amplification product of approximately 400 bp as depicted in figure 3.36A. This amplicon was cloned in pGEMT-Easy vector (Promega, USA) and the resulting construct was confirmed by *Eco* RI restriction analysis (Figure 3.36B) and by sequencing (Figure 3.37).

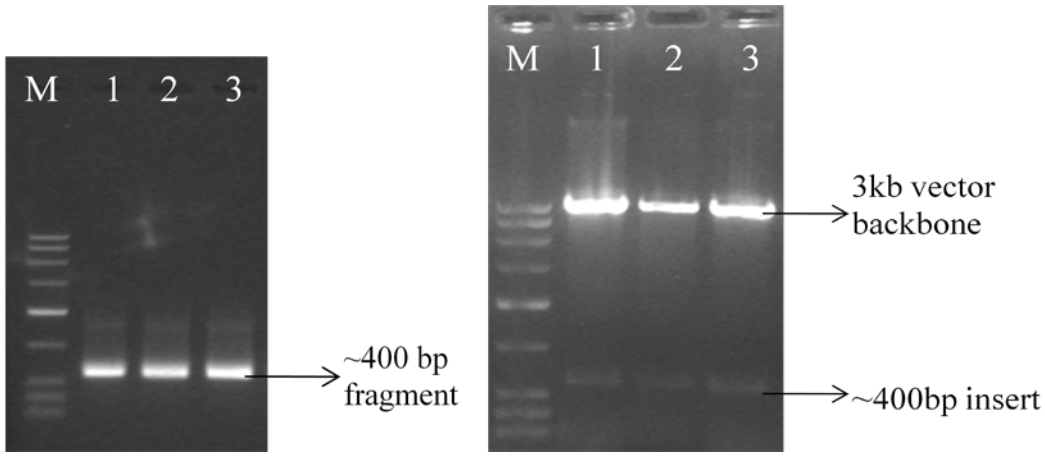


Figure A **Figure B**

Figure 3.36 Figure A: 1% agarose gel showing 5’RACE product of WsSAPGT gene. Lane M- marker, lane 1, 2 & 3 is ~400 bp 5’RACE product. **Figure B:** Lane M- marker, Lane 1, 2 & 3 *EcoRI* digested recombinant pGEMT-Easy vector and ~400 bp insert.

```
  1 GGACACTGAC ATGGACTGAA GGAGTACCGT CCTCTGAAGT TGGGTTGCCT
 51 GAAGGGATAG AAAACTTCAG CAAAGTTCCA TTCCCTGAAA TGTCTATCAA
101 AATATTTCAC GCAATTTTTC TTCTGCAGAA ACCTATGGAA GATGCAATTC
151 GTGAAATCTG TCCTGATTGT ATCTTCTCAG ATATGTATTT CCCTTGGACT
201 GTGGATATTG CTCATGAGCT GAAAATCCCT AGGCTCTTGT TCAATCAGTC
251 TGGCTACATG TGCAATTCCA TTCTACACAA TCTTAAGCTT TACAAGCCTC
301 ACGAAAAGGT CCCATCCCAA ACTACTTTCT TAGTTCCTGC TTTACCAAAT
```

```
351 AAGATCCATT TCAAGCGTTC CCAACTTACA GAGGATCTTA TAAAGTCTGC
401 CGATGAGAGG ACTGCTCATG ACCAACTGCT CGATCAAATA AGAGATTCCG
451 AGGATCGAAG CTA
```

Figure 3.37: 5` RACE nucleotide sequence.

The nucleotide sequence of 5` RACE product, which shows 5` Nested Gene Racer primer and 5` Nested Gene specific primer in red colour and underlined. The green colour sequence shows overlapping sequence with 1.2 kb. The expected 5` RACE fragment is ~500 bp, but in 5` RACE reaction only 330 bp fragment was amplified. So this sequence is truncated 5` RACE product. Still 200 bp upstream sequence is to be isolated.

3.9 PCR amplification of partial WsSAPGT gene

In 3` RACE reaction WsSAPGT fragment length is 930 bp and 5` RACE reaction fragment length is 330 bp. There is still remaining upstream sequence of ~200 bp. To amplify single partial fragment extreme primer from 5` RACE product 463 nucleotide sequence and extreme reverse primer from 930 nucleotides used for PCR amplification. 1st strand cDNA is used as a template and PCR is performed. After PCR amplification total ~1250 bp (Figure 3.38A) partial WsSAPGT clone is obtained. This fragment is cloned into pGEM-T Easy vector and the resulting construct was confirmed by *EcoRI* restriction analysis (Figure 3.38B) and by sequencing (Figure 3.39).

3.9.1 Primers designed for PCR amplification

5` CCGTCCTCTGAAGT TGGGTTGCCT 3`

5` ACCACTTTCTTTTTCGATAAAAGTAA 3`

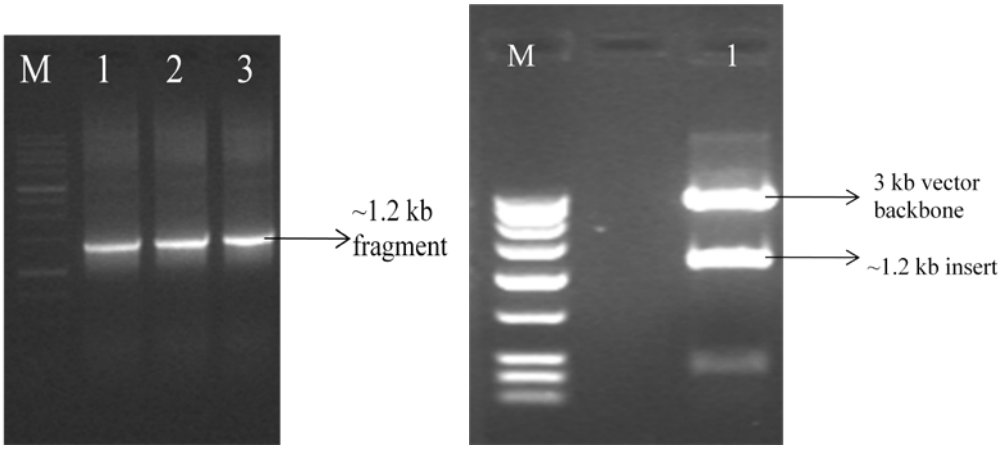


Figure A **Figure B**

Figure 3.38 Figure A: 1% agarose gel showing partial WsSAPGT gene amplicon. Lane M- marker, lane 1, 2 & 3 is ~1250 bp amplicon. **Figure B:** Lane M- marker, Lane 1 *Eco* RI digested recombinant pGEMT-Easy vector and ~1250 bp insert.

```
   1 CCGTCCTCTG AAGTTGGGTT GCATGAAGGG ATAGAAAACT TCAGCAAAGT
  51 TCCATTCCCT GAAATGTCTA TCAAAATATT TCACGCAATT TTTCTTCTGC
 101 AGAAACCTAT GGAAGATGCA ATTCGTGAAA TCTGTCCTGA TTGTATCTTC
 151 TCAGATATGT ATTTCCCTTG GACTGTGGAT ATTGCTCATG AGCTGAAAAT
 201 CCCTAGGCTC TTGTTCAATC AGTCTGGCTA CATGTGCAAT TCCATTCTAC
 251 ACAATCTTAA GCTTTACAAG CCTCACGAAA AGGTCCCATC CCAAACTACT
 301 TTCTTAGTTC CTGCTTTACC AAATAAGATC CACTTCAAGC GTTCCCAACT
 351 TACAGAGGAT CTCATAAAGT CTGCCGATGA GAGGACTGCT TATGACCAAC
 401 TGCTCGATCA AATAAGAGAT TCCGAGGATC GAAGCTATGG CATCGTTCAC
 451 GACACATTTT ATGAGCTGGA ACCTGCCTAT GCTGACTACT ATCAGAAGAT
 501 GAAGAAAACA AAGTGTTGGC AAATTGGTCC TATCTCTCAT TTCTCTTCCA
 551 AATTAATCCG AAGAAAAGAG CTAATTGATG CCTCCGATGA CGTTAACTCG
 601 TGTGAGATTG ATAAATGGTT GAACAAGCAG GGGCAACGGT CGGTCCTCTA
 651 CATCTCTTTC GGGAGCCTTG TCAGATTCCC AGAGGACCAA CTCACGGAAA
 701 TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC CTTTTGTTTG GGTAATGAGG
 751 AAGGACCAAT CAGCACAAAC CACGTGGTTA CCCGATGGTT TCAAGGAAAA
 801 AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG GTGGGCACCG CAACAGACCA
 851 TCTTGGACCA TTCAGCGGTT GGAGGATTCA TCACTCACTG TGGTTGGAAT
 901 TCCGTACTCG AAGCTATCGT TGCTGGCGTG CCGATGCTGA CATGGCCACT
 951 ATTCGCAGAT CAATTCTACG ACGAGAAGCT CGTAGAGGTT TTGGGTTTGG
1001 GAGTGAAAGT GGGGTCAGAG GTATGTAGTT TGGTAGGTGT TGACATAATG
1051 GGTCCTATAA TAGGGAGTGA AAAGATTAAA GAAGCAATAC ACCAATTGAT
1101 GAGTGGTGGT TCCAAGGAAA GGGAAAATAT TAGGGAGAAA TCAATGGTTA
1151 TGAGTAAGAT GGCGAAAAAG GCAACCGAAG GAAATGGATT TTCATGCAAC
1201 AGTCTCACAG CGCTCATTGA TGATATCAAG AATTTTACTT TTATCGAAAA
1251 AGAAAGTGGT
```

Figure 3.39 Partial nucleotide sequence of WsSAPGT gene.

Partial 1260 bp nucleotide sequence of WsSAPGT gene, forward and reverse primers shown in red coloured and underlined.

3.10 Characterization of WsSAPGT gene

The partial cDNA sequence of WsSAPGT has been submitted to NCBI GenBank database under accession no.FJ560881 (Figure 3.39). The WsSAPGT sequence contains a 1260 bp open-reading frame, which codes for a protein of 420 amino acids with a molecular weight 46.2 kDa (Figure 3.41). The sequence analysis shows that RACE amplification has been well extended downstream to the stop codon (TAA) and has yielded a 285 bp 3'UTR and a poly-A tail. The 5` upstream region with UTR region need to be isolated. Near poly-A tail an ubiquitous polyadenylation (poly-A) signal sequence '**AATAA**' was present in the 3'UTR, which is present in most eukaryotes and forms a complex with U4 SnRNP for primary cleavage site selection in pre- mRNA (Joshi, 1987).

Nucleotide sequence of WsSAPGT gene (FJ560881)

```
   1 CCGTCCTCTG AAGTTGGGTT GCATGAAGGG ATAGAAAACT TCAGCAAAGT
  51 TCCATTCCCT GAAATGTCTA TCAAAATATT TCACGCAATT TTTCTTCTGC
 101 AGAAACCTAT GGAAGATGCA ATTCGTGAAA TCTGTCCTGA TTGTATCTTC
 151 TCAGATATGT ATTTCCCTTG GACTGTGGAT ATTGCTCATG AGCTGAAAAT
 201 CCCTAGGCTC TTGTTCAATC AGTCTGGCTA CATGTGCAAT TCCATTCTAC
 251 ACAATCTTAA GCTTTACAAG CCTCACGAAA AGGTCCCATC CCAAACTACT
 301 TTCTTAGTTC CTGCTTTACC AAATAAGATC CACTTCAAGC GTTCCCAACT
 351 TACAGAGGAT CTCATAAAGT CTGCCGATGA GAGGACTGCT TATGACCAAC
 401 TGCTCGATCA AATAAGAGAT TCCGAGGATC GAAGCTATGG CATCGTTCAC
 451 GACACATTTT ATGAGCTGGA ACCTGCCTAT GCTGACTACT ATCAGAAGAT
 501 GAAGAAAACA AAGTGTTGGC AAATTGGTCC TATCTCTCAT TTCTCTTCCA
 551 AATTAATCCG AAGAAAAGAG CTAATTGATG CCTCCGATGA CGTTAACTCG
 601 TGTGAGATTG ATAAATGGTT GAACAAGCAG GGGCAACGGT CGGTCCTCTA
 651 CATCTCTTTC GGGAGCCTTG TCAGATTCCC AGAGGACCAA CTCACGGAAA
 701 TCGCAAAGGC ACTAGAAGCT TCGAGCGTTC CTTTTGTTTG GGTAATGAGG
 751 AAGGACCAAT CAGCACAAAC CACGTGGTTA CCCGATGGTT TCAAGGAAAA
 801 AGCAAAGAAC AAGGGTCTCC TCCTTAAAGG GTGGGCACCG CAACAGACCA
 851 TCTTGGACCA TTCAGCGGTT GGAGGATTCA TCACTCACTG TGGTTGGAAT
 901 TCCGTACTCG AAGCTATCGT TGCTGGCGTG CCGATGCTGA CATGGCCACT
 951 ATTCGCAGAT CAATTCTACG ACGAGAAGCT CGTAGAGGTT TTGGGTTTGG
1001 GAGTGAAAGT GGGGTCAGAG GTATGTAGTT TGGTAGGTGT TGACATAATG
1051 GGTCCTATAA TAGGGAGTGA AAAGATTAAA GAAGCAATAC ACCAATTGAT
1101 GAGTGGTGGT TCCAAGGAAA GGGAAAATAT TAGGGAGAAA TCAATGGTTA
1151 TGAGTAAGAT GGCGAAAAAG GCAACCGAAG GAAATGGATT TTCATGCAAC
```

```
1201 AGTCTCACAG CGCTCATTGA TGATATCAAG AATTTTACTT TTATCGAAAA
1251 AGAAAGTGGT TAAGGGATGA CAACCAGGAG GATGGGATTG GAAGCAACCA
1301 TCGGTTAGTC ATCAACAATA TAGCTTAAAT GATCGTTGGT TTGGTTTAGT
1351 TTTCCTCCTT TTCGTTATTT TTTTTTGTCA TGTTACCATG CACTAATGCT
1401 GTGTTTGTGC ACAACTACTC CATAAACAAG CATGTTATTG TATGTTGTTA
1451 AAATCTATGT ATATTGTGGG AACTAGCTTC CCTCCAGTTG CAATAAATCA
1501 AAGACATGAA AAATCAATCC TCATAGAAAT AAATAGATTA AATTTTGGAA
1551 AAAGAAAAAA AAAAAAAAAA AAAAAAAAA
```

Figure 3.40: Nucleotides sequence of WsSAPGT gene.

Deduced amino acid sequence of WsSAPGT gene (FJ560881)

P S S E V G L H E G I E N F S K V P F P E M S I K I F H A I F L L Q K P M E D A I R E I
C P D C I F S D M Y F P W T V D I A H E L K I P R L L F N Q S G Y M C N S I L H N L
K L Y K P H E K V P S Q T T F L V P A L P N K I H F K R S Q L T E D L I K S A D E R
T A Y D Q L L D Q I R D S E D R S Y G I V H D T F Y E L E P A Y A D Y Y Q K M K K
T K C W Q I G P I S H F S S K L I R R K E L I D A S D D V N S C E I D K W L N K Q G
Q R S V L Y I S F G S L V R F P E D Q **L T E I A K A L E A S S V P F V W V M R K D**
Q S A Q T T W L P D G F K E K A K N K G L L L K G **W A P Q Q T I L D H S A V G**
G F I T H C G W N S **V L E A I V A G V P M L T W P L F A D Q** F Y D E K L V E V L
G L G V K V G S E V C S L V G V D I M G P I I G S E K I K E A I H Q L M S G G S K E
R E N I R E K S M V M S K M A K K A T E G N G F S C N S L T A L I D D I K N F T F I
E K E S G

Figure 3.41: Deduced amino acid sequence of WsSAPGT gene.

Analysis of the deduced amino acid sequence showed it contained the donor binding site (green coloured bold and underlined) and the 44 bp PSPG box which is conserved in all plant glycosyltransferase genes (coloured in red). Most conserved region of PSPG box i.e, **HCGWNS** (highlighted in turquoise) is also present in the sequence. It also contained the tryptophan residue in the most conserved region of PSPG box (**HCGWNS**) which is found to be responsible for glucose specificity.

3.10.1 CLUSTAL W (1.8) Multiple sequence alignment of nucleotide sequences of WsSAPGT with *Solanum aculeatissimum*

```
WsSAPGT         ------------------------------------------------------------
Solanum         ATGGATAACGGCAGCAATCAACTACATGTCCTCTTCCTTCCTTACTTCGCCACTGGTCAT 60

WsSAPGT         ------------------------------------------------------------
Solanum         ATCATCCCATTAGTTAACGCCGCCAGGCTATTCGTCTTTCACGCCGGCGTCAAAGTCACA 120

WsSAPGT         ------------------------------------------------------------
Solanum         ATCCTCACTACCCACCACAACGCCTCCCTATTCCGATCTACTATTGACAACGACGTTGAA 180

WsSAPGT         -------------------------------------CCGTCCTCTGAAGTTGGGTTG 21
Solanum         GACGGCCATTCCGTAATCTCTATCCATACTCTTAGGTTCCCCTCCACTGAAGTTGGGTTG 240
                                                    ** *** **************

WsSAPGT         CATGAAGGGATAGAAAACTTCAGCAAAGTTCCATTCCCTGAAATGTCTATCAAAATATTT 81
Solanum         CCTGAAGGGATCGAGAACTTCAGCTCTGCCTCTTCACCTGAACTCGCGGGCAAAGTATTT 300
                * ********* ** *********   *   * *  ****** *  *   **** *****

WsSAPGT         CACGCAATTTTTCTTCTGCAGAAACCTATGGAAGATGCAATTCGTGAAATCTGTCCTGAT 141
Solanum         TACGCCATTTATCTTCTACAGAAACCAATGGAAGATAAAATTCGTGAAATCCATCCGGAT 360
                 **** **** ****** ******** *********  *************  *** ***

WsSAPGT         TGTATCTTCTCAGATATGTATTTCCCTTGGACTGTGGATATTGCTCATGAGCTGAAAATC 201
Solanum         TGTATTTTCTCTGATATGTACCTCCCATGGACTGTCAATATTGCCTTGGAGCTCAAAATC 420
                ***** ***** ********  **** ********  *******    ***** ******

WsSAPGT         CCTAGGCTCTTGTTCAATCAGTCTGGCTACATGTGCAATTCCATTCTACACAATCTTAAG 261
Solanum         CCCAGGCTATTGTTCAACCAGTCCAGCTACATGTACAATTCCATTCTGTACAATCTGAGG 480
                ** ***** ******** *****  ********* ************  ******* * *

WsSAPGT         CTTTACAAGCCTCACGAA---AAGGTCCCATCCCAAAC-----TACTTTCTTAGTTCCTG 313
Solanum         CTTTACAAACCTCACAAATCCAAAA--CTATTACTAGTACTGATAGTATCTCAGTTCCTG 538
                ******** ****** **   **    * **  * *       ** * *** ********

WsSAPGT         CTTTACCAAATAAGATCCACTTCAAGCGTTCCCAACTTACAGAGGATCTCATAAAGTCTG 373
Solanum         GTTTACCCGATAAGATCGAGTTCAAGCTATCGCAACTTACAGATGATCTGATAAAGCCTG 598
                 ******  ******** * *******  ** *********** ***** ****** ***

WsSAPGT         CCGATGAGAGGACTGCTTATGACCAACTGCTCGATCAAATAAGAGATTCCGAGGATCGAA 433
Solanum         AGGATGAGAAGAATGCTTTTGACGAATTGCTCGATCGAACTAGAGAATCCGAGGATCGAA 658
                  ******* ** ***** **** ** ********* **  ***** *************

WsSAPGT         GCTATGGCATCGTTCACGACACATTTTATGAGCTGGAACCTGCCTATGCTGACTACTATC 493
Solanum         GCTACGGCATCGTTCACGATACTTTTTACGAGCTAGAACCTGCCTACGCTGATTATTATC 718
                **** ************** ** ***** ***** *********** ***** ** ****

WsSAPGT         AGAAGATGAAGAAAACAAAGTGTTGGCAAATTGGTCCTATCTCTCATTTCTCTTCCAAAT 553
Solanum         AGAAGGTGAAGAAAACTAAATGTTGGCAAATTGGTCCCATCTCCCATTTTTCTTCCAAGT 778
                ***** ********** ** ***************** ***** ***** ******** *

WsSAPGT         TAATCCGAAGAAAAGAGCTAATTGATGCCTCCGATGACGTTAACTCGTGTGAGATTGATA 613
Solanum         TATTCCGTAGAAAAGAGCTGATTAATGCTGTCGATGAAAGTAACTCATGTGCCATTGTAG 838
                ** **** *********** *** ****   ******   ****** ****  ****

WsSAPGT         AATGGTTGAACAAGCAGGGGCAACGGTCGGTCCTCTACATCTCTTTCGGGAGCCTTGTCA 673
Solanum         AGTGGTTGAACGAGCAGGAGCATAAATCAGTCCTCTATGTCTCTTTCGGGAGCGTAGTTA 898
                * ********* ****** ***    ** ********  ************** * ** *
```

```
WsSAPGT         GATTCCCAGAGGACCAACTCACGGAAATCGCAAAGGCACTAGAAGCTTCGAGCGTTCCTT 733
Solanum         GATTCCCAGAAGCCCAACTCACTGAAATCGCAAAAGCTCTAGAAGCTTCGAGCATCCCTT 958
                ********** * ******** ********** ** ************** * ****

WsSAPGT         TTGTTTGGGTAATGAGGAAGGACCAATCAGCACAAACCACGTGGTTACCCGATGGTTTCA 793
Solanum         TCATTTGGGTAGTGAAGAAGGACCAATCGGCAGAAACCACGTGTTTGCTGGAGG------ 1012
                *  ******** *** ************ *** ********** ** *  ** *


WsSAPGT         AGGAAAAAGCAAAGAACAAGGGTCTCCTCCTTAAAGGGTGGGCACCGCAACAGACCATCT 853
Solanum         AGGAAAAATTGAAGAACAAGGGTCTGATTATTAGAGGGTGGGCGCCGCAACTAACCATCT 1072
                ********   **************  *  *** ********* *******  *******

WsSAPGT         TGGACCATTCAGCGGTTGGAGGATTCATCACTCACTGTGGTTGGAATTCCGTACTCGAAG 913
Solanum         TGGACCATTCAGCCGTGGGAGGATTCATGACACACTGCGGTTGGAATTCCATACTCGAAG 1132
                ************* ** ********** ** ***** ************ *********

WsSAPGT         CTATCGTTGCTGGCGTGCCGATGCTGACATGGCCACTATTCGCAGATCAATTCTACGACG 973
Solanum         CCATCATCGCCGGCGTCCCGTTGGTGACGTGGCCGGTGTTCGCGGAGCAATTCTACAATG 1192
                * *** * ** ***** *** ** **** *****  * ***** ** ********* * *

WsSAPGT         AGAAGCTCGTAGAGGTTTTGGGTTTGGGAGTGAAAGTGGGGTCAGAGGTATGTAGTTTG- 1032
Solanum         AAAAGCTTGTGGAGGTAATGGGGTTAGGAGTGAAGGTGGGGGCAGAAGTACATGAGTCGA 1252
                * ***** ** *****  **** ** ******** ****** **** ***  *   * *

WsSAPGT         --GTAGGTGTTGACATAATGGGTCCTATAATAGGGAGTGAAAAGATTAAAGAAGCAATAC 1090
Solanum         ATGGAGGTGTAGAGATCTCGAGCCTTGTGATAGAGAGCGAAAAGATCAAAGAAGCAATAG 1312
                  * ****** ** **   * * * * * **** *** ******** ************

WsSAPGT         ACCAATTGATGAGTGGTGGTTCCAAGGAAAGGGAAAATATTAGGGAGAAATCAATGGTTA 1150
Solanum         AGAAATTAATG---GATGACTCAAAGGAAAGCCAGAAAATTAGAGAGAAAGTAATAGGTA 1369
                *  **** ***   * **  ** ********  * ** ***** ******  *** * **

WsSAPGT         TGAGTAAGATGGCGAAAAAGGCAACCGAAGGAAATGGATTTTCATGCAACAGTCTCACAG 1210
Solanum         TGAGCGAAATGGCTAAAAATGCAGTCGAAGAAGGTGGATCTTCATGGAACAATCTCACTG 1429
                ****  * ***** ***** ***  ***** *  ***** ****** **** ****** *

WsSAPGT         CGCTCATTGATGATATCAAGAATTTTACTTTTATCGAAAAAGAAAGTGGT- 1260
Solanum         CACTTATTGATGATATCAAGAATTTTACTT----CGACGACGAACGTTTAA 1476
                * ** ************************    ***  * *** **
```

Figure 3.42: Nucleotide sequence alignment of WsSAPGT with *Solanum aculeatissimum*

Nucleotides sequence shows 79% homology with *Solanum aculeatissimum* UDP-glucose glucosyltransferase.

3.10.2 CLUSTAL W (1.8) multiple sequence alignment of amino acid sequences of WsSAPGT with *Solanum aculeatissimum*

```
WsSAPGT         ------------------------------------------------------------P 1
Solanum         MDNGSNQLHVLFLPYFATGHIIPLVNAARLFVFHAGVKVTILTTHHNASLFRSTIDNDVE 60

WsSAPGT         SSEVGLHE-------------GIENFSKVPFPEMSIKIFHAIFLLQKPMEDAIREICPD 47
Solanum         DGHSVISIHTLRFPSTEVGLPEGIENFSSASSPELAGKVFYAIYLLQKPMEDKIREIHPD 120
                              ******    **   * * ** ******** **** **
```

```
WsSAPGT         CIFSDMYFPWTVDIAHELKIPRLLFNQSGYMCNSILHNLKLYKPHEKVP--SQTTFLVPA 105
Solanum         CIFSDMYLPWTVNIALELKIPRLLFNQSSYMYNSILYNLRLYKPHKSKTITSTDSISVPG 180
                ******* **** ** *********** ** **** ** *****      *     **

WsSAPGT         LPNKIHFKRSQLTEDLIKSADERTAYDQLLDQIRDSEDRSYGIVHDTFYELEPAYADYYQ 165
Solanum         LPDKIEFKLSQLTDDLIKPEDEKNAFDELLDRTRESEDRSYGIVHDTFYELEPAYADYYQ 240
                ** ** ** **** ****  **  * * ***  * ************************

WsSAPGT         KMKKTKCWQIGPISHFSSKLIRRKELIDASDDVNSCEIDKWLNKQGQRSVLYISFGSLVR 225
Solanum         KVKKTKCWQIGPISHFSSKLFRRKELINAVDESNSCAIVEWLNEQEHKSVLYVSFGSVVR 300
                * ***************** ****** * *  *** *  *** *   **** **** **

WsSAPGT         FPEDQLTEIAKALEASSVPFVWVMRKDQSAQTTWLPDGFKEKAKNKGLLLKGWAPQQTIL 285
Solanum         FPEAQLTEIAKALEASSIPFIWVVKKDQSAETTCLLEEEKLKNKGLII--RGWAPQLTIL 358
                *** ************* ** **  ***** ** *    * * *       ***** ***

WsSAPGT         DHSAVGGFITHCGWNSVLEAIVAGVPMLTWPLFADQFYDEKLVEVLGLGVKVGSEVCSLV 345
Solanum         DHSAVGGFMTHCGWNSILEAIIAGVPLVTWPVFAEQFYNEKLVEVMGLGVKVGAEVHESN 418
                ******** ******* **** ****  *** ** *** ****** ******* **

WsSAPGT         GVDI-MGPIIGSEKIKEAIHQLMSGGSKERENIREKSMVMSKMAKKATEGNGFSCNSLTA 404
Solanum         GGVEISSLVIESEKIKEAIEKLMDDSKE-SQKIREKVIGMSEMAKNAVEEGGSSWNNLTA 477
                *        * ********  **         ****   ** *** * *  * * * ***

WsSAPGT         LIDDIKNFTFIEKESG 420
Solanum         LIDDIKNFTS--TTNV 491
                *********
```

Figure 3.43: Amino acid sequence alignment of WsSAPGT with *Solanum aculeatissimum*

Amino acid sequence shows 72% homology with *Solanum aculeatissimum* UDP-glucose glycosyltransferase.

3.10.3 Analysis of amino acid sequence of the WsSAPGT gene (FJ560881) for their conserved domain

The amino acid sequences of WsSAPGT (FJ560881) analysed for their sequence similarity. It matched with glycosyltransferases which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. The acceptor molecule can be a lipid, a protein, a heterocyclic compound, or another carbohydrate residue. The structures of the formed glycoconjugates are extremely diverse, reflecting a wide range of biological functions. The members of this family share a common GTB topology, one of the two protein topologies observed for nucleotide-sugar-dependent glycosyltransferases. GTB proteins have distinct N- and C- terminal domains each containing a typical Rossmann fold. The two domains have high structural homology despite minimal sequence homology. The large cleft that separates the two domains includes the catalytic

center and permits a high degree of flexibility. The conserved domains of WsSAPGT (Figure 3.44) are shown below.

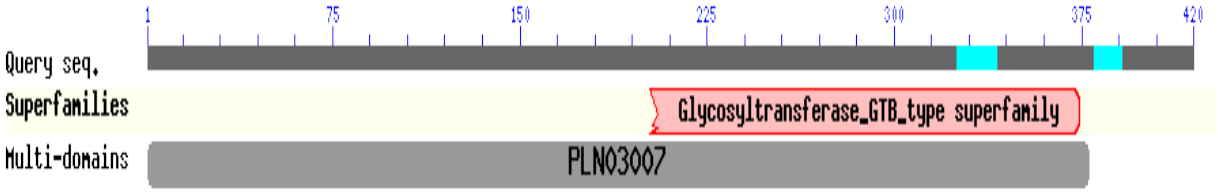


Figure 3.44: Conserved domains of WsSAPGT

3.10.4 Phylogenetic analysis of WsSAPGT (FJ560881) gene with other reported Glycosyltransferases

Phylogenetic analysis was done for WsSAPGT gene using various reported protein sequences of GT, which were retrieved from the GenBank database. Multiple alignments of the predicted protein coding sequences were performed using Clustal X. The phylogenetic tree was constructed by neighbor joining methods using Mega 4.0 software program. It is clear from the phylogenetic tree that, WsSAPGT is evolutionarily most similar with *Solanum aculeatissimum* (BAD89042 and AB182385), UDP-glucose glucosyltransferase and *Solanum tuberosum* (ABB29874 and AAB48444) UDP-glucose:solanidine glucosyltransferase.

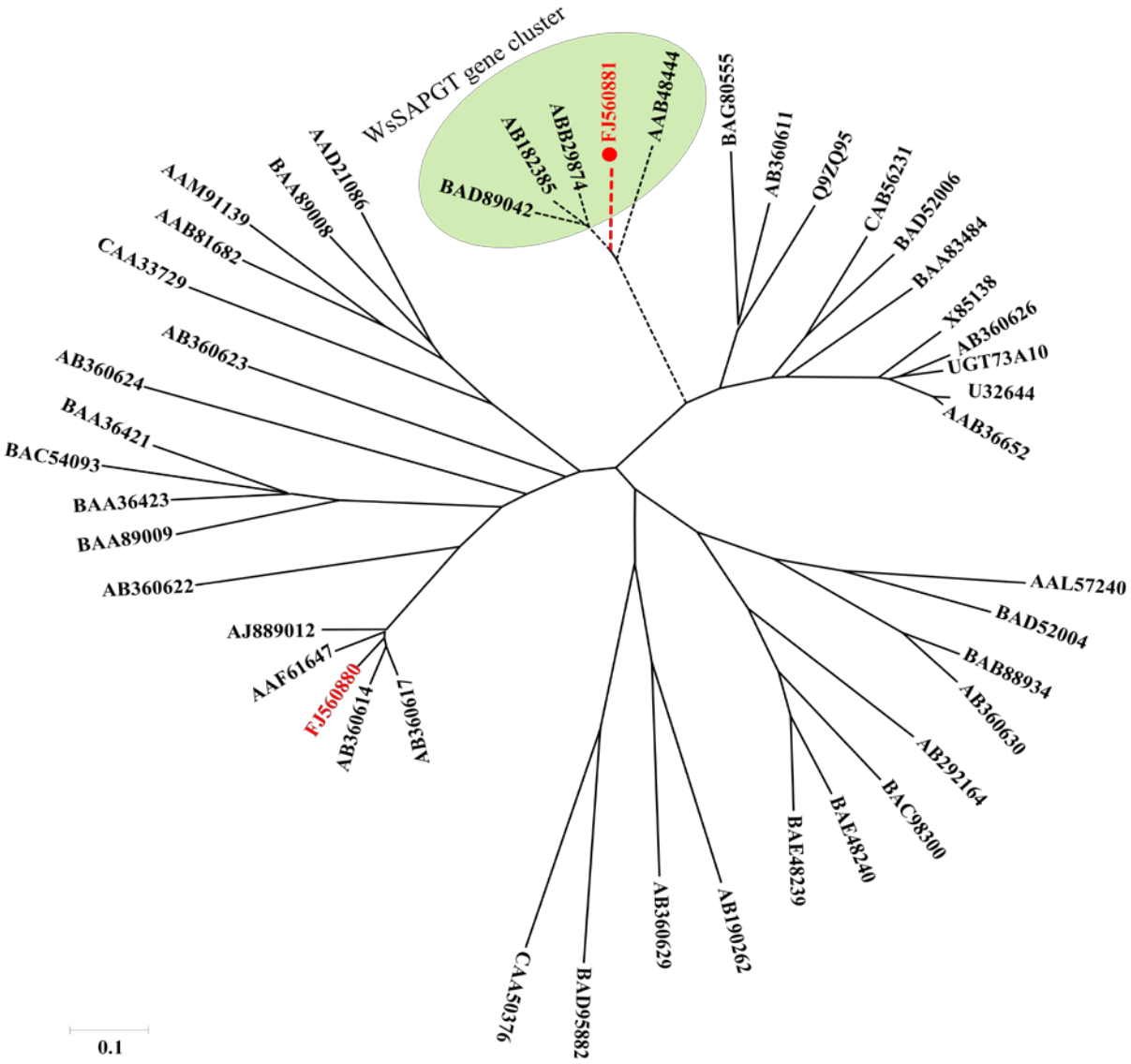


Figure 3.45: Phylogenetic tree of WsSAPGT gene. WsSAPGT (FJ560881) is marked with colour dot. The sequences used for the alignment were primarly those of glycosyltransferases that have been functionally characterized. The phylogenetic tree was constructed by neighbor joining methods using Mega 4.0 software program. The names and accession numbers of glycosyltransferases used for the alignment are as follows: *Solanum tuberosum* (AAB48444), *S. tuberosum* (ABB29874), *Solanum aculeatissimum* (BAD89042), *Solanum aculeatissimum* (AB182385), *Lycopersicon esculentum* (AJ889012), *Lycium barbarum* (BAG80555), *L.*

barbarum (AB360624), *L. barbarum* (AB360629), *Dianthus caryophyllus* (BAD52004), *Scutellaria baicalensis* (BAC98300), *Lycopersicum esculentum* (X85138), *Nicotiana tabacum* (BAB88934), *Dorotheanthus bellidiformis* (AAL57240), *Dorotheanthus bellidiformis* (BAD52006), *L. barbarum* (AB360626), *L. barbarum* (UGT73A10), *Nicotiana tabacum* (U32644), *Nicotiana tabacum* (AAB36652), *Petunia x hybrid* (BAA89009), *Torenia hybrid* (BAC54093), *Arabidopsis thaliana* (Q92Q95), *Petunia x hybrid* (CAA50376), *L. barbarum* (AB360623), *Perilla frutescins* (BAA36421), *Verbena x hybrid* (BAA36423), *Antirrhinum majus* (BAE48239), *Arabidopsis thaliana* (AAM91139), *Hordum vulgare* (CAA33729), *Forsythia x intermedia* (AAD21086), *Petunia x hybrid* (BAA89008), *Dorotheanthus bellidiformis* (CAB56231), *L. barbarum* (AB360611), *Bellis peremnis* (AB190262), *Vitis vinefera* (AAB81682), *Ipomoea purpurea* (BAD95882), *L. barbarum* (AB360630), *Linaria vulgaris* (BAE48240), *L. barbarum* (AB360622), *N. tabacum* (AAF61647), *L. barbarum* (AB360614), *L. barbarum* (AB360617), *Scutellaria baicalenis* (BAA83484), *Glycine max* (AB292164).

3.11 Conclusion

- PCR based approach was used to isolate the *Withania somnifera* glycosyltransferase genes. Two GT cDNA clones were isolated namely WsGT (Accession No.**FJ560880)** and WsSAPGT (Accession No.**FJ560881).**

- WsGT gene contain 1371 bp ORF region with 5` and 3` UTR region. The polyA tail is identified in the 3` UTR region with polyadenylation site. The gene showed 87% identity at nucleotide level and 84% identity at amino acid level with the UDP-glucose:glycosyltransferases gene from *Lycium barbarum.*

- Analysis of deduced amino acid sequence of WsGT gene showed the conserved domains for glycosyltransferase acceptor and donor binding sites. It also showed the most conserved 44 amino acids domain called as PSPG box which is found to be present in all the plant glycosyltransferase genes.

- Phylogenetic analysis of WsGT gene deduced amino acid sequence was done using MEGA 4 software program. The results show that WsGT gene is evolutionarily most similar to UDP-glucose:glycosyltransferases from *Lycium barbarum* of *Lamiaceae* family.

- The second glycosyltransferases namely WsSAPGT (Accession No. FJ560881) contain 1260 bp partial sequence with 3` UTR region. The 5` upstream region with UTR of WsSAPGT is pending. The gene showed 79% identity at nucleotide level and 72% identity at amino acid level with *Solanum aculeatissimum* UDP-glucose glucosyltransfese.

- Analysis of deduced amino acid sequence of WsSAPGT gene showed the conserved domains for glycosyltransferase acceptor and donor binding sites. It also showed the most conserved 44 amino acids domain called as PSPG box which is found to be present in all the plant glycosyltransferase genes.

- Phylogenetic analysis of WsSAPGT gene deduced amino acid sequence was done using MEGA 4 software program. The results show that WsSAPGT gene is evolutionarily most similar to *Solanum aculeatissimum* UDP-glucose glucosyltransfese.

Chapter: 4A

Heterologous expression of glycosyltransferase (genes), its purification and characterization

4.1 Introduction

Glycosyltransferases transfer sugar moieties from activated sugar donors to acceptor molecules with high efficiency and regiospecificity. Due to the small number of plant GT sequences that were available, their use in biocatalysis has been limited. In recent years, many plant GT sequences have been identified and their corresponding recombinant proteins analyzed *in vitro*. Although the enzymatic activities and substrate specificities of many GTs of plants are known through *in vitro* analysis, their biological functions *in vivo* need to be confirmed.

The biochemical characteristics of most plant GTs identified so far are consistent with their physiological ones in plants. Therefore, *in vitro* identification of enzymatic activities of plant GTs provides a good starting point for *in vivo* functional analyses.

In chapter 3, two glycosyltransferase genes WsGT and WsSAPGT were isolated and characterized. Here, in this chapter 4, both the GTs are expressed in heterologous system, purified and characterized for their substrate specificity.

4.2 Materials and methods

4.2.1 Materials

Glassware and plasticware: As discussed in Chapter 2, Section 2.2 and Section 2.3

Chemicals: As discussed in Chapter 2, Section 2.4

***Escherichia coli* and vectors used in the study:**

Escherichia coli XL-1 Blue (Stratagene, USA)

E. coli, BL 21 (DE3) (Novagen, USA)

pGEM-T Easy Cloning vector (Promega, USA)

pET30b (+) Expression vector (Novagen, USA).

Stock solutions: As discussed in Chapter 2, Section 2.6.

4.2.2 Methods

4.2.2.1 Bacterial culture conditions: As discussed in Chapter 2, Section 2.8.2.1.

4.2.2.2 Bacterial cells transformation: *E. coli* transformation and selection was done as discussed in Chapter 2, Section 2.8.2.4.

4.2.2.3 Isolation of plasmid DNA from *E. coli*: As described in Chapter 2, Section 2.8.3.1.

4.2.2.4 Restriction digestion of DNA: As described in Chapter 2, Section 2.8.3.3.

4.2.2.5 Extraction and purification of DNA from agarose gel: As described in Chapter 2, Section 2.8.5.

4.2.2.6 Polymerase Chain Reaction (PCR): As described in Chapter 2, Section 2.8.7.

4.2.2.7 Colony PCR method: As described in Chapter 2, Section 2.8.2.5.

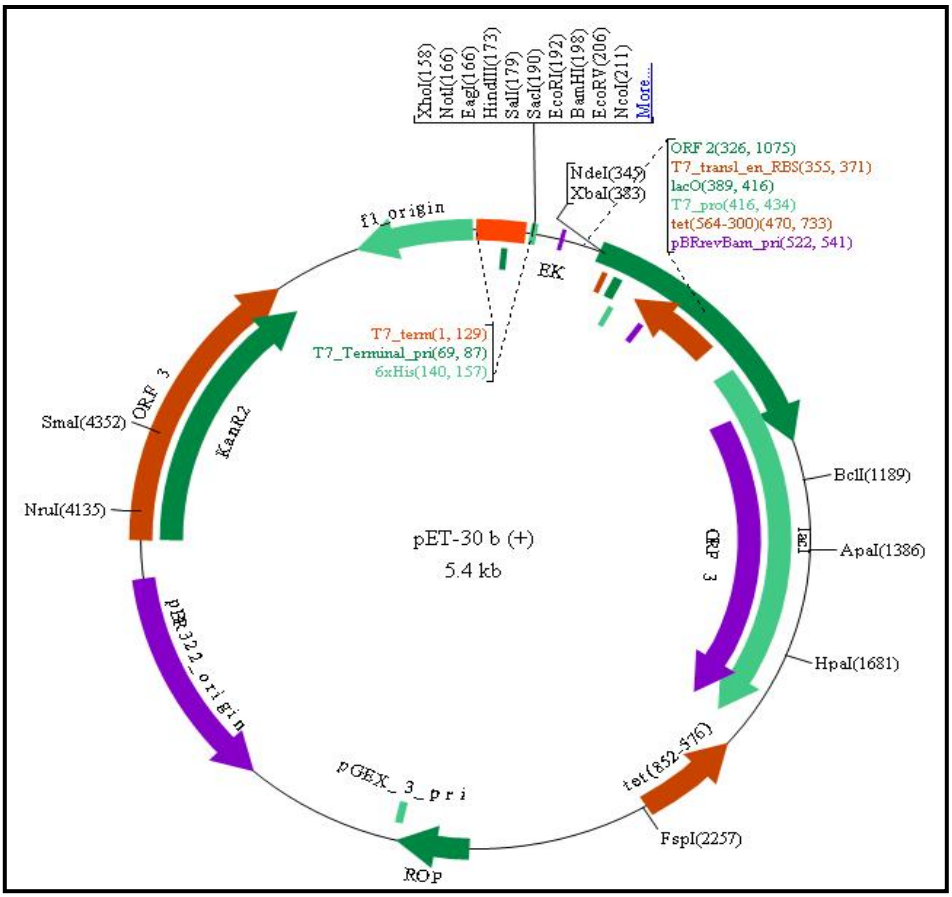
4.2.2.8 Cloning of WsGT gene in pET 30b (+) Expression vector

GT genes were cloned in pGEM-T Easy vector by incorporating appropriate restriction sites in the primers. In case of WsGT gene, *Kpn* I restriction site was added to the forward primer and *Xho* I restriction site was added to the reverse primer (namely WsGTKpn and WsGTXho). High fidelity *Taq pol* (*Pfx* Invitrogen) was used to amplify the WsGT gene using the above set of primers from the cDNA clone, sequenced and maintained in pGEM-T Easy vector as the template. The above recombinant plasmid DNA was diluted 1:50 times and 1 µL was used as the template and PCR was performed (Chapter 2, Section 2.8.7). Approximately 1.4 kb band was amplified (exactly 1369 bp coding region of WsGT gene + *Kpn* I and *Xho* I sites added into primer sequences). The band is excised from gel, purified, ligated in pGEM-T Easy vector and transformed in *E. coli* XL1 blue cells. Clones were screened by colony PCR and a few colonies were inoculated in 5 mL LB (Ampicilin 100 µg/mL) medium in tubes. Plasmid was isolated and restriction digested with respective restriction enzymes (*Kpn* I & *Xho* I) to confirm the integration of the gene.

4.2.2.9 PCR cycling conditions for WsGT gene

Temperature	Time	Cycles
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
55 ºC	45 s	
72 ºC	1.30 min	
72 ºC	7 min	1 cycle
4 ºC	hold	1 cycle

The recombinant WsGT clone with respective restriction sites were directionally cloned in pET-30b (+) vector (Figure 4.A.1). Colony PCR (Chapter 2 Section 2.8.2.5) was done to screen the recombinant pET-30b (+) clones. Integration of WsGT gene in pET-30b (+) was confirmed by digestion with respective restriction enzymes (Figure 4.A.2).



