Cloning and characterization of glycosyltransferase gene(s) (GTs) from *Withania somnifera* and establishment the structure function correlation of the GTs

A THESIS

SUBMITTED TO THE

UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

BY

Somesh Singh

Under the guidance of

DR. B. M. KHAN

HEAD

PLANT TISSUE CULTURE DIVISION

NATIONAL CHEMICAL LABORATORY

PUNE- 411008

INDIA

February 2013

Dedicated to My family.....

राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग पुणे - 411 008.भारत
NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411 008 India.

Dr. Bashir M. Khan
Head
Plant Tissue Culture Division

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Cloning and characterization of glycosyltransferase gene(s) (GTs) from *Withania somnifera* and establishment the structure function correlation of the GTs"** submitted by **Somesh Singh** 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. C. G. Suresh
(Research Co-guide)

Dr. B. M. Khan
(Research guide)

Off : 091 20 2590 2220
e-mail : bm.khan@ncl.res.in
Res. : 091 20 2589 3251
URL : www.ncl-india.org.
Fax : 091 20 2590 2645
091 20 2590 2785

DECLARATION

I hereby declare that the thesis entitled **"Cloning and characterization of glycosyltransferase gene (s) (GTs) from *Withania somnifera* and establishment the structure function correlation of the GTs"** has been carried out at Plant Tissue culture Division, National Chemical Laboratory, Pune, under the guidance of **Dr. B. 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.

Somesh singh

Date: February, 2013

Place: Plant Tissue Culture Division

National Chemical Laboratory (NCL)

Pune, Maharashtra, India

Pin- 411008

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 like to thank my research co-guide ***Dr. C. G. Suresh****, Department of Biochemical Sciences, NCL for his valuable comments and suggestions throughout this research work.*

I am immensely grateful to ***Dr. Anjan Banerji****, Assistant Professor, IISER, pune for his valuable comments and precious suggestion during the progress and completion of thesis work.*

A very special thanks to ***Dr. D. C. Agarwal*** *and his student Krunal patel to help in homology and modeling studies during completion of this work and I would like to thanks* ***Dr. Srinivas Hotha*** *and his student Asif shiekh for providing me the LC-MS facility and their analysis respectively.*

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 grateful to Dr. Rishi. K. Vishwakarma, Prashant. D. Sonawane, Dr R. J. Santosh Kumar, Dr. Ruby and Krunal Patel for their help during the progress and completion of this work.

My sincere thanks to Dr. Sucheta and Mr. S. C. Mishra for timely support and guidance during my stay at NCL.

I would like to extend my thanks to Dr. D.C. Agarwal, Dr. Urmil Mehta for allowing me to use their lab facilities and all other members of the scientific and supporting staff (Mr. Mahale, Mr. Jathar and Mr. Khamkar, of the Plant Tissue Culture Division, Biochemical Sciences Division and NCL Library. My sincere thanks to Mrs. S.

Kendurkar for suggestions and encouragement. Mr. Suryaprasad deserves special thanks for using his technical expertise for timely repairing of the equipments.

Many thank to my seniors Dr. Sameer, Dr. Sushim, Dr. Noor, Dr. Ashish, Dr. Manish, Dr. Arun, Dr. Pallavi, and Dr. Abhilash, for their encouragement and help during start of this work.

Dr. Sameer, Tonima Bhabhi jee and Saurav (Roud, Supriya bhabhi jee) deserve special thanks for their constant support and suggestions.

I am immensely grateful to seniors cum ex-labmate Dr. Ruby, Dr. R. J. Santosh Kumar, Dr. Sumita and Dr. Santosh Gupta for their valuable support and suggestions during the entire work.

I would like to thank my labmates and colleagues Rishi, Prashant, Krunal, Shakeel, Trupti, Parth, sameena, Poonam, Sumita,Uma, Santosh,Neha, Kannan, Nishi, Rajshree and Gayatri for maintaining lively and cordial atmosphere in the lab.

I thank to trainees Tripti, Shuchi, Aditi, Mehak, Henna, Zoheb, Deepali, Sonali,Mukul, Nazneen, Krishna, Sobia, Ambreen, Yasir, Rajeev for their valuable help during my work period. Many thanks too lab trainees Henna and Mehak for their help.

Thanks are due to all my friends Rishi, Sudhanshu, Prashant, Krunal, Parth, Shakeel, Ruby, Santosh R.J., , Dr. Bhuban, Neha, Pooja, Trupti, Uma, Manas,Tulika, Manisha, Pushkar, Dr. Fazal, Devdutta (DD), Arup, SangramDr. Prasad, Raju, Poonam, Yashwant, Kannan, Neha Khandelwal, Abhishek, Nishant, Manas, Deepak, Priyabrat, Ruby singh and Rajeshree for the support, best wishes and lighter moments shared together throughout my work period.

Special mention is a must for Rishi, Prashant and Parth for being my extended family during the course of my work. Those were really memorable days when, on weekend we all including Neha, Pooja, Krunal and Ruby gathered for dinner. They have given me a perfect family atmosphere and I never ever felt that I am away from my home.

*A very special thanks to member and player of **GJPL (Golden jubilee premier league)**, Ashish Tripathi, Mandeep, Varun, Abhishek, Kishore Bhaiya, Kaushal, Ganesh, Sridhar, Sameer, Sushim, Kanna,Krunal and Rama with whom I spend a memorable time on ground or outside the ground.*

I take this opportunity to thank my M.Sc. friends Rishi, Nidhi, Tripti, Gurpreet kaur, Nikki, Shashikant, Tanuj, Vishal, Ajeet and saptarshi dutta for affection, wishes and encouragement.

I owe my deepest gratitude to my parents, my didi and jijaji, my wife sonam and whole family for their blessing, constant support and unconditional affection, which helped me in building my career. Very special thanks to my nephew Abhinav and Utpal.

Thanks to all who directly or indirectly helped during my Ph. D tenure and I might have forgotten to mention those names here.

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

CSIR, India is duly acknowledged for fellowship and grants.

Last but not the least I thank ***almighty God*** *for giving me the strength and courage at every step of my life.*

Somesh

CONTENTS

Abbreviations

AA	Amino acid
AMV	*Avian Myeloblastosis Virus*
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
Ci/ mmol	Curie per milli mole
CIAP	Calf Intestinal Alkaline Phosphatase
cm	Centimeter
Da	Dalton
DEPC	Diethylpyrocarbonate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribose nucleic acid
DTT	Dithiothritol
EDTA	Ethylene Diamine Tetra Acetic acid
EtBr	Ethidium bromide
GT	Glycosyltransferase
g /L	Grams per litre
g	Gram
G	Guaiacyl
GSP	Gene Specific Primers
GSH-Px	Glutathione peroxide
Glc T	Glycosyltransferases
Gat T	Galactosyltransferases
h	Hour(s)
HPLC	High performance liquid chromatograhy
IPTG	Isopropyl ß-D-thiogalactoside

Kb	Kilobase pairs
KDa	Kilo Daltons
Kg	Kilogram
L	Litre
LC-MS	Liquid chromatography-Mass Spectrometry
LB	Luria-Bertani
m	Meter
MCS	Multiple cloning sites
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
PCR	Polymerase Chain Reaction
pI	Isoelectric point
PEG	Polyethylene glycol
pg	Picogram
pmol	Picomole
PMSF	Phenyl methyl sulphonyl fluoride
ppm	Parts per million

PVPP	Poly vinyl pyro phosphate
pmol	Pico mole
PSPG	Plant secondary product glycosyltransferase
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 PCR
s	Second(s)
SDS	Sodium dodecyl sulphate
PAGE	Polyacrylamide gel electrophoresis
SMQ	Sterile Milli Q
TAE	Tris acetic EDTA buffer
TE	Tris EDTA buffer
TEMED	Tetramethylethylenediamine
U	Units
UGT	UDP-glucose glycosyltransferases
UDP	Uridine diphosphate
UPM	Universal Primer Mix
UTR	Untranslated Region
UV	Ultraviolet
V	Volt
v/v	Volume / Volume
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 (Commonly known as Ashwagandha) belongs to the *Solanaceae* family. It is one of the important perennial medicinal crops in many states of India. Different parts of the plant have the different medicinally important functions. Its fruits and seeds are diuretic and hypnotic, also rich in saponins and can be used as substitutes of soap. In Ayurveda the roots are prescribed for gynaec disorders, bronchitis, arthritis, rheumatism, inflammation, fevers, and skin diseases, etc. Mainly, leaves contain the different type of glucosides that exhibit anti-inflammatory, anti-tumor and anti-oxidant property etc.

W. somnifera plant produces many types of secondary metabolites, including flavonoids, alkaloids and terpenoids**.** These compounds undergo modification reactions such as methylation, hydroxylation and glycosylation, which lead to the structural diversity of secondary metabolites**.** Glycosylation is one of the major modification reactions that often occur in the final step of the natural compound biosynthesis. The primary roles of glycosylation in plants are the stabilization of pigments, enhancement of solubility, storage of secondary metabolites and regulation of plant growth regulators. Enzymes, leading to glycoside formation are known as glycosyltransferases (GTs), which transfer nucleotide-diphosphate-activated sugars to low molecular weight substrates. The activated sugar form is typically UDP-glucose, but UDP-galactose and UDP-rhamnose may also be found.

In plants, sugar acceptors include all major classes of secondary metabolites, such as phenolics, terpenoids, cyanohydrins and alkaloids. Large numbers of UGTs have been recently identified which are functionally involved in the biosynthesis of saponins, flavonoids, and iso-flavonoids. But the functional characterization of glycosyltransferase gene, specific for flavonoid biosynthesis from this plant is still unknown. The present study was taken up with the objective of isolation, cloning and functional characterization of glycosyltransferase gene(s) from *Withania somnifera* which play a crucial role in glycosylation of medicinally important secondary metabolites. This work has been divided into five chapters and details are given below.

Chapter 1

General Introduction

This chapter gives general information on medicinal plants in India and background of research was done on *Withania somnifera*. A thorough literature survey of work done in the area of glycosyltransferase gene 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 in details. 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

Isolation, Cloning and Characterization of glycosyltransferase gene from *Withania somnifera*

This chapter deals with the isolation, cloning and characterization of full-length cDNA encoding flavonoid glycosyltransferase from *W. somnifera*. Primers were designed on the basis of sequence information available in NCBI GenBank database. PCR was done with different set of primers, using cDNA as a template. The partial fragment was cloned and sequenced, which were showing significant similarity with reported glycosyltransferase genes from other plants. Gene specific primers were designed and Rapid Amplification of cDNA Ends (RACE) PCR was performed to isolate full-length cNDA clones. The full length sequence reported in this chapter has been deposited in the NCBI gene bank database [Gen Bank: FJ654696]. According to the glycosyltransferase nomenclature guidelines, the systematic name of the *W. somnifera* glycosyltransferase is UGT73A16. This translated protein sequence of UGT73A16 was used to construct a phylogenetic tree with all known plant glycosyltransferases (GTs) deposited in NCBI Gen Bank database using Neighbor-joining method.

Chapter 3

(3A). Heterologous expression, Purification and functional Characterization of glycosyltransferase gene

This chapter deals with the cloning of full-length cDNA (UGT73A16) into the expression vector pET 30b (+) and its over-expression in *E. coli* BL21 (DE3). The recombinant proteins were purified by affinity chromatography using Ni-NTA agarose beads. The enzyme assay was done using UDP-glucose and UDP-glucuronic

acid as donor substrate and flavonoids as acceptor substrate. Product identification was done by hypsochromic shift, HPLC and LC-MS method. This chapter also covers the kinetic parameters of recombinant UGT73A16.

(3B). *Withania somnifera* glycosyltransferase (UGT73A16) as biocatalysts to produce plant specific glucosides

In this chapter, UGT73A16 of *W. somnifera* was used as whole cell catalyst in *E. coli.* To investigate the production of glucosides by recombinant *E. coli in vivo*, incubation of induced cell cultures of *E. coli* BL21 (DE3) carrying the recombinant UGT73A16 was done with equal concentration of each substrates (genistein, apigenin, kaempferol, naringenin, biochaninA and daidzein) for different time intervals (12 h, 16 h, 20 h, 24 h and 48 h). Approximate 21 mg/L to 22 mg/L of either apigenin 7-*O*-glucoside or naringenin 7-*O*-glucoside was produced after feeding apigenin or naringenin to *E. coli* expressing UGT73A16. More than 95% of the glucosides products were released into the medium, facilitating their isolation.

Chapter 4

***In silico* modeling and docking studies of *Withania somnifera* glycosyltransferase (UGT73A16)**

This chapter deals with structure validation, secondary structure prediction, modeling and docking studies of UGT73A16. Initially amino acid sequence of UGT73A16 was searched against PDB using BLASTP program at NCBI to find out the homologous sequences. The three dimensional coordinates of 2VG8, 2PQ6 and 3HBJ were used as templates to generate the 3D models of UGT73A16 using homology modeling software MODELLER 9v9. The modeled structures of UGT73A16 were further visualized using program PYMOL and structure validation of models was done by PROCHECK, ERRAT and DOPE score. The secondary structure of UGT73A16 was predicted by the software PSIPRED v 3.0. The docking analysis of modeled UGT73A16 with structural analogs of different acceptors (Kaempferol, naringenin, genistein and baicalein) and donor substrates (UDP-glucose and UDP- glucuronic acid) was carried out by Autodock vina docking software.

Chapter 5

Mutational analysis of *Withania somnifera* glycosyltransferase (UGT73A16)

The glycosyltransferase UGT73A16 from *W. somnifera* catalyzes glycosylation of the (iso) flavonoids, flavonols, flavonone and flavone. It can transfer glucose to hydroxyl group of (iso) flavonoids, flavonols, flavonone and flavone with 3-hydroxyl and 7-hydroxyl group as the major product. Based on the docking studies of modeled UGT73A16, site directed mutagenesis was carried out to explore the roles of amino acids involved in substrate binding. In this chapter *in-vitro* identification of amino acid was done by mutation (substitution) and compared by kinetic parameters.

Chapter 1

General Introduction

1.0 Introduction

Among ancient civilizations, the forests of India have been known to be rich repository of medicinal and aromatic plants. These medicinal plants are mainly used as raw materials for manufacture of drugs and perfumery products. Man in the pre-historic era was probably not aware about the health hazards associated with irrational therapy. With the arrival of research in medicine, it was concluded that plants contain active principles, which are responsible for curative action of the herbs. Before onset of synthetic era, man was completely dependent on medicinal herbs for prevention and treatment of diseases. With the introduction of scientific procedures, the researchers were able to understand about toxic and non-toxic principles present in the green flora.

1.1 Medicinal plants

Plants have shaped the basis of traditional medicine systems. These traditional systems have been in existence for thousands of years and continue to provide mankind with new remedies. Medicinal plant therapy is based on the experimental findings of hundreds and probably thousands of years of use. Until the beginning of modern medicine, man was fully dependent on plants for treating human and livestock diseases. The 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 (**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. Apart from healthcare, medicinal plants are mainly the alternate income-generating source of underprivileged communities (**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. About 121 (45 tropical and 76 subtropical) major plant drugs have been identified for which no synthetic drugs are currently available (**Table 1.0**).

Table 1.0 List of few important medicinal plants and their uses (**Kumar et al. 1997**)

Botanical Name	Common name	Medicinal Use
Withania somnifera	Aswagandha	Restorative tonic, stress, nerves disorder, aphrodiasiac.
Bacopa monnieri	Brahmi	Nervous, memory enhancer, mental disorder.
Azardirchata indica	Neem	Sedative, analgesic, epilepsy, hypertensive.
Ocimum sanctum	Tulsi	Cough, Cold, bronchitis, expectorand
Cinnamomum zeylanicum	Dalchin	Bronchitis, Asthma, Cardiac, Disorder, Fever
Aloe verra	Gritkumari	Laxative, Wound healing, Skin burns & care
Mesua ferrea	Nag Keshar	Asthma, Skin, Burning, Vomiting, Dysentery
Catharanthus roseus	Sadabahaar	Anticancer, Leukemia
Rauvolfia serpentina	Sarpgandha	Tranquilizer
Cinchona officinalis	Cinchona	Antimalarial, amoebic dysentery
Taxus baccata	English Yew	Breast and ovary cancer, antitumour
Digitalis sp.	Foxgloves	congestive heart failure
Papaver somniferum	Opium Poppy	Painkiller and anticough
Allium sativum	Garlic	Antifungal, amoebiasis

1.2 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, ethno botanist, ethno pharmacologist or plant ecologist who collects and identifies the plants of interest. Phytochemists (natural product chemists) prepare extracts from the plant material, subject these extracts to biological screening in pharmacologically relevant assays and initiate the process of isolation and characterization of the active compounds 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.

1.3 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 (**Kirsi et al. 2004**). 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. 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.

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 and used in the commercial applications such as drugs, dyes, flavours, fragrance, insecticides etc. One of the functions of secondary compounds is that they form biochemical defense mechanism against pathogens and predators (**Bennet and Wallsgrove 1994**). 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 culture (**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 (**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 recognized to structural changes associated with functional differentiation during the course of enzyme/gene evolution (**Fig. 1.0**) (**Pichersky and Gang 2000**).

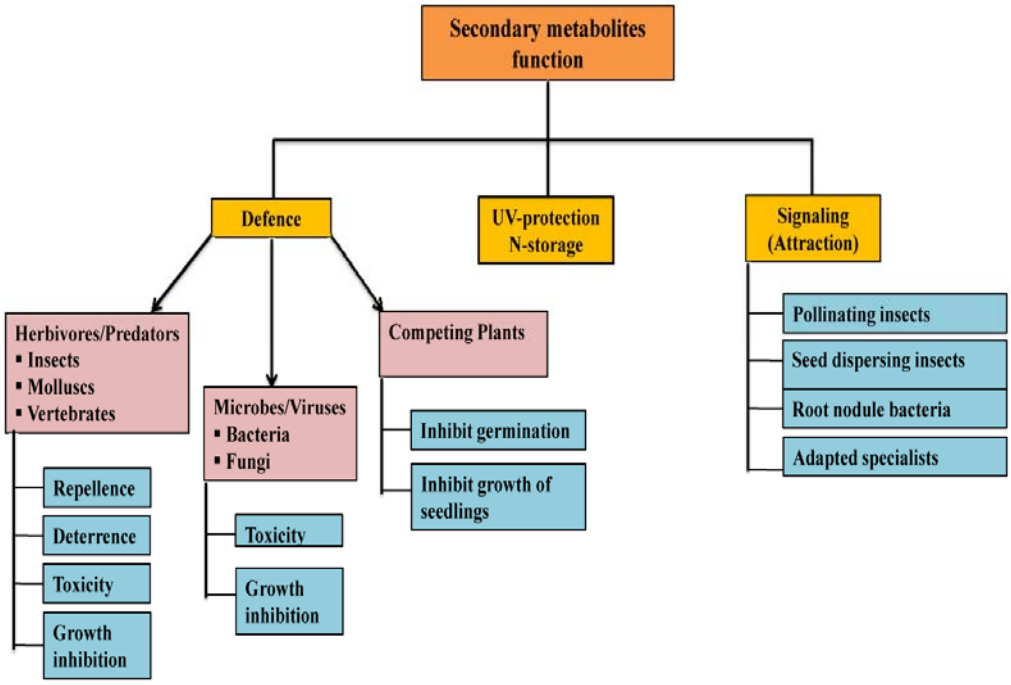


Fig. 1.0 Ecological and Physiological functions of Plant Secondary metabolites

The major roles of secondary metabolites are:

- Defense against herbivores (insects, vertebrates)
- Defense against fungi, bacteria and 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)**.

1.4 *Withania somnifera* (Ashwagandha)

Withania somnifera (L. Dunal) commonly known as 'Ashwagandha' belongs to solanaceae family, is an important medicinal plant. It is commonly used as a domestic remedy for several diseases in India as well as other parts of the world. It is described as an herbal tonic and health food in the famous book of Vedas and is considered as 'Indian Ginseng' in traditional Indian system of healing (**Dhuley 2000**). Several recent reports have demonstrated immuno-modulator and anti-tumor effect of *W. somnifera* as well (**Agarwal et al. 1999**). Moreover, various parts of the plant have been reported to

possess anti-serotogenic, anti-cancer and anabolic properties and have shown beneficial effects in the treatment of arthritis, stress and geriatric problems (**Prakash et al. 2001**).

W. somnifera is also made into dietary supplements with good nutritional properties. The pharmacological effect of the roots of *W. somnifera* is attributed to its active ingredient, withanolides (**Budhiraja and Sudhir 1987**), which has a wide range of therapeutic applications (**Udayakumar et al. 2009**).

1.4.1 Botanical description

Withania somnifera (Ashwagandha) plant is native to India and cultivated as an annual crop. The *W. somnifera* plant is erect, 30-150 cm high with fleshy roots which is whitish brown in color, gradually tapering down and pure white inside when broken. The stem is green and erect, covered with minute star shaped hair usually clothed with stellate hoary tomentum. Leaves are simple and ovate up to 10 cm long. A flower looks like lucid-yellow or greenish and it contains small berries which is orange-red in color. Mature fruits are orange-red colored enclosed in a persistant calyx, containing many seeds. Florescence occurs in fall and spring.

NOMENCLATURE

Kingdom:	Plantae
Subkingdom:	Tracheobionta
Division:	Magnoliophyta
Class:	Magnoliopsida
Subclass:	Asteridae
Order:	Solanales
Family:	Solanaceae
Genus:	***Withania***
Species:	***somnifera***

1.4.2 Geographical distribution

W. somnifera (L. Dunal) is cosmopolitan and grows throughout the drier parts and subtropical India. It is widely distributed in North-Western India, Bombay, Gujarat, Rajasthan, Madhya Pradesh, and Uttar Pradesh, Punjab plains extending to the mountain regions of Punjab, Himanchal Pradesh and Jammu, ascending to a height of 1500 m. It is cultivated in 5000 ha in Rajasthan, Madhya Pradesh, Andhra Pradesh and Uttar Pradesh

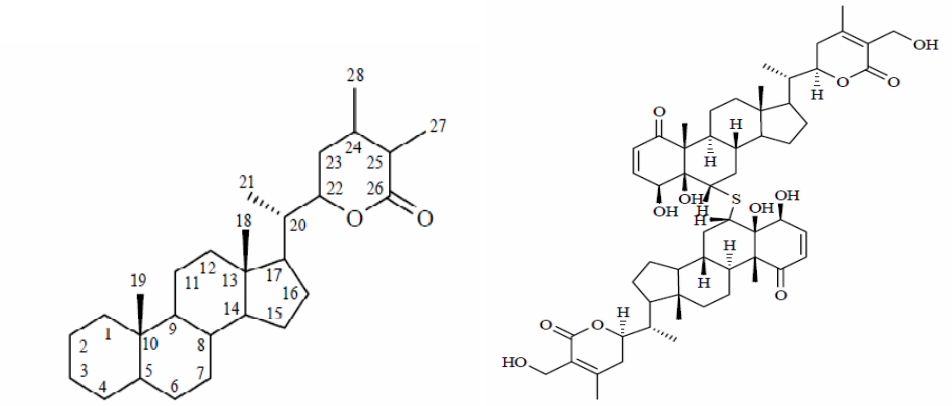
(Arun et al. 2007). The wild growth of this species has also been reported from Pakistan, Afghanistan, Palestine, Egypt, Jordan, Morocco, Spain, Canary, Island, Eastern African Cong and South Africa. As mentioned earlier, *Withania somnifera* has a very wide geographical range. In India its range extends from 230 N to 330 N and extending from 180 m to 1500 m altitude (Monograph of Ashwagandha).

1.4.3 Cultivation

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. It is a late rainy season crop. It requires relatively dry seasons, and the roots are fully developed when 1-2 late winter rains are received. Late winter rains are conducive for the proper development of the plant roots. The areas receiving 65-75 cm rainfall are best suited for its cultivation. The semi-tropical areas receiving 500-750 mm rainfall are suitable for cultivation of this rained crop. It can tolerate temperature between 20 ºC to 38 ºC and as a low temperature as 10 ºC.

1.4.4 Chemical constituents

The chemistry of *Withania somnifera* species has been extensively studied and known. Presently most of the compound such as steroidal lactones (Withanoloids, withaferins), alkaloids (isopelletierineanaferine), flavonoids, tannin and many more have been extracted, isolated and identified from aerial parts, roots and berries of *W. somnifera* (**Atta-ur-Rahman. et al. 1991; Atta-ur-Rahman. et al. 1993**). The major chemical components of *Withania somnifera* are withanoloids, which are mainly localized in leaves (**Bandyopadhyay et al. 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 (**Fig. 1.1 A**) (**Alfonso and Kapetanidis 1994**).



Withanoloide **Ashwagandhanolide**

Fig 1.1 Chemical structure of **(A)** Withanoloide **(B)** Ashwagandhanolide

Examination of *W. somnifera* roots has resulted in the isolation of a new dimeric thiowithanolide, named ashwagandhanolide (**Fig. 1.1 B**). The structure of these compounds was determined by using serotonin, glucose and long-chain hydroxyl fatty acid moieties (**Jayaprakasam et al. 2004**).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. The leaves are reported to contain five unidentified alkaloids chlorogenic acid, calystegines (nitrogen-containing polyhydroxylated heterocyclic compounds) withanone, condensed tannin and flavonoids (**Johri et al. 2005**).

1.4.5 Therapeutic use of *W. somnifera*

W. somnifera is one of the important medicinal crops and has antibiotic, anti-viral, anti-amoebic, anti-arthritic 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 which have the wide range of bioactivities like allelopathic, antimicrobial, anticholesterolemic, anti-cancer, adjuvant, immunomodulatory, antioxidant and haemolytic activities (**Bhaskara et al. 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**).

1.4.6 Pharmacological activity

From the centuries it is known that *W. somnifera* have the pharmacological value as an adaptogen, antibiotic, aboritifacient, aphrodisiac, astringent, anti-inflammatory, deobstruent, diuretic, narcotic, sedative and tonic. Ashwagandha has been found to:

- Provide potent antioxidant protection (**Abou-Douh 2002; Panda and Kar 1997**).
- Stimulate the activation of immune system cells, such as lymphocytes and phagocytes (**Wagner et al. 1994; Singh et al. 2001**).
- Counteract the effects of stress and generally promote wellness (**Singh et al. 2003**).

1.4.6.1 Anti-stress

W. somnifera has traditionally been used to stabilize mood in patients with behavioral disturbances. Research has revealed that the herb produces an anti-depressant and anti-anxiety effect in rodents comparable to the anti-depressant drug imipramine and the anti-anxiety drug lorazepam (**Archana and Namasivayam 1999**). The effect of Ashwagandha on chronic stress in rodents was done by the researchers of Institute of Basic Medical Sciences at Calcutta University (**Bhattacharya and Muruganandam 2003**).

1.4.6.2 Anti-oxidant activity

The brain and nervous system are relatively more prone 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 et al. 2000**). Researchers from Banaras Hindu University in Varanasi, India, have discovered that sitoindosides VII-X and withaferinA found in *W. somnifera* are powerful antioxidants and 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 (**Dhuley 2000**). These findings are consistent with the therapeutic use of *W. somnifera* as an ayurvedic rasayana.

1.4.6.3 Anti-carcinogenic activity

Use of *W. somnifera* on animal cell cultures causes decreased levels of the nuclear factor kappaB, suppression of intercellular tumor necrosis factor and potentiate apoptotic signaling in cancerous cell lines (**Ichikawa et al. 2006**). *W. somnifera* has the capacity

to fight cancers by reducing the size of tumour (**Prakash et al. 2002; Jayaprakasam et al. 2003**). Antitumor and radio-sensitizing effects of Withaferin were also seen in mouse Ehrlich ascites carcinoma *in vivo* (**Sharad et al. 1996**). WithaferinA from *Withania* gave a radio-sensitizer ratio of 1:5 for *in vitro* cell killing of V79 Chinese hamster cell at a non-toxic concentration of about 2 mM/L.

1.4.6.4 Anti-inflammatory activity

Research has explored the capacity of Ashwagandha to ease the symptoms of arthritis and other inflammatory conditions. These studies have proven that the herb acts as an effective anti-inflammatory agent. Its naturally occurring steroidal content is much higher than that of hydrocortisone, a commonly-prescribed anti-inflammatory (**Angalagan and Sadique 1981**). The effectiveness of Ashwagandha in a variety of rheumatologic conditions may be due in part to its anti-inflammatory properties. Rats given powdered root of Ashwagandha orally one hour before being given injections of an inflammatory agent over a three day period showed that Ashwagandha produced anti-inflammatory responses comparable to that of hydrocortisone sodium succinate (**Begum and Sadique 1988**).

1.4.6.5 Anti-aging activity

Ashwagandha was tested for its anti-aging properties in a double-blind clinical trial. A group of 101 healthy males, 50-59 years old were given the herb at a dosage of 3 grams daily for one year. The subjects experienced significant improvement in hemoglobin, red blood cell count, hair melanin, and seated stature. Serum cholesterol decreased and nail calcium was preserved. Seventy percent of the research subjects reported improvement in sexual performance (**Bone 1996**).

1.4.6.6 Cardioprotective activity

Ashwagandha has been evaluated in clinical studies with human subjects for its diuretic, hypoglycemic, and hypocholesterolemic effects (**Andallu and Radhika 2000**). Six type 2 diabetes mellitus subjects and six mildly hypercholesterolemic subjects were treated with a powder extract of the herb for 30 days. A decrease in blood glucose comparable to that which would be caused by administration of a hypoglycemic drug was observed. Significant increases in urine sodium, urine volume, and decreases in serum cholesterol, triglycerides, and low-density lipoproteins were also seen.

1.4.6.7 Immuno-modulatory activity

A series of animal studies have demonstrated that Ashwagandha have profound effects on healthy production of white blood cells, which means it is an effective immunoregulator and chemoprotective agent (**Kuttan 1996; Ziauddin et al. 1996**).

In a study using mice, administration of powdered root extract from Ashwagandha was found to enhance total white blood cell count. In addition, this extract inhibited delayed-type hypersensitivity reactions and enhanced phagocytic activity of macrophages when compared to a control group (**Davis and Kuttan 2000**). Research has also shown Ashwagandha to have stimulatory effects, both *in vitro* and *in vivo*, on the generation of cytotoxic T lymphocytes, and a demonstrated potential to reduce tumor growth (**Davis and Kuttan 2002**).

2.0 Review for glycosyltransferases

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, such as regulation of hormones homeostasis, detoxification of xenobiotics and the biosynthesis and storage of secondary compounds (**Claire et al. 2005**). Glycosylation of small molecular weight, lipophilic acceptors plays 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 (**Bowles et al. 2005**).

2.1 Glycosylation

Glycosylation is often the final step in the biosynthesis of secondary plant products resulting in the formation of an overwhelming number of natural glucosides with numerous applications. This class of products is difficult to isolate or to synthesize in pure form because of the resulting low yields. Thus, simple approaches for the generation of such glucosides would be highly beneficial. The attachment of glucose is a rather plant-specific process and only rarely catalyzed by fungi and bacteria. Glycosides are usually non-reducing compounds, on hydrolysis by reagents or enzymes yield one or more reducing sugars among the products of hydrolysis. Glycosylation is the process of addition of saccharides to proteins and lipids. It is an enzyme-directed site-specific process, as opposed to the non-enzymatic chemical reaction of glycation. Two types of glycosylation exist: *N*-linked glycosylation to the amide nitrogen of asparagine side chains and other compounds. The *O*-linked glycosylation to the hydroxy oxygen of serine and threonine side chains or other compounds (**Fig. 1.2 A & 1.2 B**).

1- Alcoholic or phenolic (aglycone): e.g., *O*-Glycoside

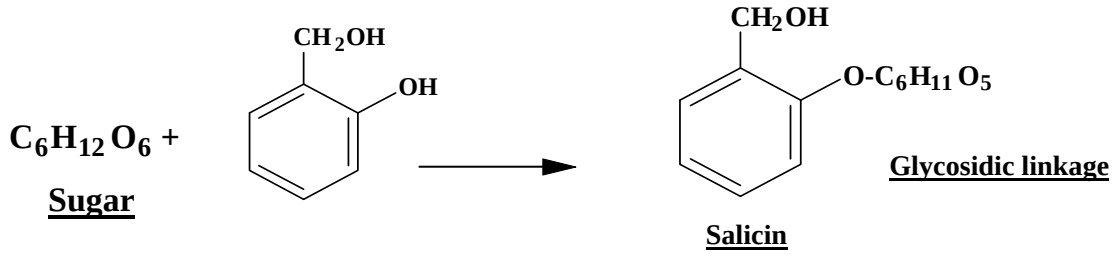


Fig. 1.2 (A) *O*-linked glycosylation

2- Nitrogen containing compounds: e.g., N-Glycoside

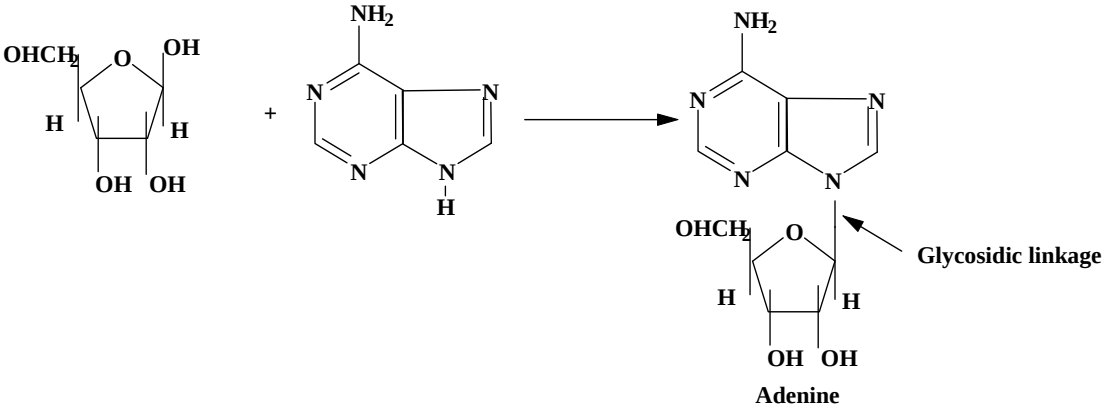


Fig. 1.2 (B) *N*-linked glycosylation

2.2 Purpose of glycosylation

Biological functions of glycosylations in plants include storage, inter and intracellular transport of metabolites, regulation of homeostasis of hormones, etc. *In vitro* studies have shown that a single gene product can glycosylate multiple substrates of diverse origins; multiple enzymes can also glycosylate the same substrate. Glycosylation alters the solubility of compounds and their movement within the cell. In higher plants, UGT catalyzed glycosylation constitutes a prominent terminal modification in the biosynthesis of secondary metabolites and generates diverse natural glycosides.

Glycosylation converts reactive and toxic aglycon into stable and non-reactive storage forms, thereby limiting their interaction with other cellular components. In general 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. Sugars are highly polar and the attachment of sugar to hydrophobic substrates, increase the water solubility (highly polar) of substrates.

The main chemical roles of glycosylation are given below;

- Detoxification (reduce reactivity)
- Solubilization (increased polarity)
- Storage of secondary metabolites
- Regulation of plant growth regulators

2.3 Glycosyltransferases

Glycosyltransferases are a highly divergent and multigene family (**Mackenzie et al. 2005**), and responsible for glycosylation reactions (attachment of a sugar residue from an activated sugar donor to a receptor molecule). The glycosyltransferase multigene family is categorized into 94 numbered families according to sequence similarity, signature motifs, stereochemistry of the glucoside linkage formed and known target specificity (**Campbell et al. 1997**). Of these 54 families, the family 1 contains the UDP-glycosyltransferases (UGTs) (**Mackenzie et al. 2005; Li et al. 2001; Ross et al. 2001**), which have been identified in plants, animals, fungi and bacteria, and also in viruses (**Campbell et al. 1997**). These glycosyltransferases are characterized by utilization of UDP-activated sugar moieties as the donor molecule, and contain a conserved UGT defining sequence motif near the C-terminus. This UGT-defining motif is often the only significant region of similarity in sequence alignments within and across phyla. The UGTs are classified based on conserved domain in their C-terminal end. Sugar donors of plant GTs are typically UDP-glucose (UDP-Glc), although UDP-rhamnose (UDP-Rha), UDP-galactose (UDP-Gal), UDP-xylose (UDP-Xyl), and UDP-glucuronic acid (UDP-GlcUA) have also been reported. Sugar acceptors include all major classes of secondary metabolites, including phenolics, terpenoids, cyanohydrins (cyanogenic glycoside precursors), thiohydroximates (glucosinolates precursors) and alkaloids (e.g. betalains). Single or multiple glycosylation of the acceptors can occur at –OH, –COOH, –NH2, –SH, and C–C groups.

2.4 Glycosyltransferases classification

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) 12, 000 sequences from different organisms have been reported. These sequences were classified into 94 distinct families (Lim 2005). Among these, family 1 has the most number of GTs which have a close relationship with plant GTs. The CAZy database proposes the continuously updated classification of glycosyltransferases using nucleotide diphospho-sugar, nucleotide

monophospho-sugars and sugar phosphates (EC 2.4.1.x) and related proteins into distinct sequence-based families. (**Campbell et al. 1997; Coutinho et al. 2003**).

2.4.1 Catalytic specificity

Glycosyltransferases can be classified as either retaining or inverting (**Fig. 1.3**) enzymes according to the stereochemistry of the substrates and reaction products (**Sinnott 1990**). By analogy with glycosidases, two main stereochemical outcomes exist for glycosyltransferases: inversion of the anomeric configuration (for instance UDP-glucose -> β-glucoside) or retention of the anomeric configuration (for instance UDP-glucose -> α-glucoside).

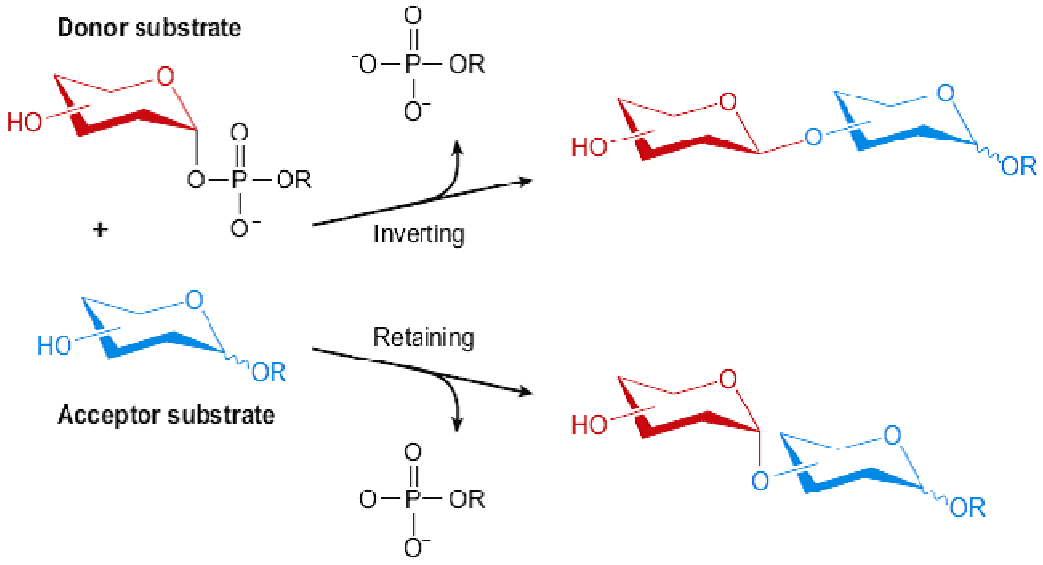


Fig. 1.3 Inverting and Retaining glycosyltransferases

Inverting glycosyltransferases most likely follow a single displacement mechanism where the acceptor makes a nucleophilic attack at C-1 of the sugar donor. On the other hand, retaining glycosyltransferases do not follow two-step mechanism involving the formation of a glycosyl-enzyme intermediate. Instead, an internal return SNi-like mechanism has been proposed, in which leaving group departure and nucleophilic attack occur in a concerted but asynchronous manner on the same face of the glycoside.

2.4.2 Structural specificity

The 3D structures of glycosyltransferases have revealed two major structural folds designated GT-A and GT-B. (**Coutinho et al. 2003; Bourne and Henrissat 2001**). The GT-A and GT-B fold 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 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 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.

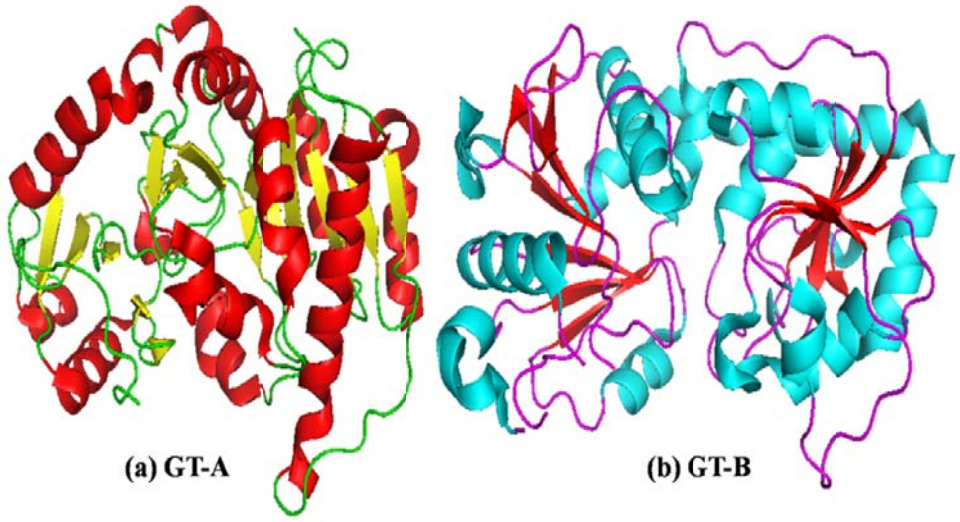


Fig. 1.4 (**A)** GT-A-fold type with a single Rossmann fold **(B)** GT-B fold type with two Rossmann folds

The two domains in the GT-A type fold cooperate to form the active- site cleft. The GT-A type fold (**Fig. 1.4 A**) is believed to be ancestral to enzymes with the GT-B type fold (**Coutinho et al. 2003**). 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 GT-B type fold (**Fig. 1.4 B**) 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 and Rini 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 and Mushegian 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. Among GT-A and GT-B super families, the overall fold of the enzyme does not dictate the stereochemical outcome of the reaction that it catalyzes, as examples of both inverting and retaining glycosyltransferase have been identified within both the GT-A and GT-B fold classes. Based on amino acid sequence, this enzyme was initially classified among the GT-A inverting family GT2; however, structural and mechanistic studies led to its reclassification among the GT-A retaining family GT78. To date, all enzymes predicted to adopt the GT-C fold belong to inverting glycosyltransferase families.

