STUDY OF GLYCOSYLTRANSFERASES AND OTHER STRESS-RELATED PROTEINS ENCODED IN CHICKPEA GENOME AND ANALYSIS OF THEIR STRUCTURES AND FUNCTIONS USING MOLECULAR MODELING AND DOCKING

THESIS SUBMITTED TO
SAVITRIBAI PHULE PUNE UNIVERSITY

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN BIOTECHNOLOGY

BY

RANU SHARMA

RESEARCH GUIDE

Dr. C.G. SURESH

DIVISION OF BIOCHEMICAL SCIENCES
CSIR-NATIONAL CHEMICAL LABORATORY
DR. HOMI BHABHA ROAD, PUNE 411008
MAHARASHTRA, INDIA

NOVEMBER 2014

DECLARATION BY THE CANDIDATE

I hereby declare that the thesis entitled "**Study of glycosyltransferases and other stress-related proteins encoded in chickpea genome and analysis of their structures and functions using molecular modeling and docking**" submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me during the period from November 2011 to August 2014 under the guidance of **Dr. C.G. Suresh, Chief Scientist** and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or other institution of higher learning.

I further declare that the material obtained from other sources has been duly acknowledged in the thesis.

Ranu Sharma

Division of Biochemical Sciences
CSIR-National Chemical Laboratory
Pune - 411 008
November 2014

Dedicated to my beloved husband
and parents

ACKNOWLEDGEMENT

This thesis was made possible by the immense and much appreciated help and support of a number of people and I take this opportunity to thank them. First and foremost, I owe my obeisance to the almighty for showering his blessings on me for successful completion of my doctoral work.

I would like to thank my research guide Dr. C.G. Suresh for his guidance, constant support, and valuable suggestions throughout the course of my doctoral candidature. His scientific knowledge and innovative ideas have inspired me profoundly. I'm really grateful to him for giving me freedom to think, plan and execute my work. His punctuality, enthusiasm, and devotion towards work set an ideal example before me, which has always given me a boost for carrying out the given task in disciplined and timely manner. He has given me all the freedom to think and plan my experiments independently. He was always available with his advice and critical suggestions whenever needed. I am very thankful to him for providing several opportunities for me to learn novel things and to test my scientific skills.

I would like to thank my co-guide Dr. B.M. Khan for his valuable suggestions. I really appreciate the timely and constructive suggestions offered by Dr. Sureshkumar Ramasamy. I am extremely grateful to Vimal, Max Planck Institute for Plant Breeding Research, Cologne for his help in RNA-seq data analysis of chickpea. My sincere thanks to Vinod Jani, CDAC, Pune for his valuable suggestions related to molecular dynamics studies. I would like to thank all my seniors and other lab members Urvashi, Nishant, Manas, Tulika, Payal, Priyabrata, Ruby, Prachi, Deepak, Manu, Ameya, Shridhar, Aditi, Tejashree, Deepanjan, Yashpal, Vijay, Debjyoti, Rutuja, Selvi, and Swati for maintaining lively and cordial atmosphere in the lab. I reserve special thanks to Tulika for her impeccable suggestions and providing moral support during tough time. I will always be grateful to Priyabrata and Manas for their constant support, advices, and timely help.

I would like to thank Dr. Sourav Pal, Director, CSIR-NCL for providing me the necessary facilities and giving me the opportunity to work in this prestigious institution. I'm grateful to Dr. J.K. Pal, Head, Department of Biotechnology, University

of Pune and Dr. Sushama M. Gaikwad for being members of my evaluation committee. Their inputs have been invaluable. I would also like to thank Indira Mam, Sheetal ,Yashashree and the SAC office members Puranik Mam and Kolhe Mam for their help in administrative work.

Nothing can be compared to the love and sacrifice of my dearest husband Vimal who has always been a passionate and patient companion to me. I am not sure if this tribute could compensate for the big time I had to be away from him but this is built up on his boundless affection, understanding and constant encouragement. He has been a fountain of inspiration throughout my life without whose help this work would have been impossible.

My friends Akanksha, Nayana, and Awanti have always stood by my side through good times and rough phases. Their support and friendship is invaluable. I would like to thank my family Papa, Mummy, my brother Nishant and sister-in-law Priya for their constant support. They have been extremely patient through all my highs and lows and provided me great encouragement. They always encouraged my dreams and will always be my strength. I would also like to thank my in-laws for their love and understanding. Their encouraging words and belief in me have helped me reach my goals. Special thanks are due to my mom-in-law for her support and affection.

I thank CSIR, India for supporting my research work through Research fellowships. Finally I would like to thank all those who have directly or indirectly supported me.

Ranu Sharma

CONTENTS

ABBREVIATIONS

Abbreviation	Long form
AB	Ascochyta blight
ABA	Abscisic acid
ABI	Abscisic acid insensitive
ABRE	Abscisic acid responsive element
ACC synthase	1-aminocarboxylase-1-cyclopropane synthase
ADH	Alcohol dehydrogenase
AFPs	Antifreeze proteins
ANFs	Anti-nutrition factors
ANPs	Anaerobic polypeptides
AP	Aspartate protease
APR	Adult plant resistance
BBI	Bowman-Birk inhibitor
BGM	Botrytis grey mold
BLAST	Basic Local Alignment Search Tool
BR	Brassinosteroids
CAM	Calmodulin
CC	Coiled-coil
CK	Cytokinin
CNL	CC-Nucelotide binding site-Leucine rich repeat
COR	COLD RESPONSIVE
CP	Cysteine protease
CPIs	Cysteine protease inhibitors
CTPs	C-terminal processing peptidases
DHQ	Dihydroquercetin
Ein2	Ethylene insensitive 2
ERFs	Ethylene response factors
EST	Expressed sequence tag
ET	Ethylene
ETI	Effector triggered immunity
ETS	Effector-triggered susceptibility
F3GTs	Flavonoid-3-O glycosyltransferases

F7GT	Flavonoid-7-O glycosyltransferase
FAO	Food and Agriculture Organization
FPKM	Fragment Per Kilobase of transcript per Million
Frs1	Freezing sensitive 1
FW	Fusarium wilt
GA	Gibberellic acid
GRAN	Granulin
GTH	Glutathione
GTs	Glycosyltransferases
HMM	hidden Markov model
HQ	Hydroquinone
HSE	Heat shock element
HSFs	Heat shock factors
HSPs	Heat shock proteins
IAA	Indole acetic acid
ICE1	Inducer of CBF Expression 1
IP3	Inositol-1,4,5,-triphosphate
IPT	Isopentenyl transferase gene
ISPs	Ice structuring proteins
JA	Jasmonic acid
KMG	Kaempferol-3-O-glucoside
LDH	Lactate dehydrogenase
LEA	Late embryogenesis abundant
LG	Linkage group
los5	Low expression of osmotically responsive genes 5
LTP	Lipid transfer proteins
LTRE	Low temperature response element
MAMPs	Microbial associated molecular patterns
MG	Methyl glyoxal
MP	Metalloprotease
NBS-LRR	Nucelotide binding site-leucine rich repeat
NCBI	National Centre for Biotechnology Information
NJ	Neighbor joining
NLR	Nucelotide binding site-leucine rich repeat receptor
NO	Nitric oxide
NPC	Nuclear Pore Complex

PAMPs	Pathogen-associated molecular patterns
PCD	Programmed cell death
PCPs	Pro-X carboxypeptidase
PDB	Protein data bank
PGIP2	Polygalacturonase-inhibiting proteins
PIP	Plasma membrane intrinsic protein
PIs	Protease/proteinase inhibitors
PLC	Phosphotidyl-inositol-4, 5- bisphosphate specific lipase C
PLCP	Papain-like cysteine protease
POP	Prolyl oligopeptidase
PPIs	Plant protease inhibitors
PR	Pathogenesis-related
PRR	Pattern recognition receptors
PSI	Plant specific insert
PSPG	Plant secondary product Glycosyltransferase
PSWM	Position-Specific Weight Matrix
PTI	PAMP triggered immunity
QTL	Quantitative trait loci
RCL	Reactive centre loop
RD21	Responsive to desiccation 21
R-genes	Resistance genes
RMSD	Root mean square deviation
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RUBP	Ribulose-1, 5 biphosphate
SA	Salicylic acid
SCP	Serine carboxypeptidase
SD	Standard deviation
SDF	Structure data file
Serpin	Serine protease inhibitor
SGTs	Sterol Glycosyltransferase
SP	Signal peptidase
SP	Serine protease
SPIs	Serine protease inhibitors
SR	Seedling resistance
SRA	Sequence read archive

STI	Soyabean trypsin inhibitor
Swarps	Systemic wound-response proteins
TIP	Tonoplast intrinsic protein
TIR	Toll-interleukin-1 receptor/ resistance
TNL	TIR-Nucelotide binding site-Leucine rich repeat
TP	Transition polypeptide
U2F	URIDINE-5'-DIPHOSPHATE-2-DEOXY-2-FLUORO-ALPHA- D-GLUCOSE
UDP	Uridine diphosphate
UGTs	UDP-glycosyltransferases
UTRs	Untranslated regions
ZPR	Z-Pro-prolinal

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Plants, being sessile, are continuously exposed to broad range of environmental stresses. The stress situations countered by the plants have been classified into two categories i.e. abiotic and biotic stresses. The abiotic stresses affecting the crops and plants have been extensively studied. It includes drought, salinity, heat, cold, chilling, freezing, high light intensity, ozone (O3), anaerobic stresses, excess or deficiency of nutrient/ mineral content in the surrounding environment and many more. Unlike in the controlled lab conditions, plants face a combination of stresses in open farms and that affect the crop productivity drastically.

A survey of all the major weather disasters in US, between year 1980 and 2012, revealed that a combination of drought and heat stress caused huge agricultural losses that amounted to around $200 billion. In the current climatic conditions there is a gradual increase in temperature, accompanied by an increase in frequency and amplitude of heat stress. Apart from this, plants also face the threat of infection by pathogen (including bacteria, fungi, viruses, and nematodes), and attack by pests and herbivore. A strong correlation is suggested between climate change and habitat range of pest and pathogen. Abiotic stress weakens the plant defense mechanism and makes them more susceptible to pathogen infection. As a result, in the near future plants are more likely to be exposed to a broad range of stress conditions single or in combination. Therefore immediate task is to generate crops with enhanced tolerance to combinations of stresses.

Plants react in three ways when countering detrimental stress conditions: resistance, avoidance, and susceptibility. Stress resistance or tolerance requires the plant to get adjusted or acclimate to the stress condition. Indeed the stress tolerance capacity of plants depends on the genetic make up to acclimate to the stress and establish a new homeostatic state over time. Avoidance mechanisms lessen the effect of a stress, even though the stress is present in the environment. In such situations, to withstand drought conditions, plants reduce dehydration of cells by minimizing water loss and maximizing water uptake. They reduce water loss by stomata closure, reducing light absorbance by rolling leaves, forming dense trichome layer to increase reflectance, and by decreasing the area of leaf canopy through reduced growth or

shedding of older leaves. In addition to that, plants maximize water uptake by increasing investment in the roots.

The signal transduction pathway is triggered when the signal is perceived by the cell membrane receptors followed by generation and release of secondary messengers like Ca^{2+}, reactive oxygen species (ROS), and inositol phosphates. These secondary messengers further influence the intracellular calcium level which is sensed by the calcium binding proteins or Ca^{+2} sensors. These sensory proteins then interact with their respective partners that lead to a cascade of phosphorylation events and target the major stress responsive genes or the transcription factors controlling these genes. The products of these stress genes ultimately help the plant to adapt and survive in such unfavorable conditions. Thus, in the beginning plant respond to stresses as individual cells followed by the transmission of signal throughout the organism and behave synergistically as a whole organism.

According to Food and Agriculture Organization (FAO) 2008, chickpea is one of the ancient and second most widely grown crops in the world [4]. Chickpea seeds are the primary source of human dietary nitrogen and are rich in proteins, carbohydrate, fibers, vitamins, and minerals. It posses zero cholesterol content [5]. Moreover various studies have shown that the cholesterol level in blood can be reduced by consuming chickpea [6]. Total losses in chickpea production caused by abiotic stress outnumber those due to biotic stress. Major abiotic stresses responsible for crop damage are drought, salinity, and cold. On the contrary the biotic agents causing severe damage to chickpea includes *Ascochyta* blight caused by *A. rabiei* (Pass) Labr.; BGM caused by *B. cinerea* Pers; *Fusarium* wilt caused by *F. oxysporum* f. sp. ciceri; Dry root rot caused by *M. phaseolina*; Collar rot caused by *Sclerotium rolfsii*; and Phytophthora root rot caused by *Phytophthora medicaginis*. Another major concern is the attack of insect pests that mainly includes Pod borer caused by *H. armigera* Hubner; leaf miner caused by *Liriomyza ciceriana* Rondani; and seed beetle caused by *Callosobruchus* spp. in major production areas [7].

In the present thesis, an *in silico* identification and characterization studies were carried out for the genes involved in stress and defense mechanisms in chickpea genome translated in high copy number. The stress genes under study includes UDP-glycosyltransferases (UGTs), nucleotide-binding site leucine rich repeat proteins (NBS-LRR), proteases, protease inhibitors, peroxidase, lipid transfer proteins, heat

shock proteins, late embryogenesis abundant, and pathogenesis-related proteins like chitinases, glucanase, and thaumatin. The close orthologous relationship of the identified gene products was established with sequenced dicot genomes like *Medicago truncatula* and *Glycine max* using Blast2Go tool. A few genes have diverged in chickpea genome to the extent that no close orthologous relationship could be discerned by comparison with other sequenced plant genomes. The phylogenetic analysis was carried out using MEGA package for the different classes of stress proteins. The dendrograms revealed substrate specific distinct clustering ~~of~~ that facilitated functional classification and annotation. A few possible recent gene duplication events were identified through analysis of sequence relation.

Gene expression studies were performed by taking advantage of the RNA-seq data available under Sequence Read Archive toolkit (SRA) in NCBI from five plant tissues viz. germinating seedling, flower, flower buds, shoot apical meristems, and young leaves and by using drought and salinity responsive EST libraries (EST number: 45038) too. The results showed a differential expression pattern for most of the genes in one or more tissue. Some of the stress genes had zero (unexpressed) or very low expression value (basal expression), but this doesn't say that they are non-functional. These lowly or unexpressed genes can express in some other plant tissues which are not checked or they may express under certain environmental or stress conditions.

The domain architecture and motifs were identified using HMMER and MEME suite to analyze the architecture of the proteins. The crystal structure of UGTs showed two distinct functional domains, the N-terminal domain binds to the acceptor substrate and the C-terminal domain accepts the sugar donor substrate. The proteins encoded by NBS-LRR gene family possess three important domains; the N-terminal either possesses homology with Drosophila Toll and Human Interleukin-1 receptors (TIR) (TNL family) or in its place a coiled-coil (CC) is present (CNL or non-TNL family). The NBS domain has several conserved motifs that bind and hydrolyze ATP or GTP. The LRR domain is involved in protein-protein interactions and thus play role in molecular recognition and specificity. The motif analysis of these three functional domains revealed four TIR/ two CC motifs, 9/ 10 signature motifs in the NBS domain, and six motifs in the LRR region in the two families. The four protease classes of enzyme namely aspartate protease, cysteine protease, serine protease, and

metalloprotease were also studied for analyzing the domain arrangement. All the members of the four classes possess a peptidase domain for performing the hydrolysis reaction but comparatively they possess more domain diversity within the class as well as between members.

The structure-function studies could be performed only for UGTs, protease, and protease inhibitor class of enzymes due to unavailability of suitable templates in the database for modeling studies. The sequence and phylogenetic analysis of UGT proteins from various plant sources were performed with the help of ClustalX tool and MEGA package. The dendrogram revealed a distinct cluster of 30 flavonoid-3-O-glycosyltransferase members, for which the 3-dimensional structures were modeled by employing Prime 3.1 utility of maestro. The docking studies with flavonoid acceptors, more specifically flavonol and anthocyanidin and sugar donor substrates were carried out in Glide utility of maestro. All the docked complexes revealed a conserved environment in the vicinity of the sugar acceptor that favored the regiospecific glycosylation of the 3-OH group of flavonoid. Molecular dynamics studies were performed in GROMACS to further validate the above finding. The trajectories of all the docked complexes were compared in terms of the root mean square deviation (RMSD) of the Cα atoms, root mean square fluctuation values of the key residues, and the intermolecular distance between the acceptor and the key residues. The above statistics and analysis showed stable and consistent binding of the ligands in the binding pocket over the complete trajectory. The above findings were utilized to assign probable functions to the UGTs in chickpea.

The *in silico* structural studies of prolyl oligopeptidase (POP) family of serine protease and protease inhibitors like cystatin and serine protease inhibitor like Bowman-Birk inhibitor (BBI), potato inhibitor-I (PI-I), Kunitz-type inhibitor, serpin, and serine protease inhibitors were also performed. The homology model of POP was built using homologous protein from *Sus scrofa* (PDB: 1E8M, identity: 55%, Resolution: 1.50 Å). Similarly the structures of PI-I, serpin, and BBI were modeled using the 3-D coordinates of potato inhibitor from barley seeds (PDB: 1C12, Identity: 30%, Resolution: 2 Å), serpin from *Arabidopsis thaliana* (PDB- 3LE2, Identity: 57%, Resolution: 2.20 Å), Bowman-Birk trypsin inhibitor from *Medicago scutellata* (PDB: 2ILN, Identity: 69%, Resolution: 2 Å). The ten Kunitz inhibitors were modelled using following crystal structures: 2QN4, 3ZC8, and 1R8N (Identity >30%, Resolution:

1.80 Å, 2.24 Å, 1.75 Å). The cysteine protease inhibitor was modeled using crystal structure of tarocystatin and papain complex (PDB-3IMA, Identity: 58%, and Resolution: 2.03 Å). The docked complex of POP with ZPR inhibitor showed two important hydrogen bonds between O9 and O16 atoms of ZPR and NH1 atom of Arg656 and NE1 atom of Trp604 with a bond length of 3.16 Å and 2.92 Å. The catalytic Ser563 and His696 were involved in mutual hydrogen bond of length 3.3 Å.

Subsequently, docking studies were performed which revealed a high Z-score (value: 1224.27) for the docked complex between cysteine protease inhibitor of chickpea and papain from *Carica papaya*. Similarly, BBI and PI-I were docked in the binding pockets of bovine trypsin with the Z-score values of 2227.32 and 1883.65. Five hydrogen bonds were observed at the interface of the two trypsin molecules and BBI i.e. F41-Y92, G216-C88, N97-N73, and H57-S65. Only two inter-molecular hydrogen bonds, between K61-D60 and S213-D54, were observed in the PI-I and trypsin docked complex.

All the five docked complexes were subjected to a simulation of 10 ns time duration to explore the changes in the structures. The Cα RMSD graph concluded that the generated structures were stable over the complete trajectory. Upon analyzing the trajectories, several changes have been observed in the hydrogen bonding network between the protease inhibitors/ proteases and their cognate target proteases/ inhibitors. After 10 ns, O16 atom of ZPR was involved in hydrogen bonds with Tyr482, Ser563 and His696 with a bond length of 2.95, 2.80, and 2.70. In the initial structure, prior to the simulation, a hydrogen bond was seen between the catalytic Ser563 and His696 residue which at the end of the simulation was lost and involved in the interaction with the inhibitor. Even the hydrogen bond with Trp604 which was observed in the initial structure was also lost. Similarly after 10 ns, the final conformations of BBI, CPI, and PI complexes revealed formation of more stable complex owing to the formation of more number of inter-molecular hydrogen bonds. The two highly conserved arginine residues in the PI-1 family are important to support the reaction site loop with respect to the main body of the inhibitor. The two arginine residues in chickpea PI-trypsin complex were involved in an intra-molecular hydrogen bonding with Asp60. The superposition of initial and final structures after MD showed an RMSD deviation of ~1Å.

Other stress genes namely heat shock proteins (HSPs), late embryogenesis abundant (LEA) proteins, peroxidases, chitinases, glucanases, thaumatin, and lipid transfer proteins (LTP), involved in abiotic and biotic stress conditions were found in large copy number. Maximum number of orthologs was detected in *G. max* and least number was found in *A. thaliana* (36). Four recent gene duplication events were seen in the seven stress gene classes. Out of the total 326 genes identified 220 were reported to show expression in one of the 5 tissues selected whereas 106 showed no expression. We observed up-regulation as well as down-regulation of some of the stress genes under drought condition. A number of genes revealed expression in root, shoot, immature seed, root & collar, and leaf tissues based on EST data analysis. Analysis of the promoter region of the identified stress genes showed overrepresentation of *cis*-regulatory elements involved in stress conditions and pathogen attack.

Using computational approaches, the identification, functional annotation, gene/ protein features, and expression studies of stress genes and its products have been demonstrated. These properties and features highlighted the importance of these stress genes in towards improving the chickpea productivity. These findings will help to fish out candidate stress genes in chickpea that will further aid in designing genetically engineered stress genes which leads to the development of an improved disease resistant and stress tolerant variety of crop with better nutritional value and higher yield.

Chapter 1

Introduction to defense mechanisms in plants including chickpea, molecules involved and plant survival under stress conditions

Plants, being sessile in nature, are often exposed to a wide range of detrimental environmental conditions which adversely affect their growth and productivity. However, they have developed distinct mechanisms to perceive environmental changes and respond to complex stress conditions which help in minimizing the damage and conserve the valuable resources required for the growth, survival and reproduction. Plants activate unique mechanisms of stress response and behave differently when exposed to multiple stress conditions [Rizhsky *et al*, 2004].

1.1 Types of stresses encountered by plants

1.1.1 Abiotic stress

The abiotic stresses are caused due to either excess or deficient physical or chemical environment of the plant. They include drought, salinity, heat, cold, chilling, freezing, high light intensity, anaerobiosis, heavy metals, gaseous pollutants, UV radiation and many such conditions [Wang *et al*, 2003; Chaves & Oliveira, 2004; Agarwal & Grover, 2006; Nakashima & Yamaguchi-Shinozaki, 2006; Hirel *et al*, 2007; Bailey-Serres & Voesenek, 2008; Suzuki *et al*, 2014] (Figure 1.1). Many agricultural areas are affected by a combination of stresses like drought and salinity, salinity and heat, drought with high light intensity or temperatures that influence the crop productivity. A survey of major weather disasters in US, between 1980 and 2012, that accounts for loss of billions of dollars showed that a combination of drought and heat stress caused extensive loss to the crop productivity estimated around $ 200 billion. The current climate prediction model predicts a gradual increase in the temperature and enhancement in the frequency and amplitude of heat intensity in near future [Ahuja *et al*, 2010; Mittler & Blumwald, 2010; Mittler *et al*, 2012; Li *et al*, 2013]. Not only this, high temperatures in combination with other weather disasters like prolonged drought condition will severely affect the crop productivity worldwide (IPCC, 2008). Hence, there is a great need to understand stress tolerance and develop stress tolerant variety of crops that can cope with such adverse climatic conditions.

1.1.2 Biotic stress

In addition to abiotic stresses, plants are also exposed to attack by pests and pathogens (bacteria, fungi, virus, and nematode) and attack by herbivore [Atkinson & Urwin,

2012] (Figure 1.1). Studies have shown that the resistance capacity of plants against pathogens gets reduced due to high temperature and it facilitates spread of pathogens [Bale *et al*, 2002; Luck *et al*, 2011; Madgwick *et al*, 2011; Nicol *et al*, 2011]. Exposure of abiotic stress leads to impairment of the defense mechanism in plants and make the plants more prone to pathogen attack [Amtmann *et al*, 2008; Goel *et al*, 2008; Mittler & Blumwald, 2010; Atkinson & Urwin, 2012].

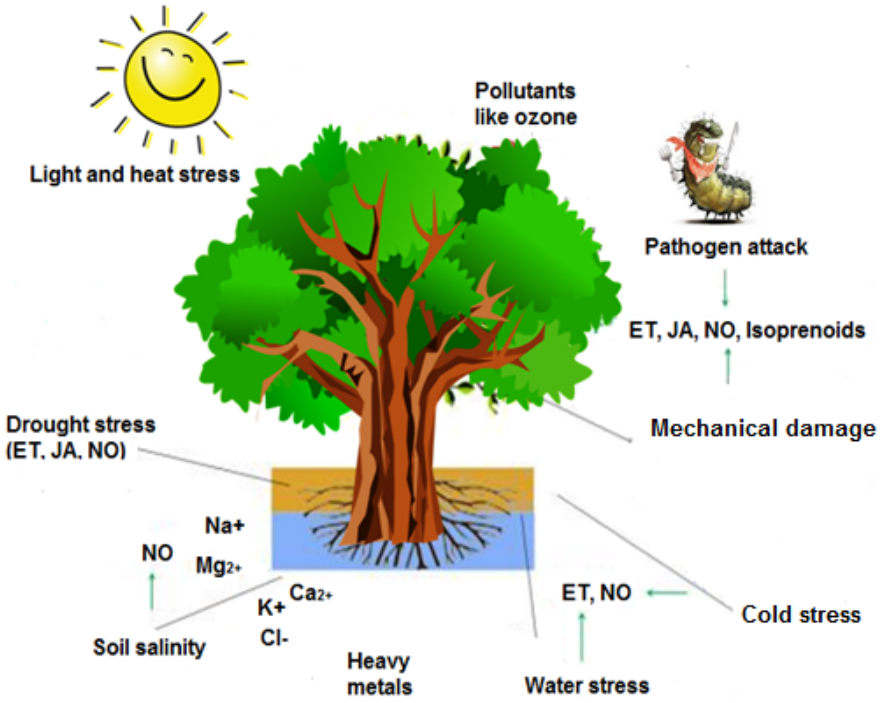


Figure 1.1: All the abiotic stresses affecting plant trigger complexe responses aiming to increase the stress tolerance. The responses include the emission of Volatile Organic Compounds (VOCs). The VOCs like ET, JA, NO, Na^+ etc produced in response to the different stresses are also shown. A common mechanism links together the different stresses: all cause oxidative stress and hamper the production of reactive oxygen species. Excess light and heat, as well as exposure to oxidizing air pollutants, cause direct accumulation of ROS which crucially contributes to initiate the stress-related signal cascades (Figure is adopted and modified from Francesco Spinelli *et al*, 2011).

1.2 Importance of countering stress by chickpea

Cicer arietinum, commonly known as chickpea, bengal gram, ceci bean, channa or garbanzo beans is a legume belonging to the plant family *Fabaceae*, subfamily *Faboideae*. It is mainly grown in temperate and semi-arid regions of Asia, Europe, Australia and North America; India happens to be leading producer contributing

approximately 66% of global production. *Kabuli* and *Desi* are two varieties of chickpea. *Desi* chickpea grains are small, thick, dark and have a jagged surface and cultivated mainly in semi-arid land. *Kabuli* variety is slightly larger than *Desi*, has a thin, bright cover grain, smooth seed surface and is grown in temperate climates (Agriculture, 2006). The two chickpea cultivars *Desi* and *Kabuli* differ primarily in content of protein, fiber, polyphenols, and carbohydrates.

According to Food and Agriculture Organization, 2008 (FAO) chickpea is one of the ancient and third most widely grown legumes in the world after dry beans and peas [Sharma *et al*, 2013]. Chickpea seeds are a primary source of human dietary protein owing to its capacity for symbiotic nitrogen fixation. It is characterized by low or zero cholesterol content and high content of vitamins, minerals, carbohydrate, protein, and fibers [Jukanti *et al*, 2012]. Moreover carbohydrate and protein content of chickpea alone constitute about 80% of the total dry seed mass. Research has shown that the consumption of chickpea seeds reduces the cholesterol level in blood [Pittaway *et al*, 2008]. It has carotenoids like β-carotene, lutein, zeaxanthin, β-cryptoxanthin, lycopene and α-carotene. Chickpea contains phenolic compounds like isoflavones, biochanin A, formononetin, daidzein, genistein, matairesinol and secoisolariciresinol [Jukanti *et al*, 2012]. The essential amino acid (except sulphur-containing) and the endogenous amino acid content are higher in chickpea flour as compared to wheat flour which also portrays its high nutritional value.

Despite its several health promoting factors, chickpea is also known for its anti-nutritional factors (ANFs) which interfere with the digestive system of animals and impair utilization of nutrients. The anti-nutritional factors can be categorized into protein and non-protein molecules. Proteinaceous ANFs include trypsin and chymotrypsin inhibitors, lectins and antifungal peptides. Non-protein ANFs includes alkaloids, tannins, phytic acid, saponins, acrylamide, and phenolics [Sharma *et al*, 2011]. Such anti-nutritional factors can be reduced or eliminated by employing different cooking techniques. Acrylamide is an anti-nutritional substance present in foods such as bread, snacks and chips. Adding chickpea flour may be a novel method to reduce acrylamide content in such products [Rachwat *et al*, 2013]. Furthermore, increased intake of chickpea helps to prevent cardiovascular disease, coronary heart disease, type 2 diabetes, digestive diseases, cancers, also helps to reduce cholesterol levels, control blood pressure and bring many more health benefits.

1.2.1 Types of stress encountered by chickpea

Production of chickpea is drastically affected by various biotic (*Helicoverpa, Aphids, Callosobruchus, Bromus and Orobanche*) and abiotic (drought, heat, cold, and salinity) constraints. Studies have shown annual salt induced chickpea yield losses of about 8 to 10% globally [Flowers *et al*, 2010]. In chickpea, the total crop loss caused by abiotic stress exceeds those due to biotic stress. Its production is affected by the attack/ infection by pod borer (*Helicoverpa armigera*), black aphid (*Apis craccivora*), *Ascochyta* blight caused by *Ascochyta rabiei*, *Fusarium* wilt caused by *Fusarium oxysporum*, root rot (*Phytophthora medicaginis*), *Botrytis* grey mould (BGM) caused by *Botrytis cinerea* Pers, stunt virus, dry root rot caused by *Macrophomina phaseolina*; Collar rot caused by *Sclerotium rolfsii* [Chérif *et al*, 2007] (www.icrisat.org/bt-pathology-fungal.htm) that leads to extensive crop damage affecting chickpea productivity. Bruchid beetles (*Callosobruchus maculates, C. chinensis*) cause significant loss during storage [Yadav *et al*, 2007]. However, in the field condition, weeds (*Bromus sp.* and *Orobanche crenata*) also cause considerable damage to winter chickpeas.

1.3 Stress effects displayed by plants

1.3.1 Wilting-browning

Wilting refers to the loss of rigidity of non-woody parts of plants owing to fall of turgor pressure below zero or due to lack of water uptake by the cells. In such a situation, the rate of loss of water is greater than the absorption of water in the plant. Wilting diminishes the plant's ability to transpire and grow which when become permanent leads to plant death.

Low water availability may be the result of:

- **Drought conditions:** Soil moisture drops below the most favorable conditions for plant functioning
- **Freezing or chilling:** Temperature falls to such a level where the plants vascular system cannot function.
- **High salinity:** It causes water to diffuse out from the plant cells and induce shrinkage.

- **Saturated soil conditions:** Roots are unable to access sufficient amount of oxygen for cellular respiration which leads to inability to transport water into the plant.
- **Biotic agents:** Microorganism, nematodes, and insects can clog the plant's vascular system.

1.3.2 Necrosis

Necrosis (death of cells or tissues) is not a disease, but rather a symptom of disease or caused by other stresses experienced by plants. The symptoms are mainly dark watery spots on leaves or fruit to dry papery spots that may be tan or black. Some regions of the plant may appear yellow or wilted, indicating a disease conditions that leads to cell death. This condition further weakens the plant and makes it more susceptible to other diseases and pests. The main reasons could be water deficiency, salt and pesticides toxicity, nutrient deficiency, pollution, temperature extremes, and biotic agents.

1.3.3 Chlorosis

Chlorosis is a condition in which there is absence or insufficient synthesis of chlorophyll which makes the leaf look yellow, pale or yellow-white instead of green. Poor drainage, damaged roots, compact roots, high alkalinity, and nutrient deficiencies in the plant are the possible causes of chlorosis. Nutrient deficiencies may occur because of their insufficient amount in the soil or the nutrients are unavailable due to high pH (alkaline soil) or the nutrients may not be absorbed due to injured roots or impaired root growth. Most importantly, iron deficiency is one of the major causes associated with chlorosis [Abadía *et al*, 2011]. Plants need iron for the production of chlorophyll which gives green color to the leaves and is necessary for the synthesis of food required for its own growth. Other elements such as calcium, zinc, manganese, phosphorus, or copper in high amounts in the soil can tie up iron so that it is unavailable to the plant. However, a shortage of potassium, manganese, and or zinc will reduce the availability of iron to the plant. Even attack of biotic agents can cause chlorosis.

1.3.4 Oedema or water soaking

Oedema often spelled as 'edema' is a physiological disorder of plants caused by the roots taking up more water than the leaves can transpire. Excess moisture gets accumulated in the plant and causes swellings that appear initially as pale-green water-soaked blisters, mostly on the undersides of leaves. The cells eventually erupt, then the spots turn yellow, brown, brownish-red, or even black. Moreover severely affected leaves may turn yellow and drop off.

1.3.5 Distortion

The most common plant abnormality seen in plants is distortion that causes leaf curling or cupped leaves. Insects, mites, pathogens, herbicides and weather events can lead to impairment of leaf, stem, flower, and fruit tissue. Weed killers, such as 2, 4-D or Weed-N-Feed, sprayed in the affected area can volatilize and drift that causes curling of the foliage on non-targeted plants. Weather events, such as cold temperatures also can cause leaf malformation. Distortion in the flowers of a plant is due to aster yellows disease caused by phytoplasma, and causes a variety of unusual and strange symptoms. Plants generally overcome the damage caused by cold injury and low doses of herbicide. However if plant is infected by virus or aster yellows, entire plant must be removed to control the disease.

1.3.6 Defoliation

The condition of premature removal of foliage by cutting or grazing is known as defoliation. Cutting grass by mowing is usually uniform and clean. However grazing animals do not defoliate plants evenly because of the different mouth structures and different chewing habits of animals. Defoliation management involves removal of plant material so as to keep the growth meristems intact for carrying out various plant functions. Extensive leaf defoliation decreases the photosynthesis rate and most of the functions in plant decline. Rapid defoliation may be caused by low temperature, herbicide, water deficit, soil air deficit, and air pollution. Therefore, proper defoliation practices allow the plant to build up carbohydrate reserves to survive the dormancy and begin to re-grow in the next season.

1.3.7 Bleeding and gumming

Bleeding and gumming is the flow of sap from wounds or injuries. Sometimes it may be associated with foul smell too. The reasons behind this condition could be deficiency of water, mechanical injury, and pathogen attack.

1.3.8 Plant galls

Galls are bizarre growths occurring on leaves, twigs, or branches. They may be simple lumps or complicated structures, plain brown or brightly colored. These tumor outgrowths develop from rapid mitosis and come in a diverse array of colors, shapes and sizes. There are 1500 species of gall producers, the majority of which are insects and mites. Many galls provide the nutrition and shelter to various species of harmless insects for breeding or laying eggs. In return, these insects provide a vital service to their host plant in the form of pollination or protection in a highly competitive environment where these plants could otherwise not able to survive. It is caused due to several abiotic factors and biotic agents like disease organism, nematodes, insects, and mites.

1.3.9 Other effects of stress

Ragged leaves are mainly the damage caused by strong wind, hail or other abrasive action. However, feeding of insects or animals could also be one of the reasons. The main causes behind this deformity could be application of systemic herbicide, chlorosis, vein clearing, necrosis or nutrient deficiency like zinc/ manganese. The tree bark becomes sunken and discolored due to sunburn, the condition become worse if accompanied by water deficiency. Tree bark may peel off when plant dies or when vigor is low due to sunburn, water deficiency, biotic agents, or lightening injury. The roots becomes shriveled and darkened due to loss of turgor. The healthy roots are white to creamy in color which when under stress becomes black, grey, blue, or tan. Discoloration indicated non-functional roots or death of roots. The factors causing this could be low aeration in soil due to flooding and over-irrigation.

1.4 Sensing of stresses by plants

Plant cells initiate the stress signal transduction by perceiving the signal through cell membrane receptors followed by generation of secondary messengers like Ca^{2+}, reactive oxygen species (ROS), and inositol phosphates. These secondary messengers further affect the intracellular calcium level which is recognized by calcium binding proteins or Ca^{+2} sensors. However, these sensors don't have enzymatic activity but they change their structure in a calcium dependent manner. These sensory proteins then interact with their respective partners followed by phosphorylation cascade and target the major stress responsive genes or the transcription factors controlling these genes. The products of these stress genes ultimately lead to plant adaptation and help the plant to survive in the unfavorable conditions [Huang *et al*, 2012] (Figure 1.2). Moreover, the stress induced changes in the gene expression pattern leads to the generation of hormones such as abscisic acid (ABA), salicylic acid (SA), jasmonic acid (JA), and ethylene (ET). These molecules may intensify the initial signal or may initiate the next round of signal by following the same route or use different signaling pathway components [Mahajan & Tuteja *et al*, 2005; Xiong *et al*, 2002; Shao *et al*, 2007]. Thus, initially plant responds to stresses as individual cells followed by the transmission of signal throughout the organism and behave synergistically as a whole organism.

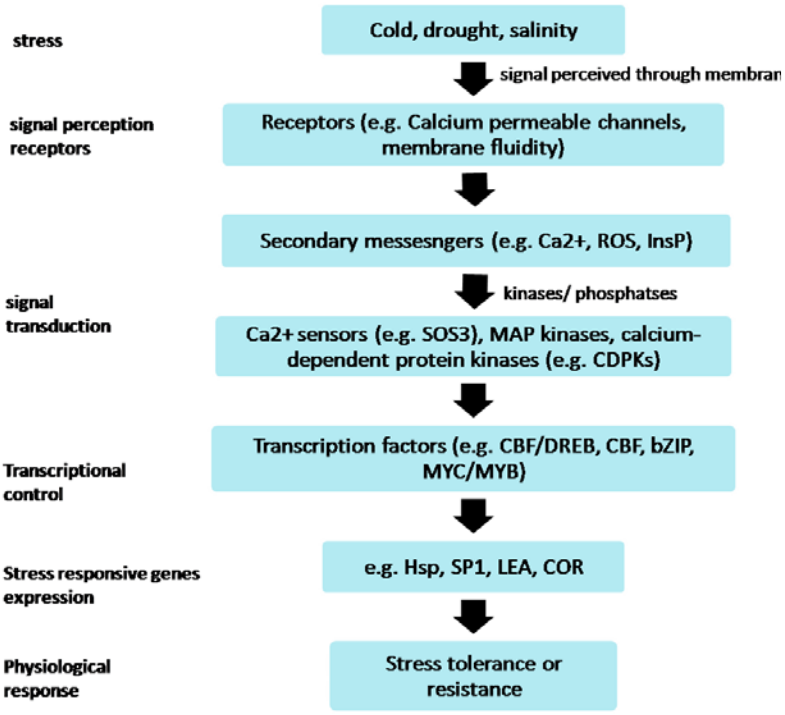


Fig. 1.2: Generic pathway for plant response to stress. The extracellular stress signal is first perceived by the membrane receptors and then activate large and complex signaling cascade intracellularly including the generation of secondary signal molecules. The signal cascade results in the expression of multiple stress responsive genes, the products of which can provide the stress tolerance directly or indirectly. Figure is adopted and modified from Huang *et al* 2012.

1.5 Mechanisms to counter stress

Plants circumvent the potentially harmful effects caused by stresses in three different ways- resistance, susceptibility, and avoidance [Kramer, 1980; Levitt, 1972] (Figure 1.3).

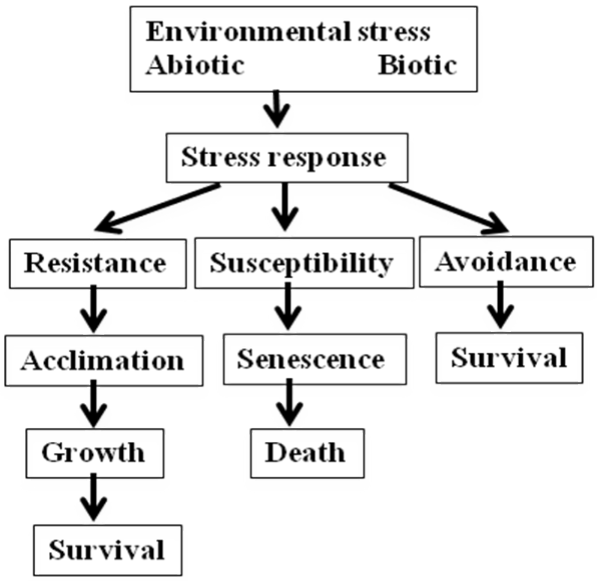


Figure 1.3: The effect of environmental stress on plant survival (*source: Hopkins W.G., Hüner N.P.A., 2009*).

Stress resistance or tolerance requires the plant to get adjusted or acclimate to the stress condition. The stress-induced modulation of homeostasis can be considered as the signal for the plant to establish a new homeostasis associated with the acclimated state [Wilson *et al*, 2006]. The stress tolerance or stress resistant capacity of plants depends on their genetic capacity to adjust or to acclimate to the stress and establish a new homeostatic state over time. Moreover, the acclimation process in stress-resistant species is usually reversible upon removal of the external stress.

Ephemeral and arctic annual plants escape the stress altogether and follow a stress avoidance mechanism to overcome adverse environmental conditions. They exhibit a high degree of developmental plasticity as they complete their life cycle before physiological deficit occur. They germinate, grow, and flower very quickly following seasonal rains/ arctic summer. Thus they rapidly complete their life cycle and form dormant seeds before the onset of the dry season/winter season. Because these plants never really experience the stress of drought or low temperature, these plants survive the environmental stress by stress avoidance.

Avoidance mechanisms lessen the effect of a stress, even though the stress is present in the environment. Plants can withstand drought conditions by minimizing water loss and maximizing water uptake. They minimize dehydration of cells by stomata closure, reducing light absorbance by rolling leaves, forming dense trichome layer to increase reflectance, and by decreasing the area of leaf canopy through reduced growth or shedding of older leaves. On the contrary, plants maximize water uptake by increasing investment in the roots. Plant breeding programs in semi-arid regions showed significant gain in crop productivity just by enhancing the depth of the roots. Additionally, shedding of older leaves will aid in water conservation and nutrient mobilization from the old leaves to the stem and young leaves.

Stress sometimes causes injuries to the plants, which imply that they exhibit one or more metabolic dysfunctions. If the stress is short term and moderate, the injury may be short-lived and the plant may recover when the stress is removed. If the stress is harsh enough, it may prevent flowering, seed formation, and induce senescence that leads to apoptosis or plant death. Such plants are considered to be susceptible.

As we have already discussed, along with the abiotic stress plants also face the threat of attack by pest and pathogens. These pathogens evade the plant cell and release the effectors or virulence factors that promote virulence and causes disease. In order to confront such situations, plants have developed a multilayered innate immune system that develops physical and chemical barriers that obstruct the pathogen entry. Moreover, they have also evolved a wide variety of inducible defense strategies like oxidative burst, expression of defense-related genes, synthesis of antimicrobial compounds, and programmed cell death (PCD) that are triggered upon pathogen attack. The basal defense mechanism or innate immune system that involves perception of microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) by host pattern recognition receptors (PRRs) results in PAMP triggered immunity (PTI) [Ausubel, 2005; Jones & Dangl, 2006]. However, successful pathogens secrete effector molecules that curb PTI and thus induce effector-triggered susceptibility (ETS). As a counter defense strategy, in order to detect the effector, plants activate effector triggered immunity (ETI) resulting in disease resistance [Jones

& Dangl, 2006]. Therefore, a synergistic activity of both PTI and ETI enhances plants' disease resistance capacity and restricts pathogen growth (Figure 1.4).

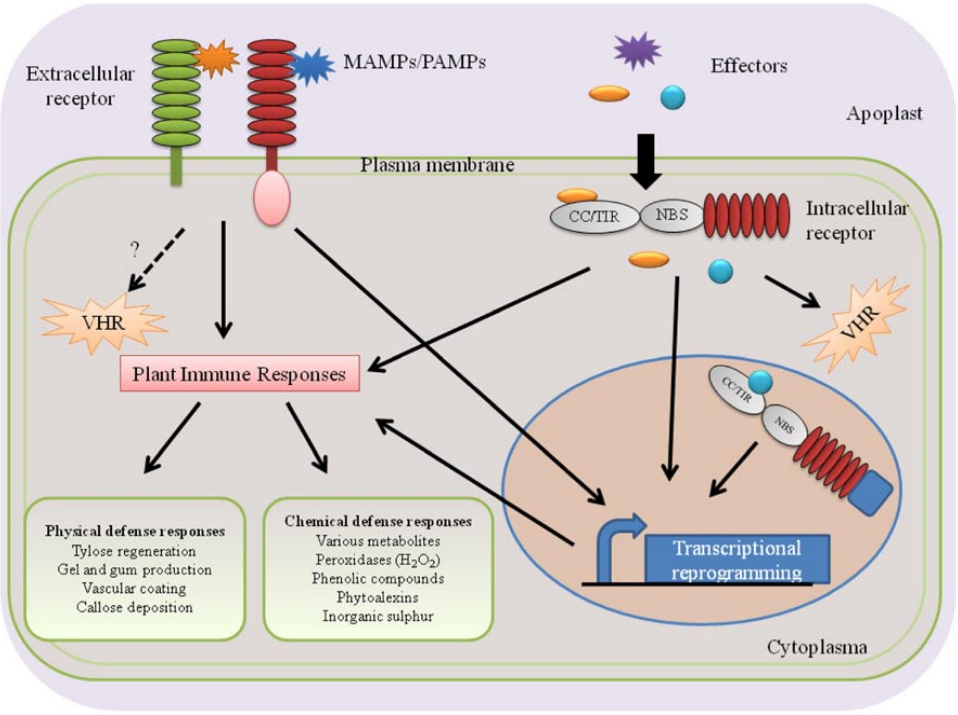


Figure 1.4: Plants perceive PAMPs/MAMPs or effector proteins using extracellular or intracellular receptors and activate immune responses. The tomato receptor-like protein Ve1 and the rice receptor-like kinase Xa21 are examples of extracellular receptors that recognize *Verticillium* Ave1 and *Xanthomonas oryzae* pv. oryzae Xa21, respectively. Tomato I-2 and *Arabidopsis* RRS1-R are examples of intracellular NB–LRR-type receptors that perceive the *F. oxysporum* f. sp. lycopersici Avr2 effector and the *R. solanacearum* effector PopP2, respectively. (Figure is adopted from Yadeta *et al*, 2013).

1.6 Stresses and stress management by plants

1.6.1 Salinity and water deficit

Salinity of the soil is one of the major hurdles in agricultural production worldwide, more specifically in arid and semi-arid areas. These regions suffer from salt movement from basal rocks to the upper layer of soil due to increased water evaporation. In nature, NaCL represent the major soluble salt causing soil salinity. About 7% of the total world agricultural land is affected drastically by salinity and this number could rise up to 20% in the near future due to land salinization, the result

of artificial irrigation and unsuitable land management. An increased concentration of salt content in the soil imbalance the water potential between plant root cell and soil solution that leads to cellular dehydration. Moreover, this osmotic difference further leads to salt penetration through plasma membrane and its accumulation in cell cytoplasm inhibiting the intracellular enzymatic activity [Munns & Tester, 2008]. In addition to salinity, the decrease in osmotic potential or osmotic effect is commonly seen in other abiotic stress as well like drought, increased evaporation rate, cold or freezing, mechanical wounding. Plants confront this osmotic imbalance by decreasing cellular osmotic potential through salt exclusion from the cell and accumulating salt ion in intracellular organelles (example: vacuoles). Not only this, plants have to face the oxidative stress too which leads to prolonged closure of stomata, decline in photosynthetic rate and reduction in plant growth [Kosová *et al*, 2011].

Plants actively counter stress by metabolism reprogramming to induce an enhanced stress tolerance capacity. Recently, considerable progress has been made in developing salt tolerance mechanisms, especially salt ion signaling and transport, by using modern genetic approaches and due to high-throughput methods of functional genomics (transcriptomics, proteomics, metabolomics, antioxidomics, etc.) [Munns, 2005; Hasegawa *et al*, 2000; Zhu, 2002; Yamaguchi-Shinozaki & Shinozaki, 2006]. In addition, proteins play an inevitable role in plant stress response since they are directly involved in acquiring an enhanced stress tolerance. Therefore, a thorough analysis of proteomics of plant abiotic stress response has a great potential in determining their key role in plant stress response and stress tolerance acquisition [Kosová *et al*, 2011]. Three major factors that determine the tolerance levels of plants to extreme environmental conditions were reported by Inan *et al*, 2004.

(1) Genomic level: Tolerant plants may possess some unique stress-responsive genes which are absent in susceptible plants (differences at genome structure level).

(2) Transcriptomic level: Tolerant plants have altered regulation of gene expression of important stress-responsive genes than susceptible plants (qualitative and quantitative differences at gene expression level)

(3) Proteomic level: Proteins involved in stress response reveal an altered activity in tolerant plants than in susceptible ones (differences in protein structure and activity level).

Plants adapt to salinity stress at all levels such as genomic, transcriptomic, proteomic, and metabolomic. At the genomic level, gene duplication event was seen to increase the copy number and adaptation of several salinity-responsive genes like transcription factor Myb24, ATPase AVP1, ion transporters: SOS1, NHX; ABC. An enhanced constitutive expression of salinity-responsive transcripts such as SOS1, SOD, P5C5, GS, INPS, cytochrome P450, and heat shock protein was observed [Dassanayake *et al*, 2011]. Transcriptomic analysis showed altered gene expression of important stress-responsive genes in tolerant plants than susceptible plants both qualitatively and quantitatively [Taji *et al*, 2004; Kant *et al*, 2006]. At the proteomic level, plant respond to stress by increased synthesis of several stress and defense proteins (LEA, redox, and PR), ion transporters, proteins involved in activation of photosynthesis (D2 protein), and activation of biosynthesis of protective compounds like lignin [Pang *et al*, 2010; Wang *et al*, 2008; Wakeel *et al*, 2011; Jellouli *et al*, 2008; Askari *et al*, 2006; Wang *et al*, 2009] (Table 1.1). Even at the metabolome level changes were noticed like altered carbohydrate metabolism that result in activation of catabolism (glycolysis, krebs cycle, and starch degradation), enhanced biosynthesis of organic osmolytes, phenolic compounds, and lignin [Gong *et al*, 2005; Pang *et al*, 2010; Yu *et al*, 2011; Sobhanian *et al*, 2010]. Accumulation of low-molecular weight organic osmolytes like sugars, proline, quaternary ammonium compounds such as glycine betaine and proteins like LEA help in adjusting the osmotic imbalance [Kant *et al*, 2006; Caruso *et al*, 2008; Liska *et al*, 2004; Katz *et al*, 2007; Mehta *et al*, 2009]. Another strategy to overcome the ionic imbalance through excluding the salt ions by increasing number of plasma membrane ion transporters (SOS1) and increasing the lignification of the xylem vessels to facilitate long distance transport [Wang *et al*, 2009]. This can also be done by intracellular transport of salt ions to vacuoles by increasing the number of tonoplast ion transporters like NHX, other ion transporters like H^+ - ATPase and FBP aldolase activity [Wakeel *et al*, 2011; Tada *et al*, 2009].

Water deficit stress is a situation in which plant water potential and turgor are reduced to the extent of interfering with normal cellular functions [Hsiao *et al*, 1973]. The moderate water stress condition leads to stomatal closure and reduced exchange of gases. Desiccation is a much more drastic loss of water which can potentially lead to great interruption of metabolism and cell structure that eventually leads to the cessation of enzymatic reactions [Smirnoff *et al*, 1993]. Water stress is indicated by

reduced water content, turgor, total water potential, wilting, closure of stomata, and reduced growth. Severe water stress may result in the arrest of photosynthesis, disturbed metabolism, and finally death of the plant [McKersie & Leshem, 1994]. Cellular water potential and turgor pressure were also controlled by the membrane permeability for water and ions. The water channels, known as aquaporins, are proteins that form a tunnel in the membrane that specifically transport water across the membrane. These channels aid water transport through osmotic or hydraulic driving force. Plasma membrane intrinsic protein (PIP) and tonoplast intrinsic protein (TIP) are water channels in the plasma membrane and tonoplast, respectively. Important proteins involved in water deficit condition are listed in Table 1.1.

Table 1.1: Proteins involved in countering salinity and water deficit stress in plants. The dependence or independence of ABA is shown in 3rd column.

Proteins involved	Role	Dependence on ABA
LEA1, 3,5,4	Salinity	Independent
Germin	Salinity	Independent
HVA 1	Salinity	Independent
Osmotin	Salinity, water deficit	Dependent
Stress associated proteins	Salinity, water deficit	Dependent
HZ-Zip	Salinity, water deficit	Dependent
LEA 2	Salinity, water deficit	Dependent
Dehydrins	Salinity, water deficit	Dependent
RAB protein	Salinity, water deficit	Dependent
D-11	Salinity, water deficit	Dependent
Nitrate reductase	Salinity	Independent
RUBP carboxylase	Salinity	Independent
Glutathione reductase	Salinity	Independent
Peroxidase	Salinity	Independent
Superoxide dismutase	Salinity	Independent
$Ca2^{+}$ ATPase	Salinity	Independent
Betaine aldehyde dehydrogenase	Salinity	Independent
Δ'-pyrroline-5-carboxylate synthetase	Salinity	Independent

Δ'-pyrroline-5-carboxylate reductase	Salinity	Independent
Vegetative storage proteins	Water deficit	Independent
Choloroplastic CDSP-32. CDSP-34	Water deficit	Independent
Glycosylated cell wall proteins	Water deficit	Independent
α- amylase	Water deficit	Independent
Sucrose synthase	Water deficit	Independent
Δ'-pyrroline-5-carboxylate synthetase	Water deficit	Independent
Protease	Water deficit	Independent
RUBP carboxylase	Water deficit	Independent
PEP carboxylase	Water deficit	Independent

1.6.2 Chilling and freezing

Cold stress, which includes chilling (< 20^o C) and/or freezing (< 0^o C) temperatures, adversely affect the growth and development of plants. It inhibits the expression of several enzymes involved in important metabolic pathways and through inhibition of water uptake (chilling), cellular dehydration (freezing), oxidative, and other stresses. Cold acclimation is a process by which plants acquire freezing tolerance upon prior exposure to low non-freezing temperatures. Plants like winter wheat, barley, oat, rye, rapeseed, etc which grow in winter have a vernalization requirement, which prevents early transition to the reproductive phase before the threat of freezing stress during winter has passed. Thus, vernalization helps plants to overcome cold stress as seedlings.

Cellular membranes have lipid bilayer structure causing fluidity, and cold temperatures can reduce their fluidity, causing increased rigidity. Plant cells can sense cold stress through low temperature-induced cell membrane rigidity, change in protein and nucleic acid conformation and/or metabolite concentration. In alfalfa and *Brassica rapa*, it has been shown that the membrane rigidification signal induces expression of *COLD RESPONSIVE* (*COR*) genes and result in cold acclimation (Orvar *et al*, 2000; Sangwan *et al*, 2001). The cold signal further leads to increased

flux of Ca^{+2} ions in the cytosol. Subsequently, calcium signal amplification and phospholipid signaling might be involved in cold-stress signaling [Vergnolle *et al*, 2005; Williams *et al*, 2005; Komatsu *et al*, 2007]. Moreover secondary signals like ABA and ROS can also induce Ca^{+2} ions flux affecting cold signaling. Other mutational studies in *Arabidopsis* showed that lack of activation of the molybdenum cofactor of abscisic aldehyde oxidase known as aba3/freezing sensitive 1 (*frs1*) [Llorente *et al*, 2000] or low expression of osmotically responsive genes 5 (*los5*) [Xiong *et al*, 2001], exhibit susceptibility to freezing stress. ROS accumulation is seen in cells affected by various abiotic stresses, and they appear to have a strong impact on regulation of cold responsive genes. *Fro1* (frostbite1) encodes the Fe–S subunit of complex I (NADH dehydrogenase) of the respiratory electron transfer chain in mitochondria, and its interruption leads to tremendous amount of ROS generation [Lee *et al*, 2002]. The *Arabidopsis fro1* mutant constitutively accumulates high levels of ROS which exhibits altered *COR* genes expression that ultimately leads to susceptibility to chilling and freezing. In *Arabidopsis* ICE1 (Inducer of CBF Expression 1) is a MYC-type basic helix-loop-helix transcription factor that bind to MYC elements in the *CBF3* promoter, which is critical for the expression of *CBF3* during cold acclimation. An altered *ice1* (mutant) expression result in defective cold induction of CBF1 followed by hypersensitivity to chilling stress and incapability towards cold acclimation. In *Arabidopsis* constitutive overexpression of *ICE1* enhances expression *CBF3*, *CBF2*, and *COR* genes during cold acclimation favoring increased freeze tolerance. Constitutive expression of *ICE1* was observed in nucleus but it induces expression of CBF only during cold stress. This conveys that post-translation modification (phosphorylation) of *ICE1* induced due to cold stress is critical for the activation of downstream genes in plants [Chinnusamy *et al*, 2003] (Figure 1.5). ICE1 is predicted to be a transcriptional inducer of *CBFs* (CBF1-CBF3), *ZAT12*, *NAC072* and the transcription factor *HOS9* in *Arabidopsis* [Benedict *et al*, 2006]. Antifreeze proteins (AFPs) or ice structuring proteins (ISPs) are synthesized by certain vertebrates, plants, fungi and bacteria that permit their survival in subzero environments. AFPs bind to small ice crystals to inhibit growth and recrystallization of ice that would otherwise be fatal for the organism [Dalal *et al*, 2001].

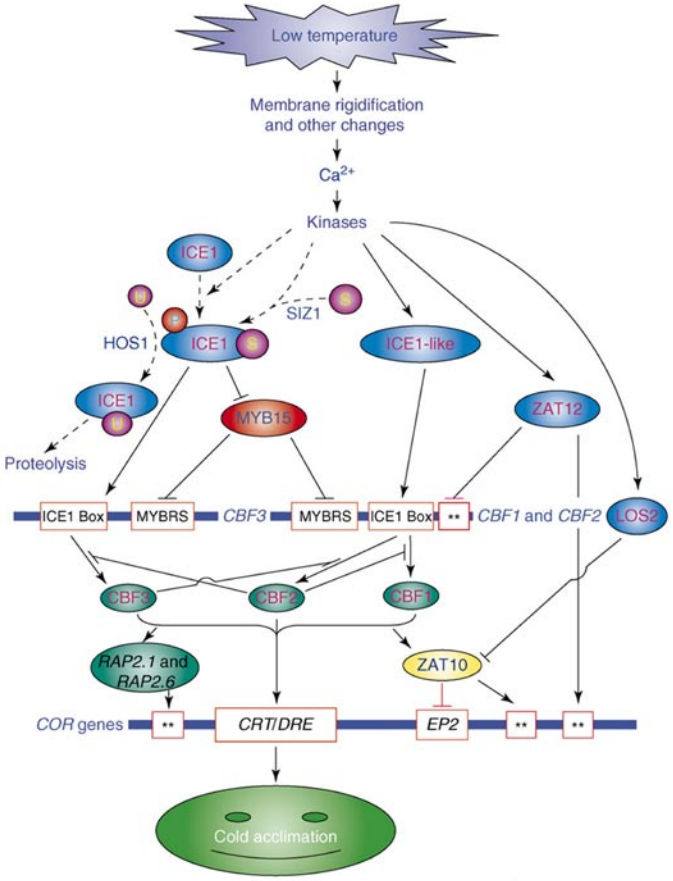


Figure 1.5: Diagram of cold-responsive transcriptional network in *Arabidopsis*. Plants probably sense low temperatures through membrane rigidification and/or other cellular changes, which might induce a calcium signature and activate protein kinases necessary for cold acclimation. Constitutively expressed ICE1 is activated by cold stress through sumoylation and phosphorylation. CBFs regulate the expression of COR genes that confer freezing tolerance. Figure is adopted from Chinnusamy *et al*, 2007.

1.6.3 Heat

Abiotic stress generally causes dysfunction of cellular macromolecules therefore maintaining them in their functional conformation and preventing the aggregation of non-native protein is vital for cell survival. Temperature above the normal optimum sensed by all living organism is known as heat stress. Heat shock alters the cellular homeostasis that causes severe retardation in growth and development that further leads to death. Higher temperature even damages the whole photosynthetic system affecting plant's productivity [Al-Khatib & Paulsen, 1990]. Suppression of photosynthetic rate is reversible at moderate temperatures which at higher temperatures can permanently damage the photosynthetic system [Berry & Bjorkman, 1980]. It also reduces the chlorophyll content, net photosynthetic rate due to defective supply of Ribulose-1, 5 biphosphate (RUBP) and conductance of stomata [Law & Crafts-Brandner, 1999]. Sudden rise in temperature causes denaturation of membrane proteins, increase in unsaturated fatty acids that leads to increased ion leakage and further loss of cellular functions [Savchenko *et al*, 2002]. Heat stress reduces the

fertility (problems in male meiosis, pollen germination, pollen tube growth, megagametophyte defects etc) [Dupuis & Dumas, 1990; Sakata *et al*, 2000; Sato *et al*, 2006; Abiko *et al*, 2005] and seed fill by effecting flower production, grain set, endosperm division, photosynthesis and assimilates transport [Prasad *et al*, 2006]. The coordination of heat shock proteins (HSPs) and heat stress transcription factors (HSFs) are assumed to play a key role in the heat stress response (HSR) that leads to thermotolerance in plants and other organisms. Moreover, HSFs play a crucial role in signal transduction pathway mediating the expression of HSPs and heat shock induced transcripts (Figure 1.6).

HSPs were first identified as those proteins that get induced upon exposure to heat shock. The five well characterized classes of HSP proteins include HSP100/ ClpB, HSP90/HtpG, HSP70/DnaK, HSP60/GroEL, and small HSP (sHSP). Although HSPs can act as molecular chaperones some are required for normal growth and development too [Nakamoto *et al*, 2007; Larkindale *et al*, 2005 A]. HSPs/ chaperones play a crucial role in protein folding, assembly, translocation, degradation of proteins and membrane stabilization, they also assist in protein refolding under stress conditions thus establishing cellular homeostasis. Furthermore, in plants, primarily they are known to be expressed during high temperature stress but also in response to other environmental assaults like water stress, salinity, osmotic, cold, and oxidative stress. The members of Hsp100/ClpB belong to AAA+ family ATPase involved in resolubilizing the protein aggregates. The sHSPs are low molecular mass chaperones (12-40 kDa) and have a conserved C-terminal domain of about 90 amino acids referred to as α-crystallin domain. Hsp90 is distinct from many other molecular chaperones as most of its substrates are signal transduction proteins such as steroid hormone receptors and signaling kinases [Young *et al*, 2001]. Along with the key role in signal-transduction other major functions of Hsp90 in assisting protein folding [Frydman, 2001; Buchner, 1999], cell-cycle control, protein degradation and protein trafficking [Young *et al*, 2001; Richter & Buchner, 2001; Pratt *et al*, 2001]. It might also play a role in morphological evolution and stress adaptation in *Drosophila* and *Arabidopsis* [Rutherford & Lindquist, 1998; Queitsch *et al*, 2002]. Hsp90 is one of the major species of molecular chaperones that requires ATP for its function. To perform its function, Hsp90 together with Hsp70 co-operates with co-chaperones, including Hip (Hsp70 interacting protein), Hop (Hsp70/ Hsp90 organizing protein), p23 and

Hsp40 (a DnaJ homolog), the immunophilins FKBP51/54 and FKBP52, and Cdc37/p50. Hsp70, together with their co-chaperones (e.g. DnaJ/Hsp40, Dnak and GrpE), assist with a wide range of protein folding processes in almost all cellular compartments. Hsp70 prevents aggregation and assist in refolding of non-native proteins under both normal and stress conditions [Frydman, 2001; Hartl, 1996]. Furthermore, they are also involved in protein import and translocation processes, and facilitate proteolysis of unstable proteins by targeting the proteins to lysosomes or proteasomes [Hartl, 1996]. Another member of Hsp70 known as Hsc 70 (70-kDa heat-shock cognate) shows constitutive expression and assist in the folding of *de novo* synthesized polypeptides and translocation of precursor proteins. Hsp60/ Chaperonins (homologous to *E. coli* GroEL) are a class of molecular chaperones present in prokaryotes and in the mitochondria and plastids of eukaryotes [Hartl, 1996; Boston, 1996]. Chaperonins are classified into two subfamilies namely GroE chaperonins (Group I) found in bacteria, mitochondria and chloroplasts (e.g. GroE and chCpn60) and CCT chaperonins (Group II) found in Archaea (e.g. trigger factor 55 and the thermosomes) and in the cytosol of eukaryotes (e.g. the TCP-1 ring complex TriC) [Ranson *et al*, 1998]. Chaperonins play a vital role in assisting wide range of newly synthesized and newly translocated proteins to achieve their native forms [Bukau & Horwich, 1998; Frydman, 2001].