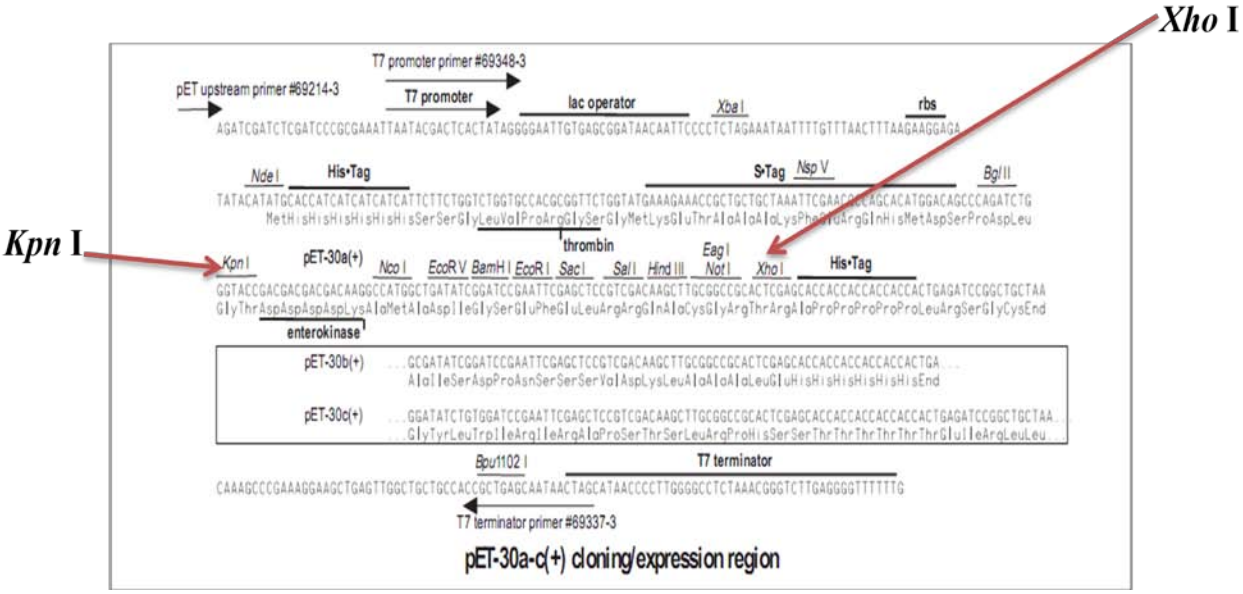


Figure 4.A.1: Vector map of pET 30b (+)

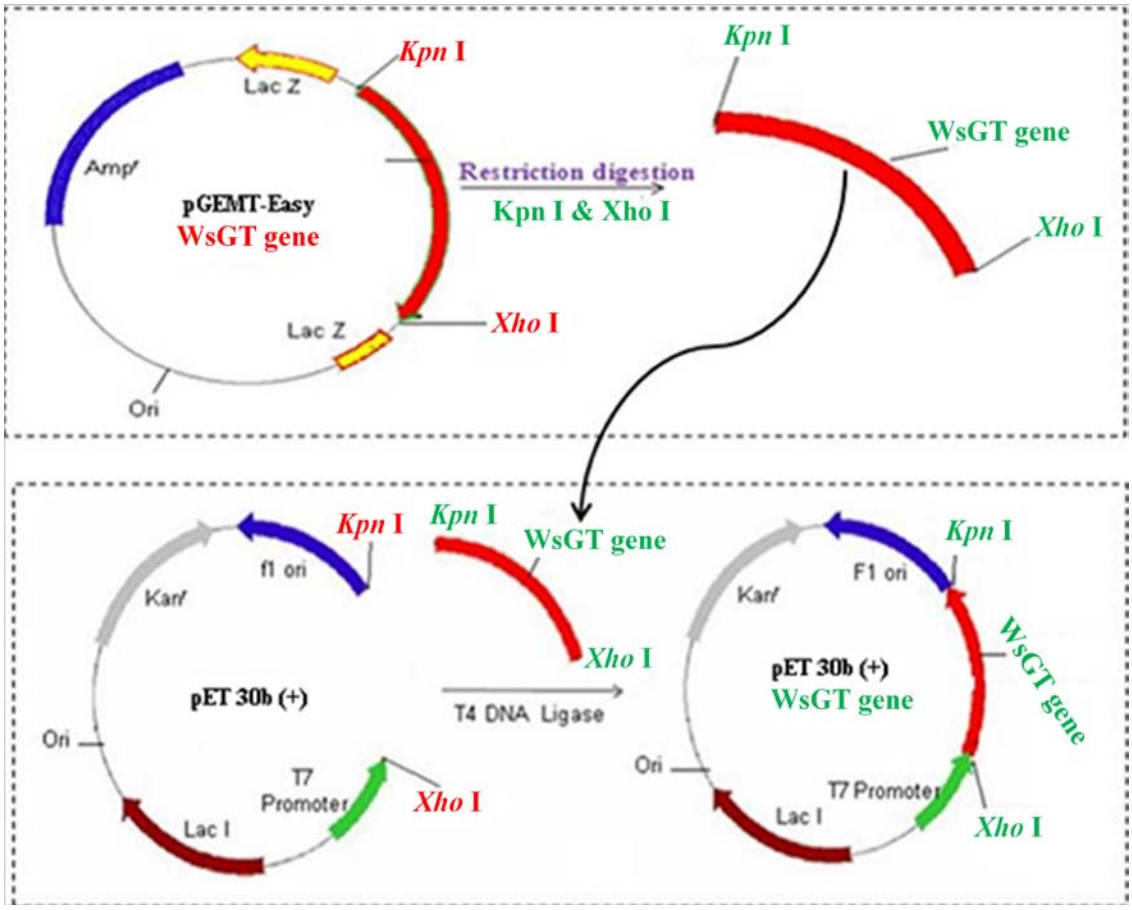


Figure 4.A.2: Strategy used for directional cloning of WsGT gene in pET-30b (+) vector

4.2.2.10 Recombinant WsGT protein expression and its purification from inclusion bodies

4.2.2.10.1 Recombinant protein expression in *E. coli* (BL21)

A single bacterial (BL21) colony carrying recombinant pET-30b (+) with WsGT gene, from freshly streaked plates (grown on LB agar medium containing 50 µg/mL kanamycin) was used to inoculate 5 mL liquid LB medium containing the same concentrations of the antibiotic. Culture was grown overnight with shaking at 200 rpm at 37 °C. One mL aliquot of over-night grown culture was used to inoculate 100 mL LB broth containing 50 µg/mL kanamycin. Once the cultures reached OD_{600} 0.6 to 0.8, recombinant protein expression was

induced by the addition of isopropyl β-D thiogalactopyranoside (IPTG), and the culture was grown for 4 to 6 h at 37 °C with shaking at 200 rpm. Recombinant protein extraction was done according to the protocol described in Section 2.12. Likewise, four positive recombinant pET-30b (+) clones were screened for WsGT protein over-expression and analysed by 10% SDS PAGE (Chapter 2, Section 2.12.3). The clone showing maximum over-expression was chosen for further studies.

4.2.2.10.2 Purification of recombinant protein (WsGT)

His-tagged recombinant protein (WsGT) was purified by metal chelate affinity chromatography. The initial stage of His-tagged protein purification is based on the remarkable selectivity and high affinity of patented Ni-NTA (nickel-nitrilotriacetic acid) resin for proteins containing an affinity tag of six consecutive histidine residues, the 6xHis-tag. NTA, which has four chelating sites for nickel ions, binds nickel more tightly than metal chelating purification systems that have only three sites available for interaction with metal ions. The extra chelation site prevents nickel ion leaching, providing a greater binding capacity and high-purity protein preparations. Purification of recombinant WsGT protein carrying a 6xHis-tag was conducted using Ni-NTA Agarose beads (Qiagen) (Chapter 2, Section 2.12.2). Purity of protein was checked on 10% SDS-PAGE (Chapter 2, Section 2.12.3).

4.2.2.10.3 Raising polyclonal antibody against purified WsGT protein in rabbit

The purified WsGT protein was used for raising polyclonal antibodies in New Zealand White rabbit.

4.2.2.10.4 Pre-treatment of serum: As described in Chapter 2, Section 2.15.1

4.2.2.10.5 Determination of titer of antibodies

ELISA was performed to determine the titer of first, second and third bleed of rabbit serum (Chapter 2, Section 2.15.2). Once the antibody titer was determined then, a fixed dilution of antibody was used for rest of the experiments.

4.2.2.11 Standardization of time and temperature for protein expression in soluble form

The information obtained from above section 4.2.2.10 was utilized for this experiment. Several flasks with 50 mL LB broth (kanamycin 50 µg/mL) were inoculated with O/N grown culture of *E. coli* BL21 harboring recombinant pET-30b (+) plasmid carrying WsGT gene. The cells were grown till A_{600} reached 0.6-0.8 and induction was done with 0.01 mM, 0.03 mM, 0.05 mM, 0.08 mM, 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM and 1 mM IPTG. The cultures were grown at different temperatures 15 ºC, 18 ºC, 20 ºC, 24 ºC, 28 ºC and different durations 6 h, 8 h, 12 h, 14 h, 18 h, 20 h for optimization of maximum expression of recombinant WsGT protein in soluble form. One flask as an uninduced control sample was used before inducing with IPTG. The soluble fraction of cell lysate was analysed on 10% SDS-PAGE.

4.2.2.12 Protein estimation

Protein estimation was done using Bradford assay (Bradford, 1976). A standard graph was made for BSA and concentration of unknown sample was determined by plotting standard graph (Chapter 2 section 2.14).

4.2.2.13 GT enzyme assay

The purified glycosyltransferases enzyme was used for performing enzyme assay studies. The control reaction contained crude lysate of only pET 30b (+) vector transformed in *E. coli* (BL21). Assay reaction was carried out as described in Chapter 2, Section 2.17. See Chapter 2 Table 2.10 substrates used for the study.

4.3 Results and discussions

4.3.1 Cloning of WsGT gene in pET-30b (+) vector

WsGT gene was cloned in pET-30b (+) expression system to get the gene expressed in its active form and for its characterization.

4.3.2 Incorporation of restriction sites

The WsGT gene cloned in pGEM-T Easy vector was amplified using gene specific primers **WsGTKpn Forward** and **WsGTXho Reverse** to incorporate the *Kpn* I site at the 5' end and *Xho* I site at 3' end of WsGT gene.

WsGTKpn Forward primer- 5' GGTACCATGACTACTCACAAAGCTCA3'

WsGTXho Reverse primer- 5' CTCGAGTGGAAATAGTAACCAACTTGG 3'

Approximately 1.4 kb (1369 bp WsGT + *Kpn* I and *Xho* I restriction sites) band was amplified. The band was excised from the gel, purified, ligated in pGEM-T Easy vector and transformed in *E. coli* XL1 cells. Clones with WsGT gene with *Kpn* I and *Xho* I restriction sites were screened by inoculating a few colonies in 5 mL LB media (ampicilin 100 μg/mL) tubes. Individual plasmids were isolated and restriction digested with *Kpn* I and *Xho* I enzymes to confirm the integration of WsGT gene.

4.3.3 Directional cloning of WsGT gene in pET-30b (+)

The above clone of WsGT gene in pGEM-T Easy vector was restriction digested with *Kpn* I and *Xho* I restriction enzymes and gene of interest was purified. pET-30b (+) vector DNA was also digested with same restriction enzymes and purified. WsGT gene was directionally cloned in purified restriction digested pET-30b (+) vector. Ligation mixture was transformed into *E. coli* XL1 competent cells and plated on LB-agar plate (kanamycin 50 μg/mL). *E. coli* colonies having recombinant plasmids were screened by colony PCR (Figure 4.A.3), with reaction cycles as shown in table below. Plasmids were isolated from PCR positive clones and

were digested with *Kpn* I and *Xho* I to confirm the integration of WsGT gene fragment in pET-30b (+) vector (Figure 4.A.4).

4.3.4 PCR cycling conditions for WsGT gene

Temperature	Time	Cycles
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
58 ºC	0.45 min	
72 ºC	1.30 min	
72 ºC	5 min	1 cycle

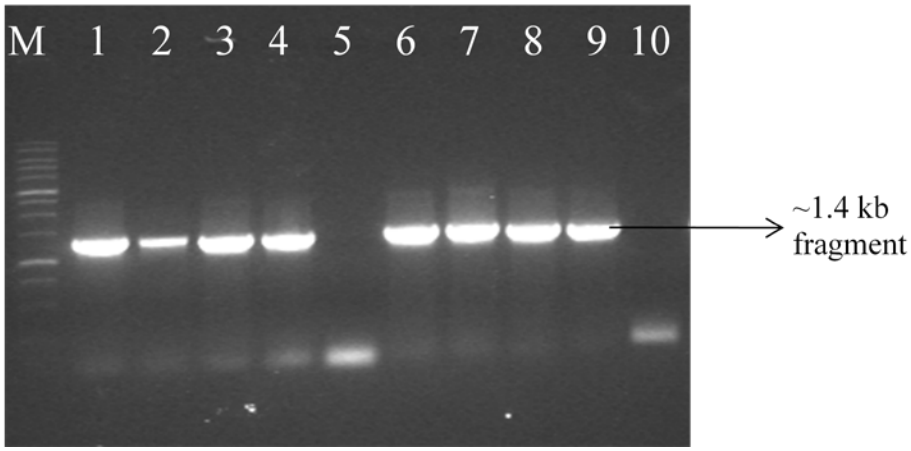


Figure 4.A.3: Colony PCR showing ~1.4 kb WsGT gene containing recombinant colonies. Lane M- marker, Lane 1-4 & 6-9 are positive clones.

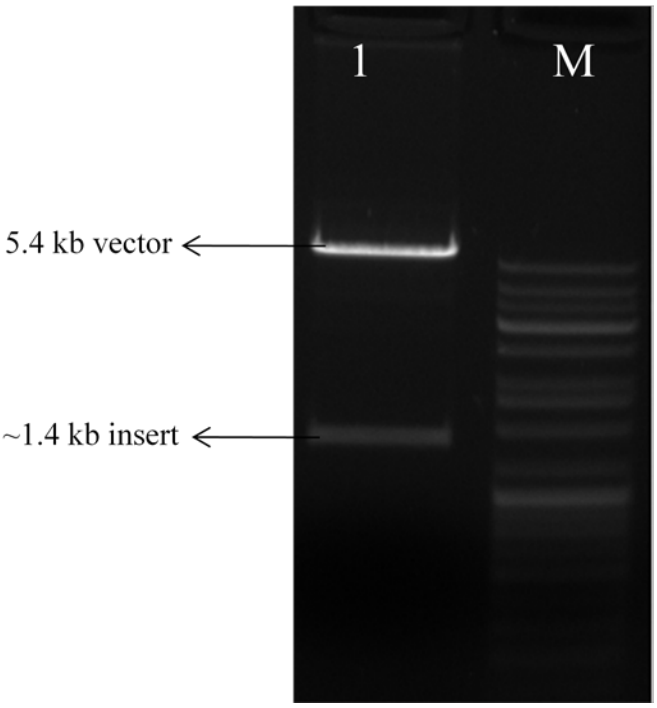


Figure 4.A.4: Restriction analysis of recombinant pET-30b (+). Lane 1 recombinant vector releasing ~1.4 kb WsGT gene and 5.4 kb vector backbone, Lane M- marker.

Integration of the WsGT gene along with restriction sites was further confirmed by sequencing. The sequence was translated using proteomic tools available at www.expasy.ch and was checked for in frame translation up to HIS tag.

4.3.5 Recombinant WsGT protein expression and purification from inclusion bodies

E. coli BL 21 (DE3) cells transformed with recombinant pET-30b (+) plasmids having WsGT gene were screened for over-expression of protein. A few positive recombinant clones were screened for recombinant WsGT protein over-expression. An approximately 58 kDa protein was expressed in all clones analyzed on 10% SDS-PAGE with varying levels of expression. Among those recombinant clones one of the clone showing maximum expression was chosen and large scale protein purification was done using Ni-NTA Agarose beads (Figure 4.A.5).

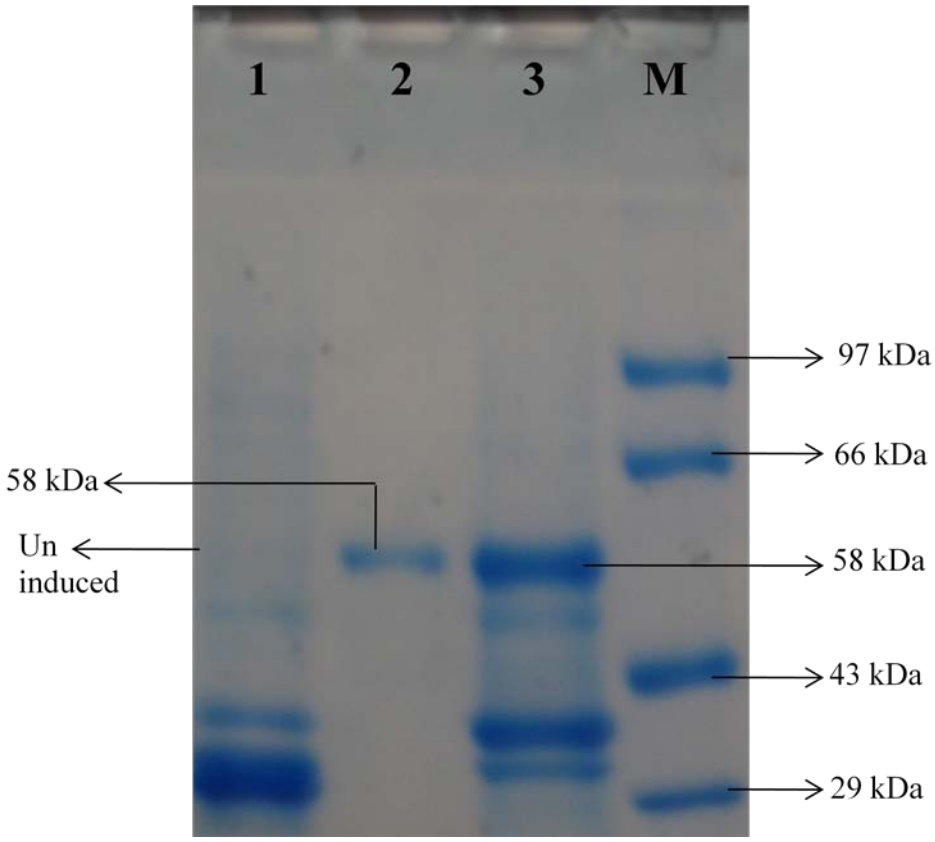


Figure 4.A.5: 10% SDS PAGE; Commassie Blue staining: Lane-1 Un induced protein, Lane-2 Ni-NTA purified WsGT Protein, Lane-3 Inclusion bodies from WsGT protein, Lane-M Protein molecular weight marker.

Purified WsGT protein was dialyzed against 10mM PBS buffer supplemented with 1 mM DTT and 0.1 mM PMSF overnight with two changes of fresh buffer. Dialyzed WsGT protein was concentrated, quantified by Bradford assay (Bradford reagent, Bio-rad, USA) and given for raising polyclonal antibodies in New Zealand rabbit.

4.3.6 Raising antibodies in rabbit against WsGT protein

300 µg of purified protein was used for first injection in New Zealand rabbit to raise antibodies. Same amount of protein was used for booster doses. Antibody titer of first, second and third bleed was determined by ELISA. Titer of third bleed is ~1:100000 (Figure 4.A.6). 3rd bleed serum dilution of 1:10000 was used for further experiments.

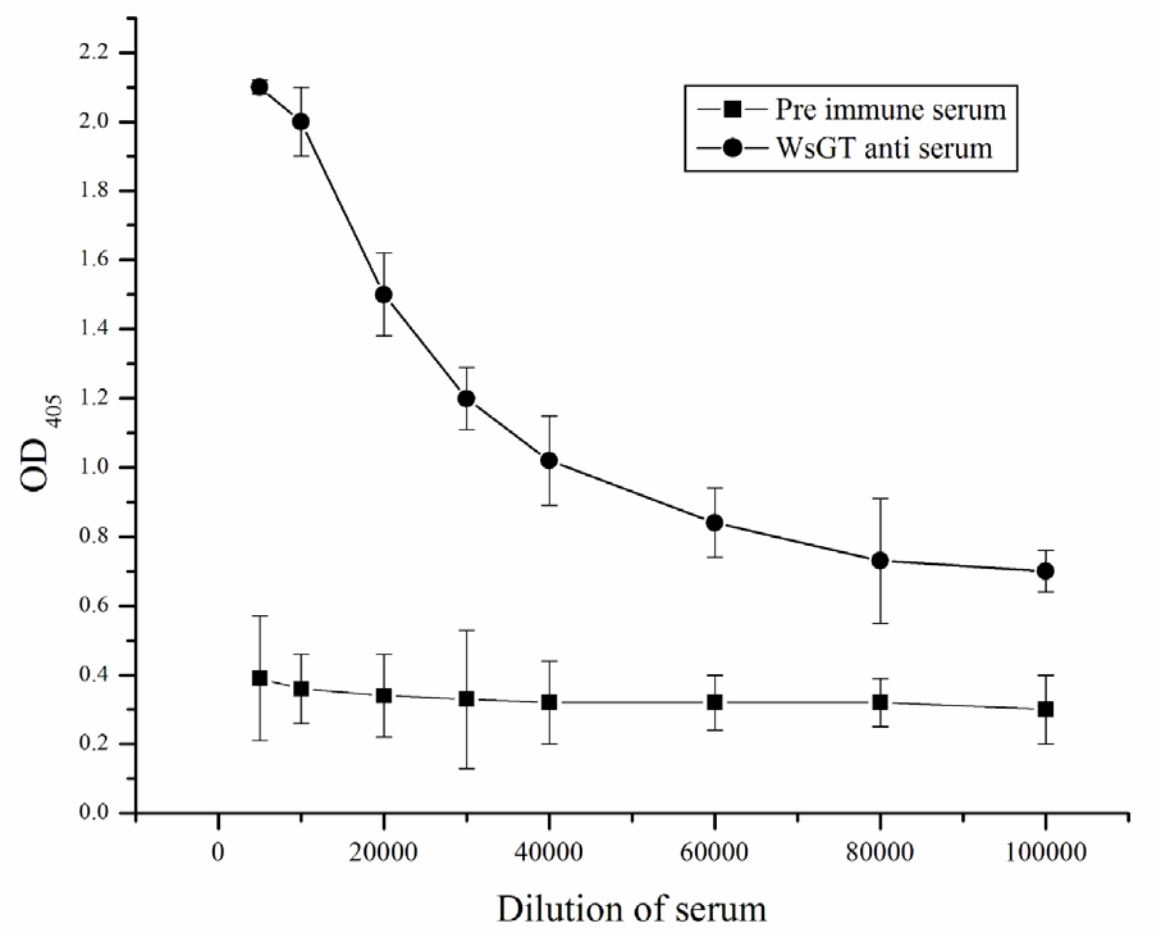


Figure 4.A.6: Determination of titre of 3rd bleed serum; Black square (■) represents OD_{405} of Pre-immune serum and black circles (●) represents OD_{405} of serum dilution.

4.3.7 Protein extraction from plant tissue

Plant tissue was harvested from *W. somnifera* and was crushed to a fine powder in liquid nitrogen. Ground tissue was homogenized in buffer (100 mM phosphate buffer, DTT 5 mM and PMSF 1 mM). Homogenized tissue was centrifuged at 12000 rpm at 4 °C and supernatant was collected in fresh centrifuge tube. Supernatant was quantified for protein using Bradford assay (Bradford reagent, Bio-rad).

4.3.8 Western blot analysis

The Purified WsGT protein was used to raise polyclonal antibodies in rabbit. The raised anti-WsGT polyclonal antibody obtained from 3rd bleed serum was diluted 1:5000 times and was used for western blot experiment. Secondary antibody i.e. goat anti-rabbit IgG conjugated with alkaline phosphatase was purchased from Bangalore Genei, Bangalore, India.

iBlot Gel Transfer System was used to transfer the protein from SDS-PAGE gel by strictly following manufacturer's instructions (Invitrogen, USA). The PVDF membrane containing transferred protein samples were processed as per standard procedure, blocking, treatment with primary antibody (anti-WsGT antibody Chapter 4, Section 4.3.6), washing followed by treatment with secondary antibody which is ALP conjugated, followed by washing and colour development using BCIP/NBT as substrate. The samples subjected for western analysis includes purified recombinant WsGT protein expressed in pET30b (+) expression system and crude protein from leaf, stem and root tissue of *W. somnifera* plant. Total crude protein used for western blot experiment was 100 µg. Purified protein of WsGT was also used as positive control. The plant extract was run on 10% SDS-PAGE and blotted on to PVDF membrane (Figure 4.A.7).

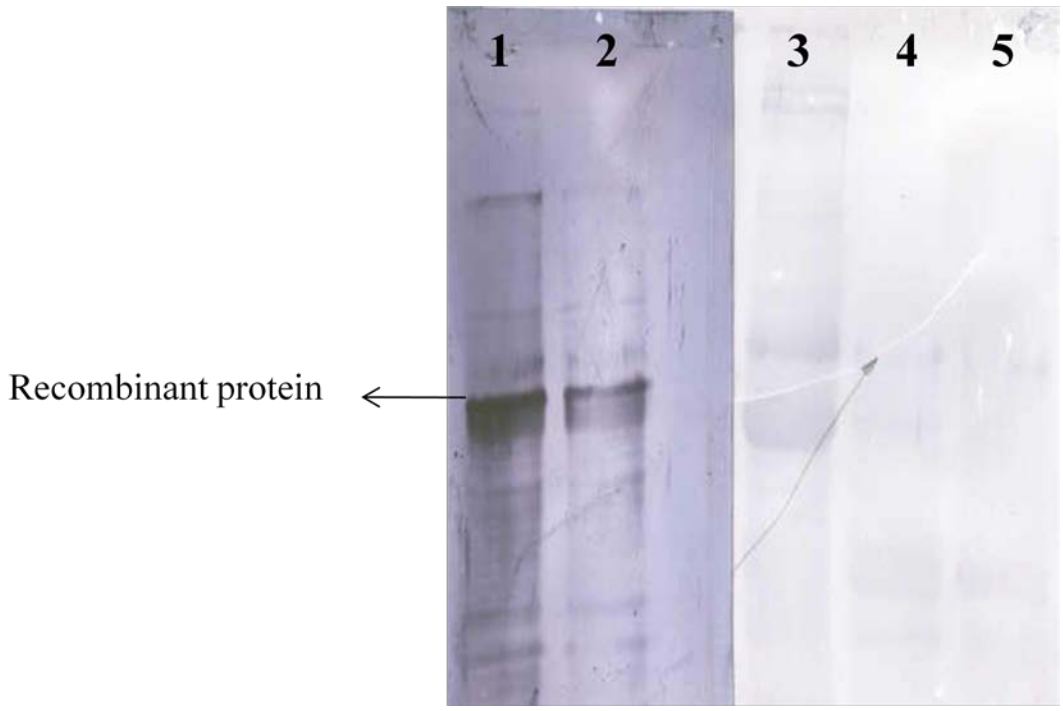


Figure 4.A.7: Western blot analysis. Lane 1&2- Purified recombinant WsGT protein; Lane-3 Leaf extract; Lane-4 Stem extract and Lane-5 Root extract.

Polyclonal antibody raised against WsGT protein was used in Western blot hybridization. In recombinant protein approximately 58 kDa fragment which is hybridized with antibodies and leaf extract also shows the band corresponds to the same molecular weight of WsGT. This band appears very lightly in stem and root extract. Among the expected size fragment the polyclonal antibody was hybridized with back ground protein. It shows that WsGT expression is high in leaf as compared to stem and root.

4.3.9 Extraction of recombinant WsGT protein in soluble form

WsGT clone which showed maximum expression as described above in section 4.3.5 was used for extraction of recombinant protein in soluble form to test the enzyme activity. Temperature and time required after induction was standardized for maximum WsGT protein expression in soluble form that is in cell lysate. The culture was grown at different temperatures ranging from 15 °C, 18 °C, 20 °C, 24 °C, 28 °C and at different time durations 6 h, 8 h, 12 h, 14 h, 18 h and 20 h for optimization of maximum expression of recombinant WsGT protein in soluble form. IPTG concentration was also checked ranging from 0.01 mM, 0.03 mM, 0.05 mM, 0.08 mM, 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM and 1 mM. The optimum parameters were as follows; after initial growth at 37 ^{o}C till A_{600} reached 0.6 to 0.8 cells were induced with 0.03 mM IPTG and grown 14 hours at 18 °C as shown in Figure 4.A.8.

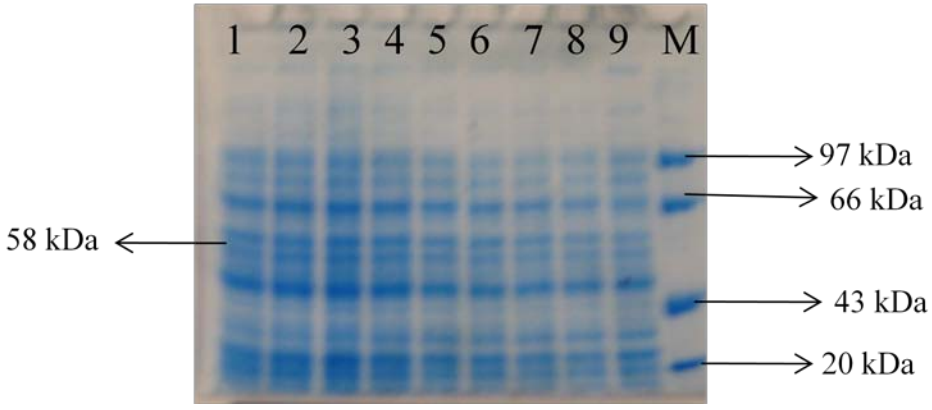


Figure 4.A.8: 10% SDS PAGE analysis of WsGT protein Commassie Blue staining. Lane M- Protein molecular weight marker, Lane 1,2,3,4 and 5- Lysate of WsGT protein grown at 8 h, 12 h, 14 h, 18 h and 20 h respectively, Lane 6, 7, 8 and 9- Inclusion bodies protein of WsGT protein grown at 8 h, 12 h, 14 h and 18 h respectively.

The expression of WsGT protein in soluble form was checked at different temperatures, incubation and IPTG concentration. But the expression of WsGT protein is very low. Also the expected size of protein band is 58 kDa, but there are two bands are present at 58 kDa. So to confirm the exact protein fragment and for enzyme assay protein was purified using Ni-NTA agarose beads. Due to very low expression of recombinant WsGT, protein purification at small scale was not successful. Therefore WsGT was expressed in large quantity and purification was processed using FPLC system (AKTA explorer, GE life science).

4.3.10 Purification of recombinant WsGT soluble protein using FPLC system

The WsGT clone selected for protein expression inoculated in 5mL LB broth and grown at 37 ºC for overnight. This culture was used as a primary culture and inoculated in one liter LB broth as a secondary culture. After inoculation culture was grown at 37 ºC till the $O.D_{600}$ reaches 0.6 to 0.8. After induction with 0.03 mM IPTG culture was grown at 18 ºC for 14 hours. Protein isolation was done from this culture according to protocol mention in chapter 2 section 2.12.1. After protein isolation total protein sample 60 ml was used for Nickel sepharose purification. The details about Nickel sepharose purification given in section 4.3.10.1. The purification chromatogram shown in Figure 4.A.9.

4.3.10.1 Ni Sepharose purification of WsGT recombinant protein

Column: Ni Sepharose 6 Fast Flow, 10 mL in XK 16/20

System: ÄKTAexplorer (GE Healthcare)

Detection: Absorbance, 280 nm

Column Equilibration

Binding buffer: 50 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazole, pH 8.0

Flow rate: 1 mL/min

Volume: 50 mL (till constant baseline for A_{280}, pH, conductivity were obtained)

Sample Application

Sample: Histidine-tagged Glycosyltransferase in Lysis Buffer

Flow rate: 1 mL/min

Volume: 60 mL (adjusted to the composition and pH of binding buffer)

Column Wash

Wash buffer: 50 mM Tris-HCl, 0.5 M NaCl, 30 mM imidazole, pH 8.0

Flow rate: 1 mL/min

Volume: 50 mL (till A_{280}= ↓1 mA.U.)

Protein Elution

Elution buffer: 50 mM Tris-HCl, 0.5 M NaCl, 250 mM imidazole, pH 8.0

Elution: Step elution (100%).

Flow rate: 1 mL/min

Volume: 20 mL

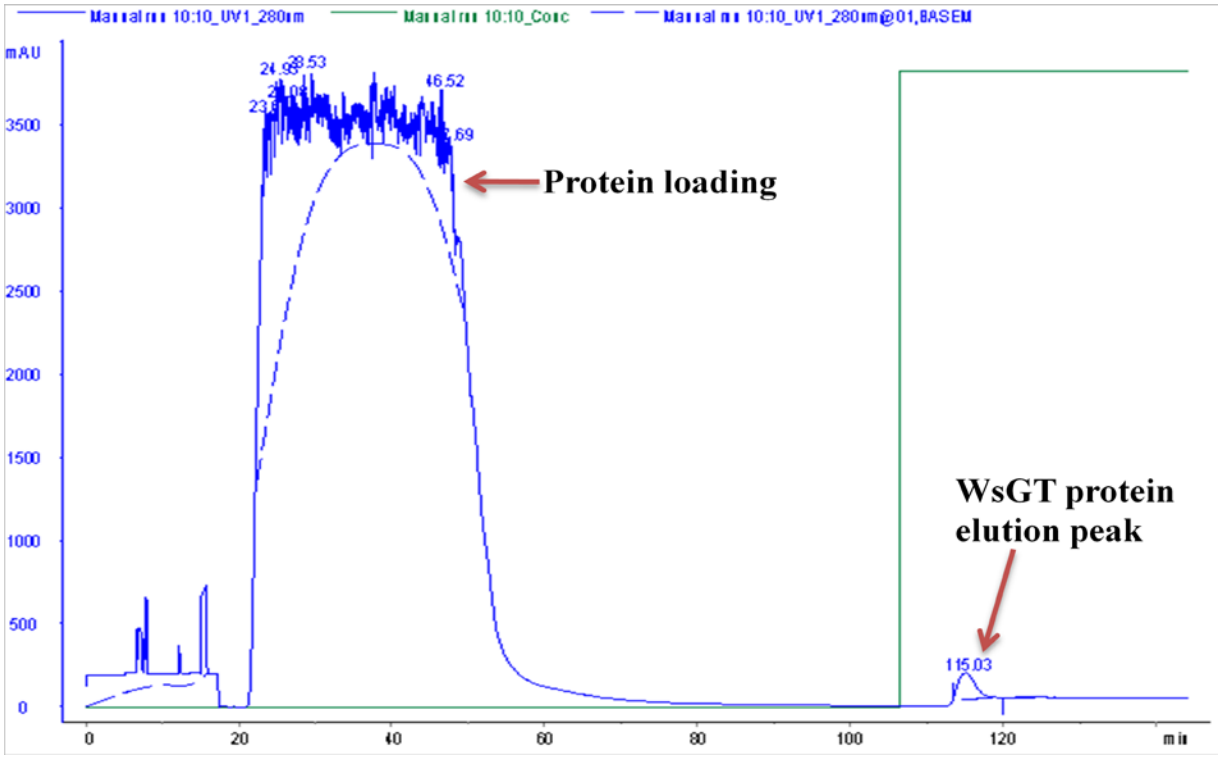


Figure 4.A.9: Purification of WsGT protein using Nickel sepharose column mounted on to AKTA explorer FPLC system.

After Nickel Sepharose purification eluted protein samples were subjected to Bradford estimation (Chapter 2 Section 2.14). Samples were checked on 10% SDS-PAGE gel (Chpater 2 Section 2.12.3) for detecting the contamination. In SDS gel it shows large number of contaminating proteins along with WsGT recombinant protein (58 kDa). To eliminate the contamination second step anion exchange purification was performed. Before loading on anion exchange column, samples were desalted to remove excess amount of salts. Desalting was performed on Hi-Trap 5 mL desalting Sephadex G25 column mounted on FPLC system. Desalting elution was done with 50 mM Tris-HCl pH-8.0 and protein chromatogram shown in Figure 4.A.10. Desalted protein elution was collected and injected on anion exchange column.

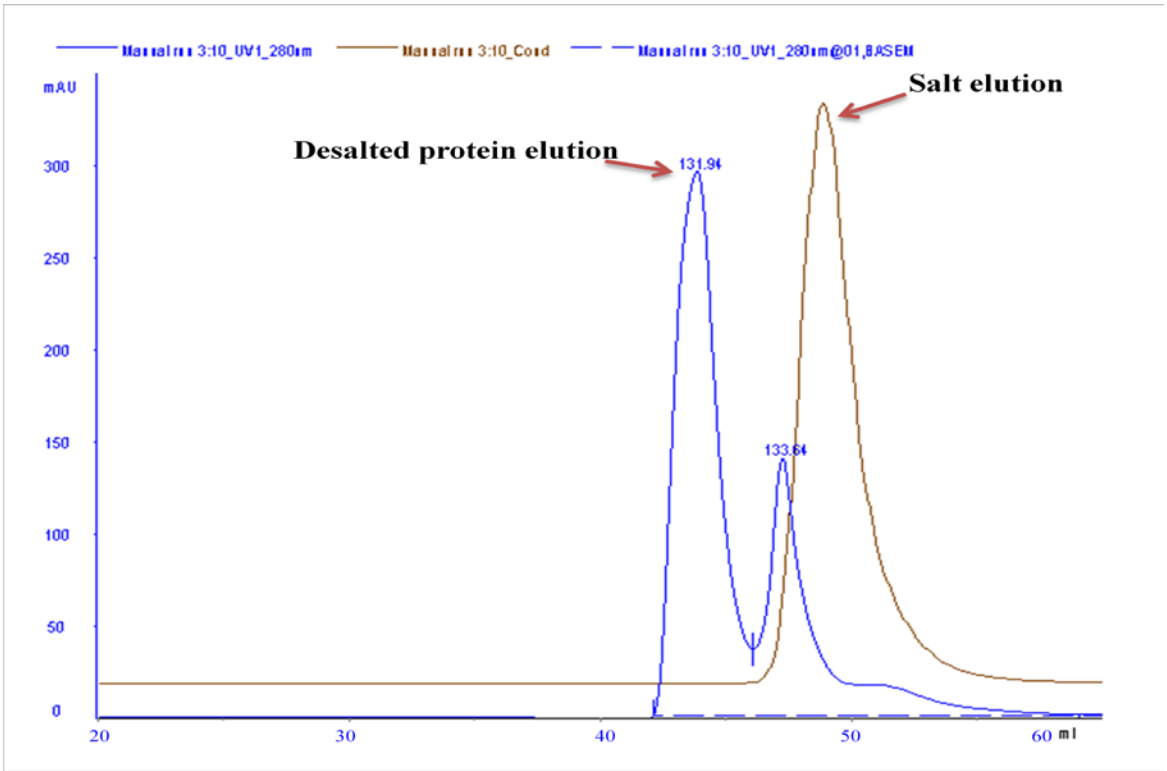


Figure 4.A.10: Desalting of WsGT Nickel Sepharose purified protein

4.3.10.2 Anion exchange purification of Nickel sepharose purified & desalted WsGT protein

Column: Q-XL HiTrap, 1 mL pre-packed

System: ÄKTAexplorer (GE Healthcare)

Detection: Absorbance, 280 nm

Column Equilibration

Binding buffer: 20 mM Tris-HCl, pH 8.0

Flow rate: 1 mL/min

Volume: 20 mL (till constant baseline for A_{280}, pH, conductivity were obtained)

Sample Application

Sample: Ni- Sepharose purified & desalted

Flow rate: 0.5 mL/min

Volume: 2 mL

Column Wash

Wash buffer: 20 mM Tris-HCl, pH 8.0

Flow rate: 1 mL/min

Volume: 20 mL (till A_{280}= ↓1 mA.U.)

Protein Elution

Elution buffer: 20 mM Tris-HCl, 1 M NaCl, pH 8.0

Elution: Gradient elution

Gradient: 0–100% elution buffer in 60 min (0–1 M NaCl)

Flow rate: 0.5 mL/min

Volume: 30 mL

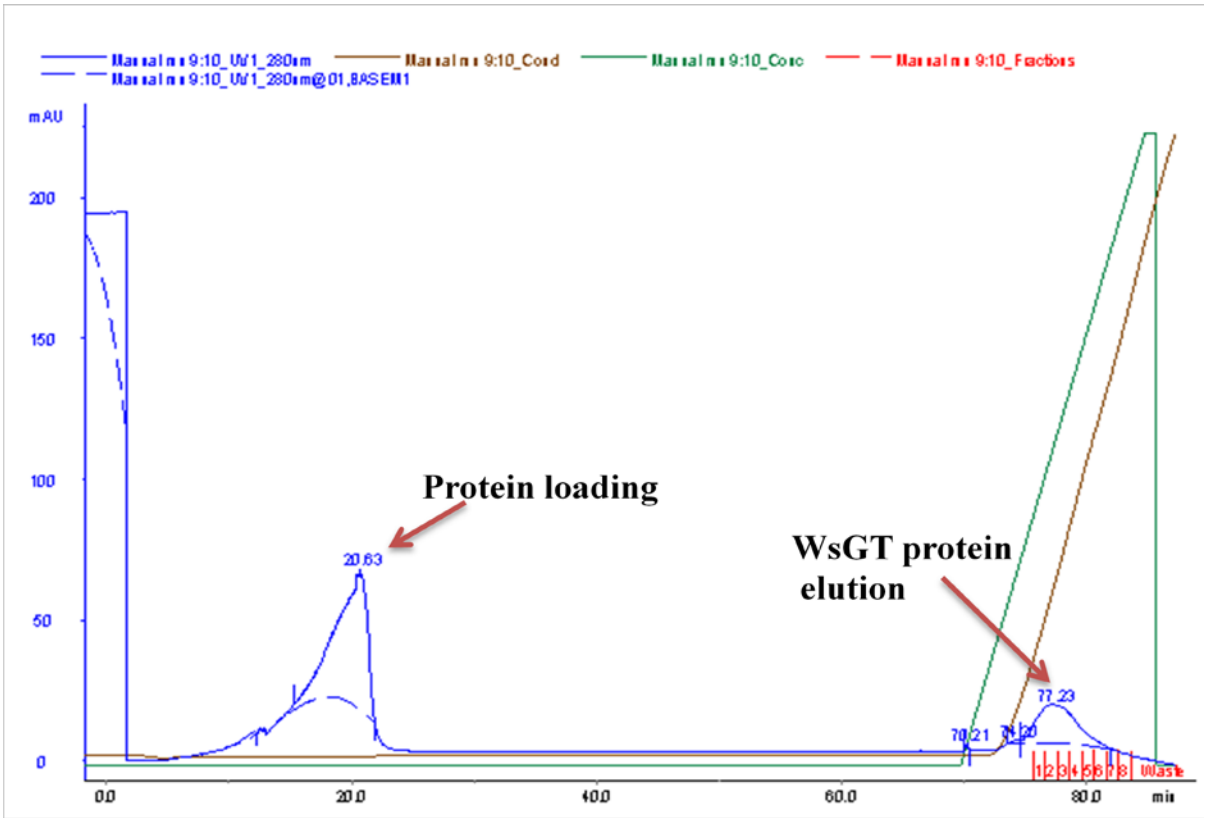


Figure 4.A.11 Second step anion exchange purification of Nickel sepharose purified WsGT protein using Q-XL HiTrap column mounted on to AKTA explorer FPLC system.

Anion exchange purified WsGT protein elution were collected and quantified using Bradford assay method (Chapter 2 Section 2.14). Protein was checked on 10% SDS-PAGE gel. Large number of contaminated proteins was washed out in anion exchange purification. The Nickel purified and anion exchange purified WsGT protein shown in silver stained gel (Figure 4.A.12).

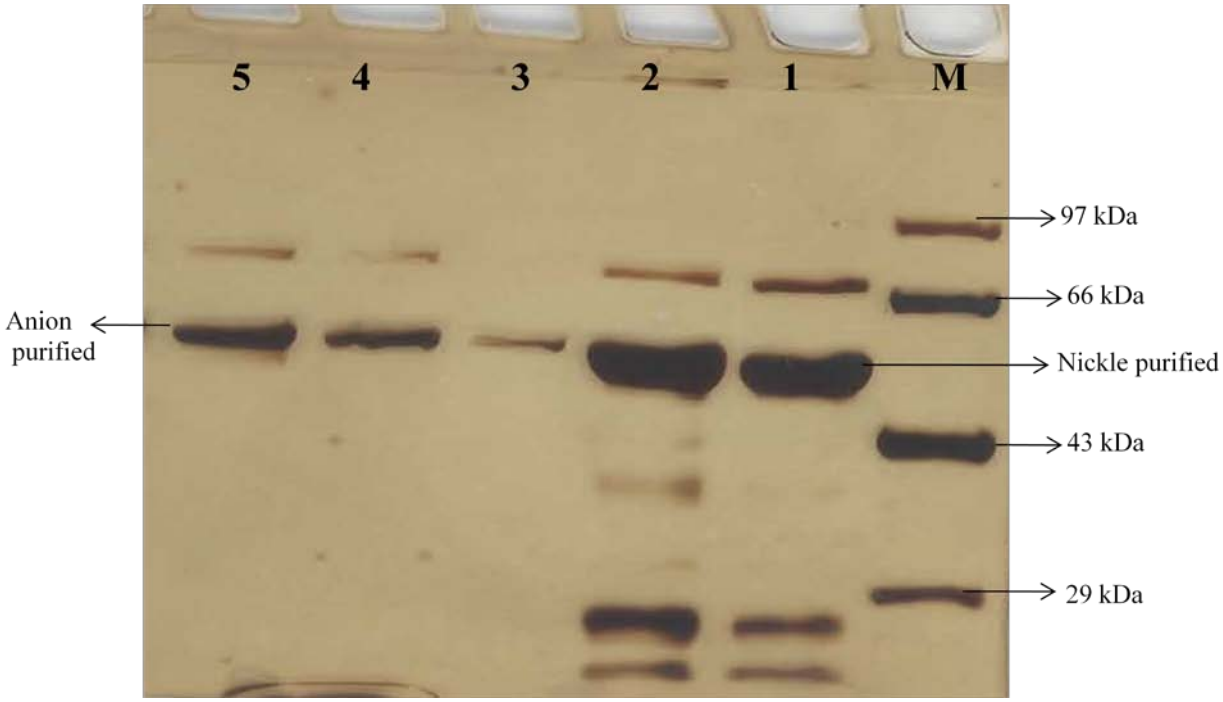


Figure 4.A.12: Nickel Sepharose purified and anion exchange purified WsGT protein 10% SDS gel. Lane-M Protein molecular weight marker; Lane 1 & 2 Nickel Sepharose purified WsGT protein (~58 kDa); Lane 3, 4 & 5 Anion exchange purified WsGT protein (~58 kDa).

4.4 2D Gel electrophoresis of WsGT protein

2-D electrophoresis is a powerful and widely used method for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples. This technique separates proteins according to two independent properties in two discrete steps.

The first-dimension step, isoelectric focusing (IEF), separates proteins according to their isoelectric points (pI); the second-dimension step, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their molecular weights (M, relative molecular mass). Each spot on the resulting two-dimensional gel potentially corresponds to a single protein species in the sample. Thousands of different proteins can thus be separated, and information such as the protein pI, the apparent molecular weight, and the amount of each protein can be obtained.

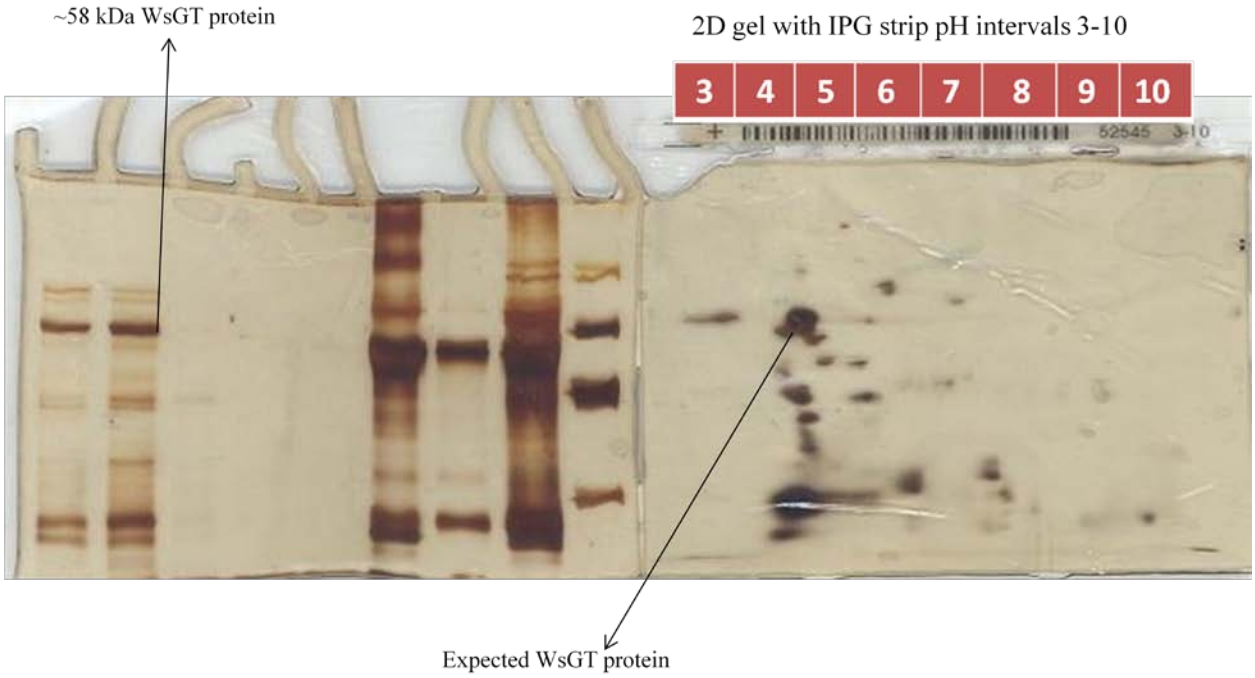


Figure 4.A.13: 2D gel of WsGT protein with IPG strip pH intervals 3 to 10

The Ni-NTA purified WsGT protein was used in 2D gel electrophoresis. The first dimension and second dimension of 2D gel electrophoresis was performed according to manual instructions (Chapter 2 Section 2.13). After second dimension the IPG strip containing gel was silver stained and measured the pI value of expected WsGT protein. The protein shows approximately 5.0 to 5.5 pI value (Figure 4.A.13).

4.5 MALDI MS/MS analysis of WsGT protein

MALDI MS/MS was done for absolutely purified protein as described in Chapter 2, Section 2.16. MS/MS analysis was done for confirmation of recombinant WsGT protein with the sequences already submitted to the Uniprot Database. The MALDI MS/MS spectra and coverage map of WsGT protein is shown in Figure 4.A.14. Ionization spectra shows mass of different peptides of the query protein and coverage map shows number of peptides of query

protein which shows exact match with the template WsGT protein present in the Uniprot database.