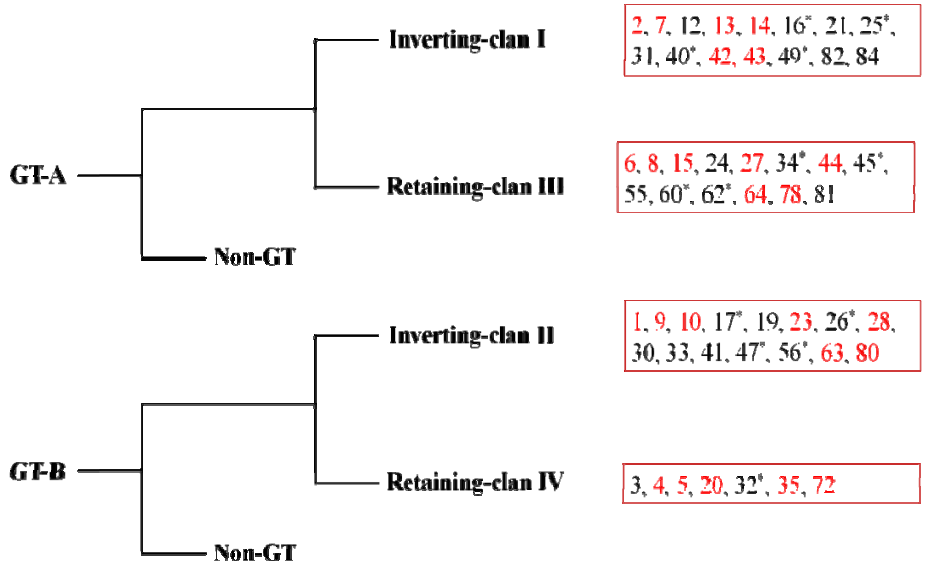


Fig. 1.5 Glycosyltransferase (GT) classification. Families are classified into clans on the basis of their fold and activity. GT family numbers belonging to each clan are indicated on the far right. Bona fide families having members with solved 3-D structures are indicated in red. The remaining families are those predicted to adopt either the GT-A or the GT-B fold. Families identified in black with an asterisk are those with structures predicted to adopt either the GT-A or the GT-B fold solely by Liu & Mushegian (19), and those in black without an asterisk have GT-A or GT-B structures as predicted by both Liu & Mushegian and the CAZY Web site.

2.4.3 Existence of consensus sequence

UGTs have been identified in plants, animals, fungi, bacteria and viruses (**Campbell et al. 1997**). These glycosyltransferases are characterized by utilization of UDP-activated sugar moieties as the donor molecule, and contain a conserved UGT defining sequence motif near the C-terminus (**Mackenzie et al. 2005**). This UGT-defining motif is often

the only significant region of similarity in sequence alignments within and across phyla. Of those 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 and is believed to be participating in binding of the activated sugar donors. Highly conserved amino acids are indicated by **':' (Fig. 1.6).**

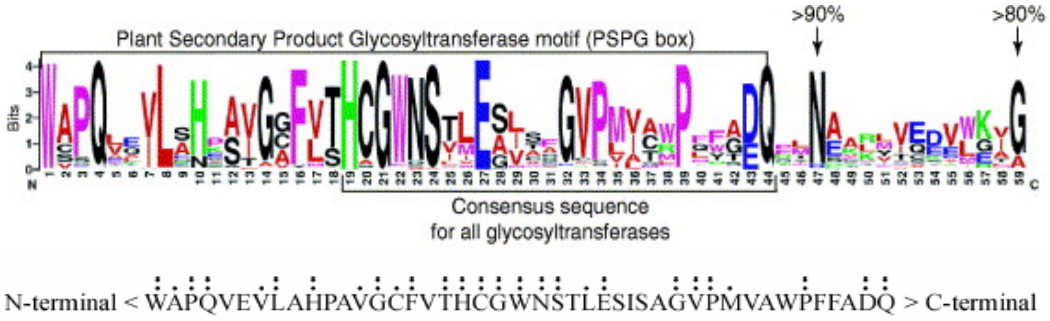


Fig. 1.6 Consensus sequence defining PSPG box

The amino-terminal regions of GTs are more variable and participate in the recognition and binding of diverse acceptors. The highly conserved regions of the PSPG box are WAPQ and HCGWNS residues. The last two amino acids of PSPG box (DQ) put the discrimination between UDP-glucose and UDP-galactose attachment (**Kubo et al. 2004**). Family 1 GTs use uridine diphosphate sugars as the sugar donors. Plants contain a large number of UDP-glycosyltransferases (UGTs) (**Jones and Vogt 2001**). *Arabidopsis thaliana* is predicted to have 107 UGTs and biological functions of some of these have been characterized (**Li et al. 2001**). Characterized UGTs include ones those glycosylate secondary metabolites such as flavonoids and alkaloids **(Jones and Vogt 2001; Bowles et al. 2005**).

2.5 Flavonoids and their glycosyltransferases

In plants, sugar acceptors include all major classes of secondary metabolites, such as phenolics, terpenoids, cyanohydrins and alkaloids (**Vogt and Jones 2000; Jones and Vogt 2001; Bowles et al. 2005)**. One of the most widely studied classes of plant glycosides is the large and heterogenic group of polyphenols. To date, an overwhelming number of polyphenolic glycosides including flavonoid glycosides have been identified.

Flavonoids are ubiquitous plant secondary products that are best known as the characteristic red, blue, and purple anthocyanin pigments of plant tissues (**Winkel-Shirley. 2001**). Flavonoids are classified into six major subgroups that are found in higher plants: the chalcones, flavones, flavonols, flavandiols, anthocyanins, and condensed tannins (or proanthocyanidins); a seventh group, the aurones, is widespread, but not ubiquitous. Some plant species also synthesize specialized forms of flavonoids,

such as the isoflavonoids that are found in legumes and a small number of non-legume plants. Flavonoids have significant activities such as anti-cancer effects and reduced heart disease when ingested by animals. They exhibit a diverse spectrum of biological functions and play an important role in the interaction between plants and their environment. These flavonoids not only protect the plant from the harmful effects of UV irradiation but also play a crucial role in sexual reproduction process. A special class of flavonoid polymers, the tannins, plays a structural role in the plant. Yet other classes of flavonoids, flavonols and anthocyanins, have been implicated in the attraction of pollinators. Certain flavonoids participate in the interaction between plants and other organisms such as symbiotic bacteria and parasites.

Flavonoids are synthesized by the phenylpropanoid metabolic pathway in which the amino acid phenylalanine is used to produce 4-coumaroyl-CoA. This can be combined with malonyl-CoA to yield the true backbone of flavonoids, a group of compounds called chalcones, which contain two phenyl rings. Conjugate ring-closure of chalcones results in the familiar form of flavonoids, the three-ringed structure of a flavone. The metabolic pathway continues through a series of enzymatic modifications to yield Flavonones → dihydroflavonols → anthocyanins (**Fig. 1.7**). Along this pathway, many products can be formed, including the flavonols, flavan-3-ols, proanthocyanidins (tannins) and a host of other various polyphenolics.

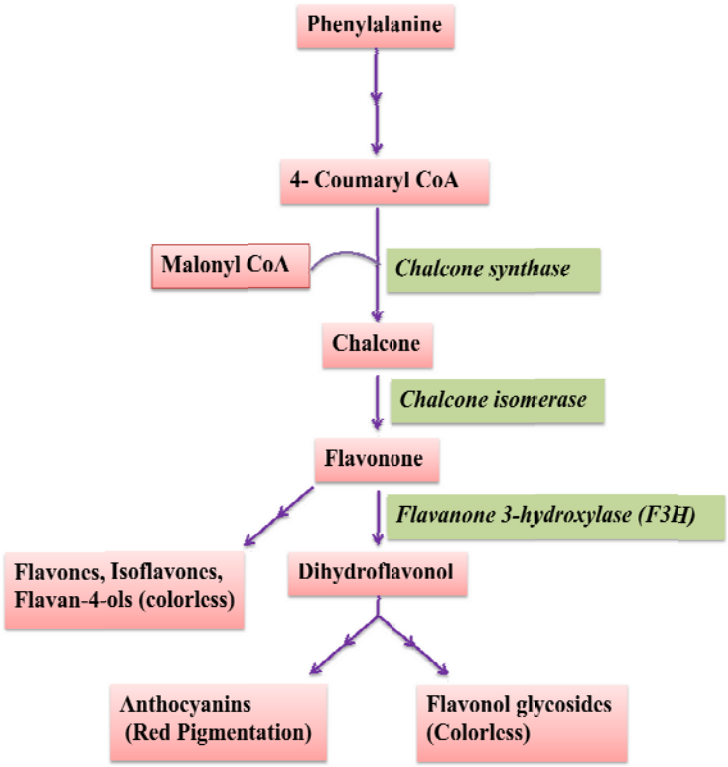


Fig. 1.7 Flavonoid biosynthetic pathway

Flavonoids are widely distributed in plants and fulfill many functions including producing yellow or red/blue pigmentation in flowers and protection from attack by microbes and insects. Flavonoids have been referred to as **"Nature's biological response modifiers"** because of strong experimental evidence of their inherent ability to modify the body's reaction to allergens, viruses, and carcinogens. They show anti-allergic, anti-inflammatory, anti-microbial and anti-cancer activity. Consumers and food manufacturers have become interested in flavonoids for their medicinal properties, especially their potential role in the prevention of cancers and cardiovascular disease. The beneficial effects of fruit, vegetables, and tea or even red wine have been attributed to flavonoids compounds rather than to known nutrients and vitamins. The flavonoids are mainly distributed in many higher plants as the glycosides of glucose, galactose and xylose. The most commonly found sugar residue of flavonoids compounds is glucose, located in the 3-*O*-position of the molecule. Although the functions of these flavonoids glycosides in plant cells remain unclear, they are assumed to represent the accumulation forms of the flavonoids and the defense compounds against potential pathogens such as fungi and bacteria.

Two kinds of structural genes involved in glycosylation of the flavonoids have been isolated from several plants. One encodes UDP-glucose: flavonoids 3-*O*-glucosyltransferase (UF3GT), which catalyzes glycosylations at the 3-*O*-position of flavonoids (**Ford et al. 1998**). The other encodes UDP-rhamnose: anthocyanidins 3-*O*-glucoside rhamnosyltransferase, which adds a rhamnose to the 3-*O*-bond glucose of the anthocyanidins 3-*O*-glucoside molecule to produce the anthocyanidins 3-*O*-rutinosides (**Brugliera et al. 1994; Kroon et al. 1994**). In addition, recently the isolation and characterization of a cDNA clone of UDP-galactose: flavonoids 3-*O*-galactosyltransferase (UF3GaT) from *Vigna mungo* has been reported (**Mato et al. 1998**). Thus, many glycosyltransferases genes have been isolated from various plant sources.

Glycosyltransferases can transfer sugars to either hydroxy group 3 or 7 of flavonols in a regioselective way have been described in a number of plants including *Petunia hybrida*, *Scutellaria baicalensis*, and *Arabidopsis thaliana* (**Hirotani. et al. 2000; Jones et al. 2003; Miller et al. 1999**). However, some glycosyltransferases from *Allium cepa* and *Arabidopsis thaliana* show no strict regioselective and glycosylate the hydroxy groups 3, 7, 3', and 4', and even form diglucosides in some cases **(Kramer et al. 2003; Lim et al. 2004**). In strawberries, flavonoids 3-*O*-glucosyltransferase activity was detected, but the respective protein was not purified **(Given et al. 1988; Halbwirth et**

al. 2006). Although a transferase with broad substrate tolerance glycosylating quercetin at position 7 was purified from strawberry fruit (**Cheng et al. 1994**), little is known about the enzymes that form the complex array of glycosylated flavonoids or their spatial and temporal regulation in *Fragaria*.

2.6 Rationale of the thesis

W. somnifera is a widely used in Ayurvedic medicine, the traditional medical system of India. It is an ingredient in many formulations prescribed for a variety of musculoskeletal conditions (e.g. arthritis, rheumatism) and as a general tonic to increase energy, improves overall health and longevity, and prevents disease in athletes, the elderly, and during pregnancy. Many pharmacological studies have been conducted to investigate the properties of Ashwagandha in an attempt to authenticate its use as a multi-purpose medicinal agent. For example, anti-inflammatory properties have been investigated to validate its use in inflammatory arthritis, and animal stress studies have been performed to investigate its use as an anti-stress agent. Several studies have examined the antitumor and radio sensitizing effect of Ashwagandha.

The major constituents of Ashwagandha are glycosides, alkaloids and steroidal lactones. It was also confirmed by researchers that anti-tumour, anti-inflammatory and anti-oxidant activity was due to the presence of phenolic glucosides. A large number of glucosides have been isolated and characterized from the different parts of this plant. The glucoside formation occurs due to the presence of glycosyltransferases, it has also been concluded by many researchers. In view of the increasing interest on this herb drug, this study involved isolation and characterization of glycosyltransferase gene which is responsible for producing several medicinally important glycosides. So far no studies have been done on glycosyltransferase genes from *W. somnifera.* This study will help in understanding the role of glycosyltransferase and its specificity towards various acceptors as well as donor molecules. The present work deals with the molecular and structural studies of glycosyltransferase gene from *W. somnifera.*

Chapter 2

Isolation, Cloning and Characterization of glycosyltransferase gene from Withania somnifera

1.0 Summary

Most of the flavonoids found in plants exist as glycosides; these secondary metabolites undergo several modification reactions, including glycosylation. This glycosylation of flavonoids is mediated by family 1 UDP – glycosyltransferases (UGT) and has a wide range of effects on flavonoids like solubility, stability and bioavailability. Glycosylation use UDP- sugars, such as UDP- glucose as the glucosyl donor and polyphenols, terpenoids, alkaloids and cyanohydrins as acceptors. Till date there are no reports on functionally characterized flavonoid glycosyltransferases from *Withania somnifera.* This chapter deals with the isolation cloning and characterization of full-length cDNA encoding flavonoid glycosyltransferase from *W. somnifera*. Primers were designed on the basis of sequence information available in NCBI GenBank Database. PCR was done with different set of primers, using cDNA as a template. The partial fragment was cloned and sequenced, which were showing significant similarity with reported glycosyltransferases genes from plants. Gene specific primers were designed and Rapid Amplification of cDNA Ends (RACE) PCR was performed to isolate full-length cNDA clones. The full length sequence reported in this chapter has been deposited in the NCBI gene bank database [Gen Bank: FJ654696]. According to the glycosyltransferase nomenclature guidelines, the systematic name of the *W. somnifera* glycosyltransferase is UGT73A16. The translated protein sequence of UGT73A16 was used to construct a phylogenetic tree with all known plant glycosyltransferases (GTs) deposited in NCBI Gen Bank Database by using Neighbor-joining method.

2.0 Introduction

Among ancient civilizations, India has been known to be rich repository of medicinal plants which produces many types of secondary metabolites such as flavonoids, alkaloids and terpenoids, etc. (**WinkM 1999**). These compounds undergo modification reactions such as methylation, hydroxylation and glycosylation, which lead to the structural diversity of secondary metabolites (**Schwab 2003**). Glycosylation is a well-known modification reaction and is the last step in biosynthesis of natural compounds (**Heller and Forkmann 1994**). It is quantitatively the most significant reaction on earth. Glycosylation of small molecules has key roles in many biological processes, including biosynthesis of various bioactive compounds, regulation of hormone activity, and detoxification of exogenous compounds and metabolism of toxins (**Bowles et al. 2005; Bowles et al. 2006**). Glycosylation of plant secondary products, such as flavonoids, coumarins, terpenoids and cyanohydrins, is generally catalyzed by plant secondary

product glycosyltransferases (PSPGs) (**Noguchi et al. 2007**), which belong to family-1 glycosyltransferases, catalyzing glycosyl transfer from nucleoside diphosphate - activated sugars (donor) to aglycon substrate (acceptor) molecules. The activated sugar form is typically UDP-glucose, but UDP-galactose, UDP-glucuronide, UDP-xylulose and UDP-rhamnose may also be found. In plants, sugar acceptors include all major class of secondary metabolites, such as phenolics, terpenoids, cyanohydrins and alkaloids etc. **(Bowles et al. 2005; Jones and Vogt 2001; Vogt and Jones 2000**).

The primary objective of present study was to isolate and clone a gene which can carry out the glycosylation of a wide variety of natural compounds. We chose *W. somnifera* as the plant source which is mostly used in Indian ayurvedic medicine. Till date there are no reports on functionally characterized flavonoid glycosyltransferases from this plant. In this chapter we report the isolation and characterization of a flavonoids glycosyltransferase gene from *W. somnifera*.

3.0 Materials

3.1 Glasswares and their preparation

Glasswares used during these studies were purchased from Borosil, India. Test tubes (25 mm x 150 mm), glass bottles (70 mm x 125 mm), conical flasks (250, 500, 1000, 2000 and 5000 mL capacity), measuring cylinders (10, 50, 100, 500, 1000 and 2000 mL), beakers (250, 500, 1000 and 2000 mL), petri-plates and funnels etc. were used during course of study. All the glasswares 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. Autoclaving of the glassware was done when mandatory at 121 ºC and 15 psi for 1 h.

3.2 Plasticwares

Sterile petridishes (55 mm and 85 mm diameter) were procured from "Axygen" and "Tarsons", India. Disposable filter sterilization units (0.22 µM) were purchased from Millipore (USA). Centrifuge tubes (1.5 mL and 2 mL capacity), microtips (10, 200, 1000 and 5000 µL capacity) and PCR tubes (0.2 mL capacity, flat cap) were obtained from "Tarsons" and "Axygen", India.

3.3 Chemicals

TRI Reagent and AMV Reverse transcriptase were purchased from Sigma (USA) and Promega (USA) respectively. PCR primers were purchased from MWG, Bangalore. PCR buffer and Taq-DNA polymerase was from Bangalore Genei, India. DNA rulers

were purchased from Bangalore Genei, India. dNTPs were from Sigma. Agarose was from Bioworld (USA), Gel elution kit was from Axygen, USA. Restriction enzymes, pGEM T Easy vector, T4 DNA ligase and RNase were from Promega (USA). Chemicals were purchased from Sigma-Aldrich (USA), Bioworld (USA), Merck, SD fine chemicals and Hi-Media.

3.4 Vectors and bacterial strains

Table 2.0

Vector	**Source**	**Features**
pGEM-T Easy	Promega, USA	T/A cloning vector
Bacterial Strains	**Source**	**Genotype**
E. coli XL-10 gold	Stratagene, USA	*Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr].*
XL1-Blue MRF′ strain	Stratagene, USA	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]*

3.5 Equipments used during the study

Table 2.1

S.No	**Equipment**	**Make**
1	Centrifuge	Sorvall/Haereus/Eppendorf
2	Gel Documentation system	Bio-Rad
3	Thermo Cycler PCR machine	Bio-Rad
5	Spectrophotometer	Perkin Elmer
7	Speed Vac concentrator	Eppendorf
8	Water purification system	TKA (Milli RO/Milli Q)
9	NanoVue (Nanodrop)	GE Healthcare

3.6 Plant Material

Withania somnifera (Solanaceae) plant was collected from the Department of Horticulture, National Chemical Laboratory, Pune, Maharashtra.

3.7 Buffers and solutions

3.7.1 Buffers and solutions for agarose gel electrophoresis

Table 2.2

Name	Components	Preparation and Storage
50 X TAE	2 M Tris Acetic acid (242 g Tris-base and 57.1 mL Glacial acetic acid for 1L) 0.05 M EDTA	pH -8.0 and stored at room temperature.
10 X TBE buffer	890 mM Tris base 890 mM Boric acid 20 mM EDTA (pH-8.0)	Room temperature
DNA loading buffer	0.25% (w/v) Xylencyanol 0.25% (w/v) Bromophenol blue 40% Sucrose 40 mM EDTA (pH 8.0)	The solutions were filter sterilized using 0.22 micron (μ) filter and stored at 2-8 ^{o}C

3.7.2 Different buffers and media used for bacterial studies

Table 2.3

Name	Components	Preparations and Storage
Luria Bertani Broth (LB)	1% Tryptone 0.5% Yeast extract 1% NaCl	pH adjusted to 7.0 with NaOH, stored at room temperature or at 4 ºC after autoclaving
SOB media	2% Bactotryptone 0.5% Yeast extract 10 mM NaCl 10 mM $MgCl_2.6H_2O$ 2 mM KCl	pH adjusted to 6.8 with NaOH, stored at room temperature or at 4 ºC after autoclaving
TB buffer	10 mM PIPES 15 mM $CaCl_2$ 250 mM KCl	pH was adjusted 6.8 with KOH. $MnCl_2$ was added to final concentration of 55 mM and filter sterilized. Stored at -20 ^{o}C

3.7.3 Stock solutions for bacterial transformation and selection

Table 2.4

Name	Components	Preparation and Storage
IPTG solution	200 mg/mL in SMQ	Filter sterilized and stored 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	Filter sterilized and stored at -20 ºC
Tetracycline	12.5 mg/mL in 70% ethanol	Filter sterilized and stored at -20 ºC

3.7.4 Buffers and solutions for plasmid DNA isolation (Alkaline lysis method)

Table 2.5

Name	Components	Preparation and Storage
Solution I (GTE Buffer)	50 mM Glucose 25 mM Tris-HCl (pH 8.0) 10 mM EDTA (pH 8.0)	Autoclaved for 10 min and Stored at 4 ºC
Solution II	0.2 N NaOH 1% SDS	Freshly prepared
Solution III	3 M Potassium acetate (pH-5.5)	Stored at 4 ºC
RNase A	10 mg/mL	Stored at -20 ºC
Other solutions or Reagents	Chloroform Absolute ethanol 3.0 M Sodium acetate 70% ethanol Deionized sterile water	Room temperature

4.0 Methods

4.1 Total RNA isolation and spectrophotometric characterization

RNase free environment was created and maintained as described by Blumberg (**Blumberg 1987**). All glass and plasticwares 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. All materials were dried in a vacuum oven. The total RNA was isolated from *W. somnifera* leaves. 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 of TRIzol reagent was added and mixed thoroughly using a vortex. Chloroform: isoamyl alcohol (300 µL) was added and mixed thoroughly using vortex. 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% TAE agarose gel.

4.2 Synthesis of cDNA first strand by reverse transcription

Complementary DNA (cDNA) was synthesized using total RNA in a reaction catalyzed by the enzyme, reverse transcriptase. The reactions were set up as per the manufacturer's guidelines. Reverse-Transcription PCR (RT-PCR) was performed on 1 µg of total RNA using AMV reverse transcription system (Promega, USA) with oligo (dT) primers in a 20 µL reaction volume. 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 and reaction was incubated at 42 ºC for up to 1-1.5 h. The reaction was terminated by enzyme denaturation at 70 ^{o}C for 15 min. The cDNA synthesized was diluted as per requirement and stored at -20 ^{o}C for further use and directly added to amplification reactions.

4.3 Partial cDNA fragment amplification

4.3.1 Primer design

The multiple sets of primer were designed and synthesized by using conserved domains of GTs from NCBI (Accession No.U32644 and AB360626). Two oligonucleotide primers, SBF1 (5′ATGATGCAAGAACCACTAGA 3′) and SBR1 (5′ GTTGAATTCCAACCACA 3′) were used to amplify a partial fragment of glycosyltransferase. Nucleotide and amino acid sequence analysis was done using software pDRAW 32, ClustalX 2.0 and online bioinformatics analysis facility available at www.justbio.com, www.expasy.org and www.ncbi.nlm.nih.gov.

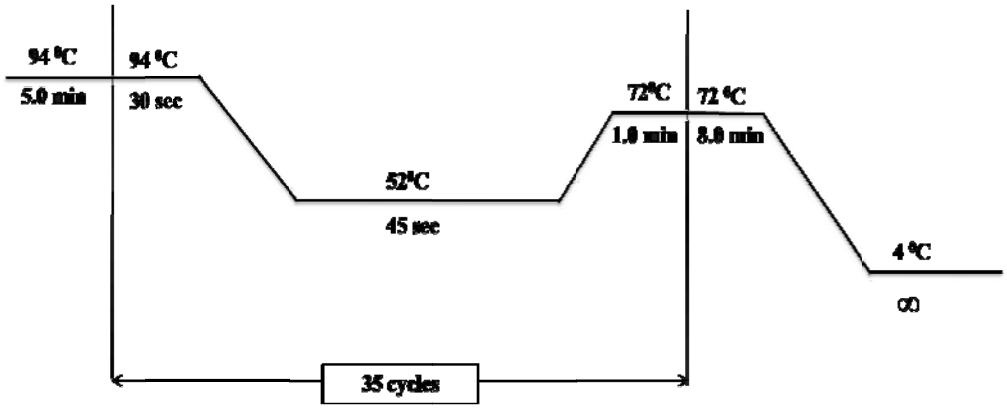
4.3.2 Polymerase Chain Reaction (PCR)

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 **(Arnheim and Erlich 1992; Mullis 1990; Mullis and Faloona 1987; Saiki et al. 1985; Saiki et al. 1988**). This technique has been successfully used for various purposes like amplifications of gene(s) from genomic DNA or from cDNA population, 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, 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 generalized cycling conditions used were as follows:

4.3.2.1 Reaction mixture

Reagent and concentration	**Volume**
Sterile deionized water	6.2 µL
Template (100 ng/ µL)	1.0 µL
Forward primer (10 pmol/ µL)	1.0 µL
Reverse primer (10 pmol/ µL)	1.0 µL
dNTP mix (0.2 mM)	4.0 µL
10 X Buffer (Mg^{+2} 1.5 mM)	1.5 µL
Taq DNA Polymerase (1 U/µL)	0.3 µL
Total volume	**15.0 µL**

4.3.2.2 PCR cycle conditions



#- Annealing temperature depends on Tm of primers * Extension time 1min/kb

4.3.2.3 Extraction and purification of DNA from agarose gels

PCR product was recovered from agarose gel by using Gel elution kit (AxygenTM GEL elution kit, Biosciences, USA.) as per manual instructions. The eluted PCR product was stored at -20 ^{0}C and was used for subsequent reactions such as ligation.

4.3.2.4. Ligation

The gel-eluted PCR amplified fragment was ligated into pGEM-T® EASY vector (Promega, USA.) as given in the protocol below and kept at 16 ^{0}C for overnight.

pGEM -T® EASY vector (50 ng/ µL)	0.5 µL
Amplicon DNA fragment (10 ng/ µL)	4.0 µL
Ligase buffer (2 X)	5.0 µL
T4 DNA ligase	0.5 µL

4.3.3 Bacterial transformation

4.3.3.1 Bacterial culture conditions

E. coli (**Table 2.0**) was grown at 37 ºC with shaking at 200 rpm in Luria Bertani (LB) broth/SOB broth (**Table 2.3**) and maintained on LB/SOB plates with 1.5% agar (**Sambrook et al. 1989**). Recombinant *E. coli* was grown in LB medium supplemented with appropriate antibiotic (s) used for plasmid DNA preparation.

4.3.3.2 Preparation of competent cells using TB buffer

A single colony of *E. coli* XL10 Gold (**Table 2.0**) was inoculated in 5 mL of LB medium containing tetracycline (12.5 mg/L) and grown overnight at 37 ^{o}C in incubator shaker at 200 rpm. Overnight grown culture (500 µL) was added to 50 mL of SOB medium and grown for 2 h at 37 ^{o}C in incubator shaker at 200 rpm (until OD600 reaches 0.5-0.7). Cells were kept on ice for 10 min and harvested by centrifugation at 5,000 rpm for 10 min at 4 ^{o}C.

The cell pellet was suspended in 10 mL ice-cold TB buffer and kept on ice for 30 min. Centrifuged for 10 min at 5,000 rpm and 4 ^{o}C. Pellet was resuspended in 5 mL TB buffer containing 7% DMSO. This was then dispensed into aliquots of 100 µL in sterile 1.5 mL eppendorf tubes, frozen in liquid nitrogen and stored at -80 ^{o}C.

4.3.3.3 Transformation and selections

The competent *E. coli* cells were transformed according to (**Sambrook et al. 1989**). Competent cells were thawed on ice and DNA (~ 50 ng /or ligation mixtures) was added gently to the (100 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. LB broth (800 µL) was added and further incubated at 37 ºC for 1 h in shaker incubator. Cells were harvested by centrifugation at 5,000 rpm for 10 min. The supernatant media was discarded and pellet was dissolved in 100 µL of LB broth. The resuspended cells were spread on LB agar medium with appropriate antibiotics, IPTG and X-gal (**Table 2.4**) as per need (**Sambrook et al. 1989**).

4.3.3.4 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 templates 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 tube containing 20 µL of sterile miliQ water. The microtip was agitated in the water to suspend the bacterial cells. Bacterial lysis was done by placing the suspension on dry bath at temperature 95 ^{o}C for 5 min. After brief spin 2–3 µL of supernatant 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. The resulting PCR products were checked on 1% agarose gel for the presence of the cloned gene of interest**.**

4.3.3.5 Plasmid isolation

The alkaline lysis method (**Table 2.5**) (**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 by vigorous pipetting, 200 µL of Solution II 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 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. 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.

4.3.3.6 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.

4.3.3.7 Cryopreservation of bacterial culture

Bacterial culture harboring plasmid with target gene was preserved for further use in glycerol at ultra low temperature. In a microfuge tube (1.5 mL) culture with 20% sterile glycerol were mixed thoroughly by pipeting and frozen in liquid nitrogen and stored at -80 ^{o}C. The stocks were revived periodically (~6 months) and fresh stocks were prepared.

4.4 Rapid amplification of cDNA ends (RACE)

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. SMART RACE cDNA Amplification Kit (BD Biosciences, Clontech, USA) was used to

perform RACE and the reactions were set up as per the manufacturer's guidelines with minor modifications. RACE PCR was used to isolate full-length WsUGT gene which was further cloned in pGEM-T Easy vector and sequenced.

4.5 Full-length genes ORF amplification by PCR

The sequence information of 5'and 3' RACE PCR product clones were used for designing of primers from start and stop codon to obtain full-length of the genes. PCR was performed with Hi-fidelity *pfx Taq* DNA polymerase (Invitrogen) by using cDNA as a template. Full-length amplicons of WsUGT were cloned in pGEM-T Easy vector and sequenced.

4.6 Phylogenetic tree construction

Nucleotide and amino acid sequence analysis was done using software pDRAW 32, ClustalX 2.0 and online bioinformatics analysis facility available at www.justbio.com, www.expasy.org and www.ncbi.nlm.nih.gov. Various physical and chemical parameters of proteins were predicted by using ExPasy tool ProtParam (http://web.expasy.org/protparam/). Multiple sequence alignments of the amino acid sequences were carried out with the ClustalW 2.0 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 500 replicates is taken to represent the evolutionary history of the taxa analyzed (**Felsenstein 1985**). 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 analysis was conducted in MEGA4.0.2 (**Tamura et al. 2007**).

5.0 Results and discussion

5.1 Total RNA isolation and cDNA synthesis

Total RNA was isolated from aerial part of *W. somnifera* plant and cDNA synthesis was done using protocol as mentioned above in materials and methods.

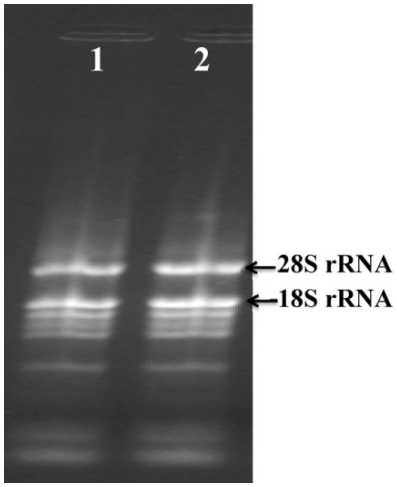


Fig. 2.1 Total RNA isolated from *W. somnifera* leaf (Lanes 1 & 2) resolved on 1% denaturing agarose gel

5.2 PCR based approach for the isolation of glycosyltransferase gene

5.2.1 Multiple sequence alignment of glycosyltransferase genes

Nucleotide sequences of glycosyltransferases (GTs) [*Nicotiana tabacum* (Accession No.U32644) and *Lycium barbarum* (AB360626)] available at NCBI GenBank database were aligned using Clustal W program. Among several set of primers, two oligonucleotide primers (**Fig. 2.2**, Red colour), were synthesized and used to amplify a partial fragment of glycosyltransferase.

CLUSTAL 2.1 multiple sequence alignment

```
Nicotiana       ATGGGTCAGCTCCATATTTTCTTCTTTCCTGTGATGGCTCATGGCCACATGATTCCTACA 60
L.barbarum      ATGGGTCAGCTCCATTTTTTCCTCTTTCCCATGATGGCTCAAGGCCACATGATTCCTACA 60
                ***************:***** ******* .**********:******************

Nicotiana       CTAGACATGGCGAAGCTCTTTGCTTCACGTGGTGTTAAGGCCACTATAATCACAACCCCA120
L.barbarum      CTTGACATGGCCAAGCTCATCGCTTCTCGTGGTGTTAAGGCCACTATAATCACTACCCCT120
                **:******** ******:* *****:**************************:*****:

Nicotiana       CTCAATGAATTCGTTTTCTCCAAAGCTATTCAAAGAAACAAGCATTTGGGTATCGAAATC180
L.barbarum      CTCAATGAATCCGTTTTCTCCAAAGCAATTCAAAGAAACAAACAGTTGGGTATCGAAATC180
                ********** ***************:**************.** ***************

Nicotiana       GAAATC------CGTTTGATCAAATTCCCAGCTGTTGAAAACGGCTTACCTGAAGAATGC234
L.barbarum      GAAATCGAAATCCGTTTGATAAAATTCCCAGCTTTGGAGAATGACTTGCCTGAAGATTGC240
                ******      ********.************ * **.** *.***.********:***

Nicotiana       GAACGCCTCGATCAAATCCCTTCAGATGAGAAGCTCCCAAACTTTTTCAAAGCTGTAGCT294
L.barbarum      GAGCGACTTGATCTCATCCCTACTGAAGCCCATCTTCCCAACTTCTTCAAAGCTGCAGCT300
                **.**.** ****:.******:*:**:*. .* ** **.***** ********** ****

Nicotiana       ATGATGCAAGAACCACTAGAACAGCTTATTGAAGAATGTCGCCCCGATTGTCTTATTTCA354
L.barbarum      ATGATGCAAGAACCATTAGAGCAGCTAATTCAAGAATGTCGCCCTGATTGTCTTGTTTCT360
                ************** ****.*****:*** ************ *********.****:
```

```
Nicotiana       GATATGTTCCTTCCTTGGACTACTGATACTGCAGCAAAATTTAACATTCCAAGAATAGTC414
L.barbarum      GATATGTTCCTTCCTTGGACAACTGATACTGCAGCTAAATTTAACATTCCAAGAATTGTT420
                ********************:**************:*********************:**

Nicotiana       TTTCATGGCACAAGCTTCTTTGCTCTTTGTGTTGAGAATAGCGTCAGGCTAAATAAGCCT474
L.barbarum      TTCCATGGTACAAACTACTTTGCCCTTTGTGTTGGAGACAGTATGAGGCGTAATAAGCCT480
                ** ***** ****.**:****** **********...* ** .* **** :*********

Nicotiana       TTCAAGAATGTGTCCTCAGATTCTGAAACTTTTGTTGTACCGGATTTGCCTCACGAAATT534
L.barbarum      TTCAAGAATGTCTCATCTGATTCTGAAACTTTTGTTGTACCGAATTTACCTCATGAAATC540
                ********** **.**:************************.****.***** *****

Nicotiana       AAGCTGACCAGAACCCAGGTGTCTCCGTTTGAGCGATCTGGGGAAGAGACGGCTATGACC594
L.barbarum      AAGCTGACTAGAACTCAGGTGTCTCCGTTTGAGCAATCGGATGAAGAGTCAGTTATGTCT600
                ******** ***** ******************.*** *. ******:*.* ****:*

Nicotiana       CGGATGATAAAAACAGTCAGGGAATCAGATTCAAAGAGCTATGGAGTTGTTTTCAACAGT654
L.barbarum      CGTGTGCTAAAAGAAGTCAGGGAATCGGATTTGAAGAGCTATGGAGTTATCTTCAACAGC660
                ** .**.*****..************.**** .***************.* ********

Nicotiana       TTCTATGAGCTTGAAACAGATTATGTTGAGCATTATACTAAGGTGCTGGGTAGAAGAGCT714
L.barbarum      TTCTATGAACTTGAACCAGATTATGTTGAACATTATACCAAGGTTCTGGGTAGAAAATCT720
                ********.******.*************.******** ***** **********.* **

Nicotiana       TGGGCTATTGGCCCTCTATCGATGTGCAACAGGGACATTGAAGATAAAGCTGAAAGAGGA774
L.barbarum      TGGGATATTGGCCCGCTTTCATTGTGCAACAGGGACATCGAAGATAAAGTTGAAAGAGGG780
                ****.********* **:**.:**************** ********** *********.

Nicotiana       AAGAAATCCTCTATTGATAAACACGAGTGCTTGAAATGGCTTGATTCGAAGAAACCAAGT834
L.barbarum      AAGAAATCCTCTATTGATAAACACGAGTGCTTGAAATGGCTTGATTCAAAGAAATCAAGT840
                **********************************************.****** *****

Nicotiana       TCCGTCGTTTACATTTGTTTTGGAAGCGTAGCGAATTTCACTGCATCACAACTGCACGAA894
L.barbarum      TCCATCGTTTACATTTGCTTTGGAAGTGTTGCAATTTTCACTGCATCCCAAATGCAAGAA900
                ***.************* ******** **:**.*:************.***.****.***

Nicotiana       CTTGCTATGGGAGTTGAAGCTTCCGGACAAGAATTCATTTGGGTTGTTAGAACAGAACTA954
L.barbarum      CTTGCTATGGGACTTGAAGTTTCTGGACAAGATTTCATTTGGGCTGTTAGAACAG-----955
                ************ ****** *** ********:********** ***********

Nicotiana       GACAACGAAGATTGGTTGCCTGAAGGATTCGAGGAAAGAACGAAAGAGAAAGGTTTAATA 1014
L.barbarum      -ACAACGAAGAGTGGCTGCCTGAAGGATTCGAGGAAAGAACGAAAGAAAAAGGATTAATA 1014
                 ********** *** ******************************.*****:******

Nicotiana       ATAAGAGGATGGGCACCCCAAGTACTAATTCTTGATCACGAATCTGTGGGAGCTTTTGTT 1074
L.barbarum      ATAAGAGGATGGGCACCCCAATTGCTAATTCTTGATCACCAAGCTGTAGGAGCTTTCGTT 1074
                ********************* *.*************** ** ****.******** ***

Nicotiana       ACACATTGTGGTTGGAATTCAACACTAGAAGGAGTTTCAGGAGGGGTTCCAATGGTAACA 1134
L.barbarum      ACTCATTGTGGATGGAATTCAACGCTAGAAGGAATATCAGCAGGGGTGCCAATGGTGACG 1134
                **:********:***********.*********.*:**** ****** ********.**.

Nicotiana       TGGCCTGTATTTGCTGAGCAATTTTTCAATGAGAAGTTAGTGACTGAGGTTTTGAAAACT 1194
L.barbarum      TGGCCTTTGTTTGCTGAGCAATTTTTTAATGAAAAGTTGGTGACTGAAGTTTTGAGAAAT 1194
                ****** *.***************** *****.*****.********.********.**.*

Nicotiana       GGAGCTGGTGTTGGTTCGATACAATGGAAGAGATCAGCTAGTGAAGGAGTGAAAAGAGAA 1254
L.barbarum      GGGGTTGGTGTTGGTTCAGTGCAATGGCAGGCAACAGCTTGTGAAGGAGTGAAAAGAGAA 1254
                **.* ************..*.******.**. *:******:********************

Nicotiana       GCAATAGCTAAGGCAATAAAGAGAGTAATGGTGAGTGAAGAAGCAGATGGATTCAGAAAC 1314
L.barbarum      GAAATAGCTAAAGCAATAAGGAGAGTAATGGTG---GATGAAGCAAAGGAATTCAGAAAC 1311
                *.*********.*******.*************   **:******.* *.**********

Nicotiana       AGAGCTAAAGCGTATAAGGAGATGGCAAGAAAGGCTATTGAAGAAGGAGGGTCATCTTAC 1374
L.barbarum      AGAGCTAAAGAGTACAAGGAAATGGCTAAGAAGGCTGTTGACGAGGGAGGATCATCTTAC 1371
                **********.*** *****.*****:*..*******.****.**.******.*********

Nicotiana       ACTGGATTGACTACTTTGTTGGAAGATATAAGTACATATAGTTCCACTGGTCATTAA 1431
L.barbarum      ACTGGTTTGACTACTTTGCTGAAAGATATAAGTACGTATAGTTTTACAAGTGATTAG 1428
                *****:************ **.*************.*******  **:.** ****.
```

Fig. 2.2 Nucleotide sequence alignment of glycosyltransferases (GTs) of *Nicotiana tabacum* (Accession No.U32644) and *Lycium barbarum* (AB360626). Highlighted regions were considered for primer synthesis

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

Among the several set of primers one primer set, i.e. SB F1/R1 (section 5.2.1) was used to amplify ~800 bp fragment of GT (**Fig. 2.4 A**) gene using first strand cDNA as a template.

The fragment is cloned in pGEM-T easy vector (**Fig. 2.3**) and confirmed by restriction analysis (**Fig. 2.4 B**). This partial amplicon will be referred to as WsUGT in the subsequent discussion.