Heat shock signal is often associated with some degree of oxidative stress. Exposure to short periods of high temperature causes a burst of H_2O_2 [Vacca *et al*, 2004]. A strong correlation exists between oxidative burst and induction of heat shock responsive genes, a process assumed to be mediated through perceiving of H_2O_2 by HSFs [Miller & Mittler, 2006; Volkov *et al*, 2006]. Previous studies have hinted at the probable role of Ca^{+2} in heat shock responsive signaling. Heat shock induced cytosolic Ca^{+2} linked through calmodulin (CAM) were measured in *Arabidopsis* and wheat [Liu *et al*, 2005]. Moreover research has shown the role of inositol-1,4,5,-triphosphate (IP3) upstream of Ca^{+2} in the heat shock response. In addition to that, phytohormones like ABA, SA, and ET are also known to be a part of heat shock signaling in different plant species [Larkindale *et al*, 2005 A; Larkindale & Huang, 2004; Larkindale *et al*, 2005]. A sharp peak of ABA level was observed in pea during heat shock response [Liu *et al*, 2006]. Even in *Arabidopsis* ABA signaling mutants abscisic acid insensitive 1 (abi1) and abi2 showed reduced survival after heat shock stress

[Larkindale *et al*, 2005 A]. During heat shock stress, an elevated level of SA was observed in many plant species [Larkindale & Huang, 2004; Liu *et al*, 2006; Clarke *et al*, 2004]. It was proposed that SA acts downstream of ABA and upstream of phosphotidyl-inositol-4, 5- bisphosphate specific lipase C (PLC) [Liu *et al*, 2006]. Mutational studies showed that ethylene resistant 1 (etr1) and ethylene insensitive 2 (ein2) are susceptible to heat shock but normally attain thermotolerance. Moreover there is no evidence for the involvement of ET in heat shock signaling [Larkindale *et al*, 2005 B].

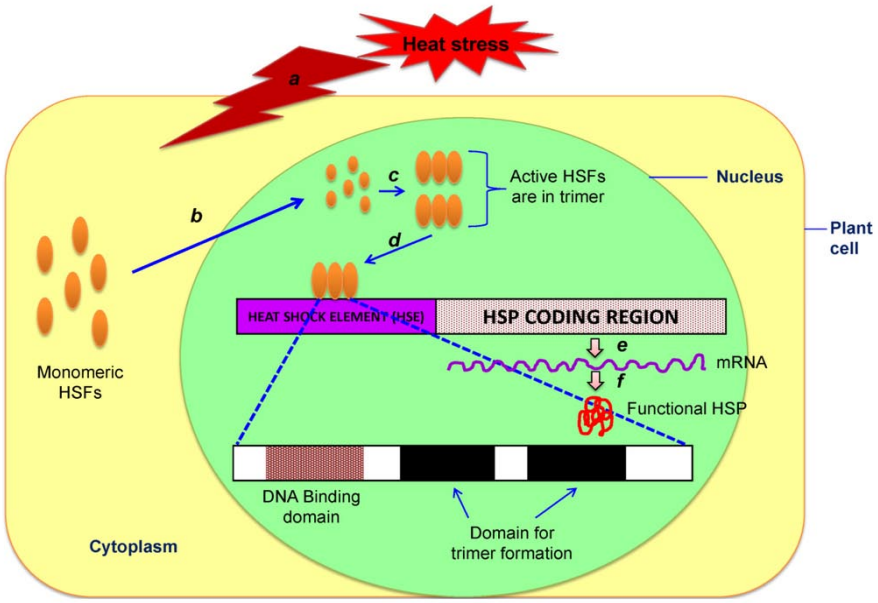


Figure 1.6: Schematic diagram showing the molecular regulatory mechanism of heat shock proteins based on a hypothetical cellular model. Upon heat stress perceived by the plant cell, (a) monomeric heat shock factors (HSFs) are entering into the nucleus; (b) from the cytoplasm. In the nucleus, HSF monomers form active trimers; (c) that will bind; (d) to the specific genomic region (promoter or heat shock element, HSE) of the respective heat shock gene (HSG). Molecular dissection of the HSF binding region of HSE showing that it is consists of one DNA binding domain and two domains for trimerization of HSFs. Successful transcription (e) translation and post-translational modification; (f) lead to produce functional HSP to protect the plant cell and responsible for heat stress tolerance. Figure is adopted from Hasanuzzaman *et al*, 2013.

1.6.4 Anaerobiosis

Anaerobiosis is caused by excessive amount of water in soil and during flooding. Plant metabolism is badly affected by low oxygen concentration in roots. However, plants acclimate to anaerobiosis condition by switching from oxidative to fermentative carbohydrate metabolism [Perata & Alpi, 1993]. Certain morphological

adaptations like formation of aerenchyma, adventitous roots or rapid shoot elongation are escape responses [Steffens & Sauter, 2010]. During anaerobic stress, normal protein synthesis is suppressed in plants due to loss of polysomes [Bailey-Serres & Freeling, 1990]. Altered genes expression leads to the synthesis of specific polypeptides known as transition polypeptides (TPs) and anaerobic polypeptides (ANPs) [Sachs *et al*, 1996]. Suppression of already existing proteins and synthesis of new proteins seems to be immediate response of anaerobiosis. There appear certain differences in the translation time and stability of TPs and ANPs. TPs are synthesized in first five hours of anoxia and remain stable, lasting long after synthesis declines. On the contrary anaerobic polypeptides are translated after 90 minutes of anoxia and last for several days and until death [Sachs *et al*, 1996]. They play a major role in maintaining ATP levels in the cells. Most of these enzymes namely alcohol dehydrogenase (ADH), aldolase, lactate dehydrogenase (LDH), enolase, glucose-phosphate isomerase, glyceraldehydes-3-phosphate dehydrogenase, pyruvate decarboxylase, and sucrose synthase are involved in glycolysis [Subbaiah & Sachs, 2003]. Out of these 8 enzymes, aldolase is well studied, whose maximum expression was seen during anoxia condition. However, there exist two ADH genes in *Arabidopsis*, out of which one ADH gene is expressed in roots during low O_2 stress while constitutive expression of another one was observed in roots and leaves.

Studies in maize have shown an increased activity of LDH after several days of hypoxia. A very interesting observation was seen in maize. After treating maize seedlings with ABA, their tolerance level against anaerobic conditions was elevated [Hwang & VanToai, 1991]. Ricard *et al*. (1991) observed increased synthesis of sucrose synthase in rice seedlings subjected to anaerobiosis. Synthesis of three more enzymes increased in response to anaerobic stress. These are 1-aminocarboxylase-1-cyclopropane synthase (ACC synthase) which catalyzes ET synthesis, xyloglucan endotransglycosylase responsible for the formation of aerenchyma during water logging and superoxide dismutase [Sachs *et al*, 1996; Artlip & Funkhouser, 1995; Dubey, 1999]. Microarray studies showed coordination among 50 genes during oxygen deficiency in several plants involved in fermentation, alanine metabolism, scavenging, and detoxification of ROS and reactive nitrogen species (RNS) and a huge number proteins with unknown function [Licausi *et al*, 2010]. Transcriptome

analysis in *Arabidopsis*, rice, and poplar revealed high expression of transcription factors important for the induction of anaerobic genes (not strictly required during hypoxia). Ethylene response factors (ERFs) are one of the major transcription factor induced during anaerobic stress that leads to adaptation towards flooding and water logging. ET plays a key role in anaerobiosis stress signaling and G-proteins are the signal transducers in several hypoxic signaling [Steffens & Sauter, 2010]. Major steps have been taken to develop tolerant varieties through the characterization of two multigenic loci that raise the endurance for complete submergence (SUBMERGENCE 1, SUB1) or rapid outgrowth of adverse partial submergence (SNORKEL, SK) [Bailey-Serres & Voesenek, 2010].

1.6.5 Heavy metals

Industrialization has caused change in the mineral content of the soil by introducing heavy metals like cadmium (Cd), lead (Pb), zinc (Zn), copper (Cu), and mercury (Hg). This increase in levels of heavy metals in soil alters plant growth as well as induction or inhibition of enzymes. It induces synthesis of metal binding cysteine rich polypeptides called phytochelatins. They have a structure like $(\gamma\text{-Glu-Cys})_n\text{-Gly}$ or $(\gamma\text{-Glu-Cys})_n\text{-}\beta\text{-Ala}$ where n=2-11 that sequester the metal ions within the plant. This situation is often associated with increase in free radicals and ROS. ROS causes decrease in photosynthesis and respiration rate and disturbs the cell metabolism. Moreover another cytotoxic compound methyl glyoxal (MG) was found to increase in heavy metal toxicity stress response. The increase in ROS species, MG and decrement of glutathione (GTH) concentration induces the oxidative stress which causes alteration of cell membrane, DNA damage, gene mutation, protein oxidation, and lipid peroxidation, inhibits the plant growth and reduces the yield that ultimately leads to cell death [Hossain *et al*, 2011]. Cd, one of the major pollutants, inhibits enzymes by interacting with SH group of the enzyme [Shah & Dubey, 1995]. In rice seedlings, heavy metals inhibit ribonuclease and acid phosphatase. Apart from phytochelatins, some more proteins were expressed under heavy metal stress (Table 1.2).

Table 1.2: Proteins synthesized upon treatment by heavy metals in some plant species are enlisted below.

Metal	Complex	Source	Reference
Cd	18 kDa Cd binding protein	Rice	Shah K & Dubey RS, 1998
Pb, Cu, NO_2^-	16 kDa polypeptide	Lupin	Przymusiński *et al*, 1995
$HgCl_2$	Glycine rich proteins, PR proteins, Chaperones & membrane protein	Maize	Didierjean *et al*, 1996

1.6.6 Gaseous pollutants

Ozone, sulphur dioxide, and nitric dioxide are the major air pollutants effecting the environment unfavorable for living organisms. Their presence in the surroundings generates ROS species which inhibit synthesis of many proteins and induce the activity of some antioxidant enzymes [Rao *et al*, 1996]. Ozone decreases the mRNA levels of gene encoding important enzymes like Rubisco, chlorophyll a/b- binding protein and a 10 kDa protein of the water-evolving complex of photosystem II [Miller *et al*, 1999]. In potato, ozone promotes senescence by inhibiting the synthesis of Rubisco small-subunit mRNA and decreasing the transcript synthesis of chloroplast proteins like glyceraldehyde-3-phosphate dehydrogenase [Glick *et al*, 1995]. Increased ozone concentration in the environment accelerates the activity of antioxidant enzymes; the most important one is cytosolic Cu: Zn superoxide dismutase. Exposure of O_3 enhances the activity of catalase, superoxide dismutase, peroxidase, glutathione reductase, and ascorbate peroxidase [Gillespie *et al*, 2011; Rao *et al*, 1996]. Moreover, a change in substrate binding affinity of glutathione reductase and ascorbate peroxidase was also seen [Rao *et al*, 1996].

1.6.7 UV radiations

Another devastating stress encountered by living organisms is the exposure to harmful UV-B radiations (280-320 mm) due to the depletion of stratospheric ozone layer. The main cause of ozone layer depletion is the increased concentration of chlorine from industrially produced chlorofluorocarbons (CFCs), halons, and selected solvents. Exposure of UV-B radiations inhibits the plant growth and protein synthesis, induces

the activity of peroxidase-related enzymes and enzymes of flavonoid-biosynthetic pathway. Moreover, it also disturbs the protein biosynthesis in leaves. However, the main target of UV-B radiation is chloroplast. The early responses to exposure includes decline in the synthesis of mRNA transcripts for the photosynthetic complexes and chloroplast proteins [Strid *et al*, 1994]. Previous studies have shown inhibition or induction of mRNA or protein synthesis upon exposure of UV-B radiations (Table 1.3).

Table 1.3: Proteins synthesized upon exposure of UV-B radiation in some plant species are enlisted below.

Plants	Effect of UV-B exposure	Reference
Pea	Inhibit protein synthesis, reduction in mRNA transcripts for the chlorophyll a/b-binding protein	Mackerness *et al*, 1998
Arabidopsis	Increased guaicol peroxidases, ascorbate peroxidase	Rao *et al*, 1996
Sunflower	Increased PR3 ad PR5	Jung *et al*, 1995
Maize	Increased membrane channel proteins and PR proteins	
Barley	Accumulation of an atypical transcript encoding a 42.3 kDa polypeptide similar to O-methyltransferase	Gregersen *et al*, 1994

1.6.8 Wounding

Plant injuries caused by mechanical damage activate a set of wound-responsive genes which helps in healing and prevention of subsequent pathogen and pest attack [Cabello *et al*, 1994; Schaller *et al*, 1996]. It increases the activity of many enzymes and proteins related to phenylpropanoid pathway, peroxidase, DHAP synthetase, glycine rich and hydroxyl-proline rich cell wall protein, protease inhibitor and 1-aminocyclopropane-1-carboxylate synthase. Moreover, enzymes performing lignification also get induced to form a wound periderm to restrict pathogenic attack. Upon wounding, expression levels of chitinase and glucanase enzymes increase in roots and stems of chickpea [Cabello *et al*, 1994]. In tobacco, a glycine rich (16 kDa) polypeptide present in cell wall is induced during wounding [Yasuda *et al*, 1997]. Furthermore several systemic wound-response proteins (swarps) have been characterized in tomato [Schaller *et al*, 1996]. Mehta *et al*, (1991) reported the synthesis of several unique proteins of molecular weights 80.0, 63.0, 33.0, 28.5, 25.5,

and 29 kDa induced during wounding response in tomato fruit tissues. These research studies propose differential protein synthesis and altered gene expression in tissues during wounding stress. The plant hormone jasmonate plays an important role in triggering the wound-induced adaptive response in plants [Koo & Howe, 2009].

1.6.9 Plant pathogenesis

Plants are most often exploited as a source of food and shelter by a diverse range of parasites like fungi, bacteria, viruses, nematodes, insects, herbivores, and even by other plants. Exposure to abiotic and biotic stresses induces expression of a set of genes encoding different proteins that activate several biochemical and physiological changes in plants such as enhancing the cell wall strength by lignifications, suberization, and callose deposition; synthesis of pathogenesis-related (PR) proteins and phytoalexins which aid in pathogen or pest attack [Bowles, 1990]. The role of PR proteins against pathogen attack or stress condition is very important. Phytoalexins are mainly synthesized by healthy tissues in the vicinity of damaged and necrotic cells. On the contrary, PR proteins are produced by infected and surrounding tissues and also by the far-flung uninfected tissues. PR protein production by the uninfected tissues helps to prevent the spread of infection any further [Ryals *et al*, 1996; Delaney, 1997]. Moreover necrosis and chlorosis also induce the synthesis of PR proteins [van Loon, 1999]. The signal transduction includes the involvement of cytosolic Ca^{2+} and H^{+} ions, ROS, jasmonate, SA, and ET that provoke the defense mechanism.

PR proteins can be acidic or basic based on the isoelectric points, although they perform similar function. Presently the PR proteins are categorized into 17 families based on their properties and function (Table 1.4) [van Loon, 1999; van Loon *et al*, 2006]. Out of all, the two most important hydrolytic enzymes are chitinases and β-1, 3-glucanases found abundant in several plants after pathogen infection. Their synthesis shoots up in plants after fungal attack because chitin and β-1, 3-glucan are the major structural components of cell wall of many pathogenic fungi. Interestingly research has shown a synergistic response of the two enzymes towards the pathogen attack. A coordinated expression of β-1, 3-glucanases and chitinase was observed during fungal infection. Moreover this synergism and co-induction of the hydrolytic

enzymes was seen in several plant species such as pea, bean, tomato, tobacco, maize, soyabean, potato, and wheat. Transgenic plants have been developed by transferring chitinases genes alone or together with β-glucanase and after expression the effect on the pathogen resistance capacity was studied. Most of the transgenic variety generated have revealed high degree of tolerance level against fungal diseases or have deferred symptom development as compared to control plants. However, research has shown the absence or less resistance against certain pathogens in plants transformed with chitinase or β-1, 3-glucanases alone as compared to transformed plants with both chitinases and β-1, 3-glucanases.

Along with the PR proteins, other disease resistance genes like nucleotide binding site-leucine rich repeat (NBS-LRR) genes also play a critical role in plant defense mechanisms. The signal transduction pathway begins with the alteration of plasma membrane permeability. The loss of membrane permeability results in increased influx of Ca^{2+} and H^{+} ions and efflux of K^{+} and Cl^{-} ions. The GTP binding proteins and protein phosphorylation/ dephosphorylation events are presumed to be involved in transferring signals from membrane receptor to calcium channels. Ion fluxes lead to production of ROS intermediates like NO, O_2^{-}, H_2O_2, $OH^{\cdot}$ (hydroxyl free radical) play a role as secondary messenger. The ROS generation causes induction of hypersensitive response that induces the expression of defense genes. Most of the defense related genes are regulated by signal transduction pathways involving at least one or more regulators like jasmonate, ET, and SA.

Table 1.4: Recognized families of pathogenesis-related proteins (adopted from van Loon *et al*, 2006).

Families	Type member	Properties	Gene symbols
PR-1	Tobacco PR-1a	Unknown	*Ypr1*
PR-2	Tobacco PR-2	β-1,3-glucanase	*Ypr2*, [*Gns2* (‘*Glb*’)]
PR-3	Tobacco P, Q	Chitinase type I, II, IV, V, VI, VII	*Ypr3*, *Chia*
PR-4	Tobacco ‘R’	Chitinase type I, II	*Ypr4*, *Chid*
PR-5	Tobacco S	Thaumatin-like	*Ypr5*

PR-6	Tomato Inhibitor I	Proteinase-inhibitor	*Ypr6*, *Pis* (‘*Pin*’)
PR-7	Tomato P69	Endoproteinase	*Ypr7*
PR-8	Cucumber chitinase	Chitinase type III	*Ypr8*, *Chib*
PR-9	Tobacco ‘lignin forming peroxidase’	Peroxidase	*Ypr9*, *Prx*
PR-10	Parsley ‘PR1’	Ribonuclease like	*Ypr10*
PR-11	Tobacco ‘class V’ chitinase	Chitinase, type I	*Ypr11*, *Chic*
PR-12	Radish Rs-AFP3	Defensin	*Ypr12*
PR-13	*Arabidopsis* THI2.1	Thionin	*Ypr13*, *Thi*
PR-14	Barley LTP4	Lipid-transfer protein	*Ypr14*, *Ltp*
PR-15	Barley OxOa (germin)	Oxalate oxidase	*Ypr15*
PR-16	Barley OxOLP	Oxalate oxidase-like	*Ypr16*
PR-17	Tobacco PRp27	Unknown	*Ypr17*

1.7 Plant hormones in countering stress

Phytohormones play a major role towards regulating the developmental and growth processes in plants and the signaling network required to activate the plant responses against various biotic and abiotic stress conditions. Plants produce a wide range of hormones, the best known group comprises of auxin (IAA), cytokinin (CK), gibberellic acid (GA), ABA, JA, ET, SA, brassinosteroids (BR), nitric oxide (NO), polyamines, peptide hormones, and strigolactone [Gomez-Roldan *et al*, 2008; Umehara *et al*, 2008]. They produce systemic signals that can transmit information over large distances (Figure 1.7). For example ABA can be transported and perform physiological roles at location far away from where it is synthesized [Sauter *et al*, 2001].

The transcriptional regulation of various stress genes by phytohormones under abiotic stresses involves various *cis*- or *trans*-acting elements. Some of the transcription factors regulated by phytohormones are ARF, AREB/ABF, DREB, MYC/MYB, NAC, and WRKY. They often alter gene expression by inducing or

dissuading the degradation of transcriptional regulators via the ubiquitin–proteasome system [Santner & Estelle, 2010]. The ability of plants to overcome wide range of environmental stresses is also finely balanced through the interaction between plant hormones and the redox signaling hub. Phytohormones generate ROS as secondary messengers in signaling cascades that transmit signals concerning changes in hormone concentrations from a single cell to the whole plant to mediate a whole range of adaptive response [Bartoli *et al*, 2012]. For example, BRs can induce plant tolerance to diverse abiotic stresses by triggering H_2O_2 generation in cucumber leaves [Cui *et al*, 2011].

The synthesis of ABA is one of the fastest responses to abiotic stress for plants. The plants under water stress synthesize ABA that triggers ABA-inducible gene expression leading to stomatal closure, thereby lowering the water loss through transpiration, and consequently, a reduced growth rate [Schroeder *et al*, 2001]. In addition to that, ABA also plays a vital role in adapting to cold temperatures. Low temperature or freezing triggers the production of ABA and even the exogenous application of ABA increases the cold tolerance of plants. ABA activates the expression of many stress-responsive genes, which makes it the most studied stress-responsive hormone.

Other plant hormones, like CK, SA, ET, and JA, also have substantial direct or indirect roles in abiotic stress responses. Being an antagonist to ABA, the level of CK usually decreases under conditions of water shortage. But transgenic tomato rootstocks expressing isopentenyl transferase gene (*IPT*, a gene encoding a key step in CK biosynthesis) with improved root CK synthesis have shown raised salinity stress tolerance [Ghanem *et al*, 2011].

As discussed above, ABA regulates stomatal actions under stress conditions. However, along with ABA; CK, ET, BR, JA, SA, and NO also regulate stomatal function [Acharya & Assmann, 2009]. The stomatal closure is induced by ABA, BR, SA, JA, and NO while CK and IAA promote stomatal opening. The phytohormone NO acts as a key intermediate in the ABA-mediated signaling network in stomatal closure. Moreover ABA regulates the BR-mediated signaling, and in turn, ABA was also shown to inhibit BR-induced responses under abiotic stress [Divi *et al*, 2010]. In

plants, the synthesis of similar set of above mentioned phytohormones is also triggered when countering an attack by pest and pathogens. The combined action of phytohormones results in synergistic or antagonist interactions, which is crucial for plants in abiotic and biotic stress responses.

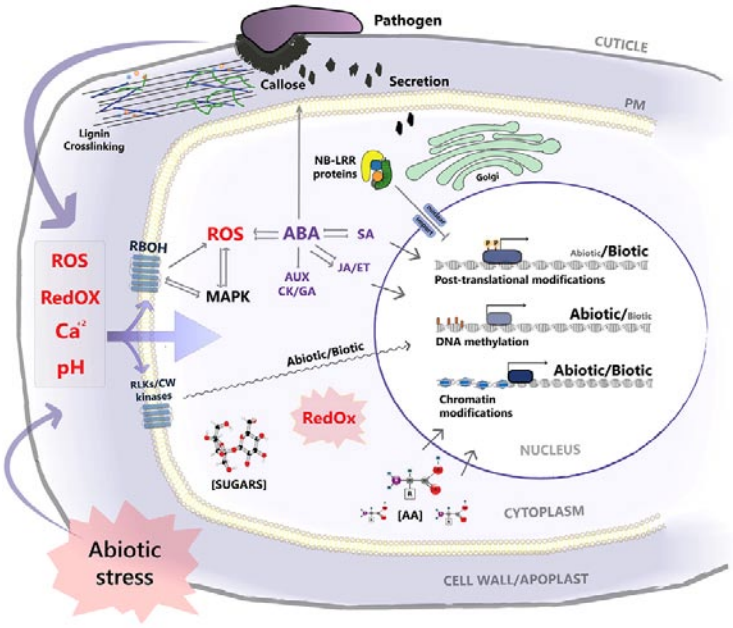


Figure 1.7: A scheme showing the interaction interface and overlapping signaling pathways of abiotic and biotic stress at the cellular level. The figure is adopted from Kissoudis *et al*, 2014.

1.8 Countering stress by chickpea

Crop breeding techniques and chemical control measures have been useful to create improved chickpea varieties that are resistant to diseases (*Ascochyta* blight and *Fusarium* wilt) and controlling crop losses due to insect pests to some extent. The main hurdle in developing the tolerant variety are *Helicoverpa*, aphids, bruchids, weeds, drought, salinity, and low methionine content in the seeds because breeding for these traits is difficult to perform due to cross incompatibility and a lack of resistant sources of germplasm. Therefore, there is an urgent need to improve such traits. Approaches to the generation of transgenic crops using gene technology to incorporate insect resistance, virus resistance, herbicide resistance and increased nutritional value have been proven suitable for many cultivated crops. Approaches to the generation of transgenic crops using gene technology to incorporate insect resistance, virus resistance, herbicide resistance and increased nutritional value have been proven suitable for many cultivated crops. In addition, modern genetic

transformation procedures offer the advantage of "stacking" or "pyramiding" multiple genes to add a single trait or multiple traits to the crops. The commercialization of stacked GM crops has increased significantly from 8% in 2003 to 25% in 2011 [James, 2010]. Several genes have been identified and used to control biotic and abiotic stresses, as well as to enhance sulfur containing amino acid content in chickpea and in other crops [Acharjee *et al*, 2013]. Several of these genes could be used for the generation of pyramided chickpeas because the incorporation of gene(s) for a single trait into chickpeas may not provide a significant yield advantage, due to the number of impediments to crop production in the field. Therefore, the generation of transgenic chickpeas with multiple genes appears to be an ideal approach to overcome both biotic and abiotic stresses, to reduce the cost per trait and to increase acceptance by the farmers.

1.9 Structure-function of proteins involved in stress

As discussed in the previous sections, transcription of a wide range of stress genes in plants gets triggered by sudden change in the physical, chemical, or biotic content of the habitat.

1.9.1 Glycosyltransferases

Glycosyltransferase (GTs; EC 2.4.x.y) is one of the important classes of stress genes involved in the glycosylation reaction by transferring sugar moiety from an activated nucleotide sugar donor to specific sugar acceptors, forming a glycosidic bond. Inspite of the wide distribution of these enzymes in plant kingdom, their biological role in abiotic and biotic stress conditions is largely unknown. Recently, several studies have thrown light over the important part played by them during dreadful stress conditions. Sun *et al*, (2013) identified a novel glycosyltransferase gene *UGT85A5* in *Arabidopsis* expressing in response to salt stress. In tobacco, the ectopic expression of *UGT85A5* enhanced the salt stress tolerance in the transgenic plants. It further increases the seed germination rates, improves the plant growth and less chlorophyll loss in transgenic lines compared to wild type plants under salt stress. Chaturvedi *et al*, (2012) recognized three members of sterol glycosyltrasferases (SGTs) gene family in *Withania somnifera*. Their possible role in defense mechanism was identified when a 10 fold increase in the expression of *sgt* genes was observed upon treatment with

SA and methyl jasmonate. In 2005, Langlois-Meurinne *et al*, identified a specific group of UDP-glycosyltransferase i.e. group D in *Arabidopsis* which are involved in stress-inducible responses in other plant species. They analyzed the expression profiles of this group in *A. thaliana* after exposing with *Pseudomonas syringae* pv tomato or after treatment with SA, methyljasmonate, and hydrogen peroxide. Their analysis showed distinct induction profiles, indicating their potential role in stress or defense responses especially for *UGT73B3* and *UGT73B5*. The above findings indicate their role in providing resistance against various abiotic and biotic stress.

The three-dimensional structure of GTs consists primarily of α/β/α sandwich structure which resembles the Rossmann-type fold, a unique structural motif found in many nucleotide-binding proteins. The structure of Rossmann-type fold is composed of six parallel beta strands connected to two pairs of alpha helices in an alternate arrangement of β-α-β-α-β. Until recently, only two structural superfamilies of GTs are known *viz.* GT-A and GT-B. The GT-A fold consists of a single Rossmann fold with a characteristic signature motif, the DxD motif, in the conserved N-terminal region and also requires a divalent cation for the activity [Breton *et al,* 1998] (Figure 1.8 A). This important signature motif is involved in the binding of sugar donor substrate by interacting primarily with the phosphate group of nucleotide donor through the coordination of a divalent cation. Contrary to this, the C-terminal region is highly variable and binds specific sugar acceptors to be glycosylated.

The GT-B fold consists of two Rossmann domains interconnected by a linker region packed in a compact manner to create the binding site between the two domains (Figure 1.8 B). A high degree of structural conservation exists between the GT-B family members, especially at the C-terminal domain which constitute the sugar donor binding domain. Moreover the UDP-glycosyltransferase (UGT) class of GT-B enzymes possesses a conserved signature sequence "PSPG" of 44 amino acid length plant secondary product glycosyltransferase [Masada *et al*, 2007] or putative plant secondary glycosyltransferase [Wang, 2009]. The conserved amino acids of this motif are involved in hydrogen bond interactions with the nucleotide sugar donor. However, variations are more pronounced in the N-terminal domain, particularly in the loops and helices of the active site to accommodate a diverse array of acceptor substrate. Plant UGTs catalyzes the transfer of glucose moiety from a sugar donor by following

a direct displacement S_N2 mechanism. At the N-terminal domain, a conserved histidine is observed in the active site that acts a catalytic base to deprotonate the acceptor substrate. A nearby conserved aspartic acid interacts with the histidine by forming hydrogen bond and balances its charge after deprotonating the acceptor. Subsequently the deprotonated acceptor attacks the C1 carbon atom of the sugar moiety of UDP-sugar resulting in direct displacement of the UDP moiety and forms the respective β-glucosidic linkage product [Wang, 2009]. Till now, 26 crystal structures of UGTs were deposited in the protein databank. Few structures from the same source like *A. thaliana* (3), *V. vinifera* (3) and *M. truncatula* (5) were determined at different resolutions. About 10 crystal structures were from bacteria and 3 from *Saccharomyces cerevisiae*. One crystal structure belonged to *Homo sapiens* (PDB-ID: 2O6L). Recently Hiromoto *et al*, (2013) reported the crystal structure of anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea*.

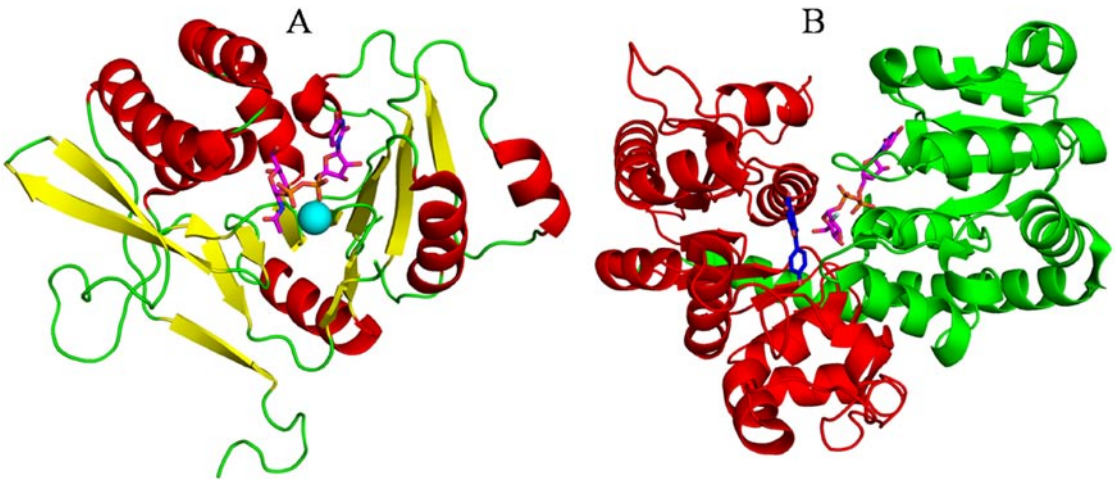


Figure 1.8: Three-dimensional structure of GT-A (A) GT-B (B) proteins. In panel B, the two Rossmann domains are shown in red and green color. The sugar donor (magenta) and acceptor (blue) are shown in stick form.

1.9.2 NBS-LRR resistance proteins

The NBS-LRR (NL) gene family is one of the largest in plant genomes that account for the largest number of known disease resistance genes. This family can be subdivided into functionally distinct TIR-domain-containing (TNL) and CC-domain-containing (CNL) subfamilies. Their precise role in pathogen recognition is still unknown; however, they are thought to monitor any attack by pathogens or the

presence of effector molecules released by pathogens within the plant cell. No crystal structure of the full protein with all three domains has been reported. However, the three-dimensional structures of individual domains are available.

The structure of the barley mildew A (MLA; PDB ID: 3QFL) revealed a distinct coiled-coil (CC) domain recognizing a distinct effector of the pathogenic powdery mildew fungus, *Blumeria graminis* f. sp. *hordei* [Maekawa *et al*, 2011]. The structure of the CC domain is mainly α-helical and contains two long antiparallel α-helices connected by a short linker loop (Figure 1.9 A), thereby forming a helix-loop-helix structure. The hydrophobic residues predominantly occupy the interior of the helices, therefore the two α-helices pack loosely against each other. As a result, the two α-helices are slightly apart connected by only marginal contacts between the N-terminal portion of α1a and the N-terminal portion of α2b. However, such interactions seem to be insufficient to stabilize the two seemingly independent α-helices, suggesting involvement of other protein-protein interactions. The studies have shown the presence of a strong dimer interface (Figure 1.9 A).

The crystal structure of TIR domain from *Arabidopsis* was determined and deposited in the database under PDB-ID: 4C6R (Figure 1.9 B) [Williams *et al*, 2014]. Studies have shown that plant and animal NBS-LRR receptors (NLRs) function in pairs to mediate immune recognition [Eitas *et al*, 2010]. Both RPS4 (resistance to *Pseudomonas syringae* 4) and RRS1 (resistance to *Ralstonia solanacearum* 1) NLRs are required in *Arabidopsis* to recognize bacterial effectors AvrRps4 (from *P. syringae* pv. pisi) and PopP2 (from *R. Solanacearum*) and also the fungal pathogen *Colletotrichum higginsianum* [Birker *et al*, 2009]. Cooperative activity is commonly seen between the immune receptor pairs in both plants and animals which might function by evolutionarily conserved mechanisms [von Moltke *et al*, 2013]. Upon infection, AvrRps4 is processed in the plant cell, and its C-terminal domain triggers RRS1/RPS4-dependent immunity response [Sohn *et al*, 2009]. RPS4 and RRS1 both carry a Toll–interleukin- 1 receptor/resistance protein (TIR) domain at their N-termini. Homo and heterotypic interactions between TIR domains are seen in Toll-like receptor signaling pathways in animals, mediating interactions between Toll-like receptors and intracellular TIR domain–containing adaptors to regulate immune signaling [Takeda *et al*, 2005].

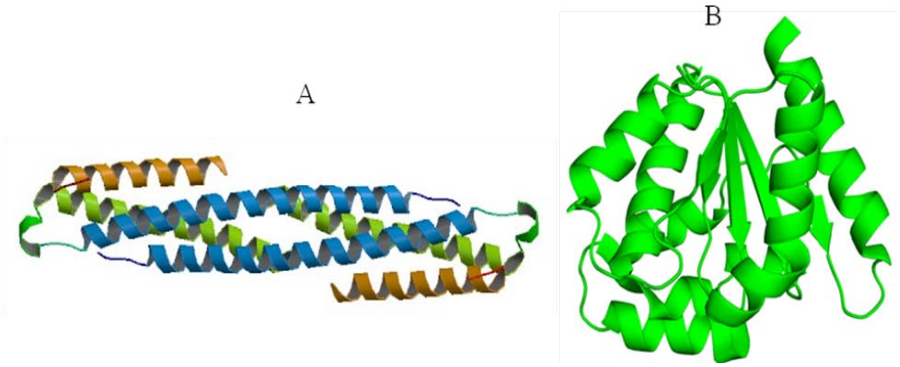


Figure 1.9: Three-dimensional structure of CC domain of *Hordeum vulgare* (PDB-ID: 3QFL) (A) TIR domain of RPS4 protein of *Arabidopsis* (PDB-ID: 4C6R).

Polygalacturonase–inhibiting proteins (PGIP2) are members of the leucine-rich repeat (LRR) protein family that play important roles in development, defense against pathogens, and recognition of beneficial microbes. The three-dimension structure revealed a typical curved and elongated shape; however, its scaffold appears more twisted than other LRR proteins (PDB-ID: 1OGQ) [Marino *et al*, 1999]. The central LRR domain, folded in a right-handed superhelix, comprises of a set of 10 tandem repeats (Figure 1.10).

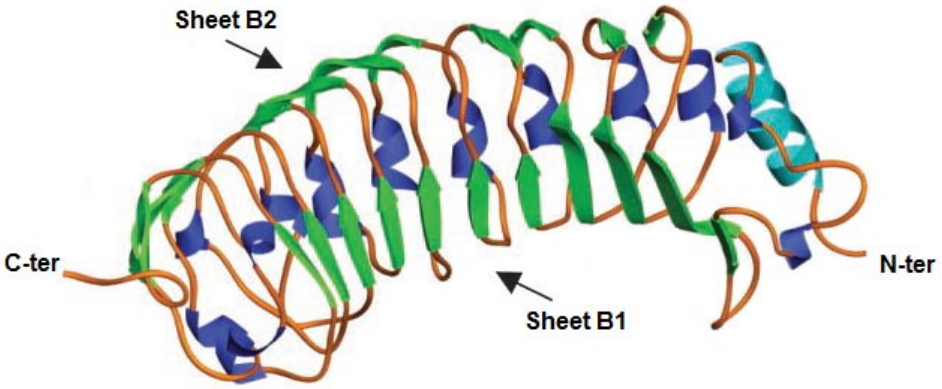


Figure 1.10: Three-dimensional structure of polygalacturonase inhibiting protein, a leucine rich protein of *Phaseolus vulgaris* (PDB-ID: 1OGQ).

1.9.3 Proteases

Along with the main function in the hydrolysis of non-functional proteins in the cell, proteases also take part in several biological processes like recognition of pathogens and pests and the induction of effective defense responses, PCD, water deficit etc. [Van der Hoorn & Jones, 2004]. Proteases are classified into six broad classes, out of which we have studied four in chickpea that are listed below.

1.9.3.1 Aspartate proteases

Till date only two crystal structures of aspartate proteases have been reported in plant kingdom. The sources of the two structures are *Hordeum vulgare* (Barley) (PDB-ID: 1QDM) and *Cynara cardunculus* (Cardoon) (1B5F) (Figure 1.11). The resolutions of the two crystal structures are 2.30 Å and 1.72 Å. The active site cleft of all APs is divided into two β-barrel domains and possesses two catalytic aspartic acid residues (numbered 32 and 215 in pepsin), require acidic pH optima for the catalytic activity, inhibition by pepstatin and preferentially cleavage of peptide bond between bulky hydrophobic side chains. The aspartic proteinase structure from barley revealed the division of the prophytepsin fold into three main elements, a 41 amino acid long propeptide region, following 338 residues constitutes the two-domain mature structure, and an independent plant-specific insert (PSI) of 104 amino acids present at the C-terminal domain of the enzyme. Therefore, a mature phytepsin is made up of two polypeptide chains, 1-247 and 248-338 and the type of fold is typical of other APs as well. It consists of two similar β-barrel domains with the active site aspartic acid resides (Asp36 and Asp223) in the interdomain cleft. The hydrophobic core of the molecule at the bottom of the cleft is shielded by a six stranded β-sheet structure with three conserved disulphide bridges stabilizing the structure. The propeptide region of the molecule wraps around the interdomain cleft in such a manner that its N-terminal takes part in the formation of six stranded β-sheet and the helical part covers the active site from the opposite side [Kervinen *et al*, 1999]. These features are conserved in the three-dimensional structure of aspartate proteases from cardoon.

Figure 1.11: Three-dimensional structure of aspartate protease from *Hordeum vulgare* (PDB-ID: 1QDM).

1.9.3.2 Cysteine proteases

In plants, PCD has been implicated in xylogenesis [Fukuda, 1996], senescence, and hypersensitive response to biotic and abiotic stresses [Greenberg *et al*, 1996]. Activation of cysteine proteases, through PCD-activating oxidative stress, constitutes the critical point in this pathway in animal and plant cells. Several cysteine proteases from different plant sources have been crystallized till date and their three-dimensional structures determined (Table 1.5).

Table 1.5: Crystal structures of plant cysteine proteases available in protein data bank.

Serial number	Source	PDB-ID
1	*Actinidia arguta*	3P5U, 3P5V, 3P5W, 3P5X
2	*Tabernaemontana divaricata*	3BCN, 2PRE, 1IWD, 1O0E, 2PNS
3	*Actinidia chinensis*	1AEC, 2ACT
4	*Zingiber officinale*	1CQD
5	*Ricinus communis*	1S4V
6	*Hordeum vulgare*	2FO5
7	*Jacaratia mexicana*	2BDZ
8	*Carica papaya*	1BP4, 1BQI, 1CVZ, 1KHP, 1KHQ, 1PAD, 1PE6, 1PIP, 1POP, 1PPD, 1PPN, 1PPP, 1STF, 2CIO, 2PAD, 3E1Z, 3IMA, 3LFY, 4PAD, 5PAD, 6PAD, 9PAP, 1YAL, 1MEG, 1PCI, 1PPO, 1GEC
9	*Pachyrhizus erosus*	2B1M, 2B1N
10	*Carica candamarcensis*	3IOQ

The crystal structure of cysteine endopeptidase from barley (EP-B2) (PDB-ID: 2FO5) in complex with the peptidic cysteine protease inhibitor (Leupeptin) was at 2.2 Å [Bethune *et al*, 2006]. The three-dimensional structure comprises of two comparably sized domains, designated as R and L, which are divided by the active site cleft (Figure 1.12). The domain R is composed of an extended N-terminal loop followed by four antiparallel β-sheets while the domain L comprises primarily of α-helical domain. The crystal structure revealed the presence of three conserved disulphide bonds between C25-C67, C59-C100, and C161-C213 that stabilizes the tertiary structure. This feature is highly conserved in other related cysteine endoprotease. The active site of the EP-B2 is located in a deep groove between the R and L domains.

The main chain of Leupeptin is bound along the length of this groove by forming two hydrogen bonds between G70 and P2 position in backbone and another one between D166 and P1 position in backbone. The nucleophilicity of the key residue C28 is enhanced by H167 mediated proton abstraction. The amide oxygen of the conserved N188 side chain orient H167 in such a manner to facilitate deprotonation of C28, thus completing the catalytic triad.

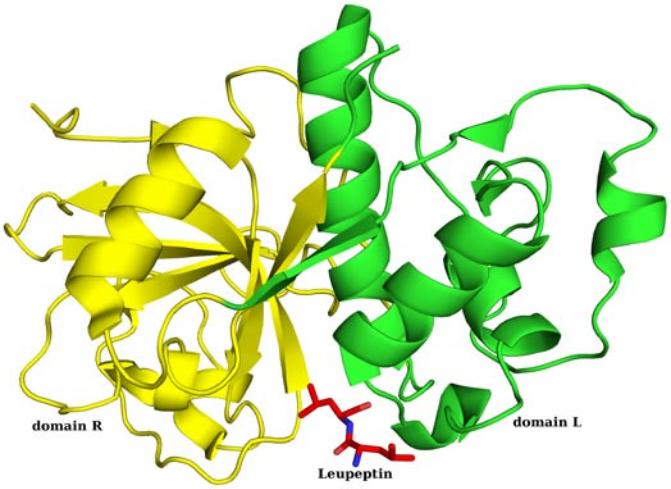


Figure 1.12: Three-dimensional structure of cysteine protease of *Hordeum vulgare* (PDB-ID: 2FO5).

1.9.3.3 Serine proteases

1.9.3.3.1 Trypsin

Trypsins constitute the largest group of proteolytic enzymes that display diverse specificities and perform endopeptidase activity. Their catalytic triad consists of the conserved Asp-His-Ser "charge relay" system. In chickpea, DegP proteases were identified which belongs to the trypsin family and have been implicated in heat shock response which combines both chaperone and proteolytic activities that switch in a temperature dependent manner. These ATP-dependent proteins are located in the thylakoid lumen and participate in high light stress responses. Previous structures of these proteases have indicated the role of C-terminal PDZ domain(s) in their regulation (PDB: 4IC6) and substrate recognition [Sun *et al*, 2013]. However, few structures were reported with no PDZ domain and the trimeric structure (PDB: 4IC5) reveals a novel catalytic triad conformation [Sun *et al*, 2013] (Figure 1.13 A).

1.9.3.3.2 Clp Endopeptidase

Clp proteases are a group of ATP-dependent serine endopeptidases consisting of a smaller protease subunit ClpP, and a larger chaperone regulatory ATPase subunit (either ClpA or ClpX) (Figure 1.13 B). Although the protease domain is capable of proteolysis on its own, ATPase subunits are essential for effective levels of proteolysis. The catalytic triad residues Ser-His-Asp are enclosed in a single cavity that allows for degradation of small peptides but precludes the entry of the large folded polypeptides.

1.9.3.3.3 C-terminal processing peptidase

C-terminal processing peptidases (Ctps) belong to MEROPS family S41. *E.coli* Tsp, a periplasmic endoprotease, was one of the first members of the family to be identified. Most family members are associated with PDZ domains that are protein-protein interaction modules important for substrate recognition [Spiers *et al*, 2002]. The C-terminal processing peptidase family consists of two subfamilies, which differ greatly in activity, sensitivity to inhibitors and molecular structure. The two subfamilies employ different mechanisms for catalysis. While, the C-terminal processing peptidase subfamily (S41A) is found chiefly in bacteria and some eukaryotes employing a Ser-Lys catalytic dyad, the tricon core protease subfamily (S41B) is largely confined to archaea with few bacterial representatives employing a catalytic tetrad consisting of Ser, His, Ser and Glu residues (Figure 1.13 C).

1.9.3.3.4 Lon protease

Like Clp proteases, Lon proteases are also a group of ATP-dependent serine proteases. The catalytic and ATPase domains of the Lon proteases reside in the same polypeptide. The three-dimensional structure consists of three functional domains: the N-terminal domain (LON), a central ATPase domain (AAA+ module) and a C-terminal proteolytic domain (Lon_C) (Figure 1.13 D). The N-terminal LON domain along with the AAA module is important for substrate specificity in Lon proteases [Rotanova *et al*, 2004]. Sequence and structure analysis indicate that Lon proteases

employ a Ser-Lys/Arg (S, K/R) catalytic dyad instead of a canonical serine protease catalytic triad [Rotanova *et al*, 2004].

1.9.3.3.5 Lys-Pro-x Carboxypeptidase

This family includes several eukaryotic enzymes such as lysosomal Pro-X carboxypeptidase (PCP), dipeptidylpeptidase II and thymus specific serine peptidase. The predicted active site residues for the members of this family (Ser, Asp, and His) occur in the same order in the sequence as that of family S10 (Figure 1.13 E).

1.9.3.3.6 Nucleoporin autopeptidase

Nucleoporin autopeptidases are a group of autocatalytic serine endopeptidases synthesised as precursors and processed by autoproteolysis prior to their association in the Nuclear Pore Complex (NPC). Various studies have shown the importance of C-terminal domain of the nucleoporins in their localization to the NPC. Mutations that inhibit proteolytic processing within the C-terminal domain prevent their association with the pore as well as inhibit the formation of NPCs [Teixeira *et al*, 1999]. Structural analysis of the proteolytic cleavage site confirms the presence of a catalytic dyad (H and S) within His-Phe-Ser motif [Hodel *et al*, 2002] (Figure 1.13 F).

1.9.3.3.7 Prolyl oligopeptidase

The members of Prolyl oligopeptidases (POPs) family of serine peptidases belong to the α/β hydrolase that includes members of different types and with distinct specificities. POPs are involved in the degradation of biologically important peptides such as peptide hormones and neuropeptides associated with learning and memory and therefore have become significant targets for drug design [Rosenblum *et al*, 2003]. The crystal structure of enzyme shows unique domain architecture with a catalytic α/β hydrolase domain and an unusual β-propeller domain. Propeller domain has radially arranged seven-fold repeat structure of four-stranded antiparallel β sheets (Figure 1.13 G). The catalytic triad (Ser, His, and Asp) is hidden and located at the interface of two domains. This unique propeller acts as a lid to hide the active site and also as a gating filter, thereby allowing only small peptides to reach active site [Fülöp *et al*, 1996].

1.9.3.3.8 Protease IV

Protease IV class of serine peptidases degrade the signal peptides that accumulate in the cytosol subsequent to their removal from precursor polypeptides by signal peptidases. Signal peptides are responsible for targeting various proteins to their respective subcellular localizations and are subsequently removed from the pre-proteins. These generated signal peptides need to be rapidly degraded since they may be harmful to cells by interfering with protein translocation or may accumulate in the membrane leading to cell lysis. The predicted active site serine residues for members of Protease IV family are Ser, Arg/His, and Asp [Rawlings *et al*, 2006] (Figure 1.13 H).

1.9.3.3.9 Rhomboid

Rhomboid proteins are serine proteases that cleave substrates within transmembrane domains [Freeman, 2003]. A proposed catalytic triad, located within a transmembrane domain, comprising of Asn, Ser and His [Urban *et al*, 2001]. However, a recent report suggests that while Ser and His along with a glycine residue (two residues away towards the N-terminal side of Ser) are essential for catalysis, Asn is not required for catalytic activity and that rhomboids are likely to function as endopeptidases with a serine-histidine dyad (Figure 1.13 I).

1.9.3.3.10 Serine carboxypeptidase

Serine carboxypeptidases catalyze the hydrolysis of the C-terminal bond in proteins and peptides. Crystal structures show that serine carboxypeptidases belong to the α/β hydrolase fold and possess a catalytic triad similar to members of chymotrypsin and subtilisin families in the order Ser, Asp and His (S, D, and H) placed in the polypeptide chain [Rawlings *et al*, 2006] (Figure 1.13 J).

1.9.3.3.11 Signal peptidase I

Signal peptidases are a diverse group of serine endopeptidases responsible for the removal of signal peptides from preproteins within the cell [Paetzel *et al*, 2002]. The failure to remove signal peptide often leads to protein inactivation and/or

mislocalization [Paetzel *et al*, 2002]. The Type I SPases, a membrane-bound serine endopeptidases, is further divided into two subfamilies: members of S26A subfamily employ a serine-lysine dyad for catalysis (Ser, Lys) and include prokaryotic, chloroplast and mitochondrial peptidases, while the members of S26B subfamily that include ER SPC and archaeal peptidases, employ a serine-histidine dyad (Ser, His) based on the mechanism of catalysis [Paetzel *et al*, 2002] (Figure 1.13 K).

1.9.3.3.12 Subtilase

The subtilase family is the second largest family of serine proteases identified in kingdoms like eubacteria, archaebacteria, eukaryotes and viruses. The three-dimensional structure reveals that subtilase utilize a highly conserved catalytic triad similar to the chymotrypsin and carboxypeptidase clans but have a different sequential order of Asp, His and Ser catalytic triad residues in the sequence with no other structural similarity [Siezen & Leunissen, 1997] Some members of the subtilase family appear to be mosaic with little or no sequence similarity to any other known proteins [Siezen & Leunissen, 1997] and with large N- and C-terminal extensions (Figure 1.13 L).

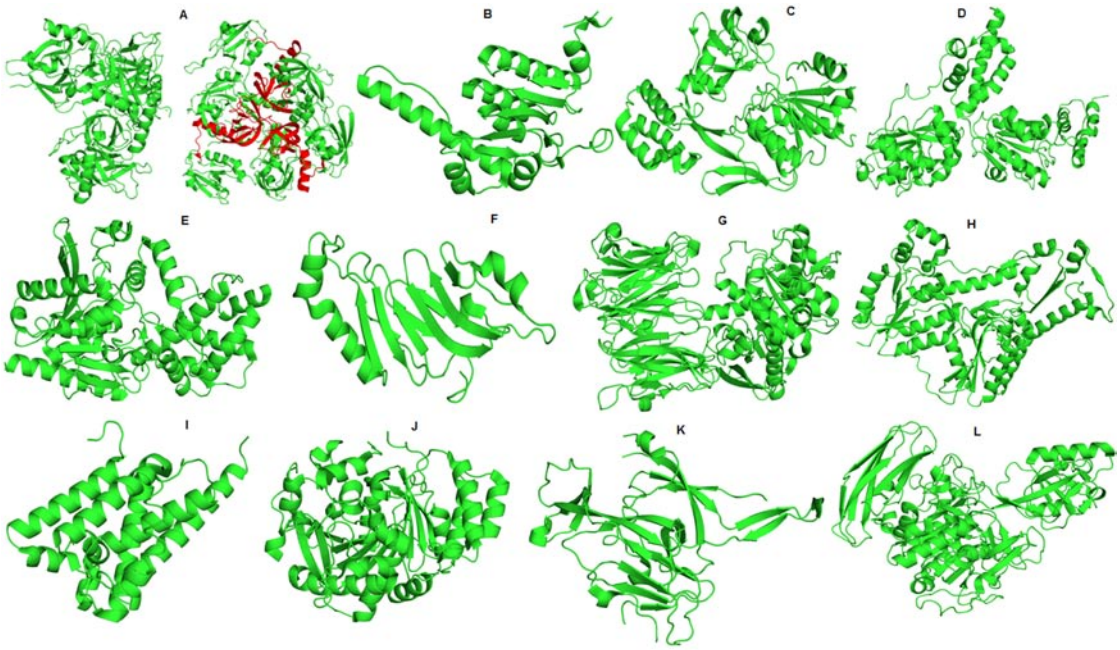


Figure 1.13: The three-dimensional structures of representative members of the 13 classes [A-M] of serine proteases. A. Trypsin (Deg5 & Deg8) from *Arabidopsis* (PDB-ID: 4IC5 & 4IC6); B. Clp Endopeptidase from *Bacillus subtilis* (PDB-ID: 3KTG); C. C-terminal processing peptidases from *Scenedesmus obliquus* (PDB-ID: 1FC6); D. Lon proteases from *Bacillus subtilis* (PDB-ID: 3M6A); E. Lys-Pro-X carboxypeptidase from *Homo sapiens* (PDB-ID: 3N2Z); F. Nucleoporin

autopeptidases from *Homo sapiens* (PDB-ID: 2Q5X); G. Prolyl oligopeptidases from *Trypanosoma brucei* (PDB-ID: 4BP8); H. Protease IV from *Escherichia coli* (PDB-ID: 3BEZ); I. Rhomboid from *Haemophilus influenzae* (PDB-ID: 2NR9); J. Serine carboxypeptidases from *Triticum aestivum* (PDB-ID: 1BCR); K. Signal peptidases I from *Escherichia coli* (PDB-ID: 1B12); L. Subtilase from *Cucumis melo* (PDB-ID: 3VTA).

1.9.3.4 Metalloproteases

No crystal structure of plant metalloprotease has been reported yet.

1.9.4 Protease inhibitors

Protease inhibitors (PIs) consist of four classes based on the protease they inhibit: cysteine PIs, serine PIs, metallo PIs, and aspartyl PIs. Serine PIs are the most well studied class of PIs with a widespread distribution in the plant kingdom. The second most well studied class of inhibitor is cysteine PIs whereas very little is known about the remaining two classes.

1.9.4.1 Cysteine protease inhibitors

Cysteine protease inhibitors are known as cystatins or phytocystatins. Based on their three-dimensional structure, they are classified into two groups. One group possess single inhibitory domain whereas PIs possessing multiple domains are included in the second group. Along with their important inhibitory role against herbivory, they are known to participate in PCD. Only four structures of cysteine PIs from the plant species, namely *Ananas comosus* (PDB-ID: 2L4V, NMR structure), *Oryza sativa* subsp. Japonica (PDB-ID: 1EQK, NMR structure), *Solanum tuberosum* (PDB-ID: 3W9P, Resolution: 2.70 Å), and *Colocasia esculentum* (PDB-ID: 3IMA, Resolution: 2.03 Å), are crystallized till date.

As compared to other plant and animal cystatins, potato multicystatin (PMC) has a high molecular mass (85 kDa/monomer); most plant protease inhibitors vary between 8 and 25 kDa [Garcia-Olmedo *et al*, 1987]. PMC can bind and inhibit several cysteine proteases simultaneously hence termed as multicystatin owing to its multiple inhibitory domains [Walsh & Strickland, 1993]. PMC consists of eight tandem cystatin domains of 10 kDa molecular mass linked by proteolytically sensitive short linkers (Figure 1.14 A & B) [Walsh & Strickland, 1993]. The soluble form of PMC (85 kDa) is distributed throughout the tuber. When ingested by insects, the acidic pH

of the midgut solubilizes the crystalline PMC, thus interfering with protein digestion and retards the larval growth [Walsh & Strickland, 1993]. Hence, protease inhibitors serve as potent insecticide [Ryan, 1990] and introduction of such genes in the plants imparts resistance against selected insect pests [Irie *et al*, 1996].

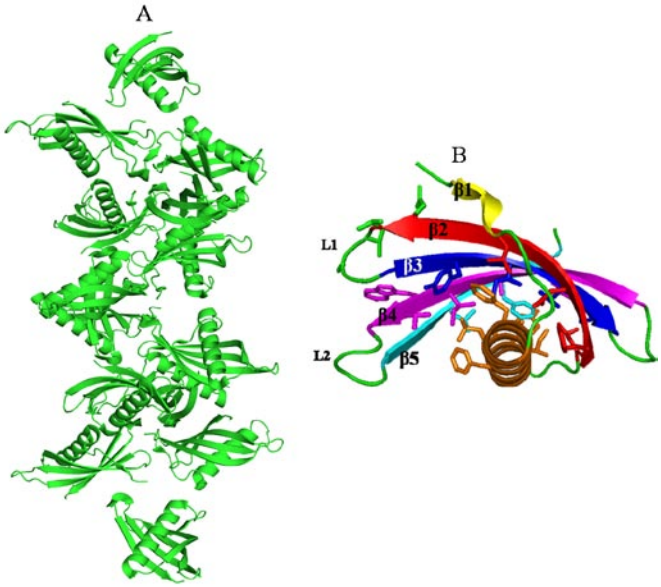


Figure 1.14: The three-dimensional structure of cysteine protease inhibitor of *Solanum tuberosum* (PDB-ID: 3W9P). Panel B is adopted from Nissen *et al*, 2009.

1.9.4.2 Serine protease inhibitors

Serine protease inhibitors is the well studied class of inhibitors. Based on their amino acid composition, conformation and structure of reactive site loop, they are classified into nine classes as follows.

1.9.4.2.1 Kunitz-Type Protease Inhibitor (KPI) Family

The three-dimensional structures of KPI were solved from twelve plant species listed in Table 1.6. The structure of *Delonix regia* trypsin inhibitor (DrTI; PDB-ID: 1R8N) comprises of a β-trefoil fold consisting of 12 antiparallel β-strands connected by long loops that form six two-stranded β-hairpins. Three of the β-hairpins form a barrel structure while the remaining three form a triangular cap on the barrel (Fig. 1.15 A). The three structural repeats marked in the figure related by a pseudo-threefold axis of symmetry oriented parallel to the barrel axis are a characteristic feature of this fold. Comparison of DrTI with other inhibitors of the Kunitz (STI) family revealed its reactive site loop from the residue Ser65 (P4) to the residue Ile73 (P4'). The superposition between the reactive site loops of DrTI and STI shows that an insertion

of one residue, Glu68, between the residues P1 and P2 that distorts DrTI reactive loop as compared to the canonical conformation.

Table 1.6: Crystal structures of plant Kunitz-type protease inhibitors available in protein data bank.

Serial number	Source	PDB-ID
1	*Psophocarpus tetragonolobus*	1WBA, 1EYL, 1FMZ, 1FN0, 1WBC, 1XG6, 2BEA, 2BEB, 2ESU, 2ET2, 2QYI, 2WBC, 3I29, 3I2A, 3ICX, 3QYD, 4WBC
2	*Bauhinia bauhinioides*	2GZB, 2GO2
3	*Murraya koenigii*	3IIR
4	*Delonix regia*	1R8N
5	*Hordeum vulgare*	3IVA, 2IWT, 3BX1
6	*Oryza sativa*	2QN4
7	*Erythrina caffrs*	1TIE
8	*Glycine max*	1AVU, 1AVW, 1AVX, 1BA7
9	*Carica papaya*	3S8J, 3S8K
10	*Lepidium virginicum*	2DRE
11	*Sagittaria sagittifolia*	3E8L
12	*Solanum tuberosum*	3TC2

1.9.4.2.2 Bowman-Birk Inhibitors (BBI-PI) Family

This class of serine protease inhibitors is widely present in legumes and cereals. The BBIs from dicot species have a molecular mass of ~ 8 kDa with double headed structure that means they can inhibit two serine proteinases at same time. There are two types of BBI in monocots. One group consists of single polypeptide chain with molecular mass of ~ 8 kDa bearing a single reactive site. Contrary to this, another group has molecular mass of 16 kDa with two reactive site. The structures of BBI from eleven plant species has been solved till date (Table 1.7). The crystal structure of BBI purified from snail medic (*Medicago scutellata*; PDB-ID: 2ILN) (MSTI) was solved in complex with two molecules of bovine trypsin to a resolution of 2 Å. Both the binding site loops of MSTI have arginine in position P1 however; there exist a difference at only one position. Six disulphide bridges are present in the structure that provides stability to the active structure (Fig. 1.15 B).

Table 1.7: Crystal structures of plant Bowman-Birk inhibitors available in protein data bank.

Serial number	Protein	PDB-ID
1	*Phaseolus angularis*	1TAB
2	*Glycine max*	1BBI, 1D6R, 1K9B, 2BBI, 2PI2
3	*Macrotyloma axillare*	1GM2
4	*Hordeum vulgare*	1C2A, 1TX6, 2FJ8, 1TX6, 1C2A, 2FJ8, 1C2A, 1TX6, 2FJ8
5	*Lens culinaris*	2AIH
6	*Medicago scutellata*	1MVZ, 2ILN
7	*Phaseolus lunatus*	1H34
8	*Vigna radiata*	1G9I, 3MYW, 1SBW, 1SMF, 3MYW
9	*Vigna unguiculatus*	2G81, 2R33
10	*Pisum sativum*	1PBI
11	*Oryza sativa*	2QN5

1.9.4.2.3 Squash Family

The available crystal structures of these inhibitors revealed lack of any regular secondary structure elements (Table 1.8). Similar to the other 'small' serine protease inhibitors, the reaction center loop protrudes from the main body in a characteristic conformation allowing an intimate contact with the cognate enzyme. In spite of different size and fold of the squash inhibitors the interaction pattern is similar in all the observed related complexes (Fig. 1.15 C).

Table 1.8: Crystal structures of plant squash protease inhibitors available in protein data bank.

Serial number	Protein	PDB-ID
1	*Cucurbita maxima*	1CTI, 1LUO, 1PPE, 2CTI, 2STA, 2V1V, 3CTI
2	*Ecballium elaterium*	1H9I, 1W7Z, 2ETI, 2IT7, 2LET, 2C4B
3	*Momordica charantia*	1F2S, 1MCT
4	*Momordica cochinchinensis*	2C4B, 1HA9, 1IB9, 2IT8, 2PO8
5	*Cucurbita pepo*	2BTC, 2STB

1.9.4.2.4 Serpin Family

A distinguishing feature of serpins is the presence of a reactive centre loop (RCL), which displays a protease target sequence as bait. RCL perform an irreversible cleavage by forming a covalent serpin-protease complex. The x-ray crystal structure of recombinant AtSerpin1 in its native stressed conformation was determined at 2.2 Å (PDB-ID: 3LE2). AtSerpin1 structure resembles those of canonical serpins comprising three conserved β-sheets and nine conserved α-helices. An additional helix is seen in the RCL, as well as three single-turn helices (Fig. 1.15 D).