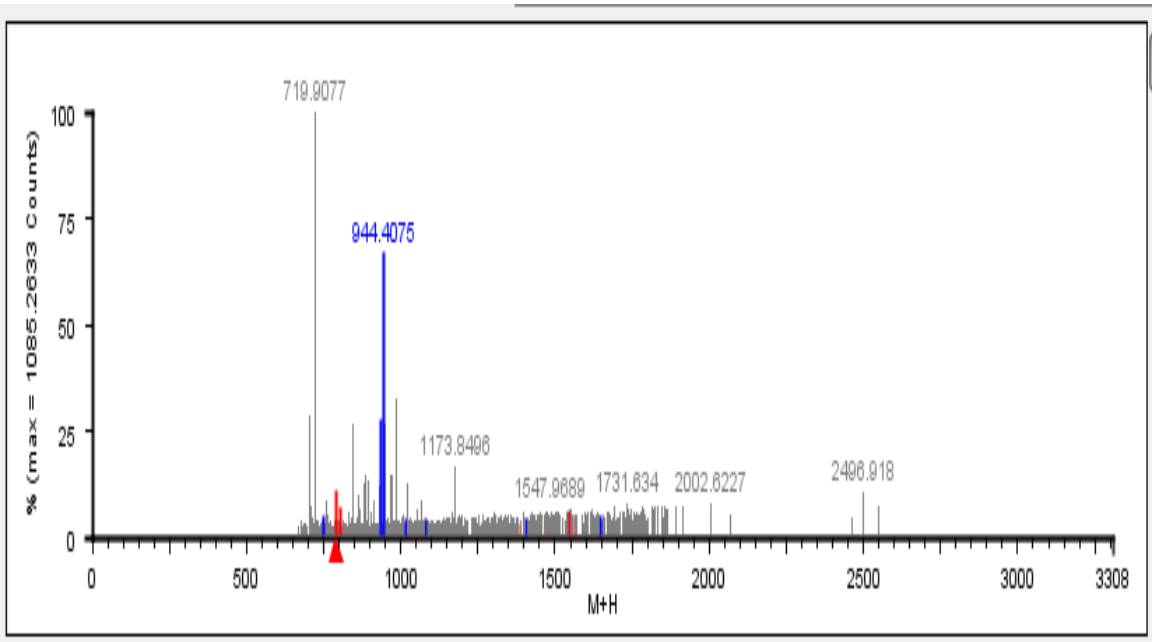


Figure A Ionization spectra

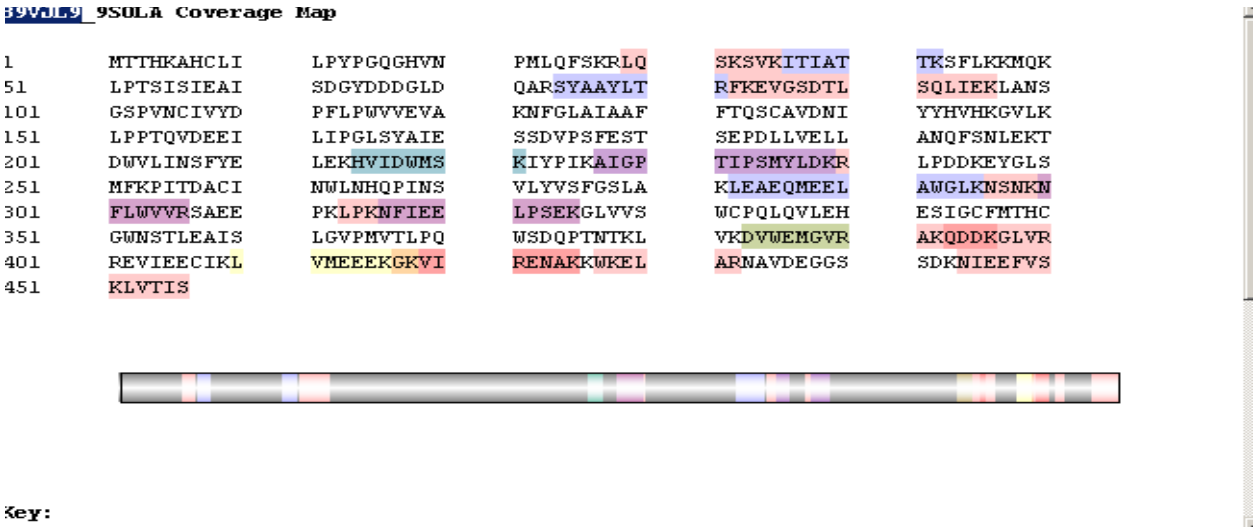


Figure B Coverage map

Figure 4.A.14: MALDI MS/MS analysis of recombinant WsGT protein (A) ionization spectra (B) coverage map of recombinant WsGT protein.

4.6 Glycosyltransferase enzyme assay

The reaction conditions for WsGT enzyme assay i.e. pH, temperature, time duration and glucose donor was optimized using different parameters. Enzyme shows maximum specificity with diadzein substrate. Therefore, diadzein substrate was used to optimize the pH, temperature, time duration and sugar donor.

4.6.1 Optimization of pH

To optimize the pH different pH i.e. 4.0 to 12.0 range was used in assay conditions. The enzyme shows maximum activity at pH 7.0 to 8.0 (Figure 4.A.15). The theoretical pI value of WsGT is 5.32. 2-D gel electrophoresis also shows the pI value range 5.0 to 5.5.

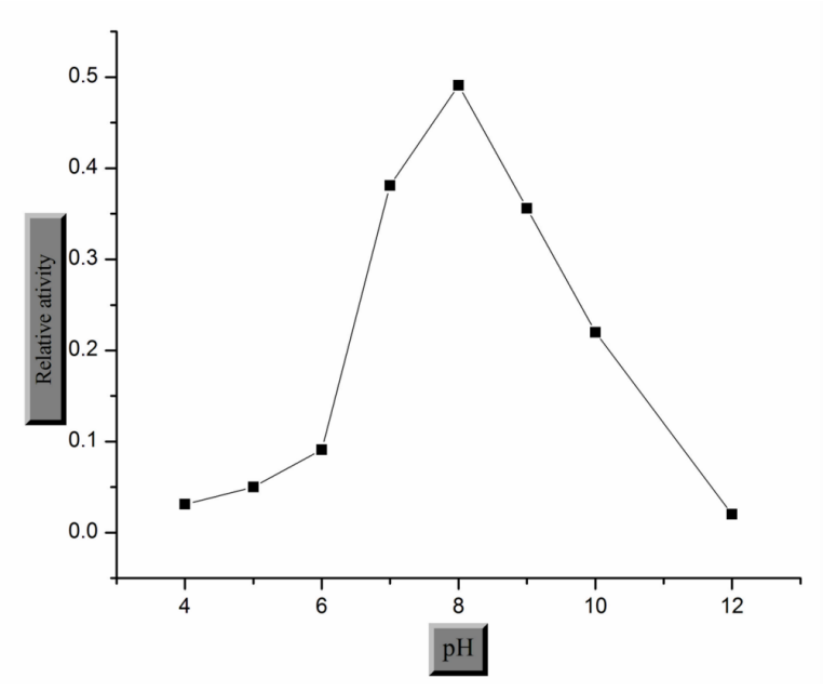


Figure 4.A.15: WsGT enzyme activity at different pH conditions

4.6.2 Optimization of temperature

To optimize the temperature WsGT enzyme assays were incubated at different temperatures from 20 ºC to 50 ºC. Standard enzyme assay conditions were used. WsGT enzyme shows maximum activity at 30 ºC to 32 ºC (Figure 4.A.16).

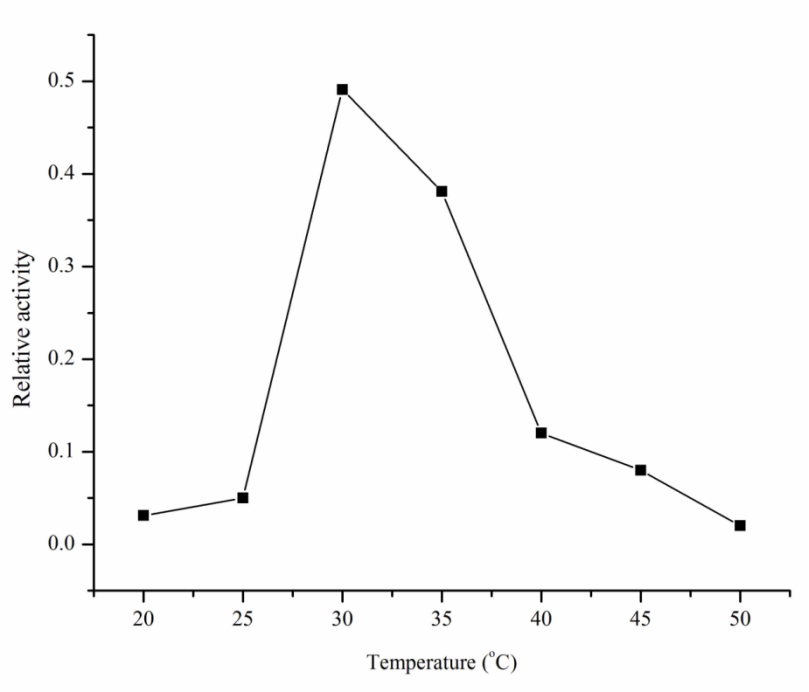


Figure 4.A.16: WsGT enzyme assays at different temperatures

4.6.3 Optimization of Time duration

WsGT enzyme assays were incubated at different time durations, to know at which time duration enzyme produce more concentration of glycosides. Assays were incubated from 30 min to 3 hours time duration. Enzyme produces enough concentration of glycosides at 3 hours time duration and shows linear range for 2h to 3h (Figure 4.A.17).

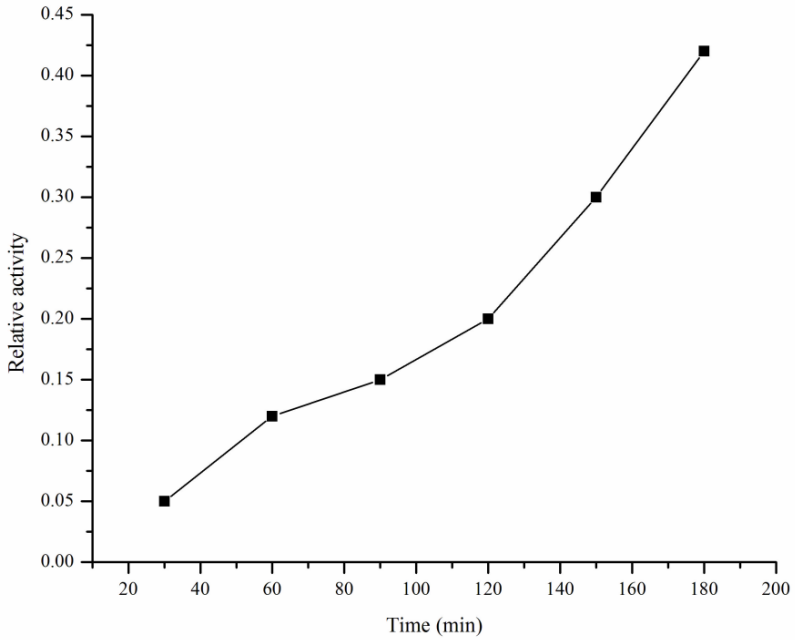


Figure 4.A.17: WsGT enzyme assays at different time duration

4.6.4 Sugar donor

Various sugar donors i.e. UDP-Glucose, UDP-Galactose and UDP-Xylose were used in assay reaction to find out which sugar donor is suitable for enzyme assay. UDP-Glucose is the suitable sugar donor for WsGT enzyme assay. The UDP-Glucose concentration used for assay reaction is 5 mM.

4.6.5 WsGT enzyme assay

The anion exchange purified protein was stored at +4 ºC and used for further enzyme assay reactions. The activity of WsGT protein was checked by glycosylation reaction of various aglycone moieties by sugar donor molecule. The assay reaction (0.5 mL) contained 20 mM Tris-HCl buffer pH 8.0, 5 mM glycosyl donor (UDP-glucose), 200 µM glycosyl acceptor and 30µg of purified WsGT enzyme. The reaction was incubated at 30 ^{o}C for 3 hours. After incubation the reaction was extracted thrice by addition of equal volume of ethyl acetate. The extracted reaction was then completely vacuum dried and dissolved in appropriate volume of

methanol. The substrates and the glycosylated products were then analyzed by LC-MS and HPLC. Same reaction was run for control (i.e. lysate of pET30b (+) vector transformed in *E. coli* BL21) with all the substrates, in order to check the background activity if any.

4.6.6 LC-MS analysis of assay products

LC-MS was done as described in Chapter 2, Section 2.18. Different substrates (Table 4.B.1) were used in order to check the specificity of the WsGT enzyme. Almost 31 different substrates belonging to different classes like flavonols, flavones, benzoic acids, isoflavones, flavanones, phenolic and sterol compounds were checked using LC-MS. But WsGT enzyme showed expected mass only with four substrates i.e. genistein, naringenin, diadzein and apigenin. All the substrates are cross checked with control reactions i.e. only pET30b (+) lysate transformed in BL21 (DE3) cells. The LC-MS profiles are given below. Naringenin profile shown in Figure 4.A.18, apigenin 4.A.19, diadzein 4.A.20 and genistein 4.A.21.

Table 4.B.1 Substrates used in LC-MS reaction

S.No	Substrates	Assay result	S.No	Substrates	Assay result
1	Genistein	+ve	16	Hesperitin	-ve
2	Apigenin	+ve	17	DK 14	-ve
3	Naringenin	+ve	18	DK 40	-ve
4	Diadzein	+ve	19	Neist 28	-ve
5	Luteolin	-ve	20	DK 114	-ve
6	Kaempferol	-ve	21	DK 172	-ve
7	Salicylic acid	-ve	22	DK 238	-ve
8	3 HF	-ve	23	Digitoxigenein	-ve
9	Myricetin	-ve	24	Sarsasapogenin	-ve
10	Iso ramanatin	-ve	25	Tomatidine	-ve
11	Methyl Vanillate	-ve	26	Thymol	-ve
12	Vanillic acid	-ve	27	Euginol	-ve
13	Diosgenin	-ve	28	Capsaicin	-ve
14	Catechin	-ve	29	Benzoid acid	-ve
15	Resorcinol	-ve	30	Curcumin	-ve
			31	4-nitro phenol	-ve

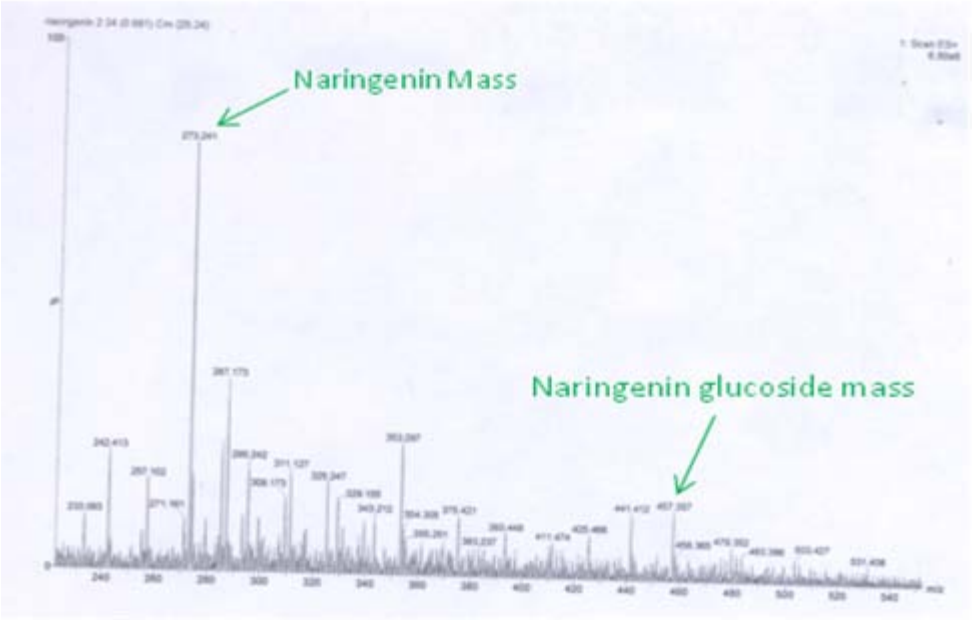


Figure 4.A.18: Naringenin LC-MS profile. Naringenin Mass 273 and naringenin glycoside 457 ($+Na^+$)

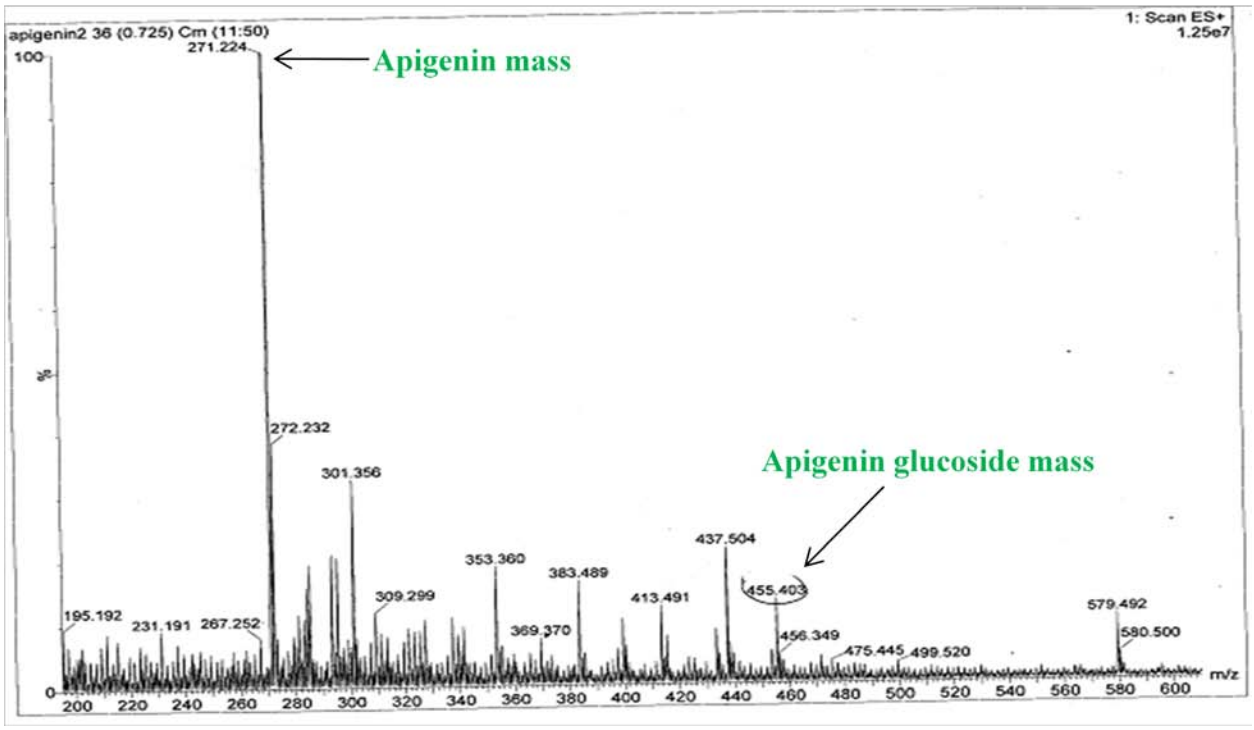


Figure 4.A.19 Apigenin LC-MS profile. Apigenin Mass 271 and apigenin glycoside 455 ($+Na^+$)

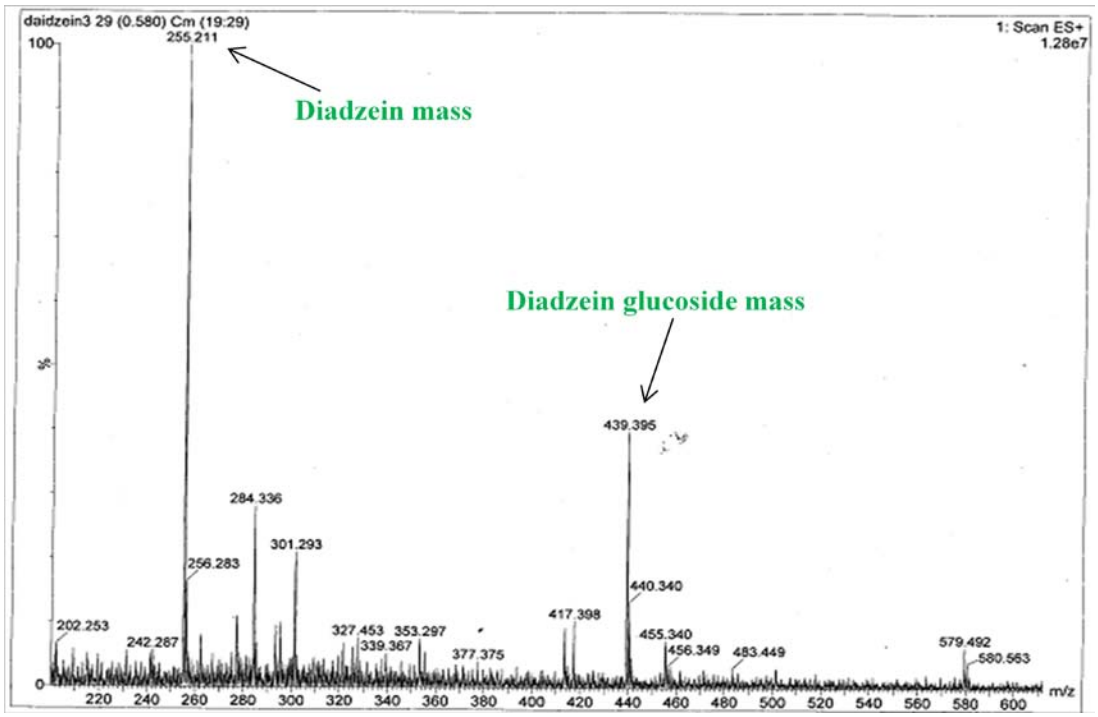


Figure 4.A.20: Diadzein LC-MS profile. Diadzein Mass 255 and diadzein glycoside 439 (+Na^+)

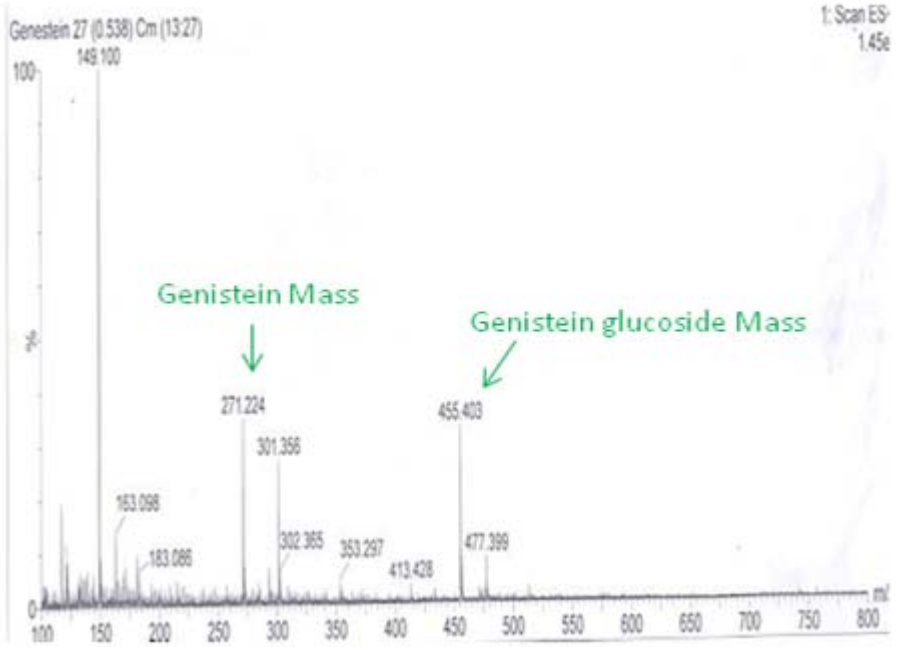


Figure 4.A.21: Genistein LC-MS profile. Genistein Mass 271 and genistein glycoside 455 (+Na^+)

4.6.7 HPLC analysis of enzyme products

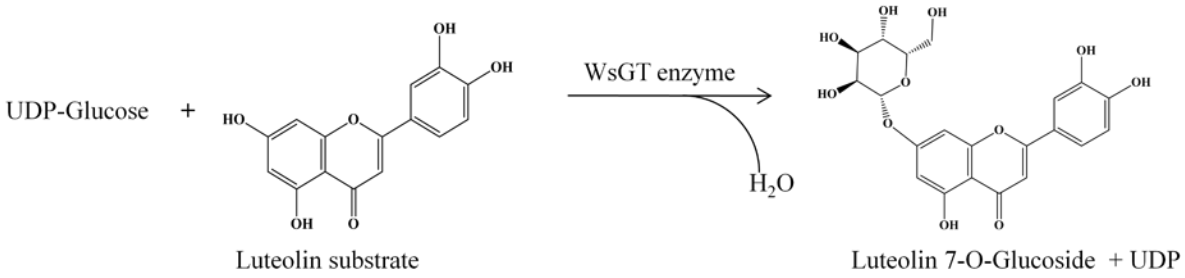
On the basis of LC-MS results genistein, naringenin, apigenin, diadzein and luteolin were used as substrates and only UDP-glucose was used as donor molecule for HPLC analysis for further confirmation. Also few other substrates were analyzed on HPLC, comparing with standard glycoside products (Table 4.B.2). Standard glycoside products were purchase from chromadex, USA. Analysis of the above mentioned substrates and their glycosylated products was done by using reversed-phase HPLC (Perkin Elmer series 200) on a Discovery supelco HS C18 column (5 µm, 25 cm x 4.6 mm, Sigma). The substrates and products were detected using a diode array detector (Perkin Elmer). The substrates and products were detected at different wavelength.

Table 4.B.2 Substrates used in HPLC analysis

Group of Flavonoids	Substrates	WsGT Activity	Glycosylation position
Flavonols	3-Hydroxy flavone	-Ve	
	Kaempferol	-Ve	
	Isorhamnetin	-Ve	
Flavononones	Hesperitin	-Ve	
	Naringenin	+Ve	7-OH
Iso-Flavones	Daidzein	+Ve	7-OH
	Genistein	+Ve	7-OH
Flavones	Apigenin	+Ve	7-OH
	Luteolin	+Ve	7-OH

4.6.8 Luteolin enzyme assay

Purified WsGT recombinant protein was used in luteolin enzyme assay. Assay conditions were given in Chapter 2 Section 2.17. The mobile phase used for separation of enzyme product and wavelengths used according to protocol given in Chapter 2 Section 2.19.1.



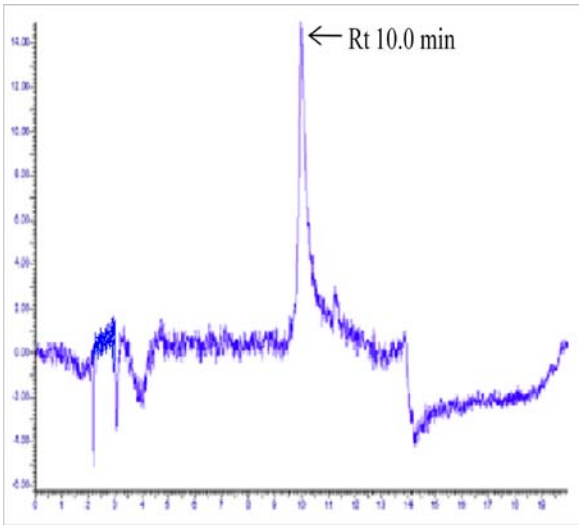


Figure 4.A.22 A Luteolin 7-O-Glucoside Standard

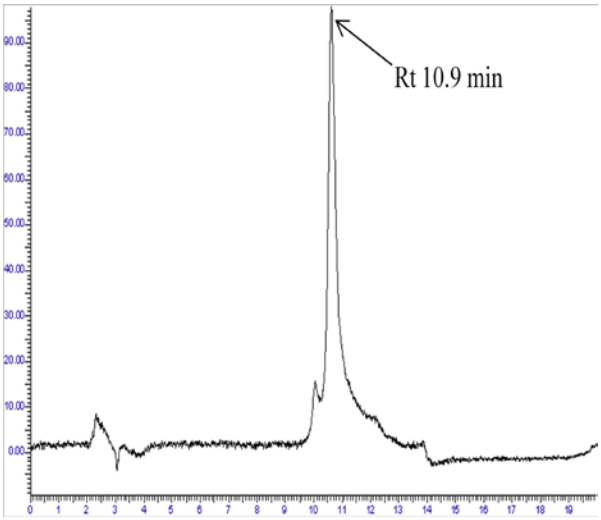


Figure 4.A.22 B Luteolin 4-O-Glucoside standard

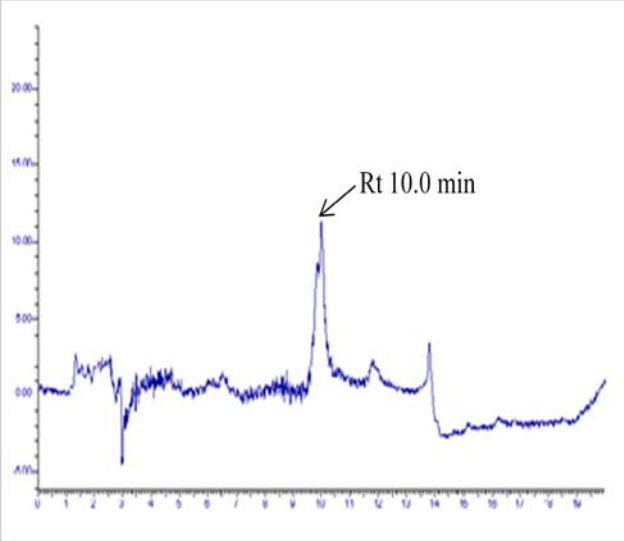


Figure 4.A.22 C Luteolin assay

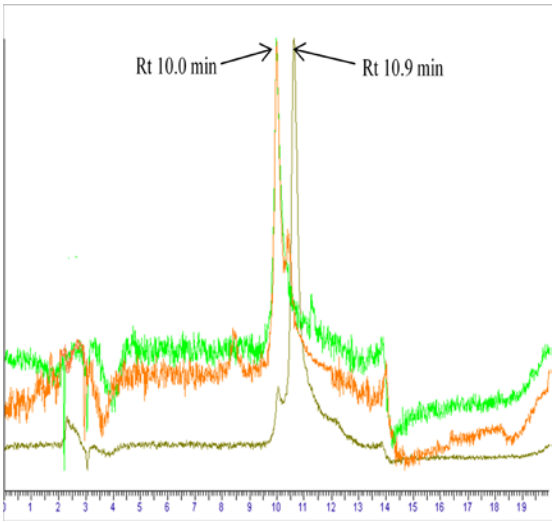


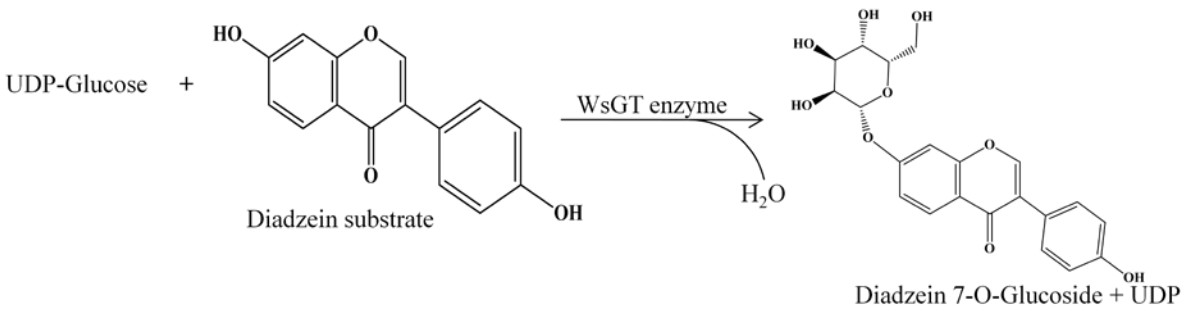
Figure 4.A.22 D Combined luteolin standard and assay

Figure 4.A.22: Luteolin HPLC profile.

Luteolin 7-O-Glucoside retention time is 10.0 min (**Figure 4.A.22 A**), and Luteolin 4-O-Glucoside retention time is 10.9 min (**Figure 4.A.22 B**), luteolin assay product retention time is 10.0 min (**Figure 4.A.22 C**). When all these chromatograms were combined it shows luteolin assay product and luteolin 7-Glucoside standards have same retention time 10.0 min (**Figure 4.A.22 D**). It means WsGT enzyme glycosylates the luteolin at 7^{th} position and it forms luteolin 7-O-Glucoside product.

4.6.9 Diadzein enzyme assay

Purified WsGT recombinant protein was used in diadzein enzyme assay. Assay conditions were given in Chapter 2 Section 2.17. The mobile phase used for separation of enzyme product and wavelengths used according to protocol given in Chapter 2 Section 2.19.2.



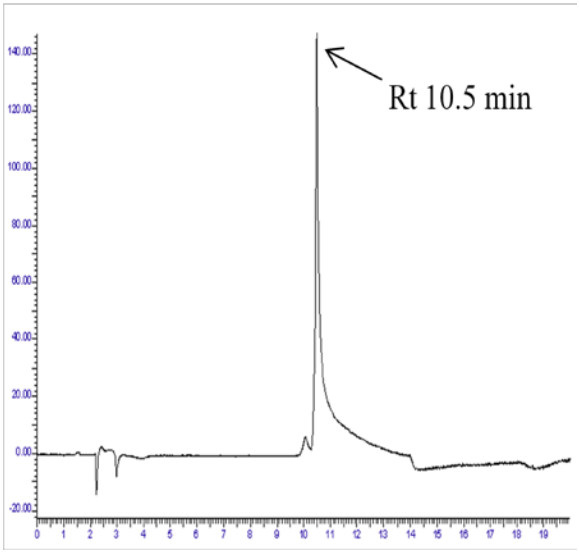


Figure 4.A.23 A Diadzein 7-O-Glucoside standard

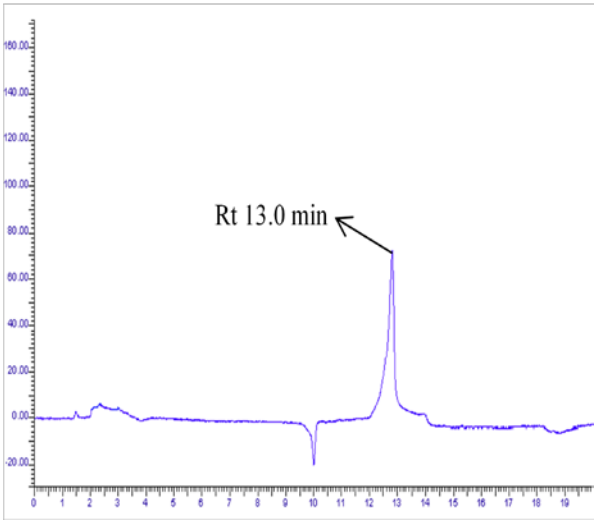


Figure 4.A.23 B Diadzein substrate

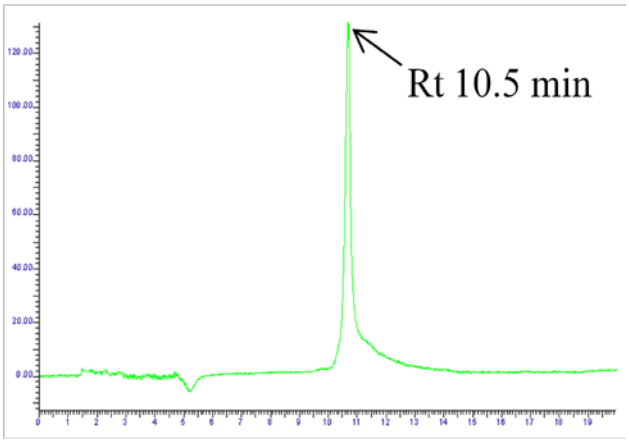


Figure 4.A.23 C Diadzein enzyme assay

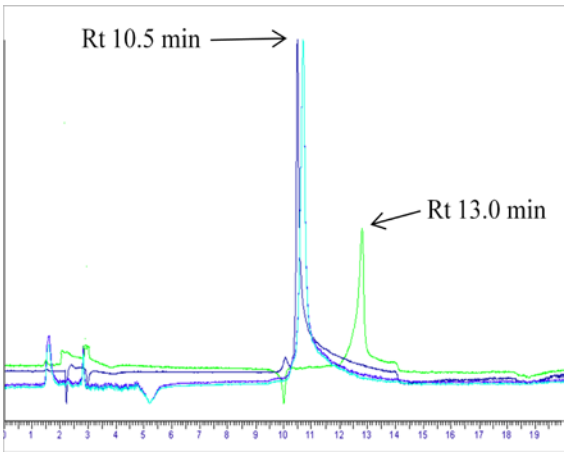


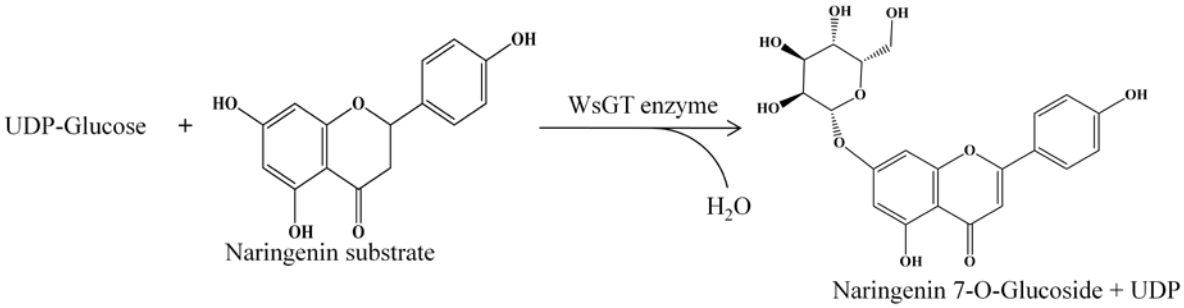
Figure 4.A.23 D Combined diadzein standard and assay

Figure 4.A.23: Diadzein HPLC profile

Diadzein 7-O-Glucoside retention time is 10.5 min (**Figure 4.A.23 A**), and diadzein substrate retention time is 13 min (**Figure 4.A.23 B**), diadzein assay product retention time is 10.5 min (**Figure 4.A.23 C**). When all these profiles were combined it shows diadzein assay product and diadzein 7-O-Glucoside standards have same retention time 10.5 min (**Figure 4.A.23 D**). It means WsGT enzyme glycosylates the diadzein substrate at 7th position and it forms diadzein 7-O-Glucoside product.

4.6.10 Naringenin enzyme assay

Purified WsGT recombinant protein was used in naringenin enzyme assay. Assay conditions were given in Chapter 2 Section 2.17. The mobile phase used for separation of enzyme product and wavelengths used according to protocol given in Chapter 2 Section 2.19.3.



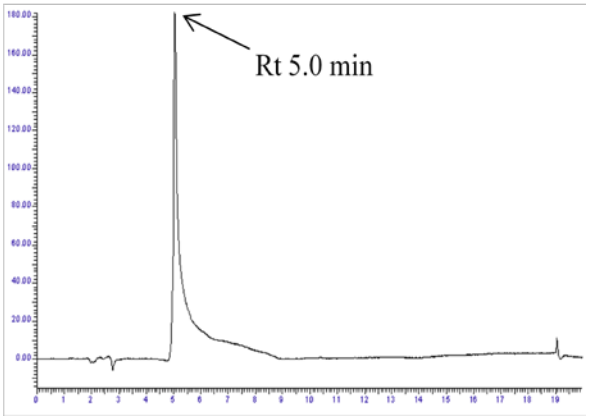


Figure 4.A.24 A Naringenin 7-O-Glucoside standard

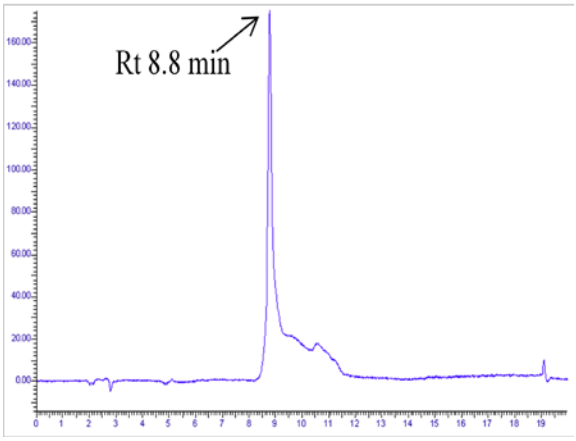


Figure 4.A.24 B Naringenin substrate

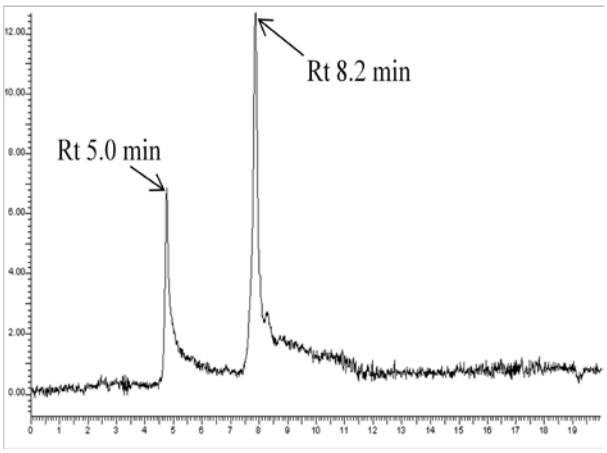


Figure 4.A.24C Naringenin enzyme assay standard

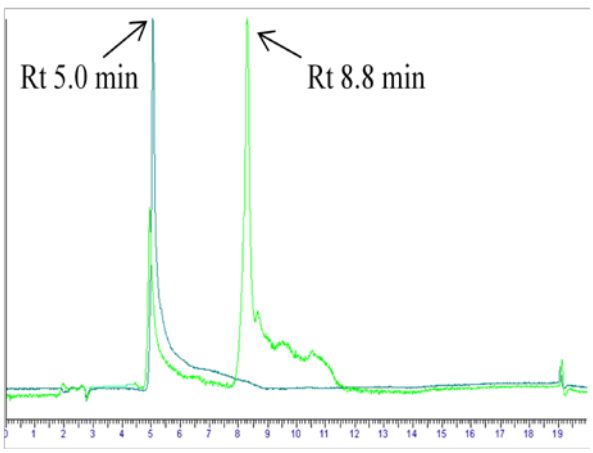


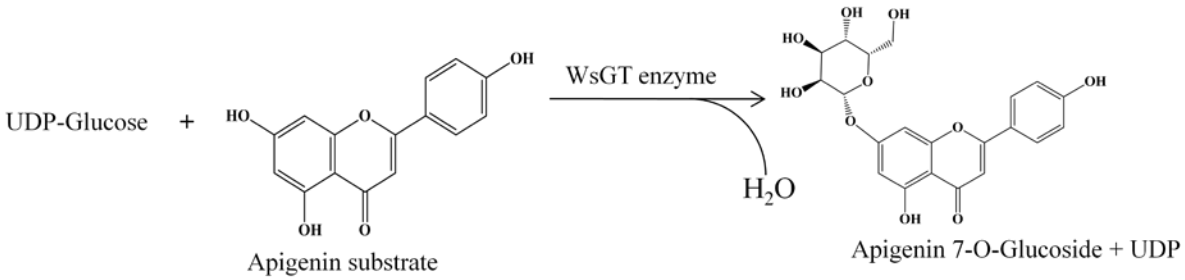
Figure 4.A.24 D Combined naringenin and assay

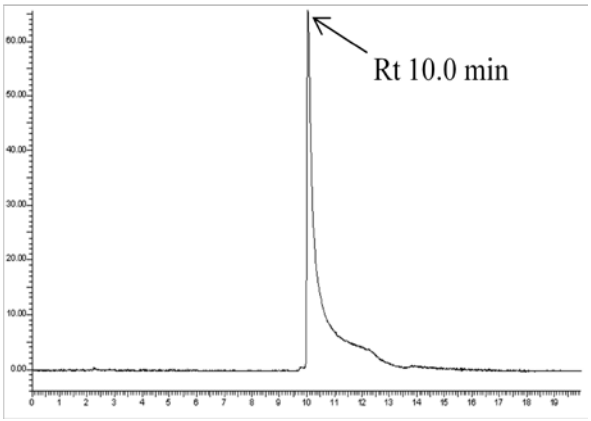
Figure 4.A.24: Naringenin HPLC profile

Naringenin 7-O-Glucoside retention time is 5.0 min (**Figure 4.A.24 A**), and naringenin substrate retention time is 8.8 min (**Figure 4.A.24 B**), naringenin assay product retention time is 5.0 min and assay also contain substrate peak at 8.2 min (**Figure 4.A.24 C**). When all these profiles were combined it shows naringenin assay product and naringenin 7-O-Glucoside standards have same retention time 5.0 min (**Figure 4.A.24 D**). It means WsGT enzyme glycosylates the naringenin substrate at 7^{th} position and it forms naringenin 7-O-Glucoside product.

4.6.11 Apigenin enzyme assay

Purified WsGT recombinant protein was used in apigenin enzyme assay. Assay conditions were given in Chapter 2 Section 2.17. The mobile phase used for separation of enzyme product and wavelengths used according to protocol given in Chapter 2 Section 2.19.4.





. **Figure 4.A.25A Apigenin 7-O-Glucoside standard**

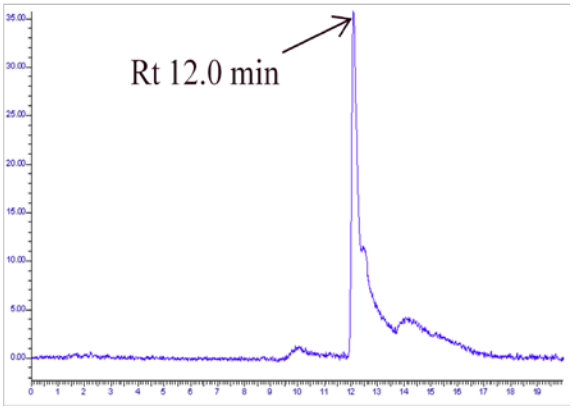


Figure 4.A.25 B Apigenin substrate

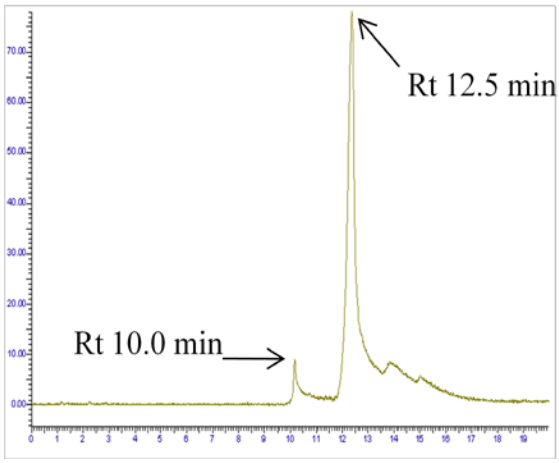


Figure 4.A.25C Apigenin enzyme assay

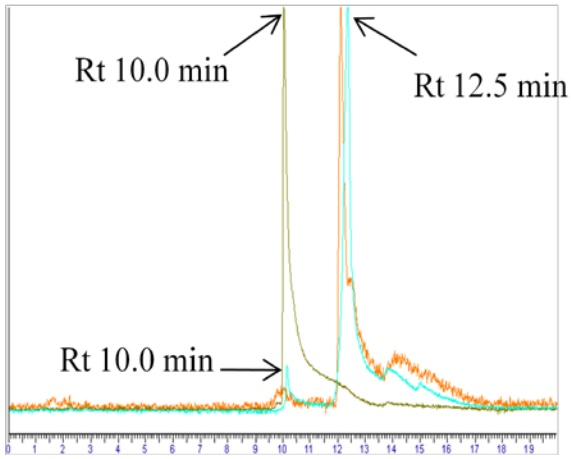


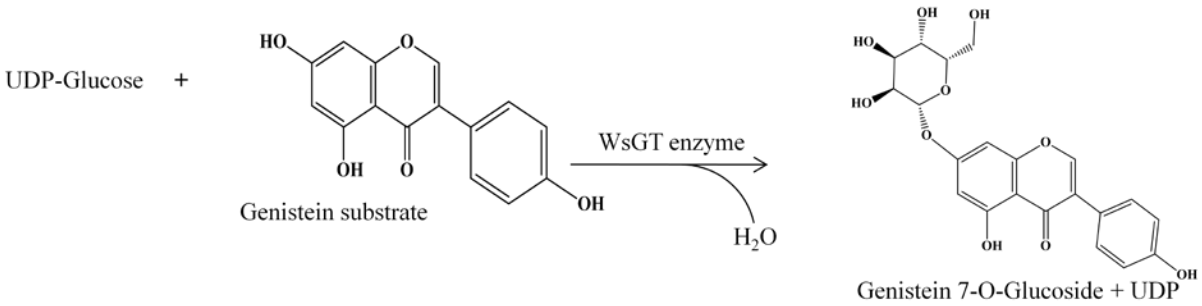
Figure 4.A.25 D Combined apigenin standard and assay

Figure 4.A.25: Apigenin HPLC profile

Apigenin 7-O-Glucoside retention time is 10.0 min (**Figure 4.A.25 A**) and apigenin substrate retention time is 12.0 min (**Figure 4.A.25 B**), apigenin assay product retention time is 10.0 min and assay also contain substrate peak at 12.5 min (**Figure 4.A.25 C**). When all these profiles were combined it shows apigenin assay product and apigenin 7-O-Glucoside standards have same retention time 10.0 min (**Figure 4.A.25 D**). It means WsGT enzyme glycosylates the apigenin substrate at 7^{th} position and it forms apigenin 7-O-Glucoside product.