5.2.2.1 Primers used for PCR amplification

SBF1: 5′ATG ATG CAA GAA CCA CTA GA 3′
SBR1: 5′ GTT GAA TTC CAA CCA CA 3′

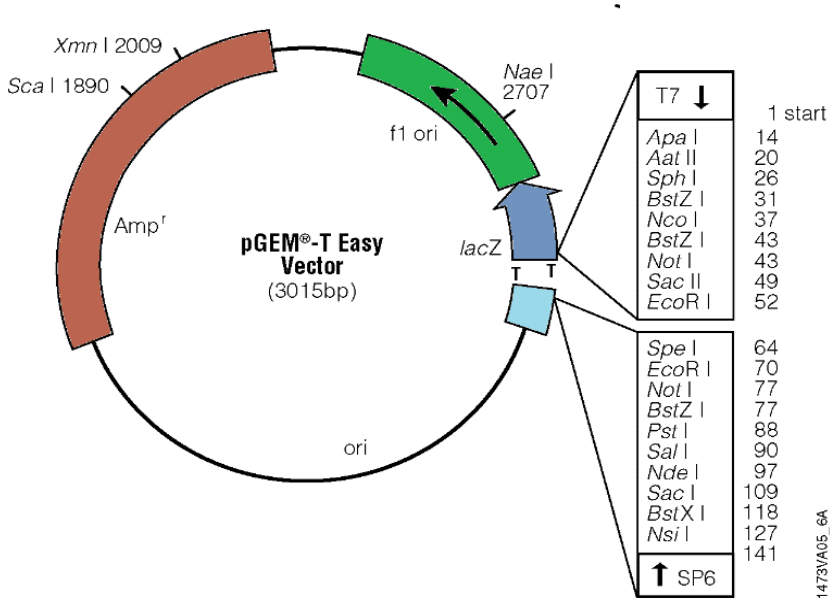


Fig. 2.3 Map of pGEM -T® Easy Vector

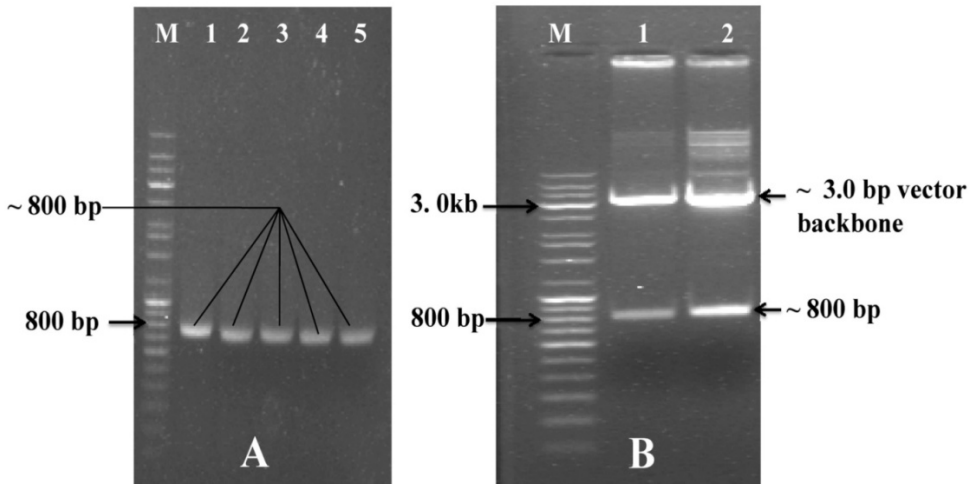


Fig. 2.4 (A) PCR amplification product of ~800 bp (SB F1/R1) fragment separated on 1% agarose gel. Lane 1-5: Amplified product and lane M- Medium range DNA ruler (**B)** Restriction digestion with *Eco*RI. Lane 1 & 2- Showing vector backbone (~3 kb) pGEM-T Easy and release of insert (~800 bp), Lane M- Medium range DNA ruler

5.2.3 WsUGT partial nucleotide sequence

After confirmation by restriction digestion with *Eco*RI, one of the positive clones was sequenced (**Fig. 2.5**) and it revealed that WsUGT fragment was 797 bp showing maximum homology with other plant GTs.

```
  1 ATGATGCAAG AACCACTAGA GCAGTTATTC AAGAAGTCGC CCCAATTGTC
 51 TGGTTTTTGA CATGTTTCTT TCCTTGGACA ACTGACACTG CTGCCAAATT
101 TAACATACCA AGAATTGTTT TCCACGGCAC AAGCTATTTT TGCCCTCTCT
151 GCCGTAGACA GCCTCAGGCT TAATAAGCCT TTCAAGAATG TTTCCTCTGA
201 TTTTGAAACT TTTGTGGTAC CGAACTTGCC GCATGAAATC AAGTTGTTCA
251 GATCGATGTT GTCTCCATTT GAGCAATTTG ATGAAGAGTC AGTTATGTCT
301 CAGATGGTAA AAGCTGTCAG GGATGCGGAT TCAAAGAGTT ATGGAGTTAT
351 CTTCAACAGC TTCTATGAGC TTGAACCAGA TTATGTGGAA CATTATACCA
401 AGGTTCTGGG CAGAAAAAAT TGGGCTATTG GCCCTCTTTC CCTGTGCAAC
451 AGGGACATTG AAGATAAAGC TGAAAGAGGG AAGAAATCCT CTATCGATAA
501 ACACGAATGT TTAAAATGGC TTGATTCAAA GAAATCAAGT TCCATTGTTT
551 ACGTTTGTTT TGGAAGTGTA GCAAATTTCA CCACATCGCA GTTGCAAGAA
601 CTTGCTTTGG GACTTGAAGC TTCTGGACAA GATTTCATTT GGGTTGTTAG
651 AACAGACAAT GAAGATTGGT TGCCTAAAGG ATTCGAGGAA AGGACGAAAG
701 GAATAGGATT AATCATAAGA GGATGGGCAC CCCAAGTGCT GATTCTTGAT
751 CACGAATCTG TTGGAGCTTT TGTGACTCAT TGTGGTTGGA ATTCAAC
```

Fig. 2.5 797 bp partial WsUGT nucleotide sequence. Forward and reverse primers are highlighted in red colour. Gene specific forward & reverse primer are highlighted in green & yellow respectively

5.3 Rapid amplification of cDNA ends (RACE)

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. The limitation of RACE PCR is partial sequence of the gene must be known to design gene specific primers. Based on the sequence information from partial fragment of WsUGT, primers for 5' RACE and 3'RACE were designed (**Fig. 2.5** highlighted in green & yellow color).

5.3.1 Primers for 5' RACE and 3' RACE

SACR – AAG CCT GAG GCT GTC TAC GGC
SACNR- GAG GCT GTC TAC GGC AGA GAG
FLONF- AAT TGG GCT ATT GGC CCG CTT TCC CTG TGC AAC
FLONNF- GCT TGA CTC AAA GAA ACC AAG TTC CAT TGT

Primers supplied with kit:

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

5.3.2 5' RACE PCR

In order to get the 5' region of WsUGT gene, 5' Rapid amplification of cDNA ends (RACE) PCR was performed as per manual instructions. Primary PCR was done with SACR & 5' UPM primers and 5'RACE ready cDNA as a template. The primary PCR product was diluted (1:50) and used as a template for secondary PCR with primers SACNR and NUP. Approximately 500 bp secondary PCR product (**Fig. 2.6 A**) was cloned in pGEM-T Easy vector, confirmed by *Eco*RI digestion (**Fig. 2.6 B**) and sequenced (**Fig. 2.7**).

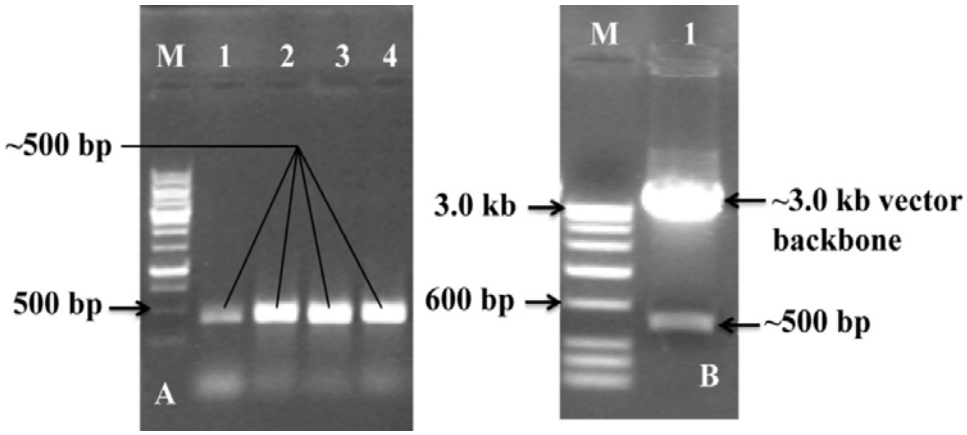


Fig. 5.6 **(A)** 1.5% agarose gel showing ~ 500 bp amplification (Lane1-4), Lane M- 1.0 kb DNA ruler **(B)** Restriction digestion with *Eco*RI. Lane 1 showing vector backbone (~3 kb) and of inserts (~ 500 bp), Lane M- Low range DNA ruler

The sequencing analysis of 5' RACE product revealed that it contains 13 bp UTR region, which is shown in red and the start codon (ATG) is highlighted in green color (**Fig. 2.7**).

```
  1 AAGCAGTGGT ATCAACGCAG AGTGAAAAAG AGATTCATGG GTCAGCTCCA
 51 TATTTTCTTC TTTCCCATGA TGGCTCAAGG TCACATGATT CCTACACTAG
101 ACATGGCCAA GCTCGTCGCT TCTCGTGGTG TGAAAGCCAC TATAATCACT
151 ACCCCTCTCA ATGAATCTGT TTTCTCCAAA GTTATTCAAA GAAACAAGAA
201 TTTGGGGATC CGATTAATCA AATTCCCAGC TGTGGAGAAC GACTTGCCAG
251 AAGATTGCGA GCGTCTTGAT CTCATTCCTT CTGATGACAA ACTCCCAAAC
301 TTCTTCAAAG CTGCAGCTAC GATGCAAGAA TCACTAGAGC AGCTTATTCA
351 AGAATGTCGC CCCAATTGTC TGGTTTCTGA CATGTTCTTT CCTTGGACAA
401 CTGACACTGC TGCCAAATTT AACATACCAA GAATTGTTTT CCACGGCACA
451 GGCTATTTTG CCCTCTCTGC CGTAGACAGC CTCAGGCTT
```

Fig. 2.7 Sequence analysis of 5' RACE PCR product of WsUGT gene. Underlined- 5' NUP sequence; 5' UTR (13 bp) shown in red colour; Start codon (ATG) highlighted with green.

5.3.3 3'RACE

To isolate 3' end and UTR of the gene two steps RACE PCR was done as per manual instruction. Primary PCR was done with gene specific primer FLONF and UPM.

The primary PCR product was diluted (1:50) and used as a template for secondary PCR with primers FLONNF and NUP. Nested PCR yielded an amplicon of ~900 bp (**Fig. 2.8 A**) which was cloned in pGEM-T Easy vector, confirmed by restriction digestion (**Fig. 2.8 B**) and sequencing (**Fig. 2.9**).

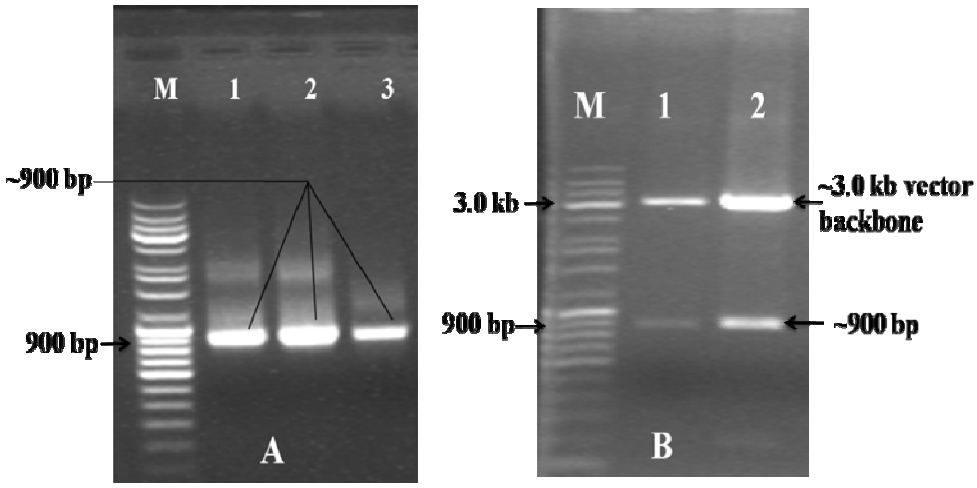


Fig. 2.8 (A) 1% agarose gel showing amplification of ~ 900 bp 3' RACE PCR product Lane M- Medium range DNA ruler, Lane 1& 2- ~ 900 bp 3' RACE PCR **(B)** Restriction digestion with *Eco*RI. Lane 1- showing vector backbone (~3 kb) and release of insert (~900 bp), Lane M- Medium range DNA ruler

Sequencing of 3'RACE clone showed stop codon (TAA, green highlighted) 3' UTR (101 bp) highlighted in turquoise color, putative polyadenylation site (Green and underlined) (**Fig. 2.9**).

```
  1 AATTGGGCTA TTGGCCCGCT TTCCCTGTGC AACAGGGATA TTGAAGATAA
 51 AGCTGAAAGA GGGAAGAAAT CCTCTATCGA TAAACACGAA TGTTTAAAAT
101 GGCTTGACTC AAAGAAACCA AGTTCCATTG TTTACGTTTG TTTTGGAAGT
151 GTAGCAAATT TCACCACATC ACAGTTGCAA GAACTTGCTT TGGGACTTGA
201 AGCTTCTGGA CAAGATTTCA TTTGGGTTGT TAGAACATAC AATGAAGATT
251 GGTTGCCTAA AGGATTCGAG GAAAGAACTA AAGGAAAAGG ATTAATCATA
301 AGAGGATGGG CACCCCAAGT GCTGACTCTT GATCACGAAT CTGTTGGAGC
351 TTTTGTGACT CATTGTGGAT GGAATTCAAC TCTAGAAGGA ATATCAGCAG
401 GTGTGCCACT GGTAACATGG CCAGTGTTTG CGGAACAATT TTTGAATGAG
451 AAGTTGGTAA CTGAGATTAT GAGAACTGGC GCTGGCGTTG GTTCAGTGCA
501 ATGGAAGAGA TCCGCTAGCG AAGGAGTGAA ACGAGAAGCG ATTGCAAATG
551 CAATAAAGAG AGTAATGATG AGTGAAGAAG CAGAAGGATT CAGAAATAGA
601 GCTAAAGCGT ATAAGGAAAT GGCAAGACAG GCTATTGAAG AAGGAGGATC
651 TTCTTACAGT GGATTGACTA CTTTGCTACA AGATATAAGT TCATATAGCT
701 CCACAAGTCA TTAATTTAGT AGTGATTGAA AAAACACGTA TGATTTCTAT
751 TCTAATAAAA GTTGCACAGG GAAAGCGGGC CAAATAGCCC CATTTTTTTC
801 TCAATGGTGG AAATT
```

Fig. 2.9 Nucleotide sequence of 3' RACE clone. Stop codon (TAA) highlighted in green, 3'UTR in red and polyadenylation site is in green underlined.

5.4 Amplification of full-length WsUGT cDNA from *W. somnifera*

On the basis of 5' and 3' RACE sequence information, primers were designed from start and stop codon to amplify the ORF of WsUGT gene as a single PCR product.

WSfullF: 5′ ATGGGTCAGCTCCATATTTTCTTC 3′

WSfullR: 5′ TTAATGACCAGTGGAACTATATGTACT 3′

PCR was performed with primers WSfull F/R and cDNA as a template with high fidelity *Taq* DNA polymerase. Approximately 1.4 kb amplicon was cloned in pGEM-T Easy vector and confirmed by restriction digestion (**Fig. 2.10 A**) and sequencing (**Fig. 2.10 B**).

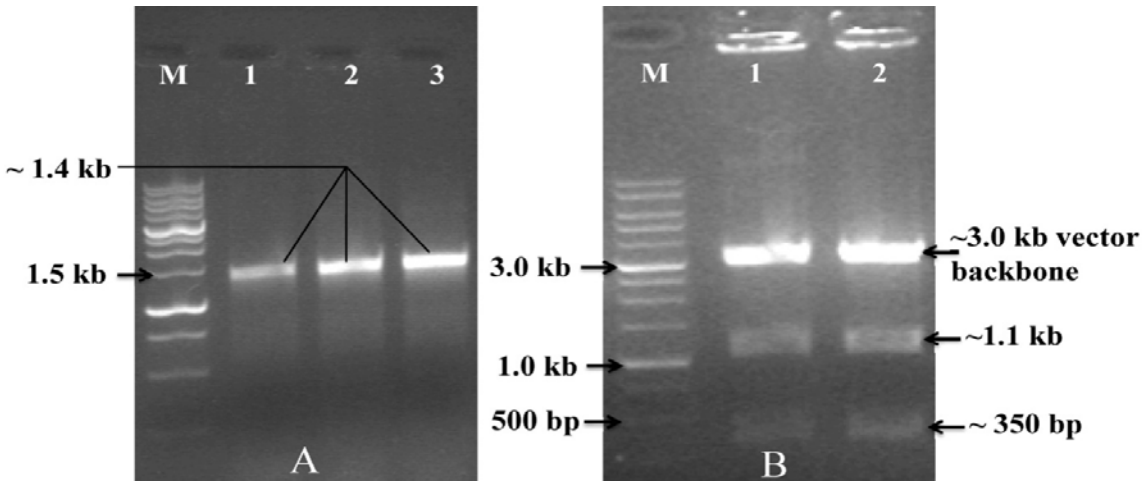


Fig. 2.10 (A) 1% agarose gel showing ~ 1.4 kb amplification (Lane 1-3), Lane M- 1.0 kb DNA ruler **(B)** Restriction digestion with *Eco*RI. Lane 1 & 2- Showing vector backbone (~3 kb) and release of insert (~1.4 kb), Lane M- 1.0 kb DNA ruler

Nucleotide sequence of WsUGT (FJ654696)

```
  1 ATGGGTCAGC TCCATATTTT CTTCTTTCCC ATGATGGCTC AAGGTCACAT
 51 GATTCCTACA CTAGACATGG CCAAGCTCGT CGCTTCTCGT GGTGTGAAAG
101 CCACTATAAT CACTACCCCT CTCAATGAAT CTGTTTTCTC CAAAGTTATT
151 CAAAGAAACA AGAATTTGGG GATCCGATTA ATCAAATTCC CAGCTGTGGA
201 GAACGACTTG CCAGAAGATT GCGAGCGTCT TGATCTCATT CCTTCTGATG
251 ACAAACTCCC AAACTTCTTC AAAGCTGCAG CTACGATGCA AGAATCACTA
301 GAGCAGCTTA TTCAAGAATG TCGCCCCAAT TGTCTGGTTT CTGACATGTT
351 CTTTCCTTGG ACAACTGACA CTGCTGCCAA ATTTAACATA CCAAGAATTG
401 TTTTCCACGG CACAGGCTAT TTTGCCCTCT CTGCCGTAGA CAGCCTCAGG
451 CTTAATAAGC CTTTCAAGAA TGTCTCCTCT GATTCTGAAA CTTTTGTGGT
501 ACCGAATTTG CCGCATGAAA TCAAGCTGAC CAGATCGAAG TTGTCTCCAT
551 TTGAGCAATC TGATGAAGAG TCAGTTATGT CTCAGATGGT AAAAGCTGTC
601 AGGGATGCGG ATTCAAAGAG CTATGGAGTT ATCTTCAACA GCTTCTATGA
651 GCTTGAACCA GATTATGTGG AACATTATAC CAAGGTTCTG GGTAGAAAAA
701 ATTGGGCTAT TGGCCCGCTT TCCCTGTGCA ACAGGGACAT TGAAGATAAA
751 GCTGAAAGAG GGAAGAAATC CTCTATCGAT AAACACGAAT GTTTAAAATG
801 GATTGATTCA AAGAAATCAA GTTCCATTGT TTACGTTTGT TTTGGAAGTG
851 TAGCAAATTT CACCACATCG CAGTTGCAAG AACTTGCTTT GGGACTTGAA
```

```
 901 GCTTCTGGAC AAGATTTCAT TTGGGTTGTT AGAACAGACA ATGAAGATTG
 951 GTTGCCTAAA GGATTCGAGG AAAGGACGAA AGGAAAAGGA TTAATCATAA
1001 GAGGATGGGC ACCCCAAGTG CTGATTCTTG ATCACGAATC TGTTGGAGCT
1051 TTTGTGACTC ATTGTGGATG GAATTCAACG CTAGAAGGAA TATCAGCAGG
1101 CGTGCCACTG GTAACATGGC CAGTGTTTGC GGAACAATTT TTGAATGAGA
1151 AGTTGGTAAC TGAGATTATG AGAACTGGGG CTGCCGTTGG TTCAGTGCAA
1201 TGGAAGAGAT CTGCTAGTGA AGGAGTGAAA AGAGAAGCGA TTGCAAATGC
1251 AATAAAGAGA GTAATGGTGA GTGAAGAAGC AGAAGGATTC AGAAATAGAG
1301 CTAAAGCGTA TAAGGAATTG GCAAGACAAG CTATTGAAGA AGGAGGATCT
1351 TCTTACAGTG GATTGACTAC TTTGCTACAA GATATAAGTA CATATAGTTC
1401 CACTGGTCAT TAA
```

Deduced Amino acid sequence of WsUGT (FJ654696)

```
         10         20         30         40         50         60
MGQLHIFFFP MMAQGHMIPT LDMAKLVASR GVKATIITTP LNESVFSKVI QRNKNLGIRL

         70         80         90        100        110        120
IKFPAVENDL PEDCERLDLI PSDDKLPNFF KAAATMQESL EQLIQECRPN CLVSDMFFPW

        130        140        150        160        170        180
TTDTAAKFNI PRIVFHGTGY FALSAVDSLR LNKPFKNVSS DSETFVVPNL PHEIKLTRSK

        190        200        210        220        230        240
LSPFEQSDEE SVMSQMVKAV RDADSKSYGV IFNSFYELEP DYVEHYTKVL GRKNWAIGPL

        250        260        270        280        290        300
SLCNRDIEDK AERGKKSSID KHECLKWIDS KKSSSIVYVC FGSVANFTTS QLQELALGLE

        310        320        330        340        350        360
ASGQDFIWVV RTDNEDWLPK GFEERTKGKG LIIRGWAPQV LILDHESVGA FVTHCGWNST

        370        380        390        400        410        420
LEGISAGVPL VTWPVFAEQF LNEKLVTEIM RTGAAVGSVQ WKRSASEGVK REAIANAIKR

        430        440        450        460        470
VMVSEEAEGF RNRAKAYKEL ARQAIEEGGS SYSGLTTLLQ DISTYSSTGH
```

Fig. 2.11 Nucleotide and deduced amino acid sequences of WsUGT

5.5 Characterization of WsUGT

The full length cDNA sequence of WsUGT has been submitted to NCBI GenBank database under accession no.FJ654696. The WsUGT sequence contains a 1413 bp open-reading frame, which codes for a protein of 470 amino acids (**Fig. 2.11**). The sequence analysis shows that RACE amplification has been well extended downstream to the stop codon (TAA) and yielded a 101 bp 3'UTR. Also it shows 13 bp 5` UTR region. The 'AATAA' signal sequence was present in the 3'UTR (**Fig. 2.11**) which is present in most eukaryotes and forms a complex with U4 SnRNP for primary cleavage site selection in pre- mRNA.

5.5.1 Amino acid analysis

The deduced amino acid sequence of WsUGT shared 93.1%, 92.5% and 91.0% similarity and identities of 85.5%, 85.5% and 82.6% with *Nicotiana tabacum* (U32643), *Lycium barberum* (AB360612) and *Lycium barberum* (AB360626) respectively.

The calculated molecular mass and predicted pI value of WsUGT were 51.81 kDa and 6.07 respectively. According to the glycosyltransferase nomenclature guidelines (**Mackenzie et al. 2005**), the systematic name of the *W. somnifera* glycosyltransferase (WsUGT) is UGT73A16 and used in following chapters.

Analysis of the deduced amino acid sequence showed it contained conserved PSPG box (44 amino acid), which is found in all plant glycosyltransferases genes (highlighted in yellow color) and participate in binding of donor substrate. Most conserved region of PSPG box i.e, **HCGWNS** (**Fig. 2.11** highlighted & underlined) is also present in the sequence.

5.5.2 Multiple sequence alignment of UGT73A16 with other plant glycosyltransferases and CDD search

The amino acid sequence of UGT73A16 was aligned with UDP-glucose: glycosyltransferase of *Lycium barberum* (AB360612 & AB360626) and immediate-early salicylate-induced glycosyltransferase of *Nicotiana tabacum* (U32643) given below in **Fig. 2.12 A**. Conserved domain database search on NCBI server showed specific hits with UDP- glucose glycosyltransferase of GT-B type superfamily. The members of this family share a common GT-B topology, one of the two protein topologies observed for nucleotide-sugar-dependent glycosyltransferases. GT-B 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 WsUGT (**Fig. 2.12 B**) are shown below.

CLUSTAL 2.0.11 multiple sequence alignment

```
withania        MGQLHIFFFPMMAQGHMIPTLDMAKLVASRGVKATIITTPLNESVFSKVIQRNKNLG--- 57
Nicotiana       MGQLHFFFFPVMAHGHMIPTLDMAKLVASRGVKATIITTPLNESVFSKSIQRNKHLGIEI 60
Lycium10        MGQLHFFLFPMMAQGHMIPTLDMAKLIASRGVKATIITTPLNESVFSKAIQRNKQLGIEI 60
Lycium12        MGQLHFFLFPMMAQGHMIPTLDMAKLIASRGVKATIITTPLNESVFSKAIQRNKQLGIEI 60
                *****:*:**:**:************:******************** *****:**

withania        ---IRLIKFPAVENDLPEDCERLDLIPSDDKLPNFFKAAATMQESLEQLIQECRPNCLVS 114
Nicotiana       --EIRLIKFPAVENGLPEECERLDLIPSDDKLPNFFKAVAMMQEPLEQLIEECRPNCLVS 118
Lycium10        EIEIRLIKFPALENDLPEDCERLDLIPTEAHLPNFFKAAAMMQEPLEQLIQECRPDCLVS 120
Lycium12        EIEIRLIKFPALENDLPEDCERLDLIPTEAHLPNFFKAAAMMQEPLEQLIQECRPDCLVS 120
                   ********:**.***:********:: :*******.* ***.*****:****:****

withania        DMFFPWTTDTAAKFNIPRIVFHGTGYFALSAVDSLRLNKPFKNVSSDSETFVVPNLPHEI 174
Nicotiana       DMFLPWTTDTAAKFNMPRIVFHGTSFFALCVENSIRLNKPFKNVSSDSETFVVPNLPHEI 178
Lycium10        DMFLPWTTDTAAKFNIPRIVFHGTNYFALCVGDSMRRNKPFKNVSSDSETFVVPNLPHEI 180
Lycium12        DMFLPWTTDTAAKFNIPRIVFHGTNYFALCVGDSMRRNKPFKNVSSDSETFVVPNLPHEI 180
                ***:***********:********.:***.. :*:* **********************

withania        KLTRSKLSPFEQSDEESVMSQMVKAVRDADSKSYGVIFNSFYELEPDYVEHYTKVLGRKN 234
Nicotiana       KLTRTQLSPFEQSGEETTMTRMIKSVRESDSKSYGVIFNSFNELEHDYVEHYTKVLGRRA 238
Lycium10        KLTRTQVSPFEQSDEESVMSRVLKEVRESDLKSYGVIFNSFYELEPDYVEHYTKVMGRKS 240
Lycium12        KLTRTQVSPFEQSDEESVMSRVLKEVRESDLKSYGVIFNSFYELEPDYVEHYTKVLGRKS 240
                ****:::******.**:.*::::* **::* ********** *** *********:**:
```

```
withania        WAIGPLSLCNRDIEDKAERGKKSSIDKHECLKWIDSKKSSSIVYVCFGSVANFTTSQLQE 294
Nicotiana       WAIGPLSMCNRDIEDKAERGKQSSIDKHECLKWLDSKKPSSVVYVCFGSVANFTASQLHE 298
Lycium10        WAIGPLSLCNRDVEDKAERGKKSSIDKHECLEWLDSKKPSSIVYVCFGSVANFTVTQMRE 300
Lycium12        WDIGPLSLCNRDIEDKVERGKKSSIDKHECLKWLDSKKSSSIVYICFGSVAIFTASQMQE 300
                * *****:****:***.****:*********:*:****.**:**:****** **.:*::*

withania        LALGLEASGQDFIWVVRT--DNEDWLPKGFEERTKGKGLIIRGWAPQVLILDHESVGAFV 352
Nicotiana       LAMGIEASGQEFIWVVRTELDNEDWLPEGLEERTKEKGLIIRGWAPQVLILDHESVGAFV 358
Lycium10        LALGLEASGLDFIWAVRA--DNEDWLPEGFEERTKEKGLIIRGWAPQVLILDHESVGAFV 358
Lycium12        LAMGLEVSGQDFIWAVRT--DNEEWLPEGFEERTKEKGLIIRGWAPQLLILDHQAVGAFV 358
                **:*:*.** :***.**:  ***:***:*:***** **********:*****::*****

withania        THCGWNSTLEGISAGVPLVTWPVFAEQFLNEKLVTEIMRTGAAVGSVQWKRSASEGVKRE 412
Nicotiana       THCGWNSTLEGVSGGVPMVTWPVFAEQFFNEKLVTEVLKTGAGVGSIQWKRSASEGVKRE 418
Lycium10        THCGWNSTLEGISAGVPMVTWPVFAEQFFNEKLVTQVMRTGAGVGSVQWKRSASEGVEKE 418
Lycium12        THCGWNSTLEGISAGVPMVTWPLFAEQFFNEKLVTEVLRNGVGVGSVQWQATACEGVKRE 418
                **********:*.***:****:*****:******::::.*..***:**: :*.***::*

withania        AIANAIKRVMVSEEAEGFRNRAKAYKELARQAIEEGGSSYSGLTTLLQDISTYSSTGH- 470
Nicotiana       AIAKAIKRVMVSEEAEGFRNRAKAYKEMARKAIEGGGSSYTGLTTLLEDISTYSSTGH- 476
Lycium10        AIAKAIKRVMVSEEAEGFRNRARAYKEMARQAIEEGGSSYTGLTTLLEDISSYESLSSD 477
Lycium12        EIAKAIRRVMV-DEAKEFRNRAKEYKEMAKKAVDEGGSSYTGLTTLLKDISTYS-FTSD 475
                 **:**:**** :**: *****: ***:*::*:: *****:******:***:*.
```

A

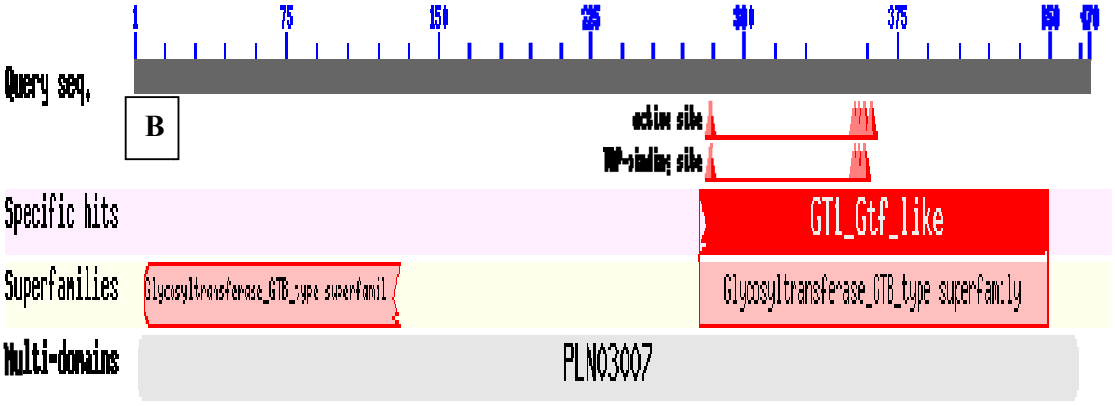


Fig. 2.12 (A) ClustalW alignment of amino acid sequences of *W. somnifera* (UGT73A16) with *Lycium barbarum* & *Nicotiana* **(B)** CDD search on NCBI server

5.5.3 Hydropathy index of UGT73A16

The hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain (**Kyte and Doolittle 1982**). Each amino acid is given a hydrophobicity score between -4.5 and 4.5. A score of 4.5 is the most hydrophobic and a score of -4.5 is the most hydrophilic. The amino acid sequence of UGT73A16 was analyzed using Kyte-Doolittle hydropathy plot of JustBio hosted tool (http://www.justbio.com/index.php?page=plots) at window size 9. Strong negative peaks indicate possible surface regions of globular proteins and score greater than 1.8 indicate possible transmembrane regions. In UGT73A16, one transmembrane region at C-terminal was observed (**Fig. 2.13**).

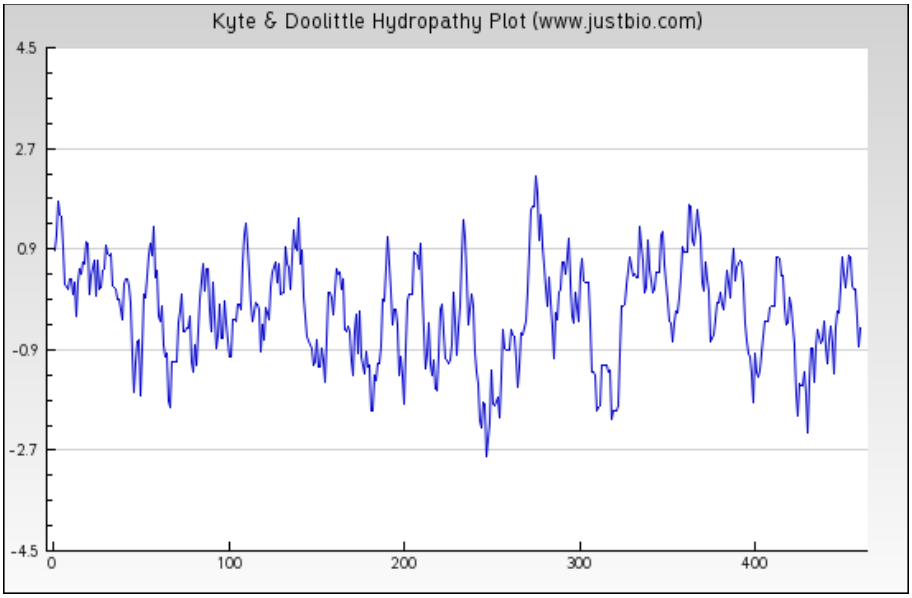


Fig. 2.13 Kyte-Doolittle Hydropathy plot

5.5.4 Phylogenetic tree analysis

The function and specificity of UGT73A16 cannot be fully predicted based on the sequence information alone, thus a Phylogenetic tree was constructed for the UGT73A16 along with the different UGTs by means of the Neighbor-joining method (**Fig. 2.14**). However, the phylogenetic tree of flavonoid UGTs divided in to three clusters; the cluster I contained flavonoid 7-*O*-glycosyltransferase, However, cluster I also contains glycosyltransferase that are capable of salicylic acid induced glycosyltransferase (Accession number: AAK28303) and UDP glucose: flavonoid 3-*O*-glucoside 7-*O*-glucosyltransferase (Accession number: Q9ZQ95). Cluster II composed of UGTs displaying diverse regioselectivity and Cluster III consisted of flavonoid 3-*O*-glycosyltransferase. UGT73A16 was most similar to cluster I, with strong correlation for its regioselectivity. Phylogenetic analysis suggests that the UGT73A16 has diverse specificity and have the potential to glycosylate a wide range of flavonoids including chalcone.

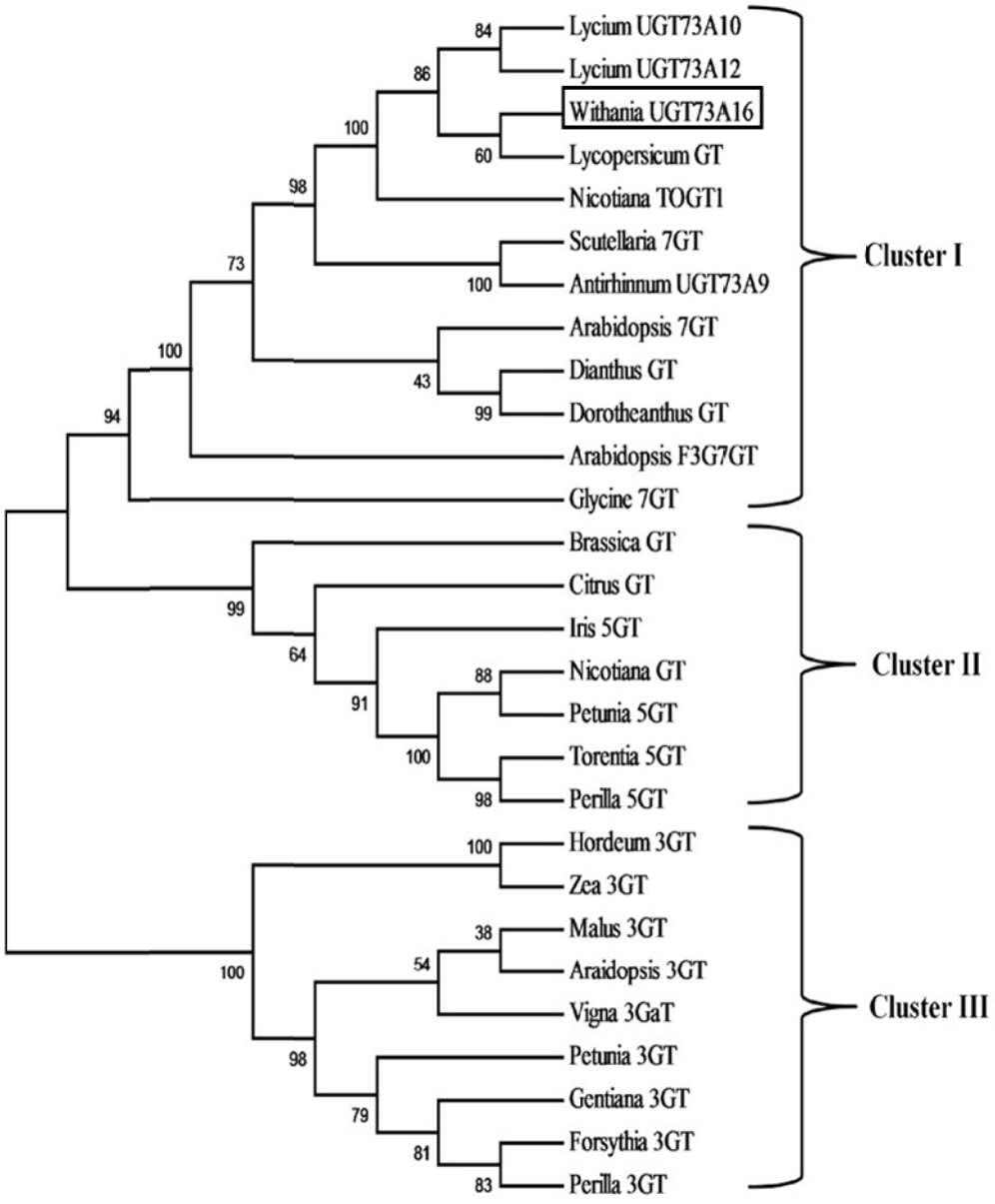


Fig. 2.14 The Phylogenetic tree of UDP-dependent glycosyltransferase, 7-*O*-Glycosyltransferase from cluster I, UGTs showing diverse regioselectivity from cluster II and 3-*O*-Glycosyltransferases from cluster III. Gene bank accession numbers of the UGTs: *Lycium* UGT73A10 (Accession number: AB360612), *Lycium* UGT73A12 (Accession number: AB360626), *Lycopersicum* GT (Accession number: X85138), *Nicotiana* TOGT1 (Accession number: AAK28303), *Scutellaria* 7GT (Accession number: AB031274), *Antirhinnum* UGT73A9 (Accession number:), *Arabidopsis* 7GT (Accession number: AAL90934), *Dianthus* GT (Accession number: BAD52006), *Dorotheanthus* GT (Accession number: CAB56231), *A. thaliana* AtF3G7GT,

(Accession number: Q9ZQ95), *Glycine* 7GT (Accession number: AB292164), *Brassica* GT (Accession number: A62529), *Citrus* GT (Accession number: AB033758), *Iris* 5GT (Accession number: AB113664), *Nicotiana* GT (Accession number: AB072919), *Petunia* 5GT (Accession number: AB027455), *Torentia* 5GT (Accession number: AB076698), *Perilla* 5GT (Accession number: AB013596), *Hordeum* 3GT (Accession number: X15694), *Zea* 3GT (Accession number: AY167672), *Malus* 3GT (Accession number: AF117267), *Arabidopsis* 3GT (Accession number: AAM91139), *Vigna* 3GaT (Accession number: AB009370), *Petunia* 3GT (Accession number: AB027454), *Gentiana* 3GT (Accession number: D85186), *Forsythia* 3GT (Accession number: AF127218), *Perilla* 3GT (Accession number: BAA19659)

6.0 Conclusions

- UGT glycosyltransferase gene was isolated from *Withania somnifera* (WsUGT) using PCR based approach and submitted to NCBI gene bank (**FJ654696**).
- WsUGT gene contains 1527 bp nucleotide sequence with 13 bp of 5' UTR and 101 bp of 3' UTR region and 3' UTR region contains a specific polyadenylation signal sequence AATAA.
- The WsUGT gene showed 87% identity at nucleotide level and 86% identity at amino acid level with the UDP-glucose: glycosyltransferases gene from *Lycium barbarum*.
- Analysis of deduced amino acid sequence of WsUGT gene showed the conserved 44 amino acids domain called as PSPG box which is found to be present in all the plant glycosyltransferase genes.
- Phylogenetic analysis revealed that WsUGT gene is evolutionarily most similar to UDP glucose: glycosyltransferases from *Lycium barbarum* of *Lamiaceae* family.
- According to the glycosyltransferase nomenclature guidelines, the systematic name of the *W. somnifera* (WsUGT) glycosyltransferase is UGT73A16 and to be further used in following chapters.

Chapter 3A

Heterologous expression, Purification and functional characterization of glycosyltransferase gene

1.0 Summary

This chapter deals with cloning of full-length cDNA (UGT73A16) into the expression vector pET 30b (+) and its over-expression in *E. coli* BL21 (DE3). The recombinant UGT73A16 was purified by affinity chromatography using Ni-NTA agarose beads in denatured conditions and analyzed on SDS-PAGE. All the conditions were optimized (IPTG 0.08 mM, temperature 22 ^{o}C and time 20 h) to express the recombinant UGT73A16 in soluble form. The recombinant protein was purified from lysate using affinity chromatography and further used for enzyme assay. The enzyme assay was done using UDP-glucose, UDP-galactose and UDP-glucuronic acid as a donor substrate and flavonoids as acceptor substrate. Product identification was done by hypsochromic shift, HPLC and LC-MS method. HPLC analysis and hypsochromic shift indicated that UGT73A16 transfers a glucose molecule to several different flavonoids but no activity with UDP-galactose. Analysis of reaction products also showed that glycosylation of substrates occurred on the 7- and 3- hydroxyl group of the aglycon. Recombinant UGT73A16 also displayed regiospecific glucosyl transfer activity towards 3- hydroxy flavone (chemically synthesized), which is the backbone of all flavonols, not found naturally. This section also covers the kinetic parameters of recombinant UGT73A16 (Wild) proteins. Based on kinetic parameters, UGT73A16 showed more catalytic efficiency towards naringenin followed by genistein and baicalein. The present study generates essential knowledge and molecular as well as biochemical tools that will allow the verification of UGT73A16 in glycosylation.

2.0 Introduction

Till date, various flavonoid UGT genes have been cloned and characterized (**Hirotani. et al. 2000; Kim et al. 2006; Kramer et al. 2003; Willits et al. 2004**). Characterized UGTs include ones that glycosylate secondary metabolite such as flavonoids, alkaloids, terpenoids and tannins. 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 flavonoid glycosides have been identified. Flavonoids are an important group of polyphenolic natural product. They are typical phytochemicals having an impact on human (**Cornwell et al. 2004; Usha et al. 2005**) and are synthesized via the phenylpropanoid pathway. A wide range of different biological activities, including antibacterial, antithrombotic, vasodilatory, anti-inflammatory, and anti-carcinogenic effects mediated by different mechanism, are associated with flavonoid compounds (**Middleton et al. 2000**). *In vitro* studies indicate

considerable difference in the anti-oxidative potential of different flavonoid subgroups, depending upon their chemical structure, bioavailability, distribution, and metabolism (**Rice-Evans et al. 1996**). In previous chapter (Chapter 2) isolation, cloning and characterization of gene encoding UGT73A16 from *W. somnifera* were described. This chapter deals with heterologous expression of UGT73A16 gene in *E. coli*, its purification and functional characterization.