1.9.4.2.5 Ragi seed Trypsin/ α-Amylase Inhibitor/ Lipid transfer protein Family

The three-dimensional structure of the adducted lipid transfer protein (LTP) from barley was solved and deposited in the database under PDB-ID: 3GSH. The structure consists of four helices separated by three short loops and concluded by a long C-terminal end without a definite secondary structure. The four disulphide bridges that correspond to the LTP1 topology were observed in the structure (Fig. 1.15 E).

Table 1.9: Crystal structures of plant Ragi seed Trypsin/α-Amylase Inhibitor/ Lipid transfer proteins available in protein data bank.

Serial number	Plant source	PDB-ID
1	*Capparis masaikai*	2DS2
2	*Helianthus annuus*	1S6D
3	*Ricinus communis*	1PSY
4	*Arachis durarensis*	3OB4
5	*Arachis hypogaea*	1W2Q
6	*Triticum aestivum*	1HSS, 1BW0, 1CZ2, 1GH1
7	*Eleusine coracana*	1B1U, 1BIP, 1TMQ
8	*Zea mays*	1BEA, 1BFA, 1AFH, 1FKO, 1FK1,1FK2, 1FK3, 1FK4, 1FK5, 1FK6, 1FK7, 1MZL, 1MZM
9	*Hordeum vulgare*	1BE2, 1JTB, 1LIP, 1MID, 3GSH
10	*Oryza sativa subsp indica*	1BV2, 1UVA, 1UVB, 1UVC
11	*Oryza sativa subsp japonica*	1RZL
12	*Prunus persica*	2ALG, 2B5S
13	*Nicotiana tabacum*	1T12
14	*Vigna radiata*	1SIY
15	*Brassica napus*	1SM7

1.9.4.2.6 Potato Type I PIs (Pin-I)

Potato inhibitor I from buckwheat seeds, BWI-1 (Buckwheat Inhibitor 1) was crystallized at 1.84 Å resolution and the coordinates were available under PDB-ID: 3RDY (Fig. 1.15 F). The rBTI protein sequence is composed of 69 amino acid residues. Its main secondary structural elements comprise of a single α-helix (α1), a central parallel β-sheet with two strands (β1), a binding site loop and two irregular structures at the N-terminus and C-terminus. A hydrophobic core is present among α1, β1, β2 and two short loops. A disulphide bond exists within the N-terminus and stabilizes the binding loop by connecting it. The binding site loop of rBTI is a convex loop sandwiched between β1 and β2.

Table 1.10: Crystal structures of plant Pin-I proteins available in protein data bank.

Serial number	Plant source	PDB-ID
1	*Hordeum vulgare*	1CIQ, 1CQ4, 1CIR, 1CIS, 1COA, 1LW6, 2SNI, 1YPA, 1YPB, 1YPC, 2CI2, 3CI2, 1TM5, 1TM1, 1TM3, 1TM4, 1TM7, 1TMG, 1TO1, 1TO2, 1Y1K, 1Y33, 1Y34, 1Y3B, 1Y3C, 1Y3D, 1Y3F, 1Y48, 1Y4A, 1Y4D
2	*Linum usitatissimum*	1DWM
3	*Cucurbita maxima*	1HYM, 1MIT, 1TIN
4	*Momordica charantia*	1VBW
5	*Fagopyrum esculentum*	3RDY, 3RDZ

1.9.4.2.7 Potato Type II PIs (Pin-II)

Tomato inhibitor-II (TI-II) is a member of the Potato II (Pot II) proteinase inhibitor family of serine proteinase inhibitors (PIs). The three-dimensional structure of TI-II was determined by molecular replacement technique to 2.15 Å resolution, available in the database under PDB code: 1PJU (Fig. 1.15 G). There are four copies of TI-II (designated as A, B, C, and D) in the asymmetric unit and the overall conformation of each copy is similar in most regions of TI-II from the complex. TI-II consists of two structurally similar inhibitory domains (Domains I and II) and each inhibitory domain adopts the fold determined previously for the single domain Pot II inhibitors (48–51). The structure consists of only a small amount of regular secondary structure in the

form of a small antiparallel β-sheet and irregular loop regions. Both domains also contain one turn of a 3_{10} helix. An inhibitory reactive site loop is found at opposite ends in each domain allowing a single inhibitor to bind to two protease molecules simultaneously.

Table 1.11: Crystal structures of plant Pin-II proteins available in protein data bank.

Serial number	Plant source	PDB-ID
1	*Solanum lycopersicum*	1OYV, 1PJU
2	*Solanum tuberosum*	4SGB
3	*Nicotiana alata*	1TIH, 1YTP, 2JYY, 1FYB, 1CE3, 1QH2, 2JZM

1.9.4.2.8 Kazal Family

The crystal structure of no Kazal inhibitor is reported from any plant species.

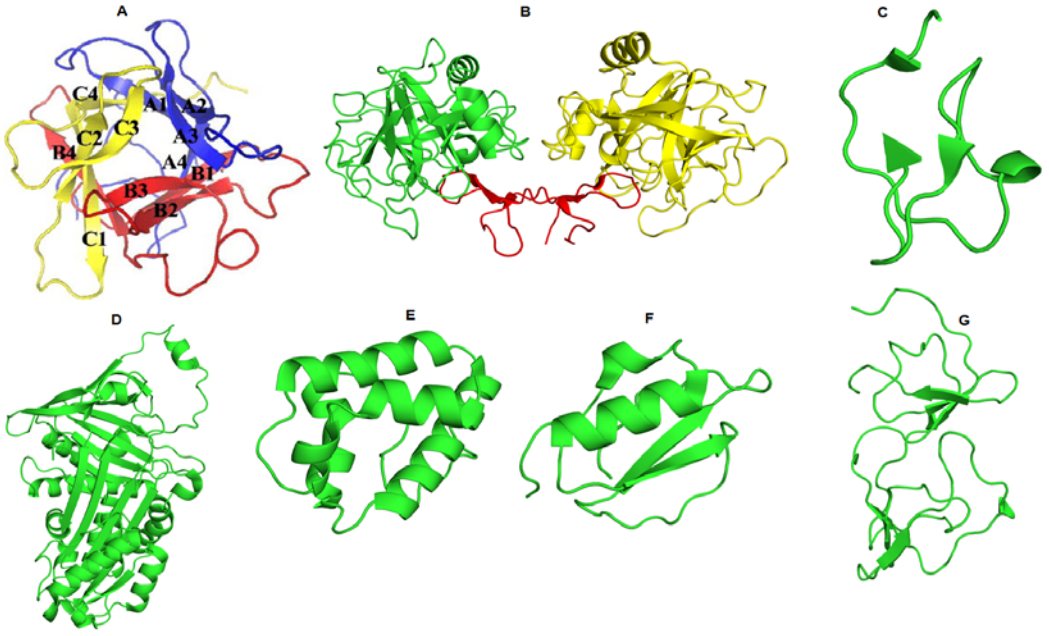


Figure 1.15: The three-dimensional structures of serine protease inhibitors [A-G]. A. Kunitz inhibitor from *Delonix regia* trypsin inhibitor (PDB-ID: 1R8N); B. Bowman- Birk inhibitor from *Medicago scutellata* (PDB-ID: 2ILN); C. Squash inhibitor from *Cucurbita pepo* (PDB-ID: 2BTC); D. Serpin from *Arabidopsis* (PDB-ID: 2ILN); E. Ragi seed Trypsin/ α-Amylase Inhibitor/ Lipid transfer protein from *Hordeum vulgare* (PDB-ID: 3GSH); F. Pin-I from *Fagopyrum esculentum* (PDB-ID: 3RDY); G. Pin-II from *Solanum lycopersicum* (PDB-ID: 1PJU).

1.10 Genome-wide study of chickpea to improve stress tolerance

In the present study, we have identified various classes of genes involved in stress and defense responses in chickpea. The results from tissue specific expression studies of the identified genes were analyzed. In addition to that, drought stressed RNA-seq reads libraries were utilized to identify the up- and down-regulated genes. A number of genes from each class were shown to have an altered expression upon exposure to drought condition. Several studies have shown that the increased synthesis of transcription factors like DRE, AREB1, ABF2 etc. seems to be a promising approach in the development of drought resistant/tolerant transgenic plants when compared to engineering individual functional genes [Bartels & Hussain 2008]. Increasing the number of stress-specific *cis*-regulatory elements and by inducing the synthesis of such genes that overexpress upon exposure to stress could be another strategy to deal with such adverse situations.

Chapter 2

Materials and Methods

In the research accounted here, several methodologies have been used for the gene identification and to study the structure-function relationship of the stress and defense proteins in chickpea. The materials and methods utilized in the entire study are enlisted below.

2.1 Sequence analysis

The protein sequences were retrieved from UniProtKB/ Swiss-Prot database and aligned using multiple sequence alignment program ClustalX 2.1 (Thompson *et al*, 1997) to analyze the extent of conservation among the sequences. The presence and probable location of signal peptide were also predicted using SignalP 4.1 Server (Petersen *et al*, 2011).

2.2 Phylogenetic analysis

Prior to phylogeny building, protein sequences were aligned using **MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation (MUSCLE) (Edgar, 2004), the program that is part of the MEGA. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based clustering method was used to generate the alignment keeping the gap open and gap extension penalties 22.9 and 0. This alignment was used to build the phylogenetic tree using the Neighbor Joining (NJ) algorithm of MEGA with the Dayhoff substitution matrix (PAM250) and bootstrap value set to 1000.

A different methodology was utilized to analyze the evolutionary history of chickpea UGTs by using the fast likelihood-based method in order to generate the dendrograms by approximate LRT (aLRT) method (Guindon *et al*, 2010; Anisimova *et al*, 2011) in PhyML 3.0 (Guindon *et al*, 2010). Amino acid sequences were given as input in phylip format keeping LG (Le and Gascuel, 2008) substitution model and proportion of invariable sites and number of substitution rate categories as 0 and 4. Nearest Neighbor Interchanges (NNI) algorithm was utilized in order to improve a reasonable starting tree topology. The phylogenetic trees generated were visualized and analyzed in FigTree v1.3.1 (Rambaut, 2009).

2.3 Recent gene duplication events

The gene pairs involved in recent gene duplication events were identified by calculating the percent identity among them by preparing percent identity matrix in ClustalX and with the help of Pairwise Sequence Alignment utility at EBI (Global alignment). Blosum62 matrix was used for the alignment keeping gap open and gap extension 10 and 0.5. The end gap penalty, end gap open, and end gap extension penalty was set to false, 10, and 0.5. The sequence identity cut off between the two gene pair should be ≥ 90%.

2.4 Molecular modeling

Similarity search using Basic Local Alignment Search Tool (BLAST) algorithm (Altschul *et al*, 1990) was carried out against the Protein Data Bank (PDB) to select the high resolution homologous protein crystal structures (Berman *et al*, 2000). The sequence identity cut off was set to ≥ 30% (E-value cut off =1). Comparative modeling of target protein sequences was carried out using the Composite/ Chimeric model type of Prime 3.1 [Schrödinger], using homologous template structures to analyze their structural features, binding mode and affinity with the substrates. The structures were visualized in PyMOL software [Schrödinger].

2.5 Validation of homology models and refinement

The molecular models were evaluated using Glide gscore, Emodel value, ERRAT 2.0 (Colovos *et al*, 1993), PDBsum (Laskowski *et al*, 2009), ProSA-web Protein Structure Analysis (Wiederstein *et al*, 2007), Verify3D Structure Evaluation Server (Luthy *et al*, 1992) and RMSD based on Cα overlap between target and template. GlideScore is an empirical scoring function that approximates the ligand binding free energy and affinity that consist of many terms including force field (electrostatic, van der Waals) contributions and terms rewarding or penalizing interactions known to influence ligand binding. GlideScore should be used to rank poses of different ligands, for example in virtual screening. Emodel value also uses the force field components (electrostatic and van der Waals energies) to comparing conformers, but much less so for comparing chemically-distinct species. Therefore, Glide uses Emodel to pick the best pose among all the generated poses of a ligand and then ranks these best poses

against one another with GlideScore. The ERRAT program analyzes the environment of the atoms in the protein model. This program plots error values with respect to amino acid position in the sequence by sliding a nine residue window along the sequence. The error function is calculated based on the non-bonded atom-atom interactions in the protein structure. PDBsum is a pictorial database providing a detailed stereochemical analysis of the protein structure in terms of Ramachandran plot statistics, main chain parameters, main chain bond length and bond angles, distorted geometry etc. ProSA is a tool widely used to check errors in 3D protein models. It gives an overall quality score or Z score shown in a plot where scores estimated from experimentally determined structures in PDB are plotted. The Z-score also measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The Verify3D plot showed the overall compatibility of the 3-D structure with respect to the protein sequence by determining the environment of each amino acid in the 3-D model. The RMSD deviation between template and target Cα atoms was calculated by superposing structures in PyMOL. It reveals the quality of the model, lesser is the deviation better is the model.

After validation, homology models were refined by impref minimization of protein preparation wizard (Sastry *et al*, 2013) followed by minimization using Impact 5.8 [Schrödinger]. The model refinement phase involved preprocessing of model structures by adding hydrogens, assigning bond order, and filling missing loops and side chains. Later on, the models were subjected to restrained minimization by applying the constraint to converge the non-hydrogen atoms to an RMSD of 0.3 Å using OPLS 2005 (Jorgensen and Tirado-Rives, 1988) force field. Subsequently, the models were subjected to 500 steps of steepest descent energy minimization followed by 1000 steps of conjugate gradient energy minimization using the same force field.

2.6 Retrieval or designing of ligands

The 2-D sketcher utility of Maestro 9.3 was utilized to build the 2-dimensional structures of ligands which were then converted into 3-dimensional structures. The structural coordinates of Z-Pro-prolinal (ZPR; CID 122623) ligand was downloaded from PubChem Compound database (Bolton *et al*, 2008) in 3-D structure-data file

(SDF). The sdf file was converted to mol2 file format in order to carry out further structural studies.

2.7 Protein and ligand preparation

Before carrying out the docking studies, water and other hetero atom groups from the protein structures were removed using protein preparation utility of Maestro. Later on the hydrogens were added to perform restrained minimization of the models. The minimization was done using impref utility of Maestro in which the heavy atoms were confined such that the strains generated upon protonation could be relieved. The RMSD of the atomic displacement for terminating the minimization was set as 0.3 Å. Similarly, ligands were refined with the help of LigPrep 2.5 [Schrödinger] to define their charged state and enumerate their stereo isomers.

2.8 Docking studies

2.8.1 Protein-ligand docking

The molecular models generated as already described were used to dock the small molecule ligands in the respective active site pocket by employing Glide 5.8 (Schrödinger) (Friesner *et al*, 2004). A grid was made either by taking the reference ligand or by selecting the active site residues crucial for the substrate binding. Flexible ligand docking was carried out using the standard precision option. The ligands were docked in the active site by creating a grid around the bound reference ligand or by drawing the grid around the catalytic residue. A total of 20 poses generated were scored on the basis of their glide score and E-model values. Out of these poses, the most favorable one was chosen based on glide gscore, glide Emodel value and essential interactions required for the stable substrate binding.

2.8.2 Protein-protein docking

Protein-protein docking studies were performed by using ZDOCK 3.0.2 (Pierce *et al*, 2014) server by defining the interface residues between the two protein chains. A total of 10 poses generated were scored on the basis of their ZDOCK score. The most favorable pose is the one which had high ZDOCK score value and stable binding mode. The hydrogen bond interactions between the protein and their cognate substrates were visualized using PyMOL.

2.9 Molecular dynamics simulation studies

2.9.1 Molecular dynamics simulation in Gromacs

The docked complexes were subjected to molecular dynamics simulations using the GROningen Machine for Chemical Simulations V4.5.4 (GROMACS) (Van Der Spoel *et al*, 2005; Berendsen *et al*, 1995) using GROMOS96 43a1 force field. The docked complexes were enclosed at the centre of the dodecahedron box solvated in water using SPC216 water model keeping 10 Å distance between the solute and the box. Topology files and other force field parameter files for the ligands were generated with the help of PRODRG2 server (Schüttelkopf *et al*, 2004). The system was initially energy minimized by steepest descent minimization for 50,000 steps until a tolerance of 10 kJ/mol in order to avoid the high energy interactions and steric clashes. Net charges on the docked structures were neutralized by adding equal number of counter ions to make the whole system neutral using genion program of GROMACS. After addition of ions, the system was again energy minimized by steepest descent minimization keeping identical parameters. The V-rescale, a modified Berendsen thermostat, temperature coupling (Berendsen *et al*, 1984) and Parrinello-Rahman pressure coupling (Martonàk *et al*, 2003) methods were used to keep the system stable at 300 K temperature and pressure of 1 bar. The Particle Mesh Ewald (PME) method (Darden *et al*, 1993) was selected to deal with long range electrostatic interactions. A distance cut off of 9 Å and 14 Å was set for Coulombic and van der Waals interactions. LINCS algorithm (Hess *et al*, 1997) was used to handle the rotational constraint to bonds. No positional constraints were applied on the system and periodic boundary conditions were applied in all three directions. The trajectories were visualized using Visual Molecular Dynamics program (VMD) (Humphrey *et al*, 1996).

2.9.2 Molecular dynamics simulation in Desmond

The docked complexes mentioned in table were prepared first using protein preparation wizard to check for any errors in the structure. Later on the processed complexes were subjected to molecular dynamics simulations using desmond 3.1 (Guo *et al*, 2010) of Maestro. OPLS2005 force field was applied on docked

complexes placed in the centre of the orthorhombic box solvated in SPC water model. Total negative charges on the docked structures were balanced by equal number of counter ions to make the whole system neutral (Table 2.1). The system was initially energy minimized for maximum 2000 iterations of the steepest descent (500 steps) using limited memory Broyden-Fletcher-Gold farb-Shanno (LBFGS) algorithm with a convergence threshold of 1.0 kcal/mol/Å. The short- and long-range Coulombic interactions were taken care by Cutoff and Smooth particle mesh ewald method keeping the cut off radius of 9.0 Å and ewald tolerance of $1*10^{-09}$. Periodic boundary conditions were applied in all three directions. The final run of 10 ns was applied on the relaxed system with a time step of 2.0 fs using NPT ensemble by employing a Berendsen thermostat at 300 K temperature and atmospheric pressure of 1 bar. The energies and trajectories were recorded after every 2.0 ps. The Cα atom RMSD of the complexes in each trajectory were calculated and plotted with respect to simulation time.

Table 2.1: Details of box size and number of water molecules and ions added during simulation process.

Serial number	Chickpea protein	Substrate/ Inhibitor	Box dimension	Number of water and ions added
1	POP	ZPR	59.69*82.42*64.98	19019, 18 Na^+
2	Cysteine protease inhibitor	Papain from *Carica papaya*	59.23*90.80*70.65	10614, 8 Cl^-
3	Bowman-Birk inhibitor	Bovine trypsin	112.25*69.25*70.51	9304, 4 Cl^-
4	Potato inhibitor- I	Bovine trypsin	58.60*79.52*71.99	15134, 17 Cl^-

2.10 Gene identification

Draft genome of *C. arietinum* was downloaded from Legume information system (http://cicar.comparative-legumes.org/) database. Chickpea has an estimated genome size of around 740 Mb which consists of 28,269 gene models and 7,163 scaffolds covering 544.73 Mb (over 70% of estimated genome size) (Varshney *et al*, 2013). The genes involved in biotic and abiotic stress response in chickpea were identified by following different methodologies namely blastp search, hidden Markov model

profile search, and Position-Specific Weight Matrix search explained in respective chapters. Predicted proteome of chickpea was searched for the presence of stress genes by screening using HMM-profiles of Pfam 27.0 (Pfam family: PF00201.13) (Punta *et al*, 2012) with the help of HMMER 3.0. (Eddy, 1998; http://hmmer.org/) selecting E-value cut off of 10^{-4}.

2.11 Detection of orthologs

Orthologs of all the identified chickpea stress proteins were searched in dicot plant genomes of *Medicago truncatula*, *Glycine max*, *Vigna angularis*, *Medicago sativa*, *Vitis vinifera*, *Lotus japonicas*, *Phaseolus vulgaris*, and *Arabidopsis thaliana* employing Blast2Go (Conesa *et al*, 2005) tool keeping E-value cut off 0.001 and sequence similarity ≥ 80%. These dicot plants were selected for analysis based on their reported chickpea homologous genomes and Blast2Go hits.

2.12 Analysis of gene structure

The gene architecture of stress genes was analyzed using the Gene Structure Display Server (Guo *et al*, 2007; GSDS) using the gene sequences and coding sequences. The generated output depicts the exon/intron arrangement, gene length, intron phases and position, length and position of exon and introns, and 3' and 5' untranslated regions (UTRs). The three intron phases were assigned as 0 for introns between two codons, 1 for those between first and second base of codon and 2 for introns inserted between second and third base.

2.13 Gene expression analysis

2.13.1 Tissue specific RNA-seq data analysis

RNA-seq raw read data or expression data was downloaded from Sequence Read Archive (SRA) database available at NCBI (http://www.ncbi.nlm.nih.gov/sra), for 5 different plant tissues namely, germinating seedling (GSM1047862), young leaves (GSM1047863), shoot apical meristem (GSM1047864) (Sam), flower bud (GSM1047865, GSM1047866, GSM1047867, GSM1047868) and flower (GSM1047869, GSM1047870, GSM1047871, GSM1047872) (Singh *et al*, 2013). Reads were mapped to genomic sequence of chickpea with spliced read mapper,

TopHat (Trapnell *et al*, 2009). Cufflinks tool was used to estimate and analyze the abundance of reads mapped to genes body and thus calculating fragment kilo base transcript per million (FPKM) value as proxy for gene expression in different plant tissues (Trapnell *et al*, 2010).

2.13.2 Analysis of RNA-seq data of drought stressed root tissues

RNA-seq raw read data of drought stressed root tissues from two different genotypes of chickpea, namely ICC37 and ICC506, were downloaded from SRA database available at NCBI. The RNA-seq read data are available for control and stress conditions under the following accession numbers: SRX048918, SRX048919, SRX048915, and SRX048917.

2.13.3 EST data

In addition to RNA-seq data analysis, another methodology was also followed to obtain transcriptional evidence for analyzing the expression pattern of stress genes. A blastn search was carried out using the coding sequences of each identified gene against the NCBI chickpea EST database (http://www.ncbi.nlm.nih.gov/nucest/?term=Cicer+arietinum). The number of chickpea EST till date is 46120 (from GenBank in 14-April-2013; 46120 EST sequences). The sequence identity cut off was set to > 90% to match an EST to map over a gene model keeping Expectation threshold value of 1.

2.14 Domain identification and motif analysis

The various domains present in nucleotide binding site-leucine rich repeat proteins were identified using hmmscan search in HMMER against the pfam database using gathering threshold and default E-value and Bit score parameter (http://hmmer.janelia.org/search/hmmscan). Sequence motifs in the three domains of NBS-LRR proteins were predicted with the help of MEME suite in which the minimum and maximum width of the motif was set to 15 and 50 in order to search for a maximum of 20 motifs with zoops model (Bailey *et al*. 2009; Bailey and Elkan 1994).

The pfam domain and signal peptide in the identified proteases were predicted by employing a Simple Modular Architecture Research Tool (SMART) (Jörg *et al*, 1998; Letunic *et al*, 2012). The diagram of the protein structure and domain architecture was generated with the Domain Graph 2.0 (DOG) software (Jian *et al*, 2009).

2.15 Pseudogene analysis

The pseudogenes were identified by performing a tblastn search in NCBI using the consensus sequences of TNL and non-TNL, mentioned in the gene identification section, against the chickpea chromosome assembly with an E-value cut off of 1. A gene is identified as pseudogene if tblastn translation with respect to the consensus sequence has at least one stop codon.

2.16 Promoter analysis

The 2 kb upstream regions of the NBS resistance genes and few other classes of stress genes with major function in plant defense response were selected and then screened against the PLACE database to identify the promoter regions (Higo *et al*, 1998). Four *cis*-regulatory elements namely WBOX (TGAC(C/T)), CBF (GTCGAC), DRE (G/A)CCGAC, and GCC (GCCGCC) are known to be involved in regulation during stress condition and disease resistance response were selected for detailed analysis. They were found to be overrepresented in the promoter region of disease resistance NBS genes.

The promoter analysis of other stress proteins responsible for inducing heat and freezing tolerance, water deficit situation, salinity, wounding, oxidative stress, and resistance against biotic agents was carried (Trivedi *et al*, 2013). The *cis*-elements analyzed are listed in Table 2.2.

Table 2.2: Various *cis*-regulatory elements present in the different classes of stress genes.

Stress protein	*Cis* elements	Response
Chitinase	WBOX, GCC box, ASF1MOTIFCAMV, GT1GMSCAM4	Pathogenesis, Disease resistance
Glucanase	WBOX, GT1GMSCAM4, ASF1MOTIFCAMV AGC box, GCC BOX	Pathogenesis, Disease resistance
Thaumatin	WBOX, GT1GMSCAM4, ASF1MOTIFCAMV	Pathogenesis, Disease resistance
Heat shock proteins	HSE, CCAATBOX1	Heat response
LEA	ABRE, MYB1AT, RaV1AAT, MYBCORE	Water deficit, Salinity Dehydration Cold response Drought
LTPs	WBOX, ASF1MOTIFCAMV, GT1GMSCAM4	Pathogenesis
Peroxidase	WBOX, G-box, WBOXNTERF3	Pathogenesis Oxidative stress

Chapter 3

Structure-function studies of UDP-glycosyltransferase family in plants

Glycosyltransferases catalyze the transfer of sugar moiety from an activated nucleotide sugar donor molecule to a saccharide or nonsaccharide acceptor to synthesize oligosaccharides, polysaccharides or glycoconjugates (Taniguchi *et al*, 2003) (Figure 3.1). It belongs to an essential ubiquitous multigene family present in bacteria, fungi, animals, plants etc. They perform glycosylation of important plant products which helps in their proper functioning as well as survival in adverse situations. Sugar donors are activated nucleotide molecules with a variable sugar part and an invariant UDP group. The sugar acceptors are generally flavonoids consisting of flavanones, flavanonol, flavans, flavones, flavonols, and anthocyanidins. Flavonoids are one of the classes of secondary metabolite possessing anti-cancer and anti-oxidant properties (Kanadaswami *et al*, 2005). These molecules play an important role in biotic as well as abiotic stresses. Stress conditions not only trigger the biosynthetic pathways, but also the expression of proteins, such as glycosyltransferase, involved in flavonoid transport and accumulation. Furthermore the regiospecific glycosylation of flavonoids alter their biochemical and pharmaceutical properties to develop more effective drugs and stress tolerant variety of the plant.

UDP-glycosyltransferases (UGTs) is a specific class of GTs which transfers the glycosyl group (glucose, galactose, xylose, rhamnose, etc) from a nucleoside diphosphate sugar donor (UDP sugar) to a diverse range of acceptor (Wang, 2009). In addition to the specific sugars they transfer, some UGTs are also highly specific to acceptor molecules and they catalyze glycosylation of only one or two types of molecules (Shao *et al*, 2005), whereas some glycosylate broad range of acceptors (Osmani *et al*, 2009). Flavonoids possess a polyphenolic framework bearing multiple –OH groups (7-OH, 5-OH, 3-OH, 4'-OH etc) on the phenyl rings (Figure 3.1). Some UGTs specifically glycosylate only one of these -OH groups, whereas others glycosylate multiple -OH groups. Therefore, due to such a diverse glycosylation preference of these enzymes, it is important to study the sequence and structural features of UGTs deciding the substrate specificity and selectivity pattern.

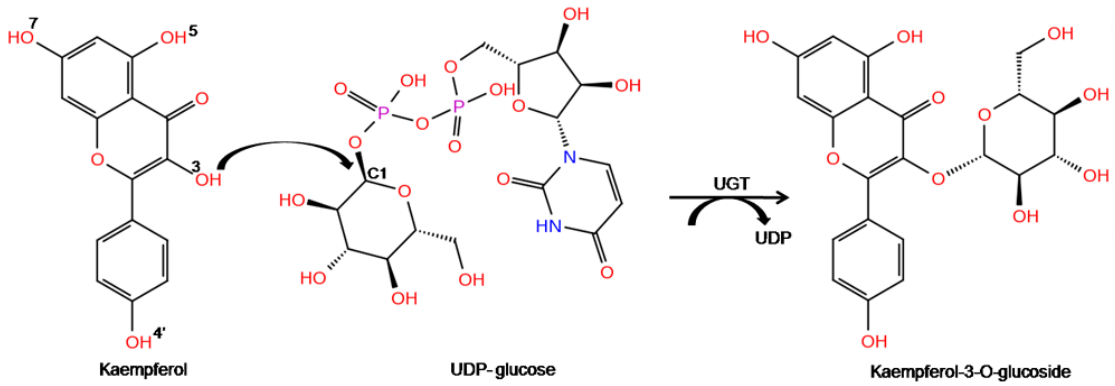


Figure 3.1: Reaction catalyzed by UGT is shown with Kaempferol and UDP-glucose as substrates that react to form Kaempferol-3-O-glucoside along with the release of UDP moiety. The numbering of the OH groups- 7-OH, 5-OH, 3-OH and 4'-OH in flavonoids is shown.

To accomplish this task, a phylogenetic, sequence and structural analysis of GT-B family of enzymes from various plant sources has been carried out to identify motifs and crucial amino acids which play an important role in the binding and recognition of specific acceptors, also to identify amino acid residues involved in maintaining the 3-D structure.

3.1 Protein sequence based phylogenetic analysis of GTs

After creating a dataset of 101 protein sequences of plant GTs, a dendrogram was built using neighbor joining method of MEGA in order to study their evolutionary relation (CD-Table S-3.1). Bootstrapping (value: 1000) was performed to examine how often a particular cluster in a tree appeared on re-sampling amino acids.

The GT sequences clustered into few distinct groups in the output of phylogenetic analysis. Out of them, the maximum number of GTs clustered in the F3GT specific clade with a significantly high bootstrap value (Figure 3.2). The members of the remaining clusters showed mixed specificity/activity for different -OH groups of acceptors. Here, the focus of study is the F3GT specific cluster, as the members of this group are closely related to each other showing high sequence identity and present in significant numbers. The percent identity matrix was prepared employing the ClustalX program. The matrix revealed a significant level of conservation with approximate 40% sequence identity among them (CD-Table S-3.2).

3.2 Molecular modeling

Protein sequences of 30 plant F3GTs were retrieved from UniProtKB/ Swiss-Prot database (Table 3.1). The blast search resulted in three homologous crystal structures from *Vitis vinifera* (2C1Z), *Medicago trucatula* (3HBF), and *Clitoria ternatea* (3WC4) of resolution 1.90 Å, 2.10 Å, and 1.85 Å respectively, sharing an identity of ≥ 40% and query coverage between the target and template was > 90%. The templates were chosen based on high sequence identity and better resolution.

Table 3.1: Details of 30 plant F3GTs are enlisted in the table below. The last column gives the SwissProt ID of the respective sequences.

Source	Protein name	SwissProt ID
Medicago truncatula	UDP flavonoid 3-O-glucosyltransferase	G7JD22
Vitis vinifera	Anthocyanidin 3-O-glucosyltransferase 2	P51094
Populus trichocarpa	UDP-glucose:flavonoid 3-o-glucosyltransferase	B9MW26
Fragaria ananassa	Anthocyanidin 3-O-glucosyltransferase 2	Q5UL10
Litchi chinensis	UDP-glucose flavonoid glucosyl-transferase	F5CET4
Prunus persica	UFGT	J7H3K6
Eustoma exaltatum subsp. russellianum	Flavonoid 3-O-galactosyltransferase	A4F1Q1
Aralia cordata	Anthocyanin 3-O-galactosyltransferase	Q76G23
Actinidia chinensis	Flavonoid 3-0-galactosyltransferase	E5D2U2
Diospyros kaki	Flavonoid 3-O-galactosyltransferase	C8KH59
Populus trichocarpa	UDP-glucose:flavonoid 3-o-glucosyltransferase	B9I672
Garcinia mangostana	UDP-glycose flavonoid 3-O-glycosyltransferase	B9UZ54
Arabidopsis thaliana	UDP-glycosyltransferase 78D2	Q9LFJ8
Rosa hybrid cultivar	UDP-glucose: anthocyanin 3-glucosylltransferase	A8CDW6
Citrus paradisi	Flavonol 3-O-glucosyltransferase	C5MR71
Dianthus caryophyllus	UDP-glucose:flavonoid 3-O-glucosyltransferase	Q60FF2
Petunia hybrida	Kaempferol 3-O-beta-D-galactosyltransferase	Q9SBQ8
Lobelia erinus	UDP-glucose:anthocyanidin 3-O-glucosyltransferase	A4F1S9

Rosa hybrid cultivar	UDP-glucose: flavonol 3-O-glucosyltransferase	Q2PGW4
Rosa hybrid cultivar	UDP-glucose:flavonol 3-glucosyltransferase	A8CDW8
Populus trichocarpa	UDP-galactose:flavonol 3-o-galactosyltransferase	B9NFI8
Forsythia intermedia	Flavonoid 3-O-glucosyltransferase	Q9XF16
Perilla frutescens	Flavonoid 3-O-glucosyltransferase	O04114
Gentiana triflora	Anthocyanidin 3-O-glucosyltransferase	Q96493
Solanum tuberosum	Flavonoid 3-glucosyl transferase	Q3YK56
Dianthus caryophyllus	Glucosyltransferase	Q60FF0
Vigna mungo	Flavonoid 3-O-galactosyl transferase	Q9ZWS2
Clitoria ternatea	UDP-glucose:anthocyanin 3-O-glucosyltransferase	A4F1R4
Iris hollandica	Anthocyanidin 3-O-glucosyltransferase	Q5KTF3
Hordeum vulgare	Anthocyanidin 3-O-glucosyltransferase	P14726

3.2.1 Molecular models of F3GT proteins

The molecular structures of F3GT proteins were modeled in order to study, from a structural point of view, the role of conserved regions in the close vicinity of acceptor and their interactions that favor formation of protein-ligand complex. BLAST search against PDB identified homologous structures to be used as templates to build the homology models (Table 3.2). Crystal structures of GTs from *V. vinifera* (PDB-2C1Z), *Medicago truncatula* (PDB-3HBF) and *Clitoria ternatea* (PDB-3WC4) of resolution 1.90 Å, 2.10 Å and 1.85 Å, respectively, shared an identity of ≥ 40% (except Iris_2 and Hordeum) and query coverage of ~ > 90% with the target proteins, and were thus used for modeling studies. Secondary structure prediction by Psipred showed the similarity between the templates and the target with respect to the arrangement of secondary structure elements. Very few gaps were present in the alignment of the target and template sequences (Figure 3.3, CD-Figure S-3.1). The bound ligands of 2C1Z i.e., kmp and U2F were incorporated in the target models during the modeling procedure. Few N-terminal residues (4-5 amino acids) couldn't be modeled for some of the target sequences because of the absence of corresponding regions in the templates (Missing residues in 2C1Z: 1-5, 3HBF: 1-11).

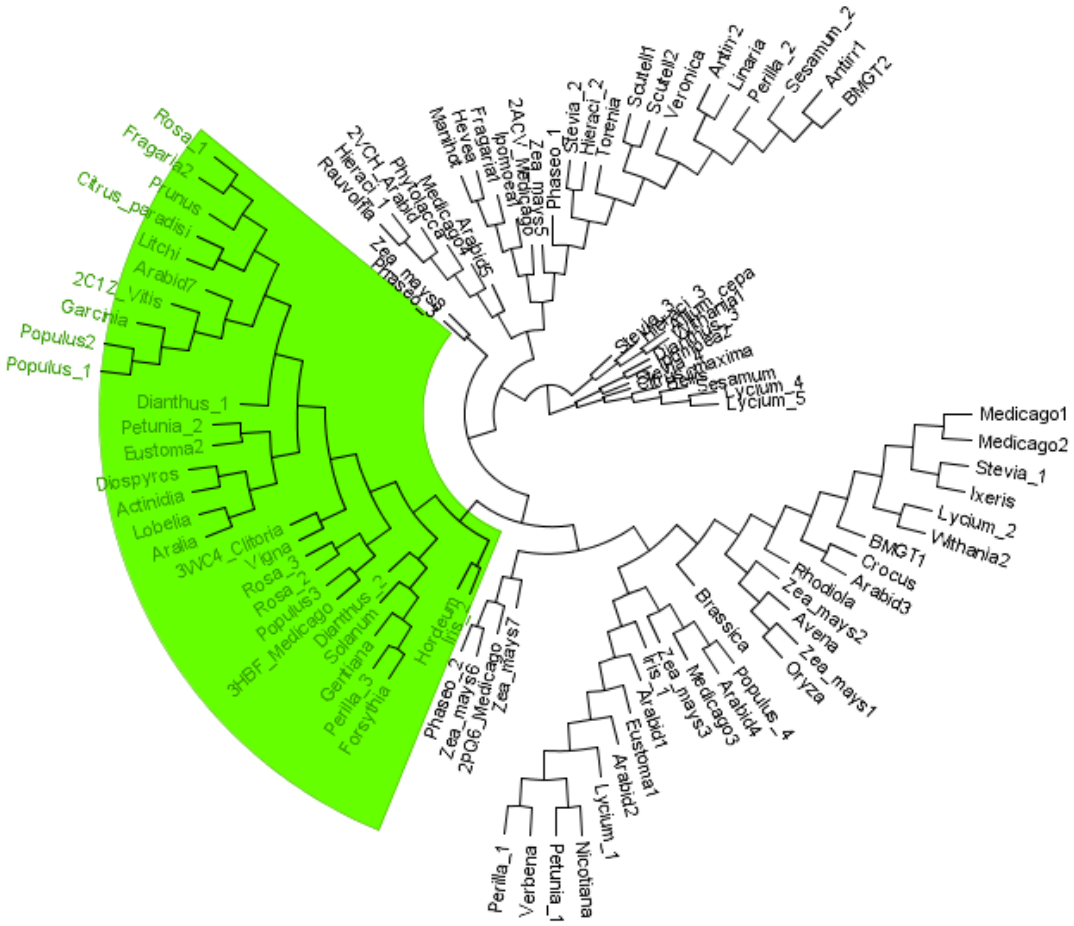


Figure 3.2: Dendrogram of 101 UGT protein sequences shows distinct clusters of F3GT members marked with an arrow. The underlined taxon names are annotated to be flavonoid-3-OH glycosyltransferase and they form a separate clade.

Table 3.2: Blast search statistics of F3GT protein sequences with their respective selected templates. T1 and T2 refer to the two templates used for modeling the structure of respective proteins.

No.	UGT	T1	% identity	E-value	T2	% identity	E-value
1	Populas2	2C1Z	61	$2.0*10^{-160}$	3HBF	49	$5.3*10^{-121}$
2	Garcinia	2C1Z	59	$2.2*10^{-154}$	3HBF	49	$2.4*10^{-120}$
3	Arabid7	2C1Z	60	$1.7*10^{-156}$	3HBF	51	$7.9*10^{-127}$
4	Rosa_1	2C1Z	56	$8.1*10^{-141}$	3HBF	44	$4.1*10^{-106}$
5	Citrus_P	2C1Z	56	$7.8*10^{-144}$	3HBF	45	$4.5*10^{-111}$
6	Dianthus_1	2C1Z	49	$3.7*10^{-120}$	3HBF	41	$8.7*10^{-94}$
7	Petunia_2	2C1Z	51	$1.8*10^{-129}$	3HBF	48	$3.7*10^{-115}$
8	Lobelia	2C1Z	54	$4.9*10^{-135}$	3HBF	48	$3.7*10^{-115}$
9	Rosa_2	2C1Z	46	$2.8*10^{-115}$	3HBF	52	$3.1*10^{-124}$
10	Rosa_3	2C1Z	40	$3.8*10^{-84}$	3HBF	44	$4.0*10^{-91}$
11	Populas3	2C1Z	45	$1.5*10^{-106}$	3HBF	49	$1.9*10^{-120}$
12	Forsythia	2C1Z	49	$1.0*10^{-123}$	3HBF	46	$4.0*10^{-104}$
13	Perilla_3	2C1Z	46	$5.0*10^{-118}$	3HBF	43	$9.4*10^{-97}$
14	Gentiana	2C1Z	47	$9.3*10^{-112}$	3HBF	46	$1.5*10^{-108}$
15	Solanum	2C1Z	47	$6.6*10^{-110}$	3HBF	45	$2.8*10^{-104}$
16	Dianthus_2	2C1Z	44	$5.2*10^{-111}$	3HBF	42	$1.4*10^{-97}$
17	Vigna	3WC4	57	$3.9*10^{-144}$	3HBF	44	$4.9*10^{-97}$
18	Populas_1	2C1Z	61	$9.6*10^{-167}$	3HBF	48	$1.8*10^{-122}$
19	Fragaria	2C1Z	58	$9.8*10^{-147}$	3HBF	47	$1.5*10^{-111}$
20	Litchi	2C1Z	57	$2.5*10^{-153}$	3HBF	45	$2.5*10^{-105}$
21	Prunus	2C1Z	55	$6.8*10^{-143}$	3HBF	45	$1.5*10^{-107}$
22	Eustoma	2C1Z	54	$2.4*10^{-140}$	3HBF	44	$6.8*10^{-105}$
23	Aralia1	2C1Z	53	$2.9*10^{-141}$	3HBF	48	$3.6*10^{-116}$

24	Actinidia	2C1Z	55	$1.2*10^{-143}$	3HBF	51	$8.8*10^{-122}$
25	Diospyros	2C1Z	50	$4.0*10^{-126}$	3HBF	44	$6.8*10^{-99}$
26	Iris_2	2C1Z	37	$1.4*10^{-88}$	3HBF	36	$4.7*10^{-83}$
27	Hordeum	2C1Z	36	$7.0*10^{-65}$	3HBF	33	$3.5*10^{-60}$

3.2.2 Model validation and refinement

The initial models were evaluated for their stereochemical parameters. More than 99% of all residues were in the allowed regions of Ramachandran plot. The goodness factor of the models, a log odd score calculated based on the stereochemical parameters, were in the range of 0 to -0.4. Verify3D plot for the models showed that more than 95% of the residues had a positive 3D-1D averaged score, which revealed that the models were folded correctly. RMSD of the C-alpha atoms of the models and the templates were close to 1Å. The z-score value calculated by ProSA server is within the z-scores of experimentally determined PDB structures. The ERRAT results also showed that the overall quality score of the models was good (Figure 3.4, Table 3.3).

The model refinement phase involved preprocessing the initial models by adding hydrogens, assigning bond order, and filling missing loops and side chains. Next, the models were subjected to restrained minimization by applying the constraint to converge the non-hydrogen atoms to an RMSD of 0.3 Å using OPLS 2005 force field. After that, the models were further subjected to 500 steps of steepest descent energy minimization followed by 1000 steps of conjugate gradient energy minimization using the same force field. These energy minimized models were further used for docking and molecular dynamics studies.

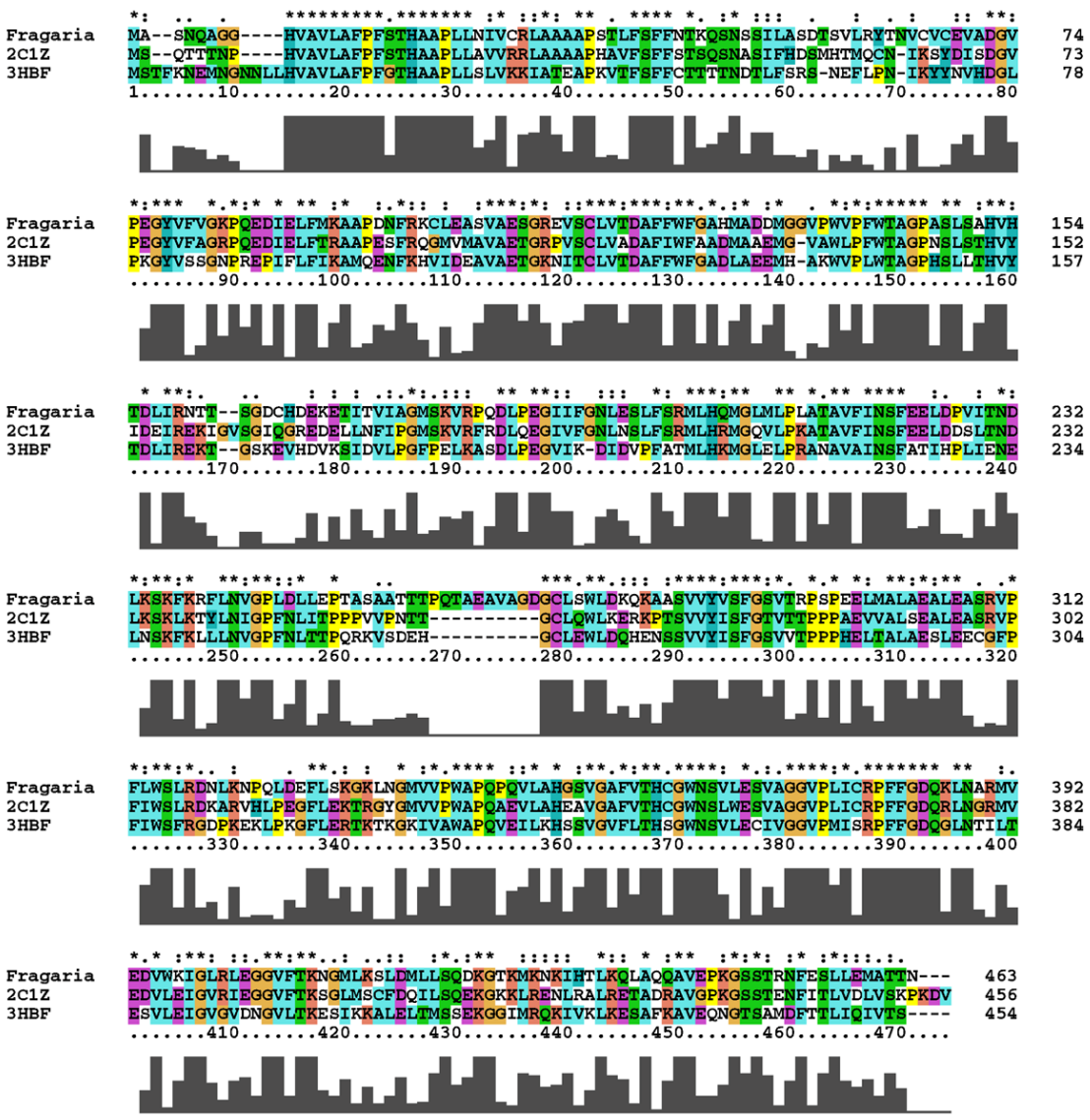


Figure 3.3: Multiple sequence alignment of *Fragaria* UGT sequence and the two templates employed for the modeling studies.

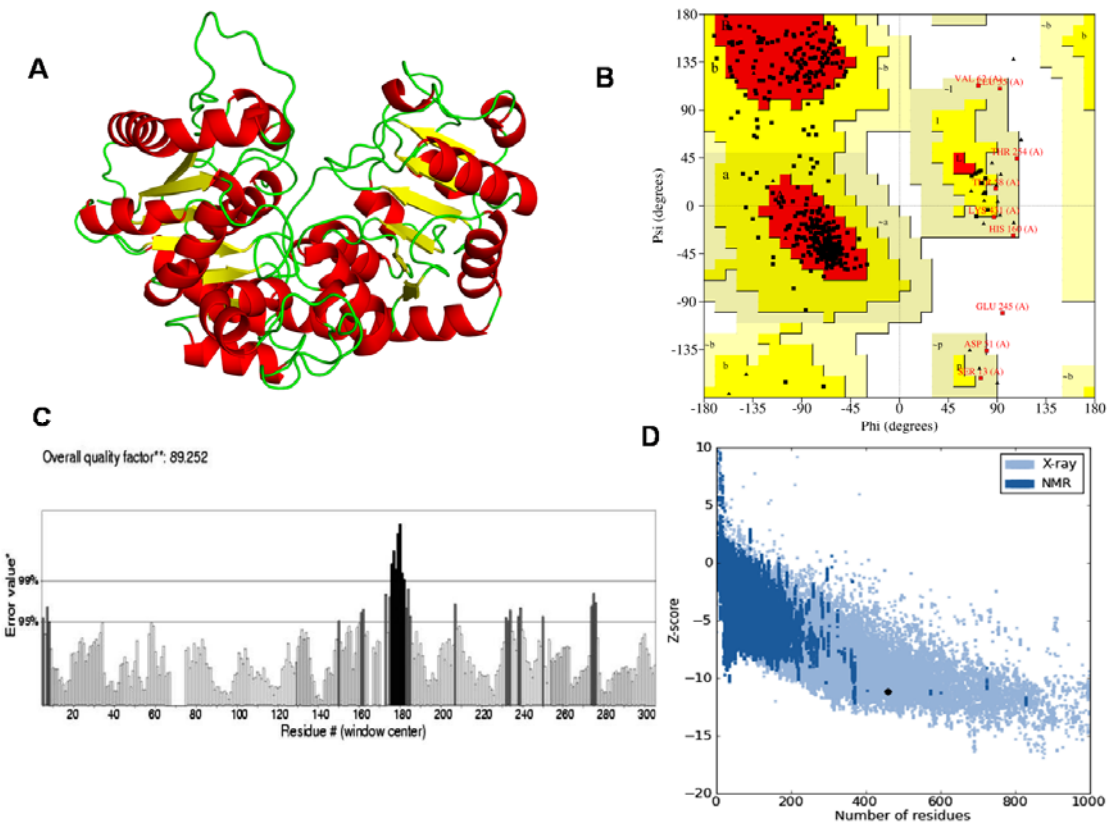


Figure 3.4: A: Molecular model of *Fragaria* UGT prepared using pdb structure 2C1Z and 3HBF as templates. B: Ramachandran plot C: ERRAT plot D: Overall model quality plot generated using ProSA showing the Z-score of the model as black dot.

Table 3.3: Model evaluation statistics of 30 F3GT protein sequences.

No	Protein source	Ramachandran plot	ProSA	ERRAT	Verify3D
1	Populas_1	100*, 0.0** (0.0)***	-10.7	85.05	98.20
2	Populas_2	100, 0.0 (-0.4)	-11.9	83.22	100.00
3	Aralia	100, 0.0 (0.0)	-10.75	90.13	99.33
4	Fragaria2	99.5, 0.5 (-0.2)	-11.17	87.67	92.37
5	2C1Z_Vitis	100, 0.0 (0.1)	-10.96	94.17	100
6	Actinidia	100,0.0 (0.0)	-11.19	84.06	100
7	Eustoma	99.5, 0.5 (0.0)	-11.47	88.86	97.33
8	Lobelia	99.7, 0.3 (-0.1)	-10.51	79.8	97.10
9	Rosa_1	98.8, 0.2 (0.0)	-11.37	90.11	91.09
10	Petunia_2	99.5, 0.5 (0.0)	-11.29	81.73	98.66
11	Litchi	99.7, 0.3 (0.0)	-11.68	81.79	97.98
12	Arabid7	99.5, 0.5 (0.0)	-10.88	99.63	97.78
13	Citrus_paradisi	99.5, 0.5 (-0.1)	-11.77	83.18	99.55
14	Rosa_2	99.5, 0.5 (-0.1)	-11.13	81.7	99.77
15	Garcinia	99.5, 0.5 (-0.2)	-10.71	89.70	96.44
16	Prunus	98, 2.0 (-0.1)	-10.72	83.90	99.33
17	Diosporus	99.7, 0.3 (0.6)	-10.79	81.75	95.31
18	Perilla_3	99.5, 0.5 (-0.1)	-11.26	81.79	97.97
19	Forsythia	98.5, 1.5 (0.0)	-12.05	84.2	100
20	Solanum	98.2, 1.8 (-0.1)	-11.13	77.93	97.48
21	Gentiana	98.7, 1.3 (0.0)	-10.85	83.52	98.43
22	3HBF_Medicago	100, 0.0 (0.3)	-11.25	95.86	97.97
23	Vigna	98.7, 0.3 (-0.1)	-10	70.27	96.17
24	Populas3	99.7, 0.3 (-0.3)	-10.62	81.10	99.10
25	Rosa_3	99.2, 0.8 (0.1)	-10.73	71.66	100.00
26	Dianthus_1	99.7, 0.3 (0.0)	-12.06	84.93	99.11
27	Dianthus_2	99.0, 1.0 (-0.1)	-10.74	88.24	98.66

28	Iris 2	98.4, 1.6 (-0.1)	-11.43	72.74	94.13
29	Clitoria	99.2, 0.8 (-0.1)	-10.13	68.66	98.87
30	Hordeum	98.7, 1.3 (-0.1)	-10.66	87.96	100.00

***Represent % of the total residue present in allowed region of Ramachandran plot.**

**** Represent % of the total residue present in disallowed region of Ramachandran plot**

*****Number in brackets denotes the value of goodness factor for the models.**

The values in the fourth, fifth, and sixth column represent the Z-score, ERRAT, and Verify3D scores of F3GT models.

3.3 Sequence conservation and structural integrity of F3GT enzymes

MSA of the F3GT proteins was carried out using ClustalX program (CD-Figure S-3.2). Sequence alignment and construction of percent identity matrix showed higher sequence conservation of both CTD and NTD (CD-Figure S-3.2 and CD-Table S-3.2). The residue numbers used in this analysis are with respect to the sequence of transferase from *V. vinifera* source (PDB-2C1Z). An analysis of the conserved residues using the *V. vinifera* GT structure revealed that these residues interacted through hydrophobic, π-π stacking and hydrogen bonding interactions for maintaining the required framework of interacting residues in the binding site.

One key conserved residue involved in catalysis is His20, which steers the deprotonation of acceptor substrates. An important interaction near the binding site is the side chain-side chain hydrogen bonding interaction between His9 and Ser41, which in turn helps Phe42 to make a π-π interaction with Phe53. Similarly, Phe40 is surrounded by hydrophobic amino acids of the adjacent α-helix and β-sheets that provide stability to the folded structure. Pro16 present near the acceptor binding site provides rigidity to the binding site loop and helps in orienting the neighboring residues Phe15 and Phe17 to provide a hydrophobic environment to the acceptor. Similarly, the conserved Gly72 residue provides flexibility to a binding site loop. Another important interaction involves residue Pro74 that orients the side chain of Glu75, making it accessible to the solvent. Phe90 is surrounded by a series of hydrophobic and aromatic amino acids such as Phe15, Pro16, Phe17, Ile69, Phe98, etc. Side chain of Asn49 interacts with Ser46 and Asp68 of neighboring β-sheet. The interaction of Asp119 with the key catalytic residue His20 balances its acquired charge after deprotonation of the acceptor. Phe121 provides hydrophobic environment to one of the phenyl rings of the acceptor molecule. Conserved Trp123 and Phe124

are involved in π-π interaction while Pro95 forms a stacking interaction with Phe124. The indole ring of Trp140 interacts with Gly143 and Ser146, its aromatic ring providing a hydrophobic environment to the acceptor substrate. His150 being a polar amino acid provides a hydrophilic environment to the 4'-OH group of the flavonoid acceptor. Arg157 connects the other loop by making hydrogen bonds with residues Gly190 and Ile191 and also interacting with the solvent molecule. Leu204 also contributes to the hydrophobicity of the acceptor binding pocket. Conserved Asn220 and Ser221 stabilize a gamma turn that folds the polypeptide chain, bringing together and allowing interactions between regular secondary structure elements (Figure 3.5a).

The CTD shows a higher level of sequence conservation and binds the nucleotide sugar donor substrate. Side chain of Tyr275 interacts with the conserved Gln335 residue and helps in maintaining the structure by connecting the β-sheet with the α-helix. In addition, conserved Phe278 is involved in hydrophobic interactions with Pro284, Phe371 and Phe397. The residues Glu288 and Arg369 are involved in ionic interactions that stabilize the structure by connecting the α-helix to a coil. Pro334 present at the beginning of the α-helix helps the neighboring Gln335 to interact with Tyr275 of a proximate β-sheet. Trp332 stacks with the uridine ring of the UDP-sugar substrate. Residues Thr19, Ala333, His350, Trp353, Asn354, Ser355, Glu358, Asp374 and Gln375 also interact with the sugar donor. Asn378 interacts with Asp374 and Asn145 thus helping in maintaining the 3-D structure by connecting two secondary structural elements (Figure 3.5b). Conserved amino acid residues with less bulky side chains provide required flexibility to the loops surrounding the binding site. These features are conserved or semi-conserved in most of the proteins of the F3GT cluster.

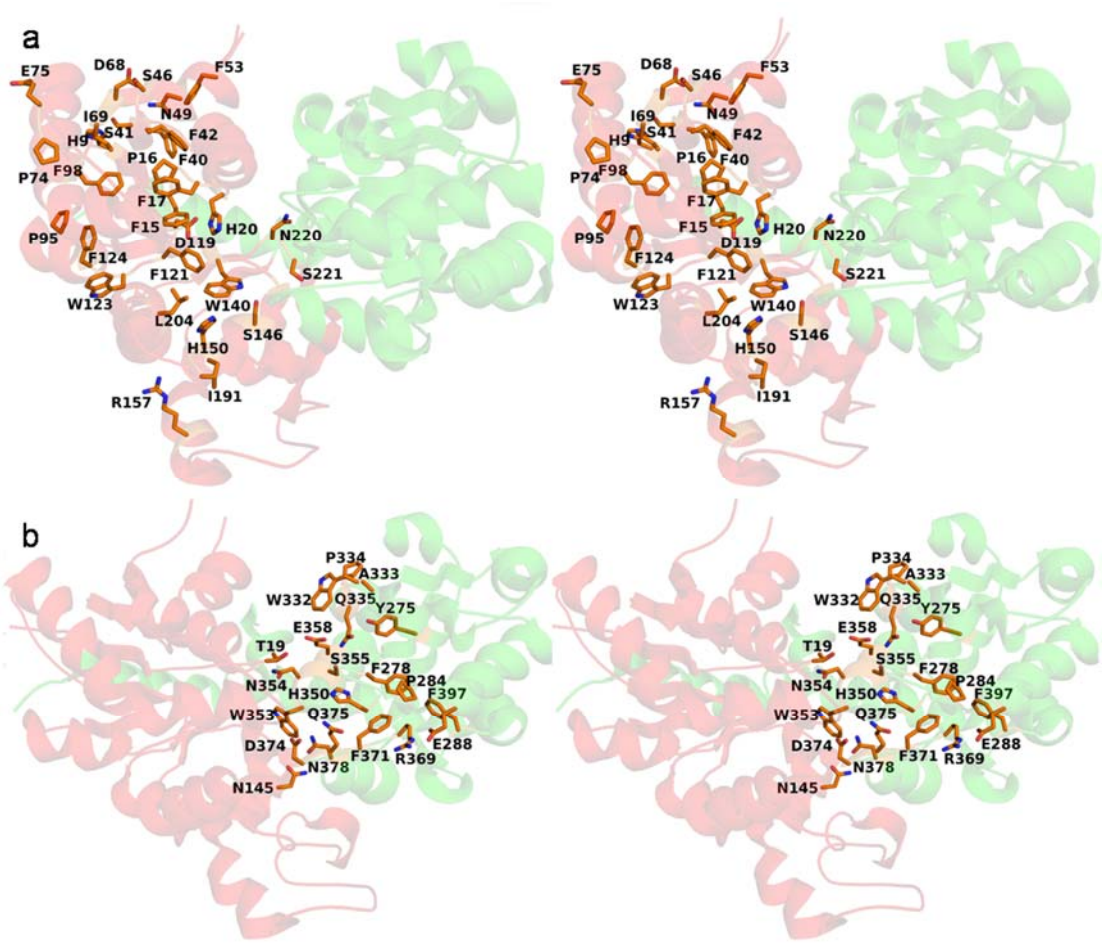


Figure 3.5: a: Stereo image representing cartoon drawing of UGT from *Vitis vinifera* with conserved amino acids of the N-terminal domain shown in stick form. b: Stereo image representing cartoon drawing of UGT from *Vitis vinifera* with conserved amino acids of the C-terminal domain shown in stick form.

3.4 Interaction studies using modeled complexes

3.4.1 Interaction between enzyme and substrates

Prior to the docking, the models and the ligands were first pre-processed to check for any problems in the structures related to missing hydrogens, side chains, improper bond orders, overlapping atoms etc. The refined sugar donor substrate was docked in the binding site cavity of the processed target by using standard precision mode of Glide. The bound ligand of template was used to mark the binding site by creating a grid around it. A total of 20 docked complexes were generated out of which the one with the highest glide score and forming all possible interactions with the conserved residues of PSPG motif was selected as the best pose (Table 3.4).

After the preparation of acceptor through LigPrep, kmp was docked in the acceptor binding pocket of the protein-UDP sugar complex by placing the grid around the reference ligand i.e. kmp. Standard precision mode of glide was used for docking studies. Similarly, 20 docked poses were generated and the one with a better glide score value and that interacted with the catalytic histidine was chosen as the best pose (Table 3.4). Docking studies have shown that the environment surrounding the acceptor is conserved or similar for all the docked structures. Some changes have been seen near the 5-OH and 4'-OH groups of kmp. In some structures 5-OH group is stabilized by the hydrophilicity provided by the side chain of Ser18 (w.r.t. PDB-2C1Z) while in some structures non-polar amino acids like glycine and alanine are present in that position. In the crystal structure 3HBF, glycine occupies this position but neighboring water molecule stabilizes the ligand molecule by forming a hydrogen bond with its 5-OH group. Amino acid Gln188 interacts with the O4' group of kmp. In the F3GT proteins this glutamine is replaced by proline (close to N6; CD-Figure S-3.2), so this interaction is lost in all the docked complexes. This is eventually compensated by another interaction formed by His150 (Region N5; CD-Figure S-3.2) (Figure 3.6). The glide score value was higher for the pose in which 3-OH group of kmp interacted with the catalytic histidine residue while the other poses in which 7-OH, 4'OH and 5-OH groups of kmp interacted with the histidine had a low glide score. Docking studies aided in the identification of eight conserved regions present in the loops and helix near the acceptor binding regions (N1-N6 regions in the NTD and C1-C2 regions in the CTD) which might decide their substrate/glycosylation preference (Figure 3.7 & 3.8). This observation suggests that the environment topology observed here favors the binding of flavonoids to facilitate the glycosylation of 3-OH group of flavonoids.

Four protein sequences of *Fragaria ananassa* (Swiss-Prot ID-Q2V6K0), *Manihot esculenta*, *Hevea brasiliensis* and *M. truncatula* (Swiss-Prot ID-Q5IFH7) were annotated as flavonoid-3-O-glycosyltransferases in the protein sequence database but they were present outside the F3GT cluster in the dendrogram (Figure 3.2). Modeling and docking studies of these proteins showed that their activity might be less towards the flavonoid-3-OH group as compared to the other annotated members of F3GT. This is an additional evidence to show that the environment

provided by the conserved acceptor binding site residues of F3GTs is more favorable for orienting the acceptor molecule for glycosylation at 3-OH group of the flavonoid molecule than sequences outside the F3GT cluster.

Table 3.4: Docking statistics (Glide score values) of sugar donor and acceptor substrates for 30 F3GT proteins.