4.6.12 Genistein enzyme assay

Purified WsGT recombinant protein was used in genistein enzyme assay. Assay conditions were given in Chapter 2 Section 2.17. The mobile phase used for separation of enzyme product and wavelengths used according to protocol given in Chapter 2 Section 2.19.5.



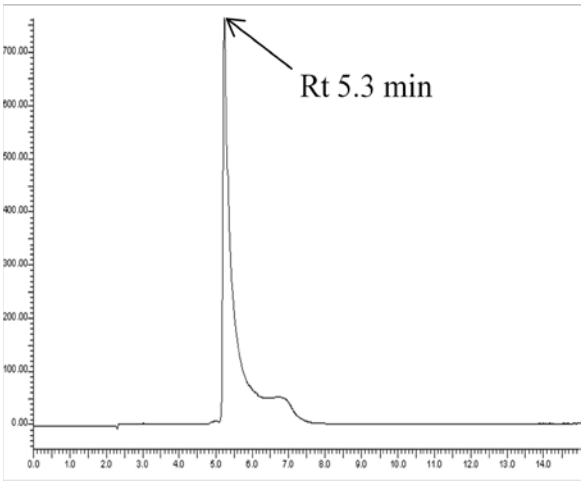


Figure 4.A.26 A Genistein 7-O-Glucoside standard

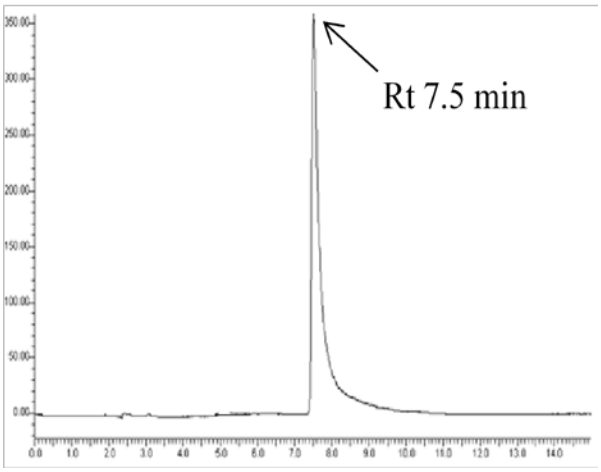


Figure 4.A.26 B Genistein substrate

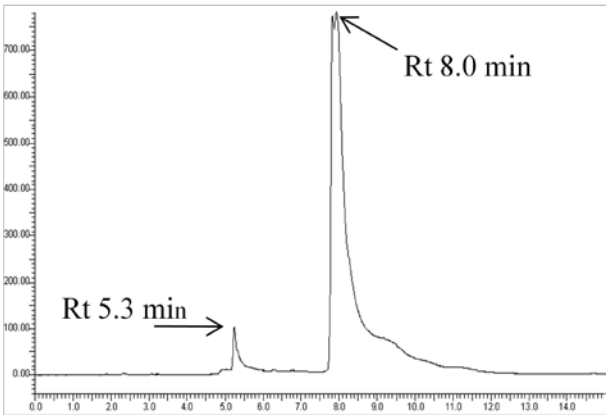


Figure 4.A.26 C Genistein enzyme assay

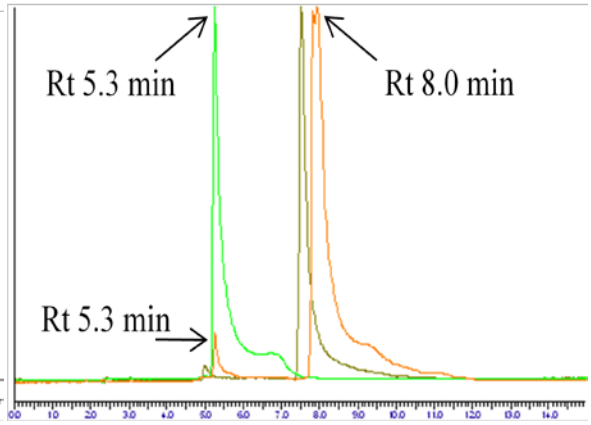


Figure 4.A.26 D Combined genistein standard and assay

Figure 4.A.26: Genistein HPLC profile

Genistein 7-O Glucoside retention time is 5.3 min (**Figure 4.A.26 A**) and genistein substrate retention time is 7.5 min (**Figure 4.A.26 B**), genistein assay product retention time is 5.3 min and assay also contain substrate peak at 8.0 min (**Figure 4.A.26 C**). When all these profiles were combined it shows genistein assay product and genistein 7-O-Glucoside standards have same retention time 5.3 min (**Figure 4.A.26 D**). It means WsGT enzyme glycosylates the genistein substrate at 7th position and it forms genistein 7-O-Glucoside product.

4.7 Enzyme Kinetics

To determine the initial velocity of WsGT, the assays were performed under steady state conditions using the standard assay system with various substrate concentrations. The standard reaction mixture (500 µL) consisted of 5 mM glycosyl donor, 20 mM Tris-Cl buffer, pH 8.0, WsGT enzyme (30 µg) and glycosyl acceptor. Reaction was incubated at 30 ºC for 3 hours. After incubation glycosylated products were extracted with ethyl acetate and analyzed using reverse-phase HPLC on a C18 column. The apparent K_m and V_{max} values for apigenin, naringenin, diadzein, luteolin and genistein were determined by fitting the initial velocity data to the Michaelis-Menten equation using nonlinear regression analysis. Substrates were used at a saturating concentration with constant enzyme concentration (30 µg).

Initially standards of different glycosylated products were used at different concentrations from 0.2 µM to 8 µM to plot a graph. Peak area was calculated for different concentrations of standards. Standard concentration on X-axis and Peak area on Y-axis plotted a graph using Origin6.1 software. Also peak areas were calculated for enzyme assays performed with saturating concentrations of substrates. Considering total product formation, concentration of enzyme and duration of assay time velocity was calculated. Using substrate concentration and velocity values Michaelis-Menten plots were constructed (Graph pad prism5 software).

4.7.1 Diadzein substrate specificity

Diadzein 7-O-Glucoside standard plot was constructed using standard concentrations and peak area. Standard concentration was used from 0.2 µM to 8 µM. Diadzein standard plot is shown in Figure 4.A.27 A. Enzyme assays were performed using different concentrations of diadzein substrate. Standard assay conditions were used with substrate concentration 4 µM to 120 µM. Finally using substrate concentration and velocity values Michaelis-Menten plot were constructed for diadzein substrate (Figure 4.A.27 B).

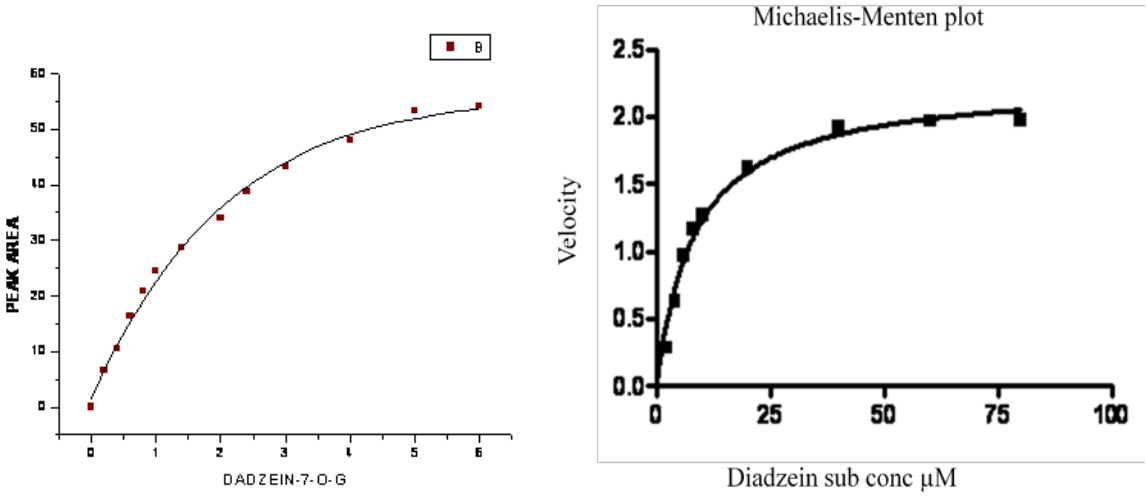


Figure A: Diadzein standard plot **Figure B: Michaelis-Menten plot for diadzein**

Figure 4.A.27: Diadzein substrate kinetics with WsGT enzyme

4.7.2 Naringenin substrate specificity

Naringenin 7-O-Glucoside standard plot was constructed using standard concentrations and peak area. Standard concentrations were used from 0.2 µM to 8 µM. Naringenin standard plot shown in Figure 4.A.28 A. WsGT enzyme assays were performed using different concentrations of naringenin substrate. Standard assay conditions were used with substrate concentrations 2 µM to 80 µM. Finally using substrate concentrations and velocity values Michaelis-Menten plot were constructed for naringenin substrate (Figure 4.A.28 B).

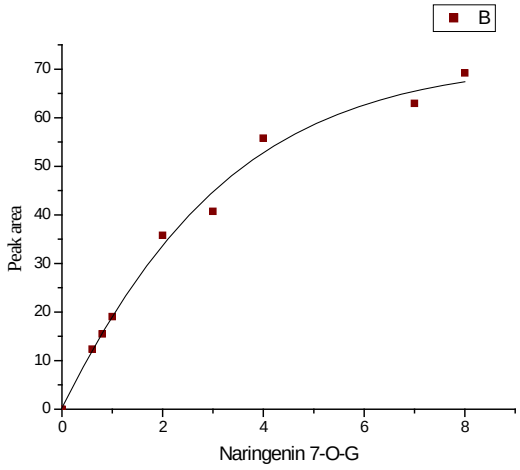


Figure A: Naringenin standard plot

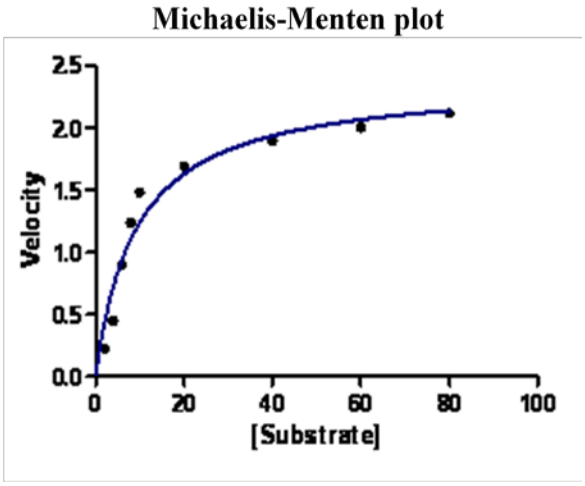


Figure B: Michaelis-Menten plot for naringenin

Figure 4.A.28: Naringenin substrate kinetics with WsGT enzyme

4.7.3 Genestein substrate specificity

Genistein 7-O-Glucoside standard plot was constructed using standard concentrations and peak area. Standard concentrations were used from 0.2 µM to 6 µM. Genistin standard plot shown in Figure 4.A.29 A. WsGT enzyme assays were performed using different concentrations of genistein substrate. Standard assay conditions were used with substrate concentrations 2 µM to 80 µM. Finally using substrate concentrations and velocity values Michaelis-Menten plot were constructed for genistein substrate (Figure 4.A.29 B).

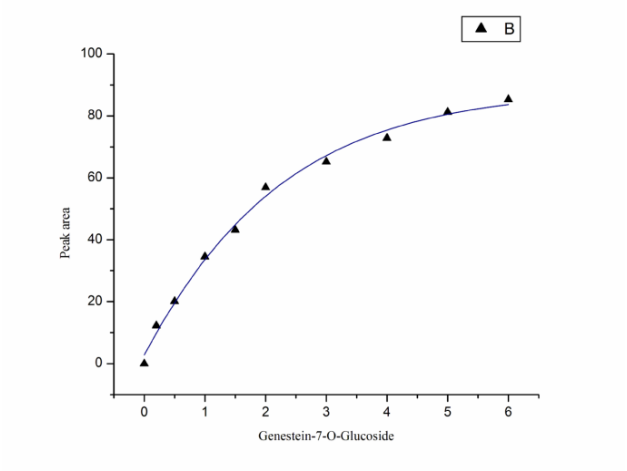


Figure A: Genistein standard plot

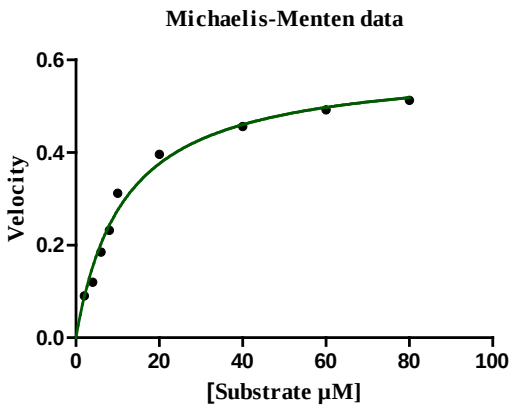


Figure B: Michaelis-Menten plot for genistein

Figure 4.A.29: Genistein substrate kinetics with WsGT enzyme

4.7.4 Luteolin substrate specificity

Luteolin 7-O-Glucoside standard plot was constructed using standard concentrations and peak area. Standard concentrations were used from 0.2 µM to 8 µM. Luteolin standard plot shown in Figure 4.A.30 A. WsGT enzyme assays were performed using different concentrations of luteolin substrate. Standard assay conditions were used with substrate concentrations 2 µM to 100 µM. Finally using substrate concentrations and velocity values Michaelis-Menten plot were constructed for luteolin substrate (Figure 4.A.30 B).

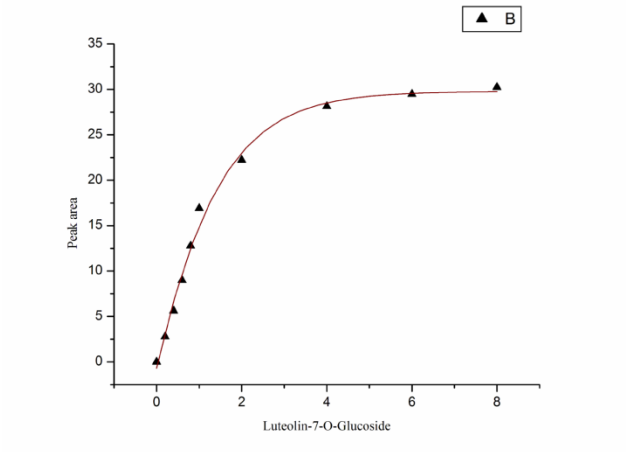


Figure A: Luteolin standard plot

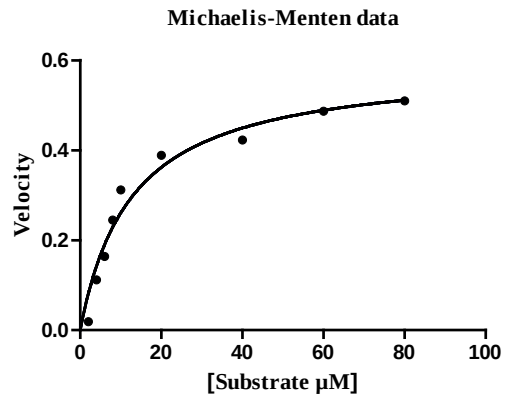


Figure B: Michaelis-Menten plot for luteolin

Figure 4.A.30: Luteolin substrate kinetics with WsGT enzyme

4.7.5 Apigenin substrate specificity

Apigenin 7-O-Glucoside standard plot was constructed using standard concentrations and peak area. Standard concentrations were used from 0.4 µM to 4 µM. Apigenin standard plot shown in Figure 4.A.31 A. WsGT enzyme assays were performed using different concentrations of apigenin substrate. Standard assay conditions were used with substrate concentrations 2 µM to 80 µM. Finally using substrate concentrations and velocity values Michaelis-Menten plot were constructed for apigenin substrate (Figure 4.A.31 B).

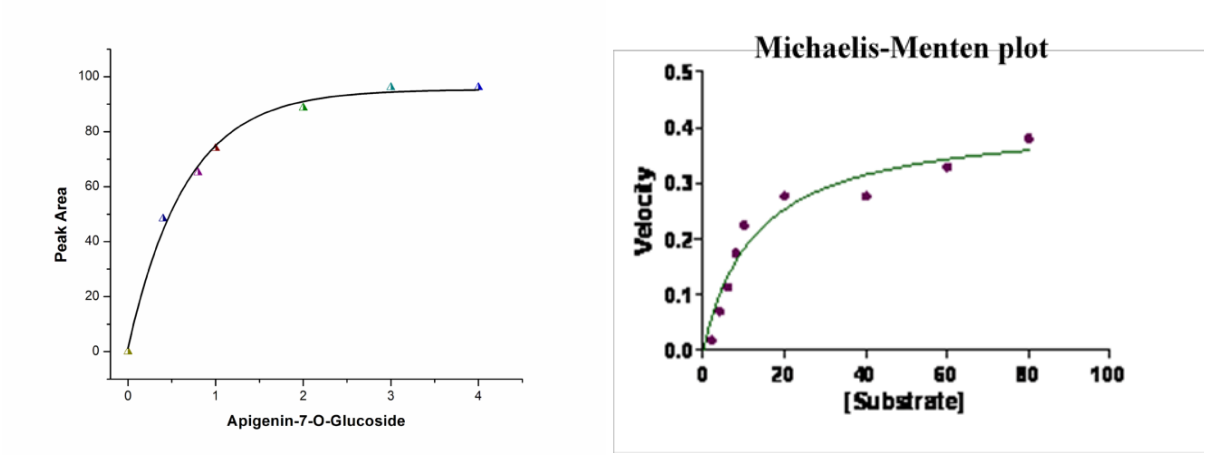


Figure A: Apigenin standard plot **Figure B: Michaelis-Menten plot for apigenin**

Figure 4.A.31: Apigenin substrate kinetics with WsGT enzyme

The relative conversion rates of variety of flavonoids substrates catalyzed by WsGT were determined. Of the substrates analyzed, diadzein was identified as the best substrate, followed by naringenin, genistein, luteolin and apigenin (Table 4.B.3). In addition, the K_m and V_{max} values of WsGT for diadzein, naringenin, genistein, luteolin and apigenin were determined using Michaelis-Menten plots. WsGT had a lower K_m value and higher V_{max} for diadzein than other substrates. According to K_{cat}/K_m ratio that reflects the enzyme catalytic efficiency, WsGT used diadzein most efficiently.

Tabel 4.B.3 Substrate specificity of the purified WsGT enzyme

Substrate	**K_m (µM)**	**V_{max} (nKat/mg)**	**K_{cat}**	**K_{cat}/K_m ($\mu M^{-1} s^{-1}$)**
Diadzein	8.488	37.79	2.190	0.258
Naringenin	9.335	39.79	2.30	0.246
Genistein	11.61	9.89	0.296	0.044
Luteolin	12.79	9.885	0.296	0.044
Apigenin	13.01	6.961	0.4	0.03

4.8 Heterologus expression of WsSAPGT gene

4.8.1 Cloning of WsSAPGT gene in pET-30b (+) Expression vector

GT genes were cloned in pGEM-T Easy vector by incorporating appropriate restriction sites in the primers. In case of WsSAPGT gene, *Nde* I restriction site was added to the forward primer and *Xho* I restriction site was added to the reverse primer (namely WsSAPNde and WsSAPXho). High fidelity *Taq pol* (*Pfx* Invitrogen) was used to amplify the WsSAPGT gene using the above set of primers from the cDNA clone, sequenced and maintained in pGEM-T Easy vector as the template. The above recombinant plasmid DNA was diluted 1:50 times and 1 µL was used as the template and PCR was performed (Chapter 2, Section 2.8.7). Approximately 1.3 kb band was amplified (exactly 1272 bp coding region of WsSAPGT gene + *Nde* I and *Xho* I sites added into primer sequences). The band is excised from gel, purified, ligated in pGEM-T Easy vector and transformed in *E. coli* XL1 cells. Clones were screened by colony PCR and a few colonies were inoculated in 5 mL LB (Ampicilin 100 µg/mL) medium in tubes. Individual plasmid was isolated and restriction digested with respective restriction enzymes to confirm the integration of the gene.

4.8.2 PCR cycling conditions for WsSAPGT gene

Temperature	Time	Cycles
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
58 ºC	45 s	
72 ºC	1.30 min	
72 ºC	7 min	1 cycle
4 ºC	hold	1 cycle

The recombinant WsSAP*GT* clone with respective restriction sites as directionally cloned in pET30b (+) vector (Figure 4.A.32 & 4.A.33). Colony PCR (Chapter 2 Section 2.8.2.5) was done to screen the recombinant pET-30b (+) clones. Integration of WsSAPGT gene in pET30b (+) was confirmed by digestion with respective restriction enzymes.

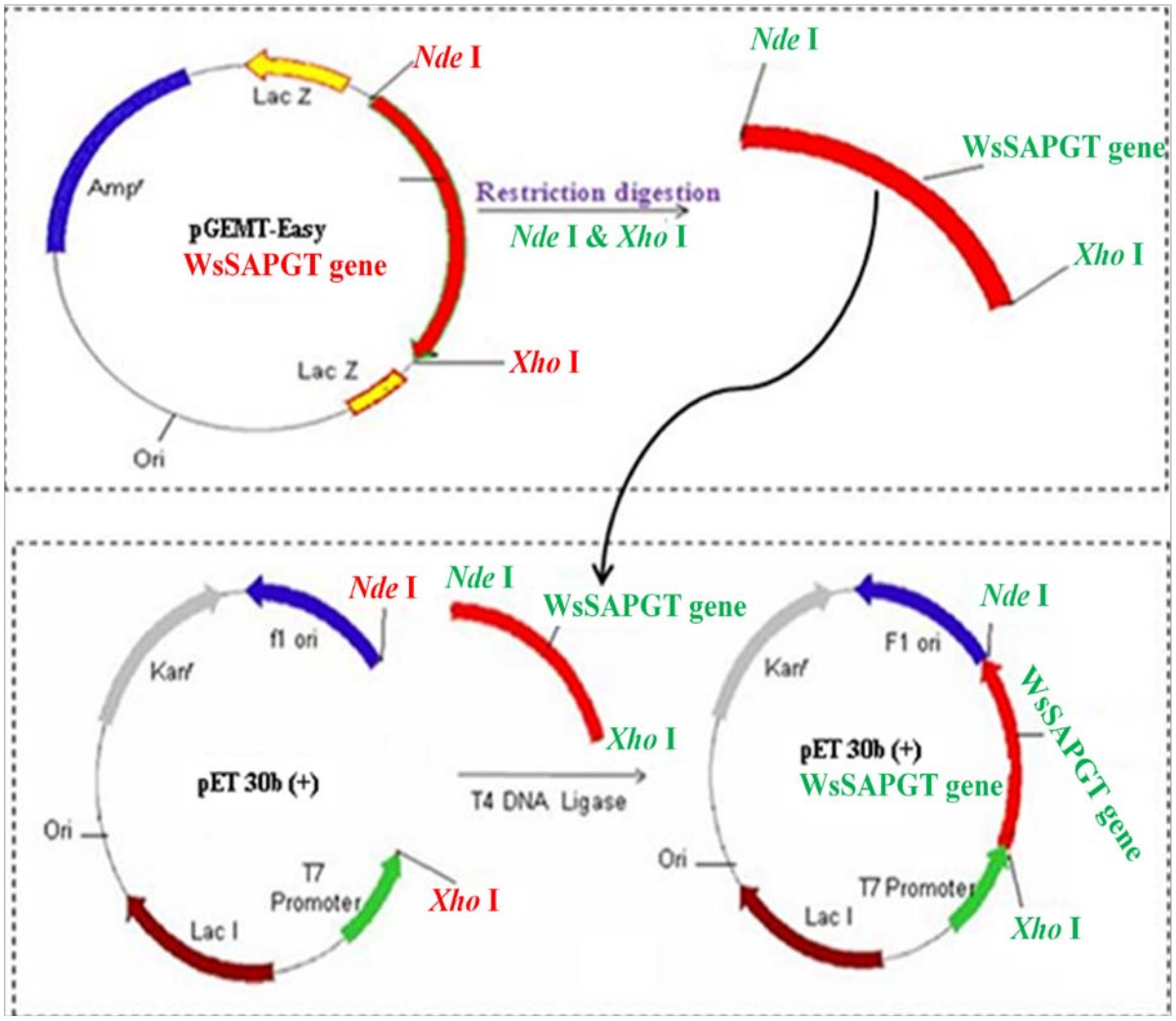


Figure 4.A.32: Strategy used for directional cloning of WsSAPGT gene in pET-30b (+) vector

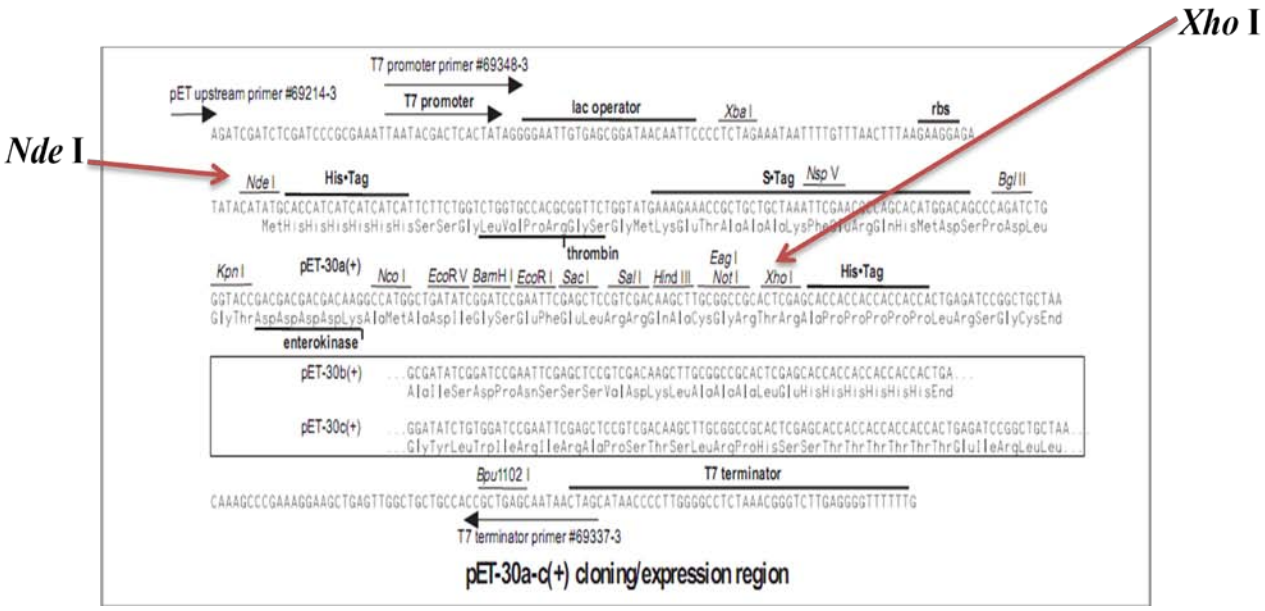


Figure 4.A.33: pET30b (+) vector map shows *Nde* I & *Xho* I sites

4.8.3 Recombinant WsSAPGT protein expression and its purification from inclusion bodies

4.8.3.1 Recombinant protein expression in *E. coli* (BL21)

A single bacterial (BL21) colony carrying recombinant pET30b (+) with WsSAPGT gene, from freshly streaked plates (grown on LB agar medium containing 50 µg/mL kanamycin) was used to inoculate 5 mL liquid LB medium containing the same concentrations of the antibiotic. Culture was grown overnight with shaking at 180 rpm at 37 °C. One mL aliquot of over-night grown culture was used to inoculate 100 mL LB broth containing 50 µg/mL kanamycin. Once the cultures reached OD_{600} 0.6 to 0.8, recombinant protein expression was induced by the addition of isopropyl β-D thiogalactopyranoside (IPTG), and the culture was grown for 4 to 6 h at 37 °C with shaking at 180 rpm. Recombinant protein extraction was done according to the protocol described in Section 2.12. Likewise, four positive recombinant pET30b (+) clones were screened for WsSAPGT protein over-expression and analysed by 10% SDS PAGE (Chapter 2, Section 2.12.3). The clone showing maximum over-expression was chosen for further studies.

4.8.3.2 Purification of recombinant protein (WsSAPGT)

His-tagged recombinant protein (WsSAPGT) was purified by metal chelate affinity chromatography. The initial stage of His-tagged protein purification is based on the remarkable selectivity and high affinity of patented Ni-NTA (nickel-nitrilotriacetic acid) resin for proteins containing an affinity tag of six consecutive histidine residues, the 6xHis-tag. NTA, which has four chelating sites for nickel ions, binds nickel more tightly than metal chelating purification systems that have only three sites available for interaction with metal ions. The extra chelation site prevents nickel ion leaching, providing a greater binding capacity and high-purity protein preparations. Purification of recombinant WsSAPGT protein carrying a 6xHis-tag was conducted using Ni NTA Agarose beads (Qiagen) (Chapter 2, Section 2.12.2). Purity of protein was checked on 10% SDS-PAGE (Chapter 2, Section 2.12.3).

4.8.3.3 Standardization of time and temperature for protein expression in soluble form

The information obtained from above section 4.8.3 was utilized for this experiment. Several flasks with 50 mL LB broth (kanamycin 50 µg/mL) were inoculated with O/N grown culture of *E. coli* BL21 harboring recombinant pET30b (+) plasmid carrying WsSAPGT gene. The cells were grown till A_{600} reached 0.6-0.8 and induction was done with 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM and 1 mM IPTG. The cultures were grown at different temperatures 15 ºC, 18 ºC, 20 ºC, 24 ºC, 28 ºC and different durations 6 h, 8 h, 12 h, 14 h, 18 h, 20 h for optimization of maximum expression of recombinant WsSAPGT proteins in soluble form. One flask as an uninduced control sample was used before inducing with IPTG. The soluble fraction of cell lysate was analysed on 10% SDS-PAGE.

4.8.3.4 Protein estimation

Protein estimation was done using Bradford assay (Bradford, 1976). A standard graph was made for BSA and concentration of unknown sample was determined by plotting standard graph (Chapter 2 section 2.14).

4.8.3.5 GT enzyme assay

The purified glycosyltransferase is used for performing enzyme assay studies. The control reaction contained crude lysate of only pET 30b (+) vector transformed in *E. coli* (BL21). Assay reaction was carried out as described in Chapter 2, Section 2.17. See Chapter 2 Table 2.10 substrates used for the study.

4.9 Results and discussions

4.9.1 Cloning of WsSAPGT gene in pET-30b (+) vector

WsSAPGT gene was cloned in pET30b (+) expression system to get the gene expressed in its active form and for its characterization.

4.9.2 Incorporation of restriction sites

The WsSAPGT gene cloned in pGEM-T Easy vector was amplified using gene specific primers **WsSAPNde Forward** and **WsSAPXho Reverse** to incorporate the *Nde* I site at the 5' end and *Xho* I site at 3' end of WsSAPGT gene.

WsSAPNde Forward primer- 5' CATATGCCGTCCTCTGAAGTTG 3'

WsSAPXho Reverse primer- 5'CTCGAGACCACTTTCTTTTTCG 3'

Approximately 1.2 kb (1272 bp WsSAPGT + *Nde* I and *Xho* I restriction sites) fragment was amplified. The fragment was excised from the gel, purified, ligated in pGEM-T Easy vector and transformed in *E. coli* XL1 cells. Clones with WsSAPGT gene with *Nde* I and *Xho* I restriction sites were screened by inoculating a few colonies in 5 mL LB media (ampicilin 100 μg/mL) tubes. Individual plasmids were isolated and restriction digested with *Nde* I and *Xho* I enzymes to confirm the integration of WsSAPGT gene.

4.9.3 Directional cloning of WsSAPGT gene in pET30b (+) vector

The above clone of WsSAPGT gene in pGEM-T Easy vector was restriction digested with *Nde* I and *Xho* I restriction enzymes and gene of interest was purified. pET30b (+) vector DNA was also digested with same restriction enzymes and purified. WsSAPGT gene was directionally cloned in purified restriction digested pET30b (+) vector. Ligation mixture was transformed into *E. coli* XL1 competent cells and plated on LB-agar plate (kanamycin 50 µg/mL). *E. coli* colonies having recombinant plasmids were screened by colony PCR (Figure 4.A.34), with reaction cycles as shown in section 4.9.4. Plasmids were isolated from PCR positive clones and were digested with *Nde* I and *Xho* I to confirm the integration of WsSAPGT gene fragment in pET30b (+) vector (Figure 4.A.35).

4.9.4 PCR cycling conditions for WsSAPGT gene

Temperature	Time	Cycles
95 ºC	5 min	1 cycle
95 ºC	1 min	35 cycles
58 ºC	0.45 min	
72 ºC	1.30 min	
72 ºC	5 min	1 cycle

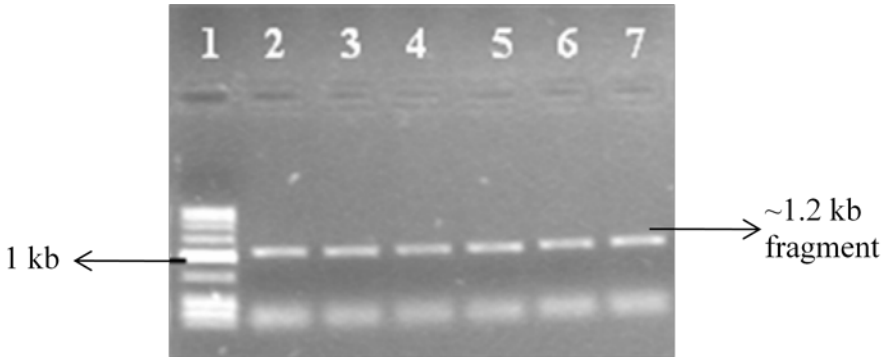


Figure 4.A.34: Colony PCR showing ~1.2 kb recombinant clones of WsSAPGT gene. Lane 1- marker, Lane 2 to 7 are positive clones.

Figure 4.A.35: Restriction analysis of recombinant pET30b (+). Lane 1 recombinant vector releasing ~1.2 kb WsSAPGT gene and 5.4 kb vector backbone, Lane M- marker.

Integration of the WsSAPGT gene along with restriction sites was further confirmed by sequencing. The sequence was translated using proteomic tools available at www.expasy.ch and was checked for in frame translation up to HIS tag.

4.9.5 Recombinant WsSAPGT protein expression and purification from inclusion bodies

E. coli BL21 (DE3) cells transformed with recombinant pET30b (+) plasmids having WsSAPGT gene were screened for over-expression of protein. A few positive recombinant clones were screened for recombinant WsSAPGT protein over-expression. An approximately 44 kDa protein was expressed in all clones analysed on 10% SDS-PAGE with varying levels of expression. Among those recombinant clones one of the clone showing maximum expression was chosen and large scale protein purification was done using Ni-NTA Agarose beads (Figure 4.A.36).

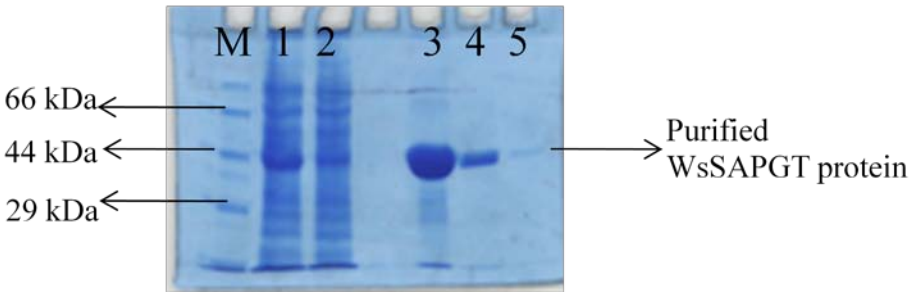


Figure 4.A.36: 10% SDS PAGE commassie blue staining. Lane-1 Inclusion bodies from WsSAPGT protein, Lane-2 lysate protein, Lane-3, 4 & 5 Ni-NTA purified WsSAPGT Protein, Lane-M Protein molecular weight marker.

4.9.6 Extraction of recombinant WsSAPGT protein in soluble form

WsSAPGT clone which showed maximum expression as described above in section 4.9.5 was used for extraction of recombinant protein in soluble form to test the enzyme activity. Temperature and time required after induction was standardized for maximum WsSAPGT protein expression in soluble form that is in cell lysate. The culture was grown at different temperatures ranging from 15 °C, 18 °C, 20 °C, 24 °C and 28 °C and at different time durations 6 h, 8 h, 12 h, 14 h, 18 h and 20 h for optimization of maximum expression of recombinant WsSAPGT protein in soluble form. IPTG concentration was also checked

ranging from 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM and 1 mM. The optimum parameters were as follows; after initial growth at 37 ^{o}C till A_{600} reached 0.6 to 0.8 cells were induced with 0.5 mM IPTG and grown 14 hours at 18 °C as shown in Figure 4.A.37.

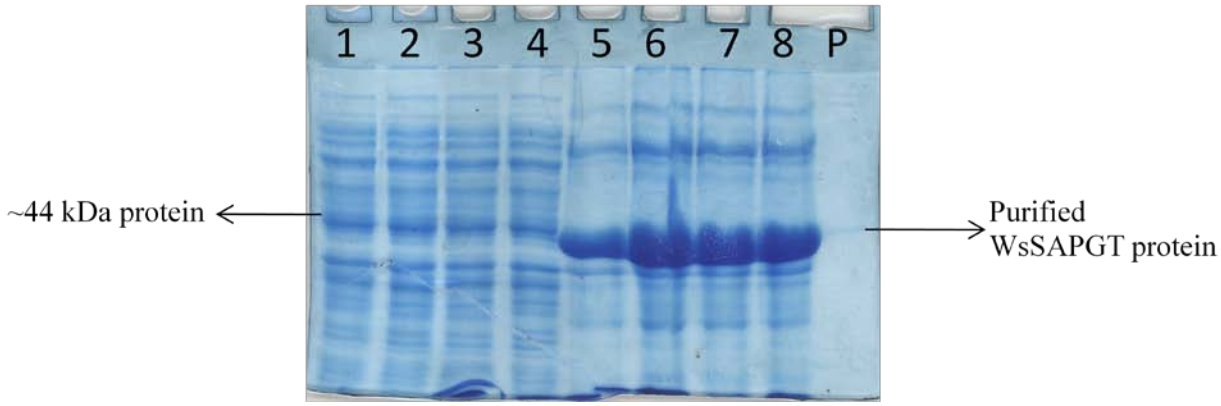


Figure 4.A.37: 10% SDS PAGE analysis of WsSAPGT protein in soluble form. Lane P- Purified WsSAPGT protein, Lane 1, 2, 3, 4 - Lysate of WsSAPGT protein grown at 8 h, 12 h, 14 h, 18 h respectively, Lane 5, 6, 7 and 8- Inclusion bodies protein of WsSAPGT protein grown at 8 h, 12 h, 14 h and 18 h respectively.

4.9.7 Purification of recombinant WsSAPGT soluble protein using Ni-NTA Agarose

The WsSAPGT clone selected for protein expression inoculated in 5mL LB broth and grown at 37 ºC for overnight. This culture was used as a primary culture and inoculated in 500 mL of LB broth as a secondary culture. After inoculation culture was grown at 37 ºC till the $O.D_{600}$ reaches 0.6 to 0.8. After induction with 0.5 mM IPTG culture was grown at 18 ºC for 14 hours. Protein isolation was done from this culture according to protocol mention in chapter 2 section 2.12. After protein isolation total protein sample was used for Ni-NTA purification. The details about affinity Ni-NTA purification given in Chapter 2 section 2.12.2.

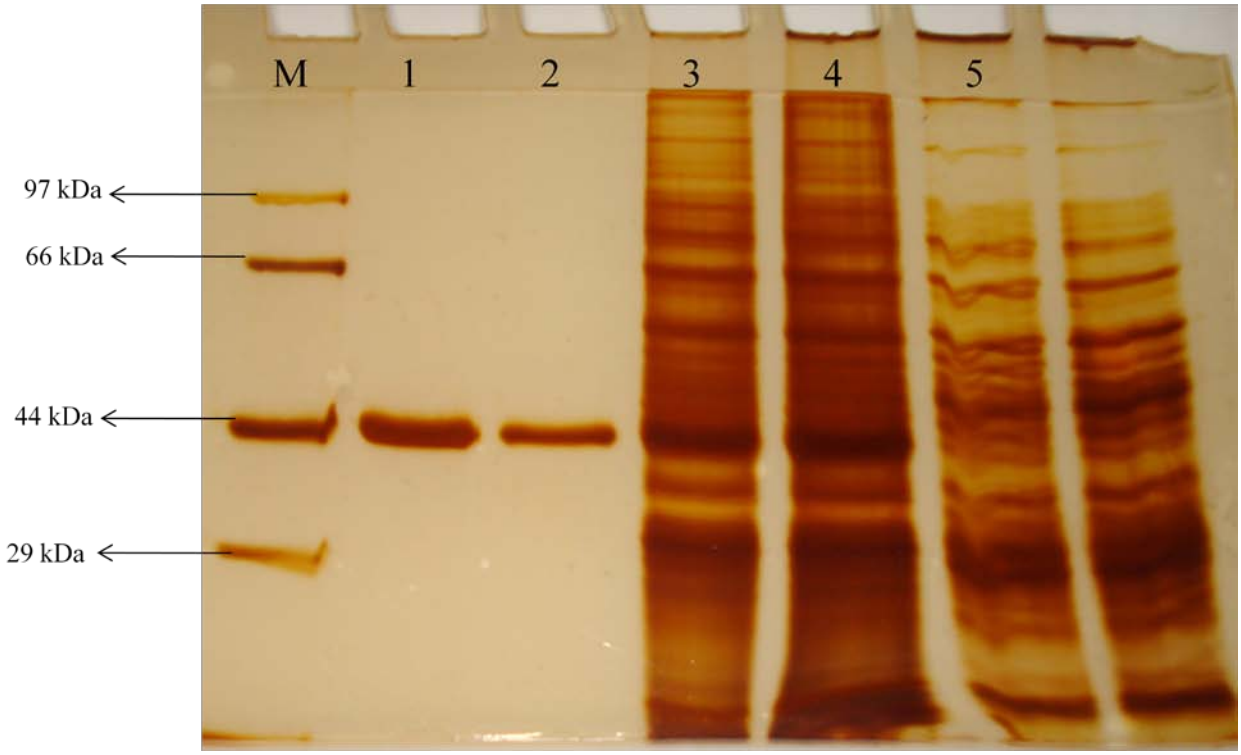


Figure 4.A.38: 10% SDS-PAGE gel of WsSAPGT affinity purification. Lane M-Protein molecular weight marker; Lane 1 & 2 affinity purified ~44 kDa WsSAPGT protein; Lane 3 & 4 lysate protein in soluble form; Lane 5 uninduced protein.

4.10 Glycosyltransferase enzyme assay

4.10.1 Optimization of pH and temperature

The reaction conditions for enzyme assay i.e. pH, temperature, substrate concentration and UDP-glucose was optimized using different parameters. To optimize the pH different pH i.e. 4.0 to 12.0 range was used in assay conditions. The enzyme shows maximum activity at pH 7.0. The theoretical pI value of WsSAPGT is 6.1. The assays were incubated at different temperatures from 25 ºC to 50 ºC. Enzyme shows maximum activity at 30 ºC. Substrate concentration varied from 10 µM to 1 mM and 100 µM to 400 µM concentration suitable for assay conditions. Finally different sugar donors i.e. UDP-Glucose, UDP-Galactose and UDP-Xylose were used in assay reaction to find out which sugar donor is suitable for enzyme

assay. UDP-Glucose is the suitable sugar donor for WsSAPGT enzyme assay. The UDP-Glucose concentration used for assay reaction is 5 mM.

4.10.2 WsSAPGT enzyme assay

The affinity purified protein was stored at +4 ºC and used for further enzyme assay reactions. Protein estimation was done using Brad ford assay method correlating with standard data. The activity of WsSAPGT protein was checked by glycosylation reaction of various aglycone moieties by sugar donor molecules. The assay reaction (0.5 mL) contained 20 mM Tris-HCl buffer pH 7.0, 5 mM sugar donor (UDP-glucose), 200 µM glycosyl acceptor and 30µg of purified WsSAPGT enzyme. The reaction was incubated at 30 ^{o}C for 3 hours. After incubation the reaction was extracted thrice by the addition of equal volume of ethyl acetate. The extracted reaction was then completely vacuum dried and dissolved in appropriate volume of methanol. The substrates and the glycosylated products were then analysed by MALDI-MS. Same reaction was run for control (i.e. lysate of pET30b (+) vector transformed in *E. coli* BL21) with all the substrates, in order to check the background activity if any.

4.10.3 MALDI-MS analysis of WsSAPGT enzyme products

MALDI-MS was done as described in Chapter 2, Section 2.16. Different substrates (Table 2.10) were used in order to check the specificity of the WsSAPGT enzyme. Four different steroidal compounds diosgenin, sarsasapogenin, tomatidine and digitoxigenin were checked using MALDI-MS (ABI voyager-DE STR Biospectrometry). Diosgenin, digitoxigenin and sarsasapogenin are known as Steroidal sapogenins. Tomatidine is known as Steroidal alkolids. But WsSAPGT enzyme showed expected mass only with steroidal sapogenin digitoxigenin. All the substrates are cross checked with control reactions (i.e. only pET30b (+) lysate transformed in BL21 (DE3) cells). The MALDI-MS profile is given below (Figure 4.A.39).

Before analysis of enzyme products substrates are analyzed on different matrices to confirm which matrix is suitable for analysis. Three different matrices used for analysis and each matrix has its own preparation in trifluoroacetic acid and acetonitrile.

10 mg of sinapic acid was preparated in 70% TFA and 30% acetonitrile. 1 µL of TFA and 699 µL of sterile milliQ water mix become 1% of 70% TFA. And 300 µL of acetonitrile in total 1 mL of volume.

The second matrix was α-cyano-4-hydroxycinnamic acid, 10 mg of CHCA dissolved in 1% of 70% TFA and 500 µl of acetronitrile in total volume of 1 mL.