3. 0 Materials

3.1 Preparation of Glasswares and plasticwares

The details of Glassware and plastic ware used in this experiment have been described previously in chapter 2 (section 3.1 & 3.2).

3.1 Chemicals

PCR primers were synthesized from MWG, Bangalore. PCR buffer and Taq-DNA polymerase was from Bangalore Genei, India. dNTPs were from Sigma. AccuTaq-LA DNA polymerase (Invitrogen). Agarose was from Bioworld (USA), Gel elution kit was from Axygen, USA. Restriction enzymes, T4 DNA ligase and RNase were from Promega (USA). Chemicals were purchased from Sigma-Aldrich (USA), Bioworld (USA), Merck, SD fine chemicals, Hi-Media. All the solvents of analytical and HPLC grade were purchased from HiMedia, Qualigens fine chemicals and E-Merck Laboratories, India. The Sucrose and glucose were obtained from Hi- Media. Bacto-Agar for microbial work was obtained from DIFCO laboratories, USA. UDP- glucose, UDP- galactose, UDP-glucuronic acid, naringenin, diadzein, genistein, apigenin, catechin, myricetin, isorhamnetin, luteolin, hesperetin, kaempferol and all the respective glycosides were also obtained from Sigma-Aldrich (USA) and Chromadex (USA).

3.3 Vector and bacterial strains

Table 3A.0

Vector	**Source**	**Features**
pGEM-T Easy	Promega	T/A cloning vector
pET-30 (b+)	Novagen	Expression vector
Bacterial Strains	**Source**	**Genotype**
E. coli XL-10 gold	Stratagene	*Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr]*

XL1-Blue MRF′ strain	Stratagene	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]*
E. coli. BL21 (DE3)	Stratagene	*E. coli B F-ompT hsdS(rB–mB–) dcm+Tetr gal λ(DE3) endA Hte*

3.4 Equipments used during study

Table 3A.1

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	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	pH-Meter	Microset
13	Water purification system	Millipore Unit (Milli RO/Milli Q)
14	Microwave oven	Electrolux
15	Fridge/ Deep freezer	Vestfrost/Leonard/Godrej
16	Magnetic rotator	REMI
17	Laminar Air Flow	Microfilt India
18	HPLC	Perkin Elmer
19	LC-MS	Waters
20	Sonicator	Misonix-XL-2000
21	NanoVue	GE life science

3.5 Buffers and solutions

3.5.1 Buffers and solutions for agarose gel electrophoresis and plasmid isolation

Buffer solution and media used for agarose gel electrophoresis, bacterial studies and plasmid isolation respectively were previously described in chapter 2 (section 3.7.1, 3.7.2 & 3.7.4).

3.5.2 Stock solutions for bacterial transformation and selection

Table 3A.2

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

3.5.3 Buffers and solutions for protein extraction and purification

3.5.3.1 Buffers used for purification of protein from inclusion body

Table 3A.3

Name	Components	Preparation and storage
Lysis buffer	50 mM Tris-HCl (pH 8.0) 1 mM EDTA 150 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)	Stored at 4 ºC
Sonication buffer	100 mM Tris HCl (pH 8.0) 50 mM Glycine	Stored at 4 ºC
Dispersion buffer	100 mM Tris-HCl (pH 8.0) 50 mM Glycine 8 M Urea	Stored at 4 ºC
Binding buffer	50 mM Tris (pH 8.0) 300 mM NaCl 20 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC
Wash buffer	50 mM Tris (pH 8.0) 300 mM NaCl 50 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC
Elution buffer	50 mM Tris (pH 8.0) 300 mM NaCl 300 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC

3.5.3.2 Buffers used for purification of protein from lysate

Table 3A.4

Name	Components	Preparation and storage
Lysis buffer	50 mM Tris-HCl (pH 8.0) 1 mM EDTA 150 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)	Stored at 4 ºC
Binding buffer	50 mM Tris (pH 8.0) 300 mM NaCl 20 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC
Wash buffer	50 mM Tris (pH 8.0) 300 mM NaCl 50 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC
Elution buffer	50 mM Tris (pH 8.0) 300 mM NaCl 300 mM imidazole	pH adjusted by adding concentrated HCl and stored at 4 ºC

3.5.4 Buffers and solutions for protein gel electrophoresis (SDS-PAGE) and staining

Table 3A.5

Name	Component	Preparation and storage
Monomer solution	29.2% acrylamide 0.8% bis-acrylamide in water	Stored at 4 ºC (in darkness)
Stacking gel	Distilled water 2.76 mL 1 M Tris-HCl (pH 6.8) 0.50 mL Acrylamide/bis 30% 0.65 mL 10% SDS 0.04 mL 10% (w/v) APS 0.04 mL TEMED 4 .0 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 (SDS-PAGE) 0.08 mL 10% (w/v) APS 0.08 mL TEMED 8.0 µL	Freshly prepared
2x Protein loading buffer	Distilled water 2.7 mL 0.5 M Tris-HCl (pH 6.8) 1.0 mL Glycerol 2.0 mL 10% SDS(SDS-PAGE) 3.3 mL β-Mercaptoethanol 0.5 mL 0.5% Bromophenol blue 0.5 mL	Stored at 4 ºC
10x SDS-	Tris base 15.1 g	Stored at 4 ºC

electrode buffer	Glycine 94.0 g SDS 0.5 g Adjust pH-8.3 *And make-up the volume up to 500 mL.*	Dilute 1:10 before use
Staining solution (Coomassie)	Coomassie-blue R 250, 0.25 g in Methanol 40 mL Acetic acid, 10 mL *Make-up volume up to 100 mL*	Stored at RT (in brown bottle)
Destaining solution	Methanol, 40 ml Acetic acid, 10 mL *Make- up volume up to 100 mL*	Freshly prepared
Silver staining		
Fixer solution I	40% Methanol (40 mL) 10% acetic acid (10 mL) *Make-up volume up to 100 mL*	Store at RT
Fixer solution II	50% Ethanol	Freshly prepared
Sensitizing solution	0.2% $Na_2S_2O_3$	Store at RT
Silver solution	0.2% silver nitrate (0.6 g) 0.01 % formaldehyde (225 µL) *Make-up volume up to 300 mL*	Prepared freshly in darkness
Developing solution	6% Na_2CO_3 (18 g) 0.02% formaldehyde (450 µL) *Make-up volume up to 300 mL*	Prepared freshly
Stop solution	1.5% Na_2EDTA (4.5 g) *Make-up volume up to 300 mL*	Store at RT

3.6 Substrates and standards used for enzyme assay

The activity of the UGT73A16 was measured using various acceptor and donor substrates and their comparison with specific standards. The list of acceptors, donors and standards are given in **Table 3A.6** and **3A.7**.

Table 3A.6 List of acceptor substrates

Acceptor	Manufacturer	Acceptor	Manufacturer
Apigenin	Sigma	Naringenin	Sigma
Daidzein	Sigma	Catechin	Sigma
Genistein	Sigma	Hesperetin	Sigma
BiochaninA	Sigma	Myricetin	Sigma
Kaempferol	Sigma	Quercetin	Sigma
Isorhamnetin	Sigma	Luteolin	Sigma
3-hydroxy flavone	Sigma	2' hydroxy chalcone (NEIST-28)	Synthesized from north-east university

Table 3A.7 List of donor substrates and standard used in HPLC

Donor substrate	Manufacturer	Standards	Manufacturer
UDP-Glucose	Sigma	Genistein 7-*O*-glucoside	Chromadex (USA)
UDP-glucuronic acid	Sigma	Naringenin 7-*O*-glucoside	Chromadex (USA)
UDP-Galactose	Sigma	Naringenin 4`-*O*-glucoside	Chromadex (USA)
		Kaempferol 3-*O*-glucoside	Chromadex (USA)
		Apigenin 7-*O*-glucoside	Chromadex (USA)
		BiochaninA 7-*O*-glucoside	Chromadex (USA)
		Diadzein 7-*O*-glucoside	Chromadex (USA)
		Isorhamnetin 3-*O*-glucoside	Chromadex (USA)

4. 0 Methods

4.1 Expression and purification of Recombinant UGT73A16

4.1.1 Directional cloning of UGT73A16 in expression vector [pET 30 b (+)]

Appropriate restriction sites (*Nde*I and *Xho*I) were introduced by PCR in UGT73A16 ORF and cloned in pGEM-T Easy vector. Inserts were excised from pGEM-T Easy (**Fig. 3A.1 A**) clone plasmids using restriction enzymes *Nde*I and *Xho*I and ligated into pET 30b (+) vector (**Fig. 3A.1 B**). Ligation mixtures were used to transform in *E. coli* XL-10 and plasmids were isolated and confirmed by digestion. Plasmids were mobilized into the expression cells *E. coli* BL21 (DE3) strain.

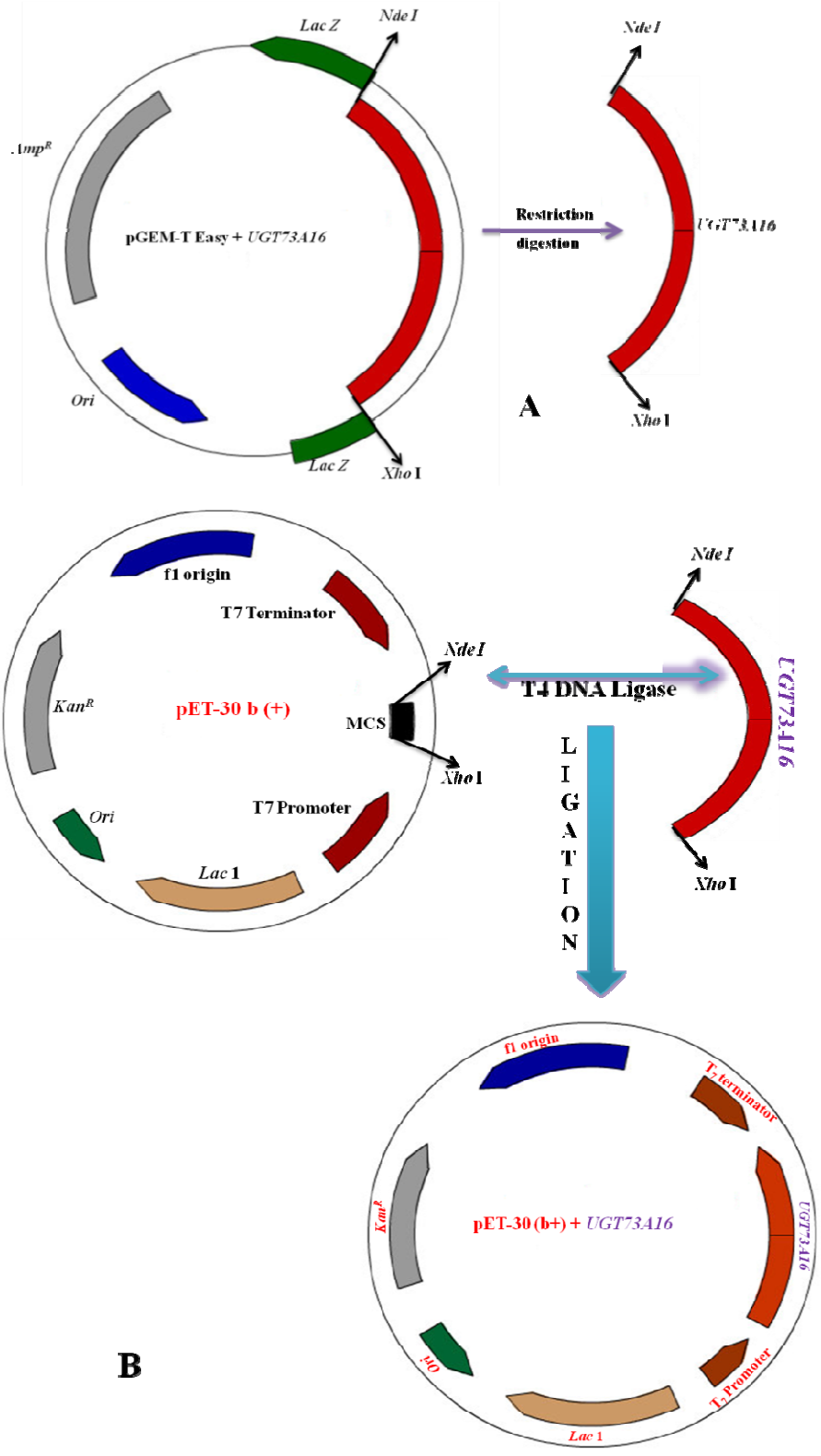


Fig. 3A.1 Strategy used for directional cloning of UGT73A16 in pET-30b (+) vector **(A)** Target gene in pGEM-T Easy vector and double digested fragment with sticky ends **(B)** Ligation in pET 30 b (+) flanked with same restriction sites and expression construct for over-expression of recombinant UGT73A16

4.1.2 Heterologous expression in *E. coli*

Heterologous expression of UGT73A16 gene was attempted in *E. coli* BL21 (DE3) host cells. 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 the culture was used to inoculate 100 mL liquid LB containing 50 µg/mL kanamycin. Once the cultures reached A_{600} 0.4 - 0.6, recombinant protein expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG), and the culture was grown for 4 to 6 h at 37 °C with shaking at 200 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 and cells disruption was done by MISONIX at 70% amplitude for 5 min. MgSO4 (10 mM) and lysozyme (100 µg/mL) was added to the disrupted cells and kept at 37 ^{o}C for 20 min. It was centrifuged at 10,000 rpm for 10 min at 4 ^{o}C 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 was checked on SDS PAGE to check heterologous expression of the proteins.

4.1.3 Affinity purification of recombinant UGT73A16 from inclusion bodies

Polyhistidine-tags are often used for affinity purification of polyhistidine-tagged recombinant proteins that are expressed in *E. coli* or other prokaryotic expression systems. This affinity matrix contains bound metal ion nickel, to which the poly histidine-tag binds with micro molar affinity. The column(s) and buffers were equilibrated to room temperature. The Ni^{+} beads column was equilibrated with binding buffer for 30 min, followed by addition of recombinant protein dissolved in dispersion buffer. The column was kept at 4 ºC for 1 h for binding of recombinant protein to Ni^{+} beads on dancing shaker. After 1 h, beads were allowed to settle down and flow through was collected. Column was then washed with 6-8 bed volume of washing buffer, until OD_{280} reaches to baseline. The washing stringency may be increase by increasing imidazole concentration and 6×His-tagged proteins are usually eluted with 4 aliquots (1 mL each) of 300 mM imidazole. Protein elution was monitored by measuring the absorbance at 280 nm of collected fractions. The eluted protein was separated and analyzed by SDS-PAGE. The gel was stained with coomassie blue and protein concentration was determined by Bradford (Bio-Rad) method using BSA as standard.

4.1.4 Polyacrylamide gel electrophoresis (PAGE)

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

4.1.4.1 Preparation of the separating gel

A vertical slab gel (GE Healthcare, Sweden) was assembled using 1.5 mm spacers. In a side armed vacuum flask, 10% separating gel solution was made as mentioned in section 3.5.4, 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 solution.

4.1.4.2 Preparation of the stacking gel

Stacking gel solution was prepared as given in section 3.5.4, 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.

4.1.4.3 Preparation of the sample

Equal parts of the protein sample and the loading buffer were mixed in a micro-centrifuge tube and kept in a boiling water bath for 3-5 min for SDS-PAGE. Then the samples were centrifuged at 10,000 rpm for 2 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.

4.1.4.4 Loading of samples and running the polyacrylamide gel

After polymerization of polyacrylamide gel, wells were flushed 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.

4.1.4.5 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 gel staining box (Tarsons, India) containing Coomassie blue staining solution and was kept for 2-3 h at room temperature on dancing shaker. Staining solution was poured off and de-staining solution was poured in. De-staining step was repeated two-three times till clear bands appeared.

4.1.4.6 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 (20 s each) and transferred to the developer till the bands developed. The gel was washed with deionized H_2O and stored in fixer.

4.1.4 Optimization of UGT73A16 expression in *E. coli*

Expression of soluble protein was optimized for different IPTG concentration (for induction), temperature (to grow the culture after induction) and duration of induction. Both IPTG concentration and temperature were reduced to increase expression of soluble protein.

4.1.5 Purification of protein from cell lysate

Cell lysate obtained as given above in section 4.1.2 was utilized to purify protein in native form. Resins and buffers were used in purification of protein from cell lysate are given above in section 3.5.3.2. Purification was performed as described in above section but all the protocols were performed at 4 °C to rescue the purified protein in active form. The eluted protein was separated and analyzed by SDS-PAGE. The gel was stained with coomassie blue stain and silver stain. The protein concentration was determined by Bradford (Bio-Rad) method using BSA as the standard.

4.2 UGT73A16 enzyme assay and product identification

The reaction mixture (200 µL) consisted of 400 µM glucose donor (UDP-glucose, UDP-galactose and UDP- glucuronic acid), 100 µM substrate, 0.1 M Tris-HCl buffer (pH 8.0) and 40 µg of purified UGT73A16. The reaction was incubated at 37 ^{o}C for 3 hours. The reaction was terminated and extracted twice by the addition of equal volume of ethyl acetate and evaporated to dryness. The dried reaction product was dissolved in methanol and analyzed by HPLC, hypsochromic shift and LC-MS.

4.2.1 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.

4.2.2 Hypsochromic shift

Hypsochromic shift is a change of spectral band position in the absorption, reflectance, transmittance, or emission spectrum of a molecule to a shorter wavelength (higher frequency). Hypsochromic shift analysis was performed on λ-650 nm scanner spectrophotometer Perkin Elmer. Usually flavonols have two absorption maxima: Band I (350-380 nm) and Band II (240-280 nm) corresponding to the B- and A- ring (**Fig. 3A.2**), respectively. Conjugation of 3-, 5-, or 4' – hydroxyl groups causes a Band I hypsochromic shift, which is larger for a 3-substitution (11-19 nm) than a 4'-conjugation (3-5 nm).

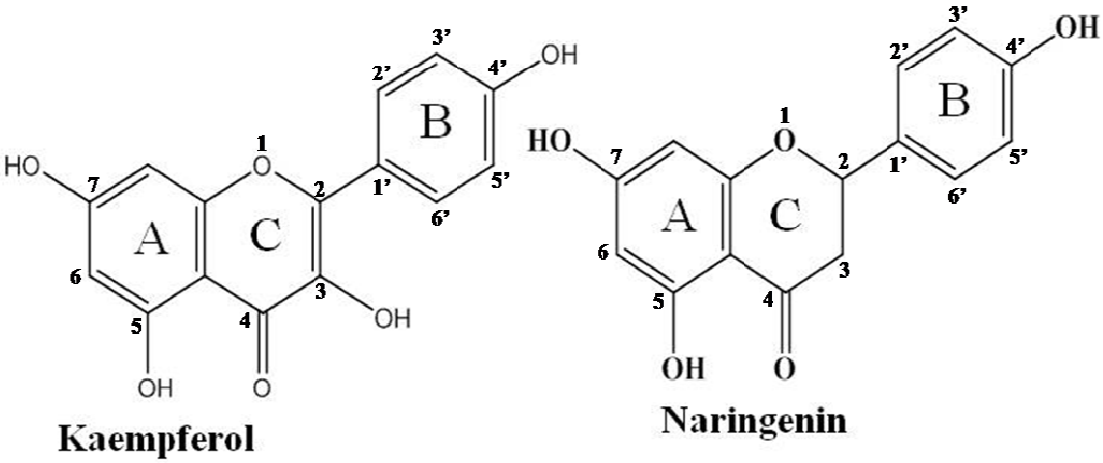


Fig. 3A.2 Structure of flavonoids and position of their OH groups

4.2.3 HPLC

The dried reaction products dissolved in methanol were analyzed by high performance liquid chromatography (HPLC, Perkin Elmer, USA) equipped with a diode array detector (DAD) and a Waters symmetry C_{18} column (5 µm particle size, 4.6 mm × 25 cm, supelco analytical) except for kaempferol and baicalein (3 µm particle size, 4.6 mm × 15 cm, supelco analytical). For generation of an analytical scale, the mobile phase consisted of sterile milliQ water (SMQ) with; either 0.1% TFA or 0.1% phosphoric acid and was programmed as follows (**Table 3A.8**). The flow rate was 1 mL/ min and UV detection was performed at 250 – 360 nm.

Table 3A.8 Mobile phase system for each substrate

Compound/UV Absorbance	Step	Time (Min)	Acetonotrile (%)	SMQ (0.1% TFA)
Apigenin (320 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	40	60
	3	5.0	70	30
	4	5.0	90	10
Daidzein (295 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	30	70
	3	5.0	60	40
	4	5.0	90	10
Genistein (262 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	30	70
	3	5.0	60	40
	4	5.0	90	10
Kaempferol (320 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	30	70
	3	5.0	75	25
	4	5.0	95	05
Isorhamnetin (320 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	35	65
	3	5.0	65	35
	4	5.0	90	10
Naringenin (280 nm)	**1**	**5.0**	**20**	**80**
	2	5.0	40	60
	3	5.0	70	30
	4	5.0	90	10
BiochaninA (260 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	30	70
	3	5.0	70	30
	4	5.0	90	10
Baicalein (260 nm)	**1**	**5.0**	**10**	**90**
	2	5.0	35	65
	3	5.0	70	30
	4	5.0	90	10
3-hydroxy flavone (320 nm)	**1**	**5.0**	**40**	**60**
	2	5.0	60	40
	3	5.0	90	10
	4	5.0	95	05
2'-hydroxy chalcone (355 nm)	**1**	**5.0**	**30**	**70**
	2	5.0	50	50
	3	5.0	70	30
	4	10.0	90	10

4.3 Kinetic studies of UGT73A16

The K_m and V_{max} values of UGT73A16 for naringenin, genistein and baicalein were determined at pH 8.0 and 37 °C by fitting the initial velocity data of Michelis-Menten equation using nonlinear regression analysis. Substrates were used at a saturating concentration with constant enzyme concentration. The reaction for UDP-glycosyltransferase (naringenin and genistein) contained 12 µg of purified UGT73A16, 100 mM Tris-Cl (pH 8.0), 100 µM of UDP- glucose and 2 µM -18 µM of substrates. In case of baicalein, reaction contained 40 µg of purified UGT73A16, 100 mM Tris-Cl (pH 8.0), 400 µM of UDP- glucuronic acid and 10 µM -100 µM of substrates. To determine the standard graph Michelis-Menten plot, peak area of HPLC chromatograms were calculated. Concentration of substrates and standards were considered on X-axis and peak area on Y-axis and graphs were plotted using Origin 6.1 software.

5.0 Results and discussion

5.1 Cloning of UGT73A16 gene in pET 30b (+) vector

5.1.1 Incorporation of restriction sites and directional cloning of UGT73A16 gene in pET-30b (+)

The cloned sequence of UGT73A16 in pGEM-T Easy was used as template using gene specific primers WSfullF and WSfullR to incorporate the *Nde*I site at the 5' end and *Xho*I site at 3' end of the gene. Approximately, 1.4 kb (1413 bp *UGT73A16* ORF + *Nde*I and *Xho*I restriction sites) amplicon was cloned in pGEM-T Easy vector. The above clone of UGT73A16 gene in pGEM-T Easy vector was restriction digested with *Nde*I and *Xho*I restriction enzymes and ~1.4 kb fragment was ligated into pET-30b (+) vector (**Fig. 3A.1 B**) flanked with same restriction sites. Ligation mixture was transformed into *E. coli* XL10 competent cells and plated on LB-agar plate (kanamycin 50 µg/mL). *E. coli* colonies having recombinant plasmids were screened by colony PCR with T7 promoter and T7 terminator primers (**Fig. 3A.4 A**). Plasmids were isolated from colony PCR positive clones and digested with *Nde*I and *Xho*I to confirm the integration of UGT73A16 gene fragment in pET-30b (+) vector (**Fig. 3A.4 B**). Further sequencing was done for confirmation of integration of the UGT73A16 gene along with restriction sites. The sequence was translated using proteomic tools for checking in frame translation up to 6x HIS tag.

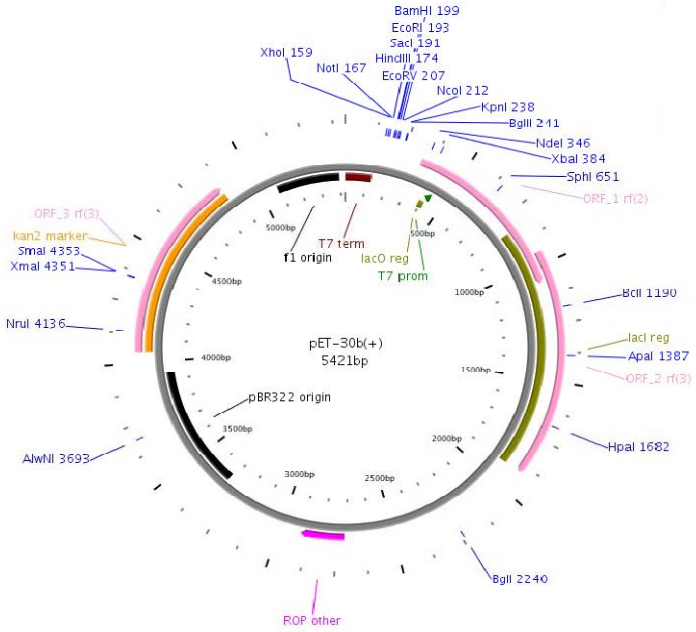


Fig. 3A.3 Vector map of pET-30b (+)

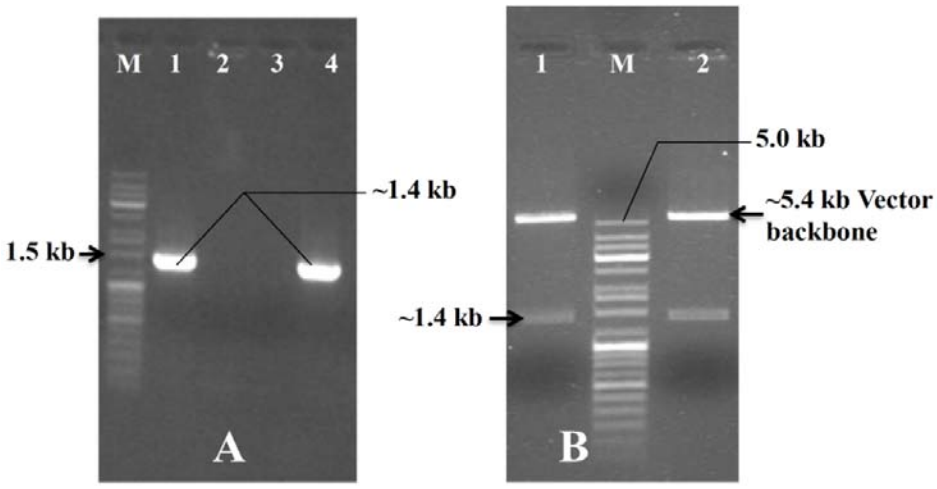


Fig. 3A.4 (A) Colony PCR showing ~ 1.4 kb UGT73A16 gene (additional amplification is due to presence of sequences between T7 Prom/Term) in recombinant colonies; Lane M: Low range DNA ruler, lane 1 & 4 are positive clones **(B)** Restriction analysis of recombinant pET 30b (+) with *Nde*I and *Xho*I: lane 1 & 2 recombinant vector releasing ~1.4 kb UGT73A16 gene and ~5.4 kb vector backbone, lane M: Medium range DNA ruler

5.2 Recombinant UGT73A16 protein expression and purification

E. coli BL 21 (DE3) cells transformed with recombinant pET 30b (+) plasmid was screened for over-expression of UGT73A16 protein. The *E. coli* BL21 (DE3) cells harboring recombinant plasmid were grown O/N in 5 mL LB (Kanamycin 30 µg/mL) tubes. The above primary culture was re-innoculated in fresh 50 mL LB (Kanamycin 50

μg/mL). As the A_{600} reached 0.5, it was induced with 1 mM IPTG solution and kept at 37 °C for 6 h. An uninduced control was taken out before inducing with IPTG. After 6 h of induction, samples were centrifuged at 10000 g for 2 min and re-suspended in 200 μL 2X SDS-Gel loading buffer. The solution was boiled for 5 min, centrifuged at 10000 g for 2 min and loaded onto a 10% SDS- PAGE (as described earlier). The gel was run at 100 V for 2 h and stained with Coomassie blue (R 250) staining solution. A ~52 kDa protein was found to be over-expressed. This over-expressed UGT73A16 protein was further purified by Ni^{+}-NTA affinity purification and analyzed on 10% SDS-PAGE (**Fig. 3A.5 A**).

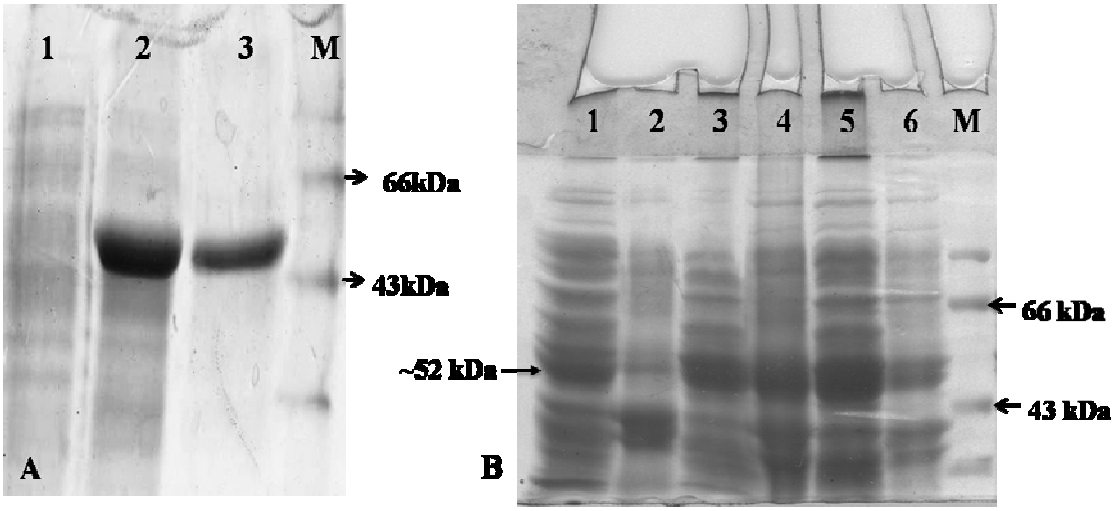


Fig. 3A.5 (A) Lane 1: uninduced recombinant UGT73A16, Lane 2**:** induced recombinant UGT73A16, Lane 3: Purified UGT73A16 ~ 52 kDa protein, Lane M: Protein Molecular Weight Marker **(B)** Over-expression of UGT73A16 protein in soluble form. The protein was extracted at 20 h after 0.08 mM IPTG induction at different temperature: Lane 1 & 2- lysate & IB at 17 °C**,** Lane 3 & 4 -lysate & IB at 20 °C**,** Lane 5 & 6- lysate & IB at 22 °C**,** Lane M- Protein Molecular Weight Marker

5.3 Standardization of IPTG concentration, time duration and temperature for protein expression in soluble form and their purification

The information obtained from above experiment was utilized for this experiment. Initially 50 mL of LB broth (Kanamycin 50 μg/mL) was inoculated with O/N grown culture of *E. coli* BL21 (DE3) harboring UGT73A16 recombinant plasmid. Induction with 0.08 mM IPTG for 20 h was done when A_{600} reached 0.4-0.6 and kept at different temperatures (17 °C, 20 °C and 22 °C). The SDS-PAGE analysis indicated that the soluble fraction of the cell lysate and pellet fraction of recombinant bacteria contained an over expressed protein with an apparent molecular mass of ~52 kDa. SDS-PAGE analysis of protein lysate obtained from harvesting cultures at different temperature

showed that amount of protein in lysate was maximum in 22 °C harvested culture (**Fig. 3A.5 B**). Lysate was used to purify protein by affinity purification using Ni-NTA (**Fig. 3A.6 A & 3A.6 B**) with the buffers given above in **Table 3A.4.**

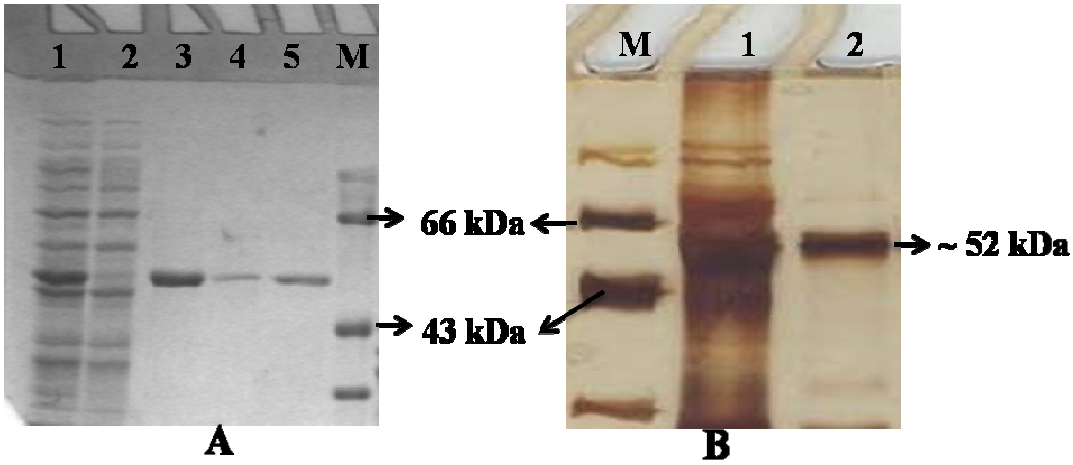


Fig. 3A.6 (A) Commassie stain: Lane 1-Induced *E. coli* UGT73A16 lysate at 22°C, Lane 2- Induced *E. coli* UGT73A16 IB at 22°C, Lane 3, 4 & 5- Purified recombinant UGT73A16, M- Protein molecular weight marker **(B) Silver stain: M-** Protein molecular weight marker, Lane 1- Induced *E. coli* UGT73A16 lysate at 22°C, Lane 2- Purified recombinant UGT73A16

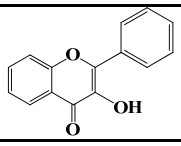
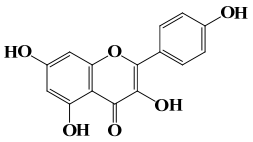
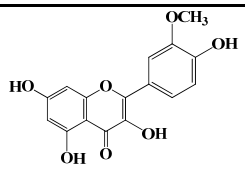
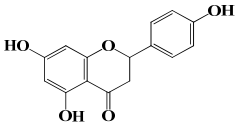
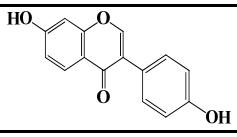
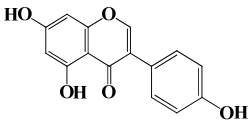
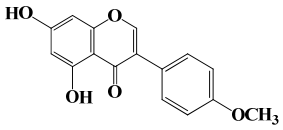
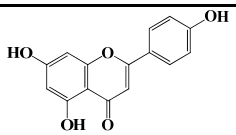
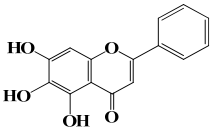
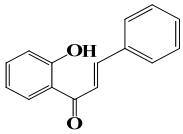
5.4 UGT73A16 enzyme assay and product identification

Affinity purified UGT73A16 protein was used for enzyme assay using UDP- glucose, UDP- galactose and UDP- glucuronic acid as a sugar donor and different flavonoids as an acceptor substrate as given above (**Table 3A.6 & 3A.7**). Activity of the UGT73A16 was checked by glycosylation reaction of various acceptor substrates by binding of sugar molecule. The glycosylation positions of reaction products were determined by three methods: comparison of HPLC retention time and UV spectra with authentic flavonoid glucosides, change in hypsochromic shift in the reaction products and liquid chromatography-mass spectrometry (LC-MS).

Increase in the molecular mass 162 Da relative to original substrate, indicates that one glucose molecule was attached during the reaction. The commercial availability of flavonoid glucoside is limited, and hence the comparison of retention time and UV - spectra was carried out only when authentic compound glucosides were available. Additionally product identification was made by hypsochromic shift data based on the observation that shift occurs only when the glycosylated positions were either at C-3 or C-4' hydroxyl group of flavone or flavonol: in contrast glycosylation at the position 7-hydroxyl group does not result in hypsochromic shift. All the acceptor substrates except

baicalein showed the activity with UDP- glucose but not with the UDP- galactose. In case of baicalein, it showed the activity only with UDP- glucuronic acid.

Table 3A.9 Substrate specificity of UGT73A16 with different substrate

Group of compounds	Donor Substrate	Acceptor substrate	Structure	UGT73A16
Flavonols	UDP-Glucose	3-Hydroxy flavone		**+Ve**
	UDP-Glucose	Kaempferol		**+Ve**
	UDP-Glucose	Isorhamnetin		**+Ve**
	UDP-Glucose	Naringenin		**+Ve**
Iso-Flavones	UDP-Glucose	Daidzein		**+Ve**
	UDP-Glucose	Genistein		**+Ve**
	UDP-Glucose	BiochaninA		**+Ve**
Flavones	UDP-Glucose	Apigenin		**+Ve**
	UDP-Glucuronic acid	Baicalein		**+Ve**
Chalcone	UDP-glucose	2' hydroxy chalcone (NEIST28)		**+Ve**

5.4.1 Identification of naringenin

Naringenin reaction generated two new peaks which has the 9.8 min and 5.5 min retention time (**Fig. 3A.7 A**). The retention time at 9.8 min (First peak) are similar to authentic compound corresponding naringenin 7-*O*-glucoside (**Fig. 3A.7 B**), which does not showed the hypsochromic shift. In the case of second peak which has the retention time 5.5 min showed ~ 6 nm hypsochromic shift, which strongly suggested that glycosylation occurs at C-4' hydroxyl group (**Fig. 3A.7 D**).

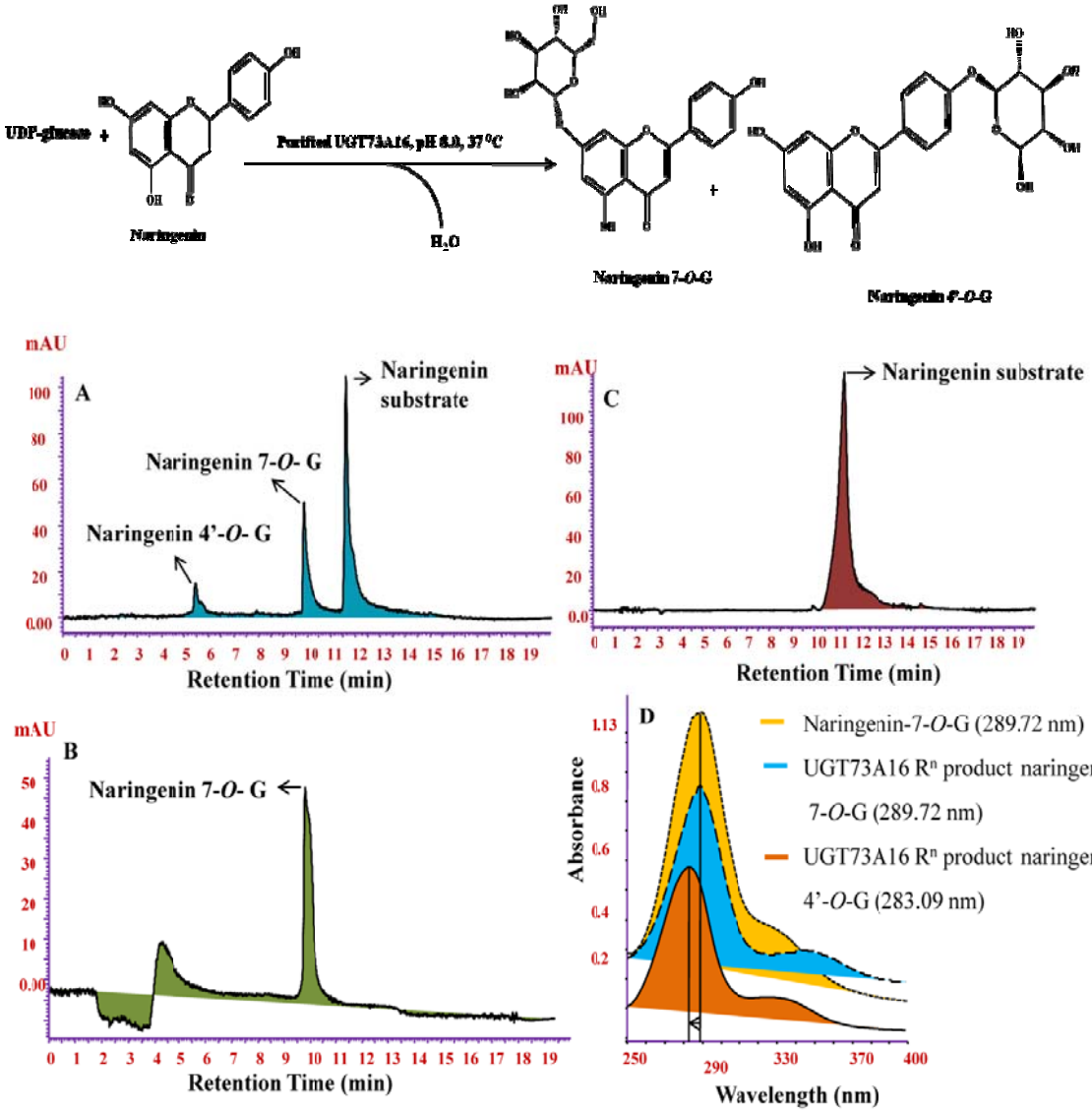


Fig. 3A.7 (A) HPLC chromatogram: UGT73A16 assay mixture of **Naringenin:** major reaction product, naringenin 7-*O*-glucoside; minor reaction product, naringenin 4'-*O*-glucoside; **(B) Standard:** naringenin 7-*O*-glucoside; **(C) Substrate:** naringenin; **(D) Hypsochromic shift**: Naringenin 7-*O*-glucoside standard in yellow, UGT73A16 Reaction product (major and minor) shown in blue and orange color respectively

5.4.2 Identification of kaempferol

The retention time of the reaction product from kaempferol was 9.5 min (**Fig. 3A.8 A**) similar to the authentic compound corresponding to kaempferol 3-*O*-glucoside (**Fig. 3A.8 B**). The reaction product showed ~18 nm hypsochromic shift (**Fig. 3A.8 D**) as compared to original substrate (13.5 min retention time), which suggested that glycosylation occurs at C-3 hydroxyl group of original compound.