No.	Protein source	Glide score	
		Sugar donor	Acceptor
1	Populas_1	-11.56	-9.07
2	Populas_2	-11.06	-8.84
3	Aralia	-9.52	-8.94
4	Fragaria2	-11.08	-9.28
5	2C1Z_Vitis	Crystal structure with bound ligands	
6	Actinidia	-10.98	-8.02
7	Eustoma2	-10.37	-8.84
8	Lobelia	-9.81	-7.63
9	Rosa_1	-10.31	-7.83
10	Petunia_2	-9.50	-9.52
11	Litchi	-10.79	-8.83
12	Arabid7	-11.40	-7.34
13	Citrus_paradisi	-11.52	-9.02
14	Rosa_2	-11.66	-7.58
15	Garcinia	-10.64	-8.40
16	Prunus	-10.96	-9.42
17	Diosporus	-9.54	-8.71
18	Perilla_3	-10.30	-8.45
19	Forsythia	-10.92	-7.02
20	Solanum	-11.25	-7.41
21	Gentiana	-10.63	-8.34
22	3HBF Medicago	-8.40	-7.34
23	Vigna	-9.78	-7.73
24	Populas3	-10.82	-8.98
25	Rosa_3	-9.32	-9.29
26	Dianthus_1	-10.51	-8.75
27	Dianthus_2	-10.84 -8.25	
28	Iris 2	-9.82	-8.28
29	Clitoria	-10.76	-7.63
30	Hordeum	-11.67	-7.89

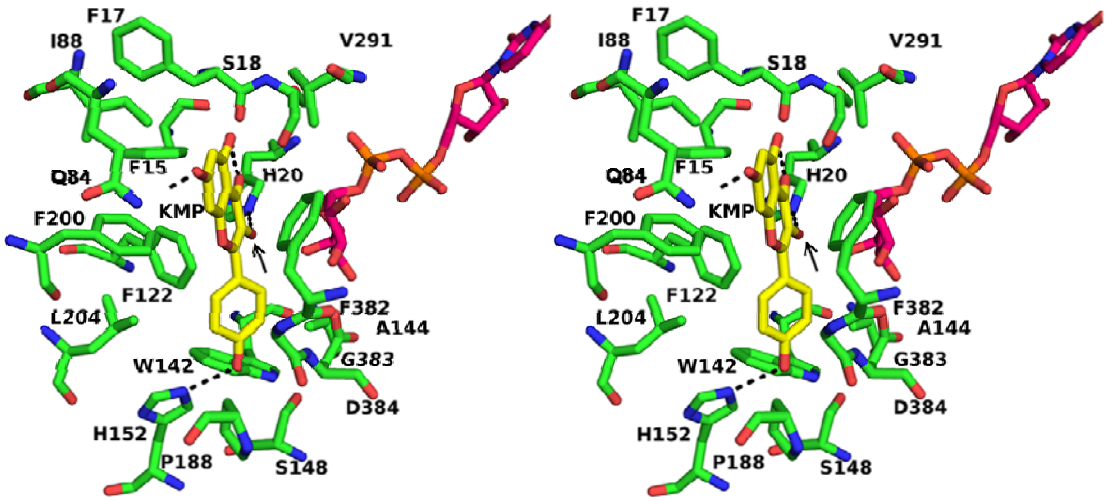


Figure 3.6: Stereo view of the docked complex of UGT from *Fragaria ananassa* with kaempferol shown in stick form. The arrow shows the 3-OH group of kaempferol which take part in glycosylation event. UDP-glucose is also shown in stick form.

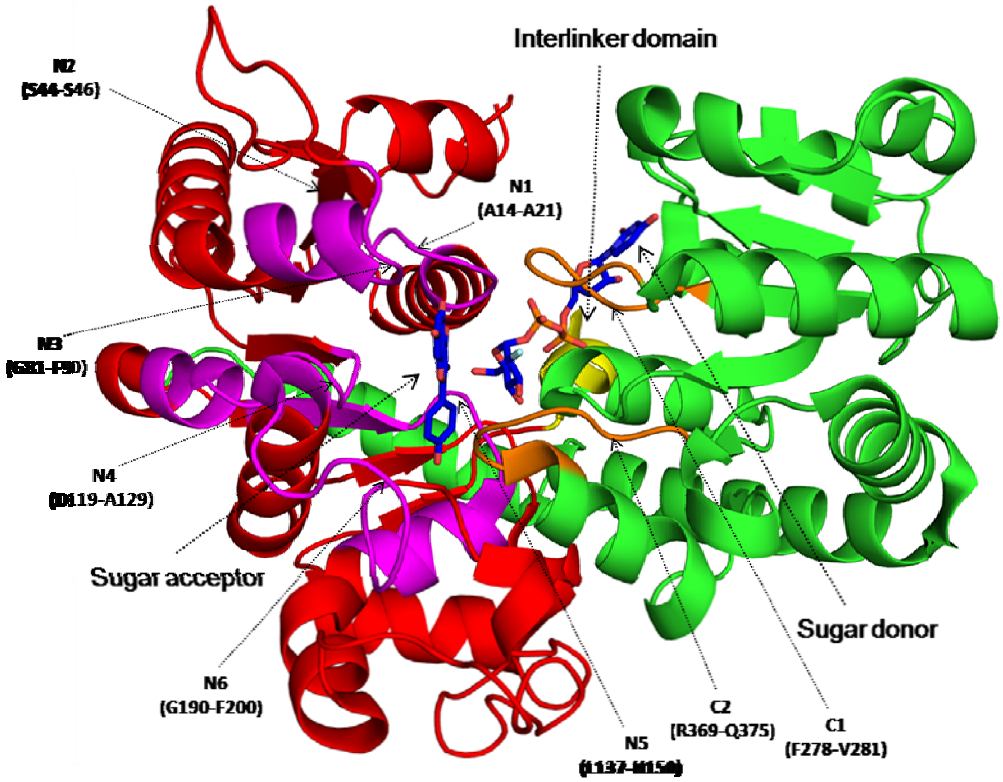
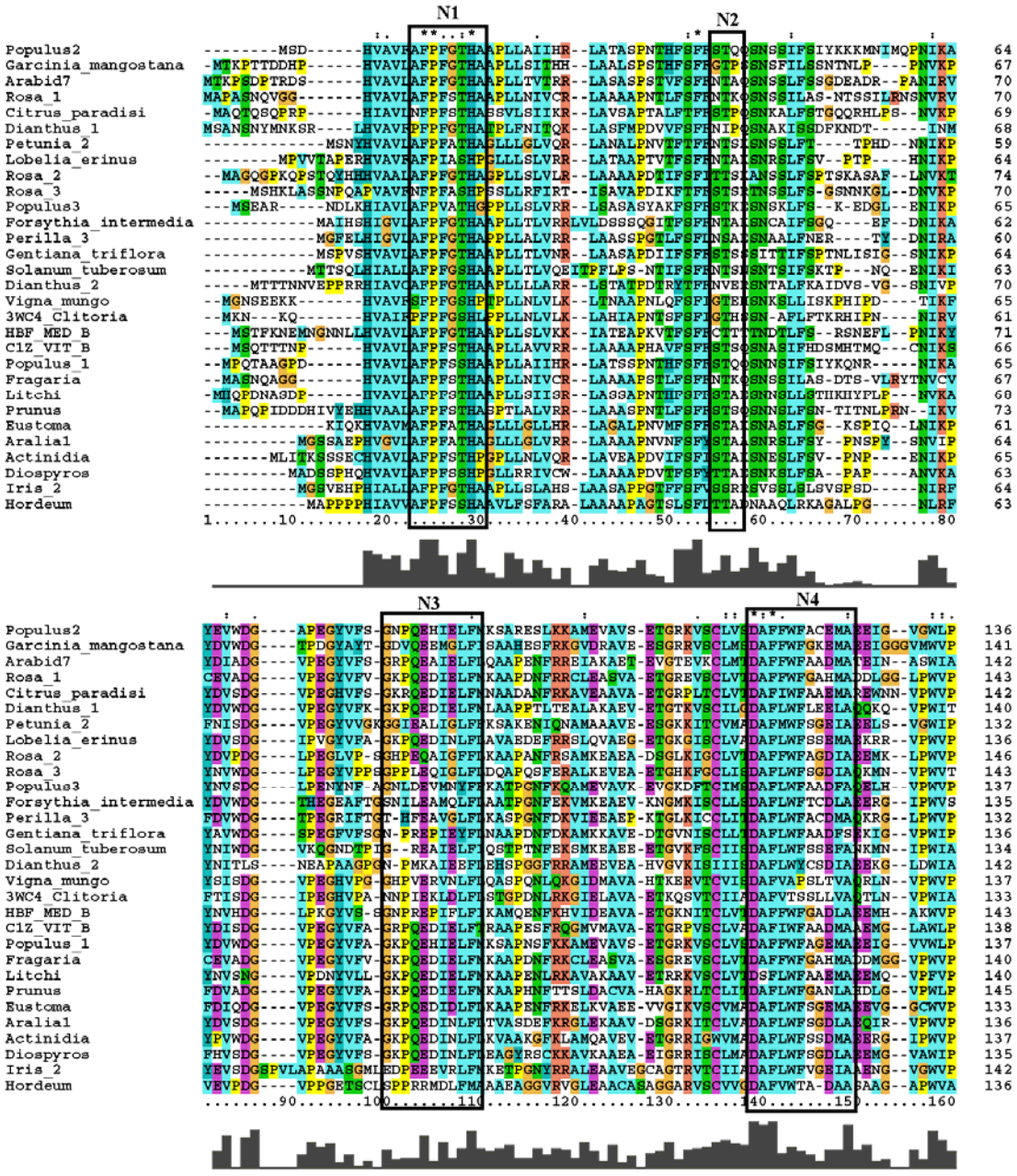


Figure 3.7: Ribbon view of UGT from *Vitis vinifera* with docked acceptor and sugar donor in stick form is shown. Six conserved regions from N1 to N6 at the NTD and two regions C1 and C2 at the CTD marked with an arrow plays a crucial role in holding the acceptor in the binding pocket.


N1
N2
N3
N4

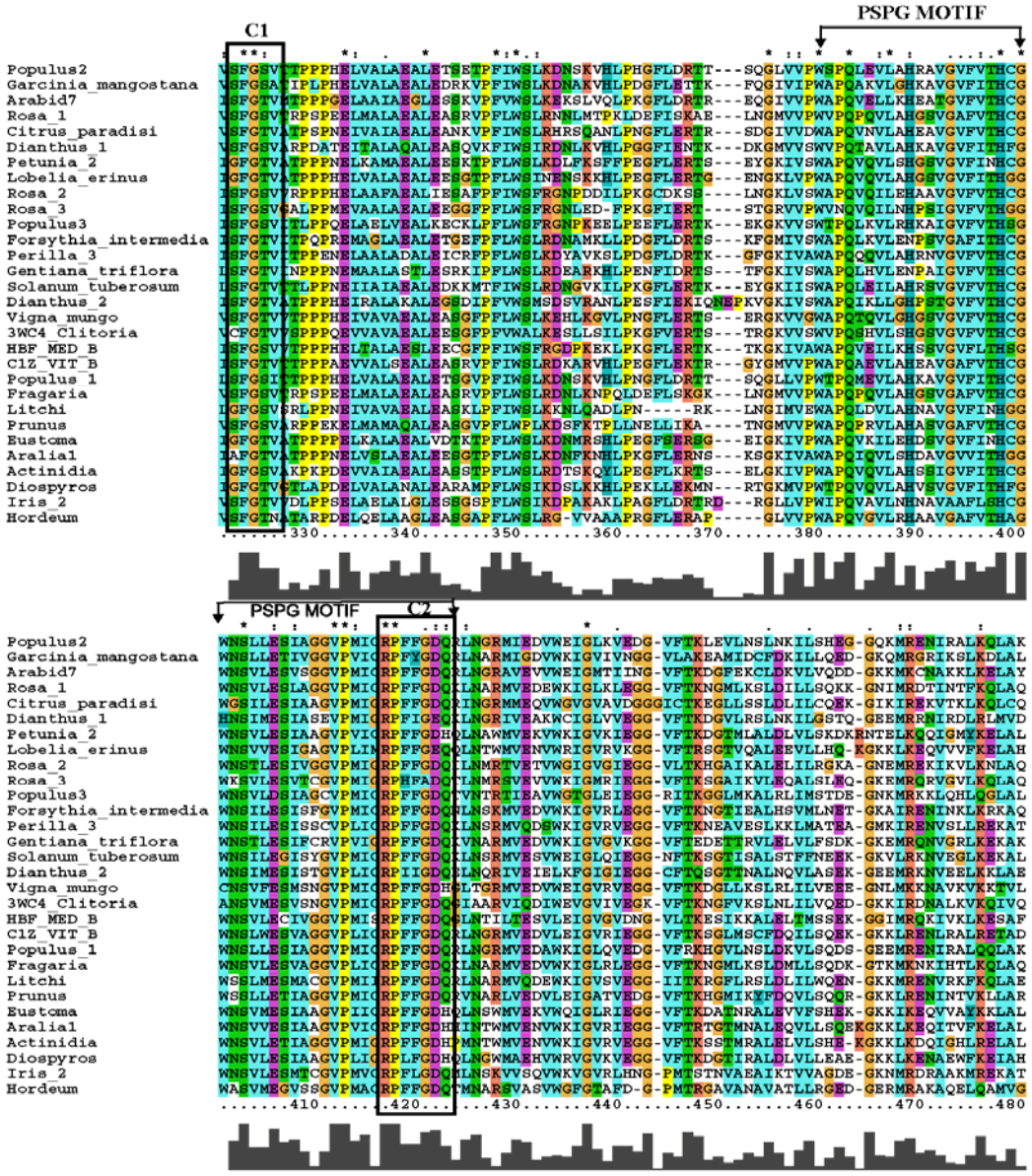


Figure 3.8: Multiple sequence alignment of 30 F3GT protein sequences to show the eight conserved regions at the N- and C- terminal domain involved in the binding of flavonoid acceptor.

3.4.2 Docking studies with reaction products: Kaempferol-3-O-glucoside and UDP

The crystal structure from *V. vinifera* (PDB-ID: 2C1Z) was taken as the target molecule to dock the glycosylated flavonoid, kaempferol-3-O-glucoside (KMG) and UDP in their respective binding pockets. Grid was created around the reference ligands, kmp (for KMG) and U2F (for UDP). The best pose for UDP and KMG had glide scores of -9.43 and -6.20 (Figure 3.9). The sugar moiety of KMG interacts with the sugar specificity determining amino acids of the PSPG motif.

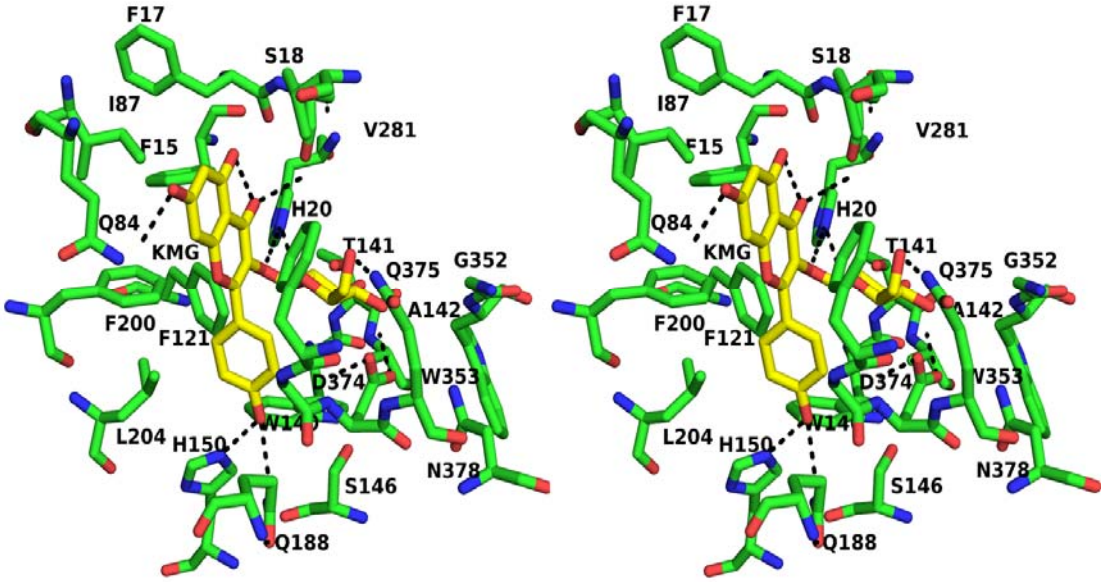


Figure 3.9: Stereo view of docked complex of UGT from *Vitis vinifera* (2C1Z) with Kaempferol-3-O-glucoside and UDP shown in stick form.

3.5 Molecular dynamics simulation of the docked complexes

Molecular dynamics simulations were performed on 31 docked complexes in order to check their stability. Details of box size and number of water molecules and ions added during simulation process are given in Table 3.5. Their trajectories were visually inspected for the binding of ligands and it was observed that the substrates were firmly bound in their respective binding pockets (CD-Video S1). Here we have explained the interactions and the stability patterns of the ligands for one of the F3GT members i.e. *F. ananassa* GT (Swiss-Prot ID-Q5UL10). The RMSD graph was drawn by assigning time in ps on the X-axis and RMSD values of the Cα atoms in nm on the Y-axis (Figure 3.10). From the graph, we can conclude that the RMSD of Cα atoms of

the generated structures over the complete trajectory is stable. To check the fluctuations of kmp in binding site, the distance between kmp and three key residues H20, Q84 and H152 which form hydrogen bonds with 3-OH, 5-OH and 4'OH group of kmp was calculated using g_mindist program of GROMACS. The graph showed that kmp fluctuates in the binding pocket within a distance of only 3 Å to 5 Å with respect to three key residue considered in the analysis for the last 2 ns of the trajectory (Figure 3.11). The life of hydrogen bond between the catalytic H20 and D120 remained for 70% of the total simulation time after equilibration. The root mean square fluctuation (RMSF) plot showed negligible fluctuations of the catalytic and binding site residues of the acceptor and donor (Figure 3.12). A highly fluctuating loop region, not part of the binding site region, was observed in the plot. Molecular dynamics simulation showed that the final products also form a stable complex (CD-Video S2).

Table 3.5: Details of box size and number of water molecules and ions added during simulation process

Serial number	F3GTs	Box dimension	Number of water molecules added	Number of ions added
1	*Populas_1*	90.78*90.78*90.78	15262	8
2	*Populas_2*	93.02*93.02*93.02	16627	5
3	*Aralia*	94.03*94.03*94.03	16935	7
4	*Fragaria2*	92.87*92.87*92.87	16478	10
5	*2C1Z_Vitis*	94.05*94.05*94.05	16975	6
6	*Actinidia*	91.14*91.14*91.14	15417	3
7	*Eustoma*	90.43*90.43*90.43	15150	7
8	*Lobelia*	95.64*95.64*95.64	18323	9
9	*Rosa_1*	91.78*91.78*91.78	15777	8
10	*Petunia_2*	90.42*90.42*90.42	15196	15
11	*Litchi*	94.80*94.80*94.80	17030	3
12	*Arabid7*	92.55*92.55*92.55	16267	9
13	*Citrus_paradisi*	93.54*93.54*93.54	16747	6
14	*Rosa_2*	91.47*91.47*91.47	15649	10

15	*Garcinia*	92.72*92.72*92.72	16427	12
16	*Prunus*	97.50*97.50*97.50	19458	3
17	*Diospyros*	90.47*90.47*90.47	15168	3
18	*Perilla_3*	92.12*92.12*92.12	16039	10
19	*Forsythia*	91.69*91.69*91.69	15680	8
20	*Solanum*	93.09*93.09*93.09	16651	9
21	*Gentiana*	93.45*93.45*93.45	16830	13
22	*3HBF_Medicago*	89.49*89.49*89.49	14538	9
23	*Vigna*	92.46*92.46*92.46	15982	7
24	*Populas3*	91.11*91.11*91.11	15435	8
25	*Rosa_3*	89.28*89.28*89.28	14419	3
26	*Dianthus_1*	90.92*90.92*90.92	15296	11
27	*Dianthus_2*	90.07*90.07*90.07	14897	18
28	*Iris_2*	89.79*89.79*89.79	14718	1
29	*3WC4_Clitoria*	94.35*94.35*94.35	15786	4
30	*Hordeum*	89.54*89.54*89.54	14697	8
31	Final Product	94.05*94.05*94.05	16975	7

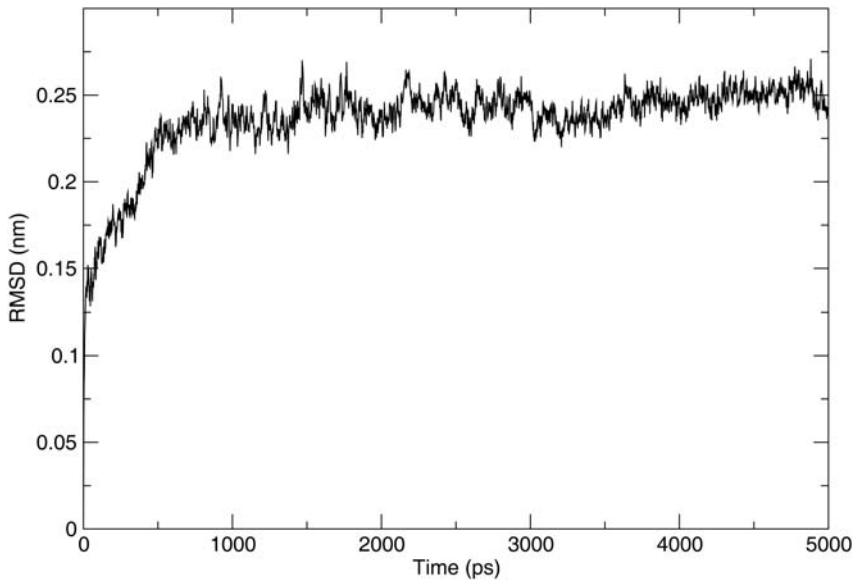


Figure 3.10: RMSD plot of Cα atoms of generated structures of *Fragaria ananassa* UGT with time shown in ps along X-axis and RMSD values in nm along Y-axis.

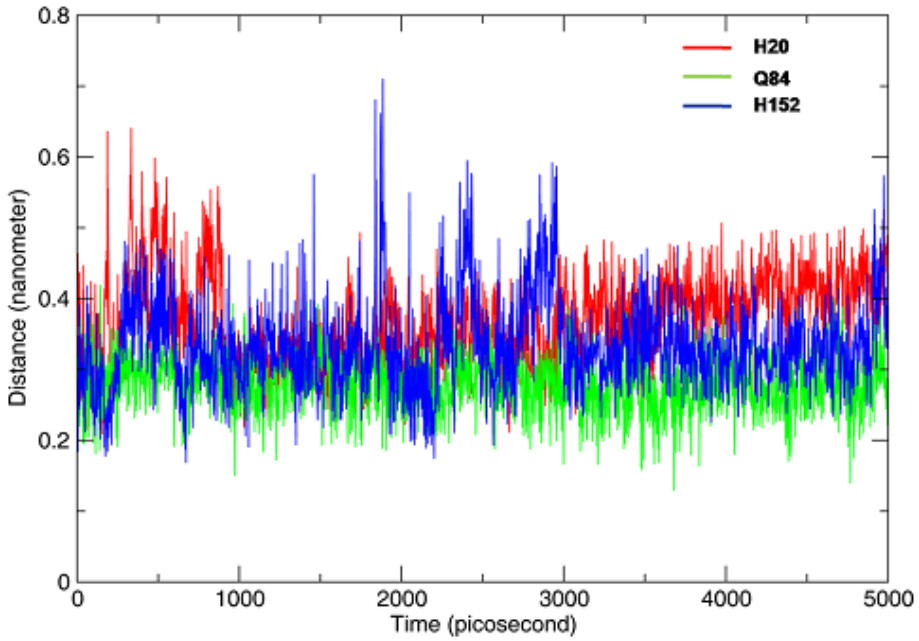


Figure 3.11: Minimum distance plot between Kaempferol and acceptor binding residues of *Fragaria ananassa* UGT.

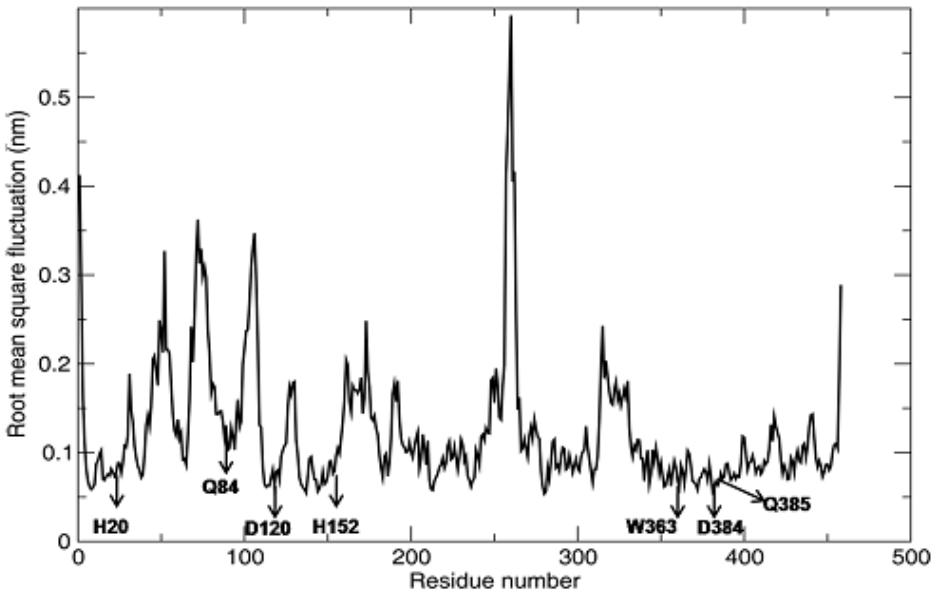


Figure 3.12: RMSF plot *Fragaria ananassa* UGT shows negligible fluctuation of catalytic and binding site residues of acceptor and sugar donor substrate.

3.6 Comparison between experimental evidence and reliability of specificity prediction

Experimental evidence for the flavonoid 3-OH activity is known for UGTs enzymes of *Dianthus_1, Diospyros, Petunia_2, Gentiana, Iris_2, Vigna, Actinidia*, and *Perilla_3* included in the analysis (Ogata *et al*, 2004; Ikegami *et al*, 2009; Miller *et al*, 2009; Tanaka *et al*, 1996; Yoshihara *et al*, 2005; Mato *et al*, 1998; Montefiori *et al*, 2011; Gong *et al*, 1997). Out of them, UGTs of *Dianthus_1, Diospyros and Petunia_2* were considered for analyzing their specificity against various flavonoid substrates. Few positive controls (3-OH group present: Kmp, Myricetin, Quercetin, and Fisetin) and negative control (3-OH group absent/ present in different orientation: Apigenin, Naringenin, Dihyrdroquercetin (DHQ), Catechin, and Epicatechin) were selected to dock in the acceptor binding pocket of the protein. It has been observed that wherever the favorable binding pose was attained the flavonoid had its 3-OH present in close proximity of the catalytic histidine and sugar donor which is known to facilitate the nucleation during glycosylation reaction. However, the neighboring regions of the protein in the vicinity of catalytic base were not conducive for binding any other hydroxyl group of flavonoid (Figure 3.13, 3.14 and 3.15).

In order to further explore the importance of these crucial binding regions, flavonoid 7-O glucosyltransferase (F7GT) of *Scutellaria baicalensis* (Swiss-Prot ID: Q9SXF2), for which the flavonoid specificity was known, was selected for the study (Hirotani *et al*, 2000). The three-dimensional structure of F7GT was modeled (Template: 2VCE; Resolution 1.9Å, Identity: 30%) and three flavonoids namely baicalein, scutellarein, and kmp were chosen for the docking studies. The hydroxyl group is present on the C3 carbon atom of kmp but absent in other two. In spite of this difference the docked complexes showed that all the three ligands attained the favorable pose having only 7-OH closer to the catalytic histidine and sugar donor (Figure 3.16). Considering the flavonoid kmp, irrespective of it's having both 7-OH and 3-OH, as per our classification F3GT favorably binds for 3-OH glycosylation and F7GT binds favoring 7-OH glycosylation. These results show how the acceptor binding site of UGTs decides the favorable binding mode of flavonoids and the nature of the glycosylation product formed.

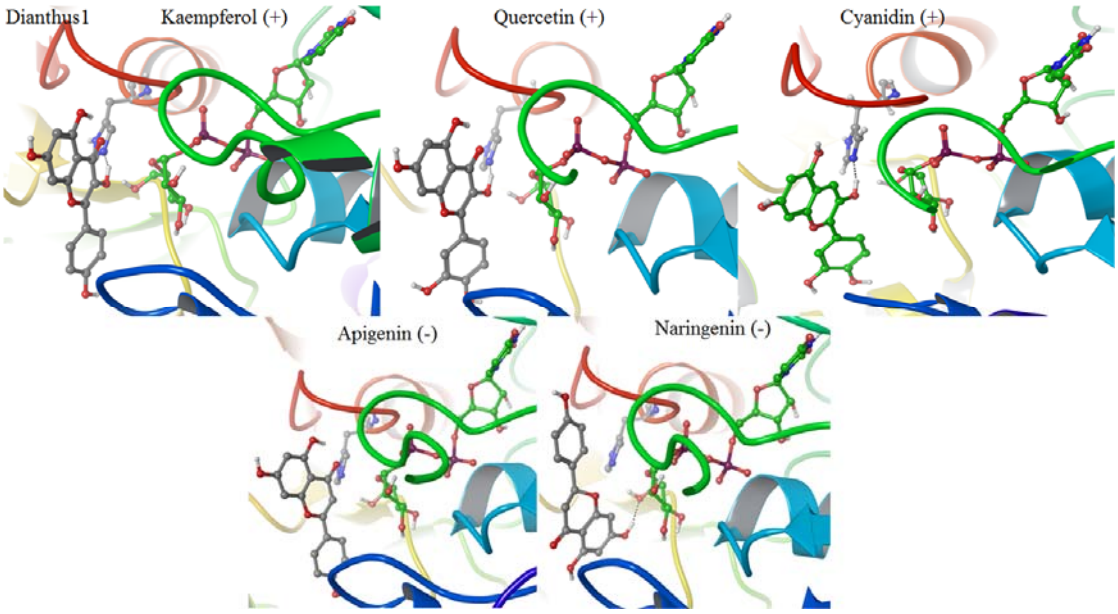


Figure 3.13: Image showing docked complexes of three positive and two negative control ligands in the acceptor binding pocket of *Dianthus caryophyllus* (*Dianthus_1*).

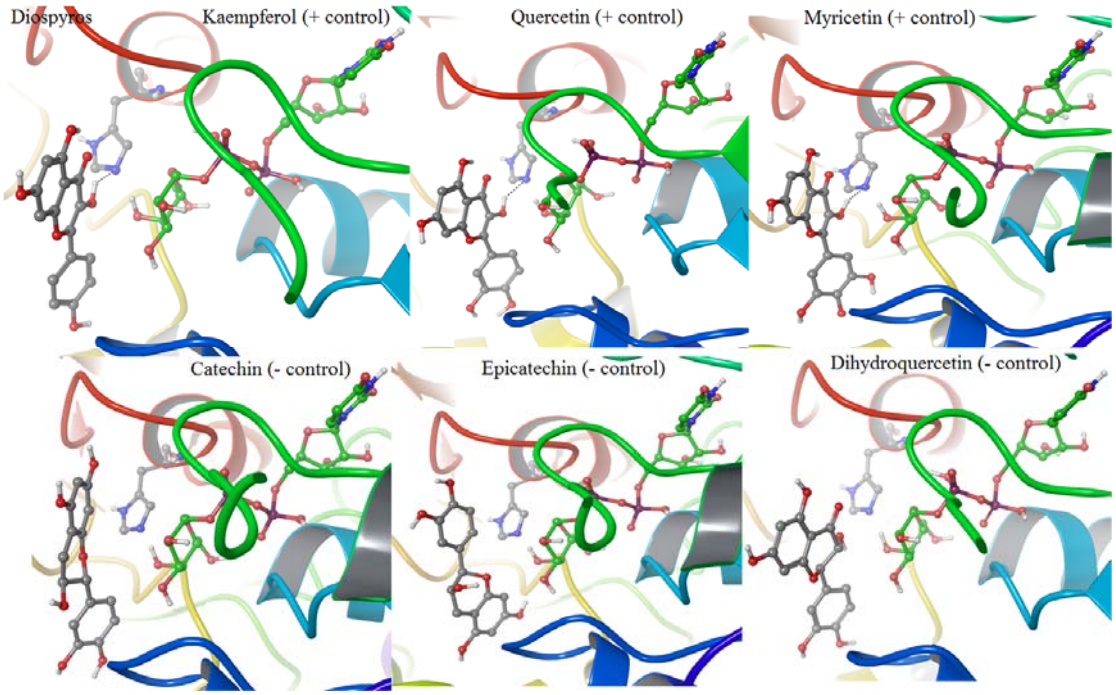


Figure 3.14: Image showing docked complexes of three positive and three negative control ligands in the acceptor binding pocket of *Diospyros kaki* (*Diospyros*).

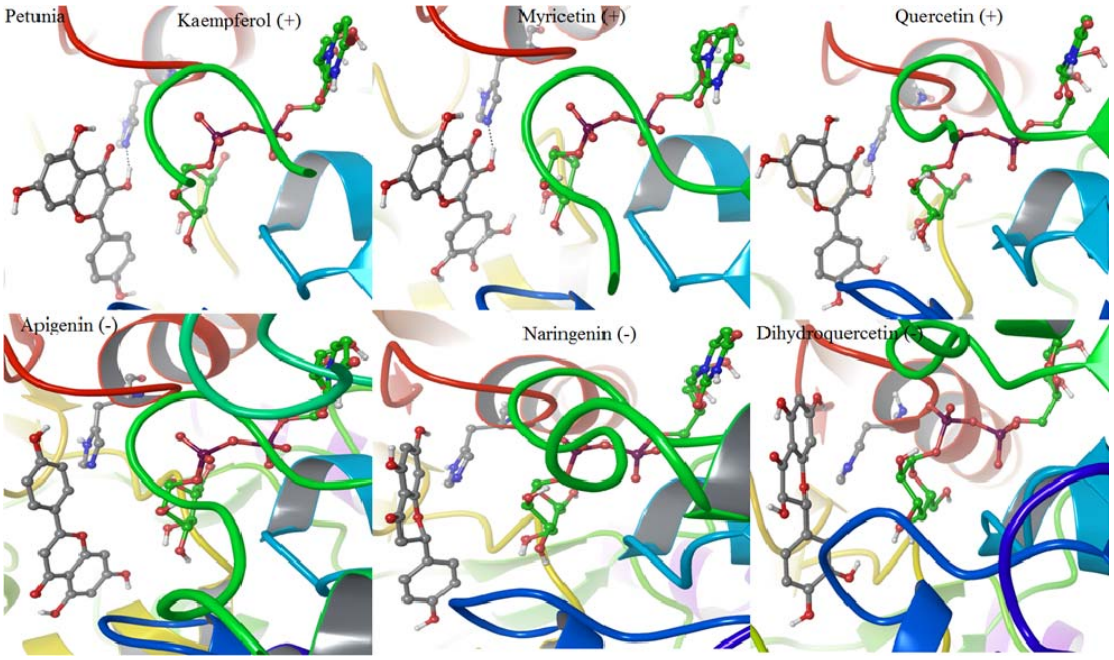


Figure 3.15: Image showing docked complexes of three positive and three negative control ligands in the acceptor binding pocket of *Petunia hybrida* (*Petunia_2*).

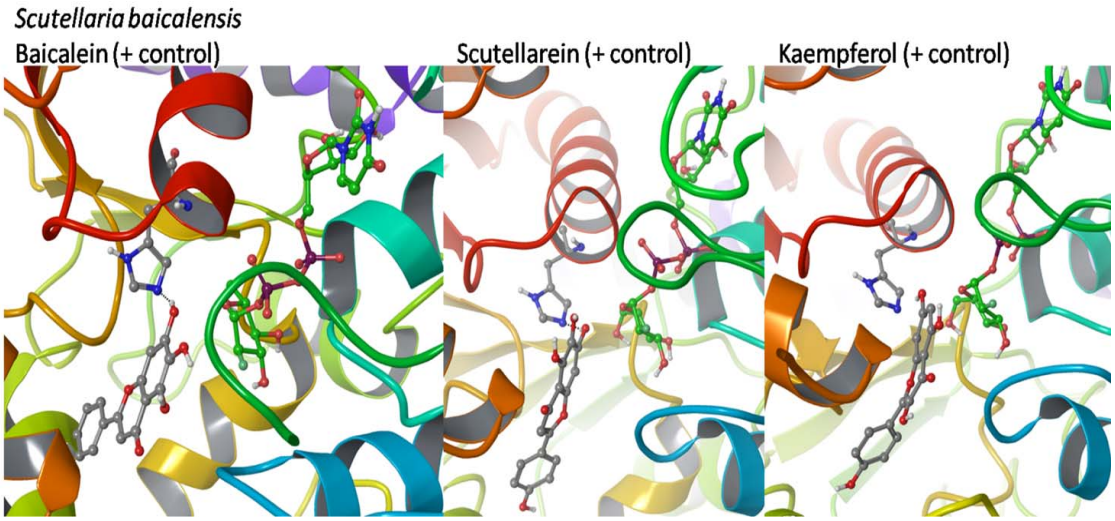


Figure 3.16: Image showing docked complexes of three positive control ligands in the acceptor binding pocket of *Scutellaria baicalensis*.

3.7 Conclusion

Apart from different roles UGTs play in plants such as cell cycle regulation and involvement in stress our detailed study of structure-function relation will help in

utilizing these enzymes in various applications. Homology modeling and docking studies of one of the phylogenetic set, F3GTs, showed that the environment in the close vicinity of the acceptor binding site is conserved in all the members of this group and that the participating amino acids will favor the binding of a flavonoid such that the glycosylation of the 3-OH group is preferred over other hydroxyl groups. Eight regions surrounding the acceptor binding site were identified as important for acceptor specificity. Alterations in these regions may affect the type of molecule to be glycosylated and change the acceptor specificity from one molecule to another or from one side group for glycosylation to another side group. These observations have been further confirmed using molecular dynamics simulations of the docked complexes. The binding site information was successfully exploited to predict UGTs with experimentally known 3-OH specificity as positive control and 7-OH specific UGTs along with different ligands as negative control. This knowledge can be implemented to design engineered GTs which glycosylate flavonoids in a regiospecific manner. In this way, the biochemical and pharmaceutical properties of flavonoids can be altered to prepare more effective drug molecules in a desired chemical form.

Chapter 4

Genome-wide identification and tissue specific expression studies of UDP glycosyltransferases gene family in *Cicer arietinum* (chickpea) genome

A diverse range of bioactive compounds exist in plants which provide several economic and health benefits, therefore it is very important to study the genes involved in their biosynthesis. In plants glycosylation is the last step towards the biosynthesis of terpenoids, phenylpropanoids, cyanogenic glucosides, and glucosinolates produced during biotic and abiotic stresses that changes their activity, sub-cellular location and modulates chemical properties like stability and solubility (Jones & Vogt, 2001). The researchers are benefited by genome sequencing projects which help to analyze the new data and get useful information out of it. Keeping these objectives in mind, a genome-wide study of UDP glycosyltransferases gene family in the newly sequenced *Cicer arietinum* (chickpea) genome was carried out in this chapter.

4.1 Identification of chickpea UGT proteins

Chickpea UGTs were identified by following three methodologies namely Blastp, Position-Specific Weight Matrix (PSWM) guided search and hidden Markov model (HMM) profile search, respectively. A blast search was carried out using PSPG signature motif of UGT from *Vitis vinifera* (PDB-2C1Z) as a query against the chickpea predicted proteome keeping the Expectation value (E-value) cut off of 1. SUPERFAMILY server (Gough *et al*, 2001) was employed to identify the superfamily to which the predicted UGTs belong to.

The hits were re-confirmed by de novo searching a conserved signature motif by creating and using a dataset of 89 plant UGT protein sequences. Multiple EM for Motif Elicitation v4.9.0 suite (MEME) (Bailey *et al*, 2009; Bailey and Elkan, 1994) with zoops (zero or one occurrence) was used for searching conserved motif (E-value 2.5*10-2741). Accurate length of the motif was confirmed by considering bit score and relative entropy. A PSPG motif of these sequences was then extracted to create PSWM. This PSWM was used to identify the UGTs using Motif Alignment & Search Tool v4.9.0 (MAST) of MEME suite (Bailey and Gribskov, 1998).

Chickpea UGTs from its predicted proteome data were identified using a stand-alone blastp search of PSPG motif as query against 28, 269 chickpea gene models. 125 UGT sequences with lengths ranging from 126 to 596 amino acids could be identified this way. However, 15 of these sequences showed comparatively higher E-value. Family 1 UGTs utilizes low molecular weight compounds as acceptors

bound in the N-terminal domain and possess a highly conserved carboxy terminal signature motif (PSPG motif), involved in the binding of sugar donor (UDP sugar) in the binding site (Figure 4.1). Taking these features into account, 96 sequences possessing both the domains and length variation 410-596 amino acids were selected for further analysis, for which the following GenBank Accession numbers were assigned [GenBank ID: KC990643, KF000375-KF000405, KF006942-KF006953, KF018245-KF018279, KF039755-KF039769 and KF843731-KF843732] (CD-Table S-4.1, CD-Figure S-4.1). These 96 sequences were taken forward for the further analysis.

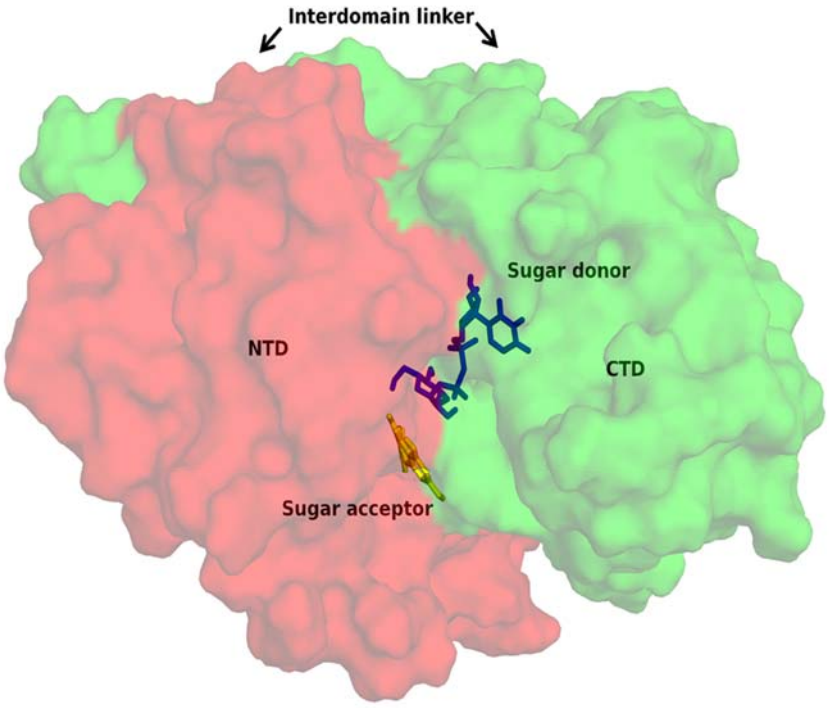


Figure 4.1: Surface representation of UGT88E9 with bound quercetin (Yellow) and UPG (blue) shown in stick form. The NTD and CTD are shown in red and green color with the interdomain linker marked by arrows.

To further confirm the above result, PSWM was created using PSPG motif from 89 protein sequences of various plant sources with the help of MEME below E-value $2.5*10^{-2741}$ (Figure 4.2, Table 4.1). Length of PSPG motif (44 amino acids) was re-confirmed by comparing bit score and relative entropy of protein motif of long length (100 amino acids) identified by MEME. PSWM was then given as input to MAST of MEME suite to screen the *Ca*UGTs in the whole predicted proteome of

chickpea. This method identified 124 *Ca*UGTs (with E-value below $9.9*10^{-0.9}$), out of which 123 sequences matched with the previously identified *Ca*UGTs using blastp. The sequence not predicted by blast (Ca_06153) was identified in the next methodology too but it was found to be a false positive hit.

Table 4.1: Dataset of 89 plant UGTs utilized to build the position specific weight matrix using the conserved PSPG motif.

Serial number	Protein	Swiss-Prot ID	Serial number	Protein	Swiss-Prot ID
1	*Nicotiana tabacum*	Q8RU71	46	*Garcinia mangostana*	B9UZ54
2	*Petunia hybrida*	Q9SBQ2	47	*Arabidopsis thaliana*	Q9LFJ8
3	*Verbena hybrida*	Q9ZR25	48	*Rosa hybrid cultivar*	A8CDW6
4	*Perilla frutescens*	Q9ZR27	49	*Citrus paradisi*	C5MR71
5	*Lycium barbarum*	B6EWY2	50	*Dianthus caryophyllus*	Q60FF2
6	*Eustoma exaltatum subsp. russellianum*	A4F1Q2	51	*Petunia hybrida*	Q9SBQ8
7	*Arabidopsis thaliana*	Q9LR44	52	*Lobelia erinus*	A4F1S9
8	*Arabidopsis thaliana*	Q0WW21	53	*Rosa hybrid cultivar*	Q2PGW4
9	*Zea mays*	B6TR02	54	*Rosa hybrid cultivar*	A8CDW8
10	*Oryza sativa subsp. japonica*	Q69IU8	55	*Populus trichocarpa*	B9NFI8
11	*Avena strigosa*	C4MF47	56	*Forsythia intermedia*	Q9XF16
12	*Zea mays*	B6SRX8	57	*Perilla frutescens*	O04114
13	*Crocus sativus*	Q6X1C0	58	*Gentiana triflora*	Q96493
14	*Medicago truncatula*	A2Q5W6	59	*Solanum tuberosum*	Q3YK56
15	*Medicago truncatula*	A2Q5W9	60	*Dianthus caryophyllus*	Q60FF0
16	*Lycium barbarum*	B6EWX8	61	*Vigna mungo*	Q9ZWS2
17	*Stevia rebaudiana*	Q6VAA6	62	*Clitoria ternatea*	A4F1Q6
18	*Ixeris dentata var. albiflora*	A9X3L7	63	*Iris hollandica*	Q5KTF3
19	*Arabidopsis thaliana*	Q9SYK9	64	*Hordeum vulgare*	P14726
20	*Brassica rapa subsp. pekinensis*	C5H9P3	65	*Zea mays*	B6TY52
21	*Rhodiola sachalinensis*	A4ZZ92	66	*Phaseolus vulgaris*	Q9FUJ6
22	*Populus trichocarpa*	B9NAD3	67	*Stevia rebaudiana*	Q6VAA5
23	*Arabidopsis thaliana*	O22182	68	*Stevia rebaudiana*	Q6VAA3
24	*Zea mays*	B6UFB5	69	*Ipomoea nil*	Q53UH4

25	*Medicago truncatula*	A6XNC4	70	*Dianthus caryophyllus*	A7M6J3
26	*Iris hollandica*	Q767C8	71	*Zea mays*	B6U5U7
27	*Rauvolfia serpentina*	Q9AR73	72	*Lycium barbarum*	B6EWZ0
28	*Hieracium pilosella*	B2CZL2	73	*Lycium barbarum*	B6EWZ4
29	*Phytolacca americana*	B5MGN7	74	*Sesamum indicum*	A9ZPI0
30	*Medicago truncatula*	B9U3W6	75	*Bellis perennis*	Q5NTH0
31	*Arabidopsis thaliana*	O23205	76	*Citrus maxima*	Q8GVE3
32	*Torenia hybrid cultivar*	B7XH67	77	*Phaseolus vulgaris*	A7L745
33	*Perilla frutescens*	B2NID2	78	*Zea mays*	B4FU09
34	*Antirrhinum majus*	B2NIC9	79	*Allium cepa*	Q7XJ49
35	*Veronica persica*	C0STS8	80	*Hieracium pilosella*	B2CZL5
36	*Scutellaria baicalensis*	B9A9D5	81	*Withania somnifera*	C1JIE1
37	*Antirrhinum majus*	Q33DV3	82	*Withania somnifera*	B9VJL9
38	*Linaria vulgaris*	Q33DV2	83	*Bacopa monnieri*	B9VNU9
39	*Phaseolus angularis*	Q8S9A4	84	*Bacopa monnieri*	B9VNV0
40	*Stevia rebaudiana*	Q6VAA7	85	*Arabidopsis thaliana*	Q9M156
41	*Hieracium pilosella*	B2CZL4	86	*Medicago truncatula*	Q5IFH7
42	*Zea mays*	B6T4P0	87	*Medicago truncatula*	G7JD22
43	*Manihot esculenta*	Q40284	88	*Vitis vinifera*	P51094
44	*Ipomoea nil*	A7M6U9	89	*Medicago truncatula*	A6XNC5
45	*Populus trichocarpa*	B9I672			

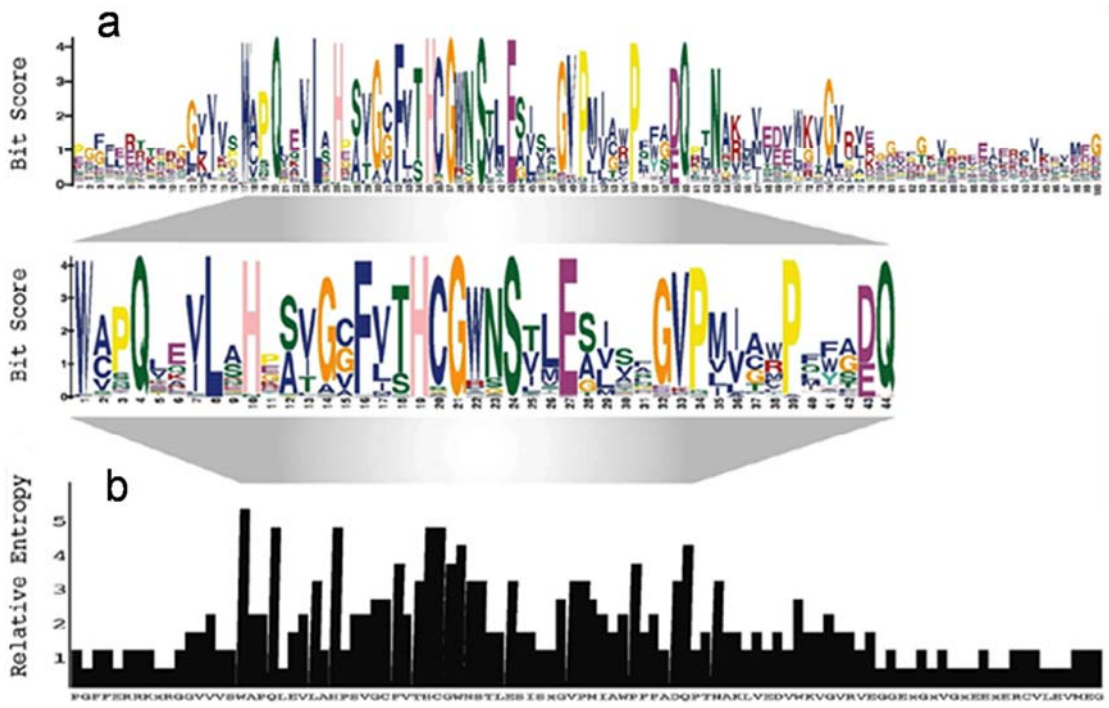


Figure 4.2: Conservation [Bit score (a) and Relative entropy (b)] of the PSPG motif of 89 UGTs from various plant sources.

Predicted proteome of chickpea was searched for the UGTs with the help of HMM-profile of Pfam Family UDPGT (PF00201). This method resulted in 129 UGT hits which matched with the hits of previous two methods. The PSPG motif of the four additional hits not predicted by BLAST was variable when compared with the remaining 125 sequences. Even these protein sequence hits (Ca_06794, Ca_06153, Ca_27131 and Ca_19130) showed slightly higher E-value as well as missing NTD or CTD and hence, were not considered in further study (CD-Figure S-4.2).

Out of total 125 predicted UGTs by BLAST, 123 sequences were identified through both PSWM and HMM profile search (Figure 4.3). These results confirm that most of the UGTs of chickpea were identified by following three different methodologies. Location and genomic distribution of each *UGT* on the genome is shown in Figure 4.4.

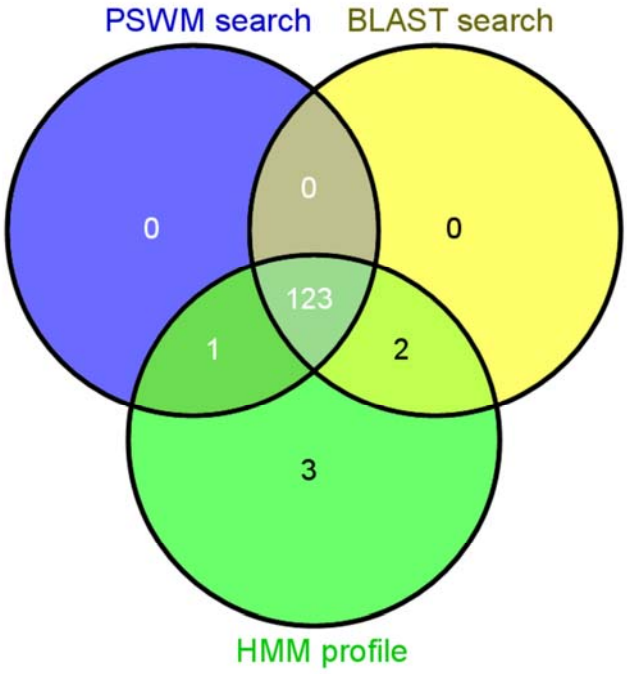


Figure 4.3: The number of *Ca*UGTs identified using various methods such as PSWM search using MEME-MAST, Blastp and HMM-profiles search shown with the help of a Venn diagram.

4.2 Phylogenetic analysis and recent gene duplication events

Nomenclature of UGTs used was that recommended by UGT nomenclature committee (Table 4.2 & 4.3). All 96 predicted UGTs of chickpea belong to the family UDPGT-like and UDP glycosyltransferase/ glycogen phosphorylase superfamily that utilize nucleotide molecule uridine diphosphate (UDP) with attached sugar molecule to perform glycosylation reaction. A phylogenetic tree of the predicted 96 *Ca*UGTs was prepared using maximum likelihood method by employing aLRT SH-like fast likelihood-based method. The long gene of UGT80B4 bearing several introns might have diverged away from the rest. Similarly, the closely related UGT85H6 & UGT85H7 (sequence identity: 96%) and UGT79B21 & UGT79B22 (identity: 98%) might be related by recent duplication events (Figure 4.5).

Table 4.2: Basis of nomenclature of UDP-Glycosyltransferase superfamily member.

UGT	UDP-Glycosyltransferase superfamily
Number from 71-100	Family which group UGT with 40% or more amino acid sequence homology
An alphabet after family	Subfamily which group UGTs with 60% or more amino acid sequence homology
A number after subfamily	Unique number specific to an individual gene

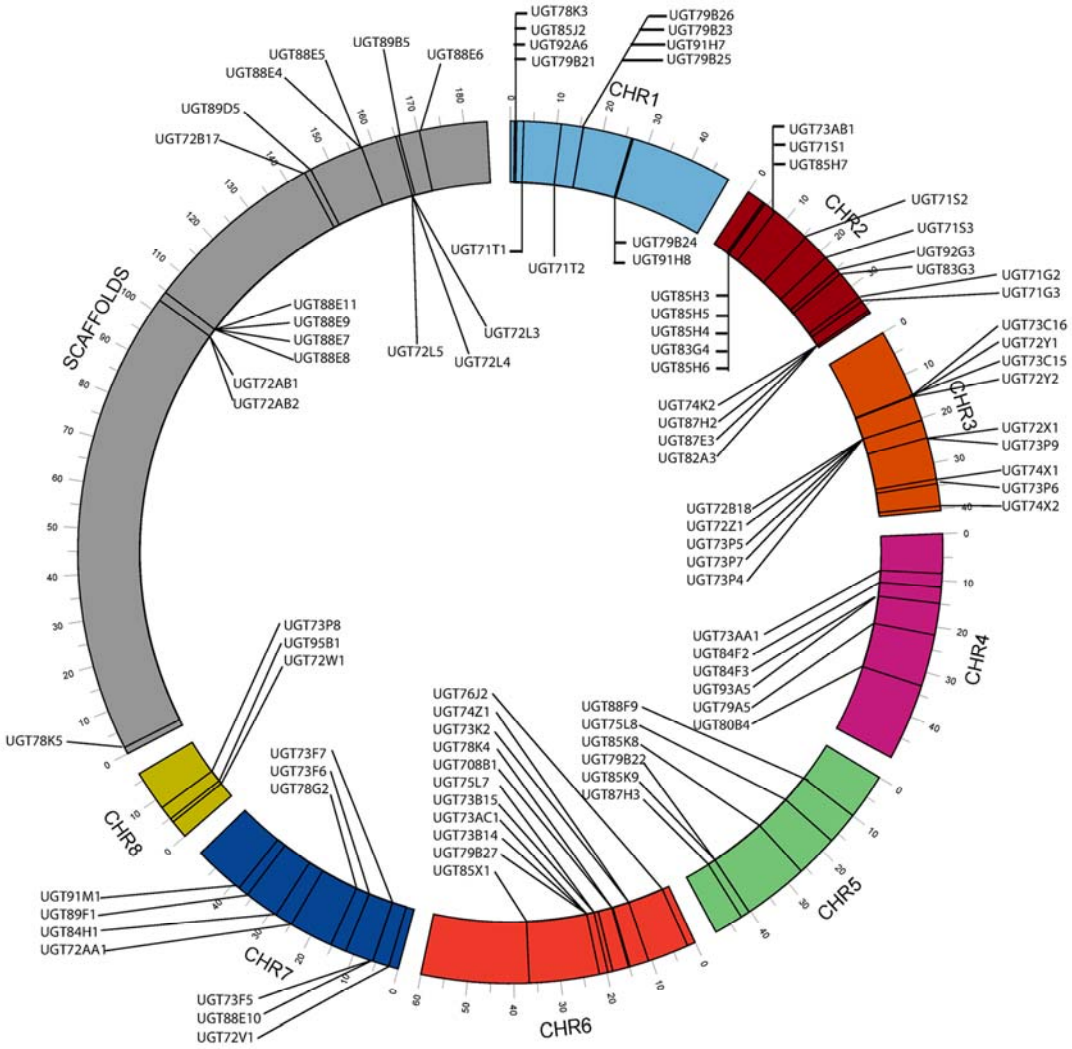


Figure 4.4: Genomic distribution of *Ca*UGTs. Chromosomal distribution of *Ca*UGTs in chickpea genome.

Table 4.3: Nomenclature provided by the UGT nomenclature committee to all the 96 *Ca*UGTs.

No.	Seq_ID	Name	No.	Seq_ID	Name
1	Ca_10301	UGT76J2	49	Ca_10248	UGT92G3
2	Ca_10468	UGT85H3	50	Ca_06582	UGT708B1
3	Ca_00166	UGT78K3	51	Ca_08527	UGT75L7
4	Ca_10472	UGT85H5	52	Ca_06480	UGT73B15
5	Ca_10469	UGT85H4	53	Ca_09314	UGT84H1
6	Ca_21108	UGT83G4	54	Ca_25712	UGT72L5
7	Ca_09924	UGT72V1	55	Ca_06479	UGT73AC1
8	Ca_10470	UGT85H6	56	Ca_21107	UGT83G3
9	Ca_19814	UGT73K2	57	Ca_22494	UGT72X1
10	Ca_21872	UGT88E4	58	Ca_06451	UGT73B14
11	Ca_21869	UGT88E5	59	Ca_11382	UGT75L8
12	Ca_19082	UGT73AA1	60	Ca_25710	UGT72L4
13	Ca_18541	UGT73AB1	61	Ca_09437	UGT73P9
14	Ca_05296	UGT78K4	62	Ca_02667	UGT92A6
15	Ca_27487	UGT88E6	63	Ca_06074	UGT74X1
16	Ca_10147	UGT71S1	64	Ca_07558	UGT85K8
17	Ca_21795	UGT88E11	65	Ca_09439	UGT73P6
18	Ca_21778	UGT88E9	66	Ca_06073	UGT74X2
19	Ca_10471	UGT85H7	67	Ca_00350	UGT79B21
20	Ca_16414	UGT88E10	68	Ca_14261	UGT71G2
21	Ca_23668	UGT73C16	69	Ca_17956	UGT79B22
22	Ca_22500	UGT72Y1	70	Ca_14260	UGT71G3
23	Ca_21263	UGT72B17	71	Ca_18599	UGT71T1
24	Ca_00728	UGT73C15	72	Ca_18617	UGT71T2
25	Ca_02074	UGT72W1	73	Ca_19894	UGT89F1
26	Ca_00112	UGT85J2	74	Ca_19131	UGT74K2

27	Ca_22982	UGT78K5	75	Ca_07557	UGT85K9
28	Ca_05608	UGT84F2	76	Ca_08995	UGT87H3
29	Ca_21788	UGT88E7	77	Ca_10596	UGT95B1
30	Ca_22496	UGT72Y2	78	Ca_26288	UGT89D5
31	Ca_21787	UGT88E8	79	Ca_17244	UGT87H2
32	Ca_01304	UGT72B18	80	Ca_11281	UGT79B27
33	Ca_06791	UGT73F5	81	Ca_05261	UGT85X1
34	Ca_17221	UGT72AB1	82	Ca_26272	UGT89B5
35	Ca_06792	UGT73F7	83	Ca_08338	UGT79A5
36	Ca_05607	UGT84F3	84	Ca_06975	UGT79B26
37	Ca_23049	UGT71S2	85	Ca_06976	UGT79B23
38	Ca_22495	UGT72Z1	86	Ca_00113	UGT91H7
39	Ca_06796	UGT73F6	87	Ca_02172	UGT73P8
40	Ca_09440	UGT73P5	88	Ca_06977	UGT79B25
41	Ca_03267	UGT78G2	89	Ca_17245	UGT87E3
42	Ca_18951	UGT72AA1	90	Ca_09825	UGT82A3
43	Ca_09436	UGT73P7	91	Ca_06978	UGT79B24
44	Ca_24017	UGT93A5	92	Ca_00114	UGT91H8
45	Ca_25711	UGT72L3	93	Ca_13719	UGT91M1
46	Ca_20510	UGT88F9	94	Ca_04376	UGT80B4
47	Ca_09824	UGT71S3	95	Ca_17222	UGT72AB
48	Ca_09438	UGT73P4	96	Ca_08677	UGT74Z1

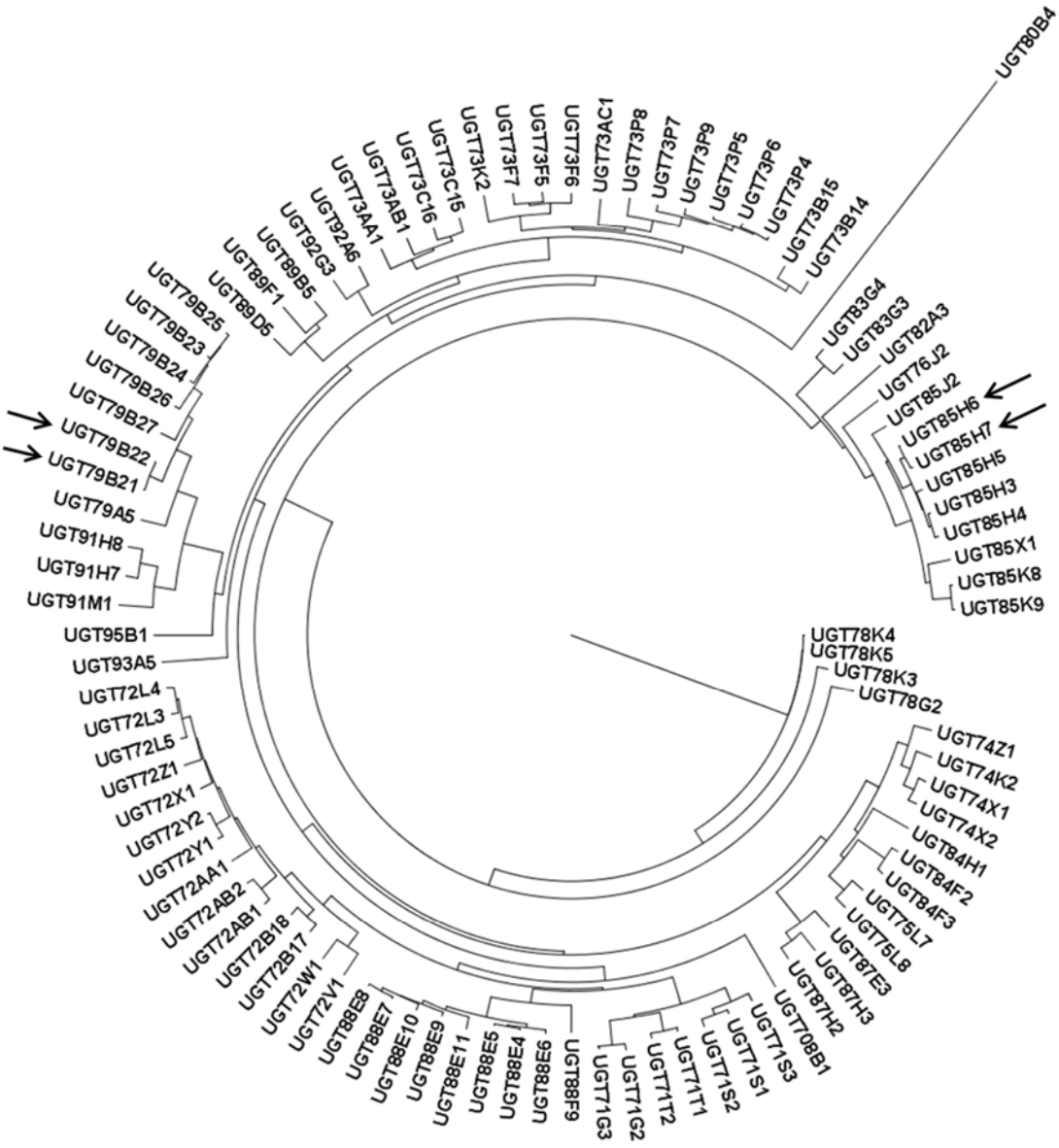


Figure 4.5: Phylogenetic analysis of *Ca*UGTs. Dendrogram showing clustering of 96 *Ca*UGTs along with two recent gene duplication events marked by arrows.