The third matrix used for analysis is 2, 5 Dihydroxy benzoic acid. 10 mg of benzoic acid in 1% of 70% TFA and 30% of acetonitrile in total volume 1 mL.

Among the three above mentioned matrices, sinapic acid is suitable matrix, which shows expected mass of substrates used in analysis. Therefore WsSAPGT enzyme products were analyzed using sinapic acid as a matrix.

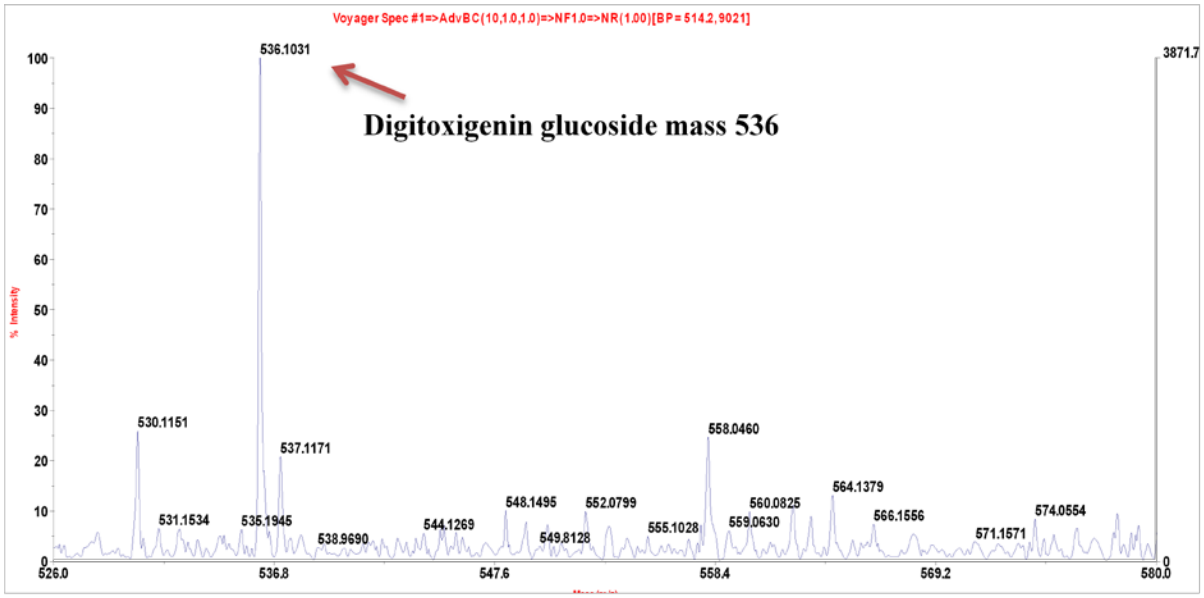


Figure 4.A.39: Digitoxigenin glucoside mass spectrometry chromatogram

4.11 Conclusion

- The WsGT and WSSAPGT genes isolated from *W. somnifera* were directionally cloned in pET30b (+) expression vector.
- Recombinant WsGT and WsSAPGT proteins were standardized for over-expression and purified from inclusion bodies.
- Only WsGT protein from inclusion bodies was used to raise polyclonal antibodies and used in further western hybridization.
- Bothe the proteins WsGT and WsSAPGT were optimized to express in soluble form for further enzyme assays and kinetics.
- WsGT protein was expressed and purified using FPLC AKTA explorer system. Nickel sepharose and anion exchange chromatography was used to purify the enzyme.
- To conform the mass and pI value MALDI-MS and 2D gel electrophoresis was done.
- Purified WsGT enzyme was used in enzyme assay with substrates like genistein, apigenin, naringenin, diadzein and luteolin. Sustrates and enzyme products were analyzed by LC-MS and HPLC.
- WsGT enzyme glycosylates the flavonoid substrates (diadzein, apigenin, naringenin, genistein, luteolin) at 7^{th} position of their structure and enzyme kinetics shows maximum substrate specificity for diadzein.
- WsSAPGT enzyme was purified with affinity chromatography. Purified enzyme was used for enzyme assays.
- Steroidal sapogenins and alkolids substrates were used in WsSAPGT enzyme assay. Only steroidal sapogenin digitoxigenin shows expected mass in MALDI-MS analysis.

Chapter 4B

Homology modeling and Docking study of WsGT protein from W. somnifera

4.12 Introduction

Glycosyltransferases (GTs) are the enzymes that synthesize oligosaccharides, polysaccharides and glycoconjugates. GTs has been grouped into 91 families on the basis of sequence similarities. Despite the fact that many GTs recognize similar donor or acceptor substrates, there is surprisingly limited sequence identity between different families. Until now only two folds have been observed for GTs: fold GT-A, consisting of one α/β/α sandwich domain and characterized by the presence of divalent cation in the binding site. GT-B fold consist of two such domains (Aram Chang, 2011).

Despite low sequence conservation, the UGTs show highly conserved secondary and tertiary structures. The sugar acceptor and sugar donor substrates of UGTs are accommodated in the cleft formed between the N- and C-terminal domains. Several regions of the primary sequence contribute to the formation of the substrate binding pocket including structurally conserved domains as well as loop regions differing both with respect to their amino acid sequence and sequence length.

Molecular modeling of GTs presents peculiar difficulties. Compared with the hundreds of different enzymes that participate in glycoconjugate synthesis, very few have been crystallized; only 20 of the 82 carbohydrate-active enzymes (CAZY) families contain at least one GT that has been crystallized. The low degree of sequence similarity within some of the CAZY families represents an insurmountable barrier for classical sequence alignment procedure, which is a prerequisite in a homology-building procedure (Maria L *et al*., 2004).

Docking of substrates also appears to be a difficult task owing to the flexibility of the nucleotide sugar and the presence of phosphate and divalent cations, for which energy parameters are not always available in modeling software although efforts have recently been made.

Molecular modeling has been recently used in the field of GTs and the models have been useful not only for rationalizing experimental data but also for designing directed mutagenesis experiments. Fold recognition coupled with multivariate analysis has been applied to a large number of sequences in order to identify putative GTs. Homology modeling, together with

docking of substrates, helps in understanding the molecular basis of specificity for human blood group A and B transferases and for plant glycosyltransferases. Although the time scale of loop opening and closing is too large to be modeled, preliminary molecular dynamics studies allow for identifying the key amino acids involved in the conformational changes.

4.13 Materials and methods

4.13.1 Secondary structure prediction of WsGT

The secondary structure of WsGT was predicted by the software PSIPRED v 3.0.

4.13.2 Homology Modeling of WsGT

The 3D models of WsGT from *W. somnifera* were built by homology modeling based on high-resolution crystal structures of homologous proteins. A basic local alignment search tool (BLAST), search for the sequence similarities with several members of the GT family was used for selecting the 3D models of the closest homologues available in the Brookhaven Protein Data Bank (PDB). The 3D model of WsGT was generated by the automated homology modelling software MODELLER 9v9 (http://salilab.org) on windows operating environment. The steps involved in model development were illustrated in figure 4.B.1. This program is used for comparative protein structure modeling that optimally satisfies spatial restraints which includes (i) homology-derived restraints on the distances and dihedral angles in the target sequence extracted from its alignment with the template structures (ii) stereochemical restraints such as bond length and bond angle preferences, obtained from the CHAR MM-22 molecular mechanics force field (iii) statistical preferences for dihedral angles and non-bonded interatomic distances, obtained from a representative set of known protein structures and (iv) optional manually curated restraints, such as those from NMR spectroscopy, rules of secondary structure packing, cross-linking experiments, fluorescence spectroscopy, image reconstruction from electron microscopy, site-directed mutagenesis and intuition. The spatial restraints are expressed as probability density functions (pdfs) for the

features restrained. The pdfs restrain C^{α}-C^{α} distances, main-chain N-O distances, main-chain and side-chain dihedral angles. The 3D model of the protein was obtained by optimization of the molecular pdf such that the model violates the input restraints as little as possible. The molecular pdf was derived as a combination of pdfs restraining individual spatial features of the whole molecule. The optimization procedure is a variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. This model building procedure is similar to structure determined by NMR spectroscopy. The modelled structures were visualized using program PYMOL.

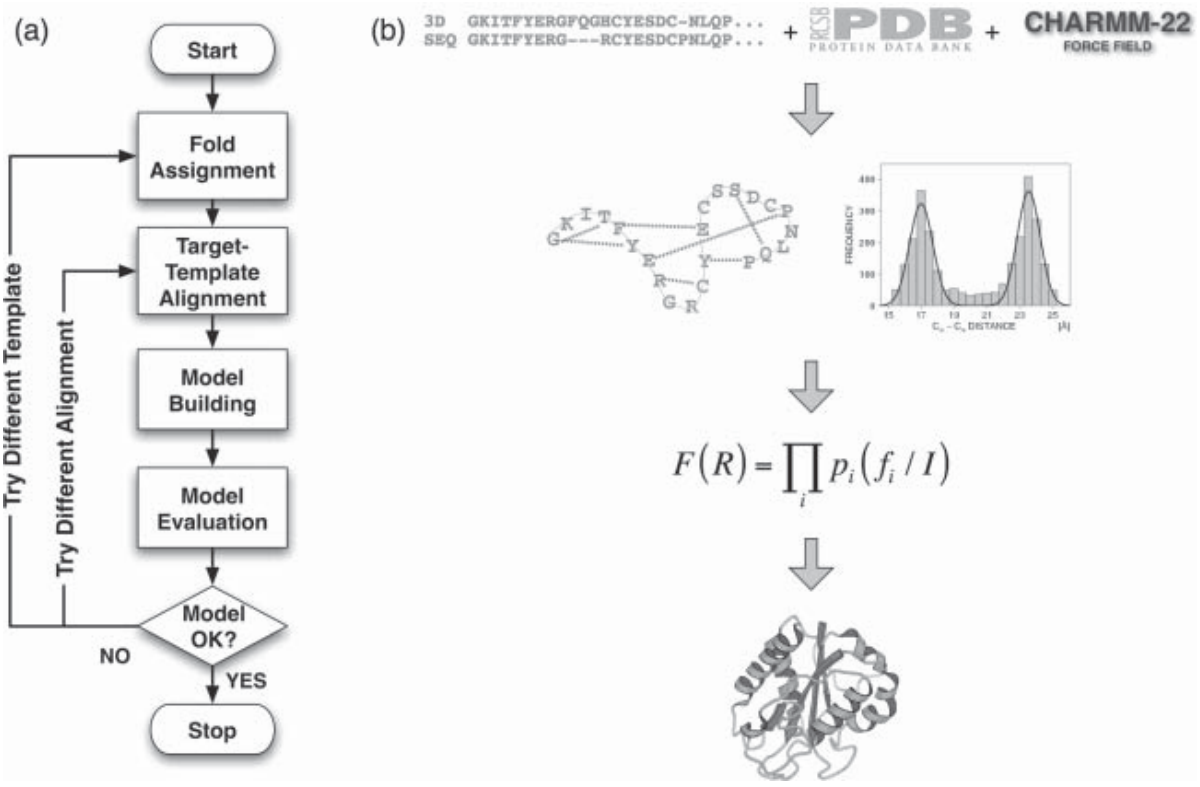


Figure 4.B.1 Protein structure Modeling

4.13.3 Molecular dynamics by FG-MD

FG-MD is a molecular dynamics (MD) based algorithm for atomic-level protein structure refinement. Given an initial protein structure, FG-MD first identifies analogous fragments from the PDB by the structural alignment program TM-align. Spatial restraints extracted from

the fragments are then used to to re-shape the funnel of the MD energy landscape and guide the MD conformational sampling. FG-MD aims to refine the initial models closer to the native structure. It can also improve the local geometry of the structures by removing the steric clashes and improving the torsion angle and the hydrogen-binding networks.

4.13.4 Structure validation of WsGT Model by PROCHECK, ERRAT and DOPE score

The PROCHECK program provides the information about the stereo chemical quality of a given protein structure by verifying the parameters like Ramachandran plot quality, peptide bond planarity, Bad nonbonded interactions, main chain hydrogen bond energy, Calpha chirality and over-all G factor and the side chain parameters like standard deviations of chi1 gauche minus, trans and plus, pooled standard deviations of chi1 with respect to refined structures.

The quality of structure was also further accessed by using ERRAT. ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement. Along with ERRAT data the DOPE (discrete optimized potential energy) score residues were computed. DOPE score is calculated by Modeller program which indicates the distance dependent statistical potential based on probabilistic theory (Raghunath Satpathy, 2011).

4.13.5 PROSAII

The ProSAII program (Protein Structure Analysis) is an established tool which has a large user base and is frequently employed in the refinement and validation of experimental protein structures and in structure prediction and modeling. This program compares Z scores between target and template structure. The Z score of model is a measure of compatibility between its sequence and structure. The model Z score should be comparable to the Z scores obtained from the template (Karim Kherraz, 2011).

4.13.6 Structural super imposition of WsGT with template

Pair wise structural superimposition of modeled WsGT was done with templates using Chimera match maker program.

4.13.7 Docking of Nucleotide Sugars and Acceptor Substrates

The docking analysis of WsGT with structural analogs of different acceptors and UDP-glucose was carried by Autodock vina docking software. Docking is the process of fitting together of two molecules in 3-dimensional space. Docking allows virtually screening a database of compounds and predicting the strongest binders based on various scoring functions. It explores ways in which two molecules, such as acceptors and an enzyme receptor fit together and dock to each other well, like pieces of a three-dimensional jigsaw puzzle.

Molecular docking was performed using Autodock vina software (http://vina.scripps.edu/). Structure Data File (SDF) of ligand molecule was downloaded from the pubchem (http://pubchem.ncbi.nlm.nih.gov). SDF format was converted to MOL2 format using Discover studio 3.1 visualizer softwar. Hydrogens were added to receptor as well as to ligand using the built-in program Add Hydrogen in Autodock vina Software. Acceptor and donor binding sites were already known so accordingly grid was made near these regions and docking was performed.

4.14 Results and discussion

4.14.1 Secondary structure prediction

The predicted secondary structure of WsGT protein contains 15 α-helices (H) and 13 β-sheets (E). Coil structure is marked in C (Figure 4.B.2).

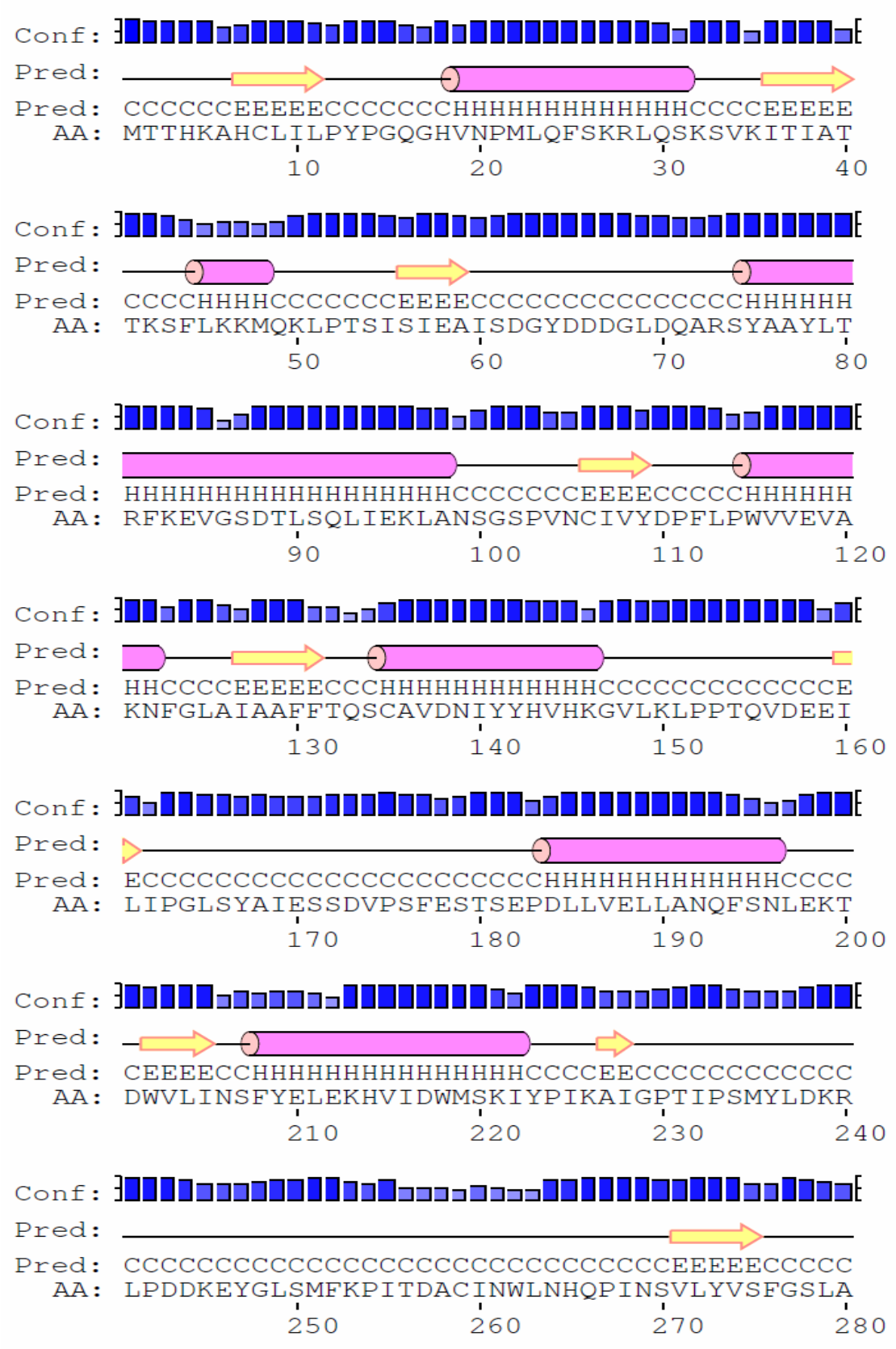

Conf:
Pred:
Pred: CCCCCCEEEEECCCCCCCHHHHHHHHHHHHHCCCCEEEEE
AA: MTTHKAHCLILPYPGQGHVNPMLQFSKRLQSKSVKITIAT
10 20 30 40
Conf:
Pred:
Pred: CCCCHHHHCCCCCCCEEEECCCCCCCCCCCCCCCHHHHHH
AA: TKSFLKKMQKLPTSISIEAISDGYDDDGLDQARSYAAYLT
50 60 70 80
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHCCCCCCCEEEECCCCCHHHHHH
AA: RFKEVGSDTLSQLIEKLANSGSPVNCIVYDPFLPWVVEVA
90 100 110 120
Conf:
Pred:
Pred: HHCCCCCEEEEECCCHHHHHHHHHHHHCCCCCCCCCCCCCE
AA: KNFGLAIAAFFTQSCAVDNIYYHVHKGVLKLPPTQVDEEI
130 140 150 160
Conf:
Pred:
Pred: ECCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCC
AA: LIPGLSYAIESSDVPSFESTSEPDLLVELLANQFSNLEKT
170 180 190 200
Conf:
Pred:
Pred: CEEEECCHHHHHHHHHHHHHHHCCCCEECCCCCCCCCCCCC
AA: DWVLINSFYELEKHVIDWMSKIYPIKAIGPTIPSMYLDKR
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEECCCCC
AA: LPDDKEYGLSMFKPITDACINWLNHQPINSVLYVSFGSLA
250 260 270 280

Figure 4.B.2 Secondary structure of WsGT protein

4.14.2 Homology Modeling of WsGT protein

In order to find out the homologous sequences in Protein Data Bank, the primary sequence of *W. somnifera* WsGT was searched against PDB using BLASTP program at NCBI server (http://www.ncbi.nlm.nih.gov/blast). Among all the homologs, glycosyltransferase from *Medicago truncatula* (PDB ID: 2pq6) was closest to WsGT, with 31% identity at the amino acid sequence level. Glycosyltransferases from *Medicago truncatula* (2acw and 3hbj) showed 30% and 27%, *Arabidopsis thaliana* (2vg8) showed 29% and *Vitis vinifera* (2clz) showed 29% identity. The three dimensional coordinates of 2pq6, 3acw, 2vg8, 2clz and 2hbj were used as templates to generate the 3D model of the WsGT using the program Modeller 9v9 (http://salilab.org) (Figure 4.B.3). All the 453 residues submitted for homology model. 15 models were generated for WsGT proten. Initially we generated model from single template (2pq6), but the dope score (-51064.63672) was higher. So we generated 30 models from multiple templates (2pq6, 3acw, 2vg8, 2clz and 2hbj). The best model generated has DOPE score was -55116.078125. Root Mean Squared Deviation (RMSD) is commonly used to represent the distance between two objects. In a structural sense, this value indicates the degree to which the three dimensional structures are similar. The lower the value, the more similar the structures are. The RMSD values between the template and our model structure were calculated using PYMOL program. The RMSD value of WsGT with templates was *i.e.* 2pq6- 1.79 Å, 2c1z - 1.89 Å, 2vg8 - 1.83 Å, 2acw - 2.13 Å and 3hbj - 2.30 Å (Table 1).

Top 5 structurally similar templates in PDB

Table 1 Top 5 structurally similar templates Tm score and RMSD values

Template	Tm Score	RMSD
2PQ6 (*Medicago truncatula)*	0.90902	1.79 Å
2C1Z (*Vitis vinifera)*	0.92189	1.89 Å
2VG8 *(Arabidopsis thaliana)*	0.88744	1.83 Å
2ACW *(M. truncatula)*	0.89778	2.13 Å
3HBJ *(M. truncatula)*	0.91289	2.30 Å

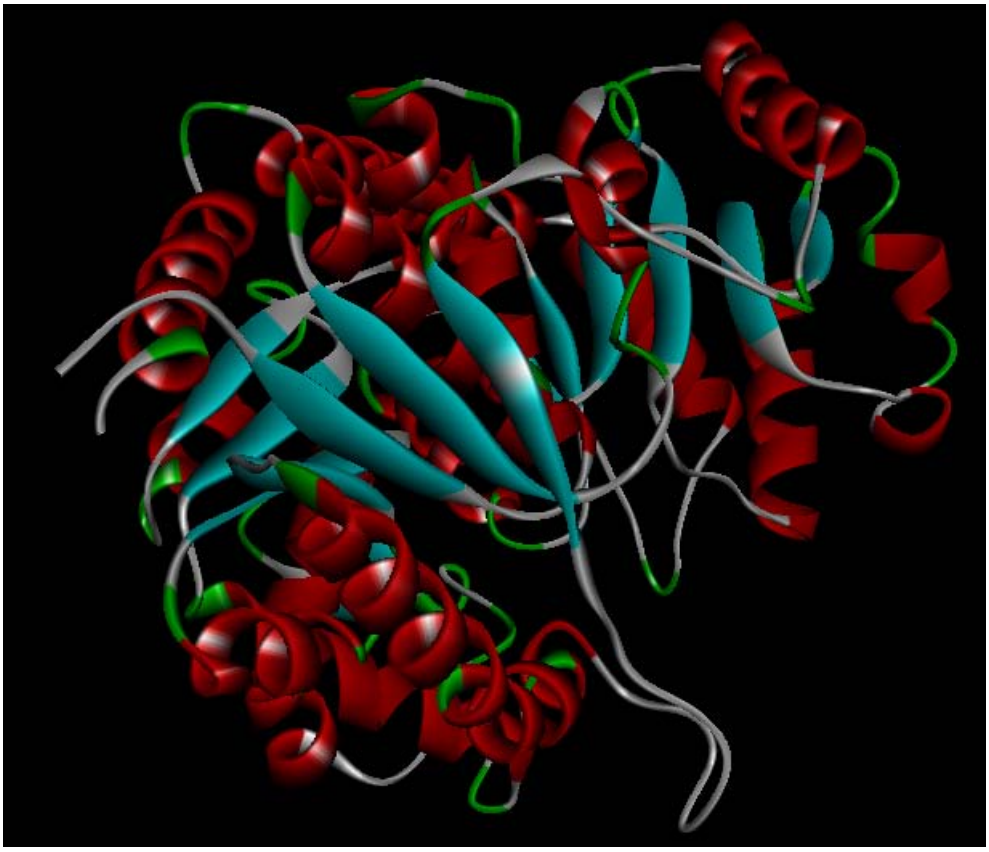


Figure 4.B.3 3D structure of WsGT enzyme. α-helices are represented in red colour and beta sheets in blue coloured ribbons

4.14.2.1 Structure validation by PROCHECK

In this study φ and ψ torsion angles were checked using the Ramachandran plot. A comparison of the results shows that one of the models generated by Modeller program is more acceptable. The best model predicted by Modeller was used for further analysis by PROCHECK (Laskowski *et al*., 1998). Ramachandran plot analysis showed that 89.2% residues were in the most favorable, 8.8% residues in the additionally allowed, 1.2% residues in the generously allowed and 0.7% residues in the disallowed regions of the Ramachandran map (Figure 4.B.4). The plot statistics are shown below.

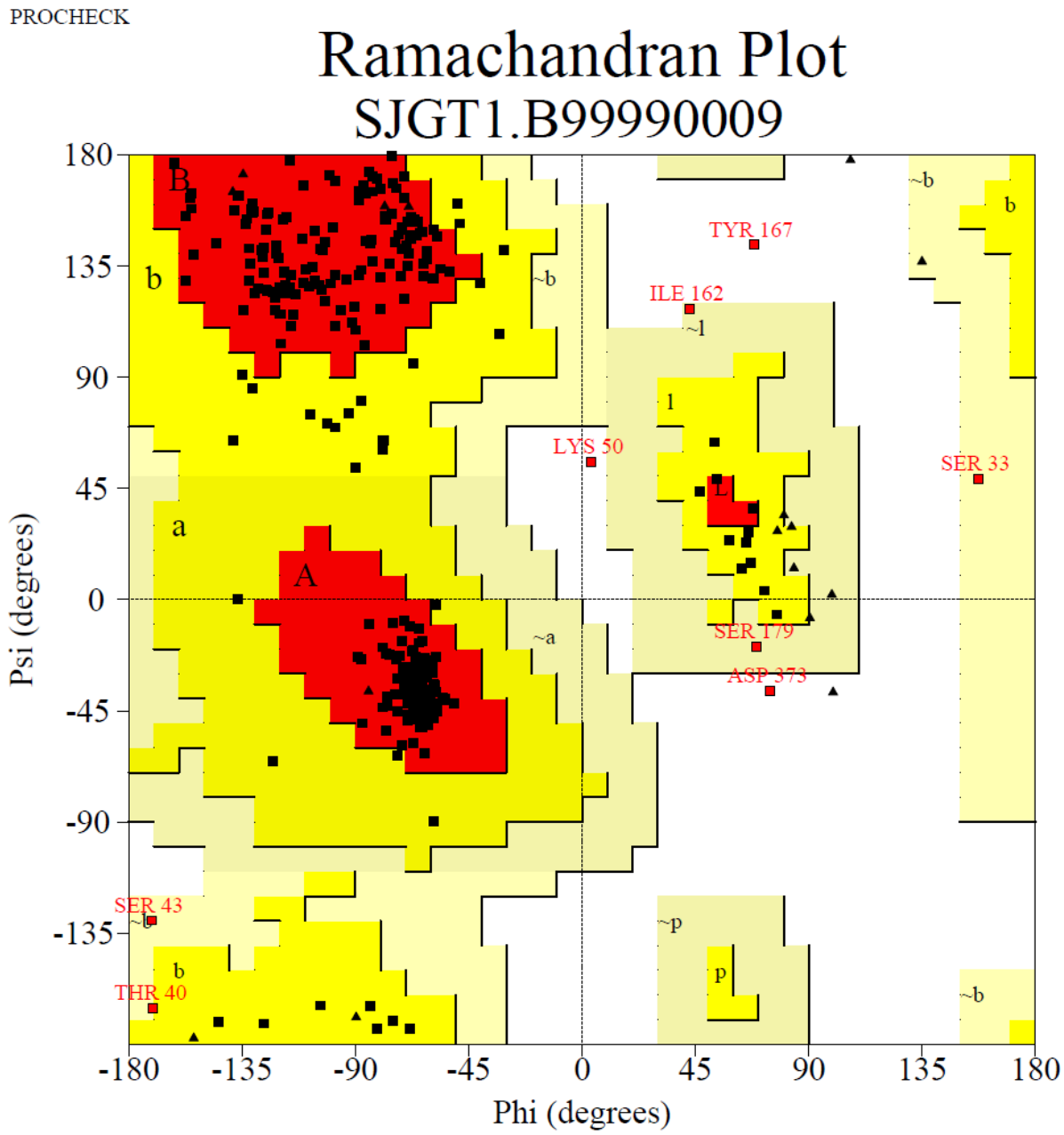


Figure 4.B.4 Ramachandran plot for the model WsGT protein

Plot statistics

Residues in most favoured regions (A,B,L)	363	89.2%
Residues in additional allowed regions (a,b,l,p)	36	8.8%
Residues in generously allowed regions (~a,~b,~l,~p)	5	1.2%
Residues in disallowed regions	3	0.7%
	---	---
Number of non-glycine and non-proline residues	407	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	22	
Number of proline residues	25	

Total number of residues	456	

4.14.2.2 Structure validation by ERRAT and DOPE score

To verity further the predicted structure of WsGT protein, the coordinates of both predicted structures were fed into the ERRAT Protein Verification Server. The overall quality factor was obtained as 72.321 which is very much satisfactory (Figure 4.B.5).

On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject rejoins that exceed that error value. Overall quality factor expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3 Å) the average overall quality factor is around 91%.

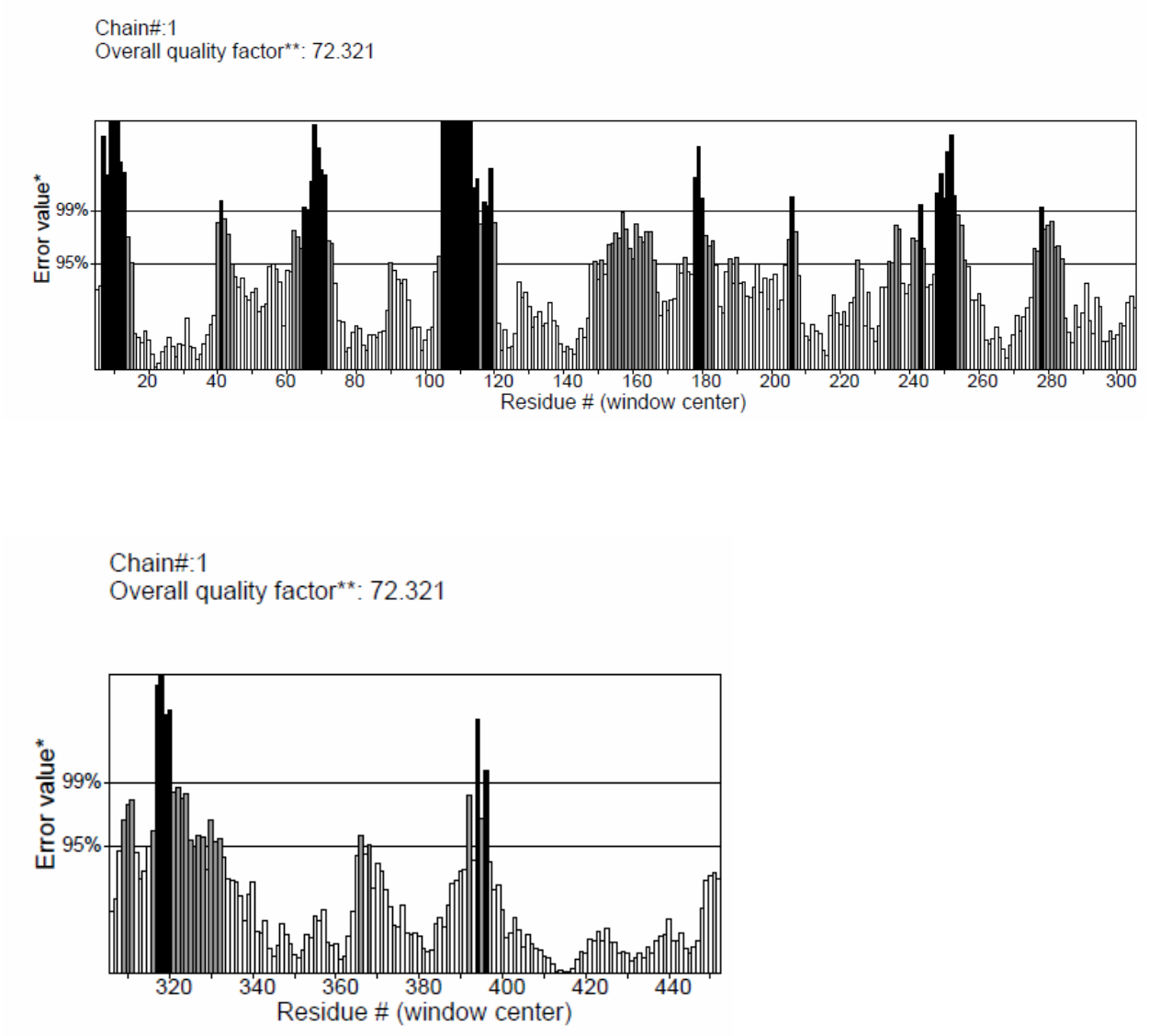


Figure 4.B.5 Showing ERRAT structural quality factor

The DOPE scores of both template and model obtained from Modeller output are shown in figure 4.B.6. The Discrete Optimized Protein Energy (DOPE) is an atomic distance dependent statistical potential based on a physical reference state that accounts for the finite size and spherical shape of proteins. The reference state assumes a protein chain consists of non-interacting atoms in a homogeneous sphere of equivalent radius to that of the corresponding protein. The DOPE potential was derived by comparing the distance statistics from a non-redundant PDB subset of 1,472 high-resolution protein structures with the distance distribution function of the reference state. By default, the DOPE score is not included in the

model building routine, and thus can be used as an independent assessment of the accuracy of the output models. The DOPE score assigns a score for a model by considering the positions of all non-hydrogen atoms, with lower scores corresponding to models that are predicted to be more accurate.

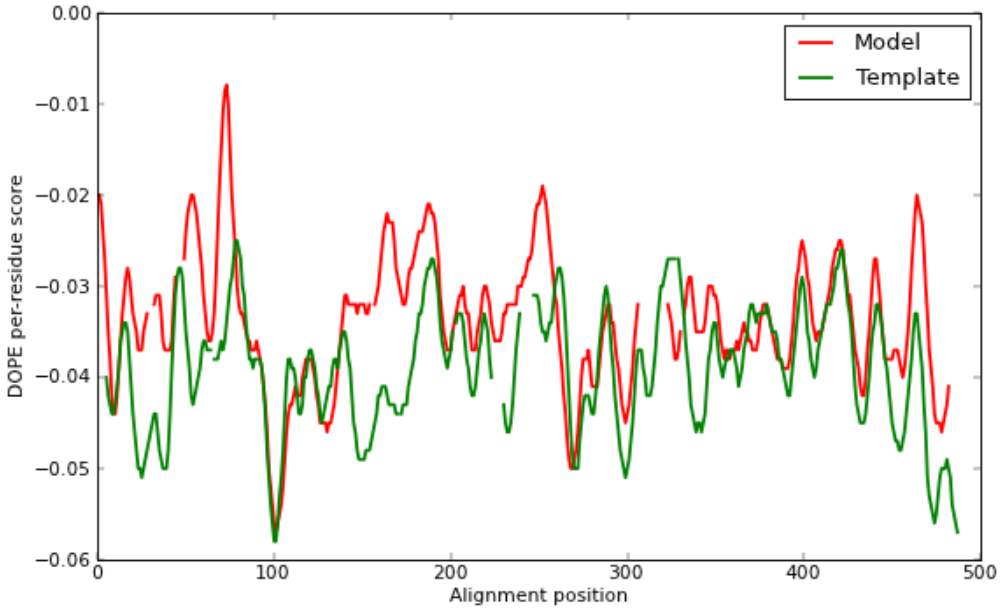


Figure 4.B.6 Comparative DOPE value of template and model

4.14.2.3 Evaluation of WsGT model accuracy by ProSA II

The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. Z-scores outside a range characteristic for native proteins indicate erroneous structures. In order to facilitate interpretation of the Z-score of the specified protein, its particular value is displayed in a plot that contains the Z-scores of all experimentally determined protein chains in current PDB. Groups of structures from different sources (X-ray, NMR) are distinguished by different colors. This plot can be used to check whether the z-score of the protein in question is within the range of scores typically found for proteins of similar size belonging to

one of these groups. The Z-score of this model is -10.28 (Figure 4.B.7). This value is extremely closed to the value of the template which suggests that the obtained model is reliable and very close to experimentally determined structures.

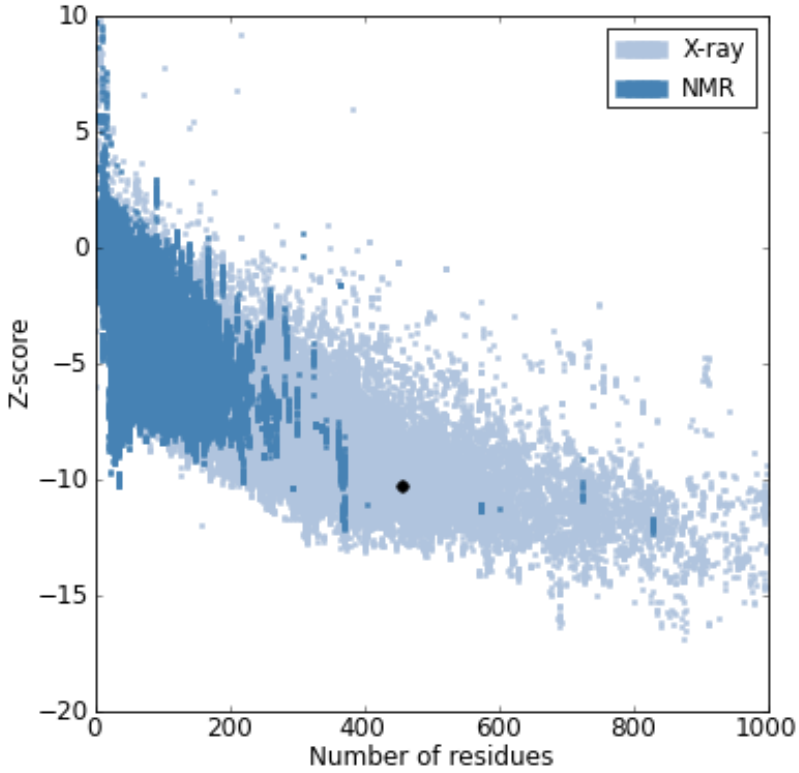


Figure 4.B.7 The plan of Z-Score shows spot of Z scores's value of protein determined by NMR (represented in dark blue colour) and by X ray (represented in light blue colour) using PROSAII program. The single black dot represents Z-Score of our model.

4.14.2.4 Energy plot of WsGT protein

The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position (Figure 4.B.8). In general, positive values correspond to problematic or erroneous parts of a model. The energy plot figure 4.B.8 shows that most of the energy scores are in negative region and proves that model quality is good.

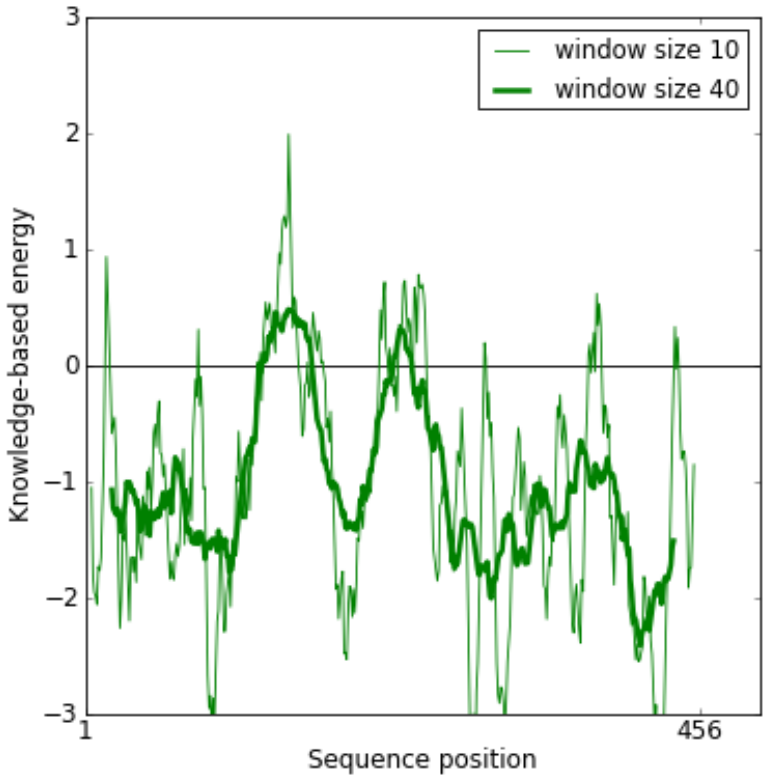


Figure 4.B.8 Energy plot of WsGT protein

4.14.2.5 Structural superimposition of WsGT with templates

Pair wise structural alignment of modeled WsGT was done with templates (2pq6, 2acw, 2clz, 2vg8 and 3hbj) using combinatorial extension algorithm at Chimera match maker. Modeled WsGT is in green colour and templates 2ACW (yellow), 2C1Z (murine blue), 2VG8 (cyan) and 3 HBJ (pale green) colour (Figure 4.B.9). The structural superimposition of WsGT with templates shows that most of the structural similarities between the templates and model.

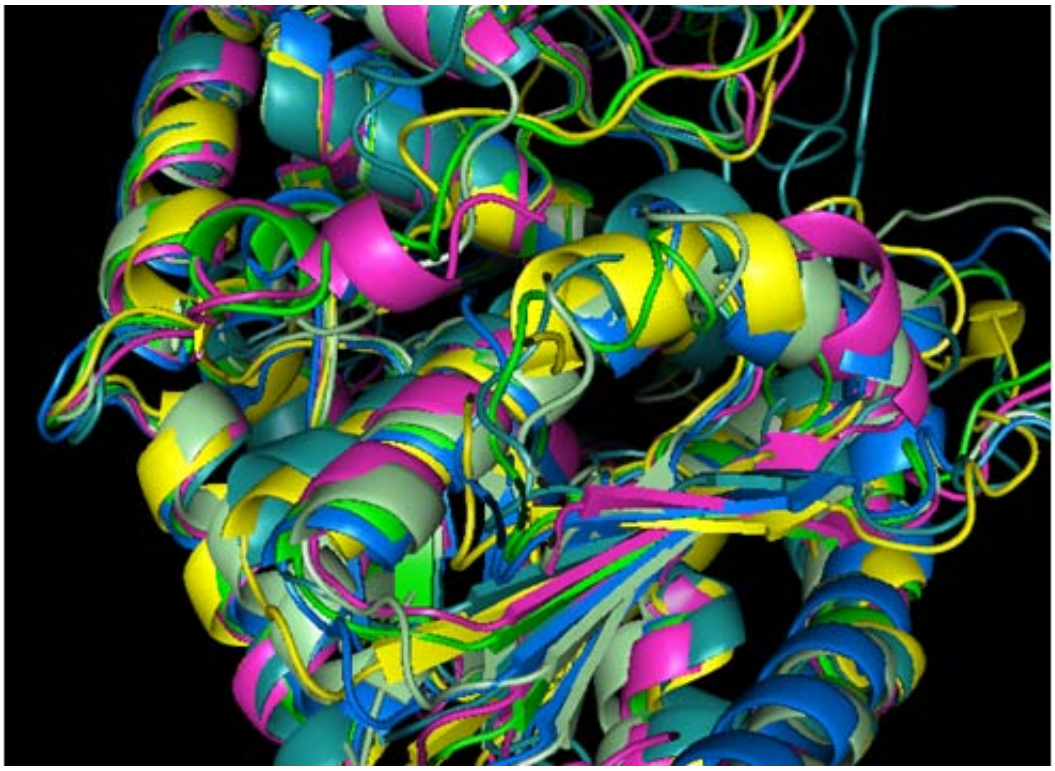


Figure 4.B.9 Structural superimposition of WsGT with templates

4.14.2.6 Main Chain Parameters

The model showed a significant resolution in various parameters with overall very low standard deviation from the standard. The Main-chain parameter results are as displayed below (Figure 4.B.10).

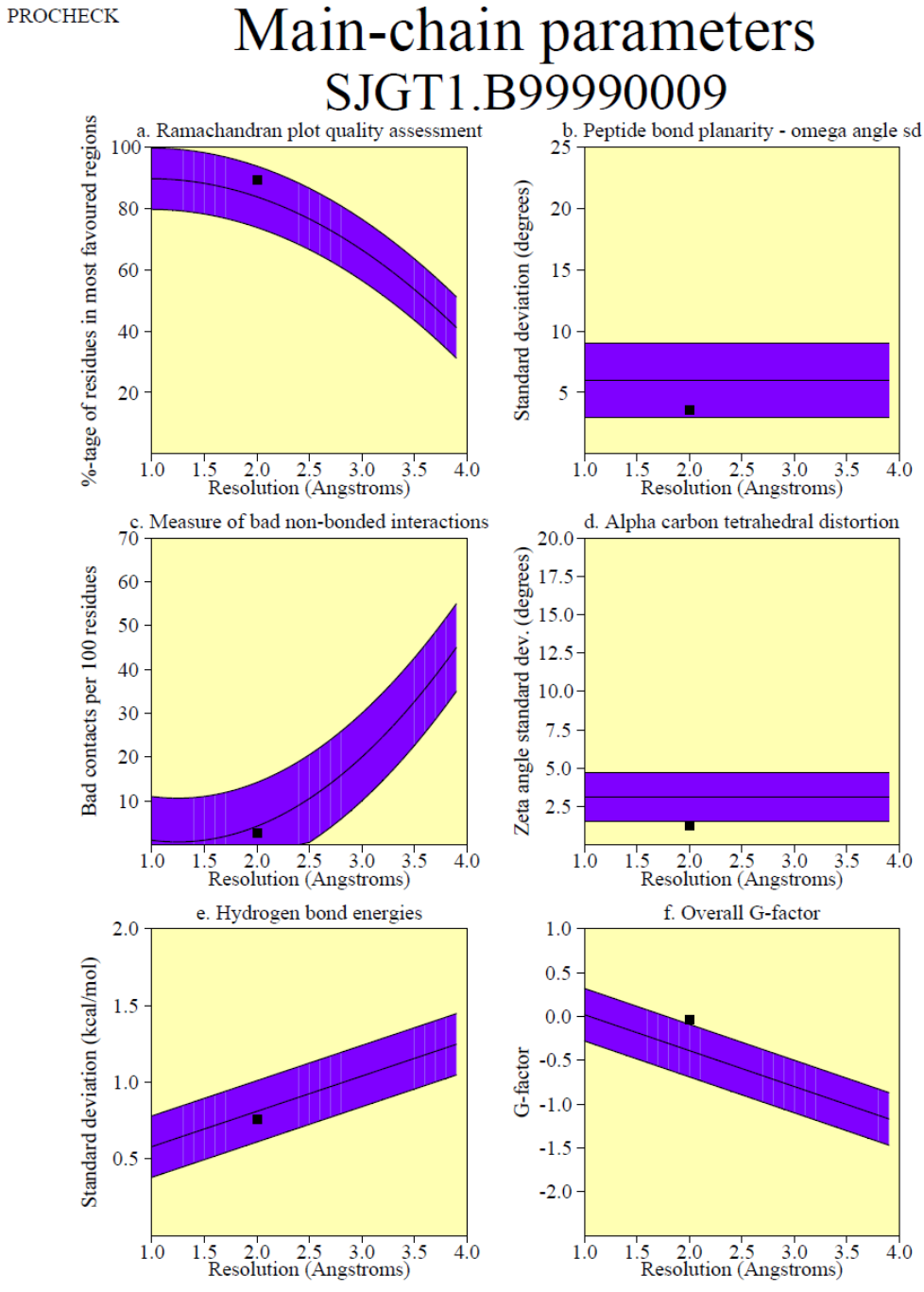


Figure 4.B.10 Main chain parameters of WsGT protein

Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Typical value	Band width	No. of band widths from mean	
%-tage residues in A, B, L	407	89.2	83.8	10.0	0.5	Inside
Omega angle st dev	453	3.5	6.0	3.0	-0.8	Inside
Bad contacts / 100 residues	12	2.6	4.2	10.0	-0.2	Inside
Zeta angle st dev	434	1.2	3.1	1.6	-1.2	Better
H-bond energy st dev	292	0.8	0.8	0.2	-0.3	Inside
Over all G-factor	456	0.0	-0.4	0.3	1.2	Better

4.15 Docking Studies

Based on the 3D model and docking investigations, the binding properties of several ligands to the proposed active site were investigated. The ligands were superimposed with the template ligand and then merged into the active site. After energy minimization the resulting protein–ligand interaction energies were calculated. WsGT docking studies were carried out with diadzein, apigenin, luteolin, naringenin, genistein and kaempferol as acceptors and UDP-glucose as sugar donor.