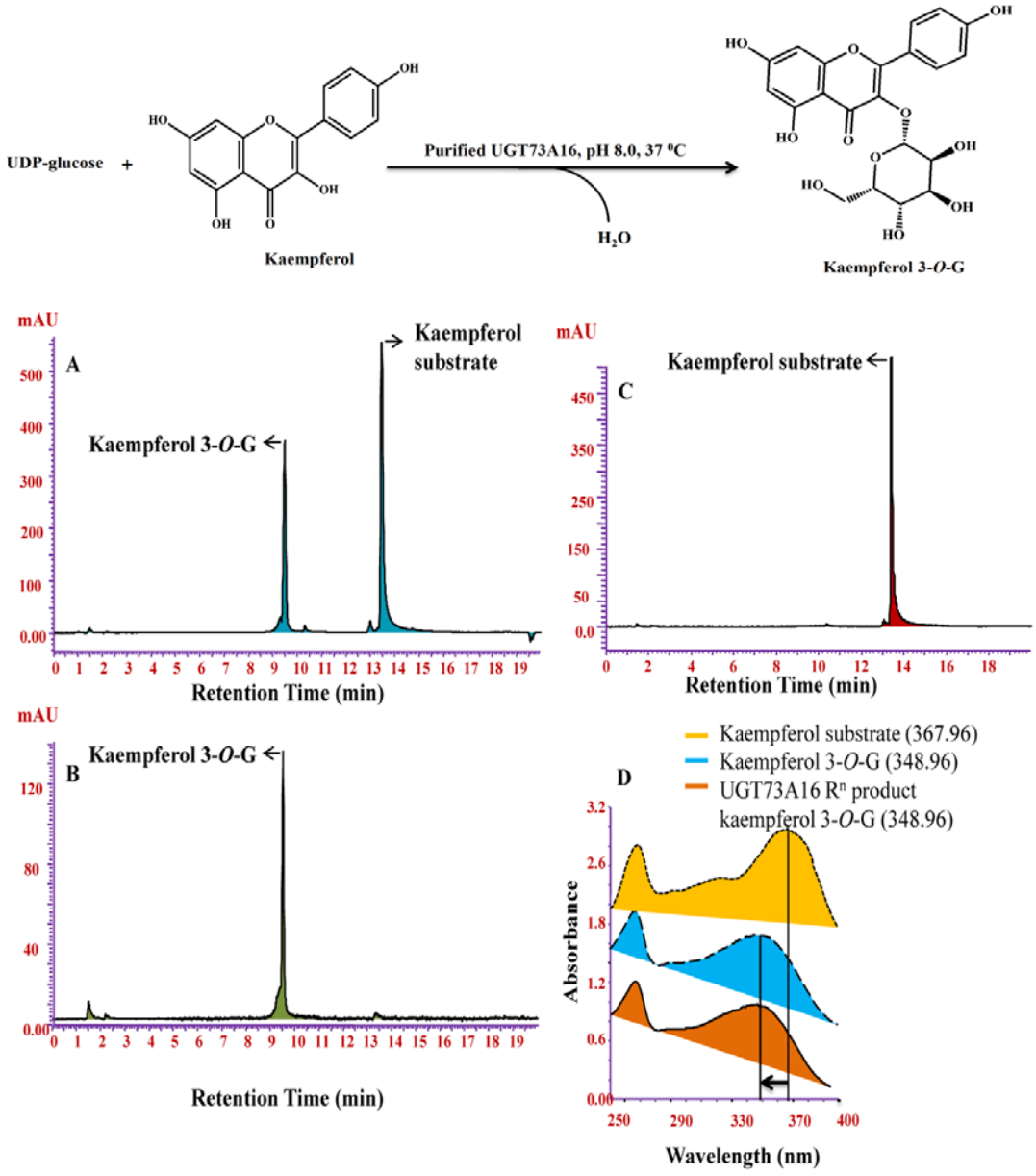


Fig. 3A.8 (A) HPLC chromatogram: UGT73A16 assay of **kaempferol:** Single reaction product, kaempferol 3-*O*-glucoside **(B) Standard:** kaempferol 3-*O*-glucoside; **(C) Substrate:** Kaemferol; **(D) Hypsochromic shift**: kaempferol substrate in yellow, kaempferol 3-*O*-glucoside standard in blue and UGT73A16 reaction product in orange color

5.4.3 Identification of Biochanin A

The retention time of the reaction product from biochaninA is 14.5 min (**Fig. 3A.9 A**) and similar to the authentic compound corresponding to biochaninA 7-*O*-glucoside (**Fig. 3A.9 B**). The lack of hypsochromic shift between substrate (17.0 min retention time) and product suggested that glycosylation occurred at hydroxyl group of C-7 (**Fig. 3A.9 D**).

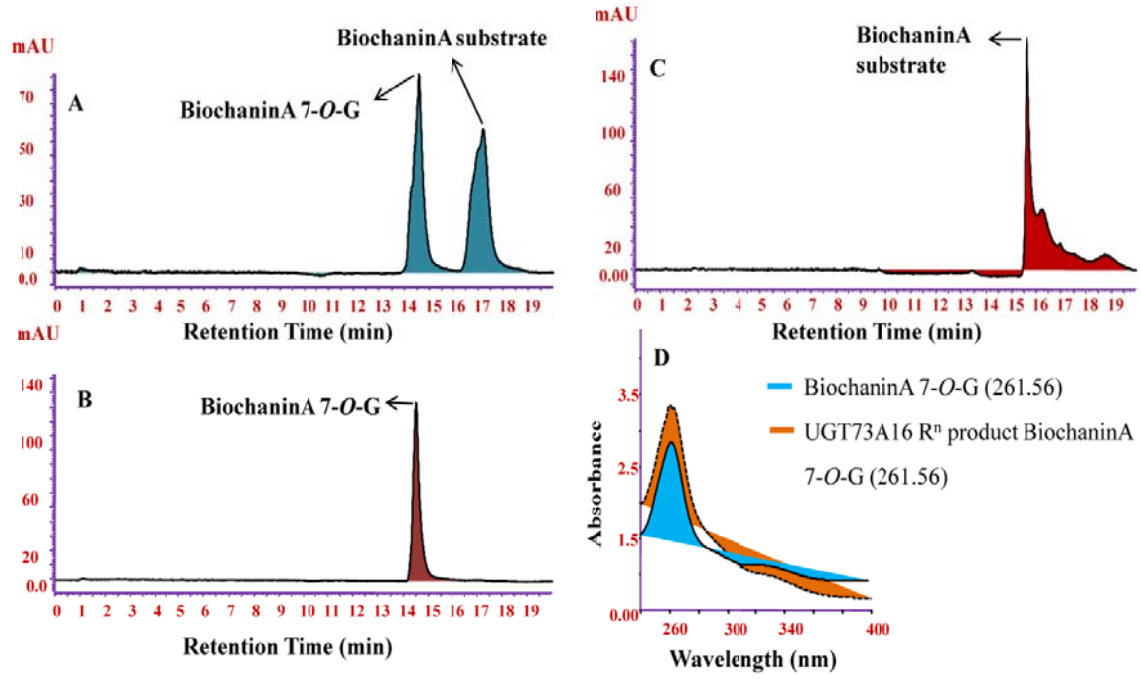
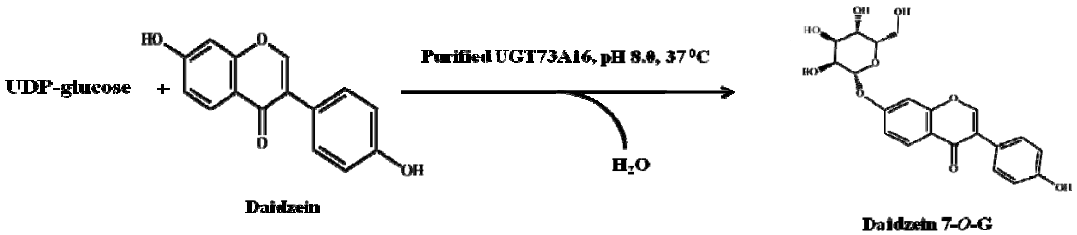


Fig. 3A.9 (A) HPLC chromatogram: UGT73A16 assay mixture of **biochanin A:** single reaction product, biochanin A 7-*O*-glucoside **(B) Standard:** biochanin A 7-*O*-glucoside **(C) Substrate:** biochanin A **(D) Hypsochromic shift**: biochanin A 7-*O*-glucoside standard in orange and UGT73A16 reaction product in orange color

5.4.4 Identification of daidzein

Daidzein reaction generated one new peak which has the 10.4 min retention time (**Fig. 3A.10 A**) and similar to authentic compound corresponding daidzein 7-*O*-glucoside (**Fig. 3A.10 B**). The lack of hypsochromic shift between substrate (14.2 min retention time) and product suggested that glycosylation occurred at hydroxyl group of C-7 (**Fig. 3A.10 D**).



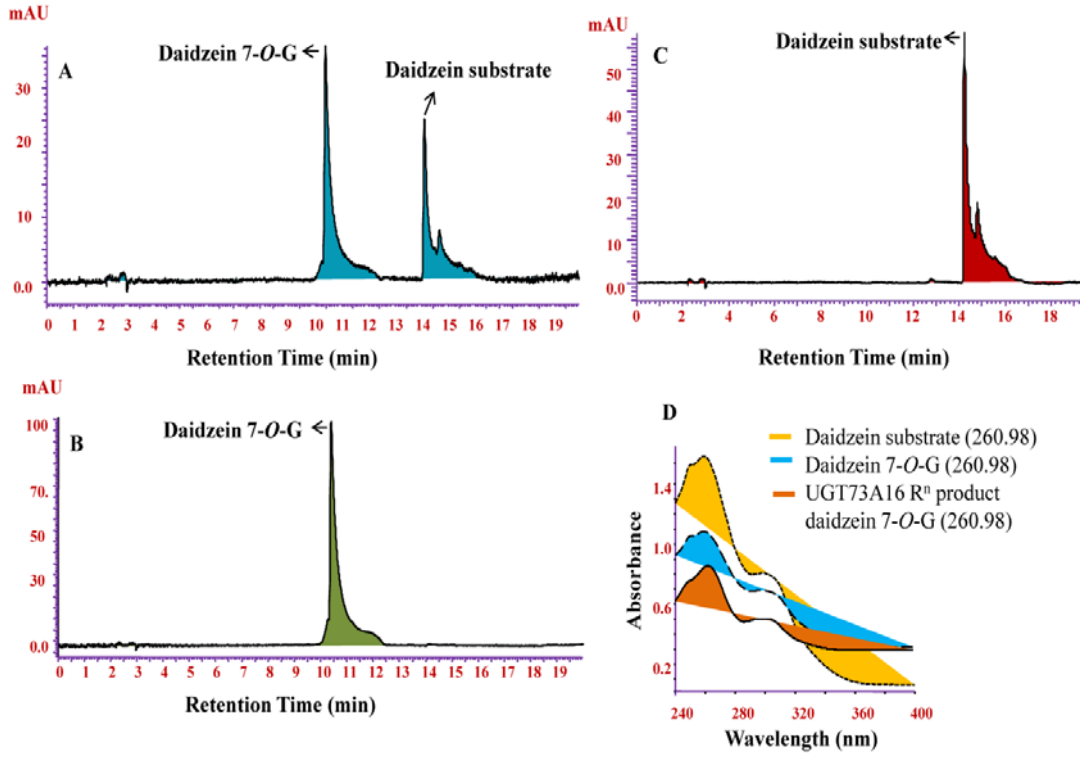
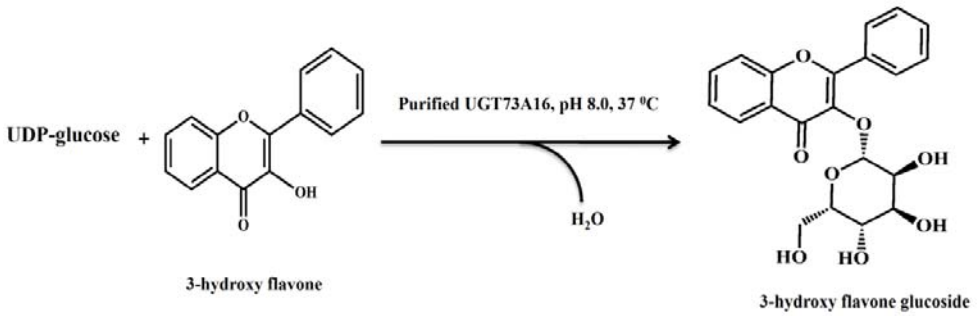


Fig. 3A.10 (A) HPLC chromatogram: UGT73A16 assay mixture of **daidzein:** single reaction product, daidzein 7-*O*-glucoside **(B) Standard:** daidzein 7-*O*-glucoside **(C) Substrate:** daidzein **(D) Hypsochromic shift**: daidzein substrate, daidzein 7-*O*-glucoside standard and UGT73A16 reaction product shown in yellow, blue and orange color respectively

5.4.5 Identification of 3-hydroxy flavone

In case of 3- hydroxy flavone (3-HF), which has the single OH group at the C-3 position of B-ring, a new peak at 6.3 min retention time on HPLC was observed (**Fig. 3A.11 A**) and further analyzed on LC-MS (**Fig. 3A.11 B**). The LC-MS analysis showed the mass of unused substrate with the majority of hydrogen ion (m/z 239.198 $[M+1]^+$) and the reaction product with majority of sodium ion (m/z 423.386 $[M+23]^+$). Increase in the mass of substrate by 162 Da with sodium ion suggested that glycosylation reaction occurred at 3- OH group of 3HF.



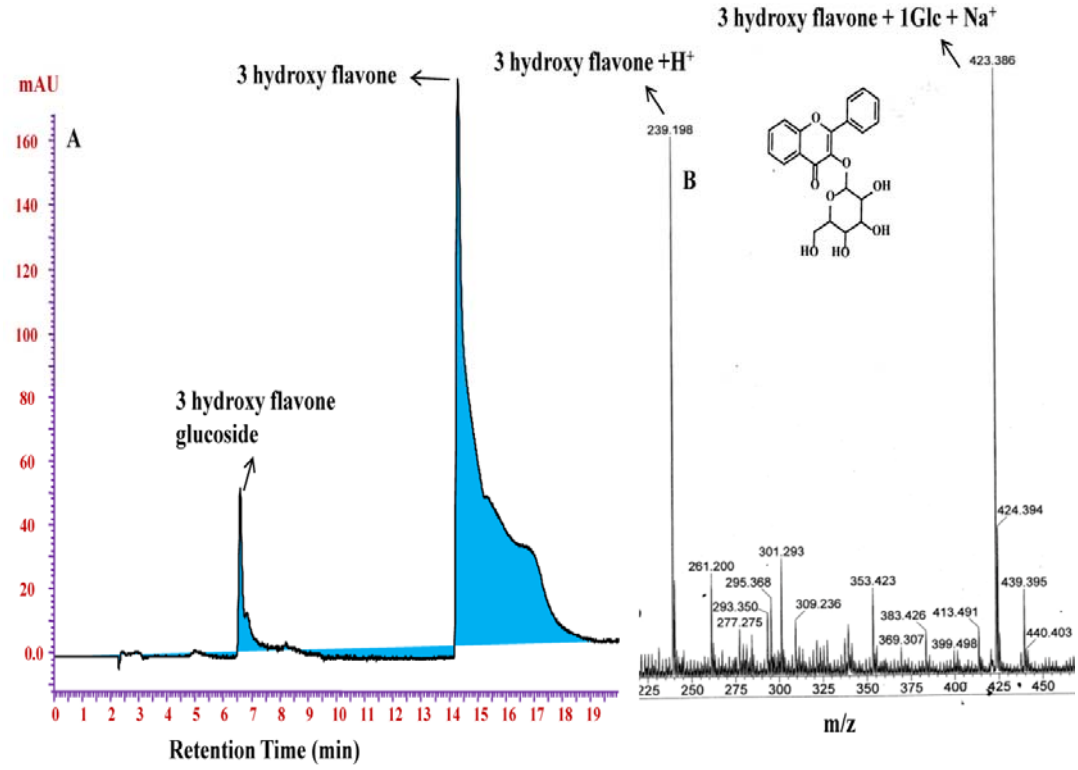
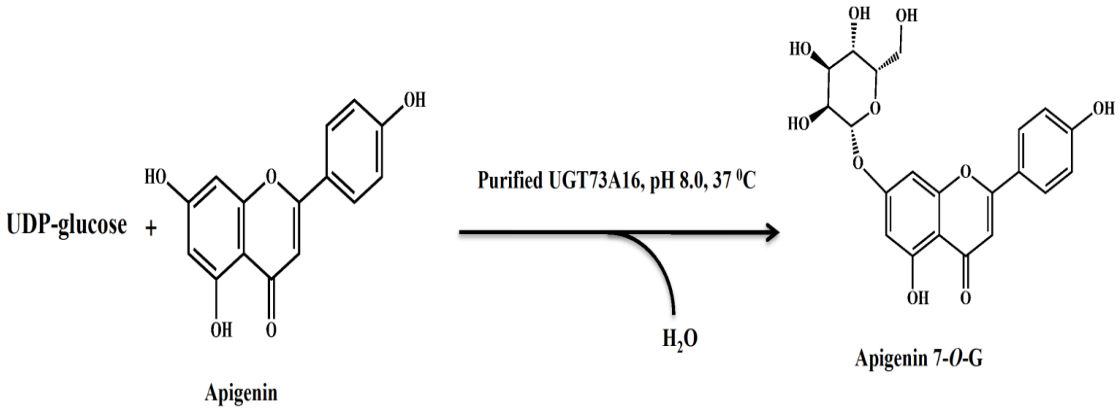


Fig. 3A.11 (A) HPLC chromatogram: UGT73A16 assay mixture of **3HF:** single reaction product, 3-hydroxy flavone glucoside. **(B)** LC-MS data of 3HF substrate and their reaction product

5.4.6 Identification of apigenin

On HPLC based analysis, apigenin substrate had 12.5 min retention time. The assay of apigenin generated a new peak with 10.4 min retention time (**Fig. 3A.12 A**), which was similar to authentic compound, apigenin 7-*O*-glucoside (**Fig. 3A.12 B**). HPLC analysis showed that the reactions of UGT73A16 with apigenin yielded a single mono-glycosylated product.



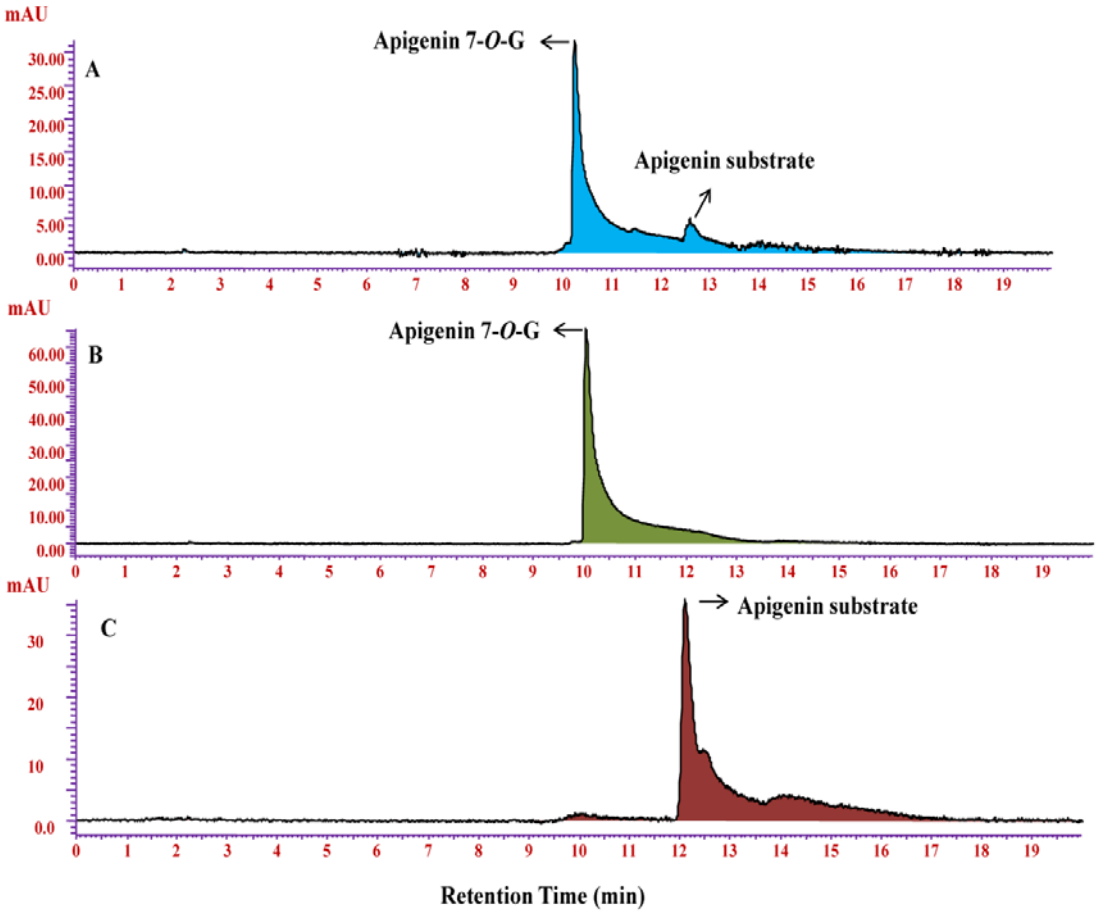
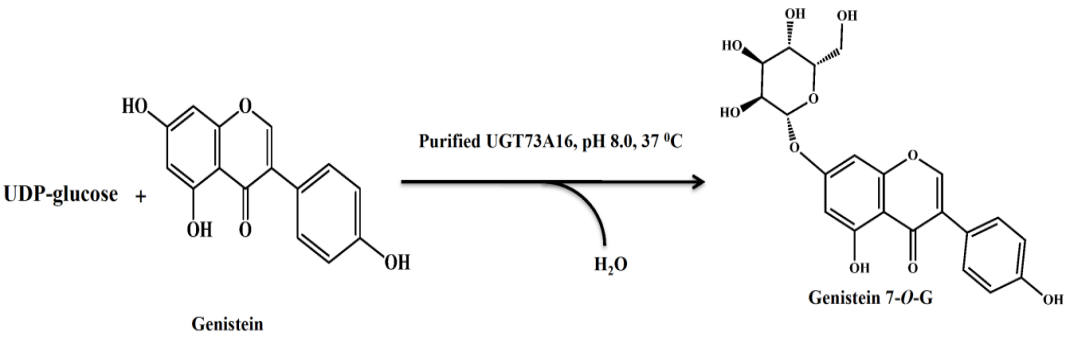


Fig. 3A.12 (A) HPLC chromatogram: UGT73A16 assay mixture of **apigenin:** single reaction product, apigenin 7-*O*-glucoside **(B) Standard:** apigenin 7-*O*-glucoside **(C) Substrate:** apigenin

5.4.7 Identification of genistein

HPLC analysis showed that the reaction of UGT73A16 with genistein yielded a single mono-glycosylated product (**Fig. 3A.13 A**). The retention time of the reaction product from genistein was 11.2 min and it was similar to the authentic compound corresponding to genistein 7-*O*-glucoside (**Fig. 3A.13 B**). The lack of hypsochromic shift between substrate (15.2 min retention time) and product suggested that glycosylation occurred at hydroxyl group of C-7 (**Fig. 3A.13 D**)



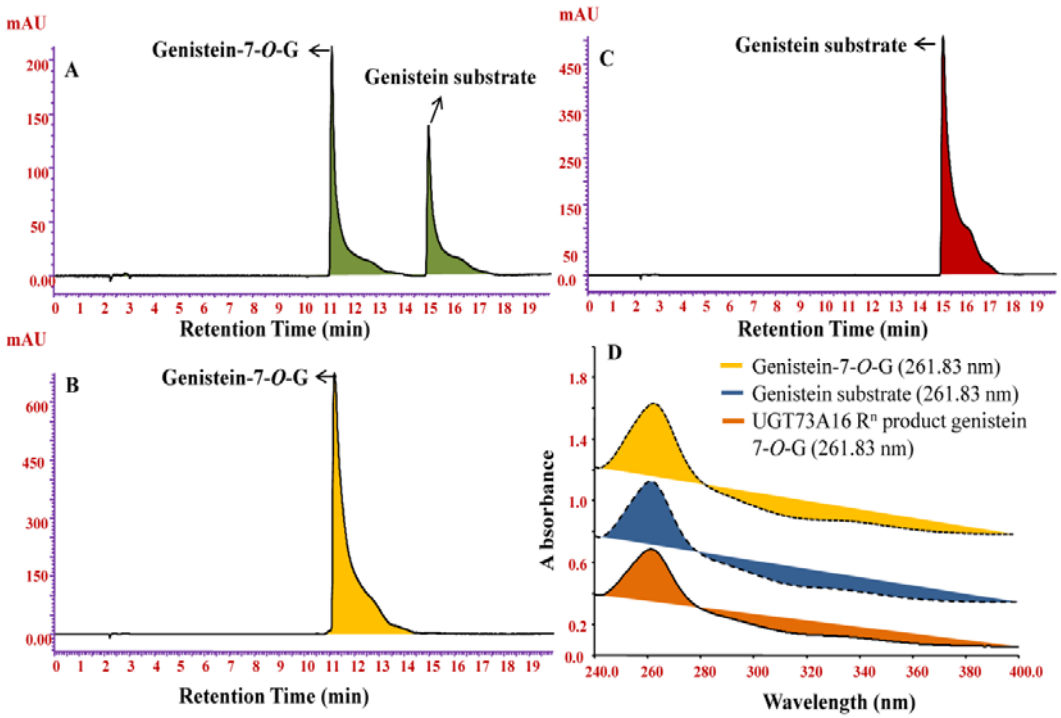
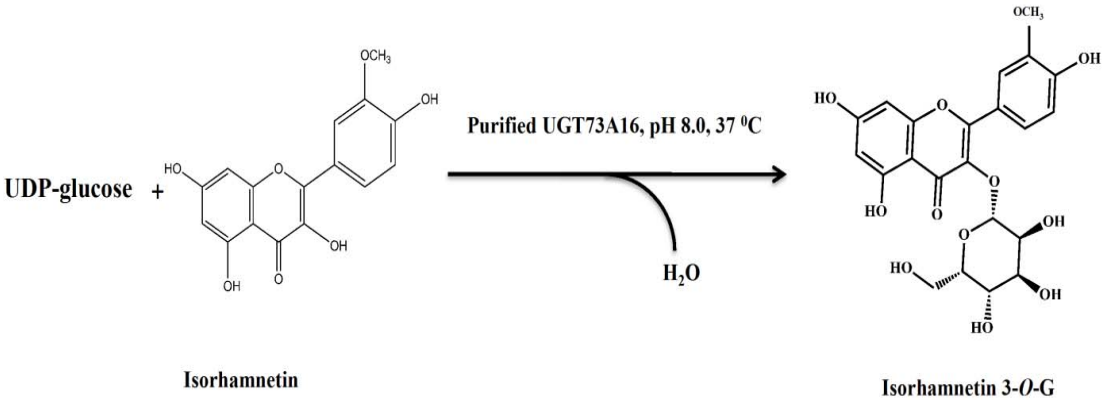


Fig. 3A.13 (A) HPLC chromatogram: UGT73A16 assay mixture of **genistein:** single reaction product, genistein 7-*O*-glucoside **(B) Standard:** genistein 7-*O*-glucoside **(C) Substrate:** genistein **(D) Hypsochromic shift**: genistein 7-*O*-glucoside standard, genistein substrate and UGT73A16 reaction product shown in yellow, blue and orange color respectively

5.4.8 Identification of isorhamnetin

The retention time of the reaction product from isorhamnetin was 9.5 min (**Fig. 3A.14 A**) and it was similar to authentic compound corresponding to isorhamnetin 3-*O*-glucoside (**Fig. 3A.14 B**), which suggested that glycosylation occurs at C-3 hydroxyl group of original compound. The reaction product showed the ~17 nm hypsochromic shift (**Fig. 3A.14 D**) as compared to original substrate (13.8 min retention time), which suggested that glycosylation occurs at C-3 hydroxyl group of original compound.



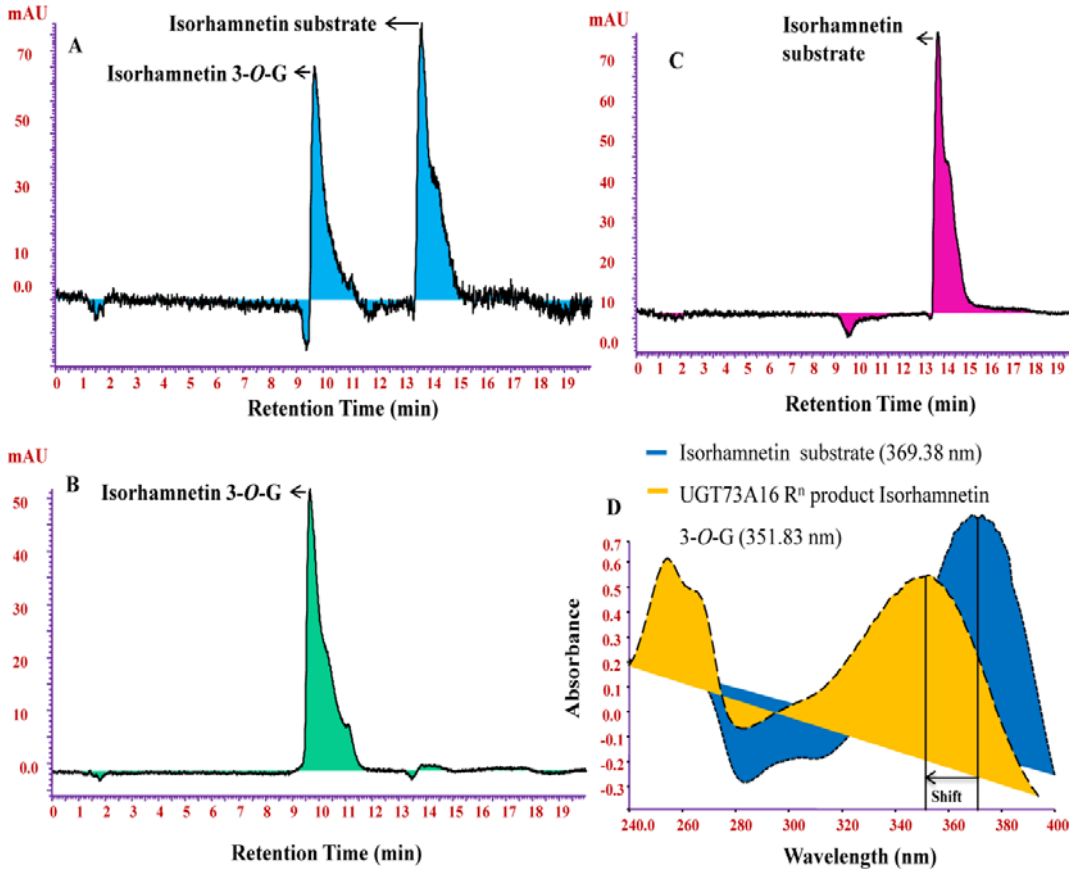
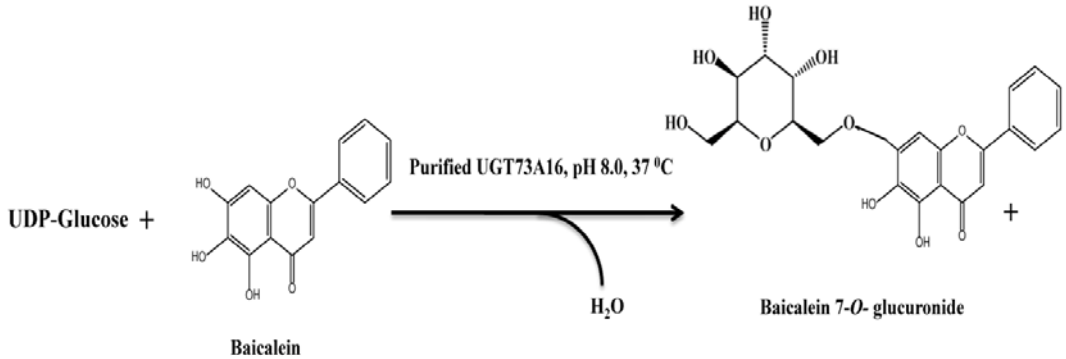


Fig. 3A.14 (A) HPLC chromatogram: UGT73A16 assay mixture of **isorhamnetin:** single reaction product, isorhamnetin 3-*O*-glucoside **(B) Standard:** isorhamnetin 3-*O*-glucoside **(C) Substrate:** isorhamnetin **(D) Hypsochromic shift**: isorhamnetin substrate and UGT73A16 reaction product shown in blue and yellow color respectively.

5.4.9 Identification of Baicalein

Baicalein reaction generated a new peak with 9.9 min (**Fig. 3A.15 A**) retention time and it was similar to authentic compound corresponding baicalein 7-*O*-glucuronide (**Fig. 3A.15 B**), which suggested that glycosylation occurs at C-7 hydroxyl group of substrate.



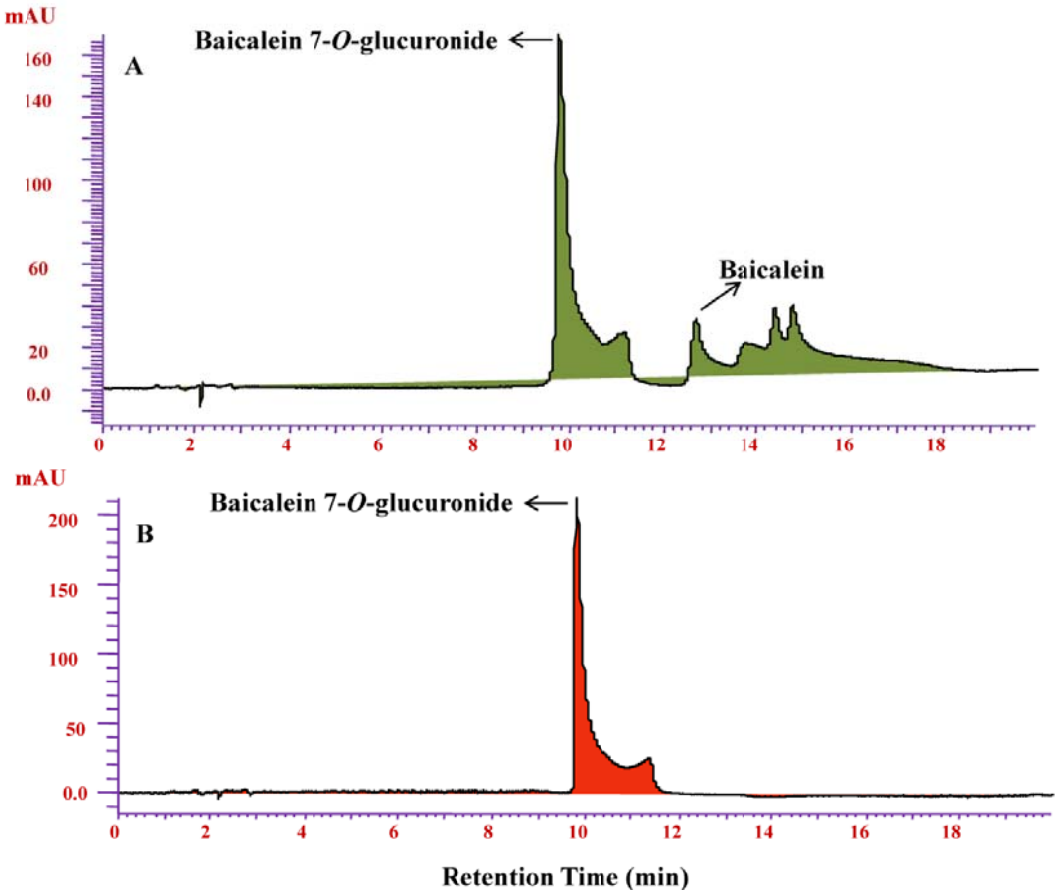
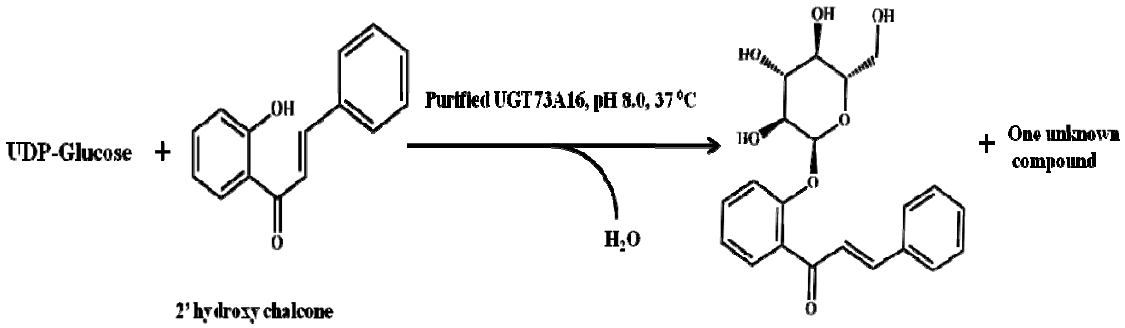


Fig. 3A.15 (A) HPLC chromatogram: UGT73A16 assay mixture of **baicalein:** single reaction product, baicalein 7-*O*-glucuronide (baicalin) **(B**) **Standard:** baicalein 7-*O*-glucuronide

5.4.10 Identification of 2' hydroxy chalcone

The 2’ hydroxyl chalcone had single hydroxyl group and showed 19.8 min retention in HPLC. The assay of this substrate generated a new peak, with 9.2 min retention (**Fig. 3A.16 A**). This peak was collected and further analyzed by LC-MS.



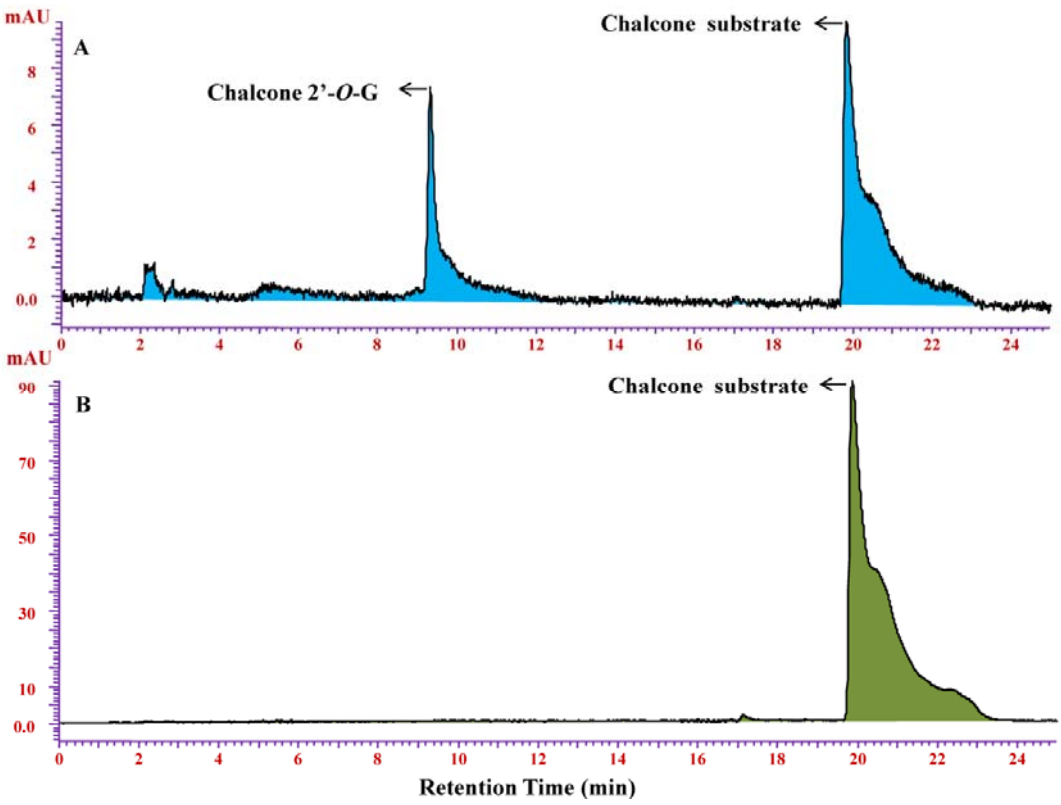


Fig. 3A.16 (A) HPLC chromatogram: UGT73A16 assay mixture of 2' hydroxy chalcone**:** single reaction product, chalcone 2'-*O*-glucoside **(B**) **Substrate:** 2' hydroxy chalcone

5.5 Kinetic studies of UGT73A16

5.5.1 Kinetic parameters of naringenin and genistein

To determine the initial velocity of UGT73A16, the assays were performed under steady state conditions using the standard assay system with various substrate concentrations. For naringenin and genistein the standard reaction mixture (200 µL) consisted of 100 µM UDP-glucose, 100 mM Tris-Cl (pH 8.0), 2 µM -18 µM of substrates and 12 µg of purified UGT73A16. The reaction was incubated at 37 ºC for 3 h and products were extracted twice with ethyl acetate and evaporated to dryness. All the reaction products were dissolved in equal volume of methanol and analyzed on reverse- phased HPLC on a C18 column. The known concentration of genistein 7-*O*-glucoside (**Fig. 3A.17 A**) and naringenin 7-*O*-glucoside (**Fig. 3A.18 A**) were used to plot a standard graph. The apparent K_m and V_{max} values for genistein and naringenin were determined by fitting the initial velocity data to the Michaelis-Menten equation using nonlinear regression analysis (**Fig. 3A.17 B & Fig. 3A.18 B**). The K_m and V_{max} values (**Table 3A.10**) of UGT73A16 for naringenin and genistein were determined at pH 8.0 and 37°C. According to k_{cat}/K_m ratio which reflects the enzyme catalytic efficiency, UGT73A16

used genistein most efficiently although a high enzymatic affinity towards naringenin was observed. The best substrates were naringenin and genistein (**Table 3A.10**).