4.3 Functional annotation of UGTs

Functional annotation and assignment of substrate specificity of *Ca*UGTs, was carried out using experimentally characterized UGT proteins with known substrate specificity from other plant species. A total of 38 such protein sequences were retrieved from Swiss-Prot database (Table 4.4). A phylogenetic tree was generated for the 96 *Ca*UGTs combined with these 38 plant UGTs to analyze the clustering pattern by keeping the parameters same as those previously used in PhyML. The clustered UGTs

in each clade were further analyzed for similarities in the regions exposed to substrate binding pocket.

Table 4.4: Dataset of 38 experimentally validated UGTs used for the functional assignment of chickpea UGTs.

No	Plant species	Protein name	Sequence id	Swiss-Prot id	Reference
1	*Zea mays*	Cis-zeatin O-glucosyltransferase 1	Cis_Zeatin_G1	Q93XP7	Martin *et al*, 2001
2	*Phaseolus lunatus*	Zeatin O-glucosyltransferase	Cis_Zeatin_G2	Q9ZSK5	Martin *et al*, 1999
3	*Glycine max*	Soyasapogenol B glucuronide galactosyltransferase	Soyasapogenol_G	D4Q9Z4	Shibuya *et al*, 2010
4	*Glycine max*	Soyasaponin III rhamnosyltransferase	Soyasaponin_R	D4Q9Z5	Shibuya *et al*, 2010
5	*Arabidopsis thaliana*	UDP-glycosyltransferase 73C5	Cytokinin_OG1	Q9ZQ94	Li *et al*, 2002
6	*Arabidopsis thaliana*	UDP-glycosyltransferase 73C1	Cytokinin_OG2	Q9ZQ99	Li *et al*, 2002
7	*Arabidopsis thaliana*	UDP-glycosyltransferase 85A1	Cytokinin_OG3	Q9SK82	Li *et al*, 2002
8	*Arabidopsis thaliana*	UDP-glycosyltransferase 76C1	Cytokinin_NG1	Q9FI99	Hou *et al*, 2004
9	*Arabidopsis thaliana*	UDP-glycosyltransferase 76C2	Cytokinin_NG2	Q9FIA0	Hou *et al*, 2004
10	*Rauvolfia serpentina*	Hydroquinone glucosyltransferase	Hydroquinone_G	Q9AR73	Arend *et al*, 2000
11	*Citrus unshiu*	Limnoid UDP-glucosyltransferase	Limonoid_UDP_G	Q9MB73	Kita *et al*, 2000
12	*Nicotiana tabacum*	Scopoletin glucosyltransferase	Scopoletin_G	Q9AT54	Gachon *et al*, 2004
13	*Rosa hybrid cultivar*	Anthocyanidin 5,3-O-glucosyltransferase	Anthocyanidin_5_3 G1	Q4R1I9	Ogata *et al*, 2005
14	*Ipomoea nil*	Anthocyanidin 3-O glucoside 2"-O-glucosyltransferase	Anthocyanidin_3_2 "G1	Q53UH4	Morita *et al*, 2005
15	*Ipomoea purpurea*	Anthocyanidin 3-O glucoside 2"-O-glucosyltransferase	Anthocyanidin_3_2 "G2	Q53UH5	Morita *et al*, 2005
16	*Verbena hybrida*	Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase	Anthocyanidin_3_5 _G1	Q9ZR25	Yamazaki *et al*, 1999
17	*Gentiana triflora*	Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase	Anthocyanidin_3_5 _G2	B2NID7	Nakatsuka *et al*, 2008

18	*Perilla frutescens*	Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase 1	Anthocyanidin_3_5_G3	Q9ZR27	Yamazaki *et al*, 1999
19	*Phaseolus angularis*	Abscisate beta-glucosyltransferase	Abscisate_beta_G	Q8W3P8	Xu *et al*, 2002
20	*Arabidopsis thaliana*	UDP-glycosyltransferase 71C1	Flavonol _7_3G1	O82381	Li *et al*, 2002
21	*Arabidopsis thaliana*	UDP-glycosyltransferase 71C4	Flavonol _7_3G2	Q9LML6	Li *et al*, 2002
22	*Arabidopsis thaliana*	UDP-glycosyltransferase 71D1	Flavonol_3_G1	O82383	Li *et al*, 2002
23	*Arabidopsis thaliana*	UDP-glycosyltransferase 71B1	Flavonol_3_G2	Q9LSY9	Li *et al*, 2002
24	*Arabidopsis thaliana*	UDP-glycosyltransferase 89C1	Flavonol_7R	Q9LNE6	Li *et al*, 2002
25	*Arabidopsis thaliana*	UDP-glycosyltransferase 89B1	Flavonol 3_7_4'G	Q9C9B0	Li *et al*, 2002
26	*Fragaria ananassa*	UDP-glucose flavonoid 3-O-glucosyltransferase 6	Flavonoid_3G1	Q2V6K0	Griesser *et al*, 2008
27	*Fragaria ananassa*	UDP-glucose flavonoid 3-O-glucosyltransferase 7	Flavonoid_3G2	Q2V6J9	Griesser *et al*, 2008
28	*Medicago truncatula*	Flavonoid 3-O-glucosyltransferase	Flavonoid_3G3	A6XNC6	Modolo *et al*, 2007
29	*Manihot esculenta*	Anthocyanidin 3-O-glucosyltransferase 1	Anthocyanidin_3G1	Q40284	Hughes *et al*, 1994
30	*Solanum melongena*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G2	Q43641	Evidence at transcript level
31	*Vitis vinifera*	Anthocyanidin 3-O-glucosyltransferase 2	Anthocyanidin_3G3	P51094	Ford *et al*, 1998
32	*Gentiana triflora*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G4	Q96493	Tanaka *et al*, 1996
33	*Zea mays*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G5	P16166	Furtek *et al*, 1988
34	*Zea mays*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G6	P16167	Furtek *et al*, 1988
35	*Zea mays*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G7	P16165	Furtek *et al*, 1988
36	*Hordeum vulgare*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G8	P14726	Wise *et al*, 1990
37	*Petunia hybrid*	Anthocyanidin 3-O-glucosyltransferase	Anthocyanidin_3G9	Q43716	Kroon *et al*, 1994
38	*Petunia hybrida*	Kaempferol 3-O beta-D-	Kaempferol_3G	Q9SBQ8	Miller *et al*, 1999

galactosyltransferase

4.4 Functional specificity of chickpea UGTs

In the combined dendrogram of 96 *Ca*UGTs and 38 selected plant UGTs, the identified UGTs clustered into 15 groups (designated A to O) (Figure 4.6, Table 4.5). Previously, we have used a strategy of comparing eight substrate binding regions of UGTs combined with clustering to identify flavonoid-3-O glycosyltransferases (F3GTs) in the database (Sharma *et al*, 2004). We have used a similar strategy here to identify UGT specificity. In the present analysis of *Ca*UGTs four clusters of F3GTs (A1 to A4) comprising glycosyltransferases specific to flavonol-3-O and anthocyanidin-3-O were observed. Significant conservation of the eight regions in the vicinity of acceptor binding site is present in each group of the dendrogram (Figure 4.7, CD-Figure S-4.3). A mixing of scopoletin glycosyltransferases and flavonoid-3-O glycosyltransferases was observed in groups A3 and J. As is known, it is possible that they indeed have mixed activity towards both the substrates (Taguchi *et al*, 2001). Similarly, a mixed preference towards different –OH group of flavonoid in group L (flavonoid 7-O, 4'O and 3-O GT) is seen. Successful functional assignment could be achieved for 74 chickpea proteins with some reliability.

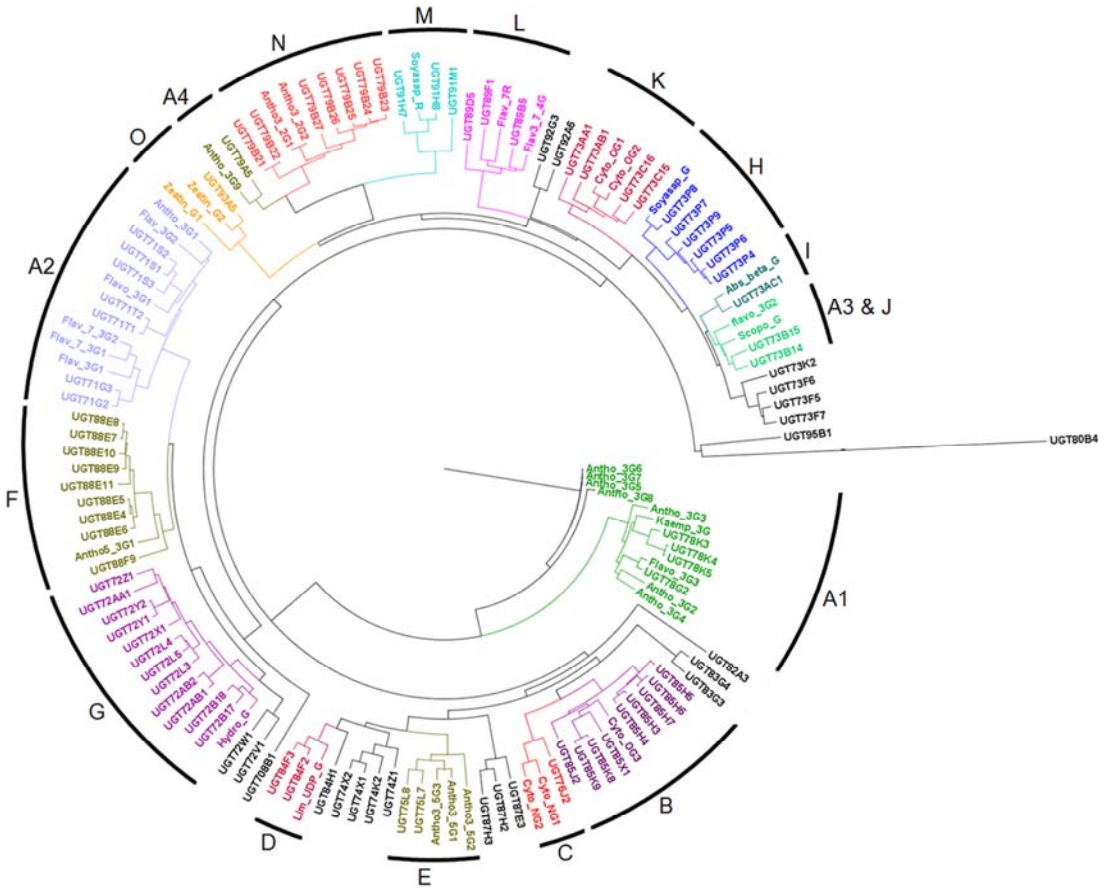
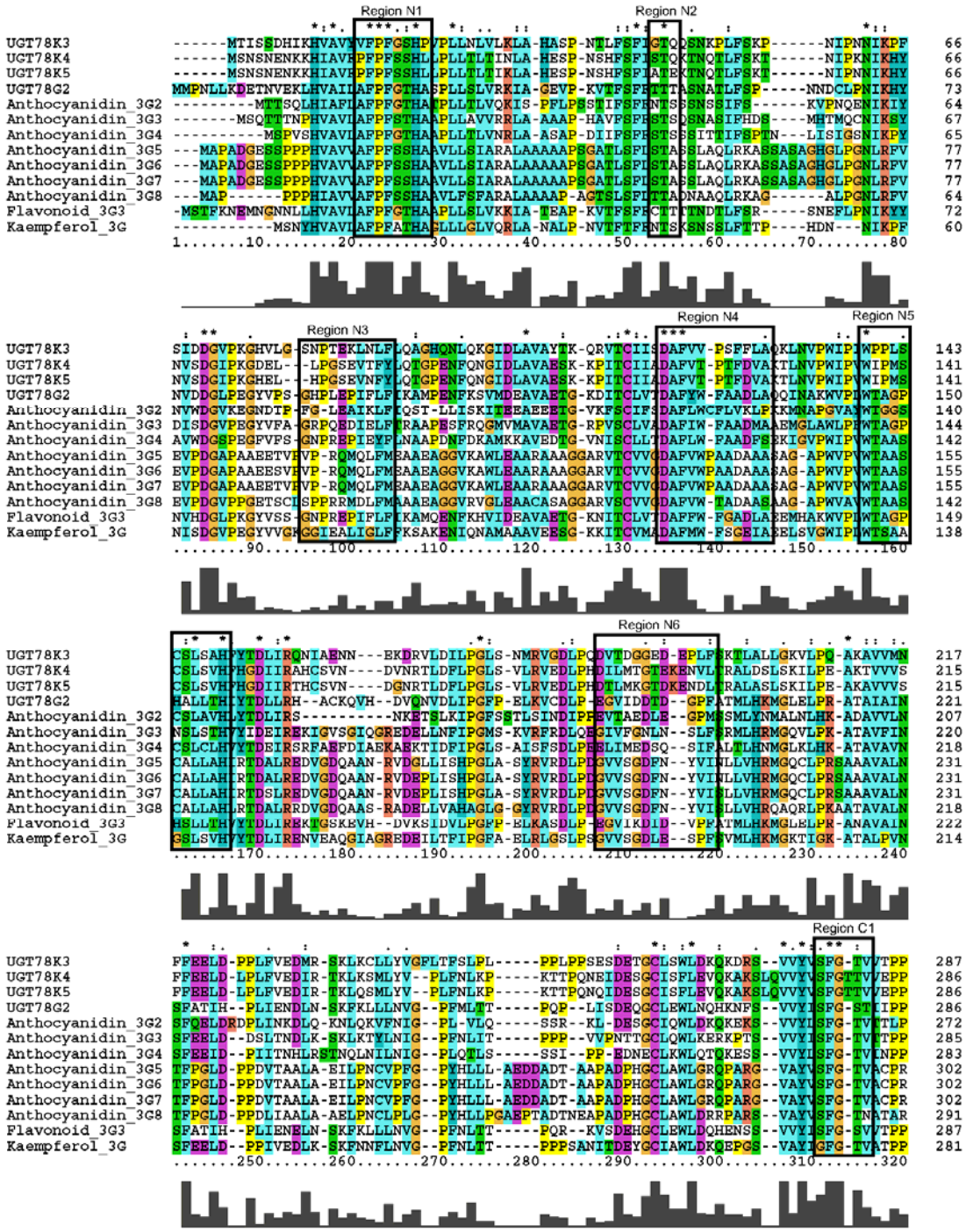


Figure 4.6: Functional annotation of *Ca*UGTs. Dendrogram showing clustering of 96 *Ca*UGTs with 38 well characterized UGT proteins from other plant species. The image shows distinct clustering of *Ca*UGTs with the functionally related UGTs.


Region N1
Region N2
Region N3
Region N4
Region N5
Region N6
Region C1
UGT78K3
UGT78K4
UGT78K5
UGT78G2
Anthocyanidin_3G2
Anthocyanidin_3G3
Anthocyanidin_3G4
Anthocyanidin_3G5
Anthocyanidin_3G6
Anthocyanidin_3G7
Anthocyanidin_3G8
Flavonoid_3G3
Kaempferol_3G

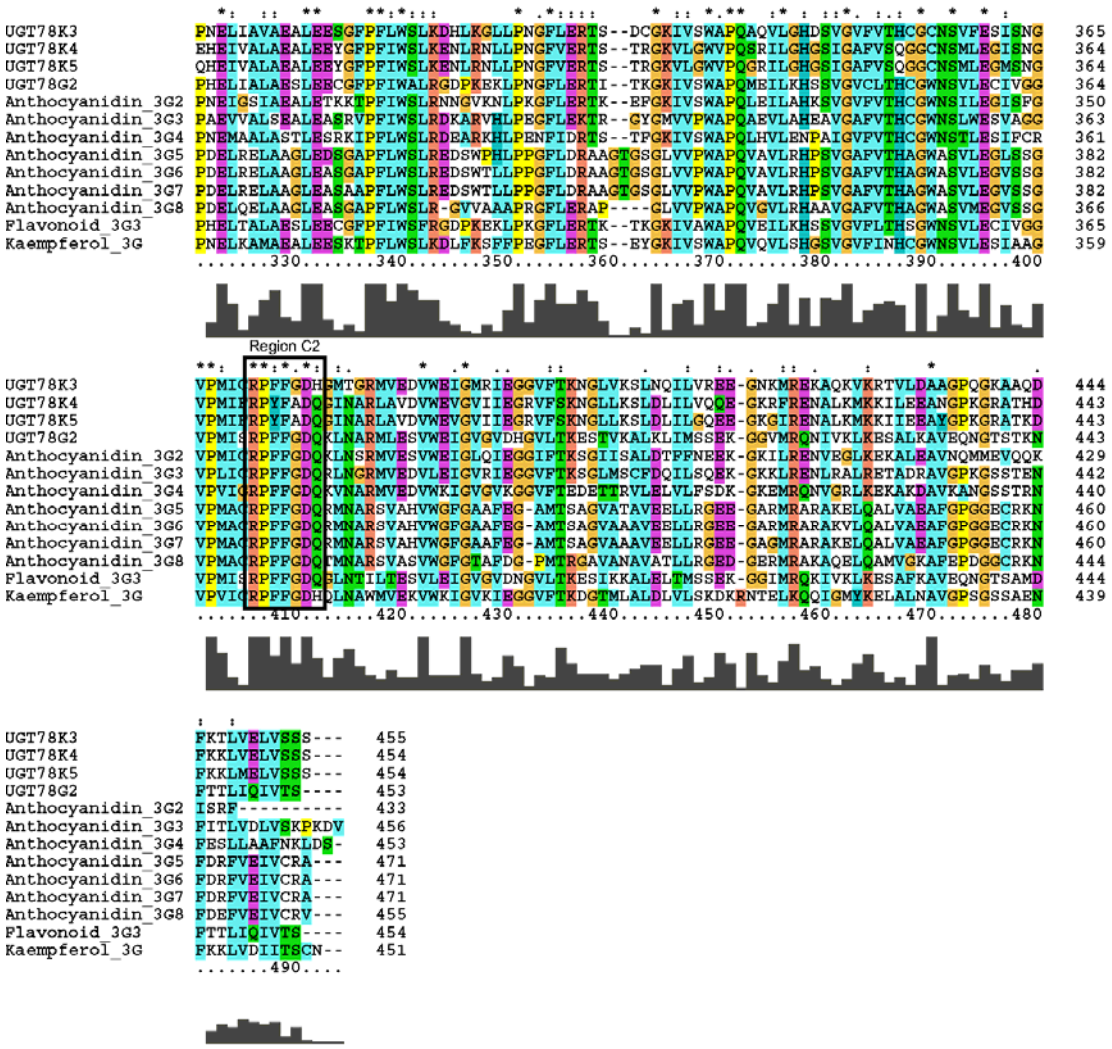


Figure 4.7: Multiple sequence alignment of members of group A1 cluster to show the eight conserved regions enclosed in boxes near the acceptor binding site.

Table 4.5: Functional annotation of the chickpea UGTs based on experimentally characterized UGTs.

No.	Group	*Ca*UGTs	Other genes clustered	Probable function
1	Group A1	UGT78K3, UGT78K4, UGT78K5, UGT78G2	Anthocyanidin_3G2, Anthocyanidin_3G3, Anthocyanidin_3G4, Anthocyanidin_3G5, Anthocyanidin_3G6, Anthocyanidin_3G7, Anthocyanidin_3G8, Flavonoid_3G3, Kaempferol_3G	Anthocyanidin 3-O-glucosyltransferase, Flavonoid 3-O glucosyltransferase, Kaempferol 3-O beta-D-galactosyltransferase
2	Group A2	UGT71G2, UGT71G3, UGT71T1, UGT71T2, UGT71S1, UGT71S2, UGT71S3	Flavonol_3G1, Flavonol_3G2, Flavonoid_3G1, Anthocyanidin_3G1	UDP-glucose flavonoid 3-O-glucosyltransferase, Anthocyanidin 3-O glucosyltransferase
3	Group A4	UGT79A5	Anthocyanidin_3G9	Anthocyanidin 3-O-glucosyltransferase
4	Group B	UGT85X1, UGT85J2, UGT85H3, UGT85H4, UGT85H5, UGT85H6, UGT85H7, UGT85K9, UGT85K8	Cytokinin_OG3	Cytokinin-O-glucosyltransferase 2, Zeatin O-glucosyltransferase
5	Group C	UGT76J2	Cytokinin_NG1, Cytokinin_NG2	Cytokinin-N-glucosyltransferase
6	Group D	UGT84F2, UGT84F3	Limonoid_UDP_G	Limnoid UDP-glucosyltransferase
7	Group E	UGT75L7, UGT75L8	Anthocynidin_3_5_G_1, Anthocynidin_3_5_G_2, Anthocynidin_3_5_G_3	Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase
8	Group F	UGT88E4, UGT88E5, UGT88E6, UGT88E7, UGT88E8, UGT88E9, UGT88E10, UGT88E11, UGTF88F9	Anthocyanidin 5,3-O-glucosyltransferase	Anthocyanidin 5,3-O-glucosyltransferase
9	Group G	UGT72Y1, UGT72B17, UGT72Y2, UGT72B18,	Hydroquinone_G	Hydroquinone glucosyltransferase

		UGT72AB1, UGT72Z1, UGT72AA1, UGT72X1, UGT72AB2 UGT72L3 UGT72L4 UGT72L5		
10	Group H	UGT73P4, UGT73P5, UGT73P6, UGT73P7, UGT73P8, UGT73P9	Soyasapogenol_G	Soyasapogenol B glucuronide galactosyltransferase
11	Group I	UGT73AC1	Abscisate_beta_G	Abscisate beta-glucosyltransferase
12	Group A3 & J	UGT73B14, UGT73B15	Scopoletin_G, Flavonoid_3G_2	Scopoletin glucosyltransferase, UDP-glucose flavonoid 3-O-glucosyltransferase
13	Group K	UGT73C15, UGT73C16, UGT73AA1, UGT73AB1	Cytokinin_OG1, Cytokinin_OG2	Cytokinin-O-glucosyltransferase, Deoxynivalenol-glucosyl-transferase 1, Zeatin O glucosyltransferase
14	Group L	UGT89F1, UGT89D5, UGT89B5	Flavonol_7R, Flavonol 3_7_4'G	Flavonol 7-O-rhamnosyltransferase, Flavonol 3-O-glucosyltransferase, Flavonol 7-O-glucosyltransferase, Flavonoid 4'O-glucosyltransferase
15	Group M	UGT91H7, UGT91H8, UGT91M1	Soyasaponin_R	Soyasaponin III rhamnosyltransferase
16	Group N	UGT79B21, UGT79B22, UGT79B27, UGT79B26, UGT79B23, UGT79B24, UGT79B25	Anthocyanidin_3_2"G1, Anthocyanidin_3_2"G2	Anthocyanidin 3-O glucoside 2"-O-glucosyltransferase
17	Group O	UGT93A5	Cis_Zeatin_G1, Cis_Zeatin_G2	Cis-zeatin O-glucosyltransferase, Zeatin O-glucosyltransferase

4.5 Experimental validation of chickpea UGTs

Out of the total 15 groups identified, UGTs of four clusters share a significant sequence identity with the crystal structure of homologous proteins. Therefore in order to study the binding affinity and specificity for the substrates, UGT78G2 of group A1, UGT71G2 of group A2, UGT85H3 of group B and UGT72B18 of group G were modeled by taking templates of high resolution and identity with the respective targets (Table 4.6). The sequence alignment of the target and the templates used for model building and secondary structure prediction by Psipred showed the similarity between the templates and the target with respect to the arrangement of secondary structure elements. Very few gaps were observed in the alignment of the target and template sequences (CD-Figure S-4.4). The structure validation parameters revealed the high quality of the generated 3-dimensional homology models (Table 4.7).

Table 4.6: Details of the templates used for the molecular modeling studies. The last two columns have the sequence identity between the target and the template structure and the resolution of the template structure.

Serial number	UGT	Group	Template	Identity (%)	Resolution(Å)
1	UGT78G2	A1	3HBF	76	2.10
			2C1Z	47	1.90
2	UGT71G2	A2	2ACV	79	2
3	UGT85H3	B	2PQ6	79	2.10
4	UGT72B18	G	2VCH	62	1.45
5	UGT72X1	G	2VCH	48	1.45

Table 4.7: Statistics of homology model assessment. * % of the total residues in the allowed region. ** % of the total residues in the disallowed region.

Serial number	UGT	Group	Ramachandran plot	Verify3D score	ERRAT value
1	UGT78G2	A1	92.3*, 0**	96.48	85.2
2	UGT71G2	A2	92.4, 0.7	96.58	76
3	UGT85H3	B	90.1, 0.5	95.46	70.33
4	UGT72B18	G	91.7, 0.5	98.52	79.52
5	UGT72X1	G	99, 1	98.93	84.43

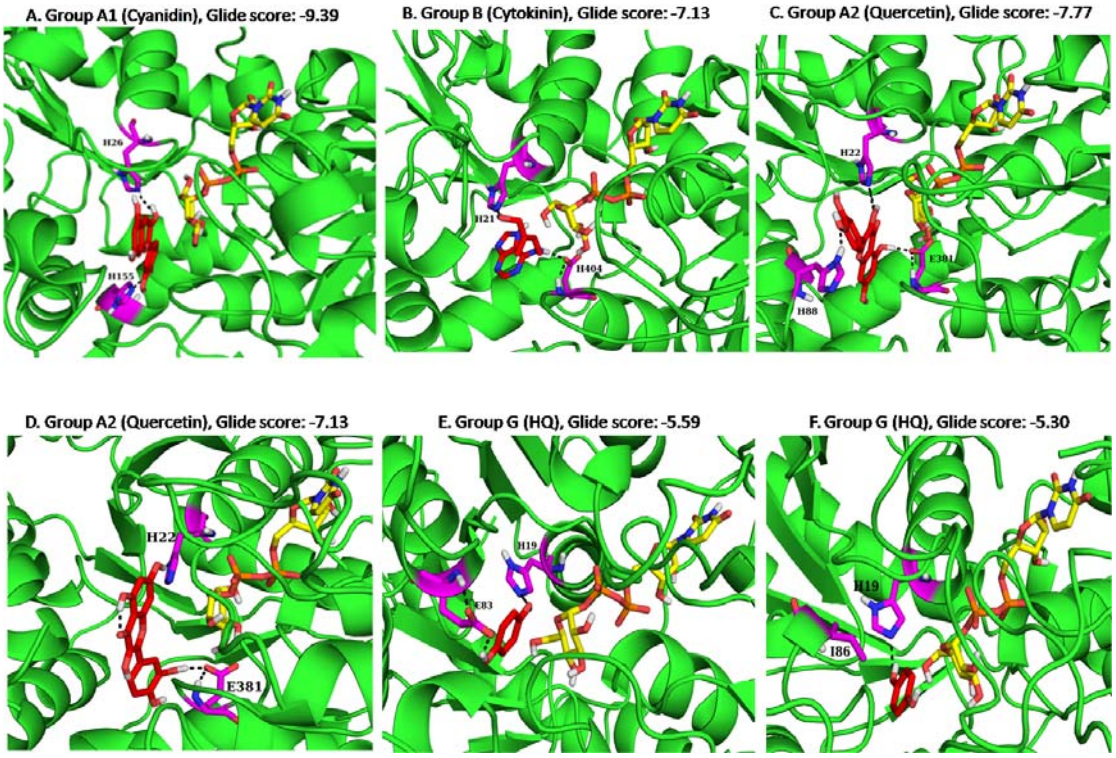


Figure 4.8: Docked complexes of *Ca*UGTs with their respective acceptor and sugar donor. A. The docked complex of *Ca*UGT of group A1 with cyanidin (shown in stick form) interacting with H26 and H155. B. The docked complex of *Ca*UGT of group B with cytokinin (shown in stick form) interacting with H21 and H404. C. The docked complex of *Ca*UGT of group A2 in which 3-OH group of quercetin (shown in stick form) interacting with H22. D. The docked complex of *Ca*UGT of group A2 in which 7-OH group of quercetin (shown in stick form) is pointing towards H22. E. The docked complex of *Ca*UGT of group E with hydroquinone (shown in stick form) interacting with H19 and E83 shown in stick form. F. The docked complex of *Ca*UGT of group G with hydroquinone (shown in stick form) interacting with H19.

The docking studies with the sugar acceptors showed higher affinity towards a specific substrate as compare to others. As shown in the dendrogram, group A1 members are shown to have high similarity with the Anthocyanidin 3-O and flavonol 3-O glycosyltransferase. The best docked complex of UGT78G2 of group A1 with an anthocyanidin named cyanidin showed high binding affinity in which its 3-OH group is interacting with the catalytic histidine (Figure 4.8A). The group B UGTs are specific towards the glycosylation of cytokinin at the oxygen (O-glycosylation). Docking studies revealed the interaction between oxygen and NE2 atom of catalytic histidine of UGT85H3 in the best docked complex (Figure 4.8B). The experimentally validated proteins of group A2 showed mixed specificity towards multiple hydroxyl groups of flavonoid, the docked complex between UGT71G2 and quercetin showed

the similar pattern of interactions (Figure 4.8C and 4.8D). Another group analyzed in this study was group G which is specific towards the glycosylation of hydroquinone (HQ). The conserved glutamic acid residue in region N3 and phenylalanine residue of region N4 involved in the hydrogen bond interaction and stacking interaction with the ring of hydroquinone identified by docking studies with *Solanum lycopersicum* UGT (Louveau *et al*, 2011) are present in UGT72 family of chickpea (Group G- glutamic acid present in UGT72B17 & UGT72B18). The docking studies have shown that the UGTs with glutamate present in the N3 region interact with the –OH group of HQ (Figure 4.8E). Contrary to this, glutamate is substituted by other residues in some of the proteins. In UGT72X1, isoleucine replaced this glutamate and the hydrogen bond at this site was lost (Figure 4.8F).

4.6 Identification of close orthologs and gene divergence

Among the four papilionideae plant genomes (*M. truncatula*, *G. max*, *V. angularis*, and *L. Japonicus*) on comparison with chickpea the maximum number of orthologs for *Ca*UGTs was detected in *M. truncatula* (143) while least number was found in *V. angularis* (only 2). Out of the 96 *Ca*UGTs, 87 had close orthologs in one of the four related dicot plants while nine UGTs seem diverged in chickpea (CD-Table S-4.2). The number of introns was found to be similar in their corresponding orthologs (except UGT83G4).

4.7 Intron incursion and deletion events

Out of the 96 *UGT* genes of chickpea 52 have no introns whereas 26 have one intron each in them (Figure 4.9 & 4.10). Two *UGTs* (UGT83G4 and UGT80B4) alone showed deviations as they contain 2 and 13 introns, respectively (CD-Table S-4.1). On the contrary ortholog of UGT83G4 (*M. truncatula*: XM_003621376) has only one intron, thus some intron gain or loss event would have occurred during the evolution. Gene length varies due to the presence of introns while the overall protein length is similar in almost all of them with an average protein length of 472 amino acids. Standard deviation (SD) of the protein length calculated showed maximum deviation for UGT95B1, UGT72AB2, UGT74Z1, UGT80B4, and UGT83G4 from mean value (Figure 4.11, CD-Table S-4.3).

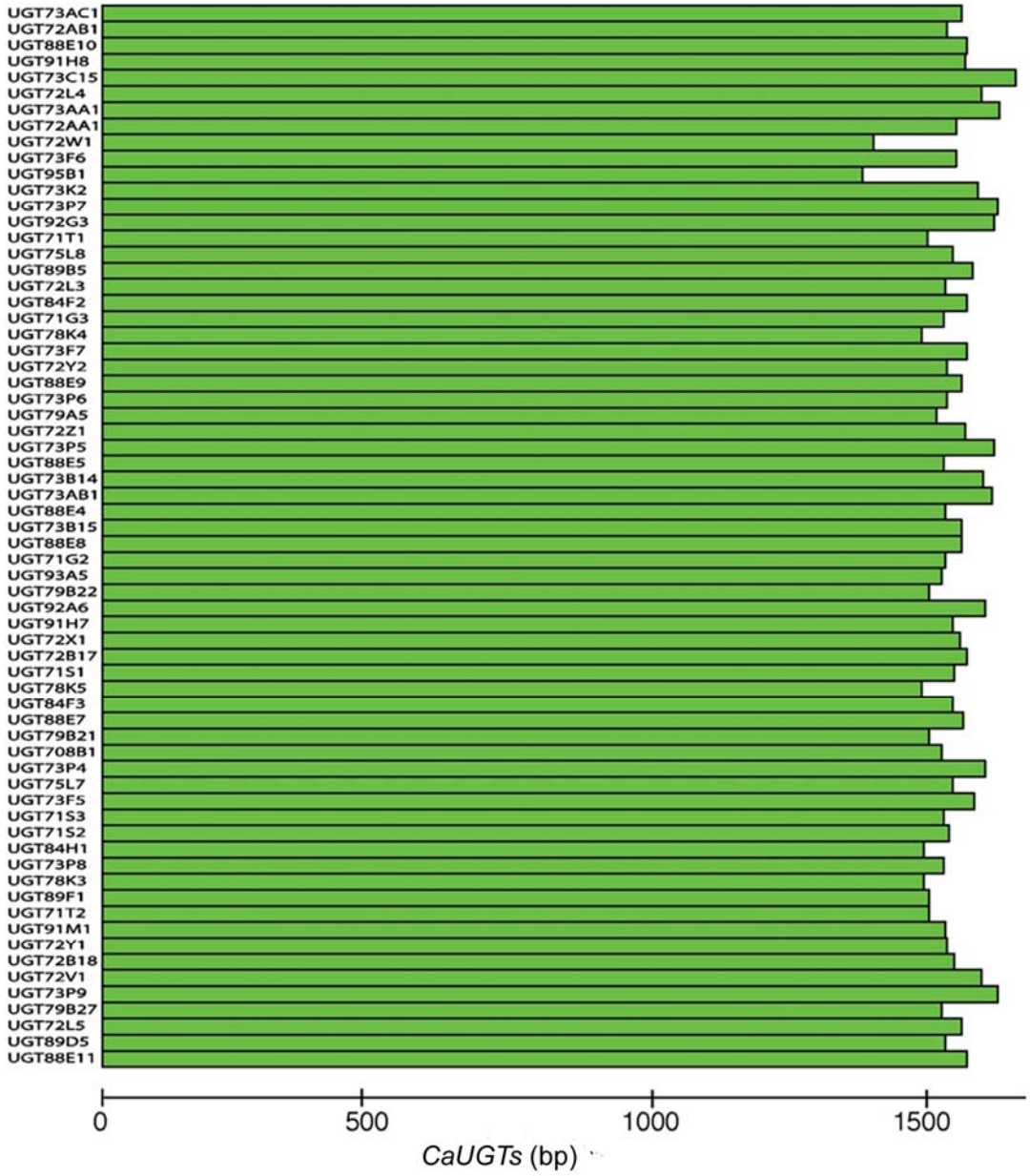


Figure 4.9: Gene architecture of 52 intronless *CaUGTs*.

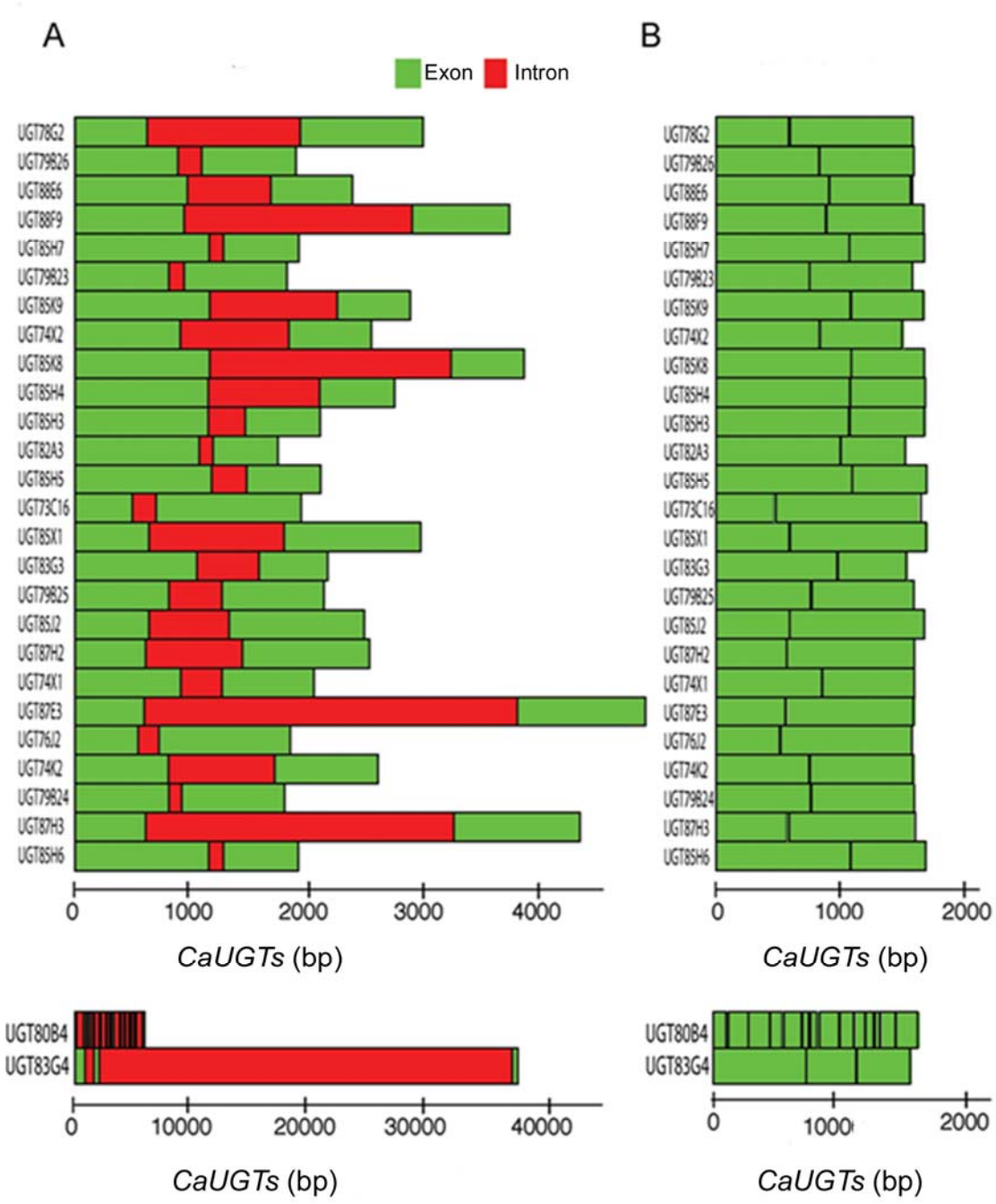


Figure 4.10: Exon-intron arrangements of *CaUGTs*. A: Length (bp) of *CaUGTs* with one or more than one introns. B: Exons length (bp) of *CaUGTs* of image A.

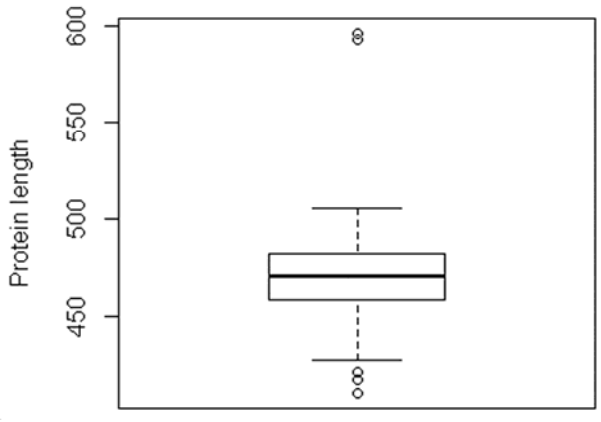


Figure 4.11: Standard deviation plot of protein length of chickpea UGTs. Five sequences deviated from the mean value.

4.8 Gene expression studies

4.8.1 *RNA-seq data*

Using RNA-seq data 84 *UGT* genes showed medium to high expression level (FPKM >= 5) in one or more tissue, whereas 10 *UGT* genes were lowly expressed (5 > FPKM > 0) and 2 (UGT72L5 and UGT87E3) showed no expression (FPKM = 0) in all the five tissues examined. Differential expression patterns were observed across the tissues with most of the *CaUGTs* showing highest expression in germinating seeds (Figure 4.12, Table 4.8). To investigate whether it is due to sample bias, as most of the genes are expressed highly in germinating seed tissue, we compared distribution of expression values (FPKM) for all the genes in five tissues considered in the study. Expression distribution didn't show any obvious bias (Figure 4.13). Upon analyzing the drought stressed RNA-seq reads from the two different genotypes of the chickpea it was seen that some of the genes were over-expressed (30 *UGTs*) as well as some are under-expressed as compared to the control samples (CD-Figure S-4.5, CD-Table S-4.4).

4.8.2 *EST data*

Gene expression for *CaUGTs* was identified also by carrying out blastn search against the chickpea EST database available at NCBI. We have used ≥ 90% sequence identity criteria to map the ESTs over gene models. Expression has been observed for 19 *UGTs* out of 96 in various tissue types such as root, shoot, stem and leaf. Out of 19 *CaUGTs*, 13 have shown expression in the root tissue of chickpea (Table 4.9). However, these 19 *UGTs* are also showing expression in the RNA-seq analysis

although the plant tissues tested happen to be different. The gene expression matches for specific genes when checked in the same tissues by using both the methods.

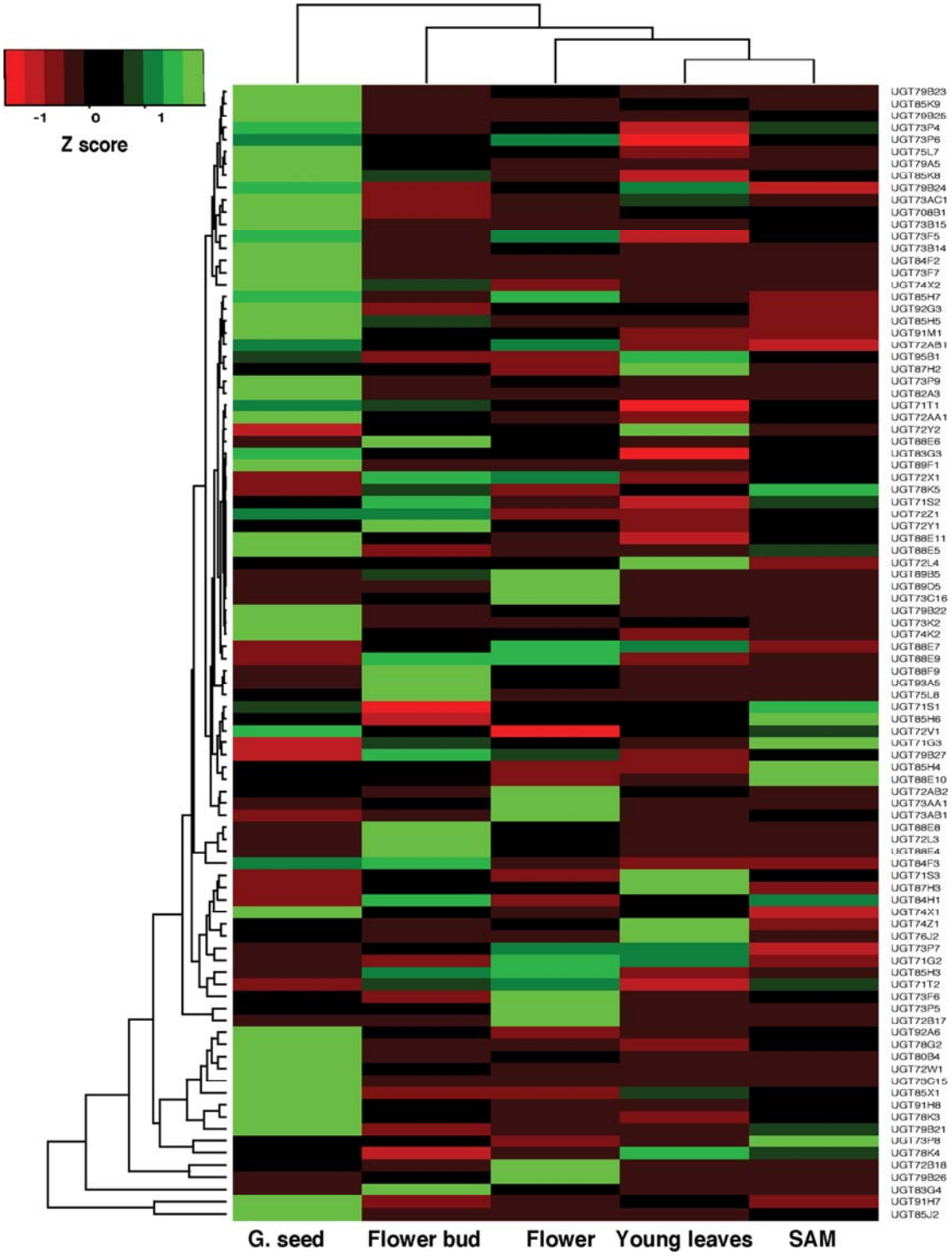


Figure 4.12: Heatmap showing relative genes expression in various tissue samples. The color scale (-1 to 1) represents Z-score, calculated by comparing fragment kilo base transcript per million (FPKM) value for UGT genes in different tissues. The *UGT* genes with FPKM > 0 are included in the analysis. Dendrogram on the top and side of the heatmap shows hierarchical clustering of tissues and genes using complete linkage approach.

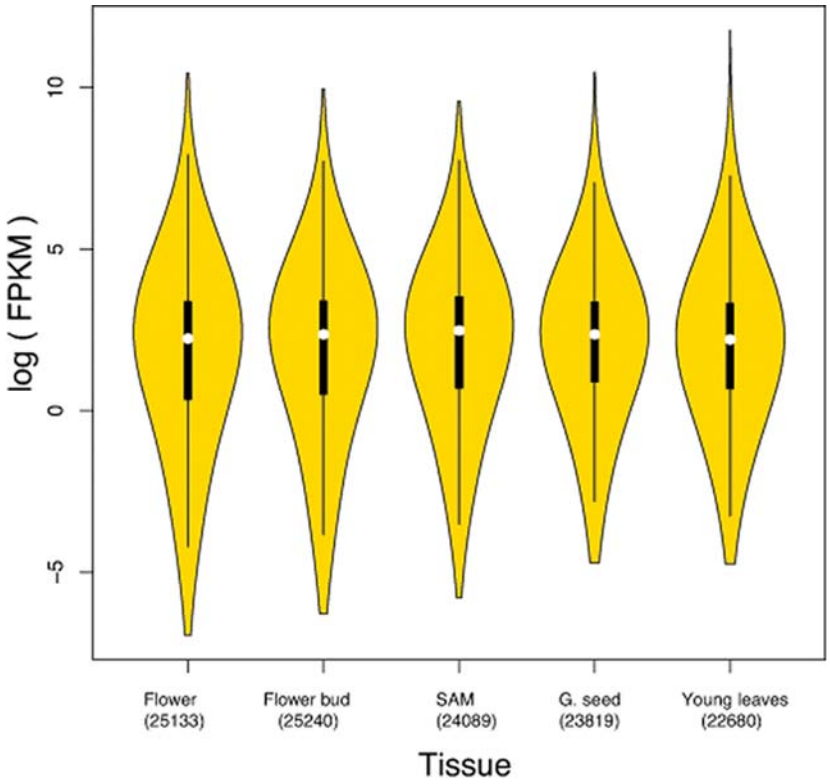


Figure 4.13: Violin plot representing distribution of FPKM values of all the expressed genes (FPKM > 0) in different tissues. Natural logarithm scale of FPKM values was plotted to reduce the range of FPKM values.

Table 4.8: Expression values (FPKM) of all the chickpea UGT genes in various plant tissues. The last two UGTs shown in bold are unexpressed in the five tissues under study.

No.	UGT	G. seed	SAM	Young leave	Flower	Flower bud
1	UGT85J2	277.91	25.88	3.08	0.95	22.3
2	UGT91H7	174.03	12.07	91.44	44.45	20.93
3	UGT91H8	118.51	26.17	2.48	2.61	17.79
4	UGT78K3	108.47	30.68	1.77	3.31	23.16
5	UGT79B21	104.43	60.6	15.5	20.12	11.42
6	UGT73C15	80.42	1.92	0.37	0.38	0.46
7	UGT72B18	67.91	13.44	16.28	179.16	28.55
8	UGT72W1	60.83	1.84	0.15	0.71	4.86
9	UGT73P8	60.31	98.93	51.92	41.71	58.9
10	UGT92A6	59.42	20.34	9.8	7.8	23.02
11	UGT78G2	58.02	19.09	2.52	5.93	9.53
12	UGT80B4	54.38	2.15	3.27	14.62	5.83
13	UGT85X1	53.81	23.03	34.32	11.08	8.92
14	UGT78K4	38.04	61.76	76.76	37	16.43
15	UGT84F3	36.86	5.95	7.67	10.02	50.53
16	UGT84F2	35.81	1.88	1.25	2.76	1.42
17	UGT74X2	32.01	4.08	2	0.59	17.05
18	UGT74X1	31.56	10.26	18.21	14.67	21.3
19	UGT73B14	30.66	1.71	0.56	14.18	1.47
20	UGT73AC1	28.24	9.33	17.13	9.55	4.85

21	UGT73B15	27.06	16.65	9.79	9.79	9.31
22	UGT708B1	26.29	11.66	12.73	5.51	2.7
23	UGT73F5	24.18	8.55	1.28	19.77	5.72
24	UGT73F7	24.18	0.79	1.27	0.96	0.63
25	UGT73F6	22.74	23.76	17.42	52.37	13.71
26	UGT79B26	22.11	11.05	8.91	125.95	47.63
27	UGT79B23	18.55	6.73	6.44	11.04	6.05
28	UGT79B25	18.51	6.31	3.87	5.53	5.33
29	UGT79B24	17.96	3.3	15.96	10.49	6.05
30	UGT85K9	16.46	6.61	8.52	7.24	6.08
31	UGT85K8	16.17	5.58	0.9	4.67	10.01
32	UGT79A5	14.23	2.92	1.84	2.08	5.86
33	UGT75L7	13.78	3.04	1.27	4.76	5.87
34	UGT74Z1	13.74	3.86	26.14	10.72	8.63
35	UGT87H3	13.73	13.68	17.31	14.61	14.77
36	UGT84H1	13.65	24.92	16.78	12.95	26.3
37	UGT73P7	13.6	7.46	34.28	34.6	21.53
38	UGT73P9	13.54	1.28	0.81	3.89	0.33
39	UGT73P4	13.16	10.9	3.59	9.01	6.64
40	UGT73P6	12.36	7.44	2.36	12.33	8.33
41	UGT73P5	12.2	6.44	3.11	83.24	11.46
42	UGT71S3	12.13	19.12	29.02	10.65	17.56
43	UGT82A3	12.12	0.07	0.02	0.06	0.05
44	UGT72V1	11.75	10.23	9.03	5.91	8.36
45	UGT71S1	10.61	12.6	9.75	8.81	4.95
46	UGT92G3	10.38	0.43	5.11	4.94	0.53
47	UGT76J2	9.38	8.04	31.37	6.93	4.51
48	UGT85H3	8.96	12.2	2.73	40.99	32.26
49	UGT85H4	8.42	14.06	3.46	2.42	6.51
50	UGT85H6	8.16	11.77	6.77	7.91	3.88
51	UGT85H7	8	2.88	3.66	8.02	3.73
52	UGT85H5	6.77	1.05	1.65	1.26	3.96
53	UGT95B1	6.18	5.05	8.8	1.21	1.21
54	UGT79B27	5.86	9.27	6.77	10.87	13.02
55	UGT75L8	5.77	2.53	1.8	2.65	17.28
56	UGT91M1	5.35	0.02	0	2.21	2.12
57	UGT71G3	5.17	17.91	8.95	9.82	13.14
58	UGT71G2	4.53	3.04	26.98	28.9	3.11
59	UGT88E10	4.3	16.77	0.77	0.58	4.3
60	UGT72AB1	4.22	1.05	1.82	4.3	3.3
61	UGT72AB2	4.13	1.23	4.42	26.74	3.01
62	UGT87H2	3.55	2.76	7.61	2.39	3.96
63	UGT79B22	3.06	0.1	0.07	0.28	0.2
64	UGT73AB1	2.83	12.17	3.55	23.79	4.63
65	UGT71T1	2.77	2.15	0.27	2.07	2.56
66	UGT71T2	2.53	21.44	2.06	25.24	20.26
67	UGT72AA1	2.5	1.63	1.24	1.26	1.46
68	UGT73AA1	2.35	1.46	0.68	16.46	4.97
69	UGT74K2	2.1	0.35	0.3	0.8	0.77
70	UGT73K2	2.01	0	0.14	0.03	0.01
71	UGT89F1	1.18	0.48	0.16	0.06	0.09
72	UGT88F9	1.12	0.55	0.15	6.2	19.46
73	UGT83G3	1.02	0.63	0.04	0.64	0.57
74	UGT83G4	0.79	1.01	1.66	14.44	245.84
75	UGT72B17	0.72	0.64	0.38	62.9	1.83

76	UGT88E9	0.58	1.36	0.15	6.84	6.88
77	UGT88E8	0.53	0.02	0.22	8.72	51.25
78	UGT88E7	0.51	0.64	6.76	7.89	2.67
79	UGT88E11	0.46	0.21	0	0.08	0.19
80	UGT88E5	0.42	0.24	0.02	0.03	0
81	UGT88E4	0.37	0.81	0.41	3.07	36.97
82	UGT72X1	0.26	0.46	0.24	0.68	0.76
83	UGT72Z1	0.24	0.11	0	0.01	0.23
84	UGT72Y2	0.22	0.76	2.52	1.55	1.18
85	UGT72Y1	0.19	0.17	0.04	0.15	0.53
86	UGT78K5	0.18	0.85	0.41	0.12	0.57
87	UGT71S2	0.1	0.2	0	0.05	0.28
88	UGT73C16	0.09	0.02	0	0.99	0.12
89	UGT93A5	0.06	0.17	0.08	2.96	17.61
90	UGT72L4	0.06	0.02	0.18	0.08	0.06
91	UGT72L3	0	0	0.04	9.83	45.69
92	UGT89B5	0	0	0	0.11	0.07
93	UGT89D5	0	0	0	0.05	0
94	UGT88E6	0	0.45	0.04	0.41	4.3
95	**UGT72L5**	**0**	**0**	**0**	**0**	**0**
96	**UGT87E3**	**0**	**0**	**0**	**0**	**0**

Table 4.9: Description of *Cicer arietinum* EST BLAST hits against the chickpea dbEST in NCBI. QC: Query coverage

No	UGT	Accession	Max score	Total score	QC (%)	E-value	Identity (%)	Tissue
1	UGT85H3	GR400167.1	520	520	25	$5*10^{-147}$	91	Root
2	UGT85H4	GR400167.1	520	520	25	$5*10^{-147}$	91	Root
3	UGT73K2	GR406264.1	407	407	15	$8*10^{-113}$	99	Root
4	UGT71S1	FE669840.1	783	783	30	0	100	Leaf
5	UGT88E9	GR399168.1	710	710	27	0	100	Root
		GR402050.1	688	688	26	0	99	Root
		GR401482.1	688	688	26	0	99	Root
		GR401481.1	578	578	22	$2*10^{-164}$	99	Root
6	UGT85H7	GR400167.1	614	614	26	$3*10^{-175}$	95	Root
7	UGT88E7	GR406820.1	320	320	27	$9*10^{-87}$	91	Root
8	UGT88E8	GR406820.1	407	407	16	$8*10^{-113}$	99	Root
9	UGT72B18	FE672275.1	964	964	37	0	100	Leaf
		GR406581.1	890	890	34	0	100	Root
		CV793607.1	654	654	34	0	90	Stem, leaf
		JG292965.1	331	331	12	$5*10^{-90}$	100	Shoot
10	UGT72AB1	HO066771.1	268	268	13	$5*10^{-71}$	90	Shoot
11	UGT71S3	FE670249.1	547	547	22	$4*10^{-155}$	99	Leaf
12	UGT73AC1	HS108799.1	805	805	31	0	99	Leaf
13	UGT83G3	GR403205.1	614	614	25	$3*10^{-175}$	100	Root
		GR404644.1	309	309	12	$2*10^{-83}$	100	Root

14	UGT75L8	GR392042.1	1272	1272	50	0	99	Root
		FE672018.1	1059	1059	41	0	100	Leaf
		GR390934.1	637	637	25	0	99	Root
15	UGT74K2	GR404104.1	654	654	26	0	100	Root
		GR403152.1	612	612	24	$1*10^{-174}$	100	Root
		GR402336.1	482	482	19	$1*10^{-135}$	99	Root
16	UGT89D5	HO066872.1	329	368	13	$2*10^{-89}$	98	Shoot
		HO066868.1	322	357	13	$3*10^{-87}$	97	Shoot
17	UGT87H2	GR407308.1	612	612	24	$1*10^{-174}$	100	Root
18	UGT79B26	GR408726.1	522	522	21	$1*10^{-147}$	99	Root
		GR397872.1	416	416	16	$2*10^{-115}$	100	Root
19	UGT72AB2	HO066771.1	340	340	16	$1*10^{-92}$	95	Shoot

4.9 Conclusion

Glycosyltransferases are part of an essential multigene family present in all species including bacteria, fungi, animals, plants etc. In plants, they perform glycosylation of important plant products which helps in their proper functioning as well as survival in adverse situations. Genome sequencing projects help the researchers to analyze the new data and get useful information out of it. Gene identification methods based on biochemical studies and characterization are difficult as well as time consuming, therefore in the present research identification of novel *UGT* genes of chickpea was carried out by screening the signature motif of UGTs as well as by aligning HMM profile of UDPGT family with the predicted proteome. Very few sequences were identified exclusively by MEME-MAST and HMM profile search but not by blast search. None of these sequences possess the key features of UGTs therefore might be considered as false positive hits**.** Two possible recent gene duplication events and nine diverged *Ca*UGTs were found. Maximum number of *CaUGT* genes has only one intron in them while two *CaUGTs* have two introns each and one has thirteen introns. The phylogenetic tree can be useful to deduce the structure-function relationship of these predicted UGTs and further assist in their functional analysis. The phylogenetic analysis carried out combining with the UGTs of known specificity helped us to achieve functional assignment of 74 chickpea UGTs.

Our results are consistent with the previous findings that expression of *UGTs* was localized to regions of rapidly dividing cells (Woo *et al*, 2007). High expression of *UGTs* coinciding with tissues involved in intense cell division (germinating seeds,

flower etc.) indicates possible involvement in cell cycle regulation. Gene expression analysis not only confirmed that 84 (out of 96) *CaUGTs* significantly expressed but also revealed tissue-specific role as possible explanation for their high content.

Phylogenetic tree generated by exploiting the activity information of other experimentally validated proteins revealed distinct clustering for all the 15 identified groups. The eight regions, identified by us in our study presented in previous chapter, in the proximity of the sugar acceptor were found to be highly conserved, which shows their selectivity towards specific sugar acceptors. The above findings were further supported by the docking simulation studies. These findings are very useful in assigning the putative functions to the identified chickpea UGTs and can be further validated by experimental approaches.

UGT class of enzymes constitutes approximately 0.4% of the total predicted chickpea proteome, which is quite a significant number for one particular class of enzyme. If we consider other sequenced genomes like *Arabidopsis thaliana* and *Oryza sativa*, similar pattern of occurrence of UGT genes has been observed (Li *et al*, 2000; Cao *et al*, 2008). Such high abundance of ubiquitous GT family in any plant genome must have indispensable role in the glycosylation of diverse array of acceptor substrates and perform distinct functions. Previous studies have shown the role of higher duplication rate behind the expansion and high content of UGT gene family in a genome (Yonekura-Sakakibara & Hanada, 2011; Caputi *et al*, 2012). The phylogenetic analysis of chickpea UGTs can be beneficial for understanding the structure-function relationship and might further assist in their functional analysis. Identification of novel chickpea UGTs helps in developing genetically modified genes and their products with improved properties and thus to develop plants that react efficiently to adverse or stress conditions.

Chapter 5

Genome-wide identification and structure- function studies of proteases and protease inhibitors of chickpea

In chickpea, total crop loss caused by abiotic stress exceeds those due to biotic stress. Drought, salinity, and cold are the major environmental factors causing crop damage. Along with the above mentioned abiotic stress, several biotic factors like *Ascochyta* blight, *Botrytis* grey mould, *Fusarium* wilt etc are also one of the major rate limiting factors affecting production of chickpea (Yadav *et al*, 2007). In order to cope up with such adverse conditions, plants posses a large proteolytic machinery which along with the hydrolysis of non-functional proteins of the cells, also take part in several biological processes like recognition of pathogens and pests and the induction of effective defense responses, PCD, during water deficiency stress (Van der Hoorn & Jones, 2004; Xia *et al*, 2004; Izuhara *et al*, 2008). Plant protease inhibitors (PPIs) play an important role in plant defense mechanism in response to attack by insects or pathogenesis and wounding by inhibiting the proteases present in the insect gut that results in reduced amount of amino acids necessary for their growth and development (Habib & Fazili KM, 2007). In 2014 Yan *et al.* reported the genome-wide analysis of regulatory proteases in *Taenia solium* genome. In this chapter, an *in silico* search of both proteases and PIs in chickpea genome was conducted. A significant number of members belonging to four protease families (aspartate protease, cysteine protease, serine protease, and metalloprotease) and PIs (cysteine and serine protease inhibitors) were identified in the chickpea proteome (1.28% of the total chickpea genome encodes proteases out of which 0.8% are with catalytic residues). The binding studies of the PIs with their targets were carried out to get more insight about the mode of binding and affinity.

5.1 Identification of proteases and protease inhibitors

The protease and protease inhibitors were searched against the chickpea proteome using the HMM profiles provided in pfam 27.0 database of respective families following similar methodology explained in materials and methods (Table 5.1 & 5.2).

Table 5.1: The HMM profiles of proteases and protease inhibitors employed for the gene identification study.