The docking studies were carried out with Autodock vina and molecular dynamics simulations were carried out to find an optimal docking arrangement of UDP-glucose and substrates. The so-called PSPG-box motif (Hughes and Hughes, 1994) is considered to represent the nucleotide diphosphate-sugar-binding site. In our model, this motif is found in a α/β/α-folding unit (Judith Hans, 2004). This type of fold has already been characterized for proteins that bind diphosphate-containing cofactors, such as NAD (H) (Rossmann *et al*., 1974). This fold has been observed as the glycosyl donor-binding domain of GT-B family members (Hu and Walker, 2002). Highly conserved amino acid of the PSPG-box and HCGWNS-motif within the box always considered as a potential candidate to be involved in enzymatic catalysis (Kapitonov and Yu, 1999).

4.15.1 Docking of diadzein acceptor and UDP-Glucose sugar donor

Diadzein substrate molecule and UDP-Glucose molecule was docked to the WsGT binding sites by Autodock vina. The structure of the model takes a GT-B-fold conformation comprising two distinct domains of the N- and C-terminal parts, which form a deep cleft accommodating the sugar donor and sugar acceptor substrates. The sugar acceptor is positioned interacting with the amino acid residues in the N-terminal domain while the sugar donor mainly interacts with the residues in the C-terminal region.

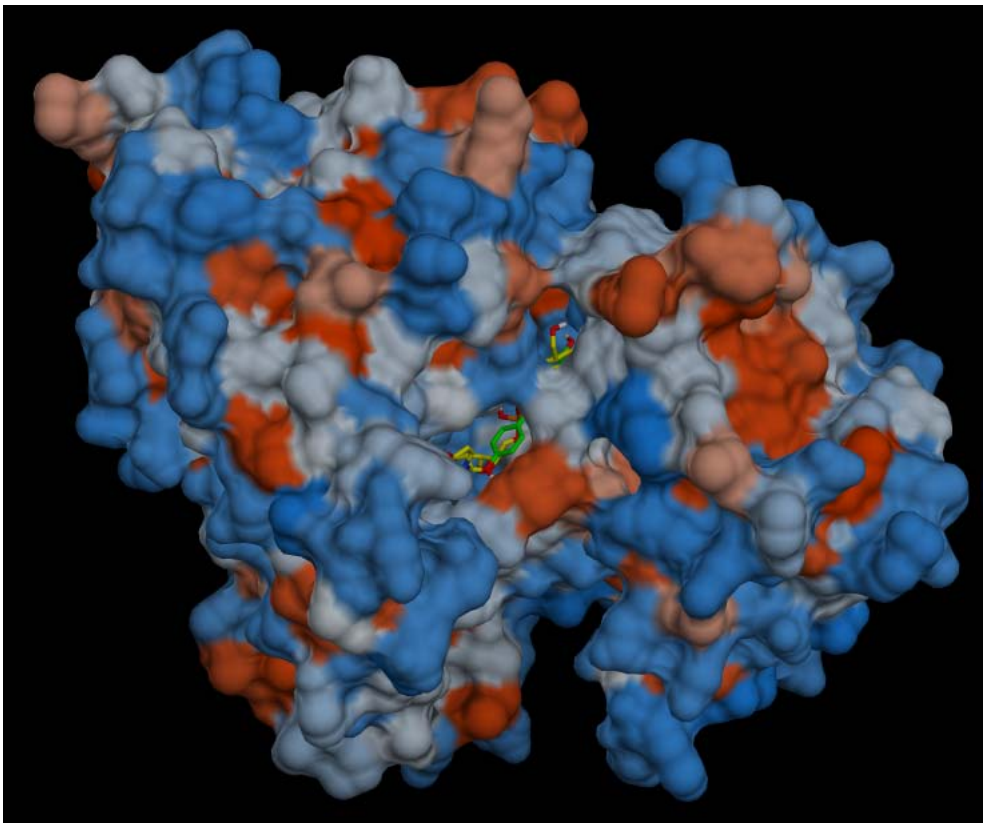


Figure 4.B.11 Binding pocket of the diadzein substrate and UDP-glucose

The interaction energy of diadzein with WsGT was -8.4 kcal/mol. Receptor ligand interactions are observed in PYMOL. In the template, active sites for substrate binding comprise Glu 357, Trp 371, Ser 278 and Ser 354. Glu 357 (OE2) interacts with 7-OH group, Trp 371 (NE1) interacts with 4`-OH group, Ser 278 (OG) interacts with 4-O group and Ser 354 (OG) interacts with 1-O group of diadzein molecule (Figure 4.B.12). All the hydrogen bond distances between WsGT and diadzein complex were observed within the range of 2.2 Å to 3.3 Å.

In the template, active sites for UDP-Glucose comprise His 18, Asn 20, Gln 133, Lys 245, Ser 278, Gln 334, His 349, Asn 353, Ser 354, Glu 357 and Trp 371. All the hydrogen bond distances between WsGT and UDP-Glucose complex were observed within the range of 2.1 Å to 3.4 Å.

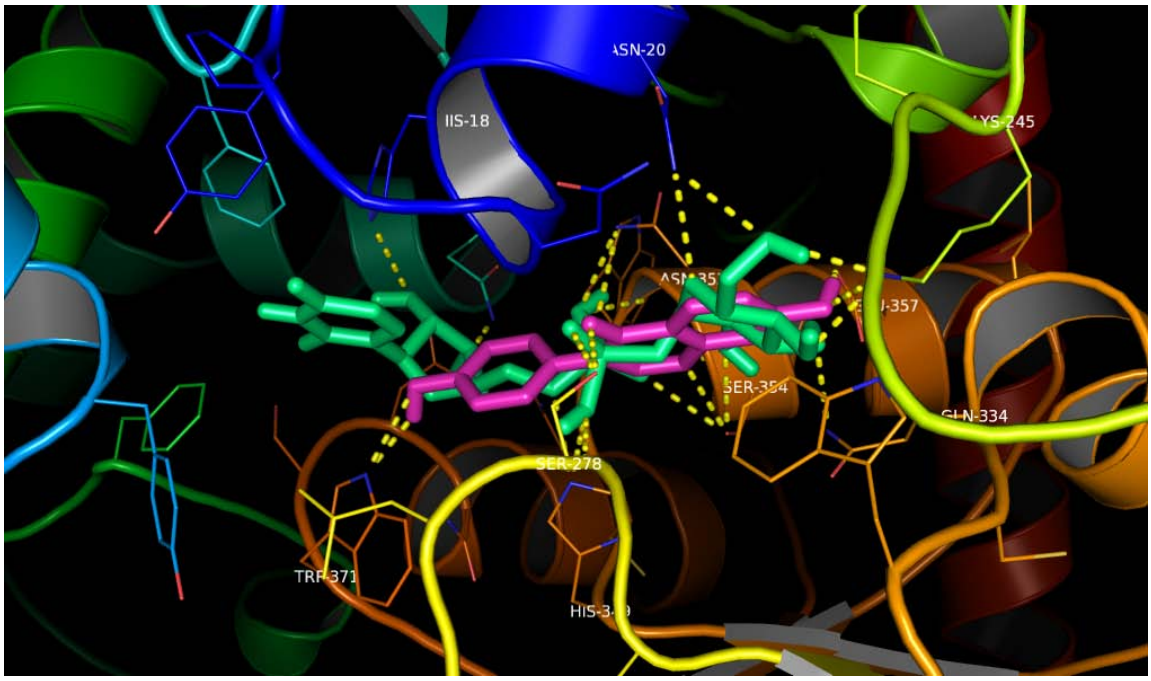


Figure 4.B.12 Interaction of catalytic residues of WsGT with substrate diadzein (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.15.2 Docking of naringenin acceptor and UDP-Glucose sugar donor

The interaction energy of naringenin with WsGT was -9.2 kcal/mol. Receptor ligand interactions are shown in PYMOL. In the template, active sites for substrate binding comprise Trp 352, Tyr 13, His 18, Asn 353 and His 349. His 18 (NE2) interacts with 1-O group, Tyr 13 (UDP-G) interacts with 4`-OH group, Trp 352 (N) interacts with 7-OH group, Asn 353 (ND2 & N) interacts with 7-OH group and His 349 (NE2) interacts with 5-OH group of naringenin molecule (Figure 4.B.14). All the hydrogen bond distances between WsGT and naringenin complex were observed within the range of 2.1 Å to 3.2 Å.

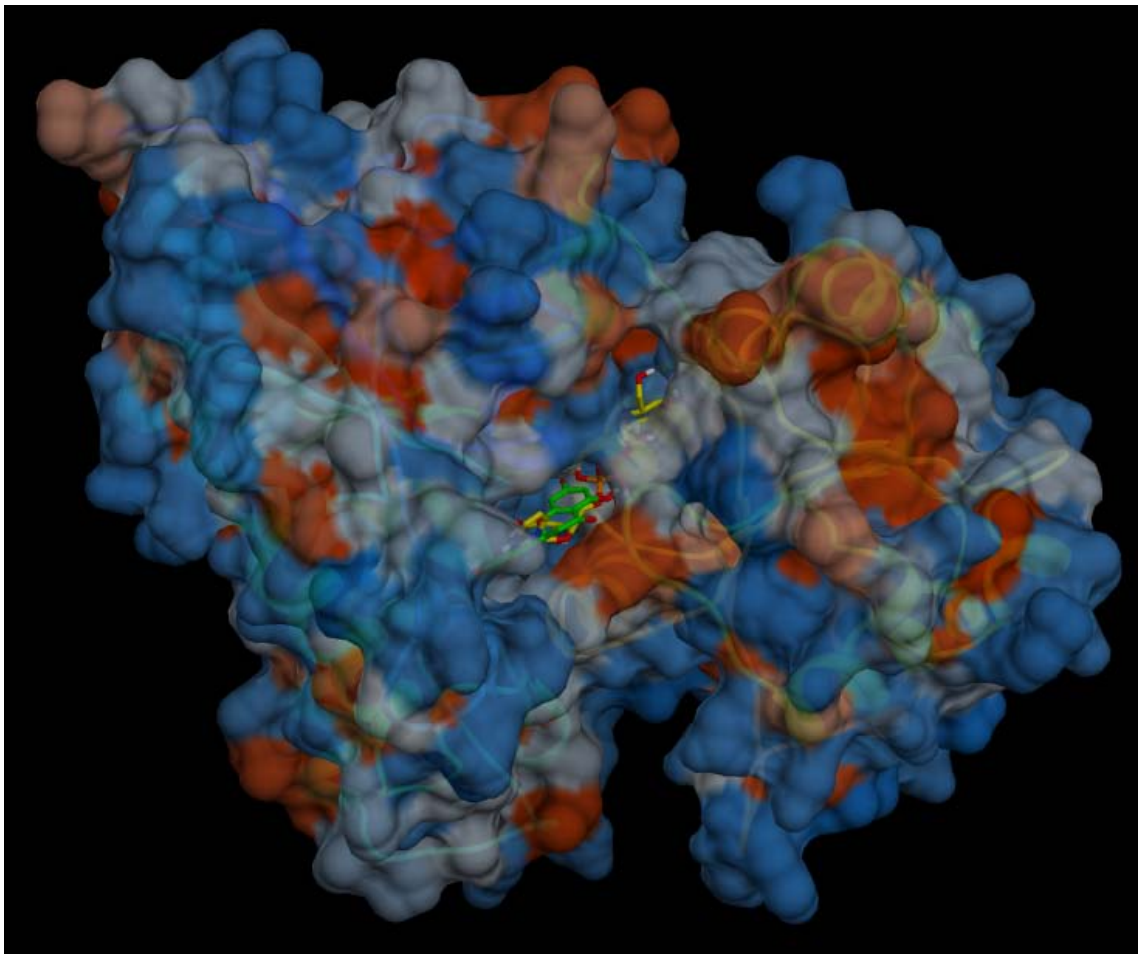


Figure 4.B.13 Binding pocket of the naringenin substrate and UDP-glucose

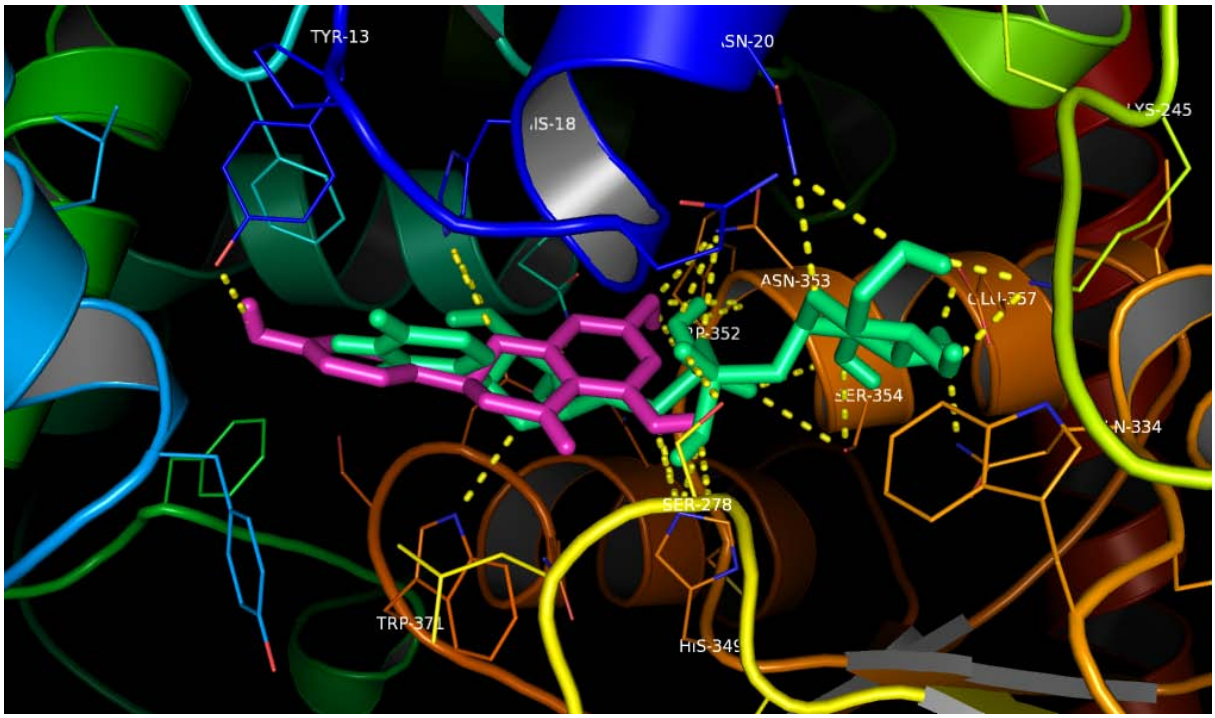


Figure 4.B.14 Interaction of catalytic residues of WsGT with substrate narigenin (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.15.3 Docking of luteolin acceptor and UDP-Glucose sugar donor

The interaction energy of luteolin with WsGT was -9.5 kcal/mol. Receptor ligand interactions are shown in PYMOL. In the template, active sites for substrate binding comprise Asp 244, Gln 24 and Asn 353. Gln 24 (NE2) interacts with 3`-OH group, Asp 244 (O) interacts with 3`-OH group and Asn 353 (ND2) interacts with 4-O group of luteolin molecule (Figure 4.B.16). All the hydrogen bond distances between WsGT and luteolin complex were observed within the range of 2.1 Å to 3.1 Å.

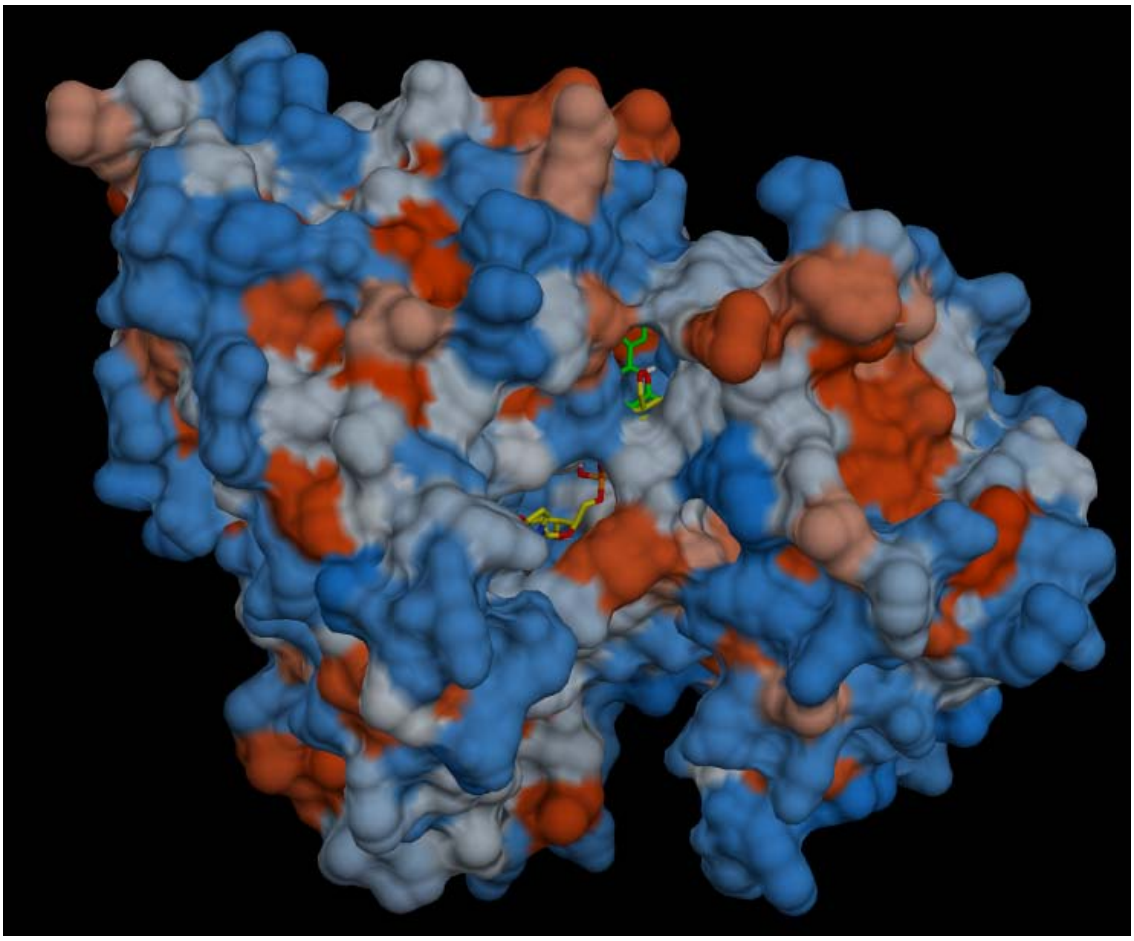


Figure 4.B.15 Binding pocket of the luteolin substrate and UDP-glucose

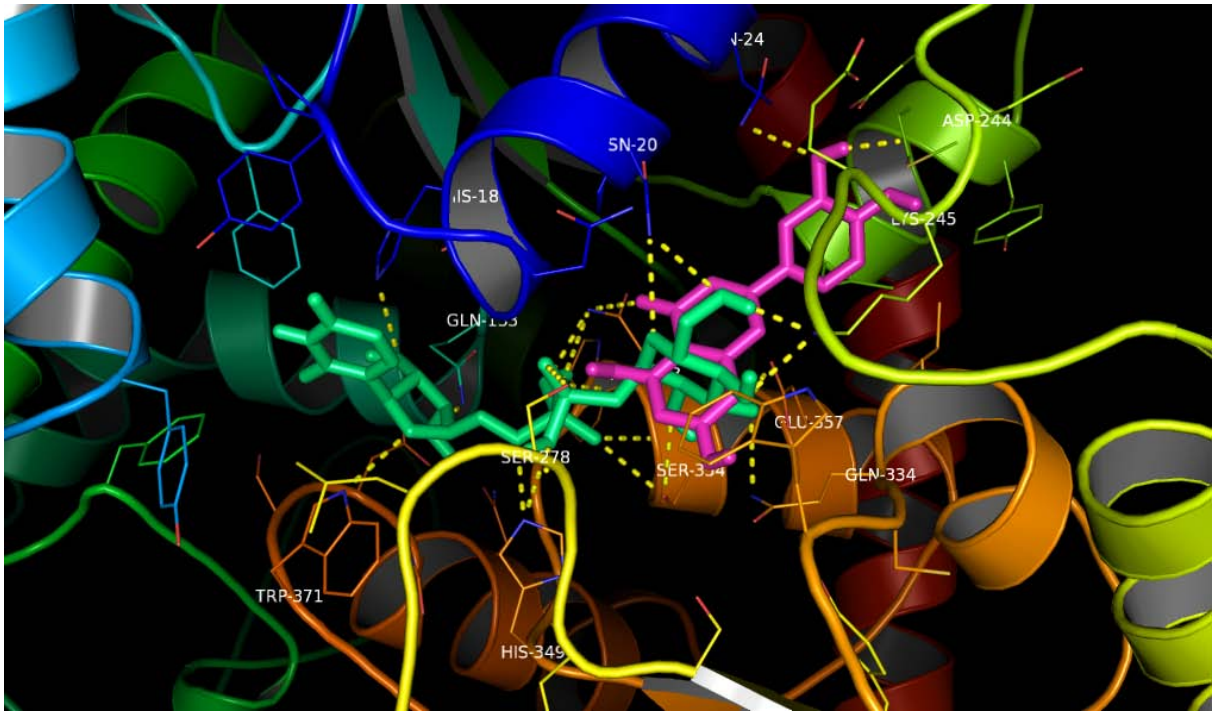


Figure 4.B.16 Interaction of catalytic residues of WsGT with substrate luteolin (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.15.4 Docking of genistein acceptor and UDP-Glucose sugar donor

The interaction energy of genistein with WsGT was -8.4 kcal/mol. Receptor ligand interactions are shown in PYMOL. In the template, active sites for substrate binding comprise Trp 352, Asn 353, Glu 357 and Trp 371. Trp 371 (NE1) interacts with 7-OH group, Trp 352 (N) interacts with 5-OH group, Asn 353 (ND2 & N) interacts with 4-O group and Glu 357 (OE1) interacts with 4`-OH group of genistein molecule (Figure 4.B.18). All the hydrogen bond distances between WsGT and genistein complex were observed within the range of 2.3 Å to 3.2 Å.

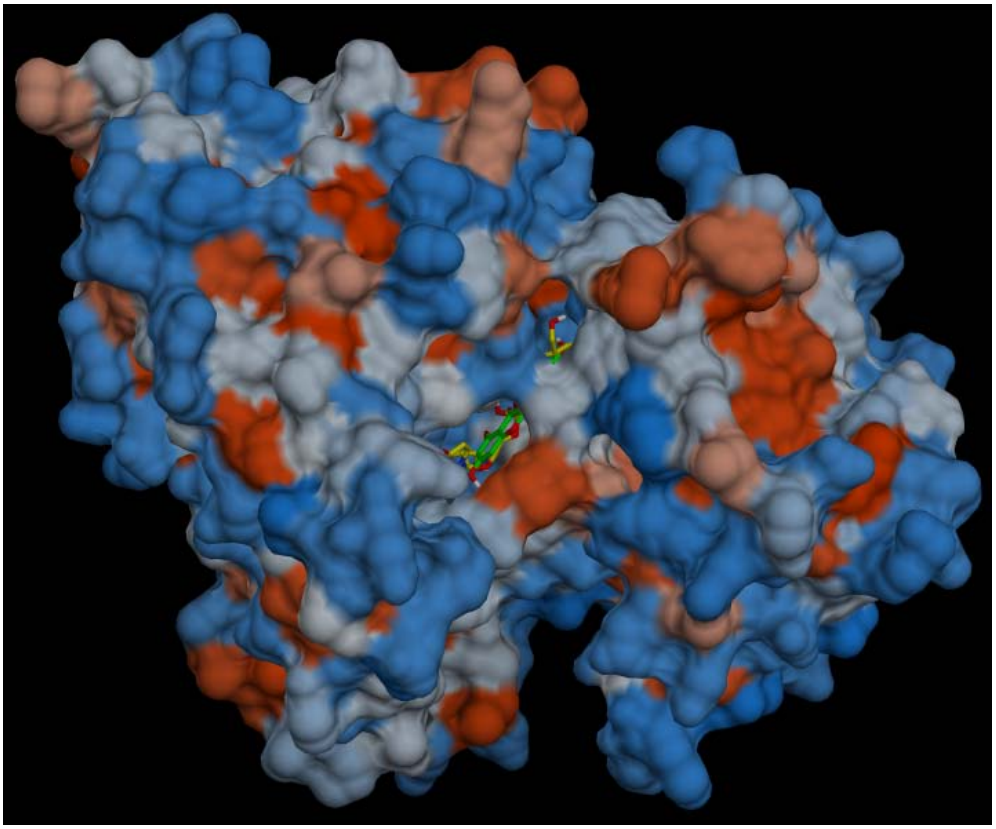


Figure 4.B.17 Binding pocket of the genistein substrate and UDP-glucose

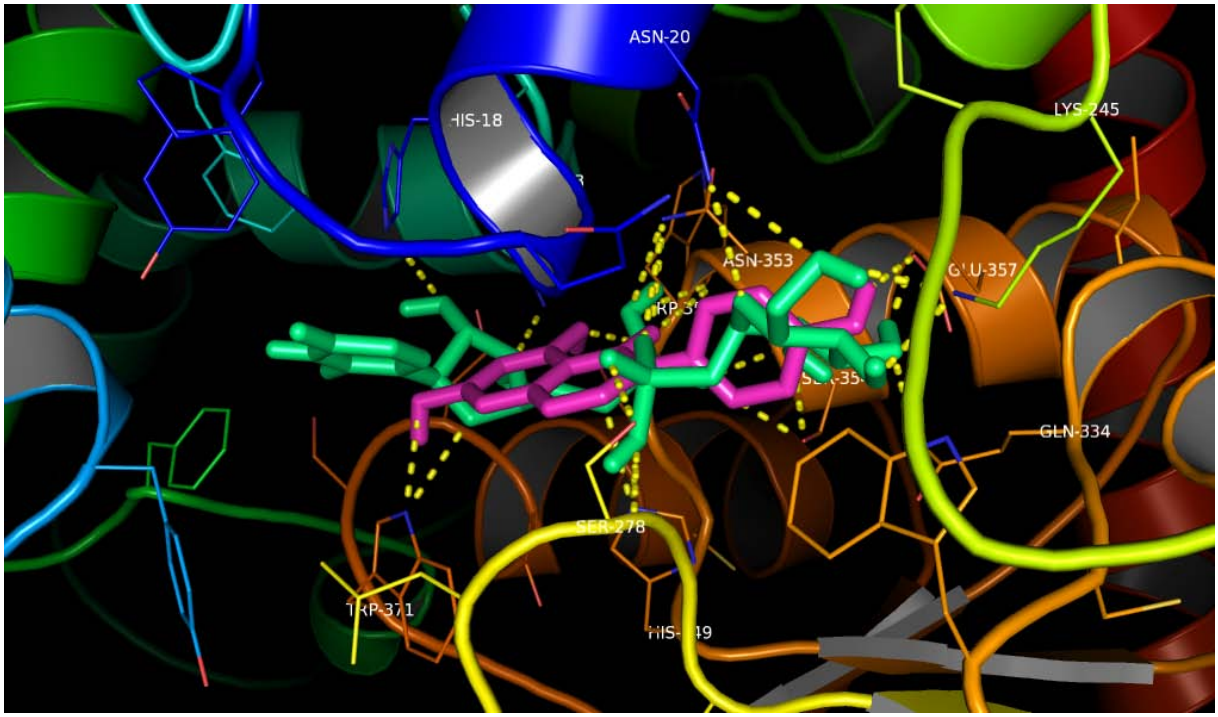


Figure 4.B.18 Interaction of catalytic residues of WsGT with substrate genistein (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.15.5 Docking of apigenin acceptor and UDP-Glucose sugar donor

The interaction energy of apigenin with WsGT was -9.2 kcal/mol. Receptor ligand interactions are shown in PYMOL. In the template, active sites for substrate binding comprise Trp 352, Asn 353, Tyr 13, His 18, Gln 133 and His 349.

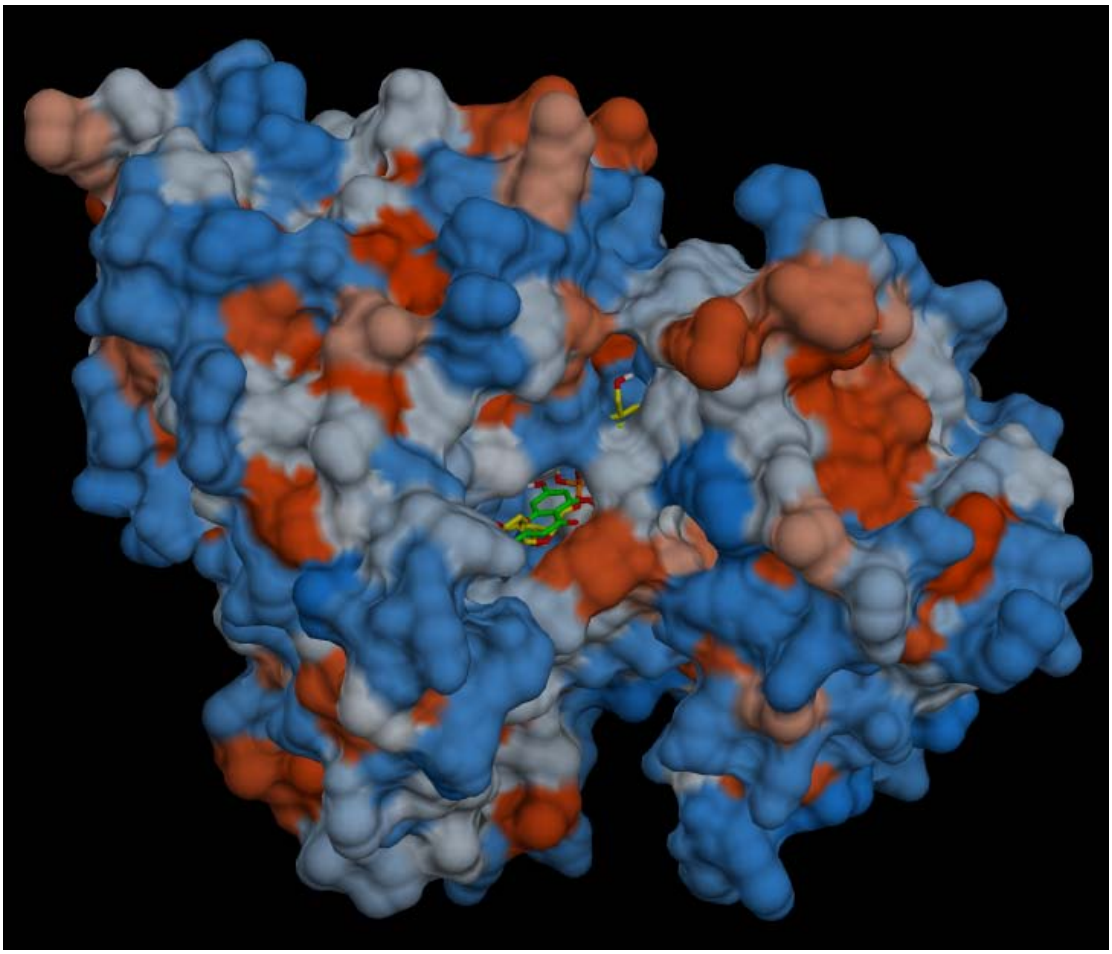


Figure 4.B.19 Binding pocket of the apigenin substrate and UDP-glucose

Trp 352 (N) interacts with 7-OH group, Asn 353 (N) interacts with 7-OH group, Gln 133 (OE1) interacts with 7-OH group, His 18 (NE2) interacts with 1-O group, His 349 (NE2) interacts with 5-OH group and Tyr 13 (UDP-Glucose) interacts with 4`-OH group of apigenin molecule (Figure 4.B.20). All the hydrogen bond distances between WsGT and apigenin complex were observed within the range of 2.1 Å to 3.4 Å.

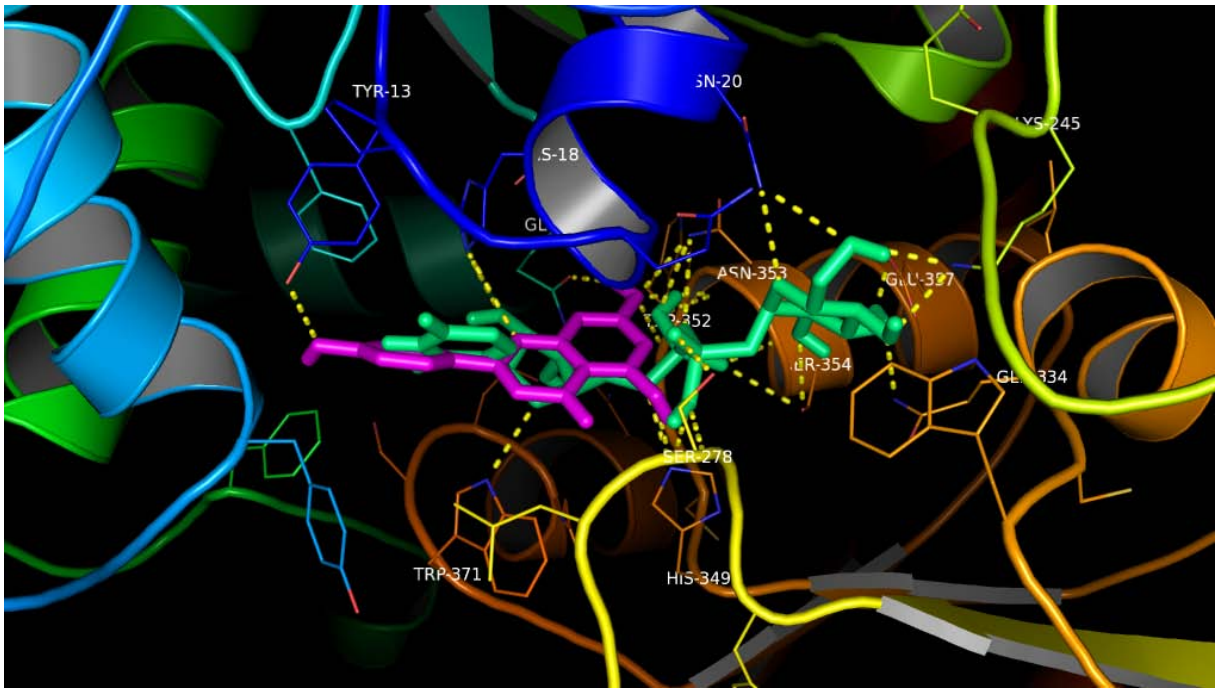


Figure 4.B.20 Interaction of catalytic residues of WsGT with substrate apigenin (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.15.6 Docking of Kaempferol acceptor and UDP-Glucose sugar donor

The interaction energy of kaempferol with WsGT was -9.2 kcal/mol. Receptor ligand interactions are shown in PYMOL. In the template, active sites for substrate binding comprise Trp 352, Asn 353, Tyr 13, His 18, Gln 133, His 349 and Trp 371.

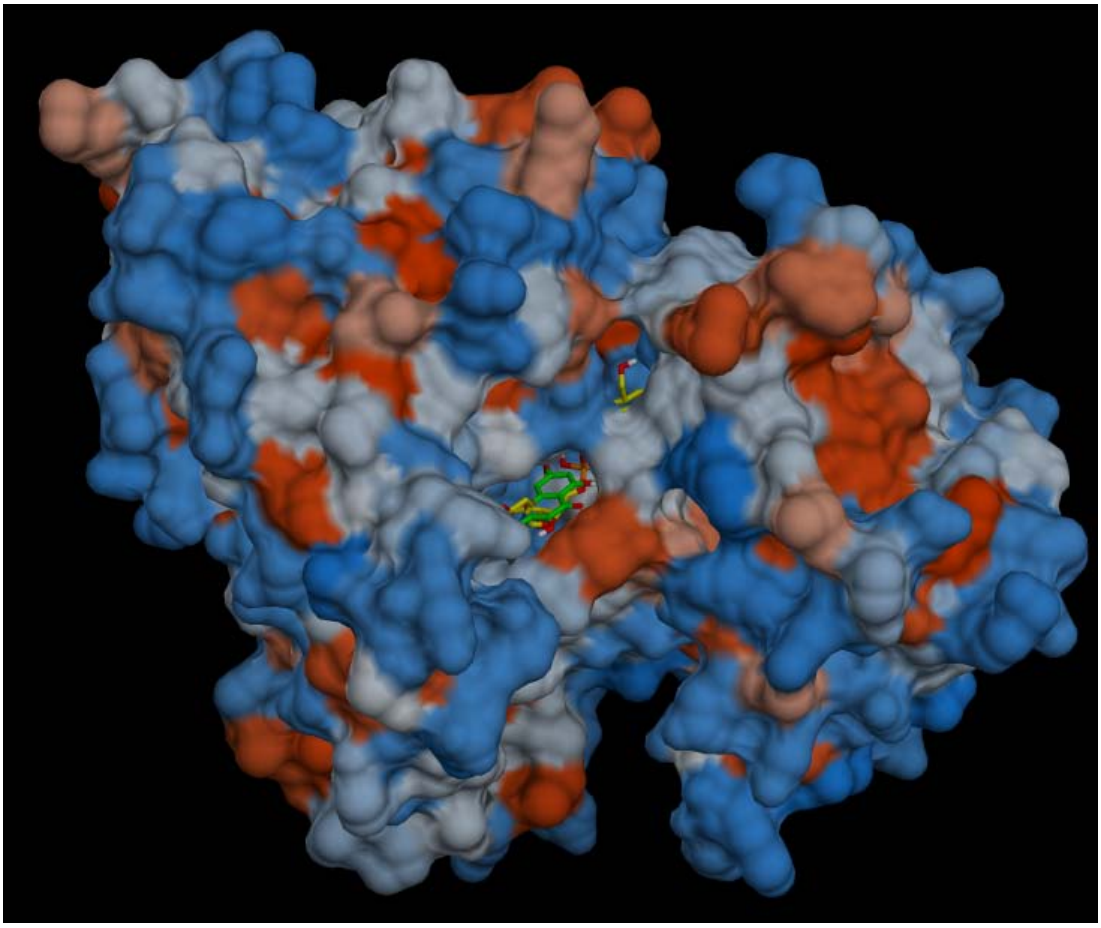


Figure 4.B.21 Binding pocket of the kaempferol substrate and UDP-glucose

Trp 352 (N) interacts with 7-OH group, Asn 353 (ND2 & N) interacts with 7-OH group, Gln 133 (OE1) interacts with 7-OH group, His 18 (NE2) interacts with 1-O group, His 349 (NE2) interacts with 5-OH group, Tyr 13 (UDP-Glucose) interacts with 4`-OH group and Trp 371 (NE1) interacts with 3-OH group of kaempferol molecule (Figure 4.B.22). All the hydrogen bond distances between WsGT and kaempferol complex were observed within the range of 2.5 Å to 3.3 Å.

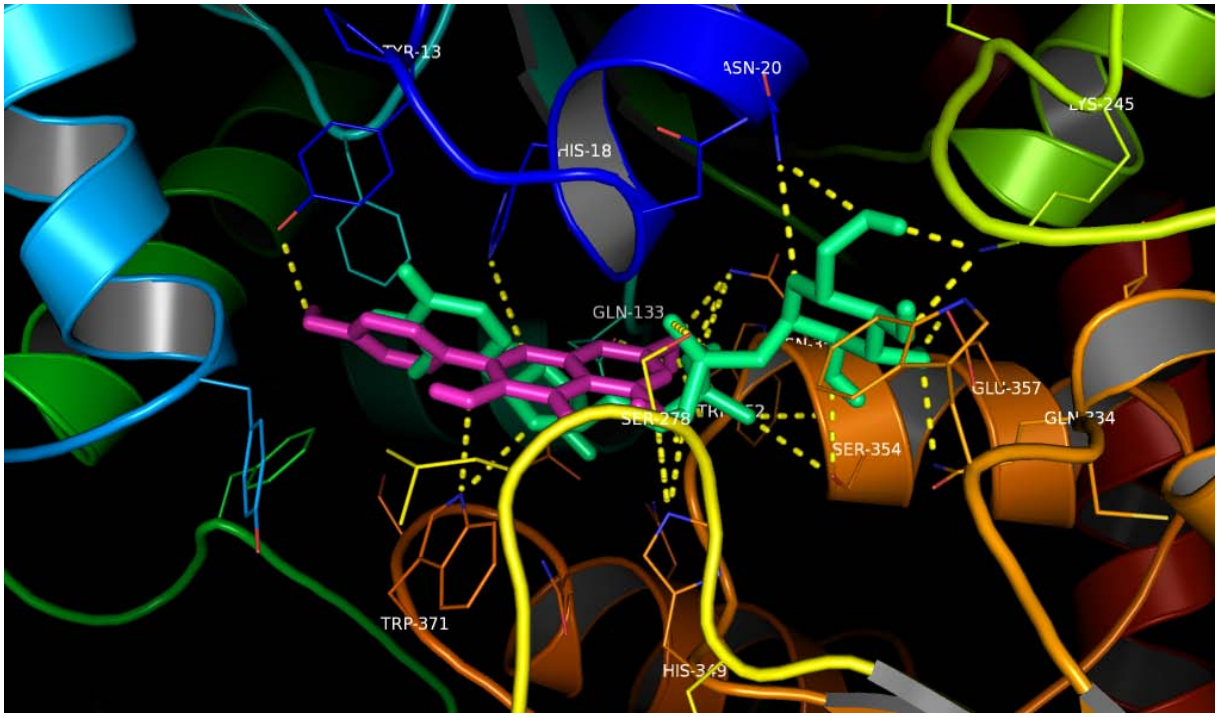


Figure 4.B.22 Interaction of catalytic residues of WsGT with substrate kaempferol (pink colour) and UDP-glucose (green colour). Hydrogen bonding interactions are indicated by yellow dashed lines.

4.16 Conclusion

- Homology modelling and docking studies of WsGT were done to validate the experimental data on substrate specificity. Homology model of WsGT showed 89.2% residues in the most favorable region.

- Homology modeling of WsGT protein was done with Modeller 9v9 and structure validation was done with PROCHECK and Pro SAII.

- Docking studies of WsGT was done with Autodock vina using UDP-glucose (donor) and acceptor (diadzein, luteolin, genistein, naringenin, kaempferol and apigenin) molecules. Docking studies confirmed the experimental results that WsGT glycosylate flavonol substrates at 7-OH position and it form 7-O-glucosides.

- Docking studies also show different amino acid residues interactions with UDP-glucose (sugar donor) and various substrates.

Chapter: 5

Tissue specific expression of glycosyltransferase gene-Real time PCR analysis

5.1 Introduction

Plants can synthesize many different low-molecular-weight compounds, defined as secondary plant metabolites. Part of this diversity arises from decoration with glycosyl-groups by glycosyltransferases. In general, many glycosyltransferases are involved in glycosylation of hormones such as auxins and cytokinins, secondary plant metabolites, such as flavonoids, plant growth regulators, such as ethylene and foreign compounds. When the plants are stimulated by a wound and pathogen in the environment, they respond to stress through the synthesis of defensive secondary metabolites and other defense-related proteins, and then the expression of glycosyltransferase homologue are up regulated.

Real-time reverse transcription PCR (real-time RT-PCR) is an established technique for quantifying mRNA in biological samples. Benefits of this procedure over conventional methods for measuring RNA include its sensitivity, large dynamic range, and the potential for high throughput as well as accurate quantification. Furthermore, many of the key proteins are found in such low abundance that real-time RT-PCR quantification of their mRNAs represents the only technique sensitive enough to measure reliably their expression *in vivo* (Huggett *et al*., 2005).

In the present chapter, *W. somnifera* leaves (young and mature), stem and roots were analyzed for relative expression of WsGT transcript using real time PCR. The modulation of expression of WsGT by salicylic acid (SA), methyl jasmonate (MJ) and heat shock was studied by real time PCR.

5.2 Materials and methods

5.2.1 Plant material

The surface sterilized seeds were inoculated on MS1/2 solid medium under aseptic conditions. Cultures were incubated at temperature 26±1°C under 16 hour photoperiod at 11.7 µmol/m^2/sec light intensity/8 h dark cycles for about 15 days. The sprouting was observed after 15 days and the seedlings were grown up to 3-5 cm within 4 weeks after inoculation. For rooting, seedlings were transferred on half strength MS without supplementing gelling agent. The seedlings were supported by 2 cm^2 whattman filter paper no.1 in culture tubes. The seedlings were grown up to 9-10 cm within 15 days after shifting. Well grown cultures were used for further stress experiments.

5.3 Methods

5.3.1 Stress treatment for Quantitative expression

In vitro seedlings were used for study of different stress treatments like methyl jasmonate, salicylic acid, cold and heat. Leaf, stem and root tissues were collected and incubated in MS1/2 liquid medium for different time intervals (2 h, 4 h, 6 h and 8 h) at different concentrations of methyl jasmonate and salicylic acid (50 µM and 100 µM). Approximately 0.2-0.4 gm of tissue was inoculated on the medium containing methyl jasmonate, salicylic acid. Tissues were also incubated for heat (42 ºC) treatment for 0 min to 90 min time period. After incubation period RNA was isolated and cDNA was prepared for Quantitative PCR analysis.

5.3.2 Total RNA extraction and its cDNA synthesis

Total RNA was isolated from experimental *in-vitro* leaf, stem and root cultures (stress cultures) according to the RNA isolation kit (Sigma). First strand cDNA synthesis, primed with an oligo $(dT)_{15}$ primer, was performed with Avian Myeloblastosis Virus Reverse

Transcriptase (AMV-RT) according to the manufacturer's protocol (Promega, USA) (Chapter 2; section 2.8.6.3).

5.3.3 Quantitative Real-Time PCR (QRT-PCR)

Quantitative Real Time PCR is a powerful tool for gene expression analysis and was first demonstrated by Higuchi *et al.* (1992, 1993). Q-PCR quantifies the initial amount of the template most specifically and is a preferable alternative to other forms of quantitative reverse transcriptase PCR that detect the amount of final amplified product at the end point (Freeman *et al.,* 1999). Q-PCR monitors the fluorescence emitted during the reaction as an indicator of amplicon production during each PCR cycle (*i.e.*, in real time) as opposed to the endpoint detection. In the present study the SYBR Brilliant® II QPCR Master Mix (Stratagene, USA) was used. This kit supports quantitative amplification and detection systems. The kit supports PCR amplifications and detection of a variety of DNA targets, including genomic DNA, plasmid DNA, and cDNA. The SYBR Brilliant II QPCR master mix includes SureStart® *Taq* DNA polymerase, a modified version of *Taq2000*™ DNA polymerase with hot start capability. A passive reference dye (an optional reaction component) is provided so as to control the final dye concentration, increasing the flexibility of the reagents for use with different platforms. The fluorescent dye SYBR Green I in the master mix enables the analysis of many different targets without having to synthesize target-specific labeled probes. Real time PCR can be divided into four major phases: the linear ground phase, early exponential phase, log-linear (also known as exponential) phase, and plateau phase. During the linear ground phase (usually the first 10–15 cycles), PCR is just beginning, and fluorescence emission at each cycle has not yet risen above background. Baseline fluorescence is calculated at this time. At the early exponential phase, the amount of fluorescence has reached a threshold where it is significantly higher (usually 10 times the standard deviation of the baseline) than background levels. The cycle at which this occurs is known as Ct in ABI Prism® literature (Applied Biosystems, Foster City, CA, USA) or crossing point (CP) in LightCycler® literature (Roche Applied Science, Indianapolis, IN, USA). This value is representative of the starting copy number in the original template and is used to calculate experimental results. During the log-linear phase, PCR reaches its optimal amplification

period with the PCR product doubling after every cycle in ideal reaction conditions. Finally, the plateau stage is reached when reaction components become limited and the fluorescence intensity is no longer useful for data calculation. In general, lower Ct value indicates higher initial copies.