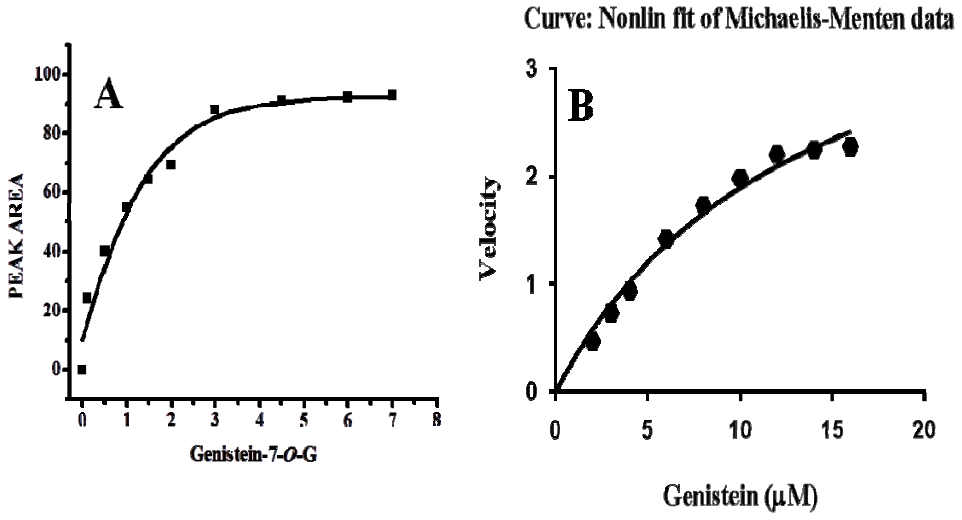


Fig 3A.17 Kinetic parameters of genistein with purified UGT73A16 **(A)** Genistein 7-*O*-glucoside standard plot **(B)** Michaelis-Menten plot

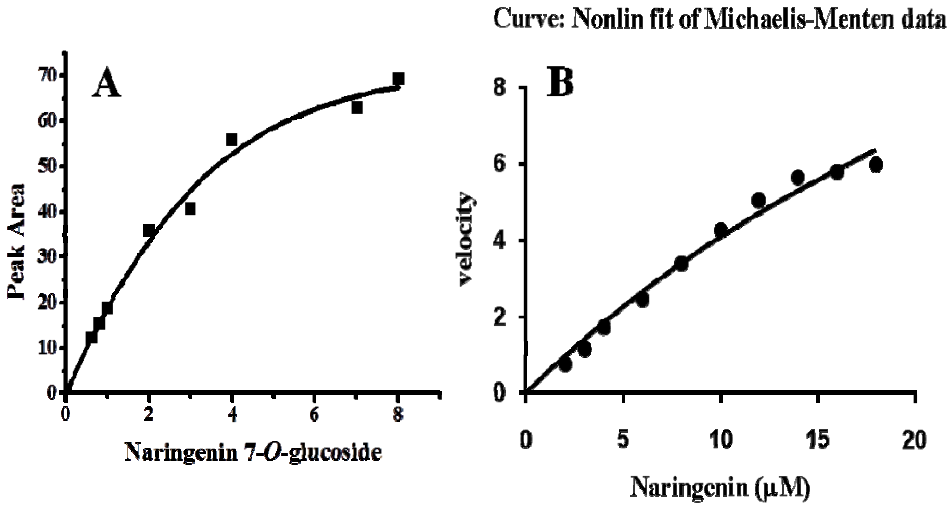


Fig 3A.18 Kinetic parameters of naringenin with purified UGT73A16 **(A)** Naringenin 7-*O*-glucoside standard plot **(B)** Michaelis-Menten plot

Table 3A.10 Substrate specificities of the purified UGT73A16 recombinant protein

Substrate	K_m **(µM)**	V_{max} **(nkat/mg)**	V_{max}/K_m	K_{cat}/K_m **(µM^{-1}/sec^{-1})**
Naringenin	42.51	357.40	8.40	0 .430
Genistein	13.63	83.123	6.09	0.317
Baicalein	47.08	21.50	0.45	0.023

5.5.2 Kinetic parameter of baicalein

To determine the initial velocity of UGT73A16, the assays were performed under steady state conditions using the standard assay system with various substrate concentrations. The standard reaction mixture (200 µL) consisted of 400 µM UDP-glucuronic acid, 100 mM Tris-Cl (pH 8.0), 10 µM -100 µM baicalein and 40 µg of purified UGT73A16. The reaction was incubated at 37 ºC for 3 h and processed as given above (section 5.4.1). The known concentration of baicalein 7-*O*-glucuronide (**Fig. 3A.19 A**) was used to plot a standard graph. The apparent K_m and V_{max} values for baicalein (**Table 3A.10**) were determined by fitting the initial velocity data to the Michaelis-Menten equation using nonlinear regression analysis (**Fig. 3A.19 B)**.

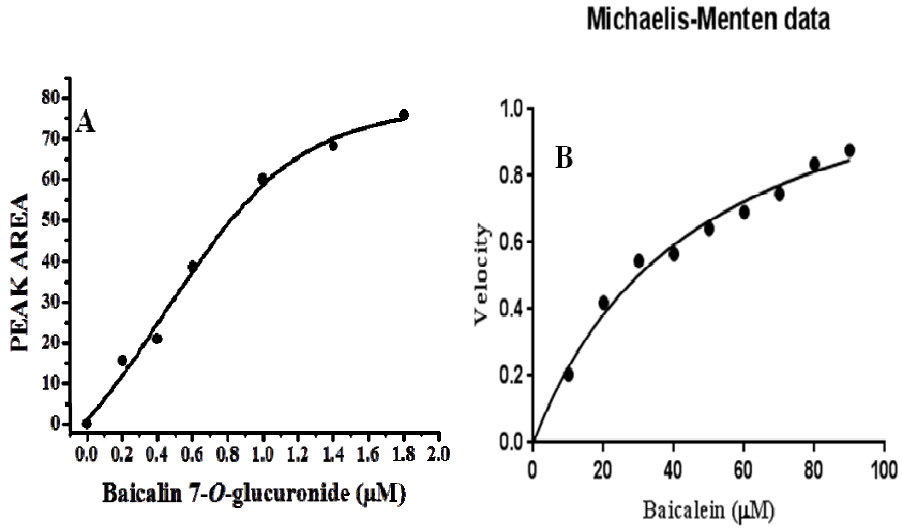


Fig. 3A.19 Kinetic parameters of baicalein with purified UGT73A16 **(A)** Baicalein 7-*O*-glucuronide standard plot **(B)** Michaelis-Menten plot

6.0 Conclusions

- An open reading frame of UGT73A16 from *Withania somnifera* was directionally cloned into pET 30b (+) expression vector.
- Recombinant UGT73A16 was transformed into *E. coli* BL-21(DE3) cells and heterologously expressed.
- Over expressed UGT73A16 protein was purified from inclusion bodies using Ni-NTA affinity chromatography.
- Optimization was done to express the recombinant UGT73A16 protein in soluble form (0.08 mM IPTG, 22 ^{0}C temp and 20 h time duration) and purified by affinity chromatography using Ni-NTA agarose.

- Enzyme assay was done using UDP-glucose and UDP-glucuronic acid as donor and genistein, daidzein, apigenin, naringenin, kaempferol, biochanin A, baicalein, 3-hydroxy flavone and 2' hydroxy chalcone as acceptor substrate.
- Product identification was done by high performance liquid chromatography (HPLC), hypsochromic shift and LC-MS.
- Kinetic parameters showed that UGT73A16 used genistein most efficiently and high enzymatic affinity was observed towards naringenin.

Chapter 3B

Withania somnifera glycosyltransferase (UGT73A16) as biocatalysts to produce plant specific glucosides

1.0 Summary

Plant secondary metabolites glucosides are very difficult to isolate or chemically synthesize in pure form because of low yield. Thus simple approaches for the production of such glucosides would be highly beneficial. In this regards we explored UGT73A16 of *Withania somnifera* as whole cell catalyst in *E. coli*. We used flavonoids (genistein, apigenin, kaempferol, naringenin, biochaninA and daidzein) as substrates for this study because flavonoids and their glucosides have important medicinal properties and they provide a complex structure for regioselective glycosylation. Approximate 21 mg/L to 22 mg/L of either apigenin 7-*O*-glucoside or naringenin 7-*O*-glucoside was produced after feeding apigenin or naringenin to *E. coli* expressing UGT73A16. More than 95% of the glucosides products were released into the medium, facilitating their isolation.

2.0 Introduction

Plant derived glucosides have attracted great attention due to their widespread applications. The isolation and purification of significant amount of specific flavonoid glucosides from plant source is time consuming since plants contain wide and variable spectra of glucosides with different types of sugar attachment (**Harborne and Baxter 1999**). On the other hand chemical synthesis of specific glucoside is notoriously complicated. The flavonols aglycone has five potential glycosylation sites and to synthesize any single monoglucoside, four other hydroxyl groups must be blocked **(Bouktaib et al. 2002; Li et al. 2002**). Thus the sequential blocking and deblocking of the hydroxyl groups for regioselective glycosylation is also time consuming and complicated process and presence of some impurity also found (**Bouktaib et al. 2002**). The purpose of this study was to explore the utility of plant UGTs as regioselective biocatalyst. Nucleotide activated sugar is the essential component in small molecule glycosylation. One of the major problems for large scale application of glycosylation is the provision of UDP-glucose to the in vitro system. Although different approaches including chemical method (**Kretzschmar and Stahl 1988**), enzymatic synthesis (**Bulter and Elling 1999**) and regeneration using UDP-glucose pyrophosphorylase and pyrophosphatase for synthesis of nucleoside diphosphate sugars (**Heidlas et al. 1992; Wong et al. 1992**) have been reported, these approaches are either laborious, difficult or require sugar phosphates; phosphoenolpyruvate and 5' triphosphate and therefore involve high cost. Since UDP-glucose is a natural intermediate in cell wall synthesis in bacterial cells, using glycosyltransferases engineered in bacteria to synthesize small molecule glucosides should be an efficient system to overcome the difficulty associated

with the preparation of nucleotide sugars. Similar studies on production of flavonoid glucosides in *E. coli* have recently been reported (**Lim et al. 2004; Willits et al. 2004; Xian et al. 2008**). Our recent research on UDP-glucose: glycosyltransferase (UGT73A16) of *W. somnifera* offers a new possibility of exploring regioselective glycosylation of genistein, apigenin, daidzein, naringenin, kaempferol and biochaninA. Here we report the production of milligram quantities of the 7-*O*-glucoside of iso-flavones (genistein, daidzein and biochaninA), flavones (apigenin), flavanone (naringenin) and 3-*O*-glucoside of flavonols (kaempferol) in *E. coli* cultures expressing UGT73A16.

3.0 Materials

3.1 Preparation of Glasswares and plasticwares

The details of glasswares and plasticwares have been described previously in chapter 2.

3.2 Chemicals

All the aglycones, iso-flavones (genistein, daidzein and biochaninA), flavones (apigenin), flavanone (naringenin) and flavonols (kaempferol) were purchased from sigma (Sigma, USA) and their glucosides were purchased from chromadex (www.chromadex.com). The protein expression vector pET 30 b (+) and *E. coli* strain BL21 (DE3) were purchased from Novagen (Novagen, USA). All the solvents of analytical and HPLC grade were purchased from HiMedia, Qualigens fine chemicals and E-Merck Laboratories, India.

4.0 Methods

4.1 UGT73A16 as whole cell biocatalyst

E. coli BL21 (DE3) strains carrying the recombinant UGT73A16 construct were grown in 50 ml Luria-Bertani (LB) medium at 37 ^{o}C until OD_{600} reached 0.6-0.7. IPTG was then added to the cultures to a final concentration of 0.08 mM. *E. coli* culture harboring empty vector pET 30 b (+) was used as a control. At the same time 100 µM of each substrates; genistein (27.02 mg/L), daidzein (25.4 mg/L), biochaninA (28.4 mg/L), apigenin (27.02 mg/L), naringenin (27.2 mg/L) and kaempferol (28.6 mg/L) dissolved in dimethyl sulfoxide (DMSO) were then added to the induced cultures. After an appropriate incubation time (12 h, 16 h, 20 h, 24 h and 48 h) at 22 ^{o}C; cultures were centrifuged at 5000×g for 10 min to obtain the supernatant fraction and cell pellet fraction. The cell pellet was resuspended in ethyl acetate to extract aglycone and its glucosides and centrifuged at 10,000×g for 5 min to remove the cell debris and the supernatant fraction was also used for ethyl acetate extraction.

4.2 HPLC analysis of glycosylated products

Both the supernatant and cell fractions were evaporated to dryness and dissolved in methanol for high performance liquid chromatography (HPLC) analysis. Reverse phase HPLC analysis was carried out on a Perkin Elmer system equipped with a diode array detector (DAD) as given in chapter 3A (Section 4.2.3).

5.0 Results and discussions

5.1 Production of glucosides using UGT73A16 as biocatalysts in *E. coli*

UGT73A16 is a flavonoid glycosyltransferase from *W. somnifera*, which can transfer glucose from uridine diphosphate (UDP)-glucose to aglycone substrates (genistein, daidzein, biochaninA, apigenin, naringenin and kaempferol) to produce its glucoside in vitro (Chapter 3A). To investigate the production of glucosides by recombinant *E. coli in vivo*, we incubated induced cell cultures of *E. coli* BL21 (DE3) carrying the recombinant UGT73A16 with equal concentration of each substrates for different times. The medium and cell pellet were extracted with ethyl acetate and products were analyzed by HPLC. More than 95% of glucoside products were released into the medium extract than the cell pellet extract (**Table 3B.1**). The expected glucosides were produced and could be detected in the culture medium after 12 h (**Fig. 3B.1, Table 3B.0**).

Table 3B.0 Glucoside synthesis in *E. coli* in LB medium (100 µM of each substrate)

Glucoside	**LB medium (mg/L)**					
	12 h	**16 h**	**20 h**	**24 h**	**48 h**	**% Rate of conversion (24 h)**
Genistein 7-*O*-glucoside	5.4±0.1	6.83±0.07	11.45±0.5	14.73±0.1	16.21±0.3	54.51
Apigenin 7-*O*-glucoside	2.96±0.13	5.59±0.1	10.15±0.05	21.52±0.2	21.71±0.21	79.64
Kaempferol 3-*O*-glucoside	1.52±0.08	2.80±0.1	5.39±0.19	14.38±0.48	14.47±0.49	50.27
Naringenin 7-*O*-glucoside	8.8±0.6	15.21±0.11	18.14±0.06	21.83±0.07	22.5±0.32	80.25
BiochaninA 7-*O*-glucoside	1.02±0.18	4.6±0.1	9.58±0.08	14.26±0.31	10.68±0.52	50.21
Daidzein 7-*O*-glucoside	1.22±0.08	3.34±0.07	5.62±0.08	7.32±0.12	6.88±0.01	28.59

Throughout a time course of 24 h, the bulk of aglycone was converted to its specific glucoside and the glucoside product yield was about 7-21 mg/L culture medium with conversion rates from 28% to 80%. Most of the conversion happens during first 24 h, later conversion up to 48 h are very slow (**Table 3B.0, Fig. 3B.1**).

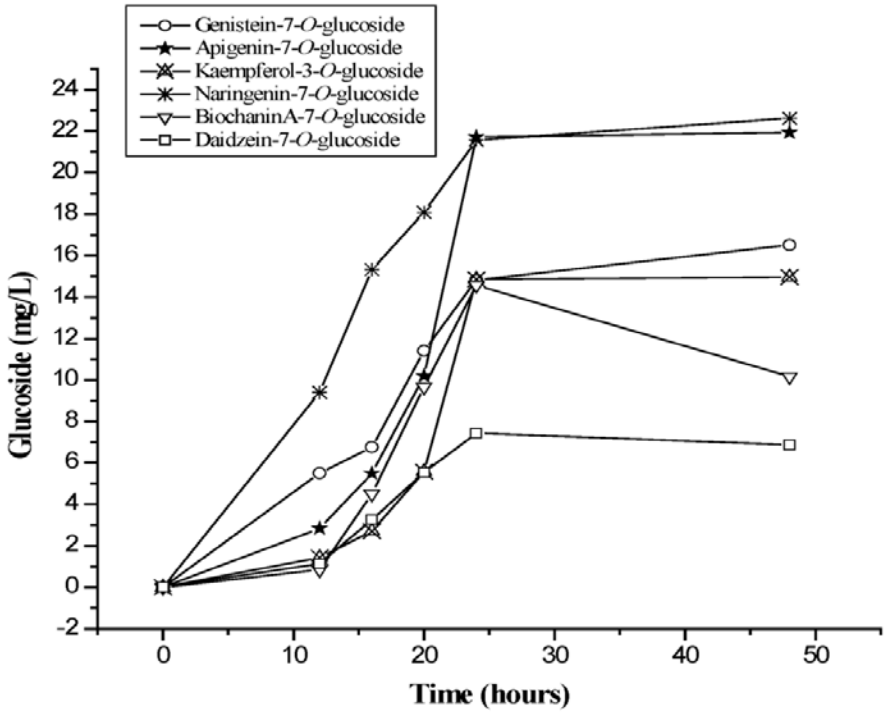


Fig. 3B.1 Production of glucosides; Time course showing levels of different glucosides in culture medium of engineered *E. coli* expressing UGT73A16 fed with different substrates

Table 3B.1 Comparison of glucoside products extracted from engineered *E. coli* cell pellet and medium at 24 h

Glucoside	Glucoside into medium (mg/L)	Glucoside into cell pellet (mg/L)
Genistein 7-*O*-glucoside	14.73±0.1	1.12±0.4
Apigenin 7-*O*-glucoside	21.52±0.2	1.32±0.3
Kaempferol 3-*O*-glucoside	14.38±0.48	1.38±0.12
Naringenin 7-*O*-glucoside	21.83±0.07	1.68±0.1
BiochaninA 7-*O*-glucoside	14.26±0.31	1.06±0.2
Daidzein 7-*O*-glucoside	7.32±0.12	Not detectable

6.0 Conclusions

- Recombinant UGT73A16 of *Withania somnifera* used as whole cell biocatalyst in *E. coli* using endogenously produced UDP-glucose as donor.
- Genistein, apigenin, kaempferol, naringenin, biochaninA and daidzeinn used as substrates for this study.
- Approximate 21 mg/L to 22 mg/L of either apigenin 7-*O*-glucoside or naringenin 7-*O*-glucosides were produced after feeding apigenin or naringenin to *E. coli* expressing UGT73A16.
- More than 95% of the glucosides products were released into the medium, facilitating their isolation.
- Result data showed that most of the conversion happens during first 24 h, later conversion up to 48 h are very slow.

Chapter 4

In silico modeling and docking studies of Withania somnifera glycosyltransferase (UGT73A16)

1.0 Summary

The glycosyltransferase UGT73A16 from *Withania somnifera* catalyzes glycosylation of the (iso) flavonoids, flavonols, flavonone and flavone. It transfers sugar molecule to hydroxyl groups of (iso) flavonoids, flavonols, flavonone and flavone with 3-hydroxyl and 7-hydroxyl group as the major product. It is well known from previous studies that the amino acid residues of glycosyltransferases are responsible for sugar donor and acceptor binding sites, which are located in the N and C-terminal ends respectively, as determined by modeling and docking studies. In this chapter modeling and docking studies of UGT73A16 was carried out to know the key residues involved in substrate recognition. A three-dimensional model of UGT73A16 was constructed based on the available crystal structures of plant UGTs. The PDB ID, 2PQ6 and 3HBJ from *Medicago truncatula* and 2VG8 from *Arabidopsis thaliana* were used as templates to construct the 3D structure of UGT73A16. The modeled structures were analysed by various tools such as Ramachandran plot, ERRAT analysis and ProSAII. The resulted models were assessed by various tools and the final refined model revealed GT-B type fold. Modeled UGT73A16 was docked with UDP-glucose, UDP-glucuronic acid, kaempferol, naringenin, daidzein, genistein and baicalein. The protein-ligand interactions showed that His 16, Asp 22, Ile 50, Asp 246, Lys 255, Ser 283, Trp 336, Ala 337, Gln 339, Val 340, Trp 357, Asn 358 and Glu 362 amino acid residues may be important for catalytic function.

2.0 Introduction

Glycosyltransferases have been grouped into more than 94 families based on sequence similarity. Plant UGTs are characterized by a highly conserved signature putative secondary plant glycosyltransferase (PSPG) motif (**Hughes and Hughes 1994**), close to the C-terminal end of amino acid sequence. This PSPG box consists of approximate 44 amino acids residues and might also define the binding of the activated sugar donors. There are several highly conserved amino acid motifs within this box including WAPQV and HCGWNS, which are highly conserved across phyla. These PSPG motifs have the low levels of sequence identity, especially within acceptor binding region, and this aspect may be crucial to account for the recognition of vast variety of acceptors and the synthesis of the large number of glycosides.

The relationship between primary amino acid sequence, substrate specificity and product regioselectivity of plant UGTs is complex and remains to be determined. The 3D glycosyltransferase database (http://www.cermav.cnrs.fr/glyco3d/) holds crystal

structures of 53 different GTs representing a total of 26 different CAZY families. Twenty six of the GTs with solved crystal structures adopt the GT-B fold and 10 of these belong to the family 1 GTs (**Fig. 4.1**) and analyzed structurally (**Buschiazzo et al. 2004; Fritz et al. 2004; Lobsanov et al. 2004; Mulichak et al. 2004**).

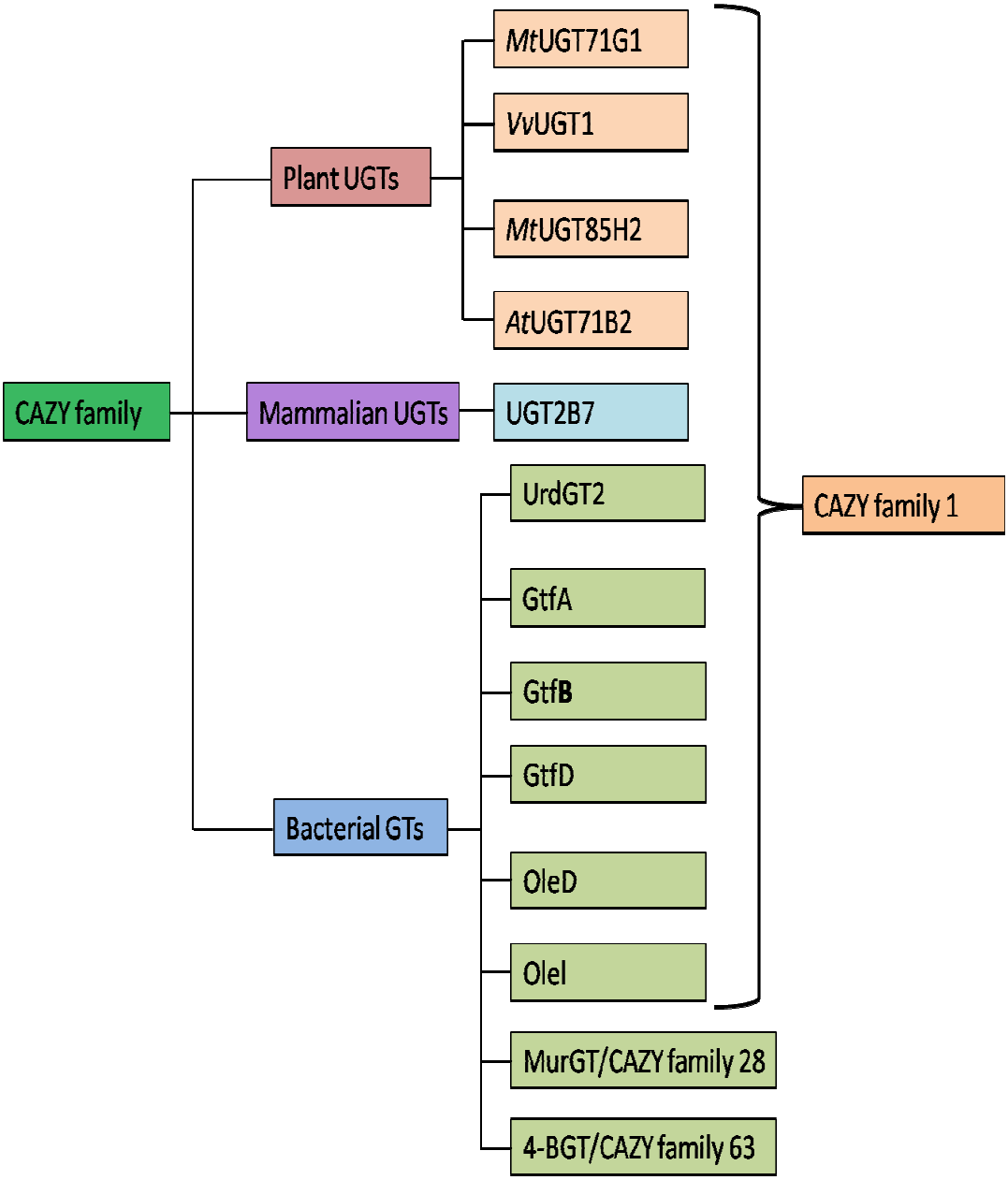


Fig. 4.1 List of GTs with available crystal structures

The crystal structures of plant UGTs from *Medicago truncatula, Mt*UGT71G1 (**Shao et al. 2005**), *Vitis vinifera, Vv*GT1 (**Offen W 2006**), *Medicago truncatula, Mt*UGT85H2 (**Li et al. 2001**) and *Aribidopsis thaliana, At*UGT72B1 (**Brazier-Hicks et al. 2007**) reveal the interactions between the protein and substrates.

These studies also unveil the role of UGT signature motif in substrate recognition and the catalytic activity of enzyme. Until now only two folds have been observed for GTs: fold GT-A, consisting of one α/β/α sandwich domain and characterized by presence of divalent cation in the binding site while GT-B fold has two such domains.

Here, we report the structural analysis of a new UGT, UGT73A16 from *Withania somnifera*. Donor and acceptor substrates screening using bioinformatics analysis showed that it has close resemblance towards UDP-glucose, UDP-glucuronic acid, flavonoids, iso-flavonoids, flavones and flavonone. The structure analysis of UGT73A16 will provide a basis for understanding the complex and often overlapping substrate specificity and regiospecificity within the plant UGT superfamily.

3.0 Materials and methods

3.1 Sequence analysis, alignment and secondary structure prediction

Glycosyltransferase protein sequence of *W. somnifera*, UGT73A16 (NCBI GenBank Accession No. FJ654696) was retrieved from NCBI GenBank database (http://www.ncbi.nlm.nih.gov). Blast algorithm against protein data bank (PDB) was used to carry out the sequence homology searches. The sequence and 3D structures of template proteins were extracted from PDB database (**Berman et al. 2000**). Multiple alignments of the amino acid sequences were carried out with clustal W1.8 program (http://www.ebi.ac.uk/clustalw/). Conserved domain in UGT73A16 was detected using Conserved Domain Database search tool (CDD) on NCBI server (http://www.ncbi.nlm.nih.gov/structure/cdd/wrpsb.cgi). The PSIPRED V.3 programme was used to predict secondary structure of UGT73A16 amino acid residues (**McGuffin et al. 2000**).

3.2 Comparative modeling of UGT73A16

An X-ray crystal structure of *M. Trunctulla* glycosyltransferase (PDB code: 2PQ6 and 3HBJ) and *A. thaliana* (PDB code: 2VG8) were used as a templates (**Mondolo et al. 2009; L. Li. et al. 2007**) to build three dimensional model of UGT73A16. UGT73A16 homology models were generated using software Modeller 9v10 (**Sali and Blundell 1993**). Out of 50 models generated, the model with the lowest DOPE scores was taken as the final model and energy minimization was done with Chimera 1.6.2 (www.ucsf.edu).

3.3 Model evaluation

Stereochemical analyses of the UGT73A16 homology model was carried out using Ramachandran plot obtained from PROCHEK (Laskowski et al. 1996), overall quality factor by ERRAT analysis (**Colovos and Yeates 1993**) and Z-score from ProSAII

(**Sippl 1993**). Superimposition and calculation of the root mean square deviations (RMSD) between model and template was made by 1.6.2 chimera software using the carbon alpha fitting method. Diagrammatic representations of the structures were made using PyMOL 1.3 software (**DeLano 2002**).

3.4 Docking studies of UGT73A16 model with acceptor and donor substrates

The structure data files (SDF) of the acceptor and donor substrates were obtained from Pub Chem Database. The structures were converted to PDB format using PyMOL software. The UGT73A16 structure and substrate ligands were prepared using AutoDock Tools v.1.5.2 software (**Sanner 1999**). Molecular docking was performed using Autodock vina 1.1.2 software (Trott and Olson 2010). The interactions of the ligands with UGT73A16 were visualized and figures were formed using PyMOL 1.3. The lowest binding energy conformation in the first cluster was considered as the most favourable docking pose. Hydrogen bonds, bond lengths and close contacts between enzyme active site and substrate atoms were analyzed.

4.0 Results and discussions

4.1 Model construction

The secondary structure prediction run at the PSIPRED server showed that the UGT73A16 consists of 40% α-helix (17 helices; 189 residues), 10% extended-beta (11 strands; 47 residues) and 50% random coil (28 coils; 234 residues) configuration (**Fig. 4.2**).

Homology modeling relies on establishing an evolutionary relationship between the sequence of a protein of interest and other members of the protein family whose structures have been solved experimentally by X-ray crystallography or NMR. UGT73A16 amino acid sequence alignment with available protein sequences showed a significant percentage of identity with *Medicago truncatula* flavonoid 3-*O*-glucosyltransferase (3HBJ A: 34%), *Medicago truncatula* UDP-glucuronosyl/UDP-glucosyltransferase (2PQ6 A: 32% identity), *Arabidopsis thaliana* hydroquinone glucosyltransferase (2VG8: 31%). These proteins were selected as templates for the modeling. In general, sequence identities of 30% are enough to construct the 3D model of target proteins through homology modeling.

Due to low sequence identity SALIGN was employed to construct multiple structure alignments of templates. The 3D models of UGT73A16 were generated by aligning the target sequence with this multiple structures-based alignment.

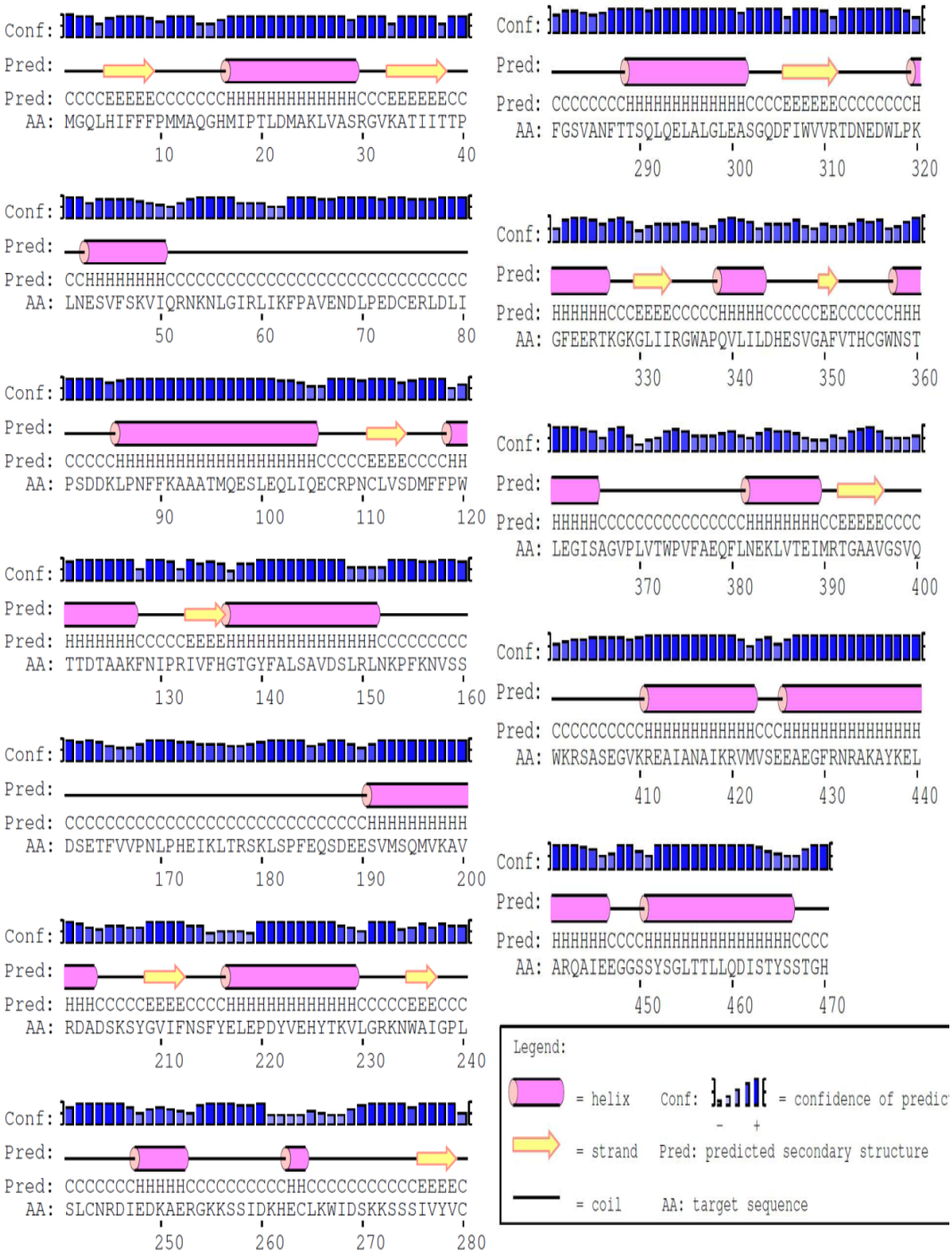


Fig. 4.2 Secondary structure prediction of UGT73A16 predicted by PSIPRED server

4.2 Validation of modeled UGT73A16

A total 50 models were generated with template 3HBJ, 2PQ6 and 2VG8 and their discrete optimized protein energy (DOPE) were calculated. The model No. 9 (UGT73A16.B99990009.pdb) having lowest DOPE score (-53072.312) was considered as the best model of the UGT73A16 followed by energy minimization. Structural analysis of modeled UGT73A16 was performed by PROCHEK, ERRAT and ProSAII.

4.2.1 Ramachandran plot analysis

Ramachandran plot provided by the program PROCHECK, assured very good confidence for the predicted protein with 87.1% residues in most favoured region, 12.4% in additional allowed region, 0.5% in generously allowed region and 0.0% in disallowed region (**Fig. 4.3 and Table 4.0**).

Table 4.0 Ramachandran plot statistics

Plot statistics	No. of residues	%
Residues in most favoured regions [A,B,L]	365	87.1%
Residues in additional allowed regions [a,b,1,p]	52	12.4%
Residues in generously allowed regions [~a, ~b, ~1, ~p]	2	0.5%
Residues in disallowed regions	0	
Number of non-glycine and non-proline residues	419	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (Shown as triangles)	29	
Number of proline residues	20	
Total number of residues	470	

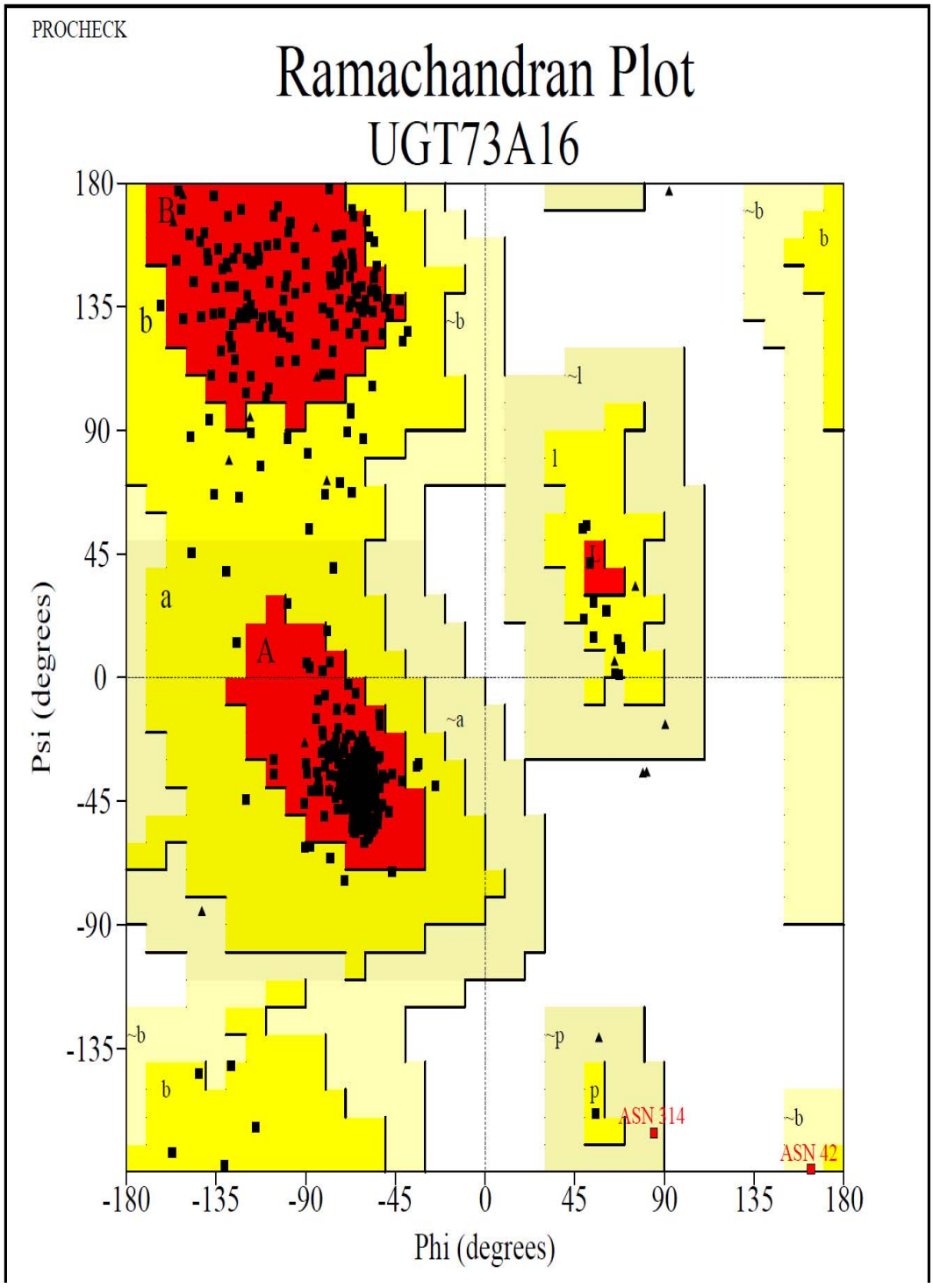


Fig. 4.3 Ramachandran plot (PROCHEK) showing the dihedral angles Psi and Phi of amino acid residues, in which residues lie in most favoured regions are in red curves (ABL) and additional allowed regions are [a, b, l, p] in dark yellow curves

4.2.2 ERRAT plot analysis

The overall quality factor of modeled UGT73A16 in ERRAT analysis was 82.340, expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit (**Fig. 4.4**).

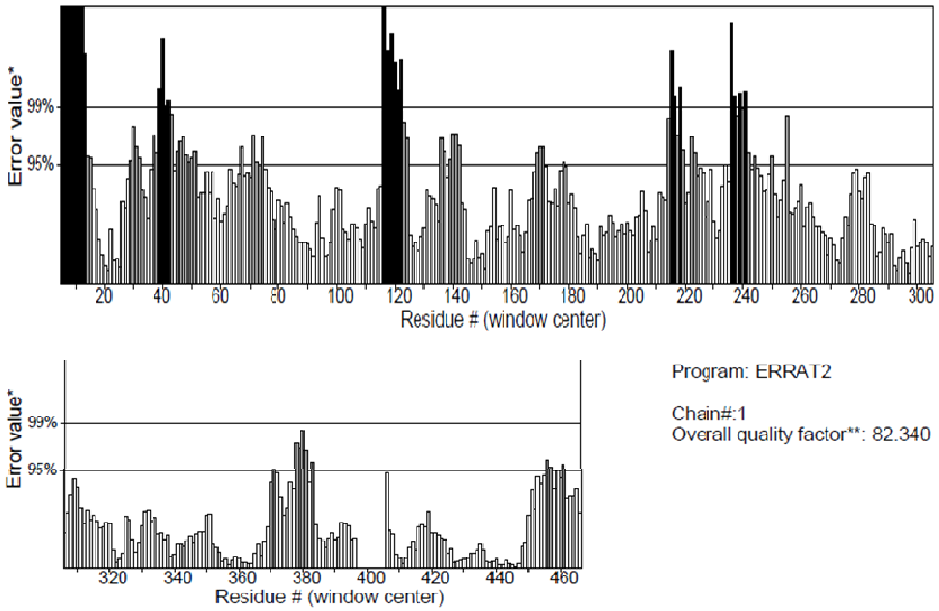


Fig. 4.4 ERRAT plot analysis of modeled UGT73A16

4.2.3 ProSAII analysis of modeled UGT73A16 with templates

The ProSAII program (Protein Structure Analysis) is an established tool, which is frequently employed in validation of experimental protein structures. It generates Z-score of model, which 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. ProSAII analysis showed that protein folding energy of our modeled structure was in good agreement with that of the template. It showed a Z-Score of -10.00, implying no more significant deviation from the templates 2PQ6 (-10.07), 2VG8 (-11.73) and 3HBJ (-11.44). The Z-score value obtained for modeled UGT73A16 (-10.00) indicated its location within the space of protein related to X-ray (Dark black point, **Fig. 4.5**). The value is quite comparable with template 2PQ6 (-10.07) suggested that obtained model was reliable. Finally the resultant energy minimized modeled UGT73A16 satisfying evaluation criteria was further used for docking analysis with ligands.

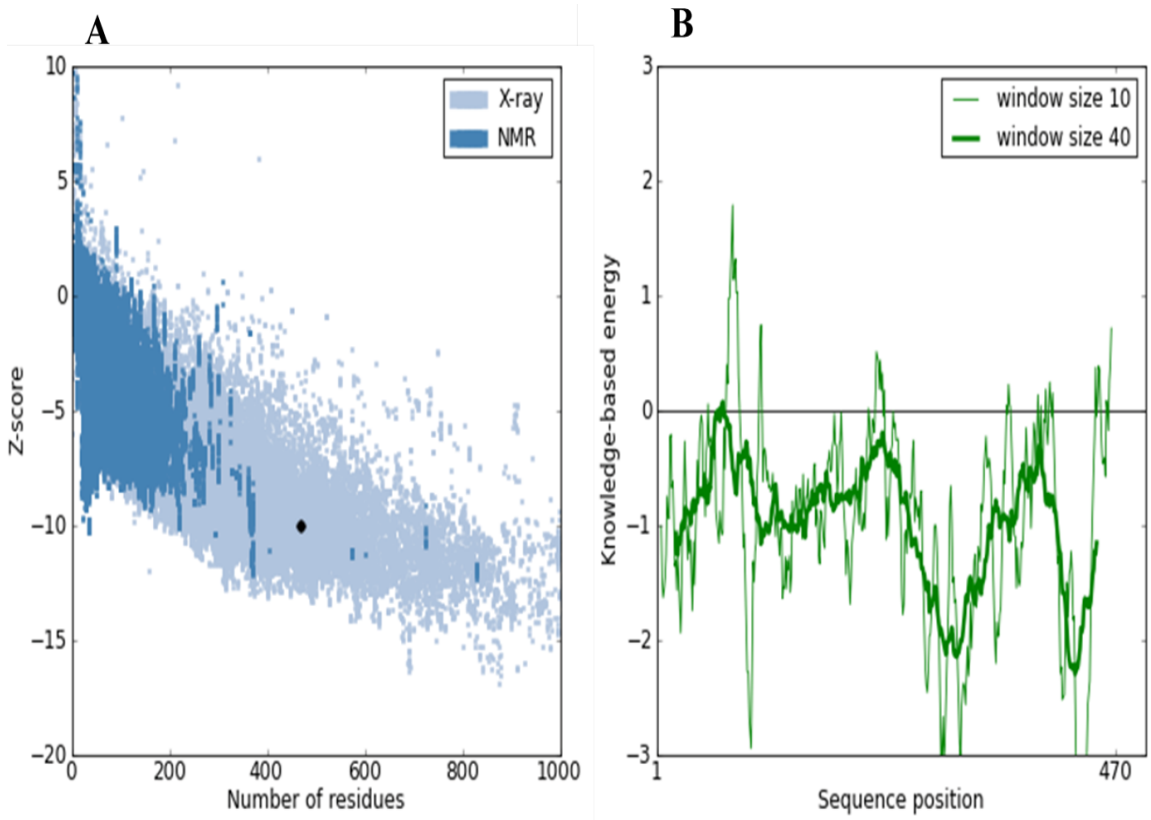


Fig. 4.5 (A) The plan of Z-Score shows spot of Z-score's value of protein determined by NMR (represented in dark blue colour) and by X-ray (represented in light blue colour) using ProSAII program. The black dot represents Z-Score of the model; **(B)** Energy plot obtained from ProSAII.