Protein	HMM profile
Aspartate proteases	PF00026
Cysteine proteases	PF00112
Cysteine protease inhibitor	PF00031
Metalloproteases	
M41	PF01434
M48	PF01435
M50	PF02163
Serine protease inhibitors	
Bowman-Birk inhibitor	PF00228
Potato inhibitor I	PF00280
Potato inhibitor II	PF02428
Serpin	PF00079
Kunitz inhibitor	PF00197
Kazal	PF07648
Squash	PF00299

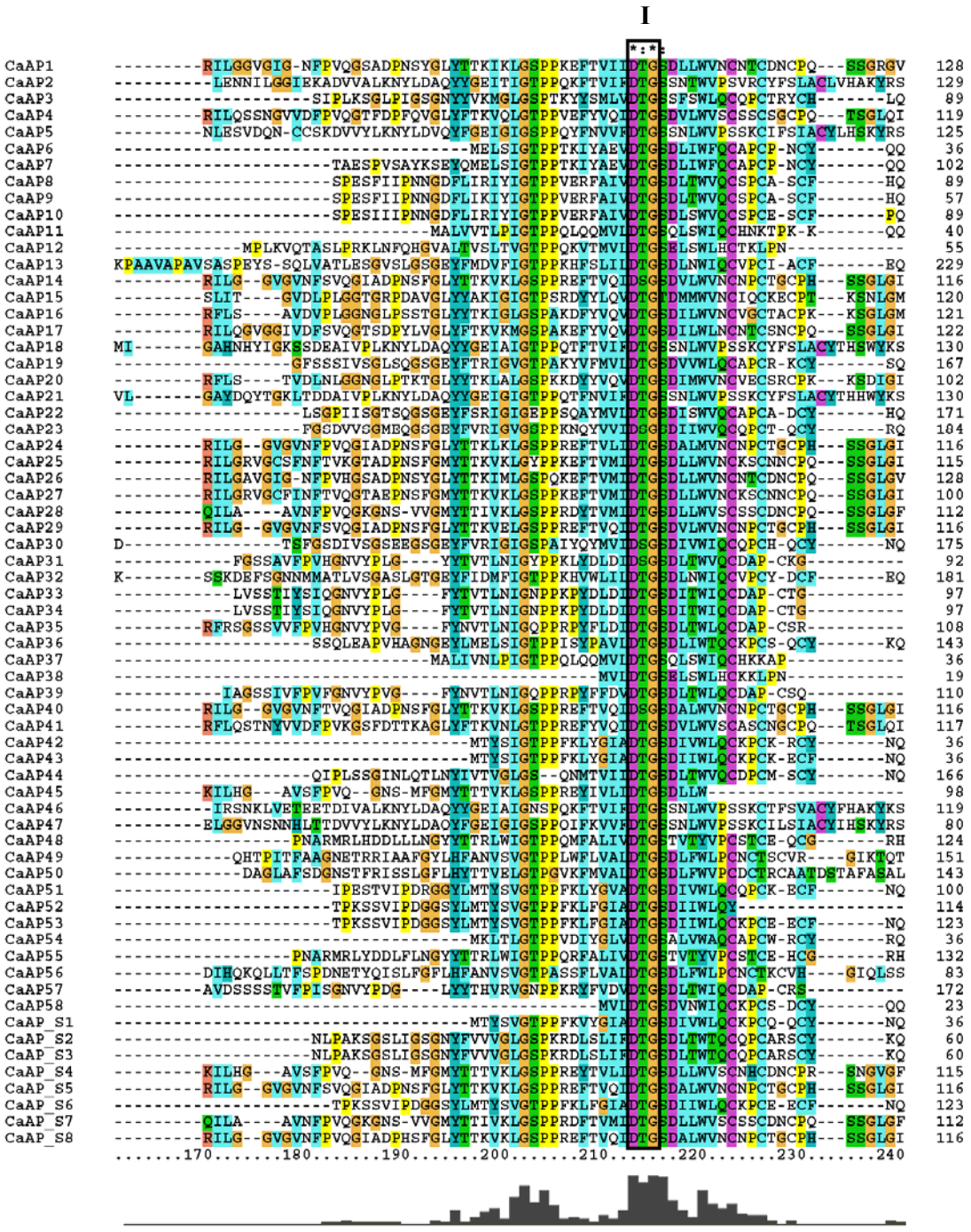
Table 5.2: Details of the HMM profiles of 13 classes of serine protease employed for the gene identification study.

Serine protease class	HMM profile
Trypsin (S1)	PF00089
Clp Endopeptidase (S14)	PF00574
C-terminal processing peptidase (S41)	PF03572
Lon protease (S16)	PF05362
Lys-Pro-x Carboxypeptidase (S28)	PF05577
Nucleoporin autopeptidase (S59)	PF04096
Prolyl oligopeptidase (S9)	PF00326
Protease IV (S49)	PF01343
Rhomboid (S54)	PF01694

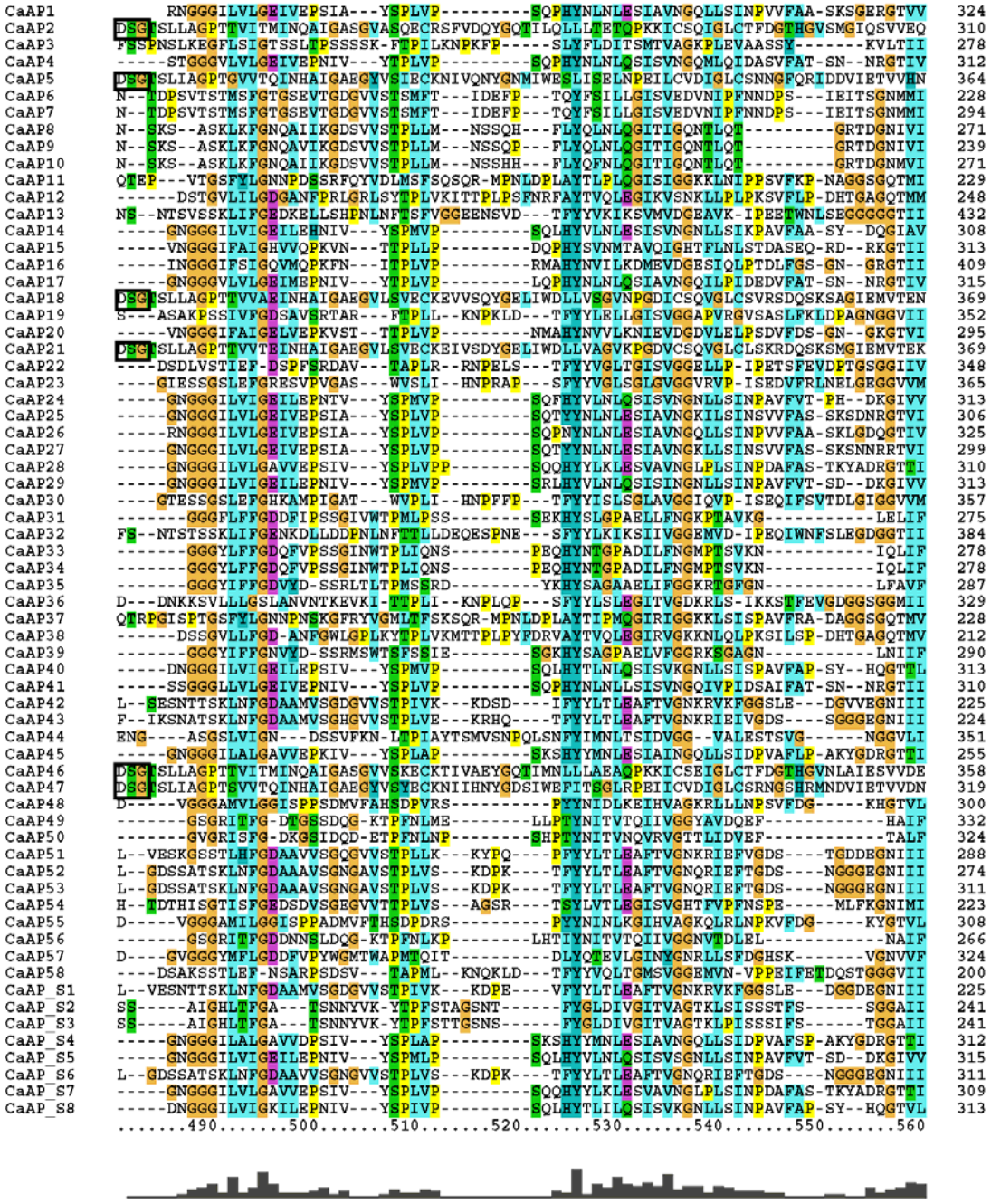
Serine carboxypeptidase (S10)	PF00450
Signal peptidase I (S24)	PF00717
Subtilase (S8)	PF00082
D-Ala-D-Ala carboxypeptidase B (S12)	PF00144

5.1.1 Aspartate protease (AP)

A hidden Markov model profile (PF00026) search using HMMER (hmmsearch) resulted in the identification of a total of 89 *Ca*APs (*C. arietinum* APs), out of which 66 sequences had the two complete Asp-Thr/Ser-Gly motifs (Figure 5.1). The remaining sequences had either an incomplete ASP domain or present in the data only a single motif. Out of these 66 sequences, genomic locations of 56 sequences were known and named from *Ca*AP1-*Ca*AP56 based on their order on the chromosomes. The remaining 8 sequences were present on scaffolds and named as *Ca*AP_S1 to S8. The GenBank accession numbers were assigned for 16 sequences [GenBank: KJ151780-KJ151790, KJ561459-KJ561463] (CD-Table S1). Almost all known aspartyl proteases are inhibited by an antimicrobial hexapeptide Pepstatin.


I
CaAP1 128
CaAP2 129
CaAP3 89
CaAP4 119
CaAP5 125
CaAP6 36
CaAP7 102
CaAP8 89
CaAP9 57
CaAP10 89
CaAP11 40
CaAP12 55
CaAP13 229
CaAP14 116
CaAP15 120
CaAP16 121
CaAP17 122
CaAP18 130
CaAP19 167
CaAP20 102
CaAP21 130
CaAP22 171
CaAP23 104
CaAP24 116
CaAP25 115
CaAP26 128
CaAP27 100
CaAP28 112
CaAP29 116
CaAP30 175
CaAP31 92
CaAP32 181
CaAP33 97
CaAP34 97
CaAP35 108
CaAP36 143
CaAP37 36
CaAP38 19
CaAP39 110
CaAP40 116
CaAP41 117
CaAP42 36
CaAP43 36
CaAP44 166
CaAP45 98
CaAP46 119
CaAP47 80
CaAP48 124
CaAP49 151
CaAP50 143
CaAP51 100
CaAP52 114
CaAP53 123
CaAP54 36
CaAP55 132
CaAP56 83
CaAP57 172
CaAP58 23
CaAP_S1 36
CaAP_S2 60
CaAP_S3 60
CaAP_S4 115
CaAP_S5 116
CaAP_S6 123
CaAP_S7 112
CaAP_S8 116
170 180 190 200 210 220 230 240

II


CaAP1 RNGGGILVLGEIVEPSIA---YSPLVP---------SQPHYNLNLESIAVNGQLLSINPVVFAA-SKSGERGTVV 324
CaAP2 DSGISLLAGPTTVITMINQAIGASGVASQECRSFVDQYGQTILQLLLTETQPKKICSQIGLCTFDGTHGVSMGIQSVVEQ 310
CaAP3 FSSPNSLKEGFLSIGTSSLTPSSSSK-FTPILKNPKFP-----SLYFLDITSMTVAGKPLEVAASSY-------KVLTII 278
CaAP4 -----STGGGVLVLGEIVEPNIV---YTPLVP---------SQPHYNLNLQSISVNGQMLQIDASVFAT-SN--NRGTIV 312
CaAP5 DSGISLIAGPTGVVTQINHAIGAEGYVSIECKNIVQNYGNMIWESLISELNPEILCVDIGLCSNNGFQRIDDVIETVVHN 364
CaAP6 N--TDPSVTSTMSFGTGSEVTGDGVVSTSMFT---IDEFP---TQYFSILLGISVEDVNIPFNNDPS---IEITSGNMMI 228
CaAP7 N--TDPSVTSTMSFGTGSEVTGDGVVSTSMFT---IDEFP---TQYFSILLGISVEDVNIPFNNDPS---IEITSGNMMI 294
CaAP8 N--SKS--ASKLKFGNQAIIKGDSVVSTPLLM---NSSQH---FLYQLNLQGITIGQNTLQT---------GRTDGNIVI 271
CaAP9 N--SKS--ASKLKFGNQAVIKGDSVVSTPLLM---NSSQP---FLYQLNLQGITIGQNTLQT---------GRTDGNIVI 239
CaAP10 N--SKS--ASKLKFGNQAIIKGDSVVSTPLLM---NSSHH---FLYQFNLQGITIGQNTLQT---------GRTDGNMVI 271
CaAP11 QTEP---VTGSFYLGNNPDSSRFQYVDLMSFSQSQR-MPNLDPLAYTLPLQGISIGGKKLNIPPSVFKP-NAGGSGQTMI 229
CaAP12 ------DSTGVLILGDGANFPRLGRLSYTPLVKITTPLPSFNRFAYTVQLEGIKVSNKLLPLPKSVFLP-DHTGAGQTMM 248
CaAP13 NS--NTSVSSKLIFGEDKELLSHPNLNFTSFVGGEENSVD---TFYYVKIKSVMVDGEAVK-IPEETWNLSEGGGGGTII 432
CaAP14 -----GNGGGILVIGEILEHNIV---YSPMVP---------SQLHYVLNLESISVNGNLLSIKPAVFAA-SY--DQGIAV 308
CaAP15 -----VNGGGIFAIGHVVQPKVN---TTPLLP---------DQPHYSVNMTAVQIGHTFLNLSTDASEQ-RD--RKGTII 313
CaAP16 -----INGGGIFSIGQVMQPKFN---ITPLVP---------RMAHYNVILKDMEVDGESIQLPTDLFGS-GN--GRGTII 409
CaAP17 -----GNGGGVLVLGEIMEPNIV---YTPLVP---------LQPHYNLNLQSIAVNGQILPIDEDVFAT-SN--NRGTIV 315
CaAP18 DSGISLLAGPTTVVAEINHAIGAEGVLSVECKEVVSQYGELIWDLLVSGVNPGDICSQVGLCSVRSDQSKSAGIEMVTEN 369
CaAP19 S---ASAKPSSIVFGDSAVSRTAR---FTPLL--KNPKLD---TFYYLELLGISVGGAPVRGVSASLFKLDPAGNGGVII 352
CaAP20 -----VNGGGIFAIGELVEPKVST--TTPLVP---------NMAHYNVVLKNIEVDGDVLELPSDVFDS-GN--GKGTVI 295
CaAP21 DSGISLLAGPTTVVTEINHAIGAEGVLSVECKEIVSDYGELIWDLLVAGVKPGDVCSQVGLCLSKRDQSKSMGIEMVTEK 369
CaAP22 ----DSDLVSTIEF-DSPFSRDAV---TAPLR--RNPELS---TFYYVGLTGISVGGELLP-IPETSFEVDPTGSGGIIV 348
CaAP23 ----GIESSGSLEFGRESVPVGAS---WVSLI--HNPRAP---SFYYVGLSGLGVGGVRVP-ISEDVFRLNELGEGGVVM 365
CaAP24 -----GNGGGILVIGEILEPNTV---YSPMVP---------SQFHYVLNLQSISVNGNLLSINPAVFVT-PH--DKGIVV 313
CaAP25 -----GNGGGILVLGEIVEPSIA---YSPLVP---------SQTYYNLNLESIAVNGKILSINSVVFAS-SKSDNRGTVI 306
CaAP26 -----RNGGGILVLGEIVEPSIA---YSPLVP---------SQPNYNLNLESIAVNGQLLSINPVVFAA-SKLGDQGTIV 325
CaAP27 -----GNGGGILVLGEIVEPSIA---YSPLVP---------SQTYYNLNLESIAVNGKLLSINSVVFAS-SKSNNRRTVI 299
CaAP28 -----GNGGGILVLGAVVEPSIV---YSPLVPP--------SQQHYYLKLESVAVNGLPLSINPDAFAS-TKYADRGTTI 310
CaAP29 -----GNGGGILVIGEILEPNIV---YSPMVP---------SRLHYVLNLQSISINGNLLSINPAVFVT-SD--DKGIVV 313
CaAP30 ----GTESSGSLEFGHKAMPIGAT---WVPLI--HNPFFP---TFYYISLSGLAVGGIQVP-ISEQIFSVTDLGIGGVVM 357
CaAP31 -------GGGFLFFGDDFIPSSGIVWTPMLPSS--------SEKHYSLGPAELLFNGKPTAVKG----------LELIF 275
CaAP32 FS--NTSTSSKLIFGENKDLLDDPNLNFTTLLDEQESPNE---SFYYLKIKSIIVGGEMVD-IPEQIWNFSLEGDGGTII 384
CaAP33 -------GGGYLFFGDQFVPSSGINWTPLIQNS--------PEQHYNTGPADILFNGMPTSVKN----------IQLIF 278
CaAP34 -------GGGYLFFGDQFVPSSGINWTPLIQNS--------PEQHYNTGPADILFNGMPTSVKN----------IQLIF 278
CaAP35 -------GGGYIFFGDVYD-SSRLTLTPMSSRD---------YKHYSAGAAELIFGGKRTGFGN----------LFAVF 287
CaAP36 D---DNKKSVLLLGSLANVNTKEVKI-TTPLI--KNPLQP---SFYYLSLEGITVGDKRLS-IKKSTFEVGDGGSGGMII 329
CaAP37 QTRPGISPTGSFYLGNNPNSKGFRYVGMLTFSKSQR-MPNLDPLAYTIPMQGIRIGGKKLSISPAVFRA-DAGGSGQTMV 228
CaAP38 ------DSSGVLLFGD-ANFGWLGPLKYTPLVKMTTPLPYFDRVAYTVQLEGIRVGKKNLQLPKSILSP-DHTGAGQTMV 212
CaAP39 -------GGGYIFFGNVYD-SSRMSWTSFSSIE--------SGKHYSAGPAELVFGGRKSGAGN----------LNIIF 290
CaAP40 -----DNGGGILVIGEILEPSIV---YSPMVP---------SQLHYTLNLQSISVKGNLLSISPAVFAP-SY--HQGTTL 313
CaAP41 -----SSGGGLLVLGEIVEPNIV---YSPLVP---------SQPHYNLNLLSISVNGQIVPIDSAIFAT-SN--NRGTII 310
CaAP42 L--SESNTTSKLNFGDAAMVSGDGVVSTPIVK---KDSD----IFYYLTLEAFTVGNKRVKFGGSLE---DGVVEGNIII 225
CaAP43 F--IKSNATSKLNFGDAAMVSGHGVVSTPLVE---KRHQ----TFYYLTLEAFTVGNKRIEIVGDS----SGGGEGNIII 224
CaAP44 ENG----ASGSLVIGN---DSSVFKN-LTPIAYTSMVSNPQLSNFYIMNLTSIDVGG--VALESTSVG------NGGVLI 351
CaAP45 -----GNGGGILALGAVVEPKIV---YSPLAP---------SKSHYYMNLESIAINGQLLSIDPVAFLP-AKYGDRGTTI 255
CaAP46 DSGISLLAGPTTVITMINQAIGASGVVSKECKTIVAEYGQTIMNLLLAEAQPKKICSEIGLCTFDGTHGVNLAIESVVDE 358
CaAP47 DSGISLIAGPTGVVTQINHAIGAEGYVSYECKNIIHNYGDSIWEFITSGLRPEIICVDIGLCSRNGSHRMNDVIETVVDN 319
CaAP48 D-----VGGGAMVLGGISPPSDMVFAHSDPVRS----------PYYNIDLKEIHVAGKRLLLNPSVFDG-----KHGTVL 300
CaAP49 -------GSGRITFG-DTGSSDQG-KTPFNLME--------LLPTYNITVTQIIVGGYAVDQEF----------HAIF 332
CaAP50 -------GVGRISFG-DKGSIDQD-ETPFNLNP--------SHPTYNITVNQVRVGTTLIDVEF----------TALF 324
CaAP51 L--VESKGSSTLHFGDAAVVSGQGVVSTPLLK---KYPQ----PFYYLTLEAFTVGNKRIEFVGDS----TGDDEGNIII 288
CaAP52 L--GDSSATSKLNFGDAAAVSGNGAVSTPLVS---KDPK----TFYYLTLEAFTVGNQRIEFTGDS----NGGGEGNIII 274
CaAP53 L--GDSSATSKLNFGDAAAVSGNGAVSTPLVS---KDPK----TFYYLTLEAFTVGNQRIEFTGDS----NGGGEGNIII 311
CaAP54 H--TDTHISGTISFGEDSDVSGEGVVTTPLVS---AGSR----TSYLVTLEGISVGHTFVPFNSPE-----MLFKGNIMI 223
CaAP55 D-----VGGGAMILGGISPPADMVFTHSDPDRS----------PYYNINLKGIHVAGKQLRLNPKVFDG-----KYGTVL 308
CaAP56 -------GSGRITFGDDNNSLDQG-KTPFNLKP--------LHTIYNITVTQIIVGGNVTDLEL----------NAIF 266
CaAP57 D----GVGGGYMFLGDDFVPYWGMTWAPMTQIT----------DLYQTEVLGINYGNRLLSFDGHSK-------VGNVVF 324
CaAP58 ----DSAKSSTLEF-NSARPSDSV---TAPML--KNQKLD---TFYYVQLTGMSVGGEMVN-VPPEIFETDQSTGGGVII 200
CaAP_S1 L--VESNTTSKLNFGDAAMVSGDGVVSTPIVK---KDPE----VFYYLTLEAFTVGNKRVKFGGSLE---DGGDEGNIII 225
CaAP_S2 SS-----AIGHLTFGA---TSNNYVK-YTPFSTAGSNT------FYGLDIVGITVAGTKLSISSSTFS------SGGAII 241
CaAP_S3 SS-----AIGHLTFGA---TSNNYVK-YTPFSTTGSNS------FYGLDIVGITVAGTKLPISSSIFS------TGGAII 241
CaAP_S4 -----GNGGGILALGAVVDPSIV---YSPLAP---------SKSHYYMNLESIAVNGQLLSIDPVAFSP-AKYGDRGTTI 312
CaAP_S5 -----GNGGGILVIGEILEPNIV---YSPMLP---------SQLHYVLNLQSISVSGNLLSINPAVFVT-SD--DKGIVV 315
CaAP_S6 L--GDSSATSKLNFGDAAVVSGNGVVSTPLVS---KDPK----TFYYLTLEAFTVGNQRIEFTGDS----NGGGEGNIII 311
CaAP_S7 -----GNGGGILVLGAVVEPSIV---YSPLVP---------SQQHYYLKLESVAVNGLPLSINPDAFAS-TKYADRGTTI 309
CaAP_S8 -----DNGGGILVIGKILEPNIV---YSPIVP---------SQLHYTLILQSISVKGNLLSINPAVFAP-SY--HQGTVL 313
.....490.....500.....510.....520.....530.....540.....550.....560

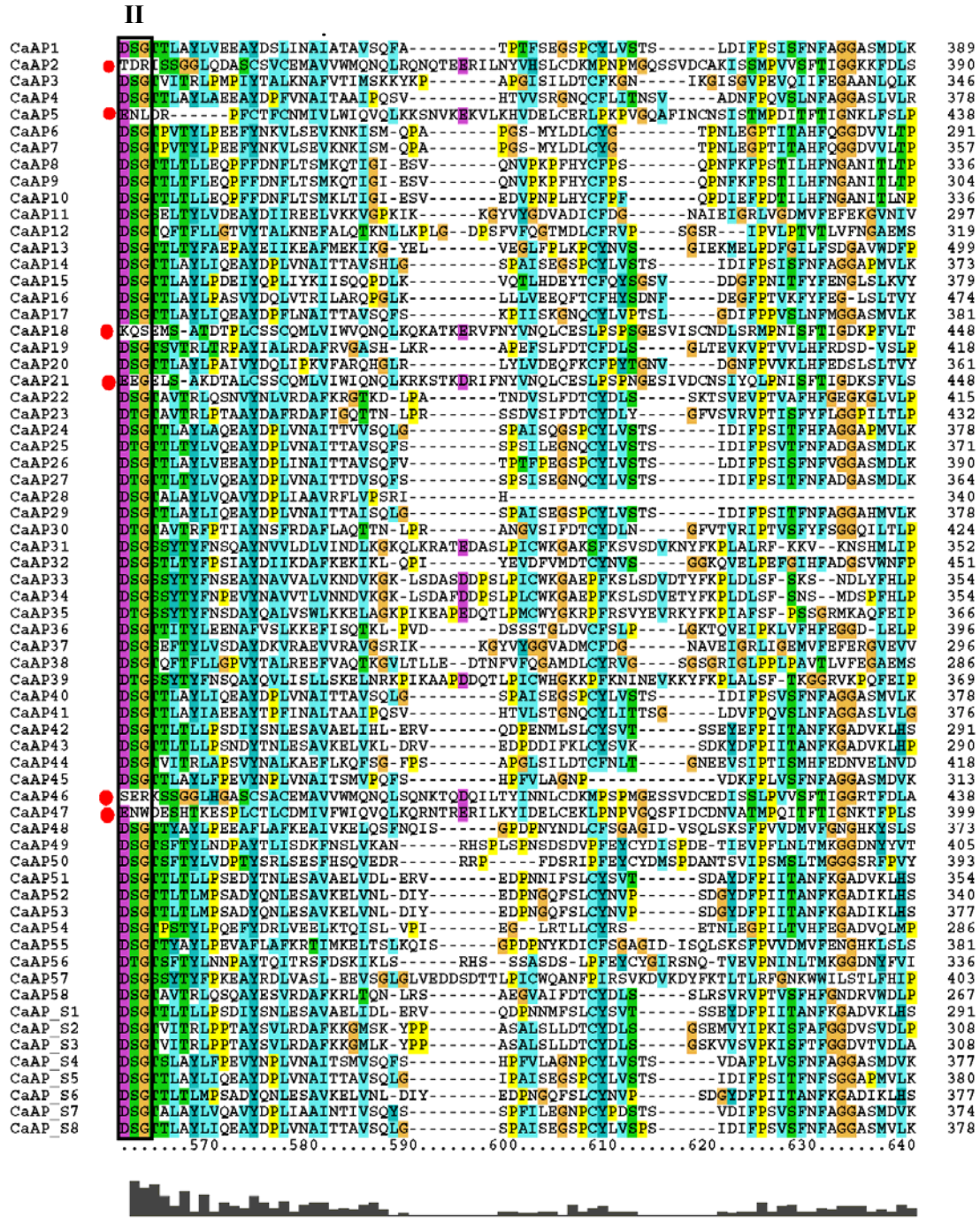


Figure 5.1: Multiple sequence alignment of aspartate proteases of chickpea. The two ASP domains are enclosed in the boxes. The sequences with the red circles marked belong to typical APs with a different position of the second ASP domain.

5.1.2 Cysteine protease (CP)

The HMM profile of cysteine proteases (PF00112) aided in the identification of 33 *Ca*CPs (*C. arietinum* CPs), out of which 28 sequences had all the three catalytic residues i.e. Cys, His, and Asn (Figure 5.2). 26 genes were mapped on chromosomes and their nomenclature was done in the similar way as done for *Ca*APs. GenBank accession numbers were assigned to 8 *Ca*CP sequences [GenBank: KJ151791-KJ151797, KJ579708] (CD-Table S-1).

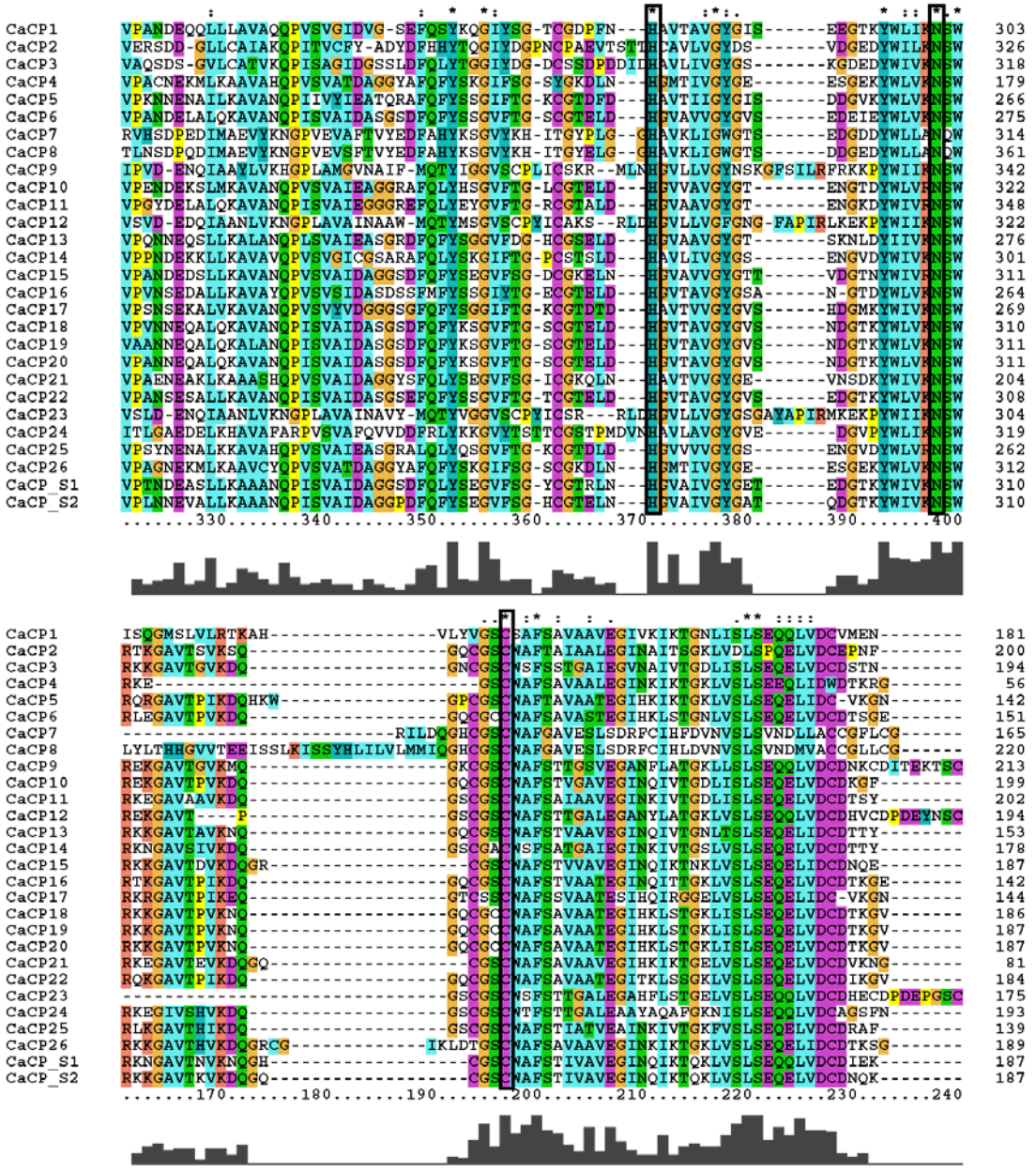


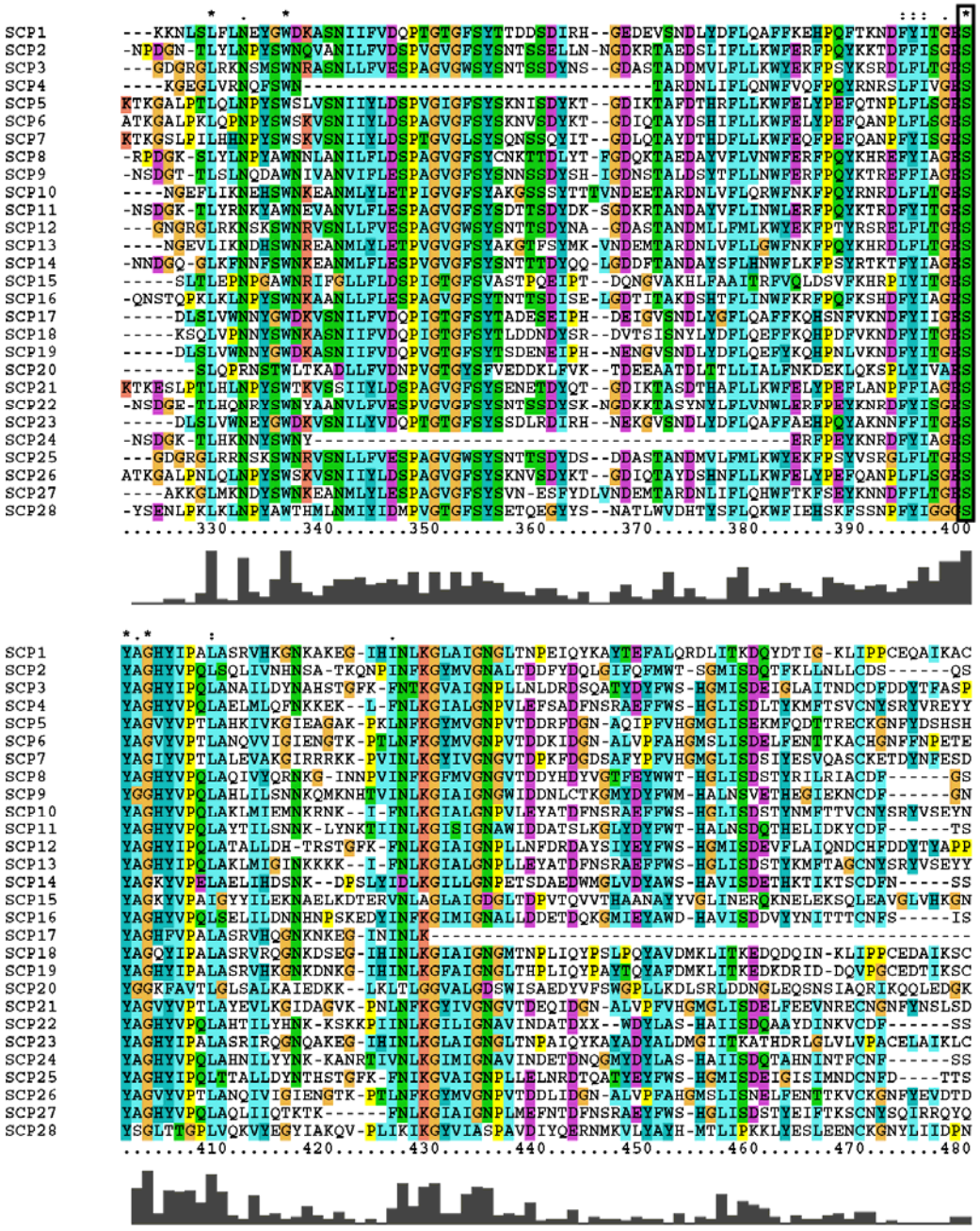
Figure 5.2: Multiple sequence alignment of cysteine proteases of chickpea. The catalytic residues are enclosed in the boxes.

5.1.3 Serine protease (SP)

Thirteen serine protease families as discussed by Tripathi *et al*, 2008 were identified in chickpea genome using HMM profile of each family (Table 5.3). Interestingly, the number of serine proteases encoded by chickpea is comparable to the number identified in *Arabidopsis* and rice genomes (206 and 222). Out of the total 220 *Ca*SP (*C. arietinum* SPs), 125 sequences possess the catalytic dyad or triad required for the catalytic reaction. The catalytic dyad or triads involved in the hydrolysis reaction are listed in Table 5.4. The GenBank accession numbers are assigned to 21 serine protease sequences [GenBank: KJ579709-KJ579715, KJ598059-KJ598071, and KJ619656] (Figure 5.3, CD-Figure S-1, CD-Table S-1).

Table 5.3: Number of hits of each of the thirteen serine protease families identified in chickpea genome using HMM profile.

Family	HMM hits	Hits with key residues present/ significant e-value/ complete sequence
Trypsin (S1)	7	5
Clp Endopeptidase (S14)	12	3
C-terminal processing peptidase (S41)	3	3
Lon protease (S16)	5	4
Lys-Pro-x Carboxypeptidase (S28)	8	4
Nucleoporin autopeptidase (S59)	2	1
Prolyl oligopeptidase (S9) (POP)	29	3
Protease IV (S49)	5	2
Rhomboid (S54)	15	6
Serine carboxypeptidase (S10)	40	28
Signal peptidase I (S24)	10	8
Subtilase (S8)	84	58
D-Ala-D-Ala carboxypeptidase B (S12)	0	0
Total *Ca*SPs	**220**	**125**



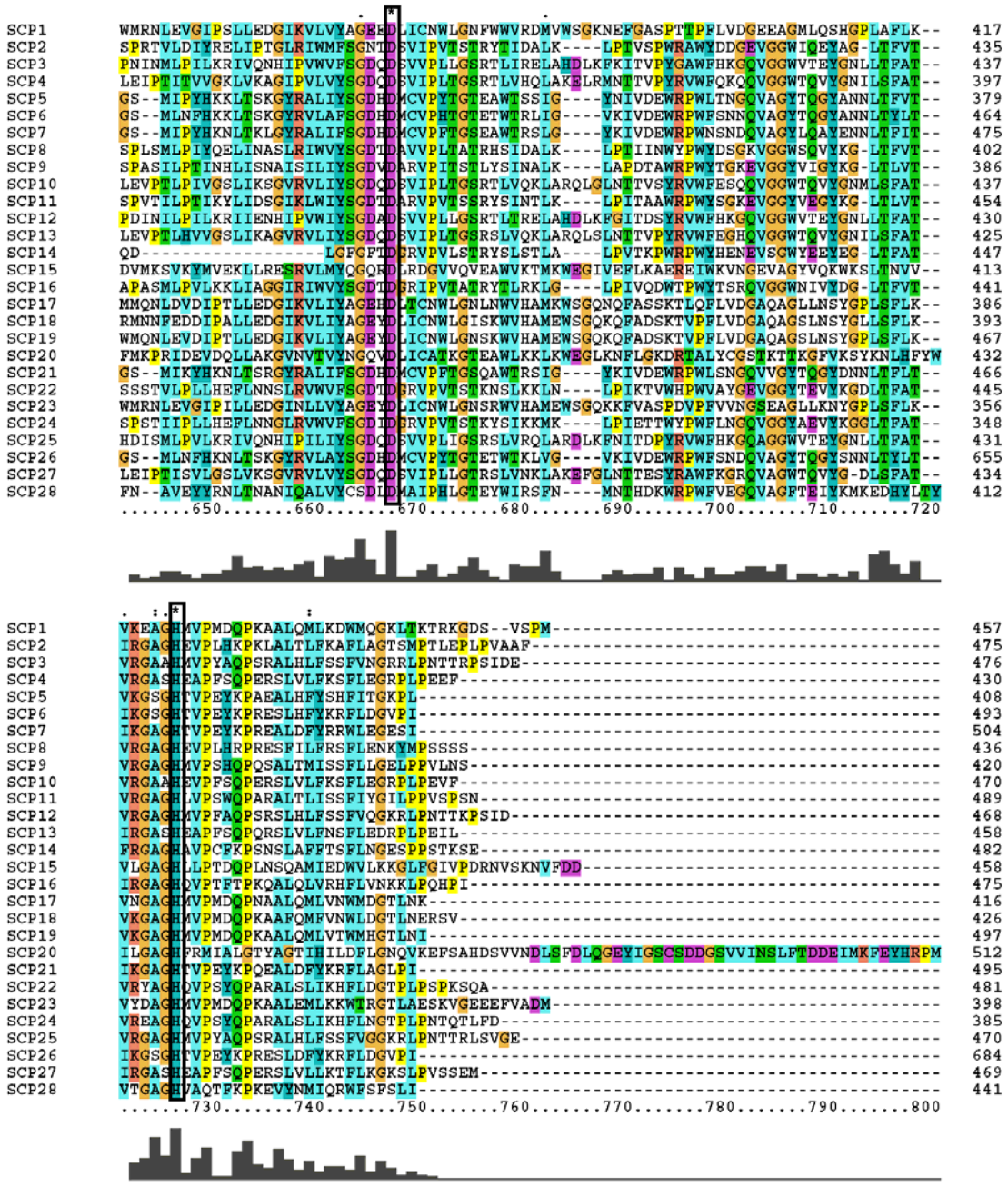


Figure 5.3: Multiple sequence alignment of serine proteases of chickpea. The catalytic residues are enclosed in the boxes.

5.1.4 Metalloprotease (MP)

The number of metalloproteases found in chickpea is not large. Only three members of metallopeptidase families i.e. M41 (PF01434), M48 (PF01435), and M50 (PF02163) were identified. The following GenBank accession numbers were assigned to 7 *Ca*MPs (*C. arietinum* MPs) [GenBank: KJ619646-KJ619652] (Figure 5.4, CD-Table S-1).

The protease and PIs with missing catalytic dyad or triad residues were not used in further analysis. The possibility that such proteins may be functional by following some other mechanism cannot be ruled out (CD-Table S-2).

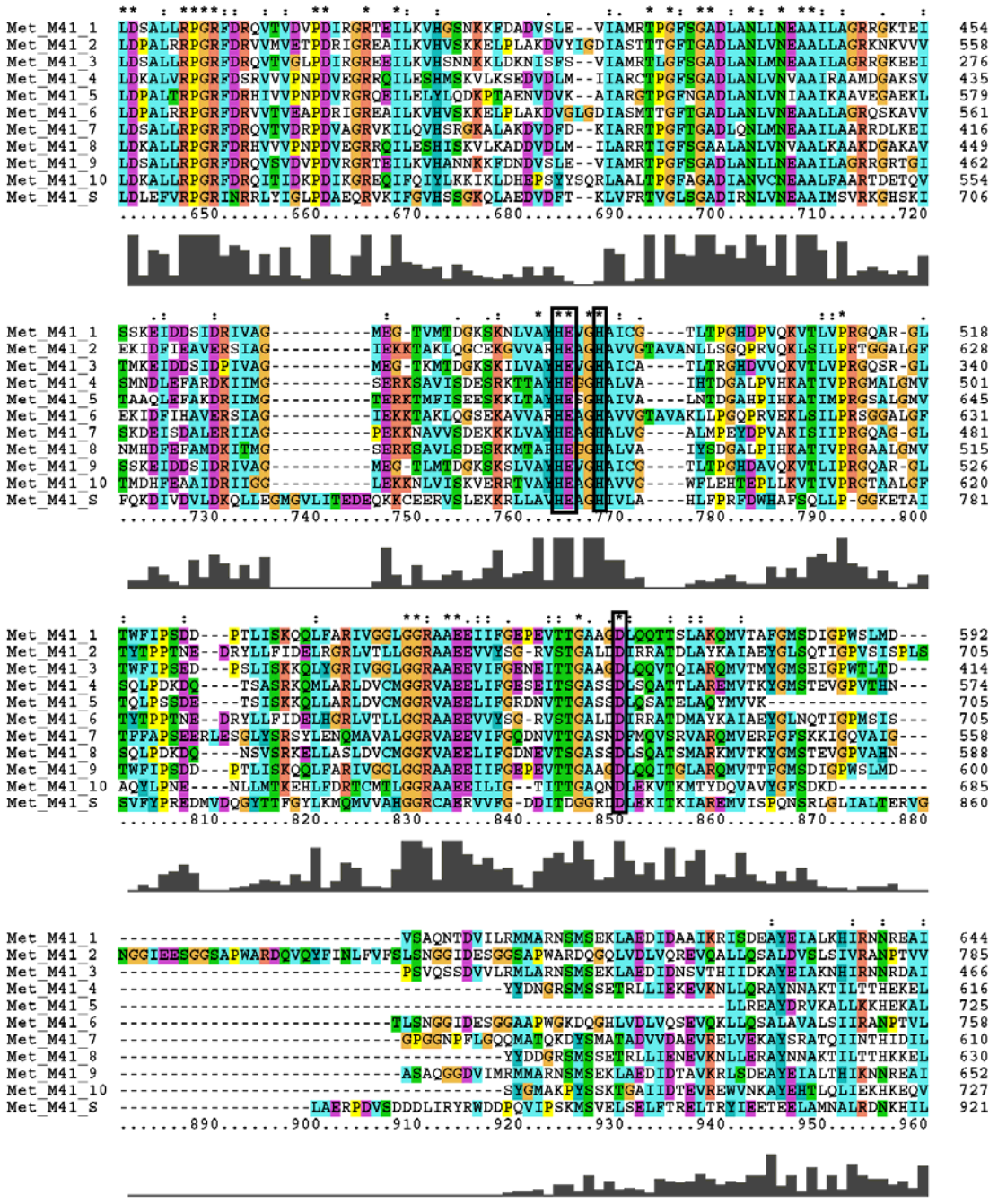


Figure 5.4: Multiple sequence alignment of metalloproteases of chickpea. The catalytic residues are enclosed in the boxes.

Table 5.4: Listed below the catalytic residues of each protease class involved in the hydrolysis reaction.

No.	Protease	Catalytic residues
1	**Aspartate protease**	Asp-Ser/Thr-Gly-------Asp-Ser/Thr-Gly
2	**Cysteine protease**	Cys, His, Asn
3	**Serine protease**	
	Carboxypeptidase	Ser, Asp, His
	CLP	Ser, His, Asp/Glu
	CTP	Ser, Asp, Lys
	Lon protease	Ser, Lys
	Lys-Pro-X	Ser, Asp, His
	POP	Ser, Asp, His
	Protease IV	Ser
	Rhomboid	Asn, Ser, His
	Signal peptidase S24	Ser, His/Lys
	Subtilase	Asp, His, Ser
	Trypsin	His, Asp, Ser
	Nucleoporin	His, Ser
4	**Metalloprotease**	
	M41	His-Glu, His, Asp
	M48	His-Glu, His
	M50	His-Glu, His

5.1.5 Cysteine protease inhibitors (CPIs)

The HMM profile search of the chickpea proteome for the presence of cystatin (PF00031) yielded 17 hits, out of which 3 hits had high e-values. The remaining 14 hits were used in further analysis. Two sequences were assigned GenBank accession number [GenBank: KJ619654-KJ619655] (Figure 5.5, CD-Table S-1).

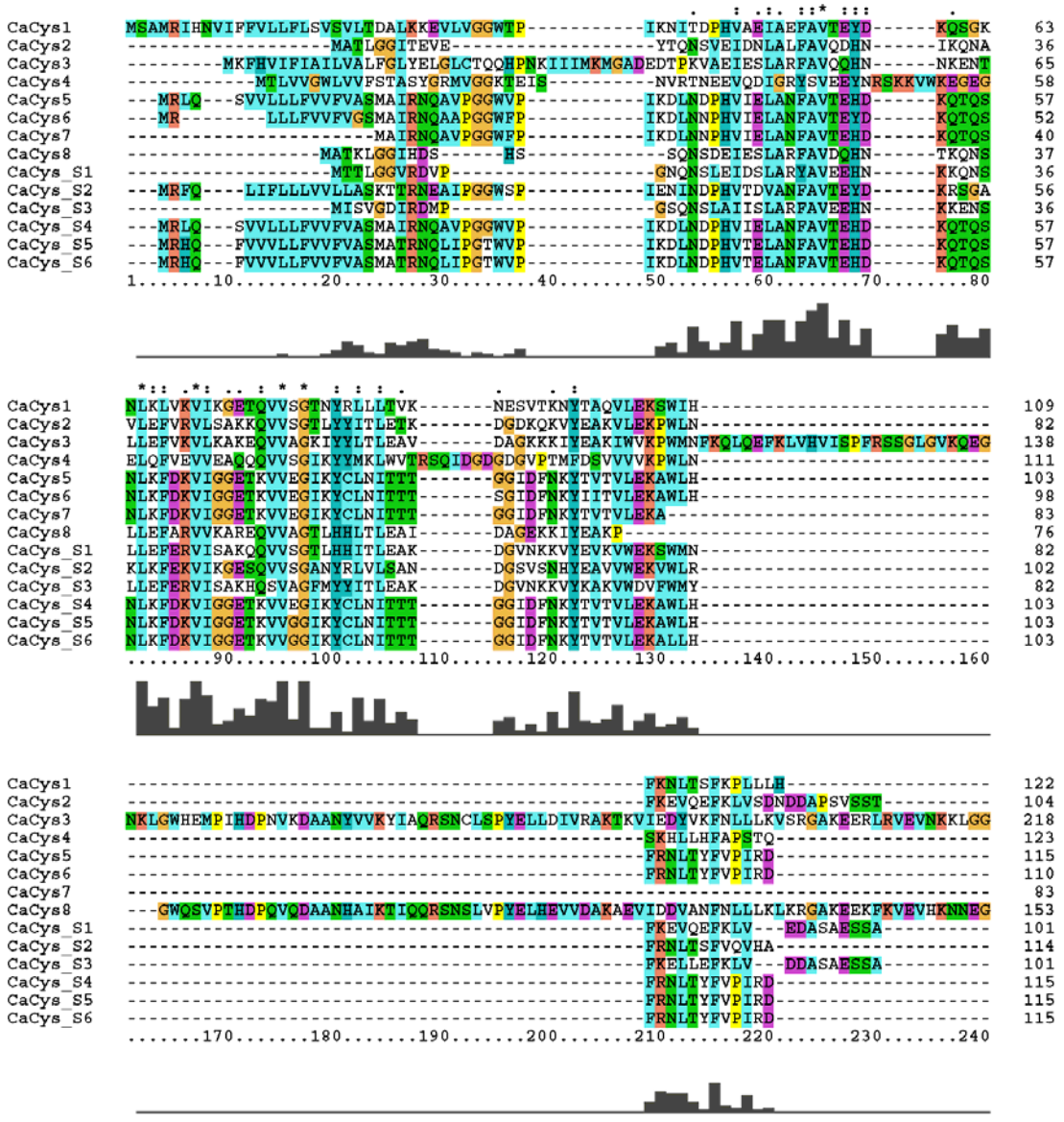


Figure 5.5: Multiple sequence alignment of cysteine protease inhibitors of chickpea.

5.1.6 Serine protease inhibitors (SPIs)

Five different classes of serine protease inhibitors were identified in chickpea namely Bowman-Birk inhibitor (PF00228) (BBI), kazal (PF07648), Kunitz-type (PF00197), squash (PF00299), potato inhibitor-1 (PF00280), potato inhibitor-2 (PF02428), and serpin (**ser**ine **p**rotease **in**hibitor) (PF00079). Ten Kunitz-type inhibitors, three serpins and one each of BBI, Kazal and potato inhibitor-I were found in chickpea. Squash inhibitor and potato inhibitor-II were not observed in chickpea proteome. GenBank accession number was assigned to one serpin sequence [GenBank: KJ619653] (Figure 5.6-5.8, CD-Table S-1).

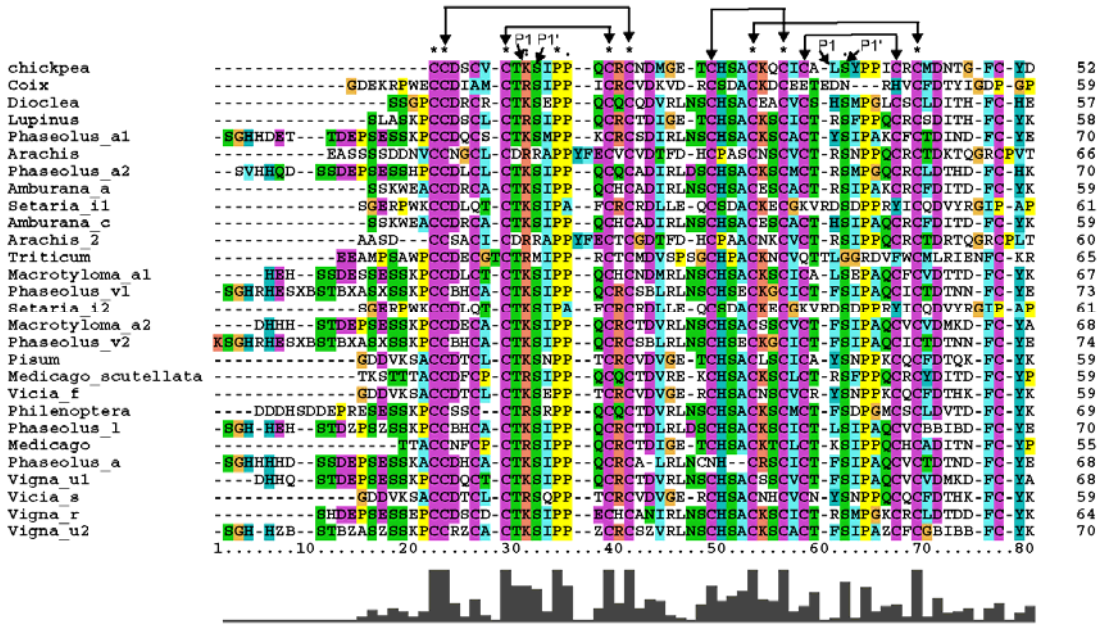


Figure 5.6: Multiple sequence alignment of Bowman-Birk inhibitor of chickpea and other plant species. The cysteines involved in the disulphide bond formation are marked with an arrow. P1 and P1' residues are also shown.

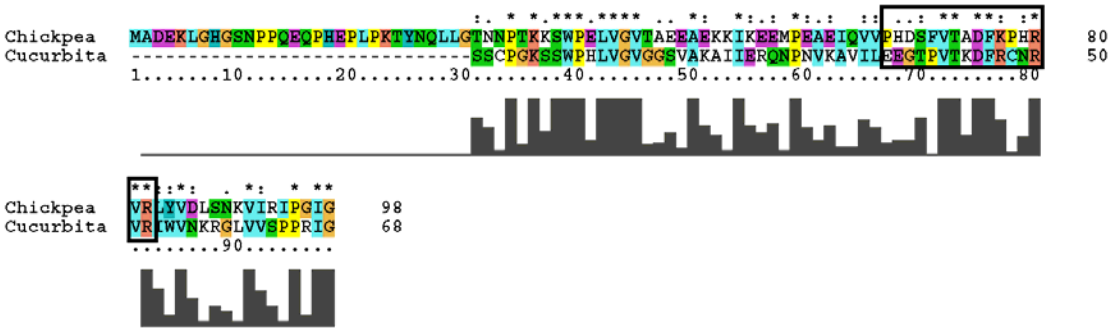


Figure 5.7: Multiple sequence alignment of potato inhibitor of chickpea and cucurbita. The residues of reactive site loop are enclosed in the boxes.

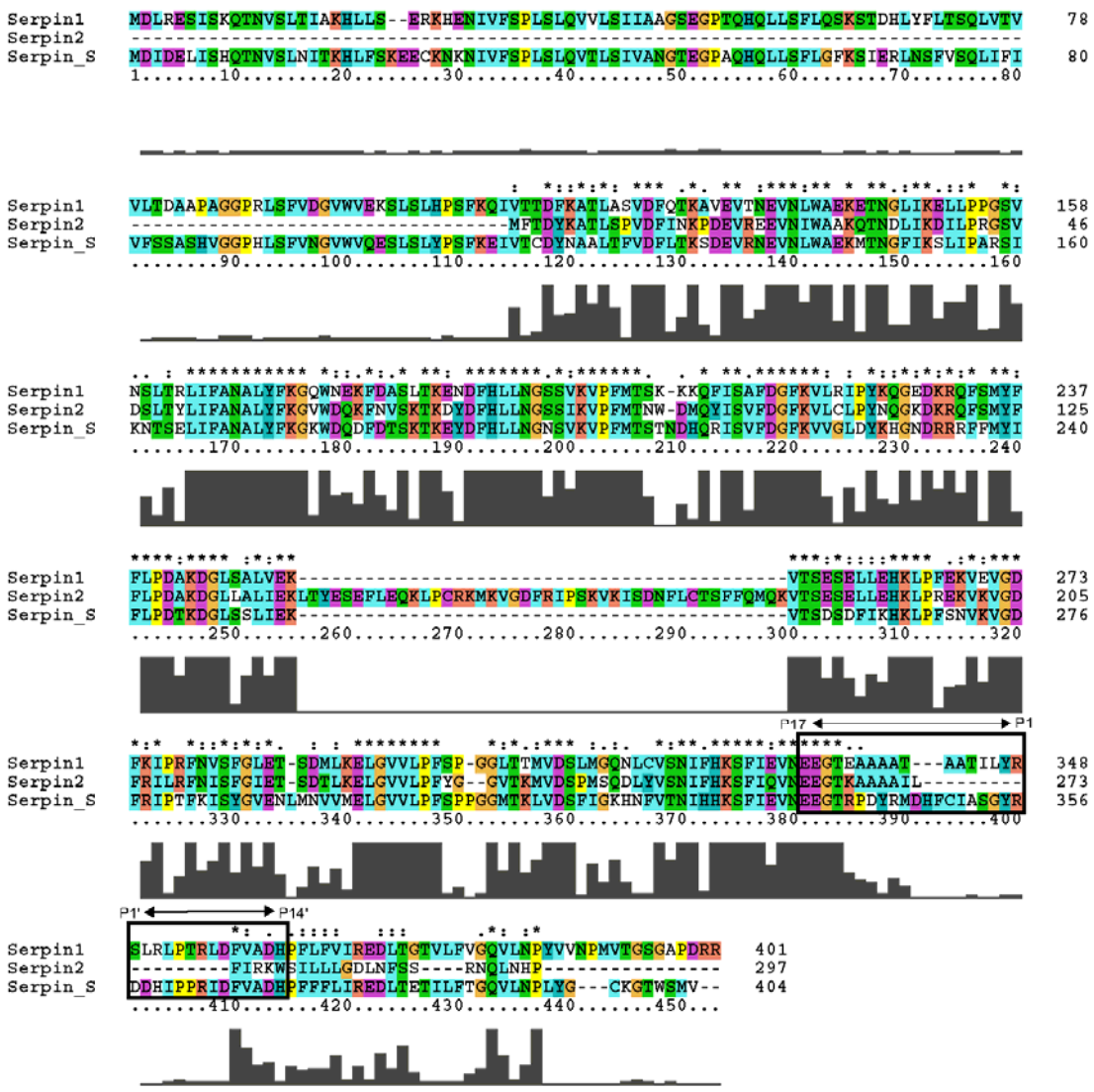


Figure 5.8: Multiple sequence alignment of serpin of chickpea. The residues of reaction centre loop are enclosed in the boxes.

5.2 Phylogenetic analysis and accounting gene duplication

Phylogenetic analysis of 66 *Ca*APs revealed their clustering into three groups i.e. typical, atypical, and nucellin-like similar to those described in *Arabidopsis* and rice. The clustering of typical APs could be supported by the presence of plant-specific insert (PSI) bearing homology with the precursor of mammalian saposins. Six sequences belong to the nucellin-like category of APs. However, maximum number of *Ca*APs (54) belongs to the atypical category. The three well characterized APs of *Arabidopsis* i.e. promotion of cell survival 1 (NP_195839, PCS1), constitutive disease

resistance 1 (NP_198319, CDR1), and aspartic protease in guard cell 1 (NP_188478, ASPG1) got clustered with the atypical chickpea APs (Figure 5.9).

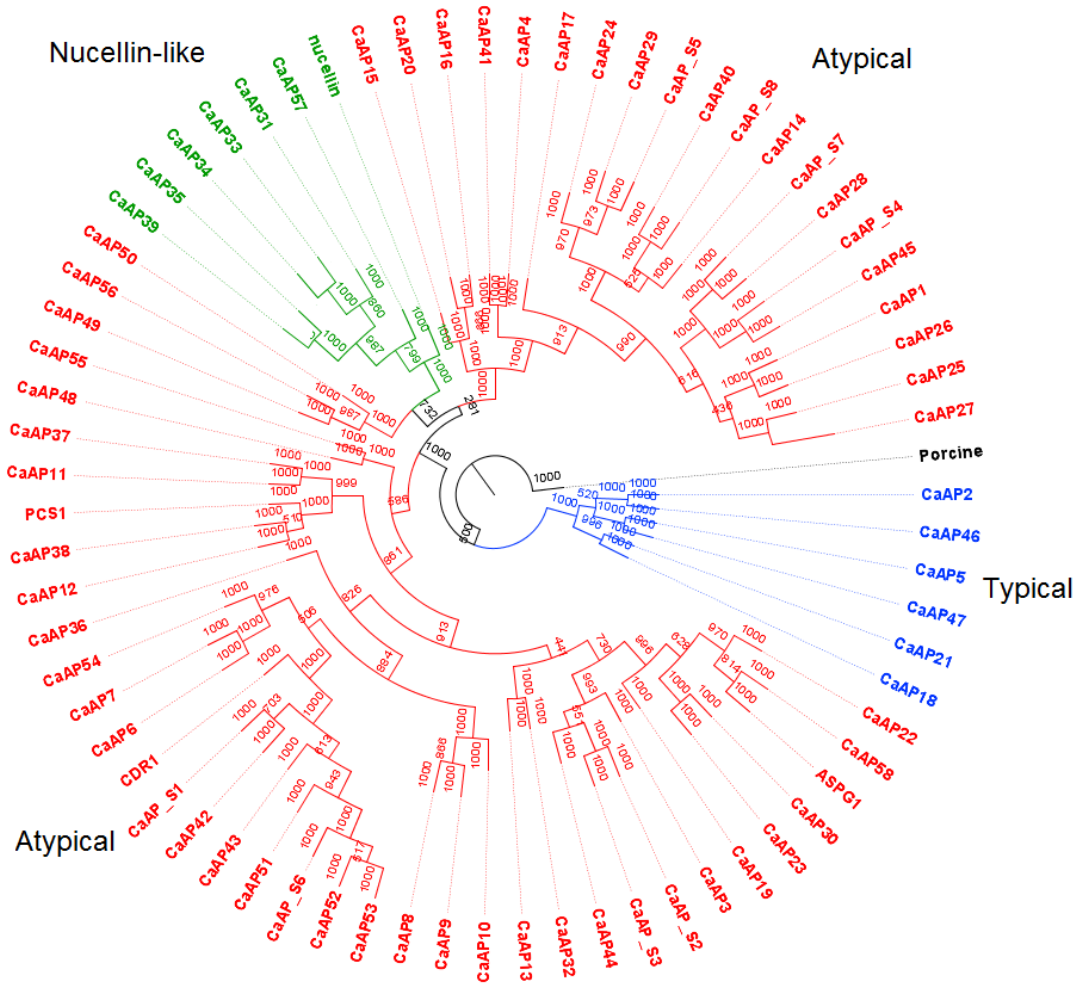


Figure 5.9: Phylogenetic analysis of chickpea APs. The dendrogram of chickpea APs with complete ASP domain clustered into four distinct clades atypical (2 clusters), typical, and nucellin like APs. Sus PGA was used as outgroup.

Phylogenetic tree of all the 125 protein sequences showed well defined clusters for each peptidase family of SPs. However, few members of signal peptidase class got diverged into two groups. The reason behind their separation might be the presence of two different peptidase domains i.e. S24 and S26 but this doesn't hold true for SP3. One member of serine carboxypeptidase (SCP20), owing to the presence of three additional domains CLH, WD40, and RING domain, behaved differently with respect

to its other members. The clustering of the different classes of SPs is supported by the presence of different domain structure in them (Figure 5.10, CD-Figure S-2).