The cDNA first strands were used as template for semi and absolute quantification of GT gene transcripts. Brilliant II™ SYBR® Green Q-PCR master mix (Stratagene, Agilent Technology, TX, USA) was used for real-time PCR. The reaction mixture was prepared as described in chapter 2; section 2.8.5.2.

5.3.4 Relative and absolute quantification methods

Relative (comparative) quantification relates the PCR signal of the target transcript in a treatment group to that of another such as untreated control.

The comparative Ct method is also known as the $2^{-\Delta\Delta Ct}$ method, where

$\Delta\Delta C_t = \Delta Ct$, sample - ΔCt, reference.

Here, $\Delta Ct_{,\ sample}$ is the Ct value for any sample normalized to the endogenous housekeeping gene and $\Delta Ct_{,\ reference}$ is the Ct value for the calibrator also normalized to the endogenous housekeeping gene.

Means,

$\Delta Ct_{,\ sample} = Ct_{,\ sample} - Ct_{,\ endogenous\ control\ (18S\ rRNA\ gene\ or\ any\ other\ housekeeping\ gene)}$

$\Delta Ct_{,\ reference} = Ct_{,\ reference/\ normal/\ untreated\ sample} - Ct_{,\ endogenous\ control.}$

For the ΔΔCt calculation to be valid, the amplification efficiencies of the target and the endogenous reference must be approximately equal. This can be established by looking at how Δ Ct varies with template dilution. If the plot of cDNA dilution versus ΔC_t is close to zero, it implies that the efficiencies of the target and housekeeping genes are very similar. If a housekeeping gene cannot be found whose amplification efficiency is similar to the target, then the standard curve method is preferred.

5.3.5 QRT-PCR considerations: See chapter 2: section 2.11.1.

5.3.6 Preparing the QRT-PCR reactions: See chapter 2: section 2.11.2.

5.3.7 Real-time quantitative PCR for *GT* genes

Total RNA was extracted individually from leaf, stem and root from *in-vitro* cultures of *W. somnifera* and also from the tissue cultures under stress. 2 µg of total RNA was used for making cDNA using ImProm cDNA synthesis kit (Promega, USA). Brilliant SYBRGreen Q-PCR kit (Stratagene, USA) and Stratagene Mx3000P real time machine were used for all reactions. The primer sequences that were designed for *W. somnifera WsGT* gene are given in Table 5.A.1. Optimal numbers of PCR cycles within the linear range of amplification for each gene were determined in preliminary experiments. QRT-PCR reactions were performed following conditions mentioned in Chapter 2: section 2.11.3 with annealing temperature of 58 ^{o}C. The reaction was run in triplicates and repeated twice. It was ensured that equal quantity of RNA template was used for each reaction (Pfaffl, 2001; Freeman *et al*., 1999; Edyta Zdunek-Zastocka, 2008).

5. A.1 Table Primers designed for QPCR analysis of WsGT gene

Primers	**Sequence 5'- 3'**	**T_m in ºC**
WsGTF	ACTTCCTCCTACTCAAGTTGATG	48.3
WsGTR	CTGGTTCAGAAGTAGACTCAAAAC	48.8
Probe	AGGTACGTCTGAACTCTCAATTGCA	50.9

5.4 Results and discussion

Expression studies were done using *in vitro* seedlings of *W.somnifera.* Stratagene Mx 3000P real time machine was used for expression studies.

5.4.1 Tissue specific expression studies of WsGT gene-Real time PCR

Tissue specific expression of WsGT gene was carried out by isolating total RNA from leaves, stem and roots of developing plants of *W. somnifera* and plants which were subjected to different stress conditions, like salicylic acid 50 µM and 100 µM; methyl jasmonate 50 µM and 100 µM; heat shock by shifting tissues from 25 to 42 ºC treatments. Plants without any stress were used as control.

5.4.2 Tissue specific expression of WsGT gene – 0 hour

Based on the Ct values, the expression patterns in different tissue types i.e. young leaf, mature leaf, stem and root were analyzed. The mean Ct value was calculated for three replicates of each sample along with the internal control (5.8S rRNA) gene. The mean Ct values obtained from all the samples along with the standard deviation were used for analysis. It was assumed that the reference gene (5.8S rRNA) does not vary in copy number or expression level in the samples of study which is evident by the very little fluctuations in the Ct values obtained from RT-PCR amplification plots.

The highest expression of WsGT gene was observed in young leaf followed by mature leaf and stem. Lowest expression of WsGT was observed in root tissue. Based on $2^{-\Delta\Delta Ct}$ values, the highest expression obtained for WsGT transcript was for young leaf tissue; which showed 65 fold expression in comparison with least expressed stem (15-fold) and root tissue (Figure 5.B.1). These results are in agreement with more accumulation of flavonoids glycosides in leaf tissue.

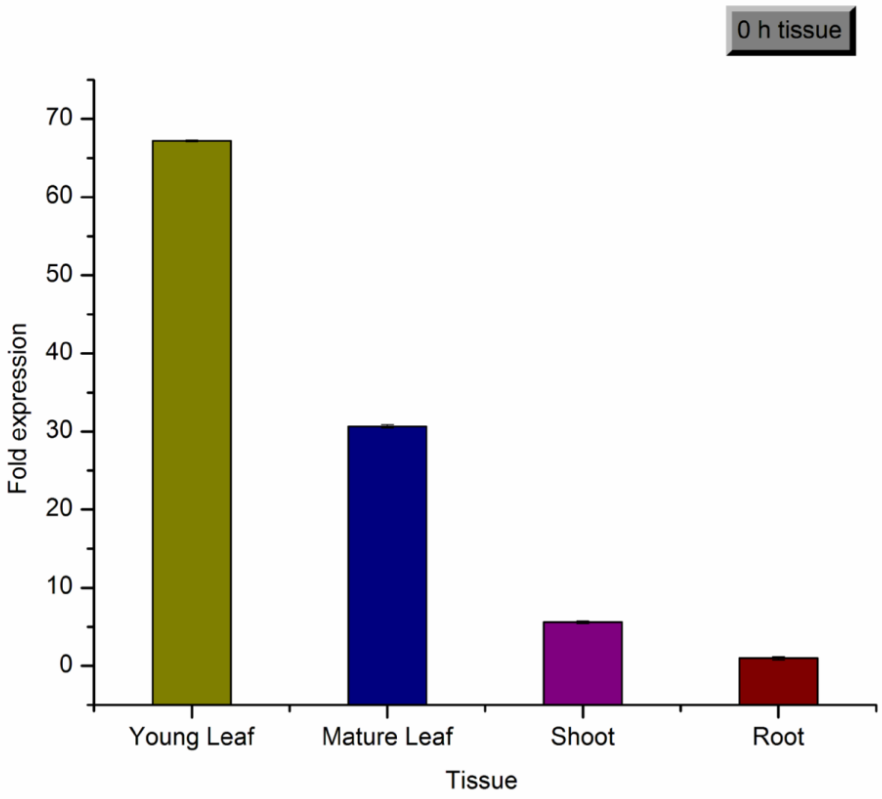


Figure 5.B.1: Fold expression of WsGT gene in different tissue types without any stress conditions

5.4.3 Real time PCR analysis for tissue specific expression of WsGT gene in leaf tissue in the presence of salicylic acid, 50 µM

W. somnifera leaf tissue was subjected to treatment with 50 µM salicylic acid for 2 h, 4 h, 6 h and 8 h. Here control tissues were used without any stress treatment. The incubation of leaf tissue in the presence of salicylic acid resulted in transient accumulation of WsGT transcript. It increased gradually to reach the maxima of 120-fold at 6 hours followed by a decline at 8 hour (Figure 5.B.2). The salicylic acid signaling is known to induce defense-related genes, particularly those encoding pathogenesis related proteins, cellular protectant glutathione-S-transferase, cytochrome P_{450}, ABC transporters and glucosyltransferases (Lokendra Kumar Sharma, 2007).

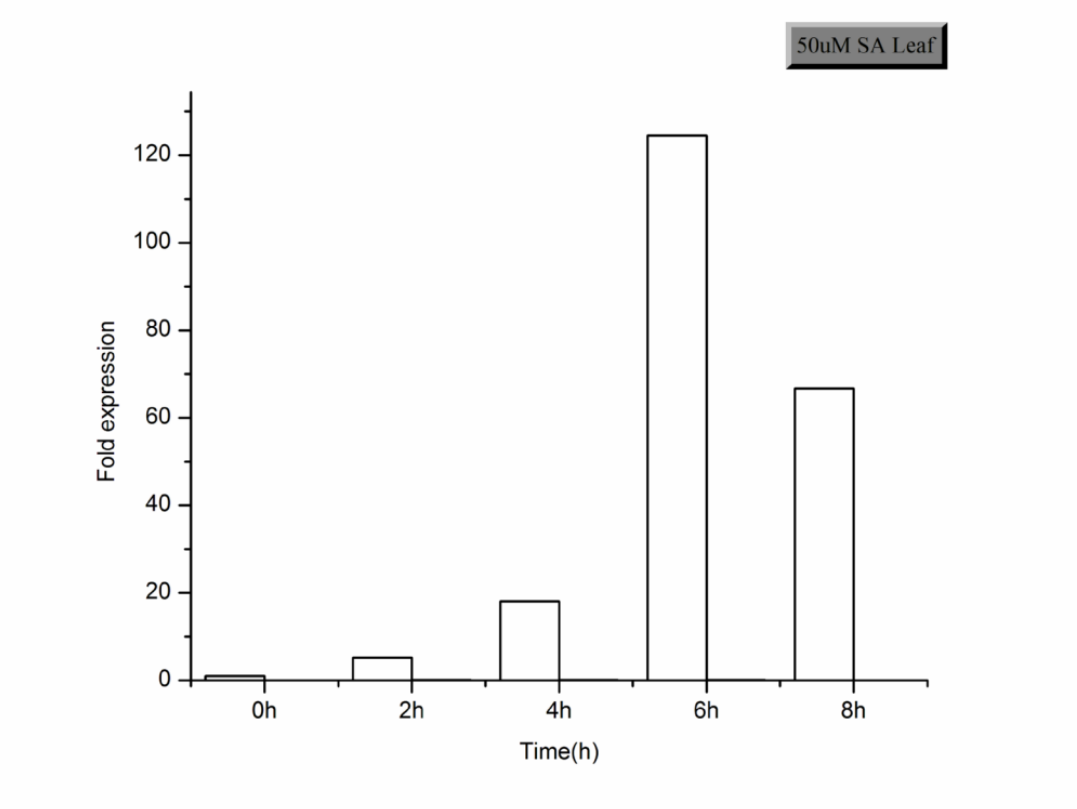


Figure 5.B.2: Relative expression in terms of mean fold expression of WsGT gene in leaf tissue under 50 µM salicylic acid stress at 2 h, 4 h, 6 h and 8 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.4 Real time PCR analysis for tissue specific expression of WsGT gene in stem tissue in the presence of salicylic acid, 50 µM

W. somnifera stem tissue was subjected to treatment with 50 µM salicylic acid for 2 h, 4 h, 6 h and 8 h. Here control tissues were used without any stress treatment. After salicylic acid stress there is little increase in fold expression in stem tissue at 6 hours (45-fold) and 8 hours (40-fold). The expression was low at 2 hour and 4 hour (15-fold) incubation and increase it increased at 6 hour and 8 hour (45-fold) (Figure 5.B.3). Stem shows WsGT expression without stress (15-fold) (Figure 5.B.1), which increases to 45-fold at 6 hour of incubation at 50 µM SA.

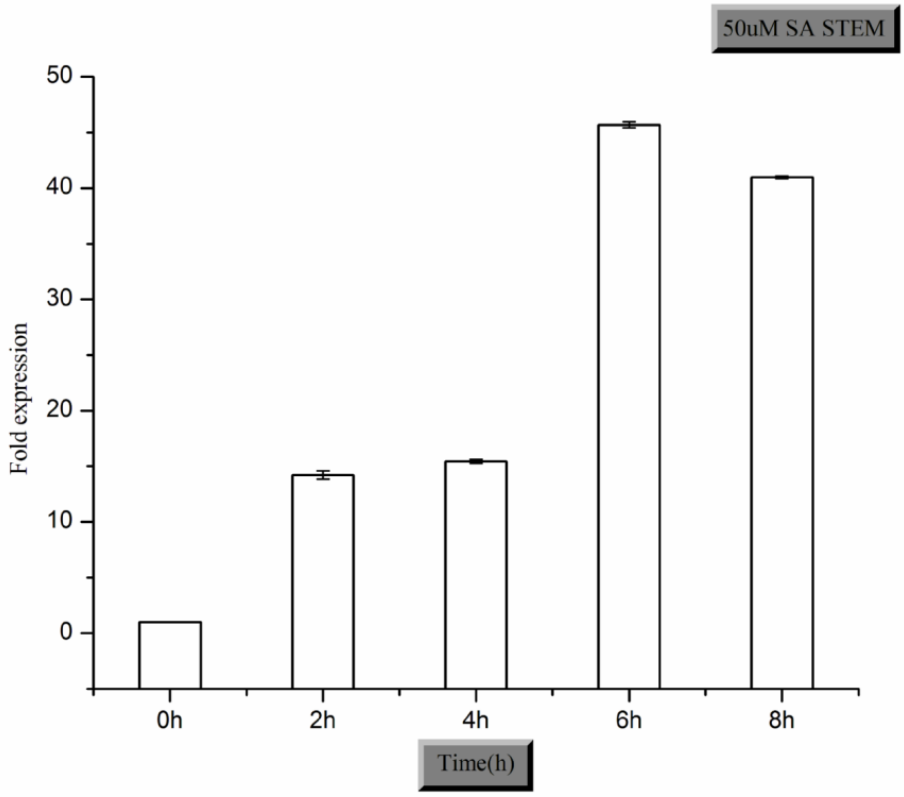


Figure 5.B.3: Relative expression in terms of mean fold expression of WsGT gene in stem tissue under 50 µM salicylic acid stress at 2 h, 4 h, 6 h and 8 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.5 Real time PCR analysis for tissue specific expression of WsGT gene in root tissue in the presence of salicylic acid, 50 µM

W. somnifera root tissue was subjected to treatment with 50 µM salicylic acid for 2 h, 4 h, 6 h and 8 h. Here control tissues were used without any stress treatment. When root tissue was subjected to salicylic acid stress there is not much increase in the WsGT expression. During 0 hour expression analysis root showed very little expression of WsGT transcript (less than 5-fold) (Figure 5.B.1). After stress incubation it increased gradually to reach maxima of 12-fold at 6 hour and slight decline at 8 hours (Figure 5.B.4).

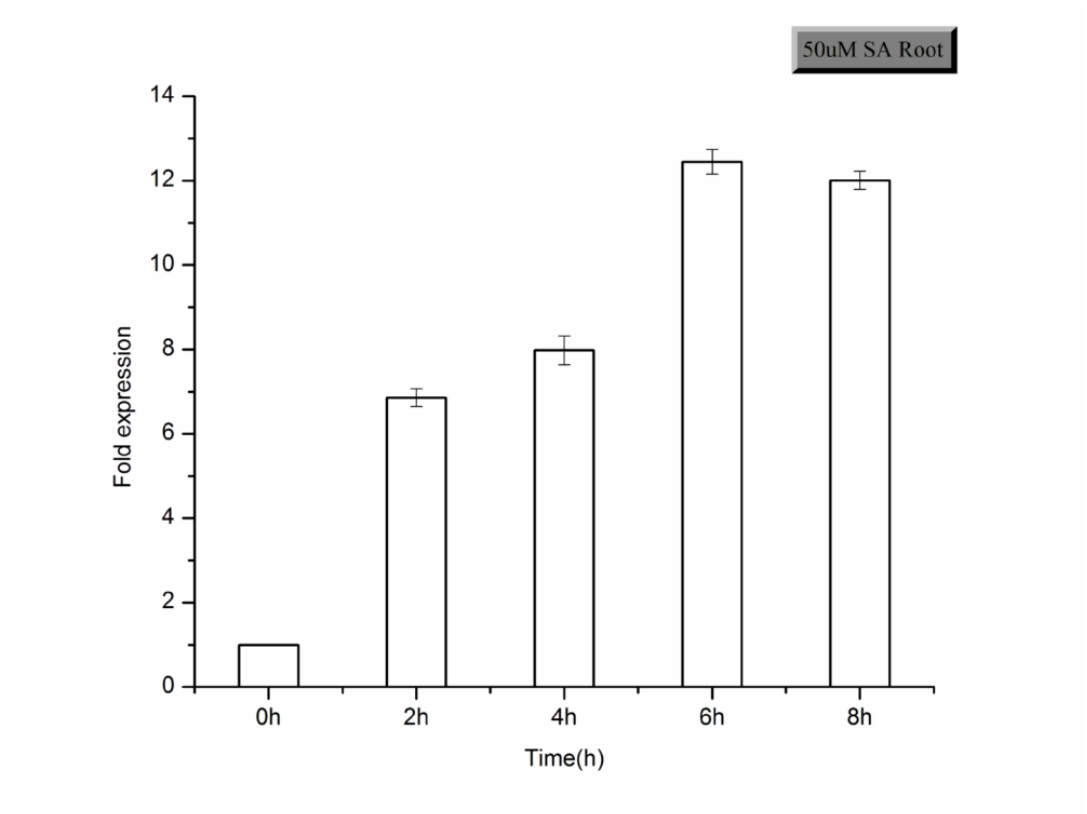


Figure 5.B.4: Relative expression in terms of mean fold expression of WsGT gene in root tissue under 50 µM salicylic acid stress at 2 h, 4 h, 6 h and 8 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.6 Real time PCR analysis for tissue specific expression of WsGT gene in leaf tissue in the presence of salicylic acid, 100 µM

W. somnifera leaf tissue was subjected to treatment with 100 µM salicylic acid for 2 h, 4 h, 6 h, 8 h and 12 h. Here control tissues were used without any stress treatment. The incubation of leaf tissue in the presence of 100 µM salicylic acid shows gradual increase of expression from 2 hours to 8 hours. WsGT revealed higher level of expression in leaf tissue compared to that in stem and root (Figure 5.B.1). After incubation at 100 µM of salicylic acid it shows 85-fold expression at 8 hour and little decrease at 12 hour incubation (Figure 5.B.5).

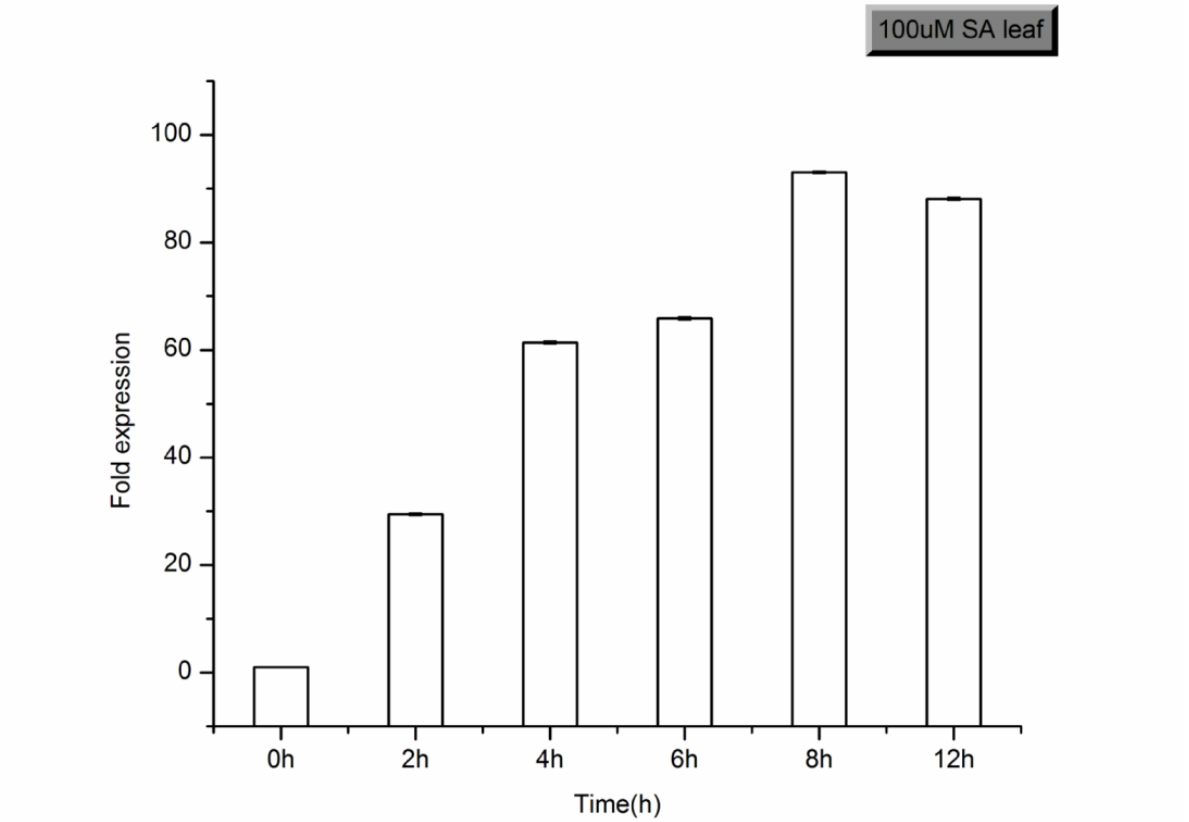


Figure 5.B.5: Relative expression in terms of mean fold expression of WsGT gene in leaf tissue under 100 µM salicylic acid stress at 2 h, 4 h, 6 h, 8 h and 12 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.7 Real time PCR analysis for tissue specific expression of WsGT gene in stem tissue in the presence of salicylic acid, 100 µM

W. somnifera stem tissue was subjected to treatment with 100 µM salicylic acid for 2 h, 4 h, 6 h, 8 h and 10 h. Here in 100 µM stress, WsGT showed gradual increase of expression up to 8 hours. In 50 µM stress at 6 hour duration expression was 45-fold (Figure 5.B.3). Even after increase of salicylic acid concentration and duration of incubation, WsGT shows 35-fold expression at 8 hour and decline at 10 hour (Figure 5.B.6). The increase of salicylic acid concentration doesn't show increase in WsGT expression in stem tissue.

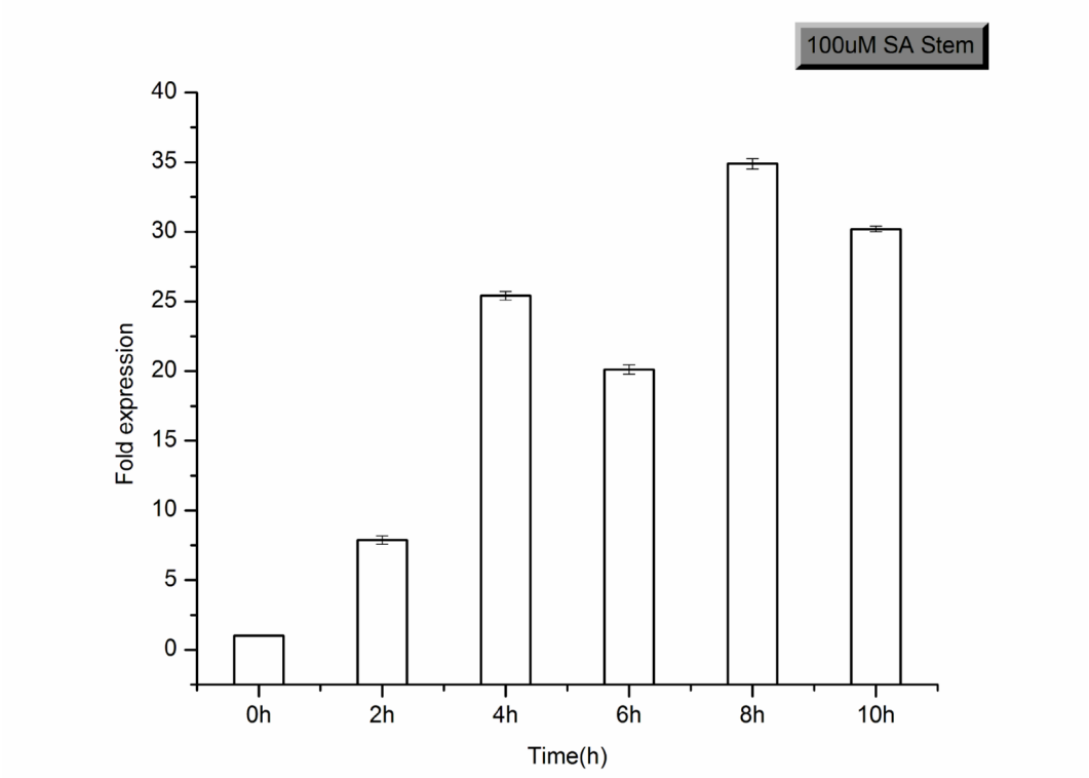


Figure 5.B.6: Relative expression in terms of mean fold expression of WsGT gene in stem tissue under 100 µM salicylic acid stress at 2 h, 4 h, 6 h, 8 h and 10 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.8 Real time PCR analysis for tissue specific expression of WsGT gene in root tissue in the presence of salicylic acid, 100 µM

W. somnifera root tissue was subjected to treatment with 100 µM salicylic acid for 2 h, 4 h, 6 h, 8 h and 10 h. The incubation of root tissue in 100 µM salicylic acid shows increase of WsGT transcript accumulation. The expression of WsGT gene in root tissue at 0 hour without any stress is very low. After incubation in 50 µM salicylic acid concentration there is little increase in expression (12-fold) at 6 hour. And in 100 µM salicylic acid concentration transcript accumulation increased gradually from 4 hour (25-fold) to reach maxima of 80-fold at 8 hour (Figure 5.B.7). These results indicate effect of salicylic acid stress on WsGT gene expression in root tissue.

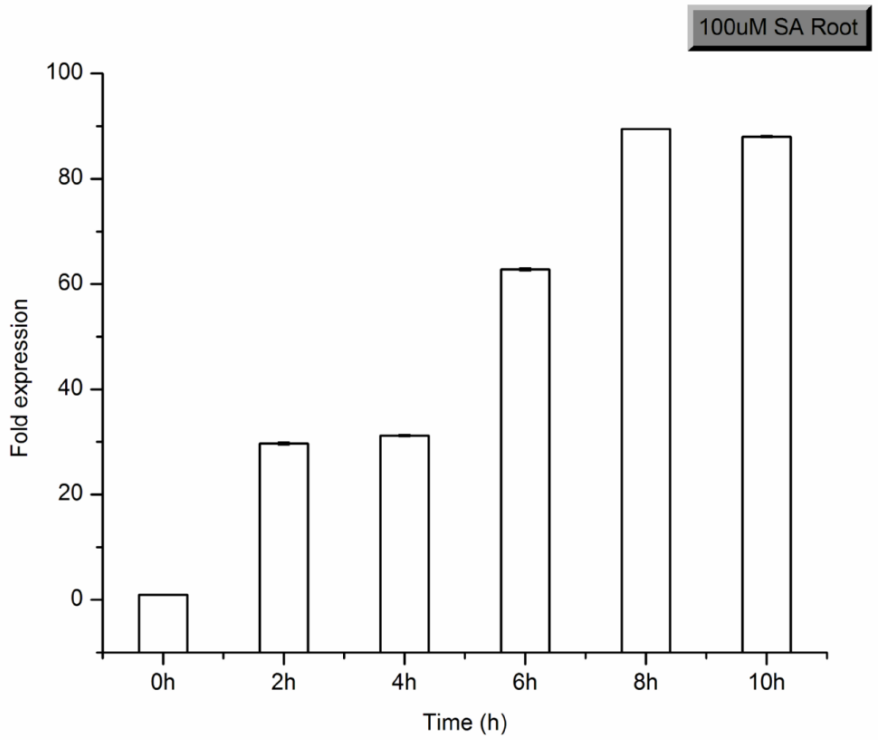


Figure 5.B.7: Relative expression in terms of mean fold expression of WsGT gene in root tissue under 100 µM salicylic acid stress at 2 h, 4 h, 6 h, 8 h and 10 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.9 Real time PCR analysis for tissue specific expression of WsGT gene in leaf tissue in the presence of methyl jasmonate, 50 µM

W. somnifera leaf tissue was subjected to treatment with 50 µM methyl jasmonate for 2 h, 4 h, 6 h, 8 h and 10 h. Comparing with salicylic acid stress (Figure 5.B.2), methyl jasmonate does not show much effect on WsGT expression in leaf tissue. Till 6 hour incubation there is only 20-fold expression and at 8 hour incubation expression increased 55-fold. In 10 hour incubation expression was declined (Figure 5.B.8).

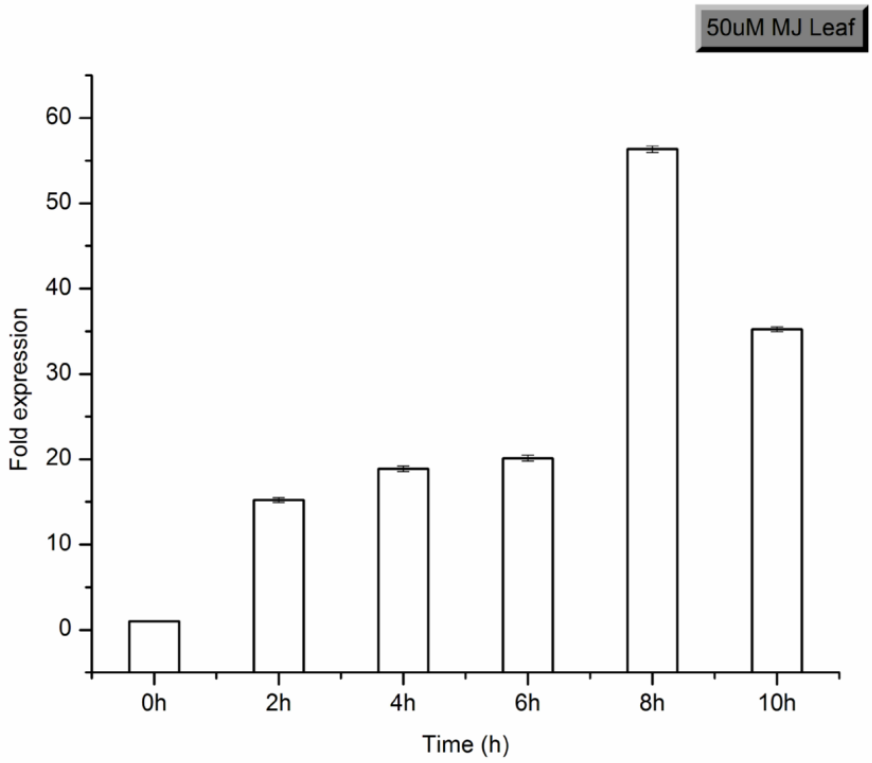


Figure 5.B.8: Relative expression in terms of mean fold expression of WsGT gene in leaf tissue under 50 µM methyl jasmonate stress at 2 h, 4 h, 6 h, 8 h and 10 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.10 Real time PCR analysis for tissue specific expression of WsGT gene in stem tissue in the presence of methyl jasmonate, 50 µM

W. somnifera stem tissue was subjected to treatment with 50 µM methyl jasmonate for 2 h, 4 h, and 6 h. The incubation of stem tissue in the presence of methyl jasmonate resulted in gradual increase of WsGT expression from 2 hour incubation and stable at 4 hour and 6 hour (Figure 5.B.9). When 50 µM salicylic acid (Figure 5.B.3) expression results compared with methyl jasmonate, due to salicylic acid stress WsGT expression was increased 45-fold, but mehyl jasmonate stress show only 20-fold expression.

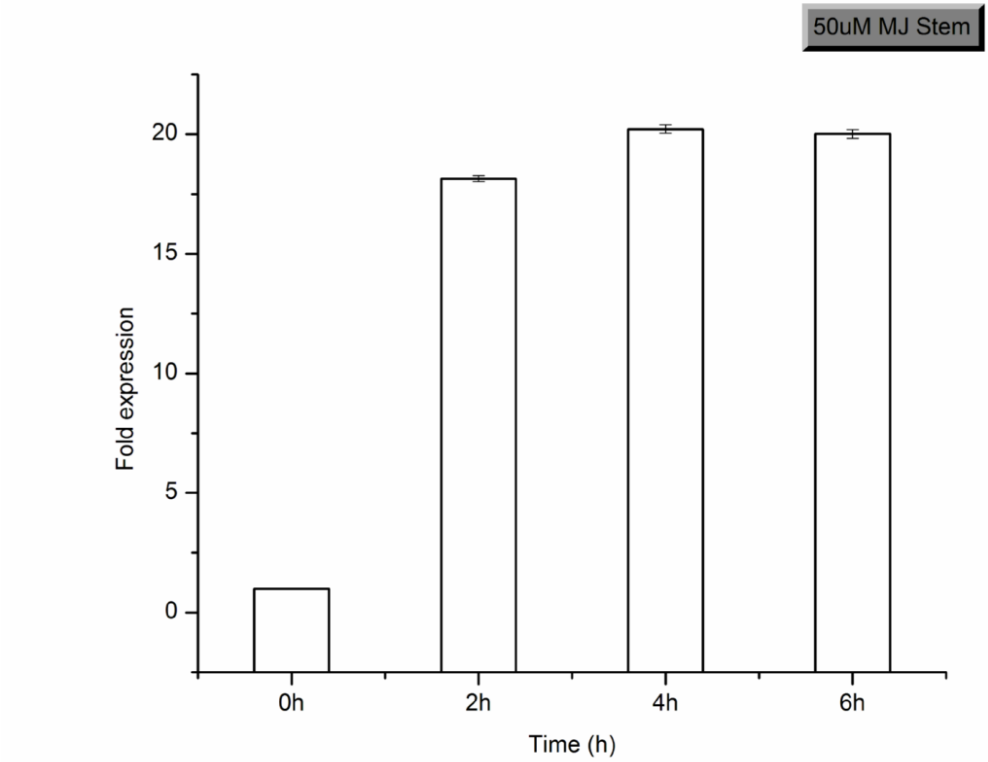


Figure 5.B.9: Relative expression in terms of mean fold expression of WsGT gene in stem tissue under 50 µM methyl jasmonate stress at 2 h, 4 h, 6 h, 8 h and 10 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.11 Real time PCR analysis for tissue specific expression of WsGT gene in root tissue in the presence of methyl jasmonate, 50 µM

W. somnifera root tissue was subjected to treatment with 50 µM methyl jasmonate for 2 h, 4 h, 6 h, 8 h and 10 h. When root tissue was analyzed for WsGT expression under methyl jasmonate stress, it showed gradual increase in expression from 2 hour to 8 hour and a little decline at 10 hour (Figure 5.B.10).

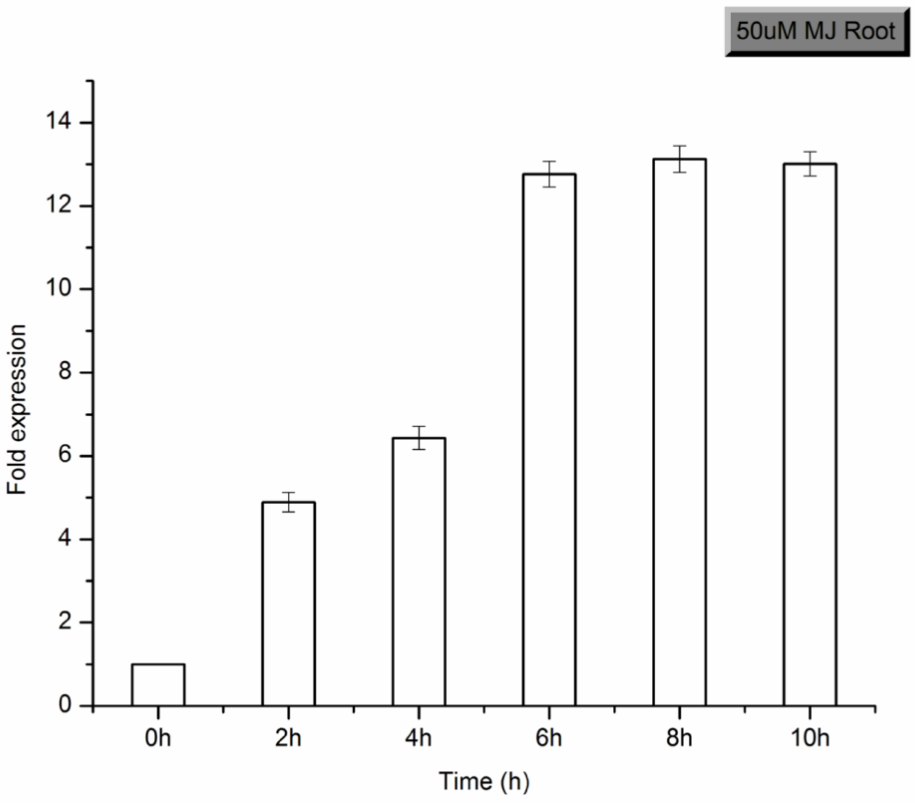


Figure 5.B.10: Relative expression in terms of mean fold expression of WsGT gene in root tissue under 50 µM methyl jasmonate stress at 2 h, 4 h, 6 h and 10 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.12 Real time PCR analysis for tissue specific expression of WsGT gene in leaf tissue in the presence of methyl jasmonate, 100 µM

W. somnifera leaf tissue was subjected to treatment with 100 µM methyl jasmonate for 2 h, 4 h, 6 h, 8 h and 12 h. When methyl jasmonate concentration was increased from 50 µM to 100 µM, WsGT expression in leaf tissue was increased from 2 hour to 8 hour incubation. In 50 µM stress WsGT expression was 20-fold till 6 hour incubation and increased 55-fold at 8 hour duration. But in 100 µM stress, in short duration (4 hour) expression was increased (20-fold), till the 8 hour duration (25-fold) and decline at 12 hour incubation (Figure 5.B.11).

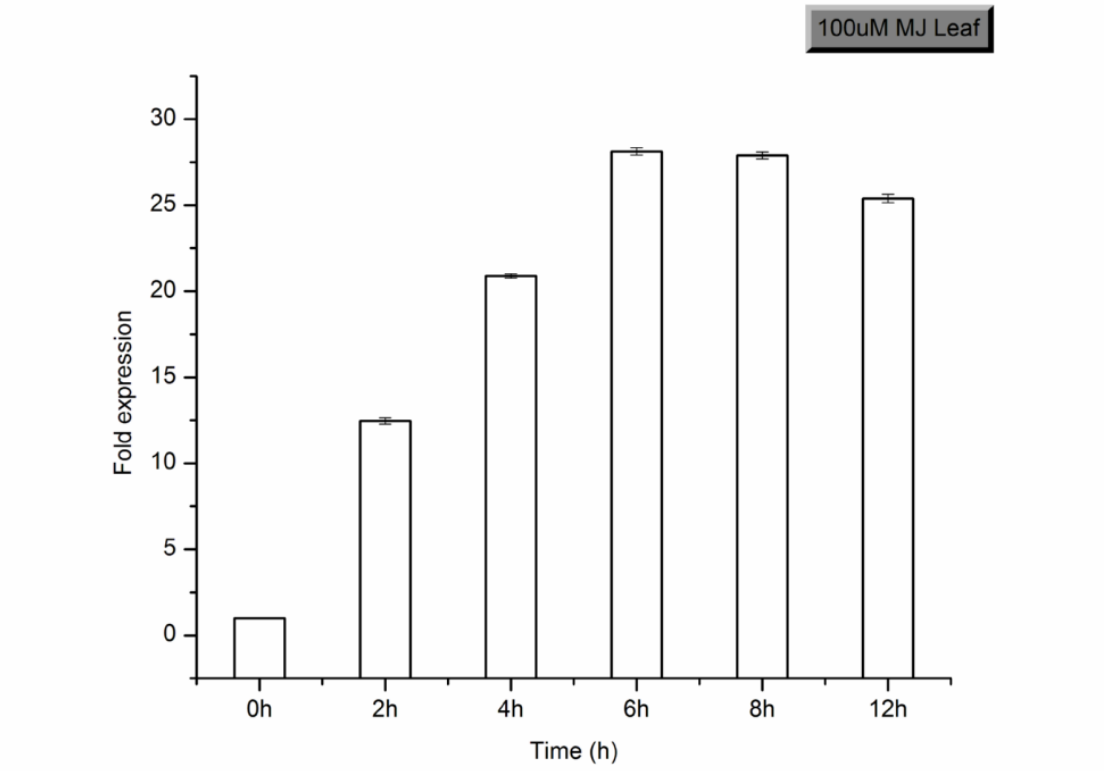


Figure 5.B.11: Relative expression in terms of mean fold expression of WsGT gene in leaf tissue under 100 µM methyl jasmonate stress at 2 h, 4 h, 6 h, 8 h and 12 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.13 Real time PCR analysis for tissue specific expression of WsGT gene in stem tissue in the presence of methyl jasmonate, 100 µM

W. somnifera stem tissue was subjected to treatment with 100 µM methyl jasmonate for 2 h, 4 h and 6 h. There is not much effect on WsGT expression in stem tissue after increase of methyl jasmonate concentration. Even in 50 µM stress, WsGT expression was 20-fold at 4 hour and 6 hour incubation (Figure 5.B.9). But in 100 µM stress, WsGT expression was 13-fold at 4 hour incubation (Figure 5.B.12), which is lower than that at 50 µM stress.

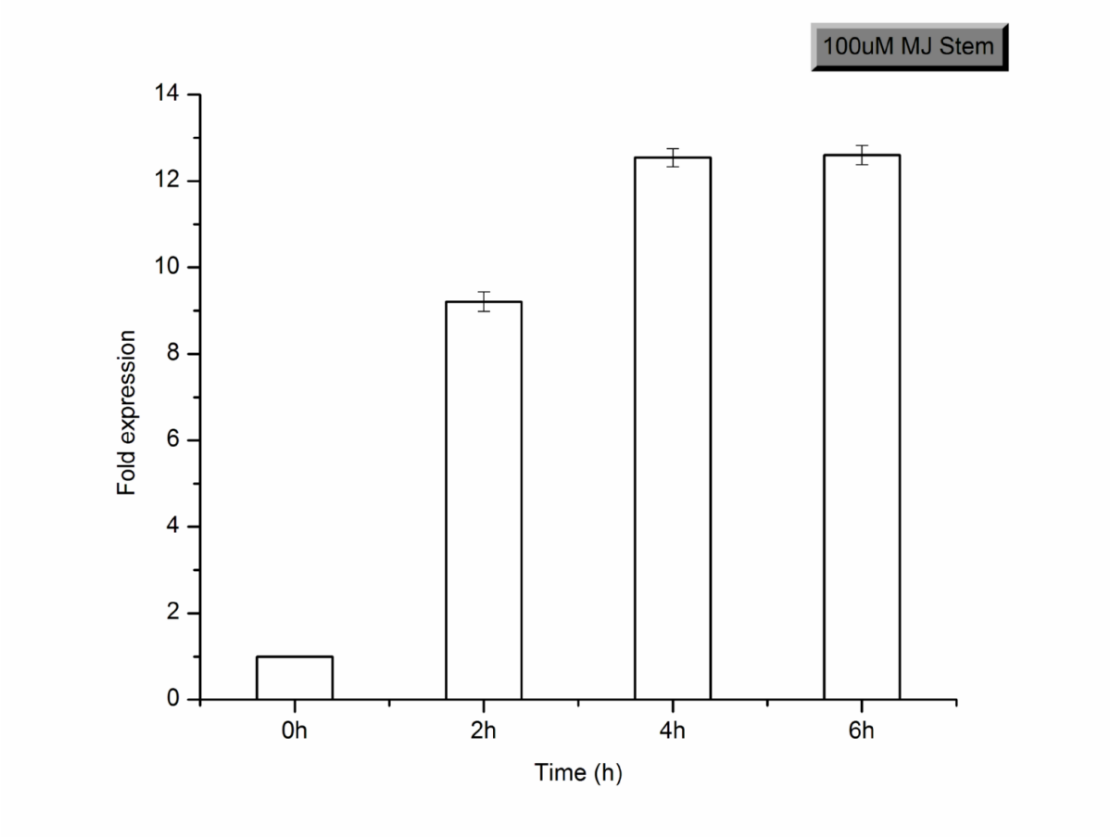


Figure 5.B.12: Relative expression in terms of mean fold expression of WsGT gene in stem tissue under 100 µM methyl jasmonate stress at 2 h, 4 h, and 6 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.14 Real time PCR analysis for tissue specific expression of WsGT gene in root tissue in the presence of methyl jasmonate, 100 µM

W. somnifera stem tissue was subjected to treatment with 100 µM methyl jasmonate for 2 h, 4 h and 6 h. Root tissue always showed less expression of WsGT gene, even in without stress or in stress conditions. Here in increased concentration of methyl jasmonate (100 µM), WsGT expression was 5-fold at 2 hour and increased 12-fold at 4 hour and 6 hour incubation (Figure 5.B.13). So even at higher concentration of methyl jasmonate, there is not much effect on WsGT transcript accumulation.

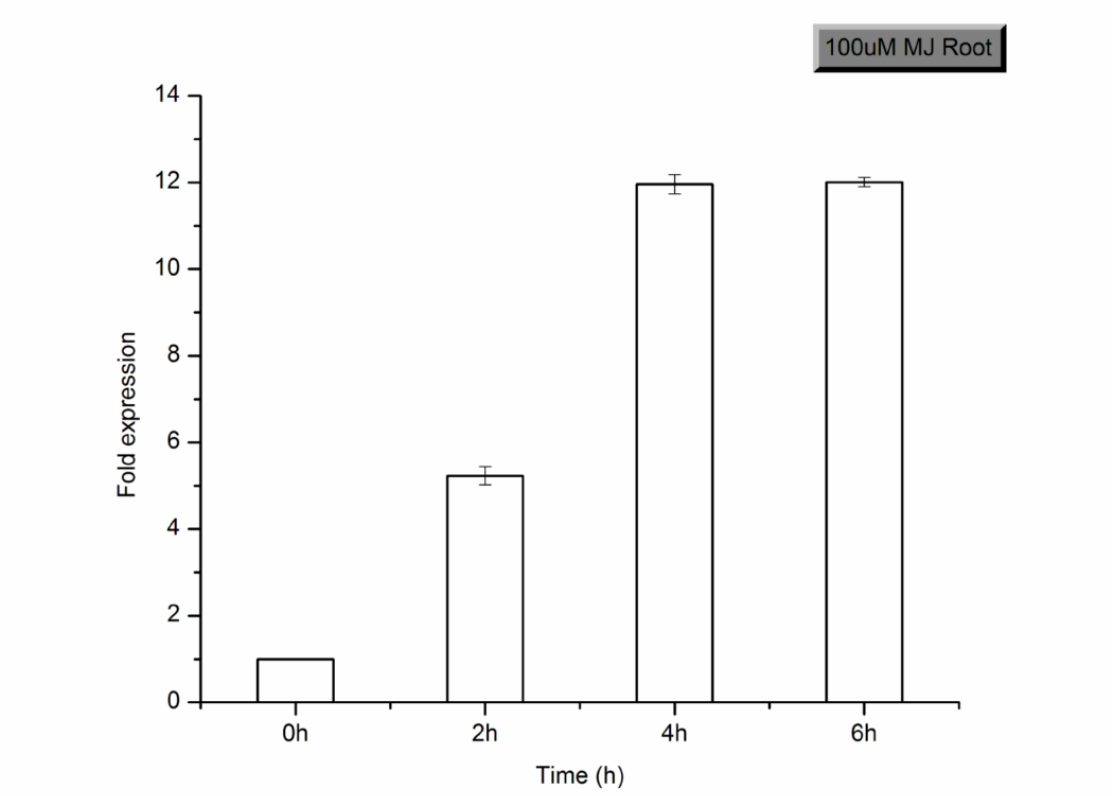


Figure 5.B.13: Relative expression in terms of mean fold expression of WsGT gene in root tissue under 100 µM methyl jasmonate stress at 2 h, 4 h, and 6 h with respect to control plant. All values are plotted with standard deviation taken into account

5.4.15 Real time PCR analysis of WsGT gene in leaf tissue by shifting from 25 ºC to 42 ºC

W. somnifera leaf tissue was subjected to heat shock by shifting the leaves from 25 ºC to 42 ºC and WsGT expression was analyzed at 15 min, 30 min, 45 min, 60 min, 75 min and 90 min. Initially WsGT expression increased to 13-fold till 15 min and then decreased transcript level was observed within 30 min. The level of WsGT transcript declined from 15 min (12-fold) to 30 min (10-fold) and undetectable at 45 min (Figure 5.B.14). It shows due to heat stress WsGT transcript accumulation was declined at prolonged period.