4.2.4 Superimposition of modeled UGT73A16

RMSD is used to make a quantitative comparison between modeled 3D structure and template, which revealed high similarity in topology and overall fold of the model with templates. The RMSD value of alpha carbon (Cα) of the modeled UGT73A16 was calculated by superimposition with templates 2PQ6, 2VG8 and 3HBJ on chimera (**Fig. 4.6**) and it was found to be 0.771 Å, 0.958 Å and 1.221 Å respectively.

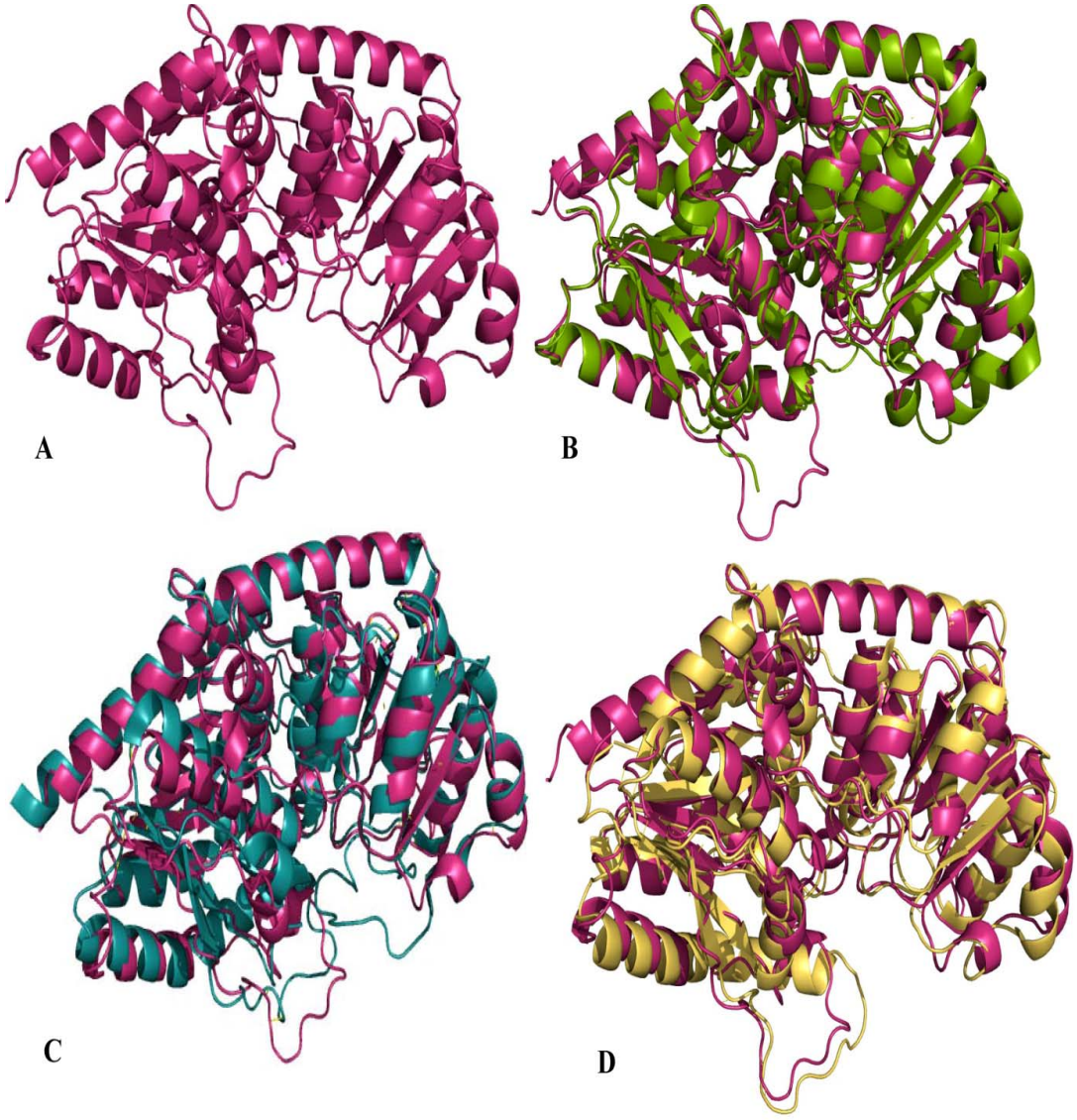


Fig. 4.6 (A) Homology model of UGT73A16 represented in ribbon diagram. **(B)** Superimposition of 3D structure of UGT73A16 (pink) and *M. truncatula* UDP-glucuronosyl/UDP-glucosyltransferase (2PQ6: green). **(C)** Superimposition of 3D structure of UGT73A16 (pink) and *Arabidopsis thaliana* hydroquinone glucosyltransferase (2VG8: Dark cyan). **(D)** Superimposition of 3D structure of UGT73A16 (pink) and *Medicago truncatula* flavonoid 3-*O*-glucosyltransferase (3HBJ: Yellow).

4.3 Molecular interactions of ligands with modeled UGT73A16

4.3.1 Docking of kaempferol and UDP-glucose with UGT73A16

Kaempferol substrate molecule and UDP-Glucose molecule was docked to the UGT73A16 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

(**Fig. 4.7**). 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.

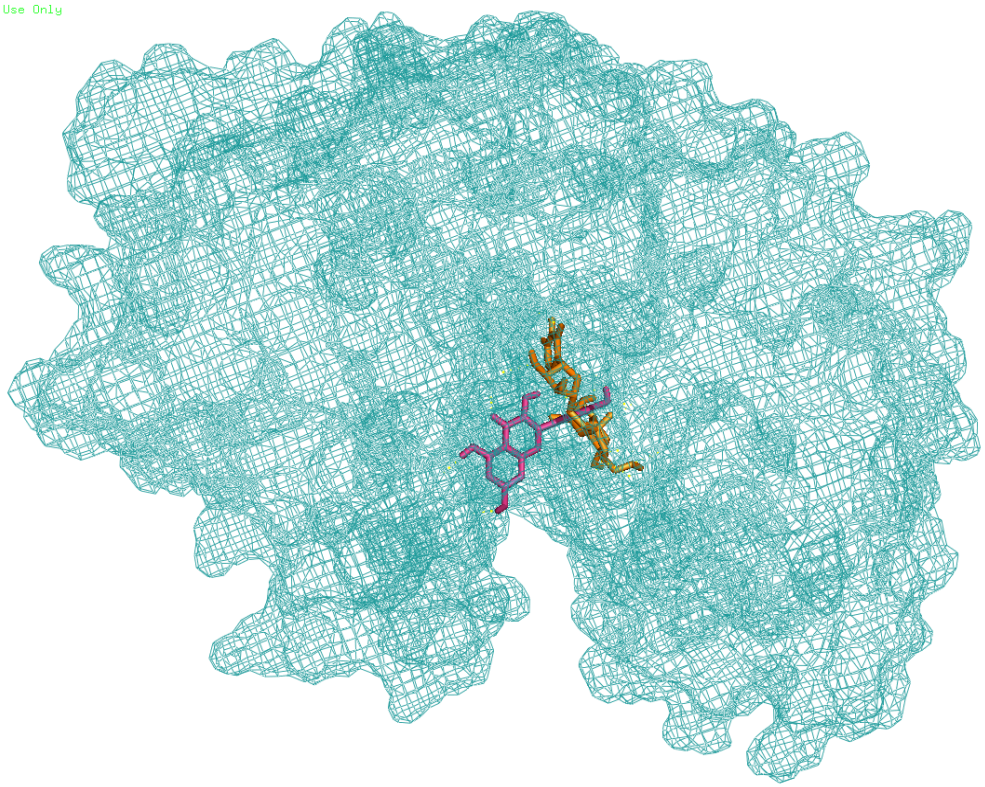


Fig. 4.7 Binding pocket of the kaempferol (pink) and UDP-glucose (dark orange)

The interaction energy of kaempferol with UGT73A16 was -7.5 kcal/mol. Receptor ligand interactions are observed in PyMOL. Amino acid residues in the active site for substrate binding comprise Asp 22, Lys 255, Trp 336, Ala 337, Gln 339 and Glu 362. Asp 22 (OD2) interacts with 3-OH group of kaempferol and UDP-glucose, Lys 255 (O) and Glu 362 (CD) interacts with 4'OH group and 4-O group of kaempferol respectively. Trp 336 (NE1), Ala 337 (O) and Gln 339 (H) interact with 7-OH group of kaempferol and UDP-glucose (**Fig. 4.8**). All the hydrogen bond distances between UGT73A16 and kaempferol complex were observed within the range of 2.1 Å to 2.7 Å.

The interaction of Asp 22, Trp 336, Ala 337 and Glu 339 with 3-OH and 7-OH group of kaempferol strongly suggested that reaction product can be either kaempferol 3-*O*-glucoside or kaempferol 7-*O*-glucoside. The interaction energy of UDP-glucose with UGT73A16 was -8.6 kcal/mol. In the template, active sites for UDP-Glucose comprise Asp 246, Lys 255, Ser 257, Tyr 278, Trp 336, Ala337, Gln 339, Val 340, Asn 358 and Glu 362. All the hydrogen bond distances between UGT73A16 and UDP-Glucose complex were observed within the range of 1.8 Å to 3.1 Å (**Fig. 4.8**).

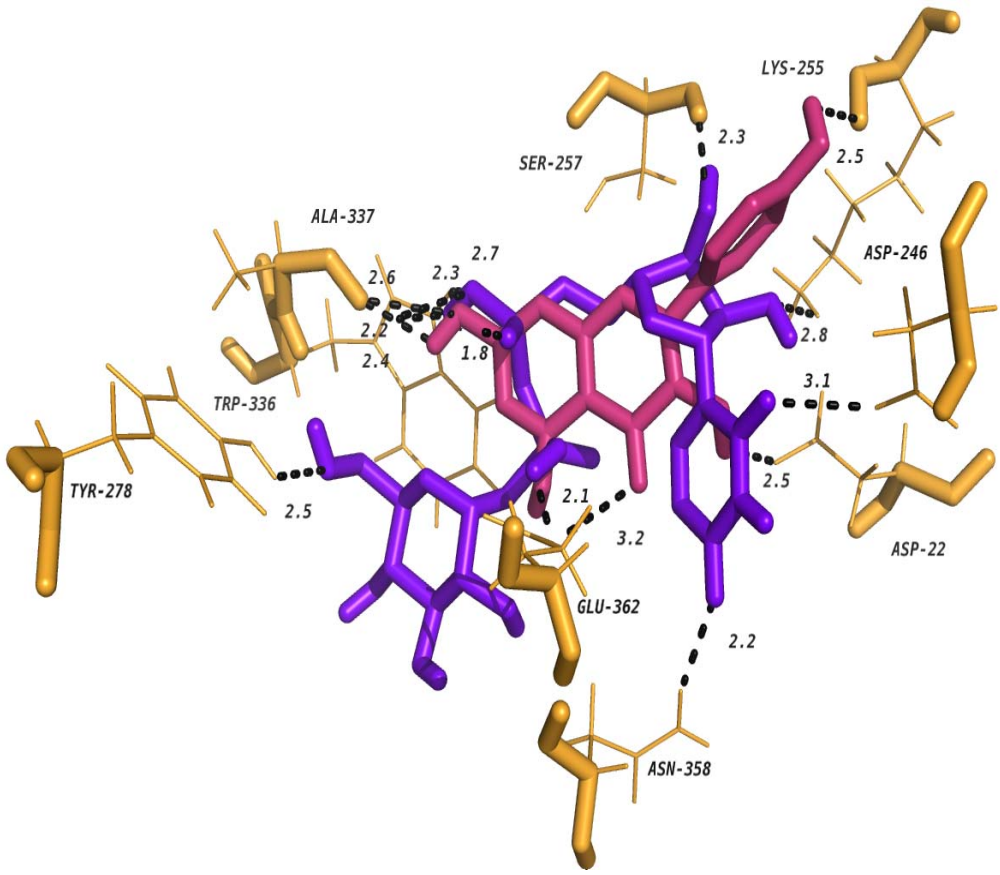


Fig. 4.8 Interaction of catalytic residues of UGT73A16 with substrate kaempferol (pink) and UDP-glucose (blue). Hydrogen bonding interactions are indicated by black dashed lines.

4.3.2 Docking of naringenin and UDP-glucose with UGT73A16

The interaction energy of naringenin with UGT73A16 was -8.6 kcal/mol. The important amino acid residues in the active site for substrate binding comprise Asp 246, Trp 336, Gln 339 and Asn 358. Asp 246 (OD2) interacts with 5-OH and Asn 358 (HD21) interacts with 7-OH group of naringenin. Trp 336 (HE1) and Gln 339 (H) interact with 4'-OH group of naringenin (**Fig. 4.10**). All the hydrogen bond distances between UGT73A16 and naringenin complex were observed within the range of 1.9 Å to 2.4 Å.

The interactions of Trp 336 and Gln 339 with 4'-OH group of naringenin as well as UDP-glucose and interactions of Asn 358 with 7-OH group of naringenin as well UDP-glucose suggesting that glycosylated product can be either naringenin 7-*O*-glucoside or naringenin 4'-*O*-glucoside or both.

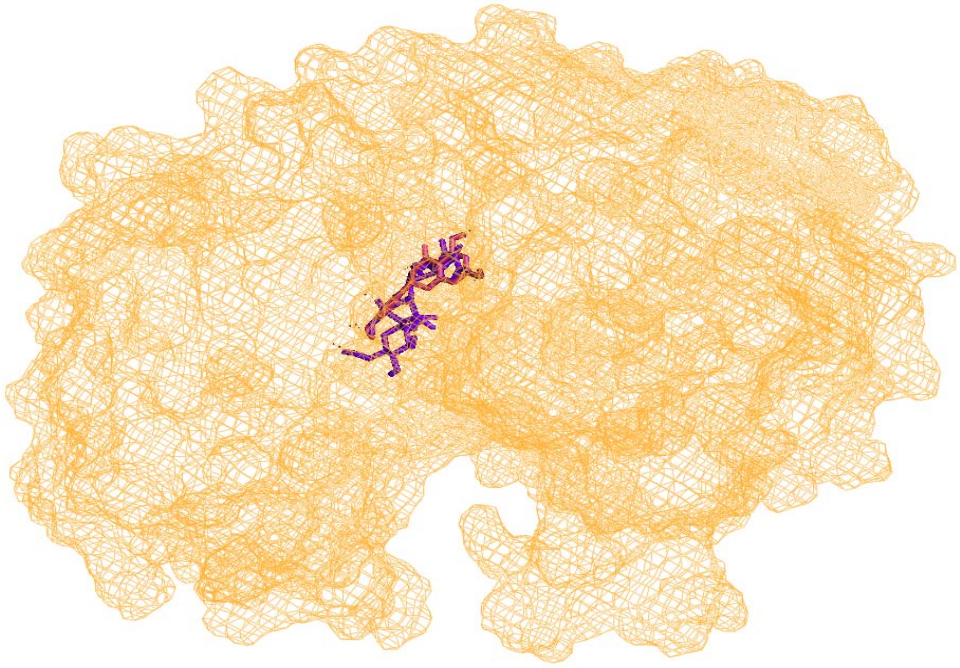


Fig 4.9 Binding pocket of the naringenin (pink) and UDP-glucose (blue)

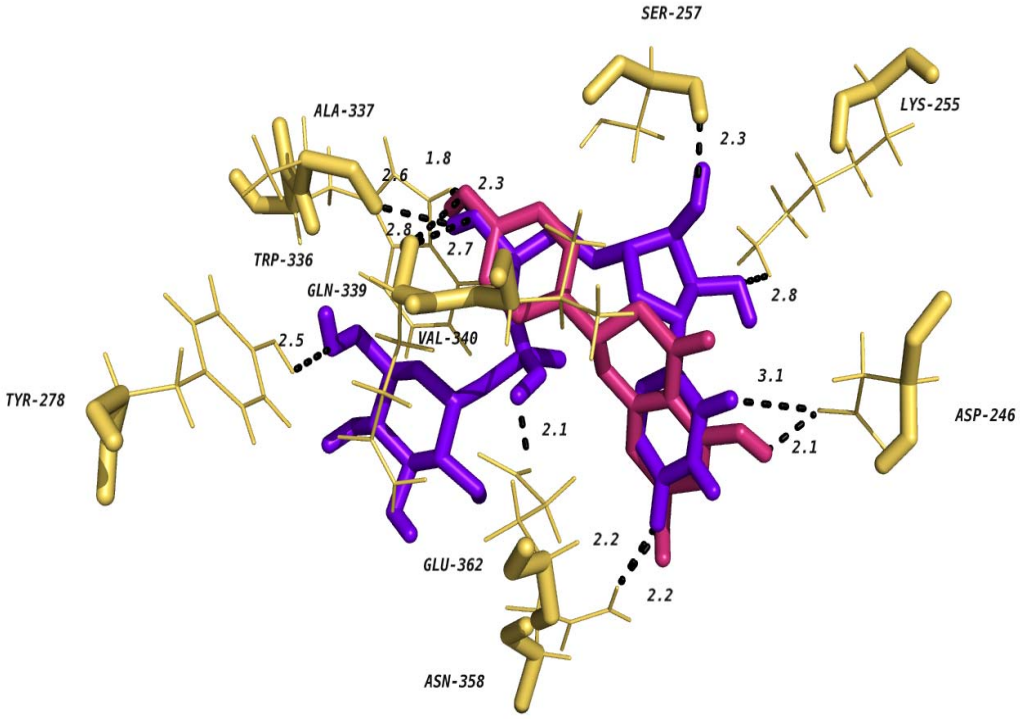


Fig. 4.10 Interaction of catalytic residues of UGT73A16 with substrate naringenin (pink) and UDP-glucose (blue). Hydrogen bonding interactions are indicated by black dashed lines.

4.3.3 Docking of daidzein and UDP-glucose with UGT73A16

The interaction energy of daidzein with UGT73A16 was -8.2 kcal/mol. In the template, active site for substrate binding comprises Glu 362. Glu 362 (CD) interacts with 7-OH

group of daidzein as well as UDP-glucose (**Fig. 4.12**). These interactions of UDP-glucose and daidzein with Glu 362 suggested that reaction product can be daidzein 7-*O*-glucoside. The other active amino residues Asp 255, Ser 257, Trp 336, Ala 337, Gln 339, Val 340 and Asn 358 interact only with UDP-glucose. All the hydrogen bond distances between UGT73A16 and daidzein complex were observed within the range of 2.1 Å to 2.4 Å.

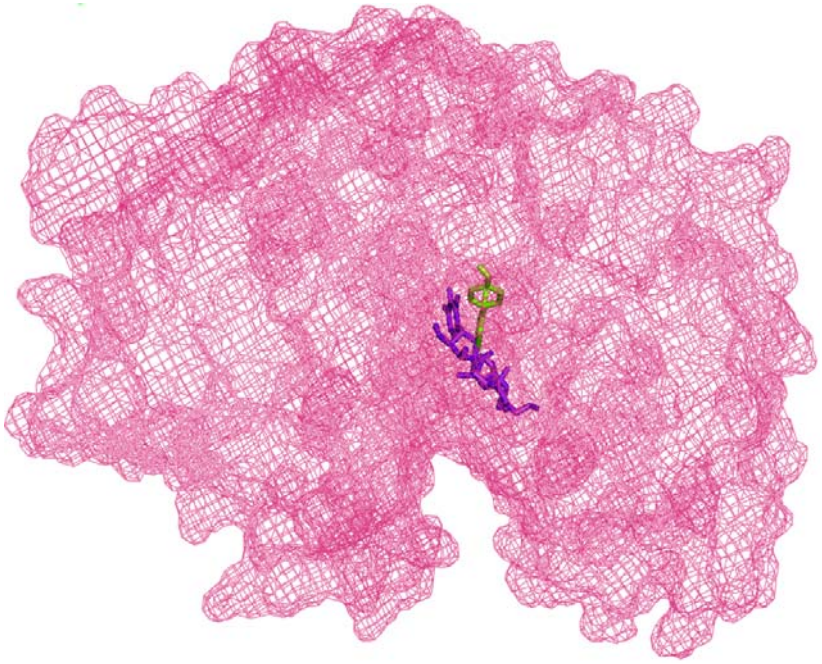


Fig 4.11 Binding pocket of the daidzein (green) and UDP-glucose (purple)

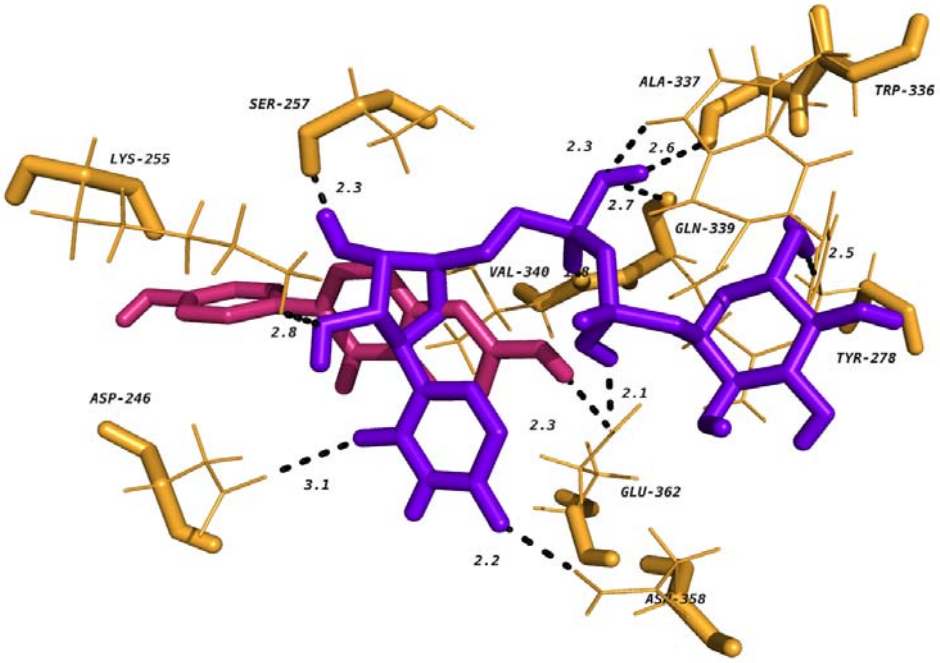


Fig. 4.12 Interaction of catalytic residues of UGT73A16 with daidzein (pink) and UDP-glucose (purple). Hydrogen bonding interactions are indicated by black dashed lines

4.3.4 Docking of genistein and UDP-glucose with UGT73A16

The interaction energy of genistein with UGT73A16 was -8.5 kcal/mol. In the template, active site for substrate binding comprises Asp 246, Lys 255 and Asn 358. Asn 358 (HD21) interacts with 4'-OH group of genistein and UDP-glucose. On the other hand Lys 255 (HZ3) and Lys 255 (HZ2) interact with 7-OH group of genistein and UDP-glucose respectively. These results showed that reaction product can be either genistein 7-*O*-glucoside or genistein 4'-*O*-glucoside or both (**Fig. 4.13** and **Fig. 4.14**).

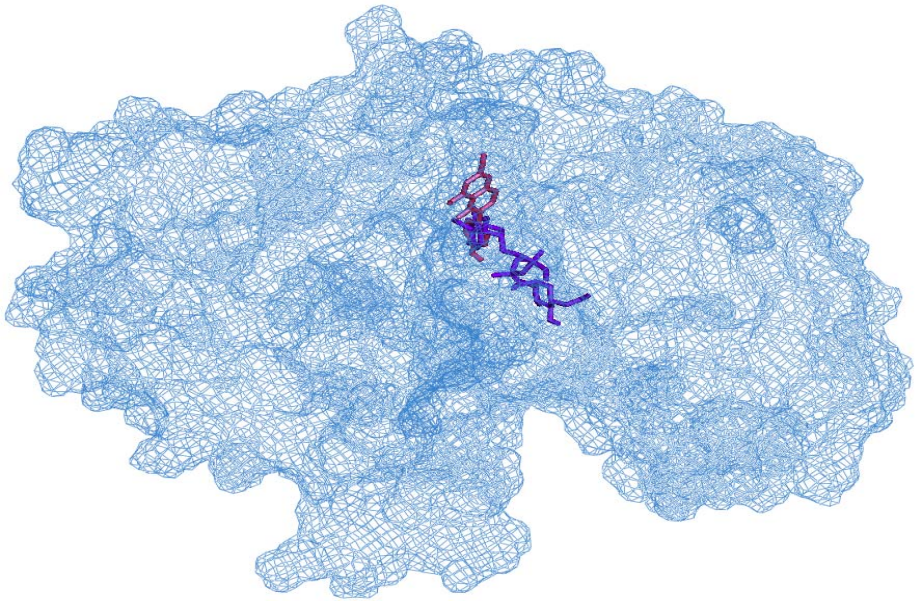


Fig. 4.13 Binding pocket of the genistein (pink) and UDP-glucose (purple)

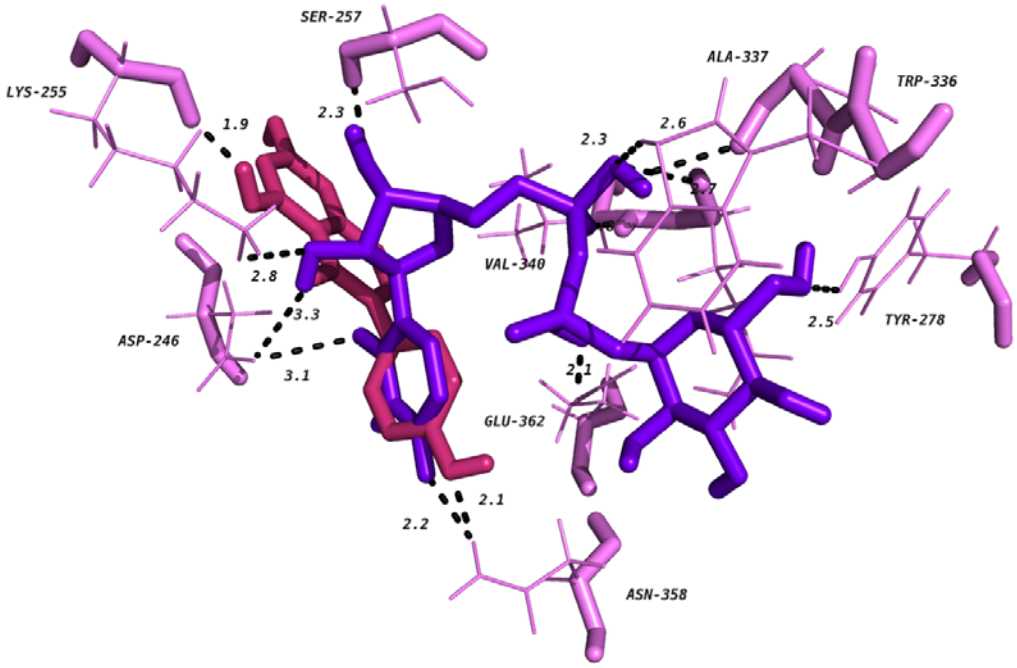


Fig. 4.14 Interaction of catalytic residues of UGT73A16 with genistein (pink) and UDP-glucose (purple). Hydrogen bonding interactions are indicated by black dashed lines

4.3. Docking of baicalein and UDP-glucuronic acid with UGT73A16

The interaction energy of baicalein with UGT73A16 was -9.0 kcal/mol. In the template, active sites for substrate binding comprise His 16, His 354 and Asn 358. HE2 of His 16 interacts with 1-O, HE2 (His 354) interacts with 5-OH group and H of Asn 358 interacts with 7-OH group of baicalein molecule (**Fig. 4.15**). All the hydrogen bond distances between UGT73A16 and baicalein complex were observed within the range of 2.0 Å to 2.3 Å.

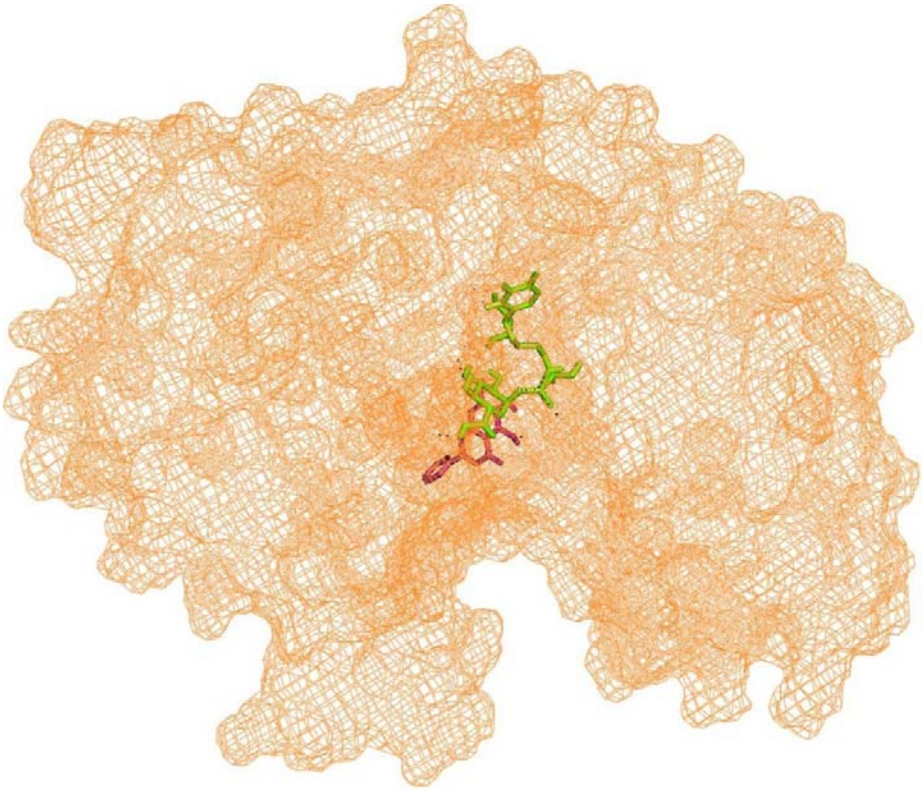


Fig 4.15 Binding pocket of the baicalein (pink) and UDP-glucuronic acid (green)

The interaction energy of UDP-glucuronic acid (UDPGA) with UGT73A16 was -8.9 kcal/mol. In the template, active site for UDP-Glucuronic acid comprises Asp 246, Lys 255, Ser 257, Val 340 and Asn 358. All the hydrogen bond distances between UGT73A16 and UDP-Glucuronic acid complex were observed within the range of 2.0 Å to 3.3 Å (**Fig. 4.16**). Asn 358 (H) interacts with 7-OH group of baicalein and UDP-glucuronic acid. This interaction suggested that reaction product can be baicalin (baicalein 7-*O*-glucuronide).

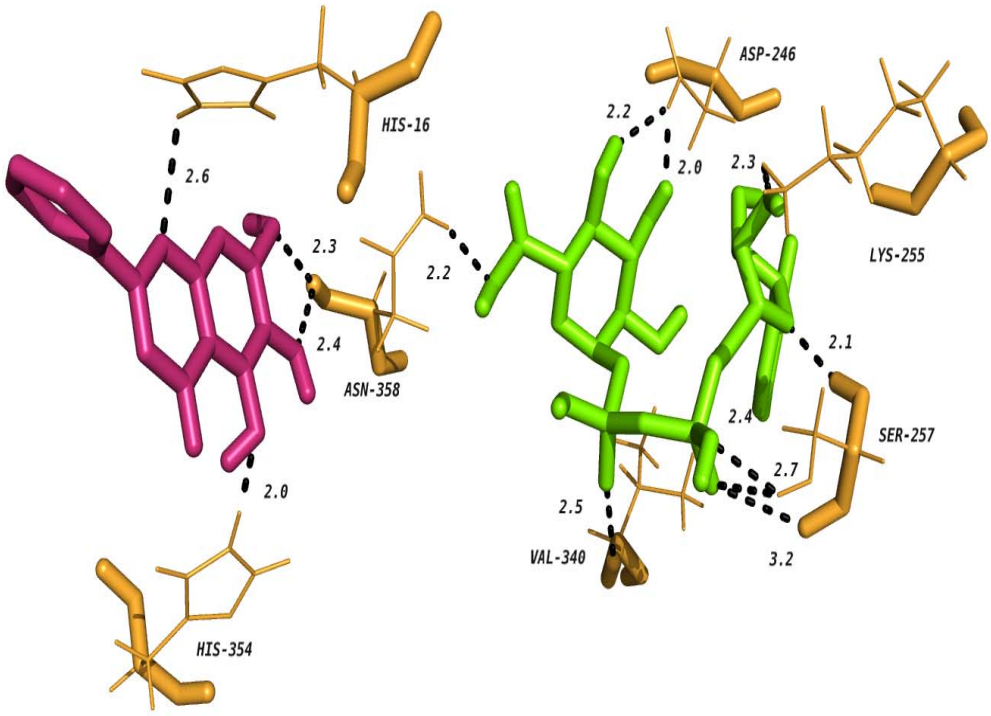


Fig. 4.16 Interaction of catalytic residues of UGT73A16 with baicalein (pink) and UDP-glucuronic acid (green). Hydrogen bonding interactions are indicated by black dashed lines

5.0 Conclusions

- ❖ Out of 50, one of the best model was chosen based on DOPE score, ERRAT plot and Z-score.
- ❖ RMSD value of modeled UGT73A16 showed that the overall conformation of the model was very similar to the template 2PQ6 and found to be 0.771 A^0.
- ❖ A homology modelling and docking study of UGT73A16 was done to validate the experimental data on substrate specificity. Homology model of UGT73A16 showed 87.1% residues in the most favourable region.
- ❖ Docking studies of modeled UGT73A16 was done with UDP-glucose and UDP-glucuronic acid as donor and kaempferol, naringenin, daidzein, genistein and baicalein as acceptor substrates
- ❖ Docking interactions of substrates and modeled UGT73A16 imply that baicalein and UDP-glucuronic acid are more specific to modeled UGT73A16.

Chapter 5

Mutational analysis of Withania somnifera glycosyltransferase (UGT73A16)

1.0 Summary

A glycosyltransferse UGT73A16 from *Withania somnifera* catalyzes the glycosylation of flavonols, isoflavonoids, isoflavone and flavonone. A 3D structure analysis based mutations were carried out to explore the roles of amino acids involved in substrate binding region. Mutations were done on acceptor and donor binding region of recombinant UGT73A16. All of the mutations were confirmed by sequencing and mutant proteins were heterologously expressed and purified as given in chapter 3A. The purified mutant (H16E, Y216H, Q379H, Q339A, W357G, A337C and S359A) proteins were incubated with all the acceptor and donor substrate as shown in chapter 3A (**Table 3A.9**). Product determination showed that mutant proteins exhibit the activity with baicalein and it also exhibited very low activity (not detectable) with kaempferol. Substitution mutant Q339A and A337C exhibited 6-7 fold and 2-3 fold more catalytic efficiency respectively with baicalein than the wild type enzyme. Other mutants H16E and Y216H showed nearly equal activity as with wild enzyme. The Q379H mutant showed large loss in enzyme activity. Our results demonstrate that Gln 339, Ala 337 and Gln379 may play an important role in enzyme activity.

2.0 Introduction

Glycosylation reaction is catalyzed by a super family of enzymes, UDP-sugar:glycosyltransferase (UGT), and is classified based on the UDP-sugar donor that is transferred, for example, glucosylation (UDP-glucose), rhamnosylation (UDP rhamnose), and glucuronosylation (UDP-glucuronic acid [UDPGA]) (**Ross et al. 2001; Lim and Bowles 2004**). Generally glucose is the most common sugar moiety in naturally occuring flavonoid glucosides. However, specific plant lineages have characteristic flavonoids conjugated with specific sugar moieties as their specialized metabolites. Flavonoid 7-*O*-glucuronide is commonly observed with flavonoid 7-*O*-glucoside in plants within the lamiales order, such as *Perilla frutescens* (dietary herb), *Antirrhinium majus* (snapdragon flower) and *Scutellaria baicalensis* (**Hirotani et al. 1998; Harborne and Baxter 1999; Yamazaki et al. 2003; Wortley et al. 2005**). Baicalin is a flavone 7-*O*-glucuronide which has anti-allergic, anti-viral, anti-inflammatory and intestinal α-glucosidase inhibitory properties (**Nishioka et al. 1998; Gao et al. 1999; Chou et al. 2003**) and produced by glucuronosylation of baicalein.

In nature baicalein is often glycosylated on one or more of the three hydroxyl groups to increase its solubility, stability and bioability. The position of conjugation has a significant impact on its biological activity and potential health benefits for humans.

The relationship between the degree of amino acid sequence identity and substrate specificity of the GTs is highly complex and this also applies to the plant UGTs. Highly diverging UGTs are often observed to share the same substrates. This may be illustrated by two *Dorotheanthus bellidiformis* UGTs, UGT73A5 and UGT71F2. These two UGTs show only 20% amino acid sequence identity (**Vogt 2002**) but glycosylate the same set of acceptor molecules **(Vogt et al. 1997; Vogt et al. 1999).** In some cases though, phylogenetically closely related UGTs also show distinct regiospecific differences towards a common acceptor. This is the case for the two *A. thaliana* UGTs, AtUGT74F1 and AtUGT74F2 that share the 82% amino acid sequence identity and glycosylate the phenolic hydroxyl group of 2-hydroxy benzoic acid. In addition, AtUGT74F2 is able to catalyze the formation of a glucose ester with this acceptor while AtUGT74F1 is not (**Lim et al. 2002; Lim 2005**). Deduction of substrate specificity not only involves specificity towards the sugar acceptor but also the UDP-sugar donor.

Different sugars may be used by UGTs within the same phylogenetically defined families and subfamilies. The prediction of substrate specificity based on Phylogenetic analyses will certainly improve as more biochemical data on the substrate preferences of UGTs within each subgroup becomes available (**Lim et al. 2003a; Vogt and Jones 2000**). However, for the time being it is evident that substrate specificity cannot be ascribed based on primary sequence analysis alone (**Modolo et al. 2007**).

Bioinformatics tools for accurate prediction of sequence features important for defining the substrate specificity of a particular UGT have not yet been developed enough. Substrate specificity of UGT73A16 cannot be predicted by phylogenetic tree analysis and amino acid sequence alone. From all of these points of view structure-based mutagenesis of UGT73A16 was carried out to explore the roles of amino acids involved in substrate binding.

3.0 Materials and methods

3.1 Site-directed mutagenesis

The details of chemicals, solvents and buffers have been given in chapter 3A (Section 3.0). Site directed mutagenesis was performed using the Quick change lightening multi site- directed mutagenesis kit (Stratagene). The UGT73A16 plasmid cloned into pET 30 b (+) vector with a hexa-histidine tag was used as template. Synthetic oligonucleotides used for mutagenesis are listed in **Table 5.0**. All of the mutation reactions are performed by manufacturer's protocol and confirmed by sequencing.

Table 5.0 Primers used for introducing amino acid changes into UGT73A16

A & D represents the mutation in flavonoid acceptor site (A) and sugar donor site (D) respectively.

Mutation	Site	Primer Sequences
H16E (CAC-GAG)	**A**	**Sense**-5'CCCATGATGGCTCAAGGT**GAG**ATGATTCCTACACTAGAC-3'
		Antisense-5'GTCTAGTGTAGGAATCAT**CTC**ACCTTGAGCCATCATGGG-3'
Y216H (CAA-CGC)	**A**	**Sense**- 5'TGGAGTTATCTTCAACAGCTT**CGC**TGAGCTTGAACCAGATTATGTG-3'
		Antisense- 5'CACATAATCTGGTTCAAGCTCA**GCG**AAGTGTTGAAGATAACTCCA-3'
Q379H (CAA-CAC)	**D**	**Sense**- 5'-GGCCAGTGTTTGCGGAA**CAC**TTTTTGAATGAAAGTTGG-3'
		Antisense-5'-CCAACTTTCATTCAAAAA**GTG**TTCCGCAAACACTGGCC-3'
Q339A (CAA-GCA)	**D**	**Sense**-5'-CATAAGAGGATGG**GCA**CCCGCAGTGCTGATTCTTGATCAC-3'
		Antisense-5'GTGATCAAGAATCAGCACTGCGGG**TGC**CCATCCTCTTATG-3'
W357G (TGG-GGG)	**D**	**Sense**-5'GCTTTTGTGACTCATTGTGGA**GGG**AATTCAACGCTAGA-3'
		Antisense-5'TCTAGCGTTGAATT**CCC**TCCACAATGAGTCACAAAAGC-3
A337C (GCA-TGC)	**D**	**Sense**-5'-AGGATTAATCATAAGAGGATGG**TGC**CCCCAAGTGCTGATTCTTGATC-3'
		Antisense-5'GATCAAGAATCAGCACTTGGGG**GCA**CCATCCTCTTATGATTAATCCT-3'
S359A (TCA-GCA)	**D**	**Sense**- 5'-GTGACTCATTGTGGATGGAAT**GCA**ACGCTAGAAGGAAT-3'
		Antisense- 5'-ATTCCTTCTAG**CGT**TGCATTCCATCCACAATGAGTCAC-3'

3.2 Expression and Purification of recombinant wild type (UGT73A16) and mutant enzymes

E. coli BL21 (DE3) cells harboring the wild type (UGT73A16) and mutated constructs were cultured in LB media at 37 ºC until A_{600} reached 0.6-0.8. For induction IPTG was added to a final concentration of 0.08 mM and cultures were incubated at 22 ºC for 20 h.

Wild type and mutated proteins were purified as given in previous chapter 3A (Section 4.1.5). Protein concentration was determined with Bradford reagent using bovine serum albumin as standard.