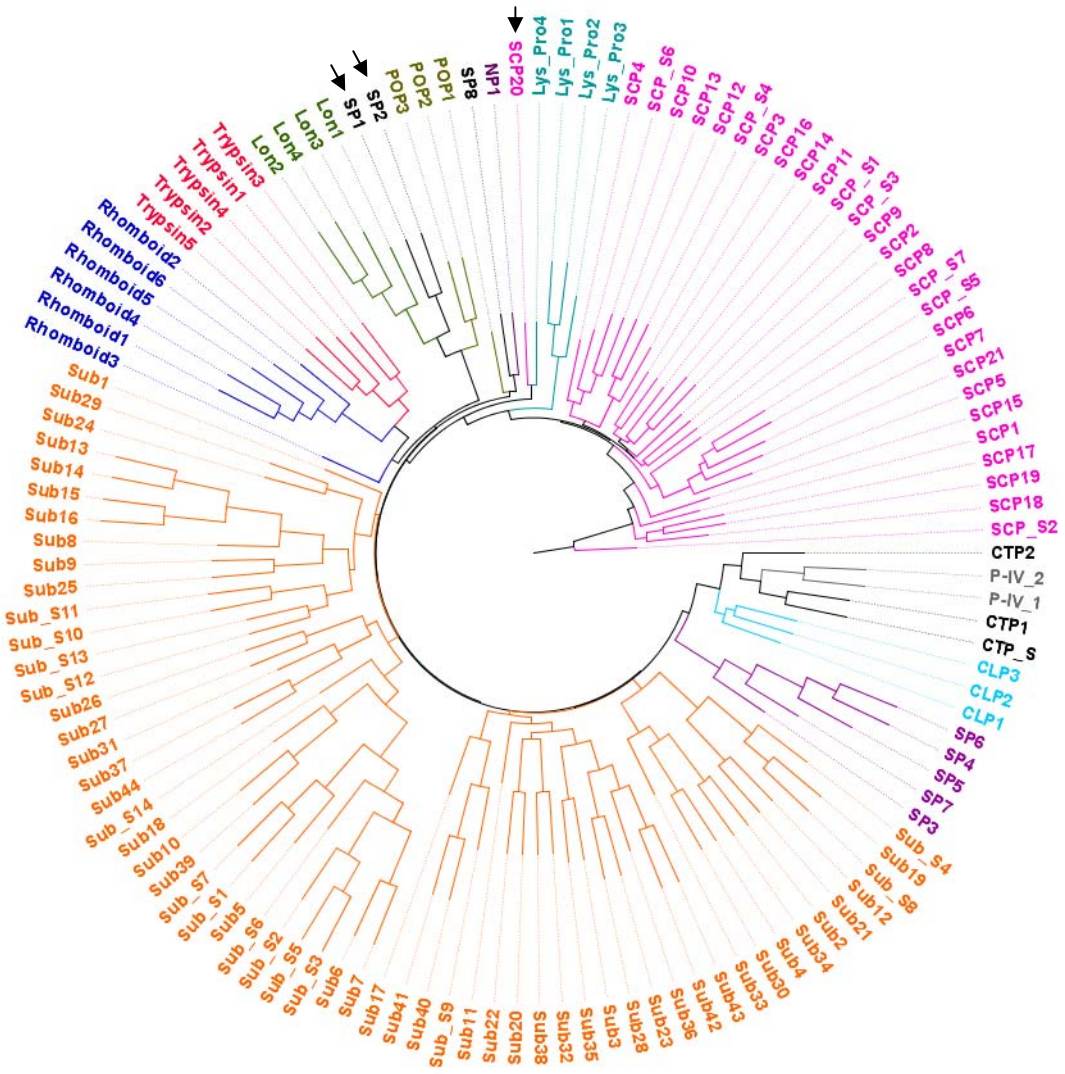


Figure 5.10: Phylogenetic analysis of chickpea SPs. The dendrogram of chickpea SPs showed distinct clusters for each serine peptidase class. Few members of signal peptidase and serine carboxypeptidase marked with the arrows were different from other members of their class.

To some extent, the members of cysteine proteases also got clustered on the basis of protein domains present in them. RD21 (responsive to desiccation 21), a papain-like cysteine protease (PLCP), is up-regulated in drought-stressed *Arabidopsis*. Their maturation process involves successive removal of signal peptide, prodomain cleavage, and removal of granulin (GRAN) domain which subsequently leads to a mature RD21. Four chickpea *Ca*CPs (*Ca*CP3, *Ca*CP10, *Ca*CP11, and *Ca*CP14)

possess the GRAN domain which clustered together in the phylogenetic tree. Three proteins (*Ca*CP15, *Ca*CP_S1, and *Ca*CP_S2) bear a carboxy-terminal amino acid sequence (KDEL) responsible for their retention in endoplasmic reticulum. The vacuolar targeting signal (NPIR) was observed in the prodomain region of only one chickpea member (*Ca*CP24) (Figure 5.11).

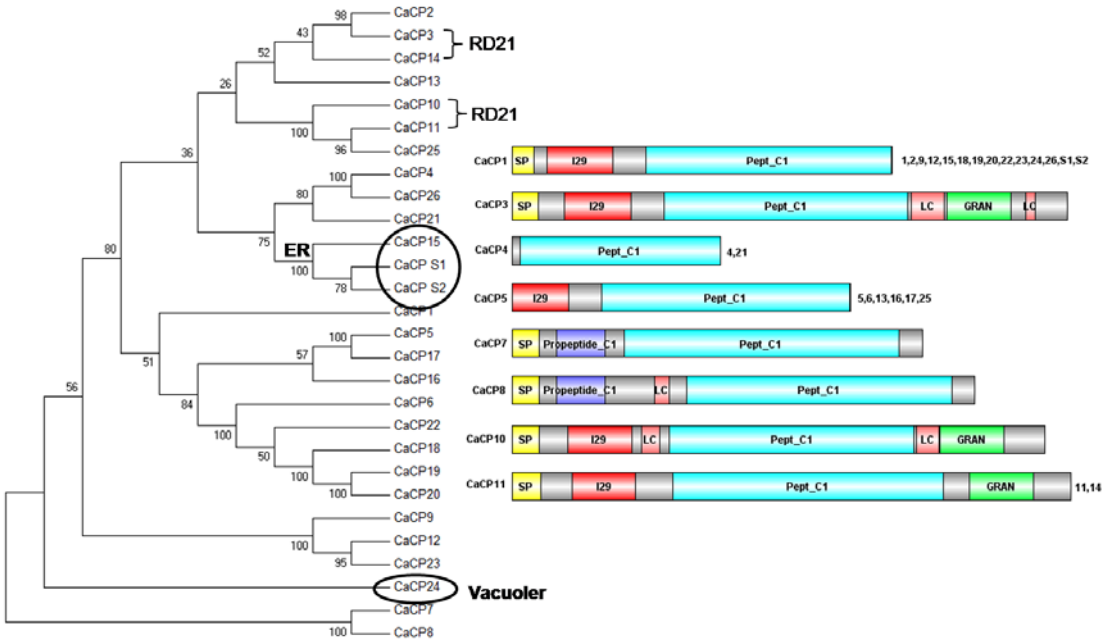


Figure 5.11: Phylogenetic and domain analysis of chickpea CPs. The dendrogram of chickpea CPs showed clustering of three endoplasmic reticulum targeting proteins. Four RD21 like proteins possess a GRAN domain. CaCP24 possess a vacuolar targeting signal.

Two possible recent gene duplication events were observed in the following gene pairs *Ca*AP8-*Ca*AP9 (90.9%) and *Ca*AP52-*Ca*AP53 (99.5%). However, only one gene pair each of cysteine proteases (*Ca*CP18 & *Ca*CP20- 92.4%) and serine proteases (subtilase: Sub40 & Sub41- 95%) was found to be involved in recent duplication event.

5.3 Sequence and structure analysis of protease genes

The domain analysis of identified APs showed the presence of a basic architecture of signal peptide, pro-peptide, and AP domain. A total of 66% of the total APs possess this structure. The transmembrane region was observed at the C-terminal region of 6

proteins while low complexity regions were seen in 19 proteins. The six typical APs had a SapB, SapB_1, and SapB_2 domain located in the PSI segment (Figure 5.12). The exon-intron architecture of AP genes was explored to get more insight about it. Our results showed correlation between the phylogeny and the exon-intron arrangement in AP genes. It means the genes possessing similar gene architecture got clustered together in the phylogenetic tree. The gene corresponding to nucellin-like proteins had number of introns ranged from 7 to 9. However, the typical AP genes had 10 and 12 introns. A varied arrangement of exons and introns was seen in atypical APs i.e. 8-11 introns with an exception of *Ca*AP28 which had only 4 introns. However, the other group of atypical APs consists of 0-1 intron with an exception of *Ca*AP48 and CaAP55 having 11 introns (Figure 5.13).

A peptidase domain present in CPs, SPs, and MPs modulates the activity of the enzyme. Each peptidase family possesses certain unique domains which are responsible for their function. Nevertheless, we couldn't identify any correlation between phylogeny, gene structure, and the domains present in them (Figure 5.11, Figure 5.12, CD-Figure S-3).

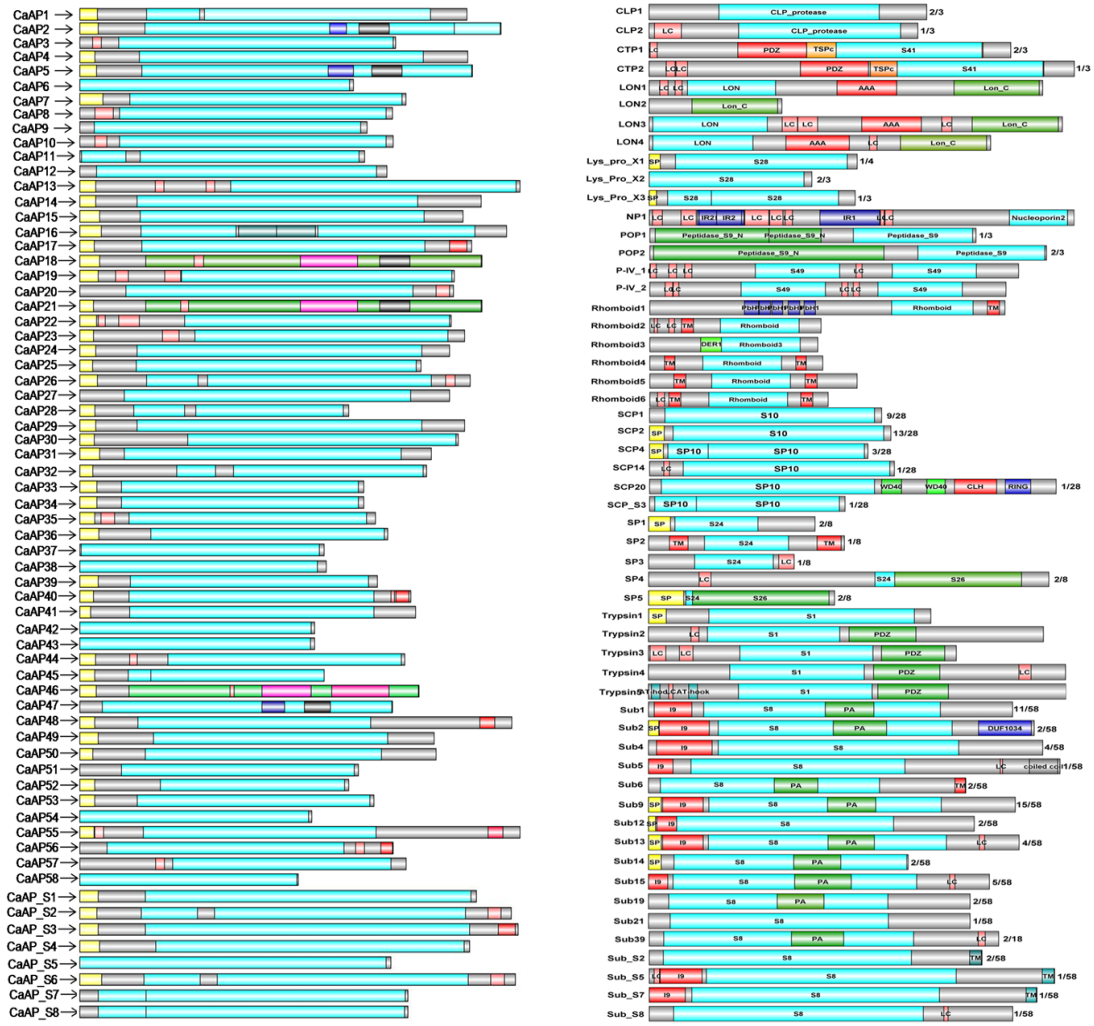


Figure 5.12: Domain arrangement of chickpea APs and SPs.

5.4 Orthologs identification and gene divergence

Orthologs were identified in some closely related and sequenced genomes like that of *M. truncatula*, *G.max*, *V. vinifera*, *A. thaliana*, *P. vulgaris*, and *L. japonicus*. The maximum number of orthologs was identified in *G. max* and *M. truncatula*. A few proteases and protease inhibitors had diverged away from the set and were unique to chickpea (CD-Table S-3).

5.5 Analysis of codon composition

The codon composition of the catalytic residues of chickpea proteases was analyzed by performing codon based multiple sequence alignment in Molecular Evolutionary Genetics Analysis 5.2 (MEGA) package (Tamura *et al*, 2011).The CDS sequences were aligned by using UPGMB based clustering keeping the gap open and gap extension penalties -400 and 0. The standard genetic code was used for the analysis. The codon composition of the key catalytic residues of proteases was analyzed to study the evolutionary conservation of a particular codon of a functionally important amino acid. The first ASP domain of most of the members of APs had amino acid threonine at second position. The codon for the conserved aspartic acid in both the ASP domains was GAT. On the contrary, the second position of another ASP domain was occupied by serine residue in most of the sequences. Moreover, the preference for GAT codon was seen in some of the serine proteases too where aspartic acid forms the catalytic site. The conserved histidine residue in CPs and SPs also showed more preference for CAT codon as compared to CAC. The similar preference for CAT and GAT codons for histidine and aspartic acid residues was also observed in MPs. The codon for the key residues cysteine and serine of CPs and SPs, respectively, in most of their members were found to be TGT and TCA (Figure 5.14). An attempt was made to identify a linkage pattern among the codons of the key residues for dyad and triads but couldn't find any such linkage (CD-Table S-1).

Figure 5.13: Image shows gene architecture and clustering of *Ca*APs. The genomic locations of aspartate proteases with IDs CaAP_S6, CaAP_S1, CaAP_S2, CaAPS3, CaAP_S7 were unknown and therefore assigned them on scaffolds.

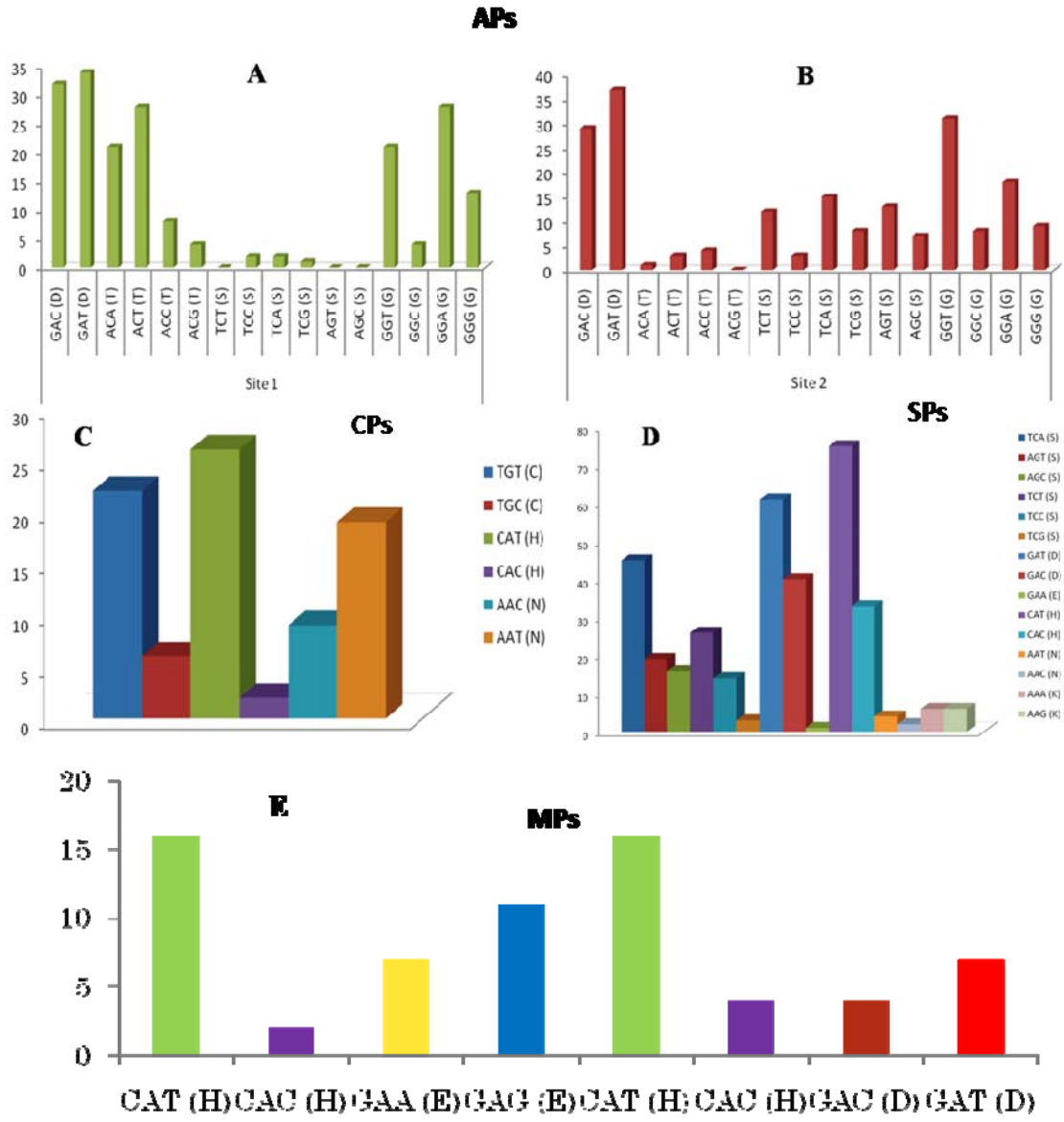


Figure 5.14: Codon composition of the identified chickpea protease sequences.

5.6 Data from gene expression studies

Gene expression for proteases and protease inhibitors were identified with the help of EST data of chickpea. 12, 9, and 32 *Ca*AP, *Ca*CP, and *Ca*SP genes showed expression in tissues such as leaves, roots, root and collar regions, and shoots. Members of various classes of serine proteases were found to express in leaves, roots, shoots, and roots and collar except trypsin, nucleoporin, and protease IV. Not a single member of M50 family of metalloprotease had support of EST data. Seven *Ca*CP and *Ca*SP inhibitor genes showed expression pattern in various plant tissues. Seven CPI

and SPI genes were observed to be expressed in the above plant tissues except Bowman-Birk and serpin (CD-Table S-4).

RNA-seq data showed that 42 out 66 *Ca*AP genes were highly expressed (FPKM ≥ 5) and 23 were lowly expressed (0 < FPKM < 5) in the five tissues taken for the study. *Ca*AP_S6 gene did not show expression in any of the five tissues. Similar analysis on *Ca*CP showed that 19 genes out of 28 were highly expressed and the remaining nine genes were expressed at low level in the plant tissues. 86 *Ca*SPs genes were highly expressed while 36 genes were lowly expressed (3 genes were unexpressed) in the five tissues under study. Eighteen *Ca*MPs genes were highly expressed while only 2 genes were lowly expressed. The expression pattern was also observed for the inhibitor genes (cysteine protease inhibitor and serine protease inhibitor), most of the genes were known to be expressed in one or more plant tissue (Figure 5.15 & 5.16, CD-Table S-5). However, if a particular gene is unexpressed in the above tissues searched, that doesn't conclude that they are non-functional. They might be expressed in some other tissues or under a specific condition. Upon analyzing the drought stressed RNA-seq reads from the two different genotypes of the chickpea it was seen that some of the genes were over-expressed (20 APs, 11 CPs, 8 MPs, 39 SPs, 4 SPIs, 5 CPIs) as well as some are under-expressed as compared to the control samples (CD-Figure S-4, CD-Table S-6).

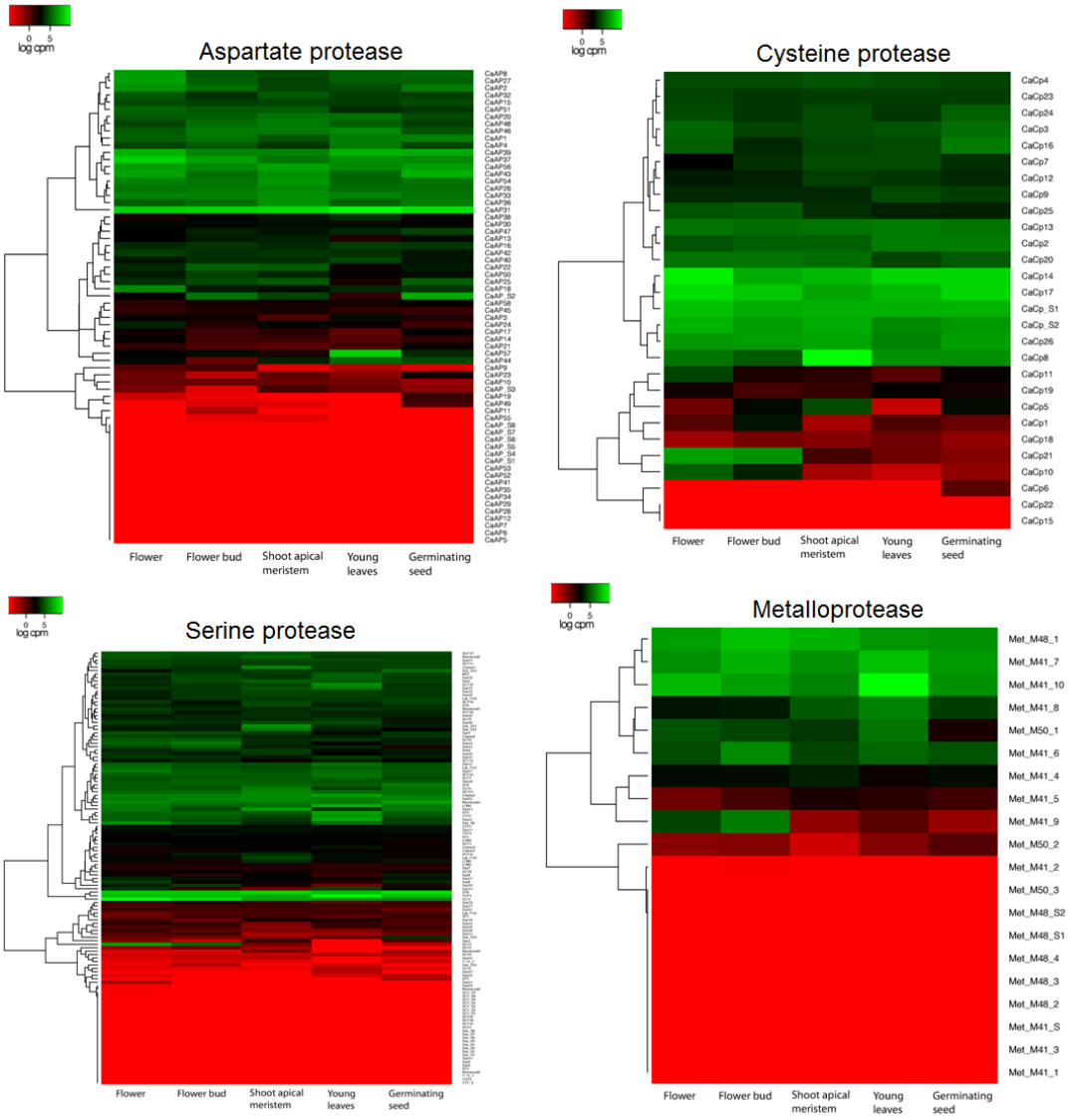


Figure 5.15: Expression level for chickpea protease genes in various tissues by RNA-seq data analysis. Heatmap showing relative gene expression in various tissue samples. The color scale represents log transformed count per million (CPM), for proteases genes in different tissues. The proteases genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

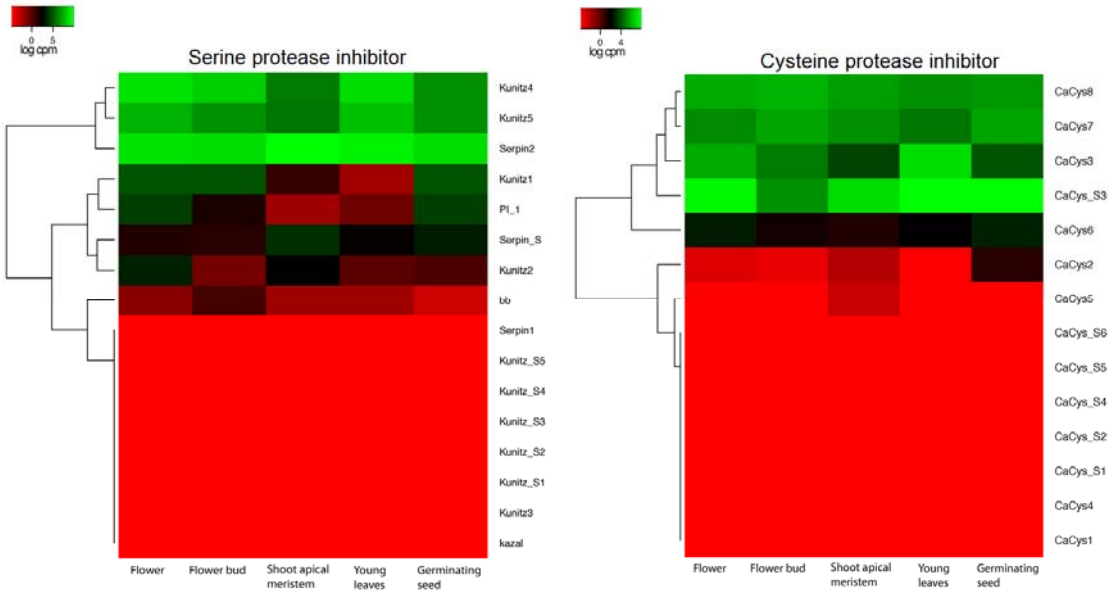


Figure 5.16: Expression level for chickpea protease inhibitor genes in various tissues by RNA-seq data analysis. Heatmap shows relative genes expression in various tissue samples. The color scale represents log transformed count per million (CPM), for protease genes in different tissues. The protease inhibitor genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

5.7 Molecular modeling, model validation and refinement

The homology models of serine protease (POP), serine and cysteine protease inhibitors were built with the help of structures of homologous proteins. The sequences for which the structural studies were possible and possess maximum identity with the template were chosen for the structural studies. The chickpea POP structure was built using homologous protein from *Sus scrofa* (PDB: 1E8M, Identity- 55%, Resolution- 1.50 Å). The structure of potato inhibitor-I was modeled using the 3-D coordinates of serine protease inhibitor from barley seeds (PDB: 1C12, Identity: 30%, Resolution: 2 Å). Crystal structure of BBI in complex with bovine trypsin was utilized to model the chickpea BBI. The structure has a resolution of 2 Å and showed an identity of 69% with the chickpea BBI sequence (PDB: 2ILN). The ten Kunitz-type inhibitors were modelled using following crystal structures- 2QN4, 3ZC8, and 1R8N (Identity >30%, Resolution- 1.80 Å, 2.24 Å, 1.75 Å). The structural alignment of 10 Kunitz inhibitor models showed highly diverse amino acid composition in the reaction site loop as stated by Patil *et al.* (2012) (Figure 5.17). The cysteine protease inhibitor from chickpea was modeled using crystal structure of tarocystatin and

papain complex (PDB-3IMA, Identity 58%, and Resolution 2.03 Å). The 3-D structure of serine protease inhibitor (SERPIN) was modeled using coordinates of serpin from *Arabidopsis thaliana* (PDB-3LE2, Identity 57%, and Resolution 2.20 Å) (Figure 5.18). The signal peptide was observed at the N-terminal of BBI and ten Kunitz-type inhibitors, which was missing in POP, serpin, cystatin, and PI-I sequences. The multiple sequence alignment of target proteins with their respective templates are shown in file CD-Figure S-5.

Figure 5.17: Homology models of 10 Kunitz inhibitors of chickpea are shown in ribbon representation. The lower panel shows the structural alignment of ten chickpea Kunitz inhibitors and Kunitz type dual inhibitor (TKI) of factor Xa (FXa) and trypsin of tamarind. The reactive loops are shown in the enclosed box.

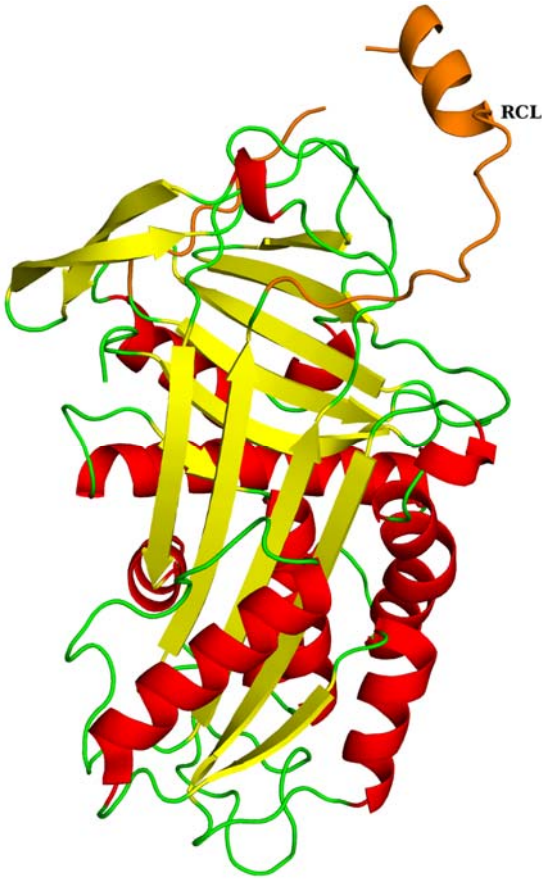


Figure 5.18: Homology model structure of chickpea serpin. The three-dimensional structure of chickpea serpin showing reaction centre loop (RCL).

The generated models were assessed for their stereochemical parameters. The Ramachandran plot revealed that more than 98% of the residues occupied the allowed region. The goodness factor (g-factor) of the models had values between -0.5 to 0.2. The overall RMSD values of the C-alpha atoms of the models with respect to the templates were close to 1 Å. The overall quality score (z-score) of the models; calculated by ProSA lies within the z-scores of experimentally determined protein structures. The overall quality score shown in ERRAT graph also proves the reliability of the generated models (Table 5.5).

The model refinement phase involved preprocessing the initial models by adding hydrogens, assigning bond order, and filling missing loops and side chains. Subsequently, the models were subjected to restrained minimization by applying the constraint to converge the non-hydrogen atoms to an RMSD of 0.3 Å using OPLS 2005 force field. After that, the models were further subjected to 500 steps of steepest

descent energy minimization followed by 1000 steps of conjugate gradient energy minimization using the same force field. These energy minimized models were further used for docking and molecular dynamics studies.

Table 5.5: Model validation statistics

Protein	Ramachandran plot	ERRAT score	RMSD (Å)	ProSA z-score
POP	98.4*, 1.6**, 0.1***	70	0	-11.08
Potato inhibitor-I	100, 0, -0.1	100	0	-5.93
Bowman-Birk inhibitor	100, 0, -0.1	90.9	1.5	-3.09
Kunitz 1	99.2, 0.8, -0.5	92.68	0.12	-3.09
Kunitz 2	100, 0, -0.1	94.93	0	-1.31
Kunitz 3	99.2, 0.8, -0.5	89.15	0.15	-3.81
Kunitz 4	98.6, 1.4, 0.2	88.88	0	-5.29
Kunitz 5	96.5, 3.5, 0.2	85.84	0	-4.97
Kunitz_S1	97.4, 2.6, 0.2	97.22	0	-5.06
Kunitz_S2	100, 0, -0.1	91.81	0	-3.64
Kunitz_S3	99.2, 0.8, -0.5	92.22	0.12	-5.41
Kunitz_S4	100, 0, -0.1	98.66	0	-1.4
Kunitz_S5	100, 0, -0.1	95.31	0	-2.12
Cysteine protease inhibitor	100, 0, 0.1	72.8	0	-3.89

*** Represent percent of the total residue present in allowed region of Ramachandran plot. ** Represent percent of the total residue present in disallowed region of Ramachandran plot. *** Represent the value of goodness factor for the models. The values in the third, fourth, and fifth column represent the ERRAT score, RMSD values, and Z-scores of the models.**

5.8 Molecular docking simulations

The docking studies of POP with ZPR, by utilizing the reference ligand of the model, generate the best pose with a glide score and E-model value of -7.76 and -65.15. The O9 and O16 atoms of ZPR interact with the NH1 atom of Arg656 and NE1 atom of Trp604 with an interacting distance of 3.16 Å and 2.92 Å. R239 provides stacking

interaction with the six carbon ring of ZPR. The catalytic Ser563 and His696 were involved in mutual hydrogen bond of length 3.3 Å (Figure 5.19).

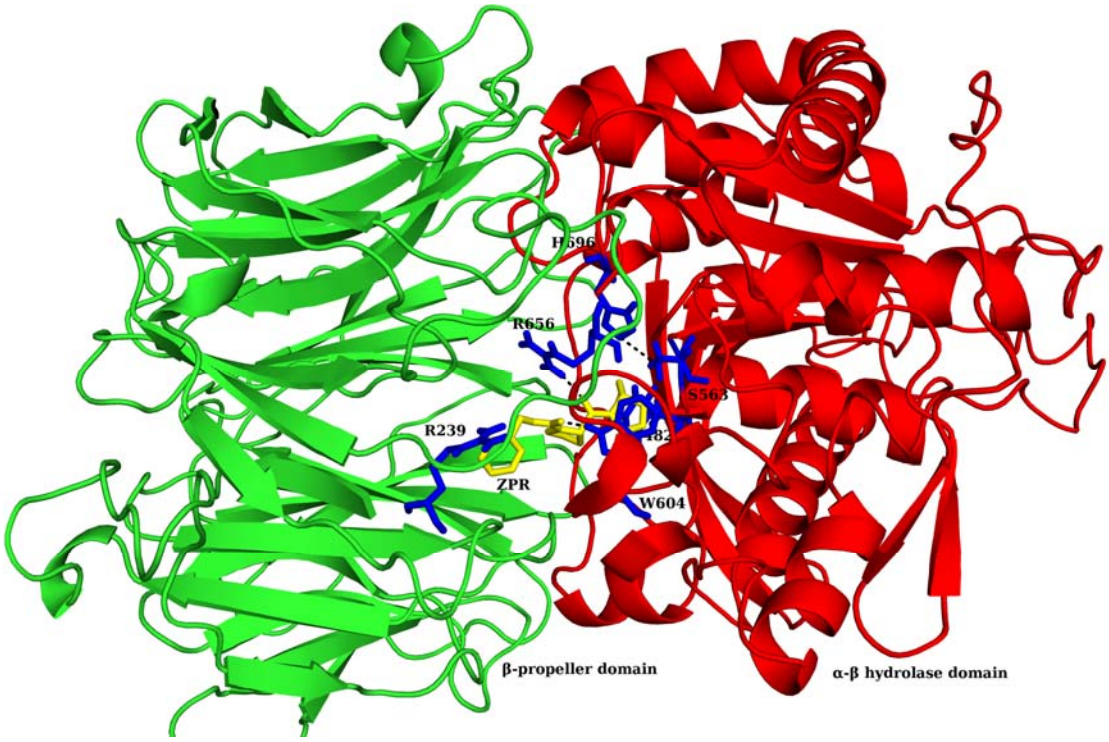


Figure 5.19: The docked complex of POP and ZPR. The docked complex of POP and drug ZPR (yellow) showing intermolecular hydrogen bond between them. The β-propeller domain is shown in green color and α-β hydrolase domain is shown in red color.

The cysteine protease inhibitor of chickpea was docked in the binding site pocket of papain from *Carica papaya* by defining their probable interface residues. The Z-score of the best pose was 1224.27 (Figure 5.20A). Similarly, BBI and PI-I were docked in the binding pockets of their respective targets i.e. bovine trypsin. The Z-scores for both BBI-trypsin and PI-trypsin docked complexes were 2227.32 and 1883.65. Only two intermolecular hydrogen bonds, between K61-D60 and S213-D54, were observed in the PI-trypsin docked complex. The residues of the inhibitor loop were involved in few intramolecular hydrogen bonding, (D60-R65, K62-R65, and S55-P52) which is essential for maintaining proper fit (Figure 5.20B). The structure of BBI was stabilized by five disulphide bonds which maintain the structure of the double-headed binding loops. Four hydrogen bonds were observed at the interface of

the two trypsin molecules and BBI i.e. F41-Y92, G216-C88, N97-N73, and H57-S65 (Figure 5.20C).

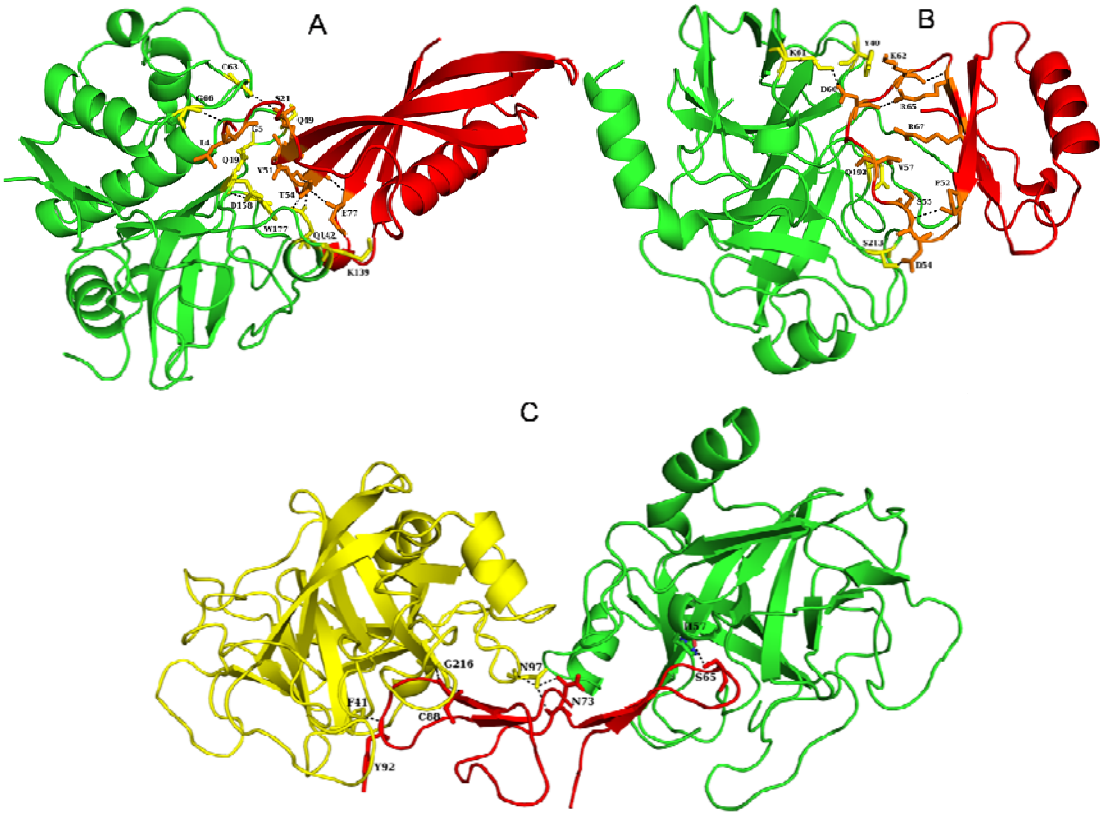


Figure 5.20: A. The docked complex of cysteine protease inhibitor (red) and bovine trypsin (green) showing inter and intramolecular hydrogen bonds at the interface. B. The docked complex of PI-I (red) and bovine trypsin (green) showing inter and intramolecular hydrogen bond at the interface. C. The docked complex of BBI (red) and papain (green & yellow) showing inter and intramolecular hydrogen bond at the interface.

In order to examine the role of protease inhibitors in chickpea, the inhibitor binding sites of bovine trypsin and papain were compared with the trypsin and cysteine proteases of chickpea. Interestingly, it was observed that chickpea trypsin was very much different from bovine trypsin with respect to its amino acid composition and tertiary structure. It belongs to a specific class of serine protease i.e. DegP which is involved in the cleavage of photodamaged D2 protein of photosystem II. Therefore, the trypsin inhibitor in chickpea may probably play a major role in defense against

pathogens and insects. On the other hand, cysteine proteases in chickpea share similar set of binding site residues and three-dimensional structure like papain and thus chickpea protease inhibitor has a role in inhibiting its own cysteine proteases too.

5.9 Molecular dynamics simulations

All the five docked complexes were subjected to a simulation of 10 ns time duration to explore the changes in the structures. The RMSD graph was drawn by assigning time in ps on the X-axis and RMSD values of the Cα atoms in nm on the Y-axis. From the graph, we can conclude that the RMSD of Cα atoms of the generated structures over the complete trajectory is stable (Figure 5.21). The Cα RMSD between the initial and final conformation after the simulation of each docked complex was ~ 1 Å (Figure 5.22).

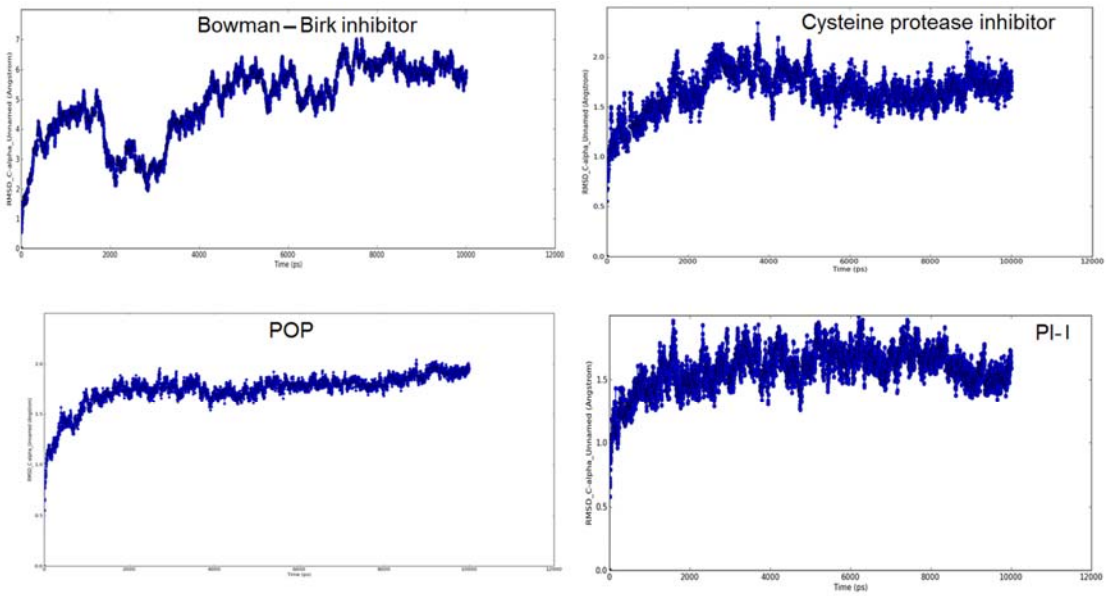


Figure 5.21: The root mean square deviation plot of C-alpha atoms of Bowman-Birk inhibitor, cysteine protease inhibitor, POP, and PI-I.

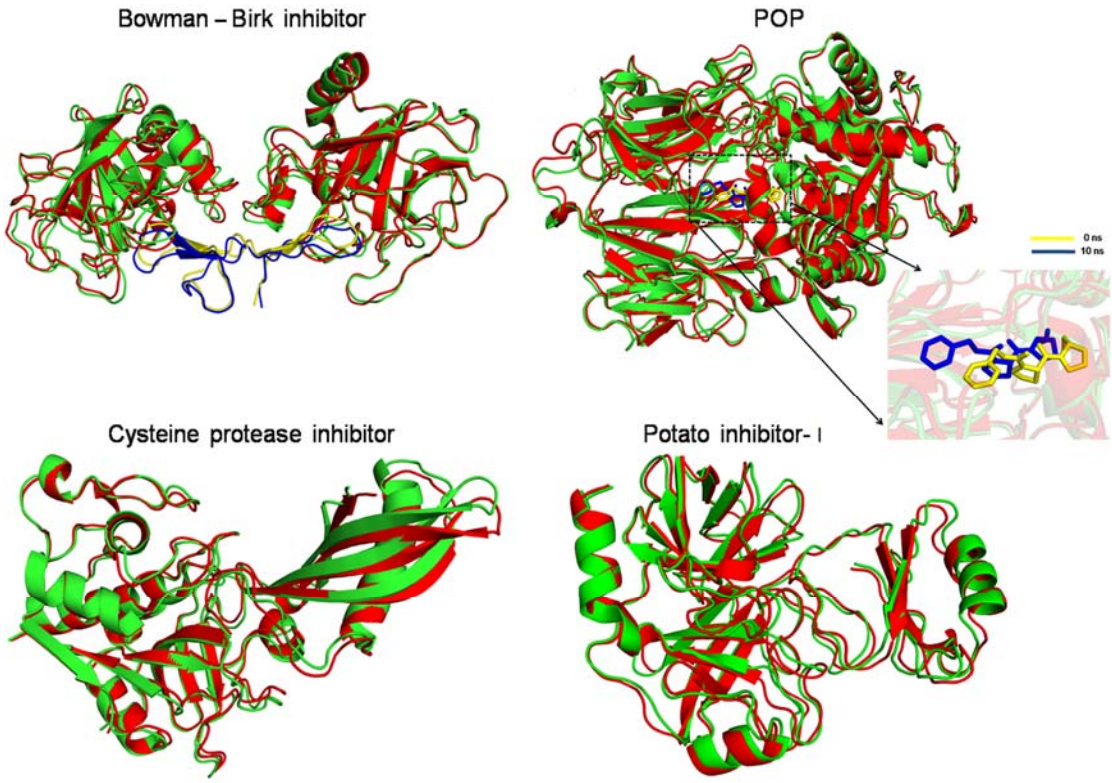


Figure 5.22: The image shows superposition of the initial structure (green) and the structure after 10 ns simulation (red).

After 10 ns, O16 atom of ZPR was involved in hydrogen bonds with Tyr482, Ser563 and His696 with a distance of 2.95, 2.80, and 2.70 Å. In the initial structure, prior to the simulation, a hydrogen bond was seen between the catalytic Ser563 and His696 residue which at the end of the simulation was lost and involved in the hydrogen bond interaction with the inhibitor. The inhibitor ZPR obstructed the binding site cavity by blocking the active site residues (Ser563 and His696). Even the hydrogen bond with Trp604, which was observed in the initial structure, was also lost (Figure 5.23).

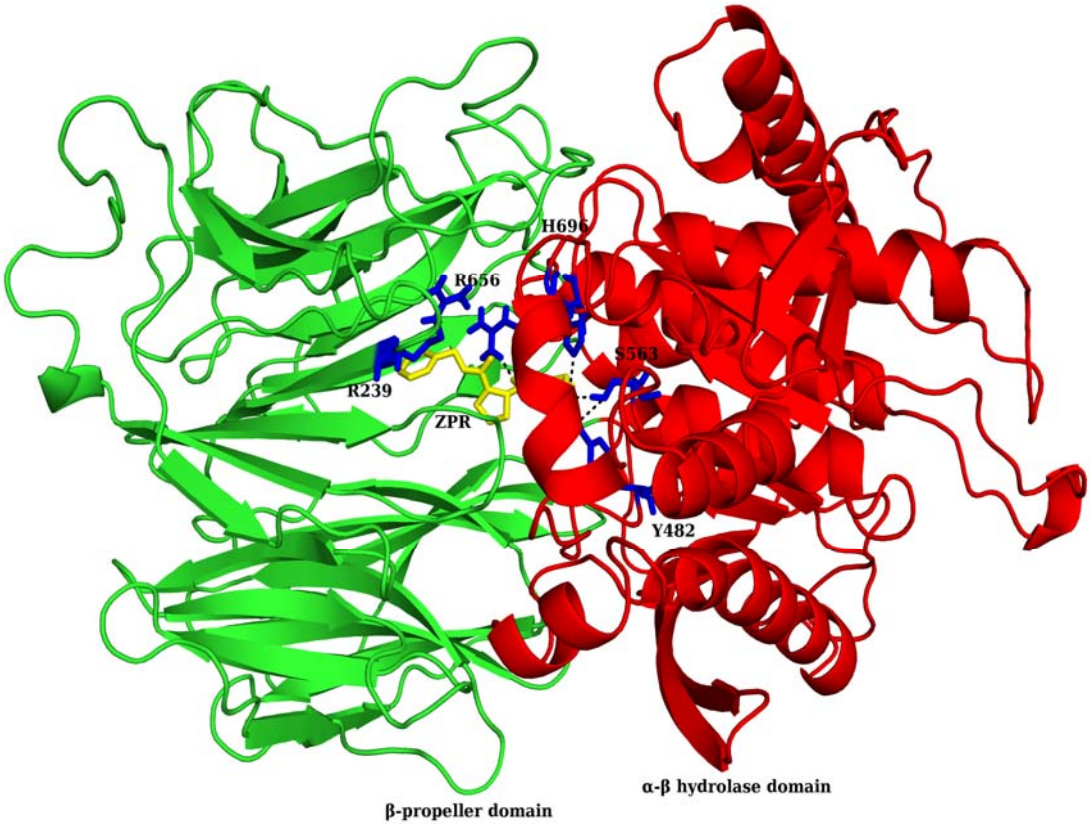


Figure 5.23: The docked POP-ZPR complex after 10 ns simulation. The docked complex of POP and drug ZPR (yellow) showing intermolecular hydrogen bond between them. The β-propeller domain is shown in green color and α-β hydrolase domain is shown in red color. After 10 ns, the drug blocked the active site residues His696 and Ser563.

Similarly after 10 ns, the final conformations of CPI, BBI, and PI-I complexes revealed formation of more stable complex as an increase in number of inter-molecular hydrogen bonds was observed (Table 5.6). The two highly conserved arginine residues in the PI-I family are important to support the reaction site loop with respect to the main body of the inhibitor. The two arginine residues in chickpea PI-trypsin complex were involved in an intramolecular hydrogen bonding with Asp60 (Figure 5.24).

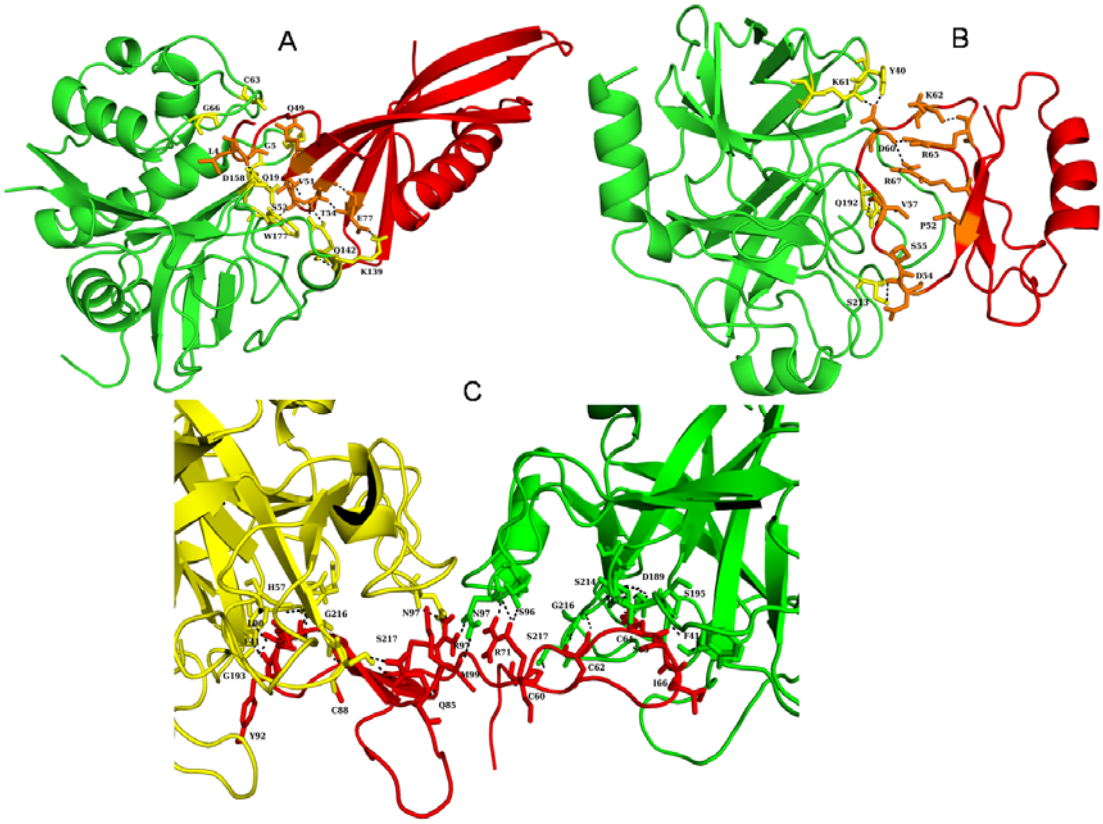


Figure 5.24: A. The docked cysteine protease inhibitor (red) and bovine trypsin (green) after 10 ns simulation. B. The docked complex of PI-I (red) and bovine trypsin (green) after 10 ns simulation. C. The docked complex of BBI (red) and papain (yellow & green) after 10 ns simulation.

Table 5.6: Analysis of intermolecular hydrogen bonding network in the docked complexes. Residues in bold indicate same hydrogen bond interaction found in the starting complex and complex after 10 ns simulation.

Complex	Initial structure	Hydrogen bond length (Å)	After 10 ns	Hydrogen bond length (Å)
POP	W604 (NE1)-ZPR (O16)	2.0	**R656 (NH1)-ZPR (O9)**	**2.7**
	R656 (NH1)-ZPR (O9)	**3.2**	S563 (OG)-ZPR (O2)	2.8
			Y482 (OH)-ZPR (O2)	3.0
			H696 (NE2)-ZPR (O2)	3.2
Potato inhibitor-I	**K61 (NZ)- D60 (OD2)**	**2.3**	Y40 (OH)- D60 (OD2)	2.5
	S213 (OG)- D54 (OD1)	3.2	**K61 (NZ)- D60 (OD2)**	**2.7**

			Q192 (NE2)- V57 (O)	3.2
			S213 (OG)- D54 ()	3.4
Bowman-Birk inhibitor	**F41 (O)-Y92 (N)**	**2.8**	**F41 (O)-Y92 (N)**	**2.1**
	G216 (N)-C88 (O)	**2.8**	G193 (N)- L90 (O)	2.8
	G216 (O)-C88 (N)	**3.0**	S195 (N)-L90 (O)	3.1
	N97 (ND2)-N73 (O)	3.0	H57 (NE2)- S91 (OG)	3.2
	N97 (ND2)-N73 (OD1)	2.8	**G216 (N)- C88 (O)**	**3.1**
	H57 (O)-S65 (OG)	3.2	**G216 (O)- C88 (N)**	**3.3**
			S217 (OG)- Q85 (NE2)	3.3
			S217 (OG)- Q85 (OE1)	3.5
			N97 (O)- R97 (NH1)	3.2
			N97 (OD1)- R97 (NH1)	2.6
			N97 (OD1)- R97(NH2)	2.8
			N97 (ND2)- M99 (O)	3.5
			N97 (O)- R71 (NH2)	2.6
			N97 (O)- R71 (NE)	3.5
			S217 (OG)- C60 (O)	2.9
			G216 (O)- C62 (N)	2.9
			G216 (N)- C62 (O)	3.1
			D189 (OD1)- K64 (NZ)	2.8
			D189 (OD2)- K64 (NZ)	2.7
			S195 (OG)- K64 (N)	3.0
			S195 (N)- K64 (O)	2.9
			S214 (O)- K64 (N)	3.3
			G193 (N)- K64 (O)	2.8
			F41 (O)- I66 (N)	2.0

Cysteine protease inhibitor	C63 (O)- Q49 (NE2)	3.4	K139 (NZ)- E77 (OE1)	2.9
	S21 (O)- Q49 (NE2)	3.1	**T54 (OG1)- Q142 (NE2)**	**2.6**
	G66 (N)- L4 (O)	3.0	T54 (OG1)- Q142 (OE1)	2.3
	T54 (OG1)- Q142 (NE2)	**2.3**	W177 (O)- S52 (OG)	2.7
	Q142 (OE1)- T54 (N)	2.4	Q19 (NE2)- S52 (O)	2.7

5.10 Conclusion

Developing disease resistant and stress tolerant varieties of plants is an important objective of breeding crops. Therefore, to help developing stress tolerant and pathogen resistant varieties, we have studied proteases and PIs in chickpea genome. A large fraction of chickpea genome encodes proteases and PI genes which provide resistance against various environmental stress conditions. A differential pattern of gene expression was reported in more than one plant tissue studied. The expression values showed that most of the genes express at basal level in normal condition though they may overexpress when encountered by adverse environmental conditions. The expression information reported here will be useful for further investigation of the functional characterization of these genes under various stress conditions. Structural studies have shown the high binding affinity and formation of more stable complex of protease inhibitor and target protein. These studies could increase our understanding of the roles of these genes in chickpea, but further functional analysis of stress-responsive proteases and protease inhibitors is required to confirm their role in stress tolerance.

Chapter 6

Genome-wide identification and tissue specific expression analysis of nucleotide binding site (NBS)-leucine rich repeat (LRR) gene family in *Cicer arietinum* (chickpea)

The disease resistance genes (R-genes) play a critical role in plant defense mechanisms and respond to attack by several pathogens and pests, including viruses, bacteria, fungi, nematodes, and insects. The signaling component required during a defense response is decided by the R-gene structure (Dong 1998).

6.1 Nucleotide binding site-leucine rich repeat (NBS-LRR) R-gene family

One of the major classes of proteins encoded by *R-gene* family possesses the nucleotide binding site-leucine rich repeat (NBS-LRR) domains. The NBS domain has several conserved motifs that bind and hydrolyze ATP or GTP (Tameling *et al*, 2002). The LRR regions are involved in protein-protein interactions and thus play role in molecular recognition and specificity (Kobe & Deisenhofer, 1995; Leister & Katagiri, 2000). Based on the structure of N-terminal domain, the NBS-LRR genes are divided into two families. The N-terminal domain of one of the families possesses homology with *drosophila* Toll and human Interleukin-1 receptors (TIR) therefore known as TIR-NBS-LRR (TNL), which is known to be involved in resistance specificity and signaling (DeYoung & Innes, 2006; Luck *et al*, 2000). The other family, where the TIR is absent or in its place a coiled-coil (CC) N-terminal domain involved in protein-protein interactions and signaling present, is known as non-TIR-NBS-LRR (non-TNL) or sometimes as CC-NBS-LRR (CNL) (Martin *et al*, 2003; Van Ooijen *et al*, 2007) (Figure 6.1). Moreover, the sequences of conserved motifs, especially those within the NBS domain, have been used extensively to identify novel disease resistance genes in the model and crop plants (Aarts *et al*, 1998; Zhu *et al*, 2002; Yaish *et al*, 2004).

6.2 Loss of crop productivity of chickpea

The fungal diseases that lead to extensive crop damage affecting chickpea productivity is already mentioned in the first chapter (Chérif *et al*, 2007) (www.icrisat.org/bt-pathology-fungal.htm). Genome sequencing of the model plants has aided genome-level investigation of NBS resistance gene family in monocot and dicot plant species such as *Oryza sativa* (Monosi *et al*, 2004; Zhou *et al*, 2004), *Arabidopsis thaliana* (Meyers *et al*, 2003; Tan *et al*, 2007), *Medicago truncatula* (Ameline-Torregrosa *et al*, 2008), *Zea mays* (Cheng *et al*, 2012), *Carica papaya* (Porter *et al*, 2009), *Cucumis sativus* (Wan *et al*, 2013), *Brassica rapa* (Mun *et al*,

2009), *Populus trichocarpa* (Kohler *et al*, 2008), *Vitis vinifera* (Yang *et al*, 2008), *Solanum tuberosum* (Lozano *et al*, 2012), *Linum usitatissimum* L. (Kale *et al*, 2013) and *Gossypium raimondii* (Wei *et al*, 2013). Previous studies have shown that NBS resistance genes constitute approximately 0.6 to 1.8% of the total genes encoded by plant genomes. Some plant genomes such as *M. truncatula*, *P. trichocarpa*, and *V. vinifera* encode for a large number of NBS resistance genes like 333, 402, and 459 each. Contrary to this, the genomes of *C. papaya* (54), *B. rapa* (92), *C. sativus* (57), and *Z. mays* (109) have lesser number of NBS resistance genes.

In this chapter the results of an *in silico* search conducted to identify NBS resistance genes present in chickpea genome are described. The findings will help to fish out candidate *R*-genes in chickpea for any type of manipulation for the betterment of crop productivity.

6.3 Identification of genes for NBS-LRR proteins in chickpea

The NBS-LRR proteins were searched in the predicted chickpea proteome dataset with the help of Basic Local Alignment Search Tool (stand-alone blastp 2.2.22) using an E-value cut off of 1. Consensus TIR-NBS-LRR (TNL) and non-TIR-NBS-LRR (non-TNL) sequences from plant extended NBS domain defined by Cannon *et al*, (2002) were used as blast query (Ameline-Torregrosa *et al*, 2008).

The hits were further reconfirmed by searching for the presence of NB-ARC domain against the chickpea proteome by using the HMM profile (Pfam-PF00931) of pfam 27.0. The E-value cut off used was 10^{-4}. The best hits obtained were used to build a chickpea-specific hidden Markov model using the module ‘‘hmmbuild’’ to check for any missing hit.

Blastp search using TNL and non-TNL consensus sequences as query was performed against the predicted chickpea proteome that consisted of 28, 269 gene models resulted in the identification of 100 NBS-LRR proteins. The nomenclature used for naming these proteins/ genes is according to the protein/ gene identifiers given in the LIS database. Out of the 100 putative protein sequences identified, 77 belonged to non-TNL and the remaining 23 sequences to TNL proteins. Another

fifteen sequences identified as NBS-LRR were not considered for further analysis due to high E-value and missing domains or motifs.

To confirm the above results further, hidden Markov model profile search was carried out against the chickpea proteome using HMMER (hmmsearch) by taking NB-ARC domain as query. A total of 124 hits were obtained, out of which 107 sequences were true positive hits. The remaining hits actually belonged to Pfam-AAA family ATPase/ tyrosine kinase. These 107 sequences were used to build a chickpea-specific hidden markov model to check for any missing hit. With this refined chickpea-specific model, a total of 201 NBS-candidate proteins were identified. Out of these 201 sequences, only 107 could be selected as true NBS and resistant candidate genes. The rest of the proteins belonged to kinase family. The seven extra sequences here, like the 15 sequences in the blast search, had missing domains or motifs and therefore were excluded from further analysis. Out of the 100 NBS-LRR gene sequences, following GenBank accession numbers were assigned to 67 [KF438075-KF438084, KF460534-KF460542, KF460544-KF460548, KF571705-KF571720, KF560321-KF560328, KF318728, KF577574-KF577584, KF550297-KF550300, KF711860, KF591102-KF591103] (CD-Table S-6.1). The remaining 33 were either located on scaffolds or contained ambiguous nucleotide bases (N) in their sequences. Out of the eight groups of plant resistance genes classified broadly on the basis of motif organization and membrane spanning regions [(1)NBS-LRR-TIR, (2)NBS-LRR-CC, (3)LRR-TrD, (4)LRR-TrD-KINASE, (5)TrD-CC, (6)LRR-TrD-PEST-ECS, (7)TIR-NBS-LRR-NLS-WRKY, (8)KINASE-KINASE-KINASE-HM1)] (Gururani *et al*, 2012) the seventh one with NBS domain (TIR-NBS-LRR-NLS-WRKY) was not found in the chickpea genome (Figure 6.2).

Varshney *et al*. (2013) reported presence of 187 disease resistance gene homologs (RGHs) in chickpea; however we could identify only 100 NBS-LRR disease resistance genes. There might be a possibility that the extra sequences identified by Varshney *et al*. belong to the kinase class of the R-gene which we have excluded in our analysis (Sanseverino *et al*, 2010). To validate this hypothesis, we checked the hits retrieved in HMM search. As expected we could identify 34 protein kinases and protein tyrosine kinase with multiple LRR domains. Along with that few sequences

falls into the category of ATPases Associated with diverse cellular Activities (AAA) family.