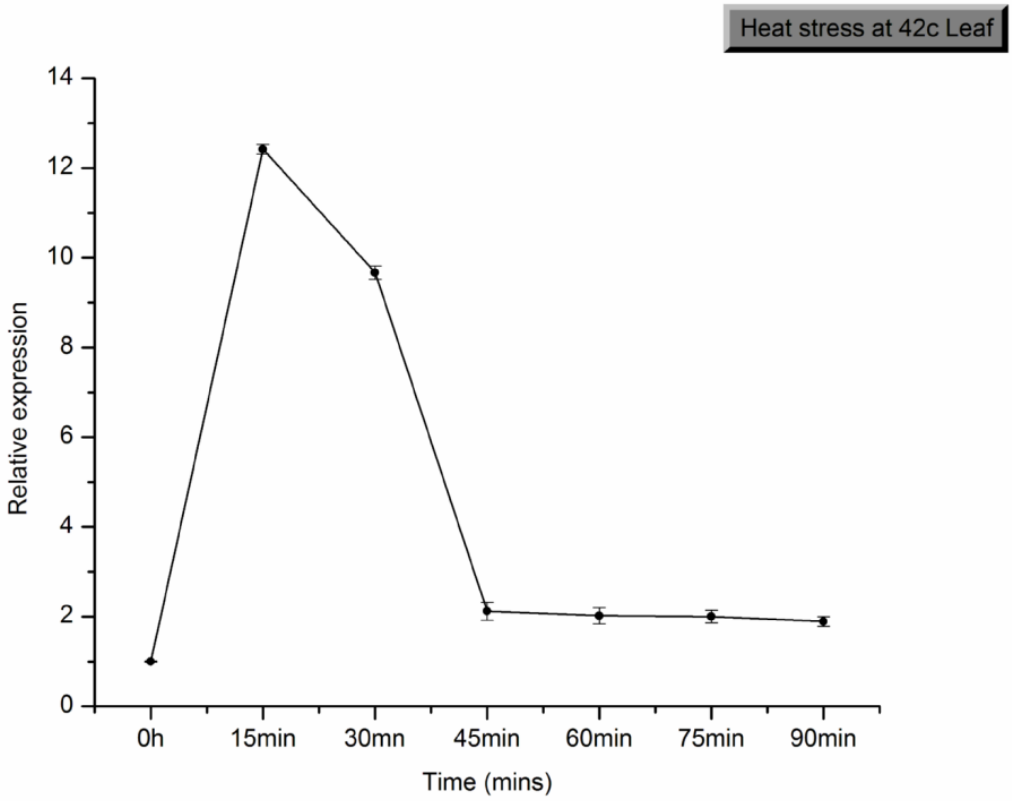


Figure 5.B.14: Relative expression in terms of mean fold expression of WsGT gene in leaf tissue subjected to heat shock by shifting from 25 ºC to 42 ºC with respect to control plant. All values are plotted with standard deviation taken into account

5.4.16 Real time PCR analysis of WsGT gene in stem tissue by shifting from 25 ºC to 42 ºC

W. somnifera stem tissue was subjected to heat shock by shifting the stem from 25 ºC to 42 ºC and WsGT expression was analyzed at 15 min, 30 min, 45 min amd 60 min. The expression of WsGT gene in stem tissue shows increase at 15 min (8-fold) to 30 min (10-fold) and then decline at 45 min and undetectable at 60 min (Figure 5.B.15).

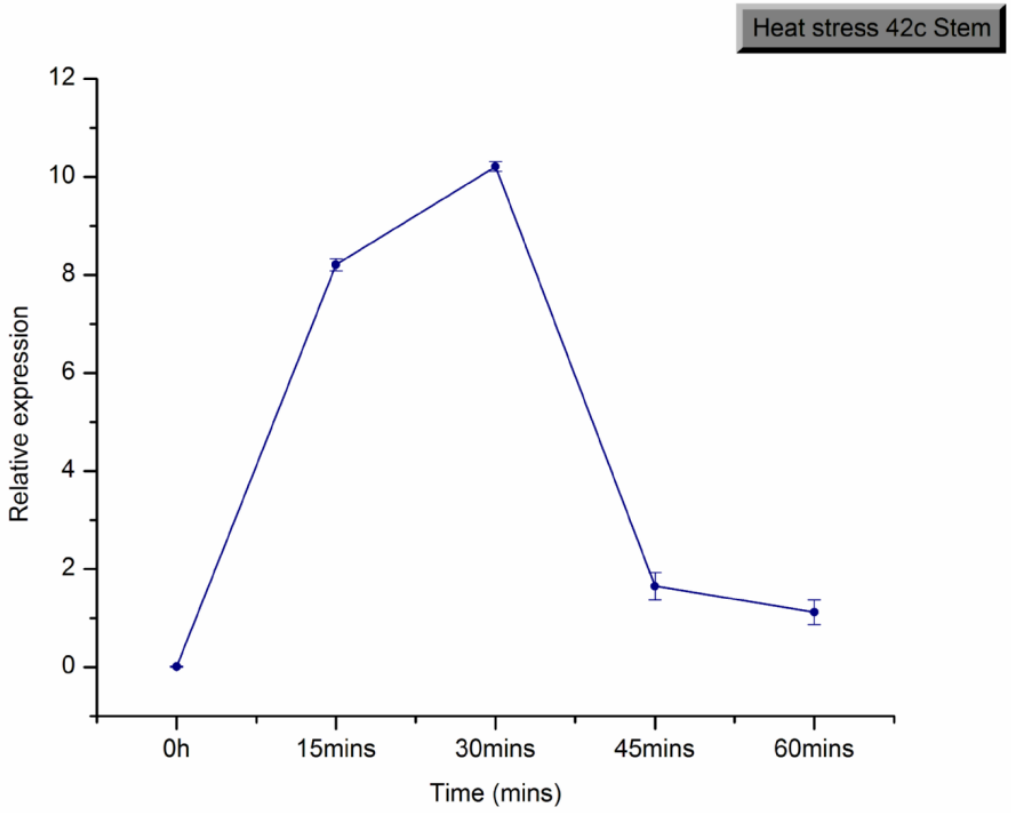


Figure 5.B.15: Relative expression in terms of mean fold expression of WsGT gene in stem tissue subjected to heat shock by shifting from 25 ºC to 42 ºC with respect to control plant. All values are plotted with standard deviation taken into account

5.4.17 Real time PCR analysis of WsGT gene in root tissue by shifting from 25 ºC to 42 ºC

W. somnifera root tissue was subjected to heat shock by shifting the roots from 25 ºC to 42 ºC and WsGT expression was analyzed at 15 min, 30 min, 45 min amd 60 min. During heat shock root tissue shows little increase of WsGT gene expression at 15 min (6-fold). After 30 min WsGT expression declined sharply and reached to undetectable at 45 to 60 min (Figure 5.B.16).

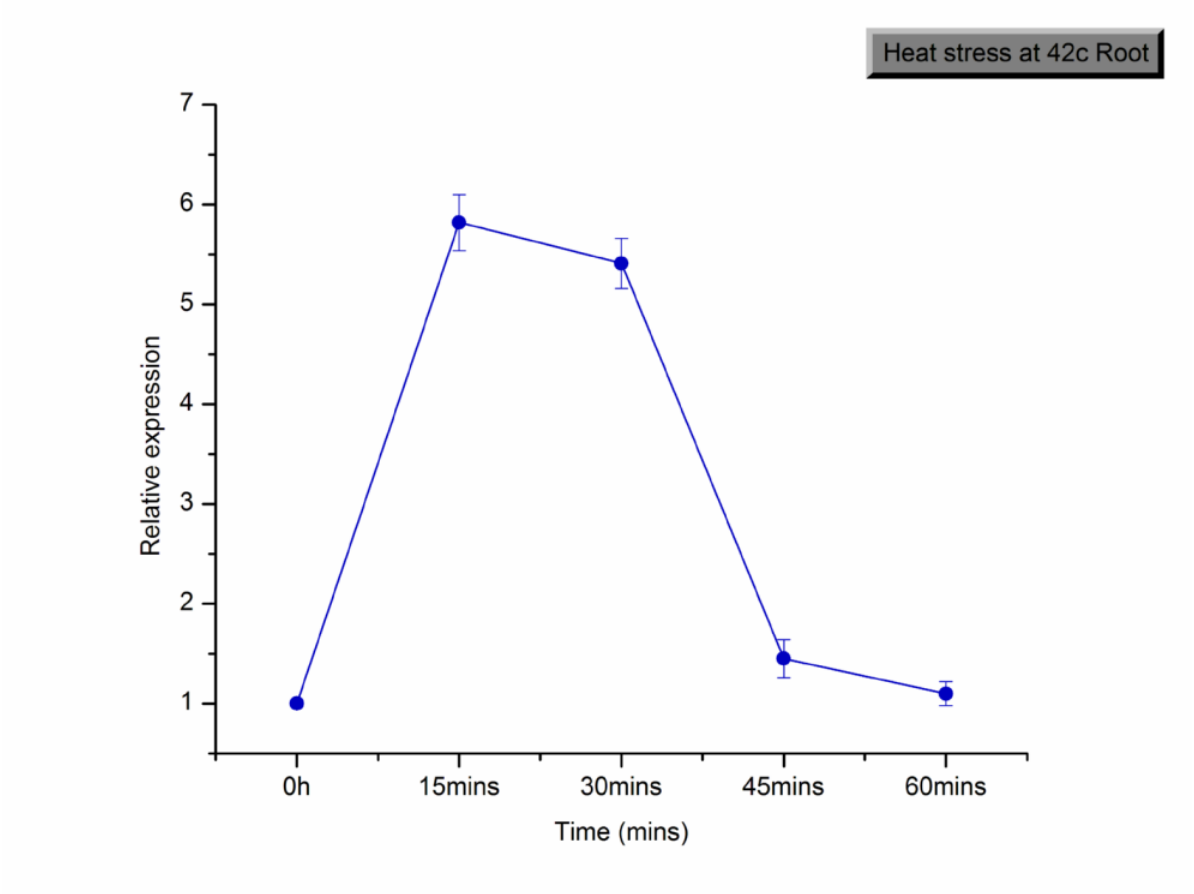


Figure 5.B.16: Relative expression in terms of mean fold expression of WsGT gene in root tissue subjected to heat shock by shifting from 25 ºC to 42 ºC with respect to control plant. All values are plotted with standard deviation taken into account

5.5 Conclusion

- Tissue specific expression of WsGT transcript was analyzed using real time PCR. In this study leaves (young and mature), stem and roots were analyzed for relative expression of WsGT gene.

- The distribution of WsGT transcript in *W. somnifera*, as determined by quantitative PCR, showed higher expression in young leaves than mature leaves. Roots showed very low level of expression.

- The expression of WsGT transcript was analyzed at different stress conditions (salicylic acid, methyl jasmonate and heat shock) following different concentrations. Expression of the WsGT transcript in the leaves of *W. somnifera* was enhanced following the application of salicylic acid.

- Over all the salicylic acid stress on leaves, stem and root showed greater enhancement of WsGT transcript accumulation, as compared with methyl jasmonate treatment.

- When leaf, stem and root tissues were subjected to heat shock at 42 ºC, expression was initially increased for 15 min and then it decreased rapidly.

- These results suggest the functional role of the WsGT enzyme in abiotic stress.

Summary

Summary

Withania somnifera (Ashwagandha) belongs to the family Solanaceae. The chemistry of *Withania* species has been extensively studied and several groups of chemical constituents such as steroidal lactones, alkaloids, flavonoids, tannin and many more have been identified, extracted and isolated. *Withania* is known for its medicinal use which is mainly due to the presence of various glycosides in the plant. Plant derived glucosides have attracted great attention due to their widespread applications. This class of products is difficult to isolate/synthesize in pure form because of the resulting low yields. The isolation and purification of such glucosides from plant sources is tedious and results in low yields. Thus, simple approaches for generation of such glucosides would be highly beneficial. Functional properties, such as solubility, physicochemical stability, bioactivity, pharmacokinetics and cellular localization of natural products, such as flavonoids, are greatly affected by glycosylation.

Glycosylation of plant secondary products, such as flavonoids, coumarins, terpenoids, and cyanohydrins is generally catalyzed by plant secondary product glycosyltransferases, which are family-1 glycosyltransferases.

These glycosides (produced by an enzyme called as glycosyltransferase) are present in the plant in very low quantity so an alternative to increase the yield of such medicinally important glycosides was to isolate the enzyme, glycosyltransferase and over express it in suitable expression system. As a first step two cDNAs of glycosyltransferase genes from *W. somnifera* were isolated and characterized. PCR based approach was used to fish out two cDNA GT gene clones, designated as WsGT (FJ560880) and WsSAPGT (FJ560881). Banding pattern in Southern hybridization suggested that at least 2 copies of *WsGT* gene are present in *Withania* genome. The sequence analysis revealed an Open Reading Frame (ORF) of WsGT to be 1371 and WsSAPGT to be 1260 bp (Partial fragment). Putative polyadenylation sites and poly A tail were identified in the 3` UTR. The predicted molecular weight and pI for WsGT and WsSAPGT were estimated to be 52 kDa / 5.32 and 46 kDa / 5.9, respectively. Deduced amino acid sequence of WsGT and WsSAPGT showed 84% and 72% identity at amino acid level with the GTs from *Lycium barbarum* and *Solanum aculeatissimum* respectively. Deduced amino acid sequences of cDNAs from *W. Somnifera* contain consensus sequences for PSPG

box which is conserved in whole of the plant GT family. Phylogenetic analysis of WsGT and WsSAPGT deduced amino acid sequences was done using MEGA 4 software program. The results show that WsGT is evolutionarily most similar to *Lycium barbarum* (AB360617) (UDP-glucose: glucosyltransferase), *Lycopersicon esculentum* (AJ889012) (UDP-xylose phenolic glycosyltransferase) and *Nicotiana tabacum* (AAF61647) (UDP-glucose: salicylic acid glucosyltransferase). Whereas WsSAPGT is evolutionary similar to *Solanum aculeatissimum* (BAD89042 and AB182385), UDP-glucose glucosyltransferase and *Solanum tuberosum* (ABB29874 and AAB48444) UDP-glucose:solanidine glucosyltransferase.

Both WsGT and WsSAPGT genes were directionally cloned into pET30b (+) expression vector system. Recombinant WsGT and WsSAPGT proteins were standardized for its over-expression and purification. Purified WsGT protein was used to raise polyclonal antibodies in New Zealand white rabbit. The candidate WsGT protein was detected in *W. somnifera* tissue by Western blotting. The active WsGT protein was purified using FPLC system. To perform enzyme assay and kinetics reaction purified proteins was used. WsGT shows activity with UDP-Glucsoe as sugar donor and apigenin, naringenin, genistein, luteolin and diadzein as a acceptor molecules. The glycosylated product was checked using LC-MS and HPLC system. Standard glycosides are used for conformation and it shows that WsGT glycosylate flavonoid substrates at 7^{th} position of substrates. Enzyme kinetics of WsGT shows more substrate specificity towards diadzein.

The active WsSAPGT protein was purified and enzyme assay was performed using different sterol sapogenins. WsSAPGT shows activity with only steroidal sapogenin digitoxigenin.

Tissue specific expression of WsGT was performed using real-time PCR analysis under normal and stress induced conditions (methyl jasmonate, salicylic acid and heat shock treatments). Real-time PCR analysis showed differential expression at transcript level with maximum expression in young leaf tissue in control plant comparing with mature leaf, stem and root. Methyl jasmonate and salicylic acid stress was used at different concentrations at different time intervels. Leaf, stem and root tissues were used for stress analysis. During

different time intervels, it shows WsGT gene expression was gradually increased and declined at different concentrations of stresses.

Publications

Publications

Research Papers under preparation

1. **R. J. Santosh Kumar**, Ruby**,** Somesh singh, R.K. Vishwakarma, Dr. B.M. Khan* (2011). "Molecular cloning, characterization and expression studies of glycosyltransferase gene from *Withania somnifera*" (Manuscript under preparation).

2. **R. J. Santosh Kumar,** Krunal patel, Disha aswar, Dr. B. M. Khan* (2011). "Molecular modeling and docking studies of *Withania somnifera* flavanol 7-O-glycosyltransferase" (Manuscript under preparation).

3. **R. J. Santosh Kumar,** Krunal patel, Somesh singh, Dr. B. M. Khan* (2011). "Isolation, cloning and characterization of saponin glycosyltransferases from *Withania somnifera*" (Manuscript under preparation).

4. Ruby, **R. J. Santosh Kumar**, R.K. Vishwakarma, Somesh singh, Dr. B.M. Khan* (2011). "Molecular cloning, characterization and expression analysis of glycosyltransferase gene from *Bacopa monniera*" (Manuscript under preparation).

5. Somesh singh, **R. J. Santosh Kumar**, R.K. Vishwakarma, Ruby, Dr. B.M. Khan* (2010). "Isolation, cloning, and characterization of Flavanoid Glucosyltransferase gene from *Withania somnifera*" (Manuscript under preparation).

6. R.K. Vishwakarma, Ruby**,** Somesh singh, **R. J. Santosh Kumar**, Dr. B.M. Khan* (2011). "Isolation, cloning, characterization and expression studies of Acetyl CoA transferase gene from *Bacopa monniera*" (Manuscript under preparation).

Abstracts / Proceedings published

1. "Isolation and Characterization of novel glycosyltransferases from *Withania somnifera*". **R. J. Santosh Kumar**, Ruby, Dr. B.M. Khan*. Abstract of the poster presented in the National Symposium on Plant Propagation, Conservation, Modification and Characterization organized at IHBT, Palampur (2009).

2. "Secondary metabolic flavonoids glycosyltransferases from *Withania somnifera*". **R. J. Santosh Kumar**, Ruby, Somesh singh, R.K. Vishwakarma, Dr. B.M. Khan*. Abstract of the poster presented in the International Symposium on Current status and Opportunities in Aromatic and Medicinal Plants held at CIMAP (2010).

3. Steriospecific glycosylation of medicinally important compounds. Somesh singh, **R. J. Santosh Kumar**, R. K. VIshwakarma, Ruby, Dr. B. M. Khan*. Abstract of the poster presented in the 1st Chemical Research Society of India (CRSI) Zonal meeting held at National Chemical Laboratory (NCL, 2011).

4. "Two different glycosyltransferases from *Bacopa monniera* involved in flavanoid biosynthesis". Ruby, **R. J. Santosh Kumar**, R.K. Vishwakarma, Somesh singh, Dr. B.M. Khan*. Abstract of the poster presented in National Symposium on recent advances in plant tissue culture and biotechnological researches in India (2011).

5. "Glycosyltransferase involved in flavonol glycoside biosynthesis in *Withania somnifera*". S. Singh, **R. J. Santosh Kumar**, R.K. Vishwakarma, Ruby, Dr. B.M. Khan*. International symposium on Aromatic and medicinal plants (AROMED), CIMAP, India. (2010).

6. "Role of squalene synthesis in withanolide biosynthesis". P. Sharma, **R. J. Santosh Kumar**, R.K. Vishwakarma, Ruby, S. Singh and Dr. B.M. Khan*. International symposium on Aromatic and medicinal plants (AROMED), CIMAP, India. (2010).

7. “Molecular study of glycosylation of polyphenolic compounds from *Withania somnifera* and its applications for human health”. Somesh singh, **R. J. Santosh Kumar**, R.K. Vishwakarma, Ruby, Dr. B.M. Khan*. (INSA 2009).

8. "An approach for genetic modification of *Leucaena leucocephala* for Eco-friendly pulp and paper production". Manish Arha, Gupta K. Sushim, Mohd Noor, Yadav Arun, Srivastava Sameer, K. Pallavi, O U. Abhilash, **R. J. Santosh Kumar**, Malini Kaul, Jay Kumar, Azfar Quraishi, Ulfat Iqbal, Ruby, Khan B. M., Rawal S. K*. Abstract of the Poster presented in the International symposium on Frontiers in Genetic and Biotechnology- Retrospect and Prospect at Osmania University (2006).

9. "Isolation, Cloning and Characterization of Caffeoty CoA 3-) methyl transferases (CCoAOMT) from *Leucaena leucocephala*". Manish Arha, Gupta K. Sushim, Mohd Noor, Yadav Arun, Srivastava Sameer, K. Pallavi, O U. Abhilash, **R. J. Santosh Kumar**, Malini Kaul, Jay Kumar, Azfar Quraishi, Ulfat Iqbal, Ruby, Khan B. M., Rawal S. K*. Absract of the Poser presented in the National Symposium on Plant Biotechnology: New frontiers held at the CIMAP (2005).

10. “Molecular approaches for production of pharmacologically active plant glycosides, A valuable strategy for finding new lead compounds”. Ruby, **R. J. Santosh Kumar**, R.K. Vishwakarma, Somesh singh, Dr. B.M. Khan*. Abstract of the poster presented in INSA Platinum jubilee international symposium on Research in Molecular medicine based on natural resources and traditional knowledge (2009).

11. “Characterization of Betaine aldehyde dehydrogenase in *Pandanus amaryllifolius* Roxb.; a plant with higher basmati aroma (2 acetyl-1-Pyrroline) contents”. Trupti Kad, Kantilal Wakte, **R. J. Santosh Kumar**, Altafhusain Nadaf* and Dr. B. M. Khan. Abstract of the poster presented in international symposium on Aromatic and medicinal plants (AROMED), CIMAP, India (2010).

References

References

Abraham A, Kirson I, Glotter E and Lavie D (1968) A chemotaxonomic study of *Withania somnifera* (L.) Dun. Phytochemistry 7, 957-962.

Akio Noguchi, Manabu Horikawa, Yuko Fukui, Masako Fukuchi-Mizutani, Asako luchi-Okada, Masaji Ishiguro, Yoshinobu Kiso, Toru Nakayama, and Eiichiro ono (2009) Local Differentiation of sugar donor speciricity of Flavonoid Glycosyltransferase in Lamiales. The Plant Cell Vol. 21: 1556-1572.

Alfonso D and Kapetanidis I (1994) Withanolides from Iochroma gesnerioides. Phytochemistry 36, 179-183.

Anjali A. Kulkarni, S.R. Thengane and K.V. Krishnamurthy (2000) Direct shoot regeneration from node, internode, hypocotyl and embryo explants of *Withania somnifera*. Plant Cell, Tissue and Organ Culture 62: 203–209.

Aram Chang, Shanteri Singh, George N Phillips Jr. and Jon S Thorson (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Current Opinion in Biotechnology, 22:1-9.

Arnheim N, Erlich H (1992) Polymerase Chain Reaction Strategy. Annu Rev Biochem 61:131-156.

Arun Kumar M, K Kaul, M K Bhan, Punit K, Khanna K, A Suri (2007) Morphological and chemical variation in 25 collections of the Indian medicinal plant, *Withania somnifera* (L.) Dunal (Solanaceae). Genet Resour Crop Evol 54:655–660.

Asthana R, Raina M K (1989) Pharmacology of *W. somnifera* (L.) Dunal - a review. Indian Drugs 26, 199-205.

Atal C k, Gupta O P, Raghunathan K, Dhar K L (1975) Pharmacognosy and phytochemistry of *W. somnifera* (Linn.) Dunal (Ashwagandha). Central Council for Research in Indian Medicine and Homeopathy, New Delhi p 6-18.

Atta-ur-Rahman, Abbas S, Dur-e-Shawar, Jamal A S, Choudhary M I (1993) New withanolides from *Withania* spp. J. Nat. Prod. 56, 1000-1006.

Atta-ur-Rahman, Jamal A S, Choudary M I, Asif I (1991) Two withanolides from *Withania somnifera*. Phytochemistry 30, 3824-3825.

Baker J T, Borris R P, Carté B, Cordell G A, Soejarto D D, Cragg G M, Gupta M P, Iwu, M M, Madulid D R and Tyler V E (1995) Natural products drug discovery and development: new perspective on international collaboration. J. Nat. Prod. 58 (9):1325-1357.

Bandyopadhyay M, Jha S and Tepfer D (2007) Changes in morphological phenotypes and withanolide composition of Ri-transformed roots of *Withania somnifera*. Plant cell Rep. 26, 599-609.

Begum V H, Sadique J (1988) Long term effect of herbal drug *Withania somnifera* on adjuvantinduced arthritis in rats. Indian J. Exp. Biol. 26, 877-882.

Bennet R N and Wallsgrove R M (1994) Secondary metabolites in plant defense mechanisms. New Phytologist 127: 617–633.

Bhaskara Reddy Madina, Lokendra kumar Sharma, Pankaj Chaturvedi, Rajender singh sangwan, Rakesh Tuli (2007) Purification and characterization of a novel glucosyltransferase specific to 27β-hydroxy steroidal lactones from *Withania somnifera* and its role in stress responses. Biochimica et Biophysica Acta 1774, 1199-1207.

Bhattacharya A, M Ramanathan, S Ghosal and S K Bhattacharya (2000) Effect of *Withania somnifera* Glycowithanolides on Iron-induced Hepatotoxicity in Rats. Phytother. Res. 14, 568-570.

Bhattacharya S K, Satyan K S, Chakrabarti A (1997) Effect of Trasina, an Ayurvedic herbal formulation, on pancreatic islet superoxide dismutase activity in hyperglycaemic rats. Indian J Exp Biol; 35:297-299.

Blumberg D D (1987) Creating a ribonuclease-free environment. Meth. Enzymol 152: 20- 24.

Bourne Y, Henrissat B (2001) Glycoside hydrolases and glycosyltransferases families and functional modules. Curr. Opin. Struct. Bio. l11:593-600.

Bradford M M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Analytical Biochemistry 72, 248–254.

Byoung Seok Hong, Jeong Ho Kim, Na Yeon Kim, Bong-Gyu Kim, Youhoon Chong and Joong-Hoon Ahn (2007) Characerization of Uridine-Diphosphate dependent Flavonoid glucosyltransferse from *Oryza sativa.* Journal of Biochemistry and Molecular Biology, Vol. 40, No. 6, 870-874.

Campbell J A, Davies G J, Bulone V, Henrissat B (1997) A classification of nucleotide diphospho-sugar Glycosyltransferases based on amino acid sequence similarities. Biochem J.326 (Pt3): 929-939.

Chandra Prakash Kala, Pitamber Prasad Dhyani, Bikram Singh Sajwan (2006) Developing the medicinal plants sector in northern India: challenges and opportunities Journal of Ethnobiology and Ethnomedicine 2:32.

Choudary M I, Abbas S, Jamal A S and Atta-ur-Rahman (1996) *Withania somnifera* - A source of exotic withanolides. Heterocycles 42, 555-563.

Chumakov K M (1994) Reverse transcriptase can inhibit PCR and stimulate primer dimer formation. PCR Methods Appl 4: 62-64.

Claire M M, Gachon, Mathilde Langlois_Meurinne and Patrick Saindrenan (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. TRENDS in plant science vol.10 No.11 November.

Collin H A (2001) Secondary product formation in plant tissue cultures. Plant Growth Regulation 34**:** 119–134.

Coutinho P M, Deleury E, Davies G J, Henrissat B (2003) An evolving hierarchical family classication for glycosyltransferases. J. Mol. Biol. 328:307-317.

Cragg G M and Newman D J (2005) Biodiversity: A continuing source of novel drug leads. Pure Appl. Chem 77 (1):7-24.

Davis L, Kuttan G (1998) Suppressive effect of cyclophosphamide-induced toxicity by *Withania somnifera* extract in mice. J Ethnopharmacol; 62:209-214.

De Silva T (1997) Industrial utilization of medicinal plants in developingcountries. pp. 38–48. In: Bodeker G, Bhat KKS, Burley J and Vantomme P. (eds), Medicinal Plants for Forest Conservation and Healthcare. Non-Wood Forest Products No. 11, FAO,Rome, Italy.

Devi P U (1996) *Withania somnifera* dunal (ashwagandha): potential plant source of a promising drug for cancer chemotherapy and radiosensitization. Indian J Exp Biol, 34:927-932.

Devi P U, Sharada A C, Solomon F E, Kamath M S (1992) *In vivo* growth inhibitory effect of *Withania somnifera* (Ashwagandha) on a transplantable mouse tumor, Sarcoma 180. Indian J Exp Biol, 30:169-172.

Dianna Bowles, Judith isayenkova, Eng-kiat Lim and Brigitte Poppenberger (2005) Glycosyltransferases: managers of small molecules. Current Opinion in Plant Biology, 8:254-263.

Dixon R A (2001) Natural products and plant disease resistance. Nature 411, 843–847.

Dodson E, Harding M M, Hodgkin D C, Rossmann M G (1966) The crystal structure of insulin.3. Evidence for a 2-fold axis in rhombohedral zinc insulin. J Mol Biol. 16:227-241.

Dörnenburg H and Knorr D (1996) Generation of colors and flavors in plant cell and tissue cultures. Critical Reviews in Plant Science 15: 141–168.

Eng-Kiat Lim (2005) Plant Glycosyltransferases: Their Potential as Novel Biocatalysts. Chem. Eur. J., 11, 5486-5494.

Farnsworth N, Akerele A O, Bingel A S, Soejarto D D, Guo Z (1985) Bull. WHO 63:965-981.

Feinbeng A P, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13.

Freeman W M, Walker S J, Vrana K E (1999) Quantitative RT-PCR: pitfalls and potential.Biotechniques 26: 112-122, 124-125.

Gachon, C M, Langlois-Meurinne, M, and Saindrenan, P (2005) Plant secondary metabolism glycosyltransferases: The emerging functional analysis. Trends Plant Sci. 10: 542-549.

Ghosal S, Lal J, Srivastava R, *et al*., (1989) Immunomodulatory and CNS effects of sitoindosides IX and X, two new glycowithanolides from *Withania somnifera*. Phytotherapy Res; 3:201-206.

Glombitza S, P H Dubuis, O Thulke, G Welz, (2004) Crosstalk and differential response to abiotic and biotic stresses reflected at the transcriptional level of effector genes from secondary metabolism. Plant Mol. Biol 54, 817-835.

Glotter E (1991) Withanolides and related ergostane-type steroids. Nat. Prod. Rep. 8, 415-440.

Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S (1989) HIV-1 isolates are rapidly evolving quasi species: evidence for viral mixtures and preferred nucleotide substitutions. J Acquired Immunol Defic Syndr 2: 344-352.

Gupta A P, Verma R K, Misra H O and Gupta M M (1996) Quantitative determination of withaferin A in different plant parts of *Withania somnifera* by TLC densitometry. J. Med. Arom. Plant Sci. 18, 788-790.

Gupta G L and Rana A C (2007) *Withania somnifera* (Ashwagandha): A Review. Pharmacog. Rev.1, 129-136.

Heid C A, Stevens J, Livak K J, Williams P M (1996) Real time quantitative PCR. Genome Res 6: 986-994.

Higuchi R, Dollinger G, Walsh P S, and Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. Biotechnology 10: 413–417.

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 11: 1026–1030.

Hisashi Matsuda, Toshiyuki Murakami, Akinobu Kishi and Masayuki Yoshikawa (2001) Structures of Withanosides I, II, III, IV, V, VI, and VII, New Withanolide Glycosides, from the Roots of Indian *Withania somnifera* DUNAL. And Inhibitory Activity for Tachyphylaxis to Clonidine in Isolated Guinea-Pig Ileum. Bioorganic & Medicinal Chemistry 9: 1499-1507.

Hu Y and Walker S (2002) Remarkable structural similarities between diverse glycosyltransferases. Chem. Biol. 9, 1287–1296.

Huggett J, Dheda K, Bustin S, Zumla A (2005) Realtime RT-PCR normalisation; strategies and considerations. Genes Immun 6:284.

Hughes J and Hughes M A (1994) Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. DNA Sequence, 5, 41–49.

Jae Hyung Ko, Bong Gyu Kim, Hor-Gil Hur, Yoongho Lim, Joong-Hoon Ahn (2006) Molecular cloning, expression and characterization of a glycosyltransferase from rice. Plant Cell Rep 25: 741-746.

Jayaprakasam B, Strasburg G A and Nair M G (2004) Potent lipid peroxidation inhibitors from *Withania somnifera* fruits. Tetrahedron, 60, 3109-3121.

Joachim Arend, Heribert Warzecha, Tobias Hefner, Joachim Stockigt (2001) Utilizing genetically engineered bacteria to produce plant specific glucosides. Biotechnology and bioengineering,Vol. 76, No. 2.

Joe Ross, Yi Li, Eng-Kiat Lim and Dianna J Bowles 2001 Higher plant glycosyltransferases. Genome Biology, 2(2).

Johri S, Jamwal U, Rasool S, Kumar A, Verma V and Qazi GN (2005) Purification and characterization of peroxidases from *Withania somnifera* (AGB 002) and their ability to oxidize IAA. Plant Sci. 169, 1014-1021.

Judith Hans1, Wolfgang Brandt and Thomas Vogt* (2004) Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-O-glucosyltransferase from *Dorotheanthus bellidiformis*. The Plant journal. 39, 319.-333.

Jun WANG, Bingkai HOU (2009) Glycosyltransferases: key players involved in the modification of Plant secondary metabolites. Front. Biol. China. 4(1): 39-46.

Kalaiselvi senthil, Neha G. Wasnik, Yu-jin Kim, Deok Chun Yang (2010) Generation and analysis of expressed sequence tags from leaf and root of *Withania somnifera* (Ashwagandha). Mol Biol Rep. 37: 893-902.

Kapoor L D (2001) Handbook of Ayurvedic Medicinal Plants; CRC Press: London, UK. pp. 337-338.

Kapitonov D and Yu R K (1999) Conserved domains of glycosyltransferases. Glycobiology, 9, 961–978.

Karim Kherraz, Khaled Kherraz and Abbelkrim Kameli (2011) Homology modeling of Ferredoxin-nitrite reductase from Arabidopsis thaliania. Bioinformation 6(3): 115-119.

Kaul K N (1957) On the origin, distribution and cultivation of Ashwagandha, the so called *Withania somnifera* Dunal, of Indian literature on materia medica. Pharmaceut and Drug Committee, Symp on the Utilization of the Indian Medicinal Plants, Lucknow, 12–14 Oct (pp 7–8).

Kikuchi N, Kwon Y D, Gotoh M, Narimatsu H (2003) Comparison of glycosyltransferases

families using the profile hidden Markov model. Biochem Biophys Res Commun. 310:574-579.

Kimmel A R, Berger S L (1987) Preparation of cDNA and the generation of cDNA libraries:Overview. Meth Enzymol 152: 307-316.

Kirsi-Marja Oksman-Caldentey and Dirk Inze (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. TRENDS in Plant Science Vol.9, No.9.

Kirson I, Abraham A, Lavie D (2007) Analysis of hybrids of *Withania somnifera* L. (Dun.) Chemotypes III (Israel) by Indian I (Delhi). *Israel J. Chem. 16*, 20.

Kirson I, Glotter E, Lavie D and Abraham A (1971) Constituents of *Withania somnifera* Dun. XII. The withanolides of an Indian chemotype. J. Chem. Soc. C, 2032-2044.

Kramer CM *et al.,* (2003) Cloning and regiospecificity studies of two flavonoids glucosyltransferase gene from adzuki bean seedlings. Plant Physiol 129, 1285-1295.

Kubo A *et al.,* (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single poing mutation. Arch. Biochem. Biophys. 429, 198-203.

Kulkarni A A, Thengane S R and Krishnamurthy K V (1996) Direct in vitro regeneration of leaf explants of *Withanis somnifera* (L.) Dun. Plant Science 119: 163-168.

Kuppurajan K *et al,*. J Res Ayu Sid 1, 247, 1980. As cited in Bone K. Clinical Applications of Ayurvedic and Chinese Herbs. Monographs for the Western Herbal Practitioner. Australia: Phytotherapy Press; 1996:137-141.

Kwok S, Kellogg D E, McKinney N, Spasic D, Goda L (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Res 18: 999-1005.

Kyte J and Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Laemmli U K (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Lakshmi-Chandra Mishra M D, PhD, Betsy B Singh, PhD, Simon Dagenais, B A (2000) Alternative Medicine Review, Volume 5, Number 4.

Lancet T (1994) Pharmaceuticals from plants: great potential, few funds. The Lancet 343: 1513–1515.

Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1998) PROCHECK: a program to check the stero chemical quality of protein structure. J Appl Cryst 26:283-291.

Lavie D, Glotter E and Shvo Y (1965) Constituents of *Withania somnifera* Dun. Part IV The structure of withaferin-A. J. Am. Chem. Soc. 30, 7517-7531.

Lee L G, Connell C R, Bloch W (1993) Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Res 21: 3761-3766.

Leet J E, Hussain S F, Minard R D and Shamma M (1982) Heterocycles. 19, 2355-2360.

Lim E K, Jackson R G, Bowles D J (2005b) Identification and characterization of Arabidopsis glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. FEBS Lett, 579: 2802-2806.

Liu J, Mushegian A (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci. 12:1418-1431.

Livak K J, Flood S J, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4: 357-362.

Lodhi M A, Guang-Ning Ye, Norman F W, Bruce I R (1994) A simple and efficient method for DNA extraction from grapevine cultivars, *Vitis* species and *Ampelopsis*. Plant Mol Biol Reporter 12: 6-13.

Malhotra C L, Mehta V L, Das P K, Dhalla N S (1965) Studies on *Withania*-ashwagandha, Kaul. V. The effect of total alkaloids (ashwagandholine) on the central nervous system. Indian J Physiol Pharmacol; 9:127-136.

Malhotra C L, Mehta V L, Prasad K, Das P K (1965) Studies on *Withania* ashwagandha, Kaul. IV. The effect of total alkaloids on the smooth muscles. Indian J Physiol Pharmacol; 9:9-15.

Marderosion A D (2001) The Review of Natural Products, Facts and Comparisons; St. Louis, MI, USA, pp. 630-632.

Maruyama K, and Sugano S (1994) Oligo-Capping: A Simple Method to replace the cap structure of Eukaryotic mRNAs with Oligoribonucleotides. Gene 138: 171 -174.

Mohammad Hossein Mirjalili, Elisabeth Moyano, Mercedes Bonfill, Rosa M Cusido, Javier Palazón (2009) Steroidal Lactones from *Withania somnifera*, an Ancient Plant for Novel Medicine. Molecules, 14, 2373-2393.

Mulichak A M, Losey H C, Lu W, Wawrzak Z, Walsh C T, Garavito R M (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic

pathway. Proc. Natl. Acad. Sci. USA. 100:9238-9243.

Mullis K B (1990) The unusual origin of the polymerase chain reaction. Sci Am 262: 56– 61.

Mullis K B, Faloona F A (1987) Specific synthesis of DNA in vitro via a polymerase catalysed reaction. Methods Enzymol 255: 335-350.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant 15: 473-497.

Nittala, S S, Lavie, D (1981) Chemistry and genetics of withanolides in *Withania somnifera* hybrids. Phytochemistry. 20, 2741-2748.

Panda S, Kar A (1998) Changes in thyroid hormone concentrations after administration of ashwagandha root extract to adult male mice. J Pharm Pharmacol; 50:1065-1068.

Pichersky E, and Gang D R (2000) Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. Trends Plant Sci 5: 439- 445.

Poppenberger B, Berthiller F, Lucyshyn D, Sieberer T, Schuhmacher R, Krska R, Kuchler K, Gloss J, Luschnig C, Adam G (2003) Detoxification of the fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from Arabidopsis thaliana. J Biol Chem, 278: 47905-47914.

Prajapati N D, Purohit S S, Sharma A K and Kumar T (2003) A Handbook of Medicinal Plants. Agribios (India), 553 pp.

Raeymaekers L (2000) Basic principles of quantitative PCR. Mol Biotechnol 15: 115-122.

Raghunath Satpathy, Rashmiranjan Behera and Rajesh Ku Guru (2011) Homology modeling and molecular dynamics study of plant defensin DM-AMP1. J Biochem Tech (Article in press).

Rao M R, M C Palada, B N Becker (2004) Medicinal and aromatic plants in agro forestry systems. *Agroforestry Systems* 61**:** 107–122.

Reischl U, Wittwer C T, Cockerill F (2002) Rapid Cycle Real-time PCR: Methods and Applications; Microbiology and Food Analysis. New York: Springer-Verlag.

Roja G, Heble M R and Sipahimalani A T (1991) Tissue cultures of *Withania somnifera*: morphogenesis and withanolide synthesis. Phytotherapy Res. 5: 185-187.

Rosen M L, Edman M, Sjostrom M, Wieslander A (2004) Recognition of fold and sugar linkage for glycosyltransferases by multivariate sequence analysis. J. Biol. Chem. 279:38683-38692.

Rossmann M G, Moras D and Olsen K W (1974) Chemical and biological evolution of a nucleotide-binding protein. Nature, 250, 194–199.

Saha B K, Tian B, Bucy R P (2001) Quantitation of HIV-1 by real-time PCR with a unique fluorogenic probe. J Virol Methods 93: 33-42.

Saiki R, Scharf S, Faloona F, Mullis K B, Horn G T, Erlich H A, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350-1354.

Saiki R K, Gelfand D H, Stoffel S, Scharf S, Higuchi R, Horn GT, Mullis K B, Erlich H A (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406-425.

Sambrook J, Fritsch E F, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn.

Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Saritha K V, C V Naidu (2007) In vitro flowering of *Withania somnifera* Dunal. An important antitumor medicinal plant. Plant science 172, 847-851.

Scharf S J, Horn G T, Erlich H A (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 233: 1076-1078.

Schippmann U, Leaman D J and Cunnningham A B (2002) Impact of cultivation and gathering of medicinal plants on biodiversity: global trends and issues. In: Biodiversity and the Ecosystem Approach in Agriculture, Forestry and Fisheries. Ninth Regular Session of the Commission on Genetic Resources for Food and Agriculture. FAO, Rome, Italy, pp. 1–21.

Schliebs R, Liebmann A, Bhattacharya SK, *et al.*, (1997) Systemic administration of defined extracts from *Withania somnifera* (Indian Ginseng) and Shilajit differentially affects cholinergic but not glutamatergic and GABAergic markers in rat brain. Neurochem Int; 30:181-190.

Sharad A C, Solomon F E, Devi P U (1996) Antitumor and radiosensitizing effects of withaferin A on mouse Ehrlich ascites carcinoma in vivo. Acta Oncol; 35:95-100.

Sharada M A, Ahuja K A, Suri S P, Vij R K, Khajuria V, Verma and A Kumar (2007) Withanolide production by in vitro cultures of *Withania somnifera* and its association with differentiation. Biologia Plantarum 51 (1): 161-164.

Shiva M P (1996) Inventory of forestry resources for sustainable management and biodiversity conservation. New Delhi: Indus Publishing Company.

Singh N, Nath R, Lata A, *et al.*, (1982) *Withania somnifera* (ashwagandha), a rejuvenating herbal drug which enhances survival during stress (an adaptogen). Int J Crude Drug Res; 20:29-35.

Singh S and Kumar S (1998) *Withania somnifera*: The Indian Ginseng Ashwagandha; Central Institute of Medicinal and Aromatic Plants: Lucknow, India.

Somasundaram S, Sadique J, Subramoniam A (1983) *In vitro* absorption of [14C] leucine during inflammation and the effect of anti-inflammatory drugs in the jejunum of rats. Biochem Med; 29:259-264.

Somasundaram S, Sadique J, Subramoniam A (1983) Influence of extra-intestinal inflammation on the in vitro absorption of 14_C-glucose and the effects of anti-inflammatory drugs in the jejunum of rats. *Clin Exp Pharmacol Physiol*; 10:147-152.

Subaraju G V, Vanisree M, Rao C V, Sivaramakrishna C, Sridhar P, Jayaprakasam B, Nair MG (2006**)** Ashwagandhanolide, a bioactive dimeric thiowithanolide isolated from the roots of *Withania somnifera*. J. Nat. Prod, 69, 1790-1792.

Taiz L and Zeiger E (2006) Plant physiology, 4th edn, Ch.13, p 315-344.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596-1599.

Tan W, Wang K, Drake T J (2004) Molecular beacons. Curr Opin Chem Biol 8: 547-553.

Terry C F, Shanahan D J, Ballam L D, Harris N, McDowell D G, Parkes H C (2002) Real-time detection of genetically modified soya using Lightcycler and ABI 7700 platforms with TaqMan, Scorpion, and SYBR Green I chemistries. JAOAC Int 85: 938-944.

Todd J A, Bell J I, McDevitt H O (1987) HLA-DQ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature 329: 599-604.

Tohge T *et al.,* (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J. 42: 218-235.

Unligil U M, Rini J M (2000) Glycosyltransferase structure and mechanism. Curr. Opin.Struct Bio. l10:510-517.

Van Seters A P (1997) Forest based medicines in traditional and cosmopolitan health care. pp. 10–16. In: Bodeker G, Bhat KKS, Burley J and Vantomme P (eds), Medicinal Plants for Forest Conservation and Healthcare. Non-Wood Forest Products No.11, FAO, Rome, Italy.

Vander Velden V H, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen J J (2003) Detection of minimal residual disease in hematologic malignancies by realtime quantitative PCR: principles, approaches, and laboratory aspects. Leukemia 17: 1013-1034.

Verpoorte R, Alfermann A W (2000) Metabolic engineering of plant secondary metabolism. Kluwer Academic Publishers, Dordrecht.

Verpoorte R, van der Heijden, Hoopen HJG and Memelink J (1998) Metabolic engineering for the improvement of plant secondary metabolite production. Plant Tissue Culture and Biotechnology 4: 3–20.

Vet J A, Marras S A (2005) Design and optimization of molecular beacon real-time polymerase chain reaction assays. Methods Mol Biol 288: 273-290.

Watt G A (1972) Dictionary of the Economic Products of India; Cosmo Publication, Delhi, India. Vol. 6, p. 309.

Williamson E M (2002) Major Herbs of Ayurveda; Churchill Livingstone: London, UK, pp. 322 - 323.

Wink M (1999) Biochemistry of plant secondary metabolism. Annual plant reviews, volume 2, Ch. 1, p 1-16.

Wittwer C T, Herrmann M G, Moss A A, Rasmussen R P (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22: 130-131, 134-138.

Wittwer C T, Ririe K M, Andrew R V, David D A, Gundry R A, Balis U J (1997) The Light cycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22: 176-81.

Xu ZJ *et al.,* (2002) Cloning and characterization of the abscisic acid specific glucosyltransferase gene from adzuki bean seedling. Plant Physiol 129, 1285-1295.

Yi Li, Sandie Baldauf, Eng-Kiat Lim and Dianna J. Bowles (2001) Phylogenetic Analysis of the UDP-glycosyltransferase Multigene Family of *Arabidopsis thaliana*. Vol. 276, No.6, 4338-4343.

Ziauddin M, Phansalkar N, Patki P, *et al*., (1996) Studies on the immunomodulatory effects of ashwagandha. J Ethnopharmacol. Feb; 50:69-76.

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, eds. Evolving Genes and Proteins, New York: Academic Press, 97-166.