3.3 Kinetic study of mutants with wild type UGT73A16

The K_m and V_{max} values of UGT73A16 for baicalein was determined at pH 8.0 and 37°C by fitting the initial velocity data of Michaelis-Menten equation using nonlinear regression analysis. Substrates were used at a saturating concentration with constant enzyme concentration. The baicalein reaction contained 40 µg of purified UGT73A16, 100 mM Tris-Cl (pH 8.0), 400 µM of UDP- glucuronic acid and 10 µM -100 µM of baicalein. Reaction was kept at 37 ºC for 3 h and the extraction was performed twice by equal volume of ethyl acetate. Peak area of HPLC chromatogram was calculated to determine the standard graph (Baicalein 7-*O*-glucuronide) and Michaelis-Menten plot of wild type and mutants. Concentration of substrates and standards was considered on X-axis and peak area on Y-axis and graph was plotted using Graph Pad Prism 5.0 software.

4.0 Results and discussion

4.1 Kinetic parameter of wild type and mutants with baicalein

To determine the initial velocity of UGT73A16, the assays were performed under steady state conditions using the standard assay system with various substrate concentrations. The known concentration of baicalein 7-*O*-glucuronide (**Fig. 5.1 A**) was used to plot a standard graph. The apparent K_m and V_{max} values for baicalein with all mutants (**Table 5.1 & 5.2**) were determined by fitting the initial velocity data to the Michaelis-Menten equation using nonlinear regression analysis.

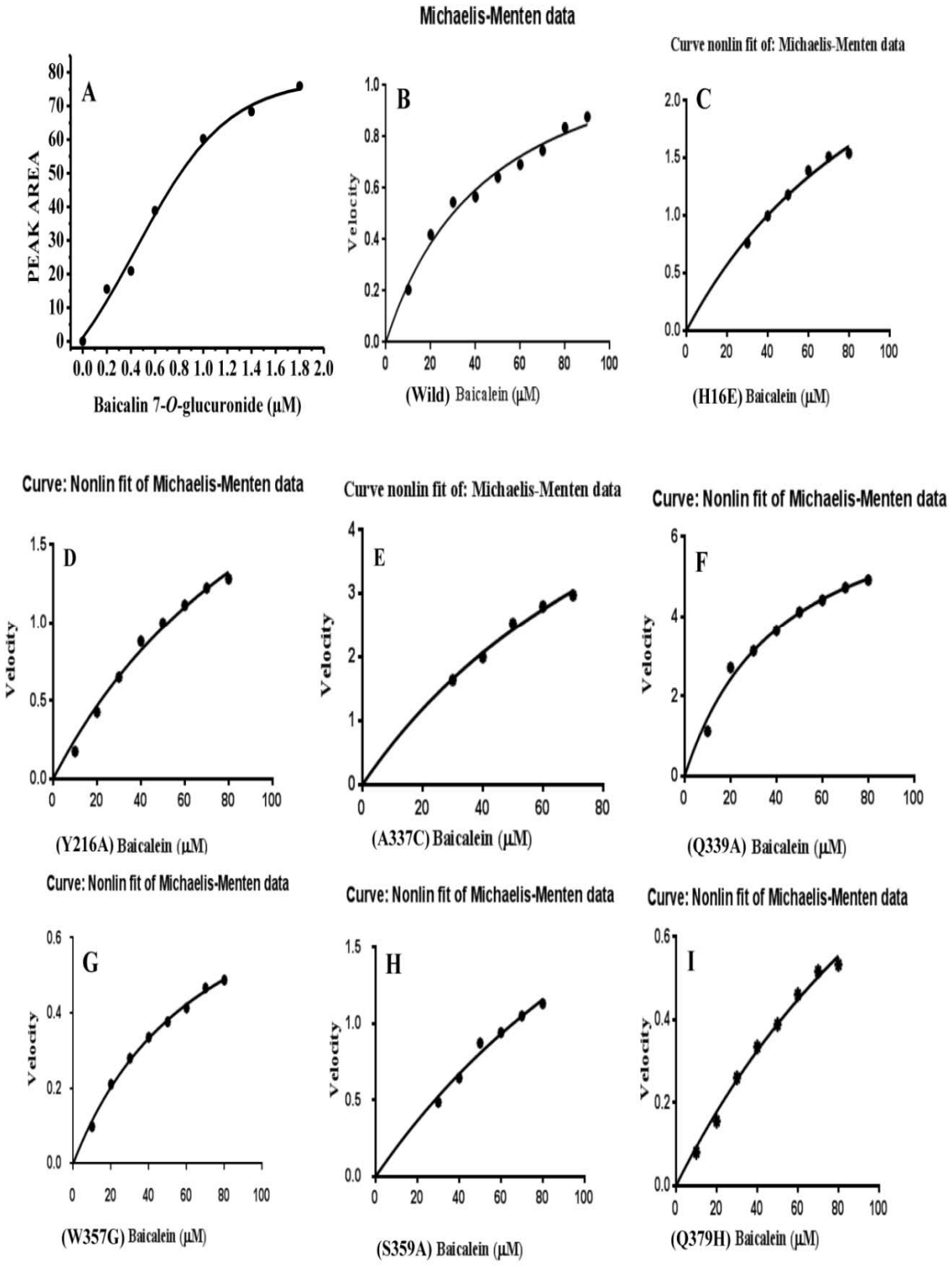


Fig. 5.1 Michaelis-Menten plot wild type UGT73A16 and mutants **(A)** Baicalein 7-*O*-glucuronide standard plot **(B)** Wild UGT73A16 (C) H16E, (D) Y216A, (E) A337C, (F) Q339A (G) W357G, (H) S359A, (I) Q379H.

4.1 Effects of mutation of acceptor binding pocket residues on enzyme activity

Under *in vitro* condition wild type UGT73A16 catalyzes the formation of baicalein 7-*O*-glucuronide (Baicalin) using baicalein as acceptor and UDP-glucuronic acid as donor. To investigate the structural basis for the substrate specificity and product regioselectivity of UGT73A16, each of amino acid side chains of the residues around the acceptor binding pockets were independently altered by site-directed mutagenesis (**Table 5.1**).

SDS-PAGE analysis confirmed that a single band of correct molecular weight was seen in the purified preparations of each mutant (H16E and Y216H). Mutant H16E showed 21% of more activity than the wild type UGT73A16. On the other hand Y216H retained approximately 98% of wild type activity. Kinetic analysis indicated that H16E and Y216H exhibited approximately 1.2 fold increased and equal catalytic efficiency respectively as compared with wild type UGT73A16 (**Table 5.1**).

Table 5.1 Kinetic parameter for wild type and mutant (Acceptor binding site) enzymes with baicalein as acceptor and UDP-glucuronic acid as donor

Substrate	K_m **(µM)**	V_{max} **(nkat/mg)**	V_{max}/K_m	K_{cat} **(sec-1)**	K_{cat}/K_m **($\mu M^{-1}/sec^{-1}$)**
Baicalein(Wild)	47.08	21.50	0.45	1.11	0.023
H16E	119.4	66.46	0.55	3.45	0.028
Y216H	125.7	56.69	0.45	2.94	0.023

4.2 Effects of mutation of donor binding pocket residues on enzyme activity

To investigate the roles of amino acid residues around the sugar binding pocket in sugar donor specificity and enzyme activity with baicalein as acceptor and UDP-glucuronic acid as donor, site directed mutagenesis was used to alter the side chains of those residue with close contacts to the sugar moiety (**Table 5.2**).

Table 5.2 Kinetic parameter for wild type and mutant (Donor binding region) enzymes with baicalein as acceptor and UDP-glucuronic acid as donor

Substrate	K_m **(µM)**	V_{max} **(nkat/mg)**	V_{max}/K_m	K_{cat} **(sec-1)**	K_{cat}/K_m **($\mu M^{-1}/sec^{-1}$)**
Baicalein(Wild)	47.08	21.50	0.45	1.11	0.023
A337C	116.8	135.09	1.15	7.022	0.060
Q339A	42.39	126.0	2.97	6.55	0.156
W357G	71.43	15.44	0.216	.802	0.011
S359A	208.1	69.33	0.33	3.60	0.017
Q379H	188.8	30.95	0.17	1.60	0.008

As shown in **Table 5.2** mutant Q339A exhibited 6-7 fold more activity than the wild type enzyme with UDP-glucuronic acid. A337C also exhibited 2-3 fold more activity than the wild type but less than Q339A. In case of W357G and Q379H the large loss of activity was observed as compare to wild type. It means W357G and Q379H may play an important role in enzyme activity.

5.0 Conclusions

- SDS-PAGE analysis confirmed that a single band of correct molecular weight was seen in the purified preparations of each mutant.
- Mutation at acceptor binding region (H16E and Y216H) showed approximate equal activity with wild type UGT73A16.
- Mutant Q339A and A337C exhibited 6-7 fold and 2-3 fold more activity than wild type UGT73A16.
- Mutation at donor binding region (W357G and Q379H) exhibit large loss of activity.

Summary of thesis

Summary of thesis

Most of the flavonoids found in plants exist as glycosides; these secondary metabolites undergo several modification reactions, including glycosylation. Glycosylation of flavonoids is mediated by family-1 UDP-glycosyltransferases (UGT) and has a wide range of effects on flavonoids like solubility, stability and bioavailability. Glycosylation use UDP- sugars, such as UDP- glucose as the glucosyl donor and polyphenols, terpenoids, alkaloids and cyanohydrins as acceptors. Till date, there are no reports on functionally characterized flavonoid glycosyltransferases from *Withania somnifera.*

A PCR based approach was used to isolate the full length of glycosyltransferase gene from *W. somnifera*. Primers were designed on the basis of sequence information available in NCBI GenBank database. PCR was done with different set of primers, using cDNA as a template. Initially, a partial fragment was cloned into pGEMT-Easy vector and sequenced. Sequence analysis was done with BLAST (NCBI) programme, which were showing significant similarity with reported glycosyltransferase genes from other plants. Gene specific primers were designed and Rapid Amplification of cDNA Ends (RACE) PCR was performed to isolate full-length cNDA clones. According to obtained sequence information from 3′ RACE and 5′ RACE, WSfullF and WSfullR was designed to obtain open reading frame. The open reading frame of 1413 bp encoded 470 amino acid residues with a calculated molecular mass of ~ 52 kDa. In addition, a 13 bp 5′ untranslated region and a 101 bp 3′ untranslated region exist in the isolated gene. The full length sequence reported in this thesis has been deposited in the NCBI GenBank database [Accession number FJ654696]. According to the glycosyltransferase nomenclature guidelines, the systematic name of the *W. somnifera* glycosyltransferase is UGT73A16. The translated protein sequence of UGT73A16 was used to construct a phylogenetic tree using bootstrapping (MEGA 4.0) with the known plant glycosyltransferases (GTs) deposited in NCBI Gen Bank database using Neighbor-joining method. Phylogenetic analysis suggests that the UGT73A16 was most similar to glycosyltransferase capable of flavonoid 7-*O*-glycosyltransferase, salicylic acid induced glycosyltransferase (Accession number: AAK28303) and UDP glucose: flavonoid 3-*O*-glucoside 7-*O*-glucosyltransferase (Accession number: Q9ZQ95).

The open reading frame of UGT73A16 was cloned into pET 30b (+) expression vector and transformed into *E. coli* BL21 (DE3) strain. The over expressed recombinant UGT73A16 was purified by affinity chromatography using Ni-NTA agarose beads in denatured conditions and analyzed on SDS-PAGE. The resulting molecular weight of the purified recombinant UGT73A16 was approximately ~ 52 kDa.

All the conditions were optimized (Induction 0.08 mM IPTG, Temperature 22 ^{o}C and Time 20 h) to express the recombinant UGT73A16 in soluble form. The recombinant protein was eluted from lysate using affinity chromatography and further used for enzyme assay. The enzyme assay was done using UDP-glucose and UDP-glucuronic acid as a donor substrate and flavonoids (Naringenin, kaempferol, isorhamnetin, 3 hydroxy flavones, daidzein, genistein, biochaninA and baicalein) as acceptor substrate. Product identification was done by hypsochromic shift, HPLC and LC-MS methods. HPLC analysis and hypsochromic shift indicated that UGT73A16 transfers a glucose molecule to several different flavonoids. Analysis of reaction products also show that glycosylation of substrates occurred on the 7- and 3- hydroxyl group of the aglycon. Recombinant UGT73A16 also displayed regiospecificity glucosyl transfer activity towards 3- hydroxy flavone (chemically synthesized), which is the backbone of all flavonols, not found naturally. Kinetic parameters of UGT73A16 show more catalytic efficiency towards naringenin followed by genistein and baicalein. The present study generates essential knowledge and molecular as well as biochemical tools that allows the verification of UGT73A16 in glycosylation.

Plant secondary glucosides are very difficult to isolate or to chemically synthesize in pure form because of low yield. Thus the simple approaches for enhanced production of such glucosides would be highly beneficial. In this regards we explored UGT73A16 of *W. somnifera* as whole cell catalyst in *E. coli*. We used flavonoids (genistein, apigenin, kaempferol, naringenin, biochaninA and daidzein) and endogenous UDP-glucose as substrates for this study because flavonoids and their glucosides have important medicinal properties and they provide a complex structure for regioselective glycosylation. Approximate 21 mg/L to 22 mg/L of either apigenin 7-*O*-glucoside or naringenin 7-*O*-glucoside was produced after feeding apigenin or naringenin to *E. coli* expressing UGT73A16. More than 95% of the glucosides products were released into the medium, facilitating their isolation. Due to its broad range of substrate and low regioselectivity, UGT73A16 may be an attractive enzyme for engineering flavonoid diversity. It will also be interesting to determine if UGT73A16 can glycosylate other structurally related and chemically synthesized compounds, because it glycosylated the 3-hydroxy flavone, which is a chemically synthesized compounds not found naturally.

The enzyme assay analysis suggested that the glycosyltransferase UGT73A16 from *W. somnifera* catalyzes glycosylation of the (iso) flavonoids, flavonols, flavonone and flavone. It transfer sugar molecule to hydroxyl group of (iso) flavonoids, flavonols, flavonone and flavone with 3-hydroxyl and 7-hydroxyl group as the major product.

It is well known from previous study that the amino acid residues of glycosyltransferases are responsible for sugar donor and acceptor binding sites, which are located in the N and C terminal respectively as determined by modeling and docking studies.

The secondary structure prediction run at the PSIPRED server showed that the UGT73A16 consists of 40% α-helix (17 helices; 189 residues), 10% extended-beta (11 strands; 47 residues) and 50% random coil (28 coils; 234 residues) configuration. UGT73A16 amino acid sequence alignment with available protein sequences showed a significant percentage of identity with *Medicago truncatula* flavonoid 3-*O*-glucosyltransferase (3HBJ A: 34%), *Medicago truncatula* UDP-glucuronosyl/UDP-glucosyltransferase (2PQ6 A: 32% identity) *Arabidopsis thaliana* hydroquinone glucosyltransferase (2VG8: 31%). These proteins were selected as templates for the modeling. A three dimensional model of UGT73A16 was constructed by MODELLER 9v9 software using 2PQ6, 2VG8 and 3HBJ as template. The modeled UGT73A16 structure was analysed by various tools such as DOPE score, Ramachandran plot, ERRAT analysis and ProSAII. The energy minimization of modeled UGT73A16 was done by Chimera 1.6.2 software. The RMSD value of alpha carbon (Cα) of the modeled UGT73A16 was calculated by superimposition with templates 2PQ6, 2VG8 and 3HBJ on chimera and it was found to be 0.771 Å, 0.958 Å and 1.221 Å respectively. The resulted model was assessed by various tools and the final refined model revealed GT-B type fold.

The structure data files (SDF) of the acceptor and donor substrates were obtained from Pub Chem Database and converted to PDB format using PyMOL software. The UGT73A16 structure and substrate ligands were prepared using AutoDock Tools v.1.5.2 software. Molecular docking was performed using Autodock vina 1.1.2 software. The interactions of the ligands with UGT73A16 were visualized and figures were formed using PyMOL 1.3. The lowest binding energy conformation in the first cluster was considered as the most favorable docking pose. Modeled UGT73A16 was docked with UDP-glucose, UDP-glucuronic acid, kaempferol, naringenin, daidzein, genistein and baicalein. The protein-ligand interactions showed that His 16, Asp 22, Ile 50, Asp 246, Lys 255, Ser 283, Trp 336, Ala 337, Gln 339, Val 340, Trp 357, Asn 358 and Glu 362 amino acid residues may be important for catalytic function.

A three dimensional structure analysis based mutations were carried out to explore the roles of amino acids involved in substrate binding region. Mutations were done on acceptor and donor binding region of recombinant UGT73A16.

All of the mutations were confirmed by sequencing. All of the mutants were heterologously expressed and purified using Ni-NTA. The purified mutants (H16E, Y216H, Q379H, Q339A, W357G, A337C and S359A) were incubated with all of the acceptor (Kaempferol, naringenin, genistein, daidzein and baicalein) and donor substrate (UDP-glucose and UDP-glucuronic acid).

Product determination showed that mutants exhibit the activity with baicalein and it also exhibited very low activity (not detectable) with kaempferol. Substitution of The Q339A and A337C mutant exhibited 6-7 fold and 2-3 fold more catalytic efficiency respectively with baicalein than the wild type enzyme. Other mutants H16E and Y216H showed almost equal activity as with wild type. The Q379H mutant showed large loss in enzyme activity. Our results point out that Gln 339, Ala 337 and Gln379 may be key residues and play an important role in enzyme activity.

References

References

Abou-Douh AM (2002) New withanolides and other constituents from the fruit of Withania somnifera. Arch Pharm 335:267-276

Agarwal R, Diwanay S, ,, Patki P, Patwardhan B (1999) Studies on immunomodulatory activity of Withania somnifera (ashwagandha) extracts in experimental immune inflammation. J Ethnopharmacol 67 (1):27–35

Alfonso D, Kapetanidis I (1994) Withanolides from Iochroma gesnerioides. Phytochemistry 36:179-183

Andallu B, Radhika B (2000) Hypoglycemic, diuretic and hypocholesterolemic effect of winter cherry (Withania somnifera) root. . Indian J Exp Biol 38:607-609

Angalagan K, Sadique J (1981) Influence of an Indian medicine (ashwagandha) on acute-phase reactants in inflammation. . Indian J Exp Biol 19 245-249

Archana R, Namasivayam A (1999) Antistressor effect of Withania somnifera. J Ethnopharmacol 64:91-93

Arnheim N, Erlich H (1992) Polymerase Chain Reaction Strategy. Annual Review of Biochemistry 61:131-156

Arun KM, Kaul K, Bhan, M. K., , Punit K, Khanna K, Suri A (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

Atta-ur-Rahman., Abbas S, Dur-e-Shawar., Jamal AS, Choudhary MI (1993) New withanolides from Withania spp. . J Nat Prod 56:1000-1006

Atta-ur-Rahman., Jamal AS, Choudary MI, Asif, I., (1991) Two withanolides from Withania somnifera. Phytochemistry 30:3824-3825

Bandyopadhyay M, Jha S, Tepfer D (2007) Changes in morphological phenotypes and withanolide composition of Ri-transformed roots of Withania somnifera. Plant cell Rep 26 (599-609)

Begum VH, Sadique J (1988) Long-term effect of herbal drug Withania somnifera on adjuvant-induced arthritis in rats. . Indian J Exp Biol 26:877-882

Bennet RN, Wallsgrove RM (1994) Secondary metabolites in plant defense mechanisms. New Phytologist 127:617–633

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindiylov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28 (1):235-242

Bhaskara R, Madina, Lokendra k, Sharma, Pankaj C, Rajender s, sangwan, Rakesh T (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 SK, Muruganandam AV (2003) Adaptogenic activity of withania somnifera: An experimental study using rat model of chronic disease. . Pharmacol Biochem Behav 75 (3):547-555.

Blumberg DD (1987) Creating a ribonuclease-free environment. Meth Enzymol 152:20-24

Bone K (1996) Clinical Applications of Ayurvedic and Chinese Herbs. Monographs for the Western Herbal Practitioner. . Australia: Phytotherapy Press;137-141.

Bouktaib M, Atmani A, Rolando C (2002) Regio- and stereoselective synthesis of the major metabolite of quercetin, quercetin-3-*O*-β-D-glucuronide. Tetrahedron Lett 43:6263-6266

Bourne Y, Henrissat B (2001) Glycoside hydrolases and glycosyltransferases families and functional modules. . Curr Opin Struct Bio 11:593-600

Bowles D, Isayenkova J, Lim EK, Poppenberger B (2005) Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 8 (3):254-263. doi:S1369-5266(05)00039-7 [pii] 10.1016/j.pbi.2005.03.007

Bowles D, Lim EK, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. Annu Rev Plant Biol 57:567-597. doi:10.1146/annurev.arplant.57.032905.105429

Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ, Edwards R (2007) Characterization and engineering of the bifunctional N- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. Proc Natl Acad Sci U S A 104 (51):20238-20243. doi:0706421104 [pii] 10.1073/pnas.0706421104

Brugliera F, Holton TA, Stevenson TW, Farcy E, Lu CY, Cornish EC (1994) Isolation and characterization of a cDNA clone corresponding to the Rt locus of Petunia hybrida. Plant Journal 5 (81-92)

Budhiraja RD, Sudhir S (1987) Review of biological activity of withanolides. J Sci Ind Res 46:488-491

Bulter T, Elling L (1999) Enzymatic synthesis of nucleotide sugar. Glycoconj J 16:47–159

Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM (2004) Crystal structure of glycogen synthase:Homologous enzymes catalyze glycogen synthesis and degradation. EMBO J 23:3196–3205

Campbell JA, Davies GJ, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J 326 (929-939)

Cheng GW, Malencik DA, Breen PJ (1994) UDP-glucose:flavonoid O-glucosyltransferase from strawberry fruit. Phytochemistry 1435 (1435-1439)

Chou TC, Chang LP, Li CY, Wong CS, Yang SP (2003) The anti-inflammatory and analgesic effects of baicalin in carrageenan-evoked thermal hyperalgesia. Anesth Analg 97:1724–1729

Claire MM, Gachon MLM, Patrick S (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. . TRENDS in plant science 10, No.11 November

Collin HA (2001) Secondary product formation in plant tissue cultures. Plant Growth Regul 34:119–134

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 9:1511-1519

Cornwell T, Cohick W, Raskin I (2004) Dietary phytoestrogens and health. Phytochemistry 65 (8):995-1016. doi:10.1016/j.phytochem.2004.03.005 S0031942204001049 [pii]

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classication for glycosyltransferases. . J Mol Biol 328:307-317

Cragg GM, Newman DJ (2005) Biodiversity: A continuing source of novel drug leads. Pure Appl Chem 77 (1):7-24

Davis L, Kuttan G (2000) Immunomodulatory activity of Withania somnifera. . J Ethnopharmacol 71:193-200

Davis L, Kuttan G (2002) Effect of Withania somnifera on CTL activity. . J Exp Clin Cancer Res 21:115-118

DeLano WL (2002) The PyMOL Molecular Graphics System. J Mol Graph Model 1:57-61

Dhuley JN (2000) Adaptogenic and cardioprotective action of ashwagandha in rats and frogs. J Ethnopharmacol 70 (1):57–63

Dodson E, Harding MM, Hodgkin DC, Rossmann MG (1966) The crystal structure of insulin.3. Evidence for a 2-fold axis in rhombohedral zinc insulin. . J Mol Biol 16:227-241

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791

Ford CM, Boss PK, Hoj PB (1998) Cloning and characterization of Vitis vinifera UDP-glucose: flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize bronze-1 locus that may primarily serve to glycosylate anthocyanidins in vivo. J Biol Chem 273:9224–9233

Fritz TA, Hurley JH, Trinh LB, Shiloach J, Tabak LA (2004) The beginnings of mucin biosynthesis: The crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. Proc Natl Acad Sci USA 101:15307–15312

Gao Z, Huang K, Yang X, Xu H (1999) Free radical scavenging and antioxidant activities of flavonoids extracted from the radix of Scutellaria baicalensis Georgi Biochim Biophys Acta 1472:643–650

Given NK, Venis MA, Grierson D (1988) Phenylalanine ammonia-lyase activity and anthocyanin synthesis in ripening strawberry fruit. Journal of Plant Physiology 133 (25–30)

Glotter E (1991) Withanolides and related ergostane-type steroids. . Nat Prod Rep 8:415-440

Halbwirth H, Puhl I, Haas U, Jezik K, Treutter D, Stich K (2006) Two-phase flavonoid formation in developing strawberry (Fragaria×ananassa) fruit. Journal of Agricultural Food Chemistry 54:1479–1485

Harborne JB, Baxter H (1999) The handook of natural flavonoids. Wiley, New Yark

Heidlas JE, Lees WJ, Whitesides GM (1992) Practical enzyme-based syntheses of uridine 5'-diphosphogalactose and uridine 5'- diphospho-N-acetylgalactosamine on a gram scale. J Org Chem 57:152–157

Heller W, Forkmann G (1994) Biosynthesis of flavonoids. In The Flavonoids Advances in Research since 1986:499-535

Hirotani M, Nagashima S, Yoshikawa T (1998) Baicalin and baicalein productions of cultured Scutellaria baicalensis cells. Nat Med 52:440–443

Hirotani., Kuroda M, Suzuki RH, Yoshikawa T (2000) Cloning and expression of UDP-glucse: flavonoid 7-Oglucosyltransferase from hairy root cultures of Scutellaria baicalensis. Planta 210:1006-1013

Hughes J, Hughes MA (1994) Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. DNA Seq 5:41–49

Ichikawa H, Takada Y, Shishodia S, Jayaprakasam B, Nair MG, Aggarwal BB (2006) Withanolides potentiate apoptosis, inhibit invasion, and abolish osteoclastogenesis through suppression of nuclear factor-kappaB (NF-kappaB) activation and NF-kappaB-regulated gene expression. Molecular Cancer Therapeutics 5 (6):1434-1445

Jayaprakasam B, Strasburg GA, Nair MG (2004) Potent lipid peroxidation inhibitors from Withania somnifera fruits. Tetrahedron 60: 3109-3121

Jayaprakasam B, Zhang Y, Seeram N, Nair M (2003) Growth inhibition of tumor cell lines by withanolides from Withania somnifera leaves. . Life Sci 74:125-132

Johri S, Jamwal U, Rasool S, Kumar A, Verma V, Qazi GN (2005) Purification and characterization of peroxidases from Withania somnifera (AGB 002) and their ability to oxidize IAA. . Plant Sci 169:1014-1021

Jones P, Messner B, Nakajima JI, Schäffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. Journal of Biological Chemistry 278:43910–43918

Jones P, Vogt T (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213 (2):164-174

Kala CP, Dhyani PP, Sajwan BS (2006) Developing the medicinal plants sector in northern India: challenges and opportunities. J Ethnobiol EthnoM 2:32

Kaul KN (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 7-8

Kikuchi N, Kwon YD, Gotoh M, Narimatsu H (2003) Comparison of glycosyltransferases families using the profile hidden Markov model. . Biochem Biophys Res Commun 310:574-579

Kim JH, Kim BG, Ko JH, Lee Y, Hur HG, Lim Y, Ahn JH (2006) Molecular cloning, expression, and characterization of a flavonoid glycosyltransferase from Arabidopsis thaliana. PLANT SCI 170 (4):897-903. doi:DOI 10.1016/j.plantsci.2005.12.013

Kirsi M, Oksman C, Dirk I (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. TRENDS Plant Sci 9 (9):433-440

Kramer CM, Prata RT, Willits MG, Luca DV, Steffens JC, Graser G (2003) Cloning and regiospecificity studies of two flavonoid glucosyltransferases from Allium cepa. Phytochemistry 64 (6):1069-1076. doi:S0031942203005077 [pii]

Kretzschmar G, Stahl W (1988) Large-scale synthesis of linkermodified sialyl Lewis X, Lewis X and N-acetyllactosamine. Tetrahedron Lett 54:6341–6358

Kroon J, Souer E, de Graaf A, Xue Y, Mol J, Koes R (1994) Cloning and structural analysis of the anthocyanin pigmentation locus Rt of Petunia hybrida: characterization of insertion sequences in two mutant alleles. Plant J 5 (69–80)

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 429:198–203

Kumar S, Shukla YN, Lavania UC, Sharma A, Singh AK (1997) Medicinal and Aromatic Plants: Prospects for India. J Med Arom Pl Sc 19 (2):361-365

Kuttan G (1996) Use of Withania somnifera Dunal as an adjuvant during radiation therapy. . Indian J Exp Biol 34:854-856

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type1 model studies. Nucleic Acids Research 18:999-1005

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157:105-132

L. Li., Mondolo LV, Escamilla-Trevino LL, Achinine L, Dixon RA, Wang X (2007) Crystal structure of *Medicago Trunctula* UGT85H2--insights into the structural basis of multifunctional (iso)flavonoid glycosyltransferase. J Mol Biol 370 (5):951-963

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685

Lakshmi C, Mishra.,, Betsy BS, Simon D (2000) Scientific Basis for the Therapeutic Use of Withania somnifera (Ashwagandha): A Review. Alternative Medicine Review, 5

Laskowski RA, Luscombe NM, Swindells MB, Thornton JM (1996) Protein clefts in molecular recognition and function. Prot Sci 5:2438-2452

Li M, Han X, Yu B (2002) Synthesis of quercetin 3-O-(2”-galloyl)-α-L-arabinopyranoside. Tetrahedron Lett 43:9467–9470

Li Y, Baldauf S, Lim EK, Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J Biol Chem 276 (6):4338-4343. doi:10.1074/jbc.M007447200 M007447200 [pii]

Lim E-K, Bowles DJ (2004) A class of plant glycosyltransferases involved in cellular homeostasis. EMBO J 23 (15):2915-2922

Lim EK (2005) Plant glycosyltransferases: their potential as novel biocatalysts. Chemistry 11:5486–5494

Lim EK, Ashford DA, Hou BK, Jackson RG, Bowles DJ (2004) Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnol Bioeng 87 (5):623-631. doi:Doi 10.1002/Bit.20154

Lim EK, Baldauf S, Li Y, Elias L, Worrall D, Spencer SP, Jackson RG, Taguchi G, Ross J, Bowles DJ (2003a) Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis. Glycobiology 13:139–145

Lim EK, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J, Bowles DJ (2002) The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. J Biol Chem 277:586–592

Liu J, Mushegian A (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. . Protein Sci 12 (1418-1431)

Lobsanov YD, Romero PA, Sleno B, Yu B, Yip P, Herscovics A, Howell PL (2004) Structure of Kre2p/Mnt1p: A yeast alpha1,2-mannosyltransferase involved in mannoprotein biosynthesis. J Biol Chem 279:17921–17931

Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW (2005) Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. Pharmacogenet Genome 15 (10):677-685

Mato M, Ozeki Y, Itoh Y, Higeta D, Yoshitama K, Teramoto S, Aida R, Ishikura N, Shibata M (1998) Isolation and Characterization of a cDNA Clone of UDP-Galactose: Flavonoid 3-*0*-Galactosyltransferase (UF3GaT) Expressed in *Vigna mungo* Seedlings. Plant Cell Physiol 39 (11):1145-1155

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404-405

Middleton EJ, Kandaswami C, Theoharides TC (2000) The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. Pharmacol Rev 52 (4):673-751

Miller KD, Jeremy GV, Evans NS, Wendy A, Shuttleworth, Loverine PT (1999) Purification, Cloning, and Heterologous Expression of a Catalytically Efficient Flavonol 3-*O*-Galactosyltransferase Expressed in the Male Gametophyte of Petunia hybrida. The Journal of Biological Chemistry 274:34011-34019

Modolo LV, Blount JW, Achnine L, Naoumkina MA, Wang X, Dixon RA (2007) A functional genomics approach to (Iso)flavonoid glycosylation in the model legume Medicago truncatula. Plant Mol Biol 64: 499–518

Mondolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. J Mol Biol 395 (5):1292-1302

Mulichak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT, Garavito RM (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. . Proc Natl Acad Sci USA 100:9238-9243

Mulichak AM, Lu W, Losey HC, Walsh CT, Garavito RM (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands Biochemistry 43:5170–5180

Mullis KB (1990) The unusual origin of the polymerase chain reaction. Scientific American 262:56-61

Mullis KB, Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase catalysed reaction. Methods in Enzymology 255:335-350

Nishioka T, Kawabata J, Aoyama Y (1998) Baicalein, an α-glucosidase inhibitor from Scutellaria baicalensis. J Nat Prod 61:1413–1415

Noguchi A, Horikawa M, Fukui Y, Masako FM, Asako LO, Masaji I, Yoshinobu K, Toru N, Eiichiro o (2009) Local Differentiation of sugar donor speciricity of Flavonoid Glycosyltransferase in Lamiales. The Plant Cell (21):1556-1572

Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T (2007) A UDP-glucose:isoflavone 7-*O*-glucosyltransferase from the roots of soybean (glycine max) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. J Biol Chem 282 (32):23581-23590. doi:M702651200 [pii] 10.1074/jbc.M702651200

Offen W MF, Yang M, Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J 25:1396–1405

Panda S, Kar A (1997) Evidence for free radical scavenging activity of ashwagandha root powder in mice Indian J Physiol Pharmacol 41:424-426

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. Trends Plant Sci 5:439- 445

Prakash J, Gupta SK, Dinda AK (2002) Withania somnifera root extract prevents DMBA-induced squamous cell carcinoma of skin in Swiss albino mice. Nutr Cancer 42:91-97

Prakash J, Gupta SK, Kochupillai V, ,, Gupta YK, Joshi S (2001) Chemopreventive activity of Withania somnifera in experimentally induced fibrosarcoma tumors in Swiss albino mice. Phytother Res 15 (3):240-244

Rice-Evans CA, Miller NJ, Paganga G (1996) Structure-antioxidant activity relationships of flavonoids and phenolic acids. Free Radic Biol Med 20 (7):933-956. doi:0891584995022279 [pii]

Rosen ML, 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

Ross J, Li Y, Lim EK, Bowles DJ (2001) Higher plant glycosyltransferases. Genome Biol 2:3004.3001–3004.3006

Saiki R, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, 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 RK, Gelfand DH, Stoffel S, Scharf S, Higuchi R, Horn GT, Mullis KB, E. HA (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

Sali and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234 (3):779-815

Sambrook J, Fritschi EF, Maniatis T (1989) Molecular cloning: Laboratory Manual. Cold Spring Harbor Laboratory Press, New York

Sanner MF (1999) Python: a programming language for software integration and development. J Mol Graph Model 1:57-61

Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 233:1076-1078

Schwab W (2003) Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62 (6):837-849. doi:S0031942202007239 [pii]

Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. Plant Cell 17 (11):3141-3154. doi:tpc.105.035055 [pii]

Sharad AC, Solomon FE, Devi PU (1996) Antitumor and radiosensitizing effects of withaferin A on mouse Ehrlich ascites carcinoma in vivo. . Acta Oncol 35:95-100

Singh B, Chandan BK, Gupta DK (2003) Adaptogenic activity of a novel Withanolide free aqueous fraction from the roots of Withania sominifera Dun (part ii). PhytotherReasearch 17 (5):531-536

Singh B, Saxena AK, Chandan BK (2001) Adaptogenic activity of a novel, withanolide-free aqueous fraction from the root of Withania somnifera. Phytother Res 15:311-318

Sinnott ML (1990) Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 90:1171–1202

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17 (4):355-362

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2:455-461

Udayakumar R, Kasthurirengan S, Mariashibu TS, Rajesh M, Ramesh AV, Kim SC, Ganapathi A, Choi CW (2009) Hypoglycaemic and hypolipidaemic effects of Withania somnifera root and leaf extracts on alloxan-induced diabetic rats. Int J Mol Sci 10:2367-2382

Unligil UM, Rini JM (2000) Glycosyltransferase structure and mechanism. . Curr OpinStruct Bio 10:510-517

Usha S, Johnson IM, Malathi R (2005) Interaction of resveratol and genistein with nucleic acids. J Biochem Mol Biol 38:198-205

Vogt T (2002) Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-*O*-glucosyltransferase from *Dorotheanthus bellidiformis*. Planta 214:492–495

Vogt T, Grimm R, Strack D (1999) Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the Solanaceae. Plant J 19 (5):509-519

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci 5 (9):380-386. doi:S1360-1385(00)01720-9 [pii]

Vogt T, Zimmermann E, Grimm R, Meyer M, Strack D (1997) Are the characteristics of betanidin glucosyltransferases from cell-suspension cultures of *Dorotheanthus bellidiformis* indicative of their phylogenetic relationship with flavonoid glucosyltransferases? Planta 203:349–361

Wagner H, Norr H, Winterhoff H (1994) Plant adaptogens. Phytomedicine 1:63-76

Willits MG, Giovanni M, Prata RT, Kramer CM, Luca DV, Steffens JC, Graser G (2004) Bio-fermentation of modified flavonoids: an example of in vivo diversification of secondary metabolites. Phytochemistry 65 (1):31-41. doi:S0031942203006289 [pii]

Wink M (1999) Biochemistry of plant secondary metabolism. Annual plant reviews 2:358

Winkel-Shirley. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology and Biotechnology. Plant Physiology 126 (2):485-493

WinkM (1999) Biochemistry of plant secondary metabolism. Annual plant reviews 2:358

Wong CH, Wang R, Ichikawa Y (1992) Regeneration of sugar nucleotide for enzymatic oligosaccharide synthesis: use of Gal-1- phosphate uridyltransferase in the regeneration of UDP-galactose, UDP-2-deoxygalactose, and UDP-galactosamine. J Org Chem 57:4343–4344

Wortley AH, Rudall PJ, Harris DJ, Scotland RW (2005) How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. Syst Biol 54:697–709

Xian ZH, Wen SL, Jack WB, Richard AD (2008) Regioselective synthesis of plant (iso)flavone glycosides in Escherichia coli. Appl Microbiol Biotechnol 80:253-260

Yamazaki M, Nakajima J, Yamanashi M, Sugiyama M, Makita Y, Springob K, Awazuhara M, Saito K (2003) Metabolomics and differential gene expression in anthocyanin chemo-varietal forms of Perilla frutescens. Phytochemistry 62:987–995

Ziauddin M, Phansalkar N, Patki P (1996) Studies on the immunomodulatory effects of Ashwagandha. . J Ethnopharmacol 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

Publications

Research papers published/Under review

1. Functional characterization, homology modeling and docking studies of β-glucosidase responsible for bioactivation of cyanogenic hydroxynitrile glucosides from *Leucaena leucocephala* (Subabul).

Noor M. Shaik[a], Anurag Misra[b], **Somesh Singh[a]**, Amol B. Fatangare[a], S. Ramakumar[b], Shuban K. Rawal[a], Bashir M. Khan[a*] (**Molecular Biology Reports**, (2013) 40: 1351-1363).

2. *In Silico* mutagenesis and Docking studies of active site residues suggest altered substrate specificity and possible physiological role of Cinnamoyl CoA Reductase 1 (Ll-CCRH1).

Prashant Dhanajirao Sonawane, Krunal Patel, Rishi Vishwakarma, **Somesh Singh**, Bashir Mohammad Khan. **Bioinformation** (Accepted, Feb 2013)

3. Molecular cloning and characterization of two differentially expressed cellulose synthase gene isoforms in *Leucaena leucocephala*: A pulp yielding tree species.

Rishi K. Vishwakarma, Sameer Srivastava, **Somesh Singh**., B. M. Khan*

(**Advances in Bioscience and Biotechnology**, 2012, 3, 92-100).

4. Molecular Cloning, Biochemical Characterization, and differential expression studies of Acetyl-CoA C-Acetyltransferase gene (AACT) from Brahmi (*Bacopa monniera*).

Rishi K. Vishwakarma; Ruby; **Somesh Singh**, Prashant D. Sonawane, Sameer Srivastava, Uma Kumari, R J Sanosh Kumar and Bashir M. Khan. **Plant Molecular Biology Reporter** (Accepted, 2 Nov 2012).

5. Molecular characterization of farnesyl pyrophosphate synthase from *Bacopa monniera* comparative modeling and docking studies.

Rishi K. Vishwakarma, Krunal Patel, Prashant D. Sonawane, **Somesh Singh**, Ruby, Uma Kumari, D. C. Agarwal and Bashir M. Khan. **Bioinformation** 2012, 8 (22): 1075-1081.

6. High frequency regeneration of *Leucaena leucocephala* –A leguminous pulpwood tree species

by: Noor Shaik, Manish Arha, A. Nookaraju, Sushim Gupta, Sameer Srivastava, Arun Yadav, Pallavi Kulkarni, O. Abhilash, Rishi Vishwakarma, **Somesh Singh**, Rajeshri Tatkare, Kannan Chinnathambi, Shuban Rawal, Bashir Khan

Physiology and Molecular Biology of Plants, Vol. 15, No. 4 (1 October 2009), pp. 311-318.

7. Functional characterization of a flavonoid glycosyltransferase gene from *Withania somnifera* and production of plant specific glucosides utilizing genetically engineered *E. coli.*

Somesh Singh, Rishi K. Vishwakarma, Santosh R.J. Prashant Sonawane, B. M. Khan. **Applied Biochemistry and Biotechnology** (Under review).

8. Molecular and docking studies of UGT73A16 glycosyltransferase from *Withania somnifera*

Somesh Singh, Krunal patel, Prashant Sonawane, Rishi K. Vishwakarma, B. M. Khan. **Journal of Molecular Modeling** (Communicated).

Book Chapter

1. "Genetic Engineering of phenylpropanoid pathway in *Leucaena leucocephala*" has been published online January 2012 in the book "Genetic Engineering- Basics, New Applications and Responsibilities", ISBN 978-953-307-671-3. INTECH-OPEN ACCESS PUBLISHER.

List of posters/Abstract

1. cDNA Cloning And Heterologous Expression of a Family 1 Glycosyl transferase From *Withania somnifera.*

Authors: **Somesh Singh**, R.J.Santosh, Ruby, Rishi K. Vishwakarma, Sameena Almas, *Dr.B.M.Khan, Presented in National Symposium at IHBT Palampur.

2. Molecular study of glycosylation of polyphenolic compounds from *Withania somnifera* and its application on human health

Authors: **Somesh Singh**, R.J.Santosh, Ruby, Rishi K. Vishwakarma, *Dr.B.M.Khan,

Presented and received **Best Poster** award in **INSA Platinum jubilee international symposium 2009** on Research in molecular medicine based on natural resources and traditional knowledge at National chemical laboratory, pune, Maharashtra.

3. Glycosyltransferase involved in flavonol glycoside biosynthesis in *Withania somnifera*

Authors: Somesh Singh, R.J.Santosh Kumar, Rishi K. Vishwakarma, Ruby, Poonam Sharma, *Dr.B.M.Khan

Presented in International Symposium on Aromatic and Medicinal plants (AROMED) on 2010 at Central Institute of Medicinal and Aromatic Plants (CIMAP), Lucknow.

4. Heterologous expression of Family 1 Glycosyltransferase involved in flavonol glycoside biosynthesis in *Withania somnifera*

Authors: Somesh Singh, R.J.Santosh Kumar, Rishi K. Vishwakarma, Ruby, Poonam Sharma, *Dr.B.M.Khan

Presented in Science day at national chemical laboratory on 2010