6.4 Disease resistant quantitative trait loci in chickpea

As already known, the productivity of chickpea crop is drastically affected majorly by two fungal diseases viz., *Fusarium* wilt (FW) and *Ascochyta* blight (AB) causing 100% yield loss in conditions favorable for infection. Sabbavarapu *et al*, (2013) identified two novel QTLs (*FW-Q-APR-6-1* and *FW-Q-APR-6-2*) for resistance against FW on linkage group (LG) 6 in FW cross ('C 214' * 'WR 315') with LOD values 8.0 and 7.6 and explained 10.4 and 18.8% phenotypic variation respectively. The AB populations were checked for both seedling resistance (SR) and adult plant resistance (APR). QTL analysis detected two QTLs viz. *AB-Q-SR-4-1* and *AB-Q-SR-4-2* for seedling resistance on LG 4 with load values 8.8 and 2.9 and explained 31.9 and 10.3 % phenotypic variation respectively. For APR, four QTLs were identified out of which two QTLs (*AB-Q-APR-6-1* and *AB-Q-APR-6-2*) were located on LG 6 with 4.8 and 5.1 LOD values explained 2.2 and 11.5% of phenotypic variations respectively. Moreover rest of the two QTLs, *AB-Q-APR-4-1* and *AB-Q-APR-5B* were detected on LG 4 and 5 with LOD values 4.3 and 3.1 explained 26.4 and 1.5 % of phenotypic variations. Similarly Flandez-Galvez *et al*, (2003) reported six QTLs for AB resistance in three regions of the genome of an intraspecific population of chickpea. Tekeoglu *et al*, (2004) reported two major and two minor QTLs conferring resistance against AB using recombinant inbred lines derived from interspecific cross *C. arietinum* * *C. reticulatum*.

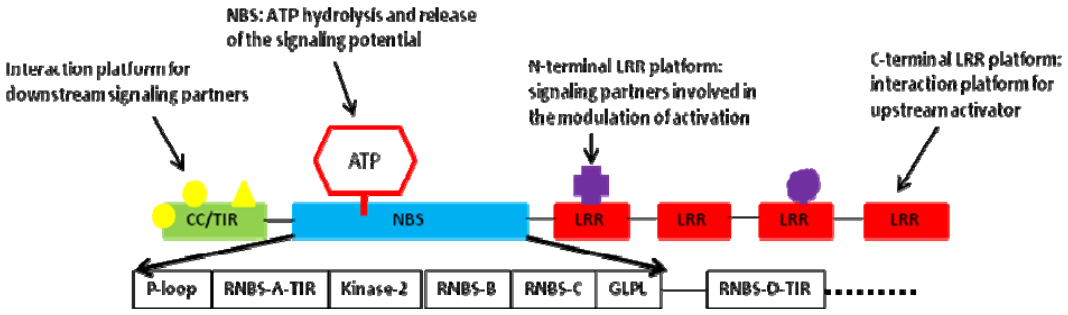


Figure 6.1: The schematic diagram shows the arrangement of domains present in the NBS-LRR proteins. The functional role of each domain is also shown.

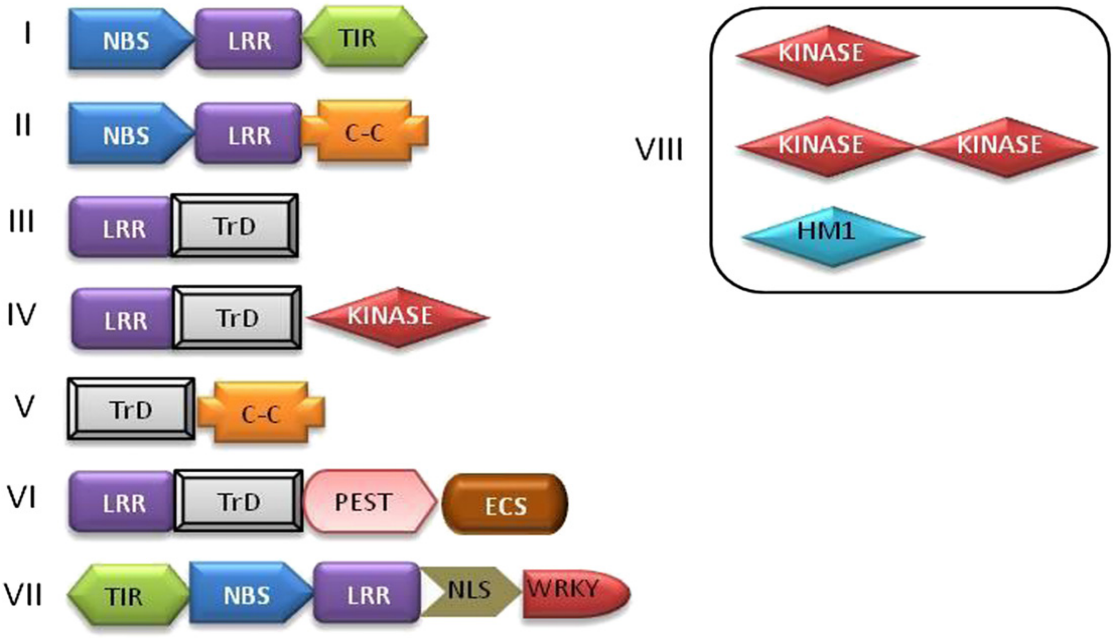


Figure 6.2: The image shows eight groups of plant resistance genes based on the motif organization and membrane spanning regions. Figure adopted from Gururani *et al*, 2012.

6.5 Phylogenetic analysis of R-gene family proteins

Generally, the 5' region preceding and the 3' region following the NBS domain have a high degree of sequence variability and therefore not considered for building the phylogenies. Meyers *et al*, (1999) reported that the phylogenetic analysis considering NBS domain classifying the sequences into non-TNL and TNL families. Therefore, NBS domains of 95 NBS-LRR proteins (74 non-TNL and 21 TNL) (P-loop to GLPL) were extracted for constructing the Neighbor Joining (NJ) phylogenetic tree (Figure 6.3 & 6.4). The remaining five sequences (non-TNL: Ca_00577, Ca_00576, Ca_12872 and TNL: Ca_10029, Ca_10030) were not considered for the phylogenetic analysis because either their NBS domains were incomplete or the signature motifs of the NBS domain were less conserved. The two families distinctly formed two separate clades in the dendrogram with high bootstrap values (Figure 6.5). Further, these families were subdivided into sister clades, CNL1-CNL4 and TNL1-TNL3. The number of non-TNL proteins was more than the number of TNL ones consisting of 77 and 23 members, respectively. This finding agrees with the distribution of NBS resistance genes in cucumber genome (Wan *et al*, 2013), however, different from the distribution in *Arabidopsis thaliana* where the TNL members outnumber non-TNL members (Meyers *et al*, 2003).

In the dendrogram two or more members from different chromosomal locations present in same phylogenetic clade could be a result of chromosomal rearrangement by transposition or by gene duplication. Six such segmental duplication events have been observed in chickpea (Figure 6.5, marked by green circles). Duplication is seen in all chromosomes except chromosome 4.

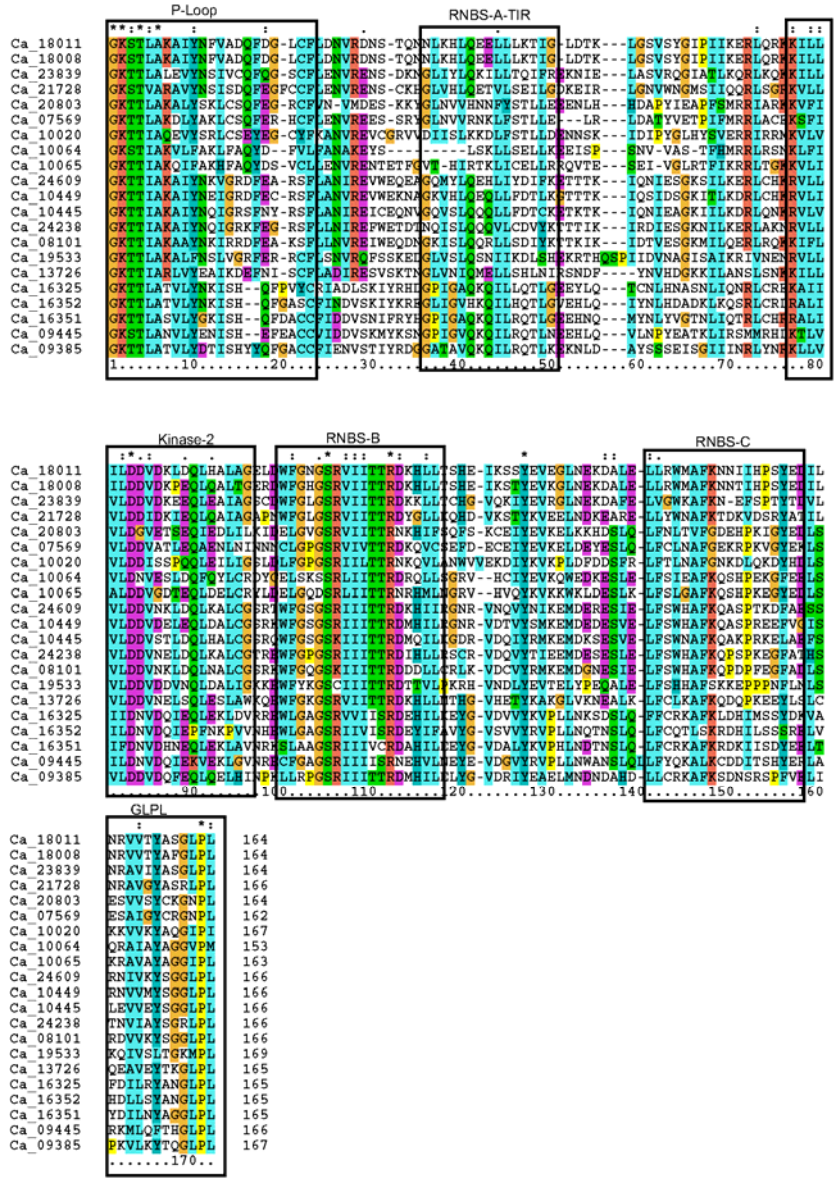


Figure 6.3: The NBS domains of TNL proteins of chickpea (from P-Loop to GLPL) were shown that were used to construct the phylogeny.

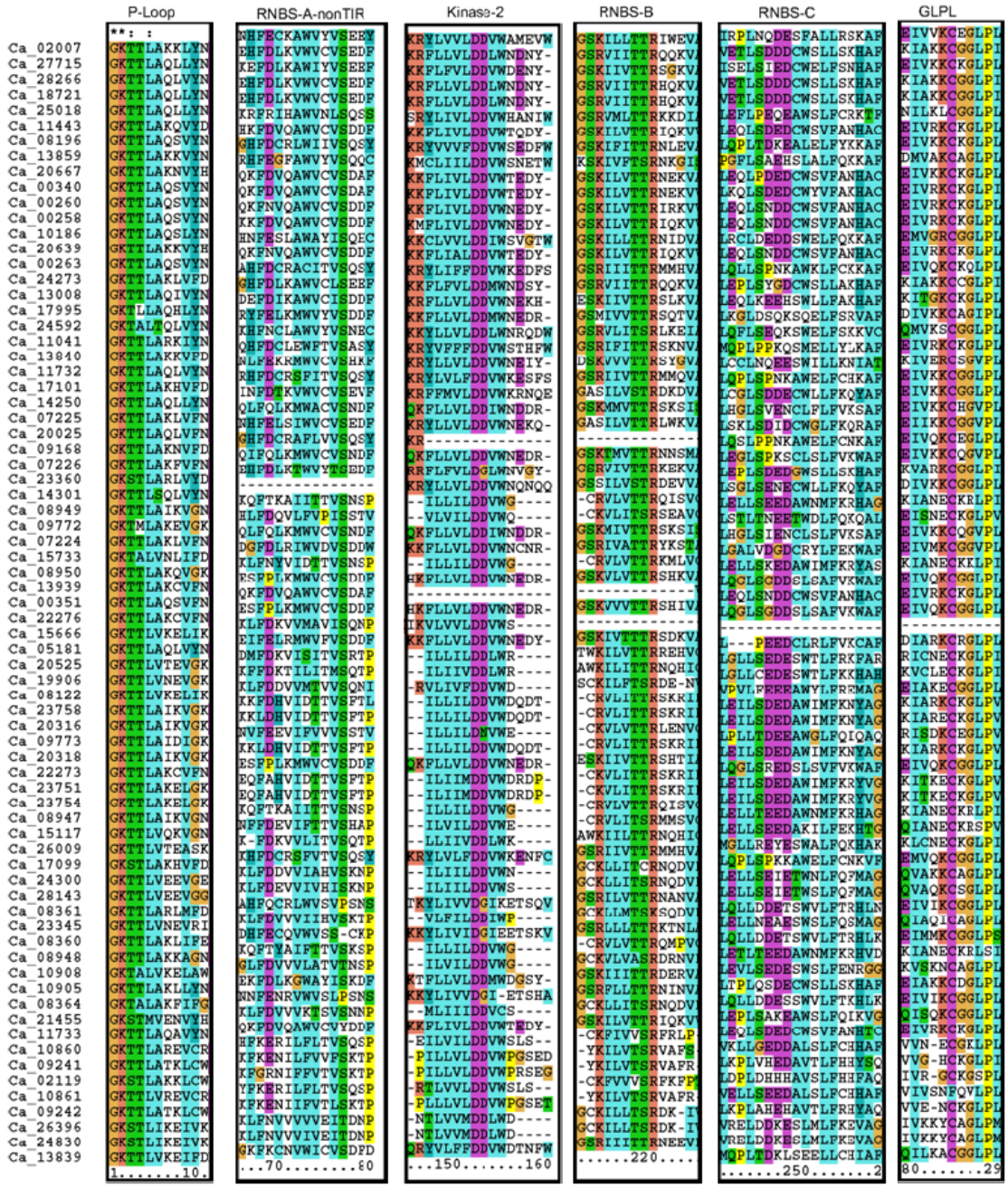


Figure 6.4: The NBS domains of non-TNL proteins of chickpea (from P-Loop to GLPL) are shown, same are used to construct the phylogeny.

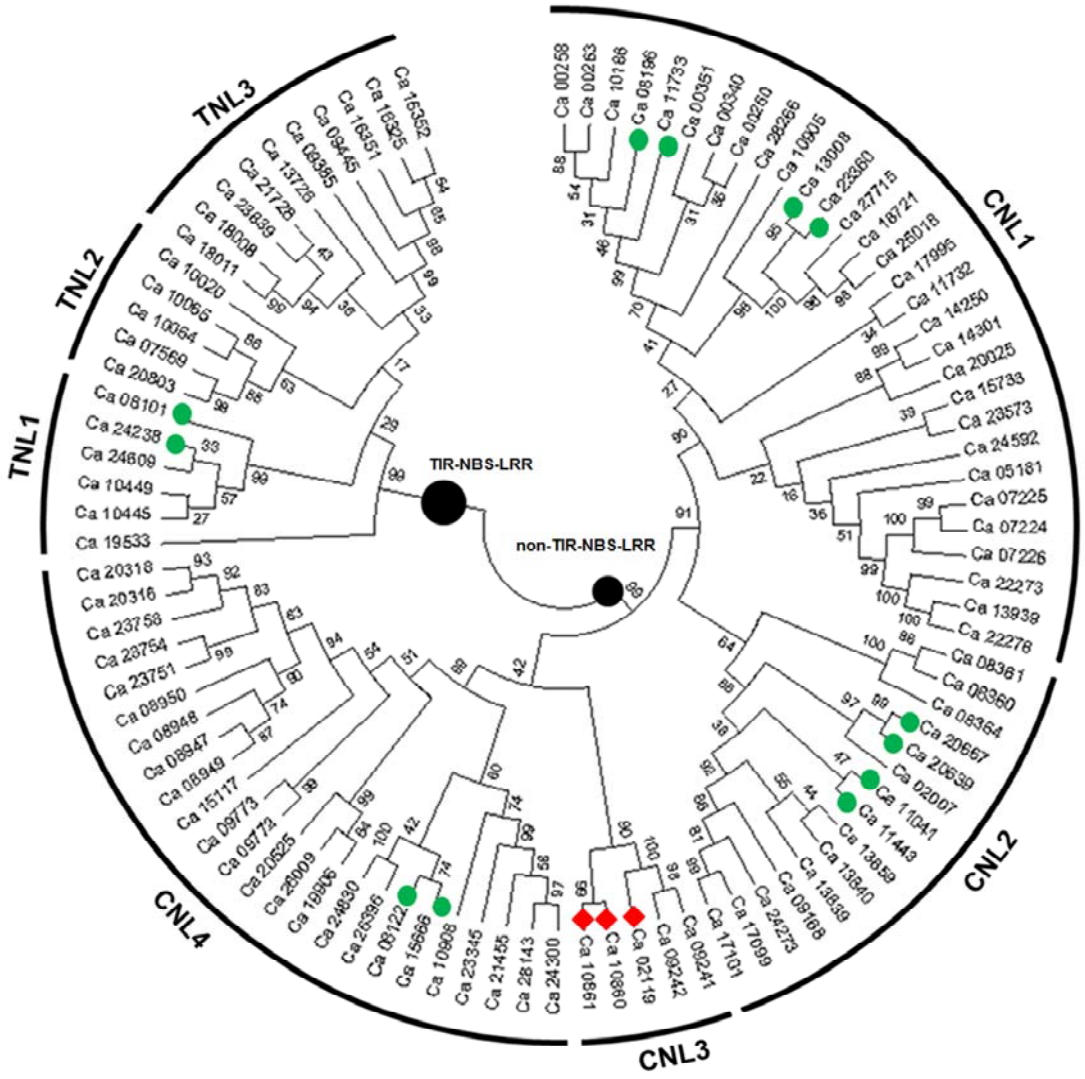


Figure 6.5: Circular representation of dendrogram reveals distinct clusters of non-TNL and TNL chickpea proteins. The two black circles show clades of two families of NBS encoding genes. The non-TNL family is divided into subfamilies CNL1 to CNL4. The TNL family is classified into subfamilies TNL1-TNL3. The diamonds represents the non-TNL proteins in which RPW8 domain fusion has occurred. The green circles depict pair of genes involved in segmental duplication events.

6.6 Distribution and clustering of R-genes

Some of the NBS resistance genes are present on chromosomes in isolation whereas the others are present as part of multi-gene clusters. The numbers per chromosome of 83 NBS resistance genes in chickpea genome distributed in Chromosomes 1 to 8 is 10, 10, 11, 9, 14, 10, 14 and 5 (Figure 6.6, CD-Table S-6.1). Mapping the remaining 17 genes on genome could not be accomplished and were therefore assigned to

scaffolds. The chromosome 8 has the least number of NBS resistance genes, in agreement with its small size.

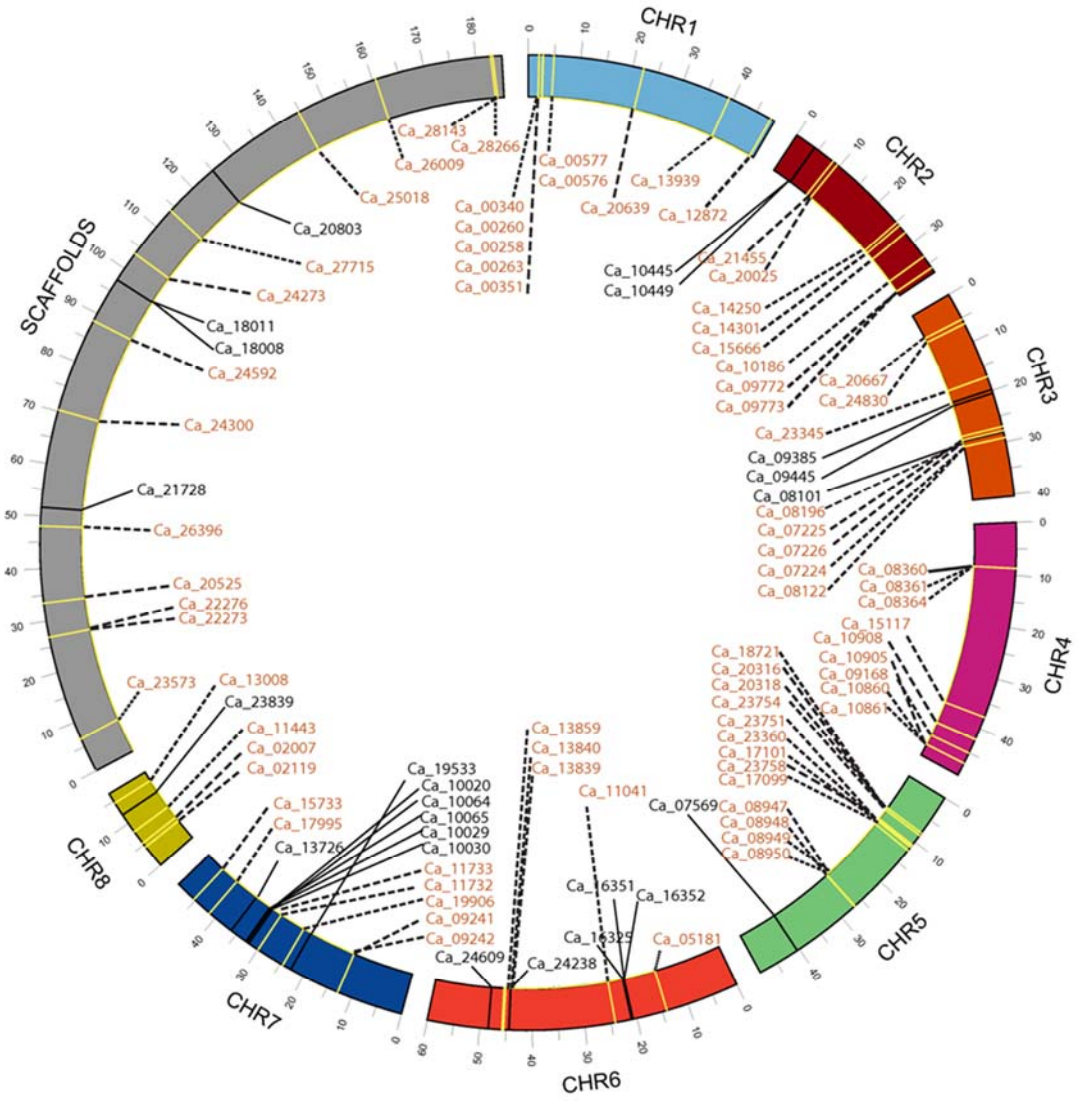


Figure 6.6: Distribution of non-TNL and TNL family members of NBS-LRR gene family on chromosome 1 to 8 and scaffolds of chickpea genome. Dashed and straight lines represent the non-TNL and TNL genes, respectively.

A gene cluster is defined if two neighboring homologous genes are not more than 200 kb apart and contain less than 8 non-NBS resistance genes between two NBS resistance genes (Meyers *et al*, 2003; Yang *et al*, 2008; Florian *et al*, 2012). Moreover, populations from a common ancestor tend to possess the same set of gene clusters that help to trace their recent evolutionary history. There are 21 gene clusters comprising of 48 NBS resistance genes in chickpea (Figure 6.6). Among these 21

clusters, four clusters were located on chromosomes 5 and 7, three on chromosomes 1 and 4, and two on chromosomes 2, 6 and scaffolds. Only one cluster was observed on chromosome 3. Chromosome 8 bears no gene cluster. Most of the gene clusters have two genes except those on chromosomes 1, 3, 4 and 5, which contained 3 to 4 genes (Table 6.1).

Table 6.1: Gene clusters are listed with genes involved in the formation of cluster along with their genomic location.

Cluster Number	Genes within the cluster	Number of genes in cluster	Chromosome/ Scaffold
1	Ca_00340, Ca_00351	2	1
2	Ca_00260, Ca_00258, Ca_00263	3	1
3	Ca_00577, Ca_00576	2	1
4	Ca_09772, Ca_09773	2	2
5	Ca_10445, Ca_10449	2	2
6	Ca_07225, Ca_07224, Ca_07226	3	3
7	Ca_08361, Ca_08360, Ca_08364	3	4
8	Ca_10908, Ca_10905	2	4
9	Ca_10861, Ca_10860	2	4
10	Ca_17101, Ca_17099	2	5
11	Ca_23758, Ca_23751, Ca_23754	3	5
12	Ca_20316, Ca_20318	2	5
13	Ca_08949, Ca_08947, Ca_08948, Ca_08949	4	5
14	Ca_13840, Ca_13839	2	6
15	Ca_16352, Ca_16351	2	6
16	Ca_11732, Ca_11733	2	7
17	Ca_09242, Ca_09241	2	7
18	Ca_10064, Ca_10065	2	7
19	Ca_10029, Ca_10030	2	7
20	Ca_22276, Ca_22273	2	scaffold242
21	Ca_18008, Ca_18011	2	scaffold1006

6.7 Orthologs identification and gene divergence

Close orthologs of chickpea NBS-LRR proteins were identified in sequenced plant genomes such as *M. truncatula*, *G. max*, *P. vulgaris* and *M. sativa*, which showed sequence similarity ≥80%. Maximum number of such orthologs was detected in *M. truncatula* (109) while the least number of orthologs was identified in *P. vulgaris* (1) and *M. sativa* (1) (Table 6.2, CD-Table S-6.2). In addition to this, half the total number of NBS-LRR genes was unique to chickpea with this criterion because orthologous relationship couldn't be established with any of the plant genomes considered (CD-Table S-6.2).

Table 6.2: Orthologs of NBS-LRR proteins of chickpea detected in other plant genomes.

Orthologs of chickpea non-TNL proteins	**Number of orthologs in genome**	**Number of chickpea orthologs**
Medicago truncatula	68	35
Glycine max	20	9
Orthologs of chickpea TNL proteins		
Medicago truncatula	41	15
Glycine max	5	3
Phaseolus vulgaris	1	1
Medicago sativa	1	2

6.8 NBS-LRR Pseudogenes

There were 27 pseudogenes with stop codon insertions in the chickpea genome identified with the help of TNL and non-TNL consensus sequences *via* tblastn query against chickpea chromosome assemblies filtered at E-value cut off of 1. The pseudogene count per chromosome from 1 to 8 was 2, 2, 6, 2, 8, 1, 5 and 1. As can be seen, the maximum number eight of the pseudogenes was present on chromosome 5. Count of pseudogenes representing non-TNL and TNL were 20 and 7. Nine pseudogenes were matched within 100 kb sequence upstream or downstream of

another predicted NBS resistance gene (CD-Table S-6.1). There is evidence from studies on *M. truncatula* that some NBS pseudogenes may get expressed (Ameline-Torregrosa *et al*, 2008). To investigate this further, the gene expression profile was checked for chickpea pseudogenes; it was observed that no chickpea pseudogene had support of gene expression data.

6.9 Domain distribution and arrangement in NBS resistance gene family

Domain arrangements in the 100 NBS-LRR proteins of chickpea were analyzed using hidden Markov model (HMM) search against Pfam database. Pfam did not predict CC motifs in the N-terminal region. Previous studies have suggested that the presence of specific signature motifs in NBS domain can be correlated with the presence or absence of CC motifs (Meyers *et al*, 1999; Meyers *et al*, 2002). We have used these signature sequences to classify the non-TNL family. The presence of CC regions was further validated with the help of HMMER search (phmmer) against UniProtKB protein database and MARCOIL server (Delorenzi and Speed 2002) using 9FAM matrix. The probability of N-terminal region showing Coiled Coil conformation in the MARCOIL plot was in the range of 0.4-0.8 for most of the predicted non-TNL proteins.

The canonical form of the domain (CNL & TNL) was observed in 19 non-TNL and 6 TNL proteins. In the non-TNL family, the predominant and unusual domain arrangement was CN type in 31 of the total 77 with missing LRR domain. Another unusual class of domains in some non-TNL proteins was the result of RPW8 domain fusing with N or CNL domain seen in Ca_10860, Ca_10861 and Ca_02119. The *RPW8* gene in *A. thaliana* provides example of a broad spectrum resistance genes against powdery mildew (Xiao *et al*, 2001). Compared to non-TNL, a more diverse arrangement of domains was found in TNL family namely TNL, TN, NL, N, L, TNTN, NN, TTNL and NTN (CD-Table S-6.1). Apart from the domain shuffling explained by Meyers *et al*, 2002, four new domains- CNNL, CNN, TNTN and NTN, were also identified here in these two families (Table 6.3).

Table 6.3: Numbers of NBS-LRR proteins with the listed predicted domain in chickpea.

Predicted protein domains	Letter code	Number of proteins
CC-NBS-LRR	CNL	19
CC-NBS	CN	31
RPW8- CC-NBS-LRR	RPW8-CNL	1
RPW8-NBS	RPW8-N	2
CC-NBS-NBS-LRR	CNNL	1
CC-NBS-NBS	CNN	4
TIR-NBS-LRR	TNL	6
TIR-NBS	TN	6
TIR-NBS-TIR-NBS	TNTN	1
TIR-TIR-NBS-LRR	TTNL	1
NBS-TIR-NBS	NTN	1
NBS-NBS	NN	1
NBS-LRR	NL	12
NBS	N	12
LRR	L	2

6.10 Motif identification

For identifying the degree of conservation and presence of signature motifs in protein families of NBS-LRR type, multiple sequence alignment and motif analysis were carried out with the help of ClustalX and MEME suite. However, these proteins generally have an N-terminal region (CC or TIR), NBS domain and LRR regions. For the sake of motif analysis protein sequences in the two families could be divided into three parts i.e. non-TIR/TIR, NBS and LRR.

6.10.1 non-TIR-NBS-LRR family

The types of motifs identified were two CC, ten NBS and six LRR in the sequences analyzed here. Presence of two CC motifs was common among the members of non-

TNL family. Also, most of the sites in these two motifs were poorly conserved. The NBS domain on the other hand showed a higher degree of conservation. The seven conserved motifs (P-loop, RNBS-A-nonTIR, Kinase-2, RNBS-B, RNBS-C, GLPL and MHDL) occurred in most of the non-TNL proteins. The RNBS-D-nonTIR motif was there in only a few sites. Apart from that, two additional motifs CNBS1 and CNBS2 were present in the NBS domain of non-TNL proteins. Similar pattern was seen in cucumber genome as well (Wan *et al*, 2013). Most of the sites in these two motifs were poorly conserved. Six LRR motifs were also found in non-TNL proteins. The pattern of occurrence of LRR motifs in these proteins was highly variable. A majority of them showed different LRR motifs. The motifs L1, L2 and L3 were widely present in most of the non-TNL proteins (Table 6.4, CD-Figure S-6.1, CD-Table S-6.3).

6.10.2 TIR-NBS-LRR family

A total of four, nine and six motifs were identified in N-terminal (TIR), NBS and LRR domains, respectively, of TNL family. Compared to non-TNL proteins, the conservation level of the NBS domain was high in TNL members (Figure 6.3 & 6.4). Except for RNBS-A-TIR motif, all the remaining eight motifs existed widely in more than half of the TNL proteins. Just like in non-TNL family, one additional motif, namely TNBS-1, is present in the members of this family. Although the same number of LRR motifs was observed in the TNL family, both the families had different LRR motifs (Table 6.5, CD-Figure S-6.2, CD-Table S-6.3). Even here most of the sites in these motifs were only weakly conserved.

Six conserved NBS motifs were present in both NBS resistance gene families (P-loop, Kinase-2, RNBS-B, RNBS-C, GLPL, and MHDL). However, RNBS-A-nonTIR and RNBS-D-nonTIR motifs were observed in only non-TNL members while RNBS-A-TIR and RNBS-D-TIR motifs were present in TNL proteins. Three additional motifs CNBS1, CNBS2 and TNBS1 were observed in chickpea which were not found even in its closest relative *M. truncatula* (Ameline-Torregrosa *et al*, 2008). These three NBS motifs were also located in NBS resistance genes of cucumber (Wan *et al*, 2013). These unique motifs can also distinguish between the two families of NBS-LRR genes.

Table 6.4: Major MEME motifs in predicted chickpea non-TIR-NBS-LRR family of proteins.

Domain	Sites	Motif	Motif sequence	E-value
CC	30	C1	L[LYT]A[IVL][EK]AVL[NL]DAE[QE]KQI[KT][ND]SA[VL][KN]NWLD[DEQ][LI]KD[AV][LV][YFS][DI]A[DE]D[LIV]LD[EH][IF][SN][TY][KE][AS][LA][RT][TCK][KQ]	$2.3* 10^{-515}$
	37	C2	WR[TR]P[ST][ST]SL[VI]x[EG]SN[VI][VIF]GR[ED]DD[KIR]E[KE][IL][IVL]xLLL	$1.8* 10^{-265}$
NBS	73	P-loop	V[IV][GP][IV]VGMGG[VL]GKTTL[AV][KQ][LEK][VL][YFG][NK]D	$6.4* 10^{-883}$
	69	RNBS-A-nonTIR	L[VK][AIV]WV[CT]VSDD[FP]D[IV]KK[IV][QT]KDI[AL]E	$1.2* 10^{-347}$
	72	Kinase-2	GK[KR][FI]L[LI][VI]LDDVWNEDYxD	$2.4* 10^{-426}$
	66	RNBS-B	AKG[SC][KR][IV][LI][VI]TTR[SN[KQ]KVASxMG[TC]	$8.3* 10^{-481}$
	76	RNBS-C	[HY]xLELLS[DE][ED][DE][ACS]WSLFx[KN]HA[FG]L	$2.2* 10^{-498}$
	75	GLPL	I[VA][KR]KC[KG]GLPLA[IA][VK][TAV][LIV][GA][GS][LS]L[RK][GS]K	$4.4* 10^{-740}$
	15	RNBS-D-nonTIR	K[ND][EK][KE]AK[RE]L[FL][LM]L[CS]S[VL]FPED[EY][EDI]I	$4.1* 10^{-253}$
	47	CNBS-1	I[LI]PAL[KR][LI]SY[DH][YD]LP[SP][YH]LKRCF[AL]Y[CF]S[LI][FY]P[KE][DG][YF]E[FI]DK[KD]DLILLW[MV]AEG[FL][LVI]Q[PS][PS]	$3.3* 10^{-1107}$
	46	CNBS-2	KTLE[ED]V[GA]EEY[FL]N[ED]L[IL]SRS[FL][FI]Q	$4.4* 10^{-339}$
	64	MHDL	[FV][KV]MHDLV[RH]DLALL[IV][AS]G[KE]EYFR	$1.6* 10^{-455}$
LRR	64	L1	x[SA]LPDSIG[EN]L[IK][HN]LRYLDLSxT	$1.8* 10^{-333}$
	75	L2	LxSLPxxLGAL[PV]SL[EKR]xL[DE]IRxCPKL[KE]S[IL]PxGIxxLxNL[EQ]xL	$3.9* 10^{-301}$
	70	L3	KSLP[DS]SI[CG][NK]LxNLQTL[DK]LxGC	$2.9* 10^{-416}$
	11	L4	FPSLE[RK]LEFD[CDN]MP[CN]W[EK][VE]W[HI]H[PY][HE][DS][SG][NE]A[YA]FPVL[KR][ST]L[VS][IL]RGCP[LK]L[RM]G[DH	$1.4* 10^{-259}$

			]LP[SN]HL	
	17	L5	QP[SA]KNLKK[LV]SI[CD]GY[GR]GT[RS]FPEW[LV]GD[PS]S[YF]SN[LM][TV][KS]L[SY][LI]S	$1.9* 10^{-279}$
	19	L6	[SK]L[SC]I[SK][KN]L[EH]NV[IS][ND]N[FIV][ED]AS[QD]A[NK][LIM][MK][DS]K[KE][YH][LI]E[EK]L[SE][LF][EQ]WSD[NQ][AT][KE]D	$1.0* 10^{-267}$

Table 6.5: Major MEME motifs in predicted chickpea TIR-NBS-LRR family of proteins.

Domain	Sites	Motif	Motif sequence	E-value
TIR	13	T1	YDVF[IVL]SFRGEDTRY[NS]FT[GS][HN]LY	$3.1*10^{-109}$
	15	T2	FY[DEK]VDP[ST[DHE]VR[HKY]Q[TEKS]GS[YF][GAE][KE]A[FL]A[KE][HL]E[EKT]R[FL][QN]xD	$1.9*10^{-145}$
	11	T3	[EDN][KM][LV]Q[RQ]W[KR][AL]ALTQ[AV]AN[LI]SG[WF][DH]	$1.8*10^{-055}$
	14	T4	[KR]G[DE][ES]I[TAS]P[SE]L[LI][EK]AIENS[RQ]I[FS][IV]V[VI][FL]SKNYASS[ST][WF]CL[RD]EL[EV][HKY]I[LIM]	$7.3*10^{-223}$
NBS	20	P-loop	IWGMGG[IM]GK[TS]T[LI]AK[AV][LIV]YNK[IL][SG][HRD]Q[FY][ED][GA]R[CS]F	$2.6*10^{-268}$
	5	RNBS-A-TIR	[GN][QK][IVM][SHY]LQ[QE][QHR][LV][LI][FCSY]D[ITV][YCFL]K	$6.5*10^{-054}$
	21	Kinase-2	[KR][VI]L[LI][VI]LD[DN]VDKL[ED]QLEAL[CA]G	$1.4*10^{-347}$
	21	RNBS-B	[SN]R[DE]W[FL]GxGSR[IV]IITTRDKH[IL]L	$1.4*10^{-347}$
	23	RNBS-C	L[FL][SC][WL][NHK]AFKQ[DN]H[PI]K[SE]G[YF][EA][ED]	$1.8*10^{-328}$
	23	GLPL	[VAI][VL]xY[AS]GGLPLA[LI][EK]VLGS[FNY]LF[GD][RK]DIxEW[KE]	$1.8*10^{-328}$
	20	RNBS-D-TIR	[YFD][DG][GL]L[DE][ED][MT][EQ]K[EN][IV]FLDIA[CF]FF[IKN]G	$2.6*10^{-155}$

	19	TNBS-1	[AG]CG[FL]F[AP][DE]IG[IL][ES][VI]L[VI][ED][KR][SA]L[IV][TK]I	$1.9*10^{-086}$
	22	MHDL	MHDL[LI][QER][DE]MGREI[VI]R[EQ][EK]SP[EK]EP	$8.8*10^{-142}$
LRR	21	L1	[KE]RSRLW[DF]P[EK]D[IV]YDVL[KS][EK]N[KT]G[TS]	$3.9*10^{-109}$
	17	L2	[VA][LI]DLS[EK]xN[EK]LxL[NS]G[DKE][AS]FKKM[KT][KN]LRLL[QDI][LI]Y	$1.8*10^{-049}$
	16	L3	L[SP][ND][EK]LR[YW]L[SR]W[DN]G[YF]P[LF][EK]SLPS[KS]FC[AP]ENLVE[LI]N[LM]P[NHY]S[NQ]	$8.2*10^{-187}$
	17	L4	W[KH]GM[QK]xL[PE]NL[KR]I[LI][DN]LSxSKNLT[EK][LIT]PD[FL]S[GK][AL][PT]NLE[VR]L[DN]L[ES]GCVSLT[QES][VI]HPS	$1.1*10^{-226}$
	21	L5	FSVS[MS]xS[LM]VxL[DN]LSx[TC]GIH[EQ]LPSSIGxLTKLExLN[LV]E[CG]C	$5.3*10^{-069}$
	18	L6	PxS[IL]KEL[SK]xLE[NY]L[DIS]LS[HN]C[RK]E[LI]Q[SE][LI]PELP[PS]S[LM][KE]TL[DT]AL[NG]Cx[SN]	$3.1*10^{-123}$

6.11 Exon-intron architecture

The detailed illustration of exon-intron arrangement in chickpea NBS resistance genes is shown in Figure 6.7 & 6.8. The number of introns ranges from 0 to 11 in both the families. Interestingly, not a single TNL gene was intronless, whereas 25 non-TNL genes had no introns. More than half of the identified proteins (47 out of 83) were either intronless or had intron number ranged from 1 to 4. Moreover, the genes in CNL1 and CNL2 subfamilies had the lowest number of introns (0-3). The gene structure of CNL3 subfamily had only 4, 5 and 6 introns. The CNL4 subfamily possesses a diverse pattern of introns that ranges from having maximum 11 introns to possessing no introns. The gene structure of TNL family members had least number of introns (1-4) in except Ca_10064, which contained 11 introns (Figure 6.7 & 6.8). These findings propose that the intron gain and loss events are progressively occurring during the structural evolution of the two families of chickpea NBS resistance genes.

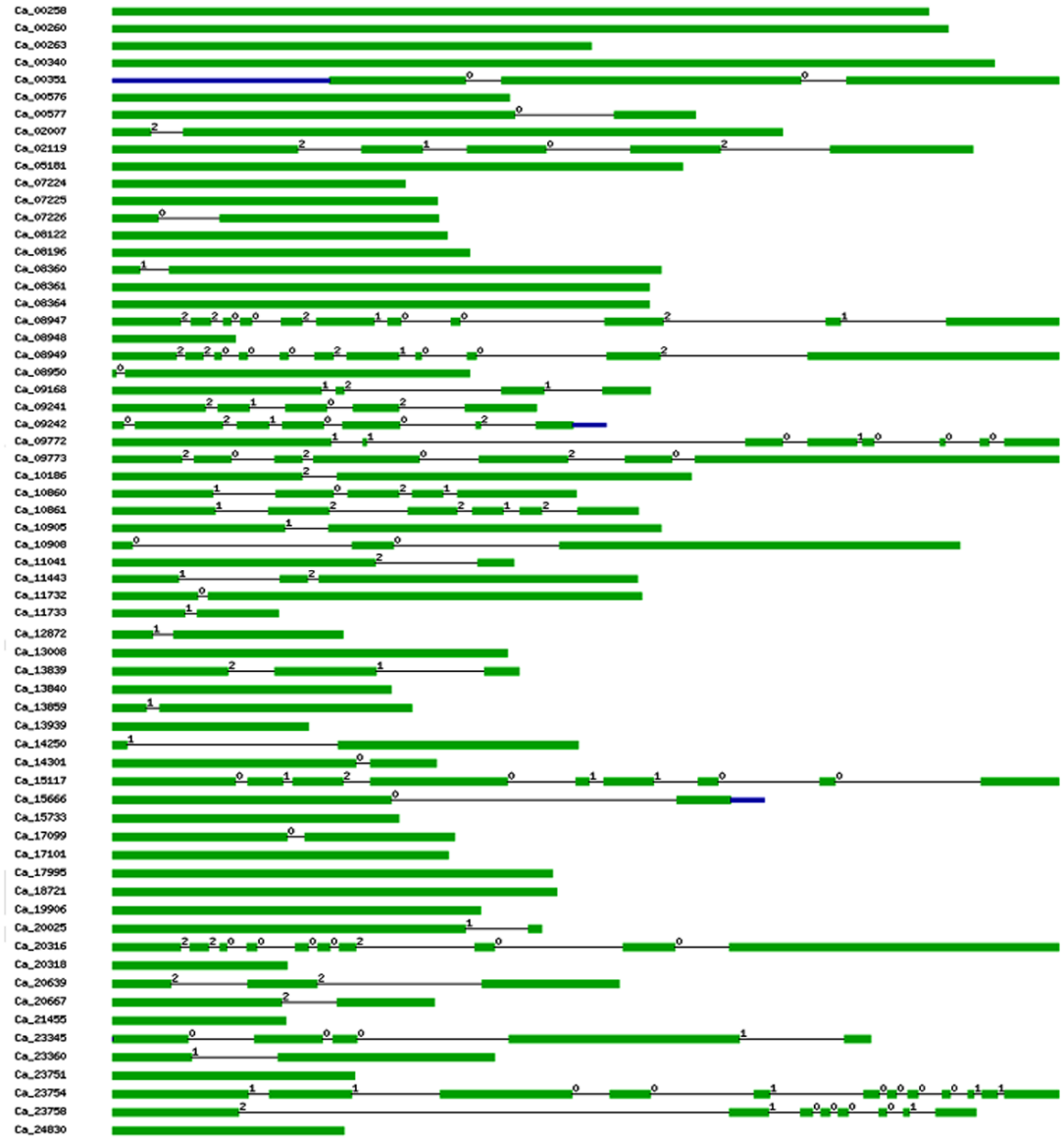


Figure 6.7: Exon-intron arrangement of non-TNL genes of chickpea.

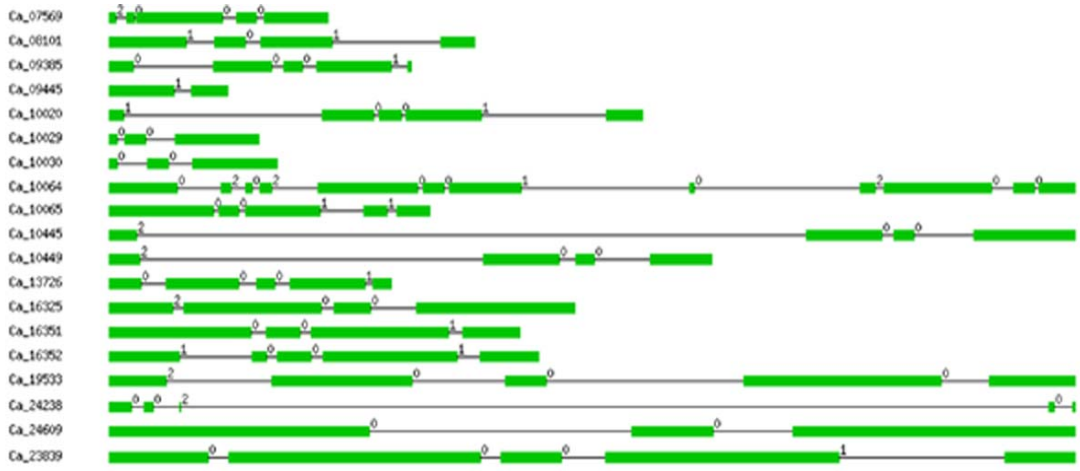


Figure 6.8: Exon-intron arrangement of TNL genes of chickpea.

6.12 Gene expression studies using RNA-seq and EST search

6.12.1 RNA-seq data analysis

Out of the total 100 NBS resistance genes, 39 (29 non-TNL and 10 TNL) showed medium to high expression level (FPKM >=5) in at least one of the 5 tissues selected (flower, flower bud, shoot apical meristem, young leaves and germinating seedling) whereas 56 (45 non-TNL and 11 TNL) showed low expression (5 >FPKM >0) and 5 showed no expression (FPKM=0) (Ca_02119, Ca_08950, Ca_09773, Ca_10020, and Ca_13726) in all the five tissues under study (Table 6.6 & 6.7). Differential expression patterns was seen across the tissues though most of the non-TNL and TNL genes showed high expression in germinating seedling and shoot apical meristem (Figure 6.9 & 6.10).

The RNA-seq data showed that more than half the NBS resistance genes expressed at low level and five genes remained unexpressed. This observation is supported by expression pattern of NBS resistance genes in *Arabidopsis* in which the expression has been at low levels and with a variety of tissue specificities (Tan *et al*, 2007).

The drought stressed RNA-seq reads from the two different genotypes of the chickpea revealed over-expression of 24 non-TNL and 10 TNL genes of chickpea as well as some are under-expressed also as compared to the control samples (CD-Figure S-6.3, CD-Table S-6.4)

6.12.2 EST data analysis

In addition to the RNA-seq data analysis, EST libraries of chickpea, available in NCBI database, were also explored to get further support for the transcriptional evidences of these predicted genes. As most of the NBS resistance genes are similar, an EST can be mapped to over more than one gene sequence; therefore only the top matches were considered by putting a sequence identity constraint of ≥ 90% between the EST data and the coding sequence. The gene expression data is available in various libraries constructed from various developmental stages, tissue types, drought and saline challenged or non challenged tissues. Very few genes had EST support due to paucity of data available in the public repository or gene expression missed due to very low level expression or only expressed under specific conditions in specific tissues. However, we could identify twelve genes (ten non-TNL and two TNL genes) which may express in root, stem and leaf tissues (Table 6.8). RNA-seq analysis also showed expression of these 12 NBS resistance genes though most of the plant tissues tested was different. The gene expression was matching for specific genes when checked in the same tissues (leaf) by using both the methods.

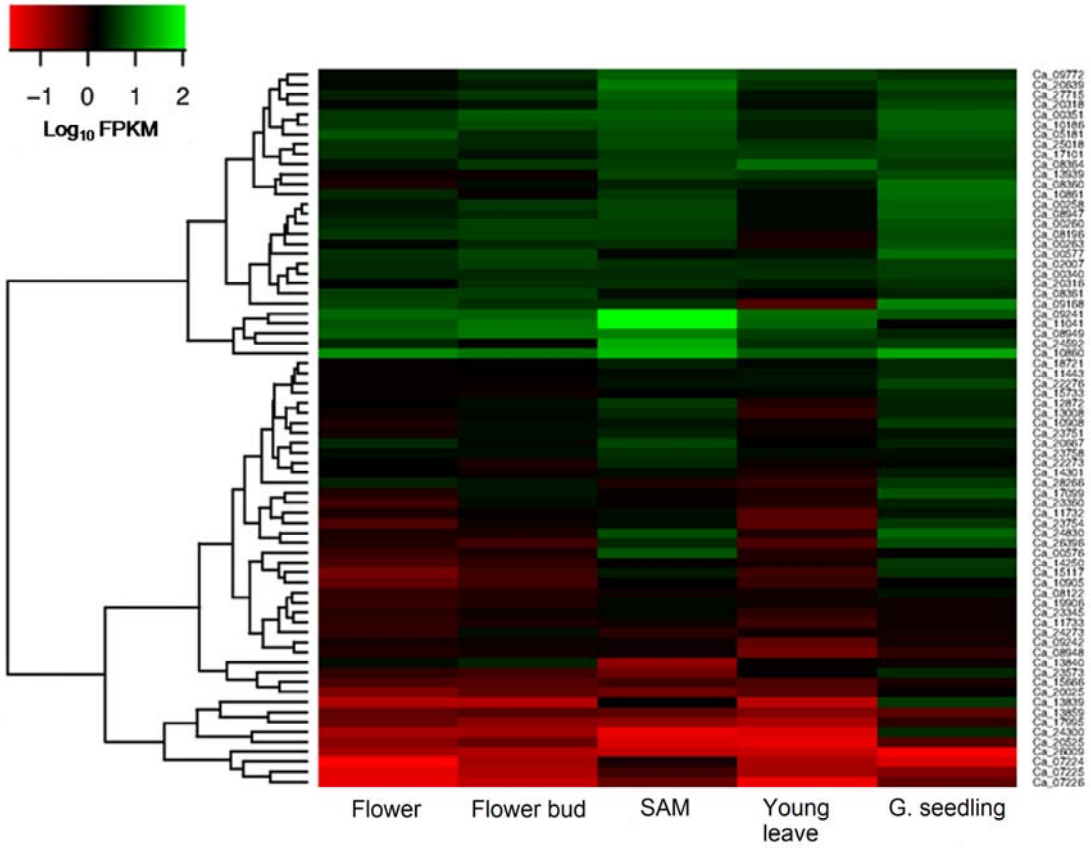


Figure 6.9: Heatmap shows relative gene expression of chickpea non-TNL genes in various tissue samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for NBS-encoding genes in different tissues. The NBS genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

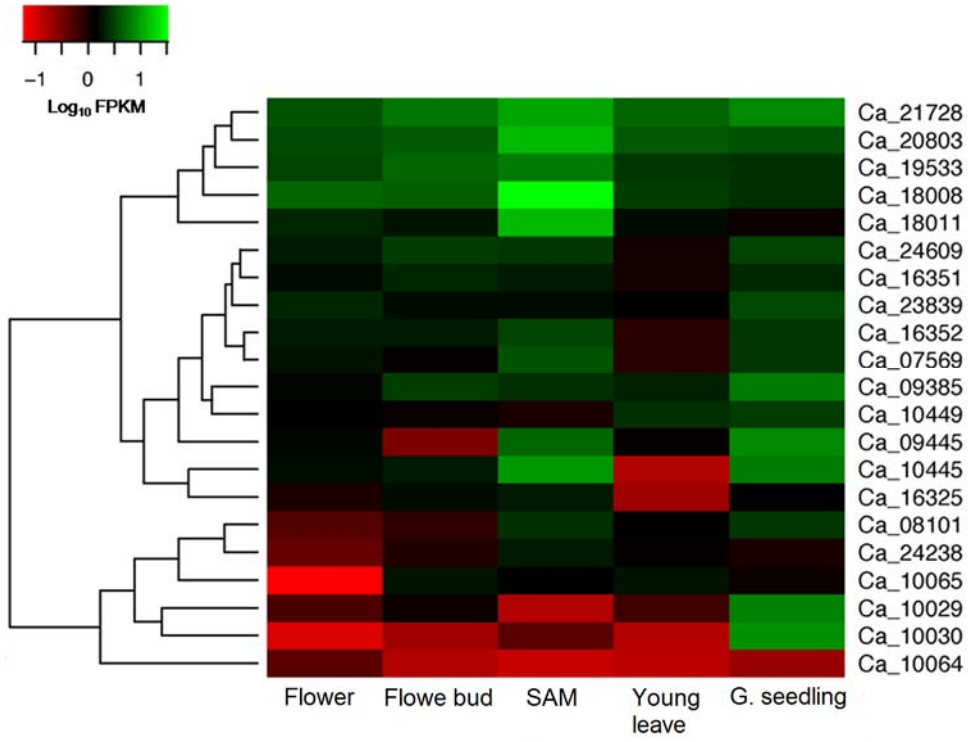


Figure 6.10: Heatmap shows relative genes expression of chickpea TNL genes in various tissue samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for NBS-encoding genes in different tissues. The NBS genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

Table 6.6: Expression values of non-TNL chickpea genes in five plant tissues.

No.	Seq id	Flower	Flower bud	SAM	Young leaves	Germinating seedling
1	Ca_28143	0.01	0.07	0	0	0
2	Ca_28266	2.82	2.20	0.90	0.75	4.0
3	Ca_00260	3.05	4.59	4.57	1.72	6.36
4	Ca_00263	1.71	3.07	3.56	0.97	5.99
5	Ca_00258	2.36	3.83	4.68	1.89	7.37
6	Ca_00340	3.24	3.29	3.07	3.59	4.19
7	Ca_00351	3.91	7.16	7.85	3.13	7.77
8	Ca_00576	0.63	0.98	6.29	0.92	1.39
9	Ca_00577	3.65	5.16	1.77	2.17	10.15
10	Ca_20639	1.79	2.47	11.66	3.37	5.49
11	Ca_13939	1.34	1.25	5.13	3.98	5.24
12	Ca_12872	1.57	1.95	3.83	0.88	2.95
13	Ca_21455	0	0.08	0	0	0.03
14	Ca_20025	0.21	0.35	0.26	0.42	1.43
15	Ca_14250	0.48	0.74	1.45	1.07	4.04
16	Ca_14301	1.52	1.24	2.11	1.07	2.51
17	Ca_15666	0.44	0.41	0.59	0.43	1.05
18	Ca_10186	3.98	5.62	6.31	2.52	7.87

19	Ca_09772	2.08	3.34	7.57	5.15	3.58
20	Ca_20667	2.98	1.75	4.79	1.67	2.81
21	Ca_24830	1.06	0.97	5.28	0.90	8.29
22	Ca_23345	0.79	1.31	1.87	0.85	1.29
23	Ca_08196	3.80	4.52	4.42	1.13	5.29
24	Ca_08122	0.69	0.89	1.33	1.11	2.15
25	Ca_07226	0.03	0.06	0.29	0.03	0.32
26	Ca_07225	0.03	0.09	0.58	0.10	0.17
27	Ca_07224	0.02	0.10	0.97	0.09	0.04
28	Ca_08360	1.06	2.08	2.96	2.53	9.88
29	Ca_08361	4.35	4.20	2.05	1.61	2.87
30	Ca_08364	2.30	3.94	4.24	9.91	3.73
31	Ca_15117	0.23	0.59	2.50	0.59	3.42
32	Ca_10908	1.00	1.99	2.22	1.33	4.19
33	Ca_10905	0.31	0.58	1.59	0.61	1.61
34	Ca_09168	4.93	3.47	3.42	0.48	13.48
35	Ca_10861	3.19	1.40	4.02	1.80	9.23
36	Ca_10860	16.10	11.43	32.56	7.59	23.50
37	Ca_18721	1.46	1.56	2.65	1.76	3.20
38	Ca_20318	1.77	2.20	6.31	1.98	5.37
39	Ca_20316	1.62	3.57	3.43	2.45	3.42
40	Ca_23751	1.08	1.96	2.71	1.23	2.20
41	Ca_23360	0.59	2.11	1.38	0.92	2.75
42	Ca_23754	0.44	1.28	1.97	0.37	4.08
43	Ca_23758	1.90	2.11	3.55	2.08	2.06
44	Ca_17099	0.87	2.25	1.37	1.00	6.30
45	Ca_17101	3.40	2.17	3.76	4.54	5.23
46	Ca_08948	0.88	1.16	1.26	0.25	0.73
47	Ca_08947	2.40	3.19	4.83	1.75	7.95
48	Ca_08949	7.66	10.50	13.32	4.63	2.68
49	Ca_05181	6.39	3.27	5.83	2.63	6.11
50	Ca_11041	5.87	10.99	90.15	8.63	1.56
51	Ca_13839	0.08	0.08	1.67	0.07	3.98
52	Ca_13840	2.09	2.69	0.14	1.36	1.28
53	Ca_13859	0.34	0.41	0.37	0.18	0.34
54	Ca_09242	1.13	1.23	1.27	0.33	1.08
55	Ca_09241	8.63	7.80	110.65	9.62	7.24
56	Ca_19906	0.67	0.88	1.85	1.26	1.29
57	Ca_11733	0.72	1.00	1.34	0.59	1.31
58	Ca_11732	0.99	1.46	1.93	0.39	2.23
59	Ca_17995	0.30	0.17	0.14	0.09	0.70
60	Ca_15733	1.40	1.21	1.56	1.74	3.20
61	Ca_02007	3.14	4.59	3.52	3.04	4.48
62	Ca_11443	1.50	1.54	1.93	2.20	3.23
63	Ca_13008	1.32	1.88	3.01	0.69	2.95
64	Ca_24273	0.71	2.02	0.80	1.10	1.14
65	Ca_27715	2.68	4.06	7.68	2.21	4.14
66	Ca_25018	3.63	3.25	5.36	4.08	4.79
67	Ca_22273	1.54	1.02	3.17	1.34	1.74
68	Ca_22276	1.37	1.48	2.17	2.36	4.68
69	Ca_20525	0.17	0.28	0.04	0.03	0.39
70	Ca_26009	0.10	0.08	0.06	0.06	0.02
71	Ca_26396	0.93	0.55	3.07	0.46	5.43

72	Ca_23573	0.75	0.58	0.27	1.60	3.25
73	Ca_24300	0.10	0.11	0.03	0.04	3.13
74	Ca_24592	3.11	1.86	26.98	3.37	4.40
75	Ca_02119	0	0	0	0	0
76	Ca_08950	0	0	0	0	0
77	Ca_09773	0	0	0	0	0

Table 6.7: Expression values of TNL chickpea genes in five plant tissues.

No	Seq id	Flower	Flower bud	Sam	Young leave	Germinating seedling
1	Ca_10449	1.36	1.22	1.02	2.56	3.07
2	Ca_10445	1.65	1.90	8.83	0.14	6.49
3	Ca_09385	1.48	2.90	2.62	2.17	6.16
4	Ca_09445	1.56	0.29	5.18	1.32	7.57
5	Ca_08101	0.51	0.79	2.52	1.38	2.70
6	Ca_07569	1.86	1.32	3.99	0.84	2.78
7	Ca_16352	1.96	1.96	3.34	0.82	2.74
8	Ca_16351	1.59	2.28	1.91	1.08	2.39
9	Ca_16325	1.02	1.68	1.99	0.19	1.37
10	Ca_24238	0.39	0.88	1.98	1.24	1.00
11	Ca_24609	1.91	2.93	2.82	1.10	3.12
12	Ca_19533	3.28	5.18	6.17	2.83	2.53
13	Ca_10064	0.47	0.15	0.12	0.14	0.21
14	Ca_10065	0.05	1.76	1.43	1.79	1.21
15	Ca_10030	0.09	0.20	0.47	0.15	8.17
16	Ca_10029	0.54	1.22	0.15	0.62	6.75
17	Ca_23839	2.35	1.65	1.72	1.36	3.47
18	Ca_18008	5.02	4.70	33.61	2.95	2.49
19	Ca_18011	2.24	1.73	13.41	1.67	1.17

20	Ca_20803	3.66	4.22	14.36	4.22	3.85
21	Ca_21728	4.01	5.79	10.57	5.01	7.59
22	Ca_10020	0	0	0	0	0
23	Ca_13726	0	0	0	0	0

Table 6.8: The result of EST search against chickpea chromosome assembly. 12 NBS-LRR genes have respective EST hits given in 3rd column. Other statistics like maximum score, length of EST matched, query coverage, sequence identity, E-values and the respective tissues in which the genes may show expression are given in the table

No	Gene ID	Accession number	Max Score	EST length	QC (%)	Identity (%)	E value	Tissue
1	Ca_10186	GR404898.1	484	524	8	92	$8*10^{-136}$	Root
2	Ca_10861	HS108365.1	874	517	23	97	0	Leaf
		HO062456.1	505	711	19	100	$6*10^{-142}$	Root
		CV793593.1	388	238	9	99	$6*10^{-107}$	Stem & leaf
3	Ca_12872	FE671101.1	453	594	12	100	$2*10^{-126}$	Leaf
4	Ca_18721	GR397596.1	1158	653	16	99	0	Root
		FE671730.1	717	465	12	95	0	Leaf
5	Ca_00258	GR404898.1	448	524	9	91	$1*10^{-124}$	Root
6	Ca_00260	GR404898.1	462	524	9	91	$4*10^{-129}$	Root
7	Ca_10860	GR393359.1	669	513	14	99	0	Root
8	Ca_27715	GR397596.1	911	653	16	93	9	Root
		FE671730.1	601	465	12	90	$7*10^{-171}$	Leaf
9	Ca_25018	GR397596.1	1114	653	16	98	0	Root
		FE671730.1	846	465	12	99	0	Leaf
10	Ca_24273	HO064439.1	534	733	11	96	$8*10^{-151}$	Root
11	Ca_18011	FE671960.1	503	346	11	93	$2*10^{-141}$	Leaf
12	Ca_18008	FE671960.1	503	346	18	93	$2*10^{-141}$	Leaf

6.13 *In Silico* promoter analysis of NBS resistance genes

A 2 kb upstream region of the NBS resistance genes was extracted and the promoter region was analyzed in detail. Four *cis*-regulatory elements, known to be involved in stress conditions and pathogen attack, were found overrepresented in the promoter region of NBS resistance genes. The four promoter regions considered were WBOX associated with WRKY transcription factor (Dong *et al*, 2003), DRE (Sakuma *et al*. 2006), CBF (Ohme-Takagi *et al*, 2000) and GCC box. WBOX elements were widely present in both the families averaging at 3.6 for non-TNL and 4.52 for TNL though few exceptions exist which had no WBOX (Ca_08949, Ca_13839 and Ca_13840). Rest of the three boxes were present in quite a low number averaging to 0.20 (DRE), 0.10 (CBF) and 0.12 (GCC). Two WBOXs was commonly seen in some twenty NBS resistance genes, nine WBOXs being the maximum. The other regulatory elements were observed only once per upstream with few exceptions of two boxes were also seen (Table 6.9). Similar arrangement and degree of occurrence of *cis*-regulatory elements was observed in *M. truncatula* (Ameline-Torregrosa *et al*, 2008).

Table 6.9: List of disease resistance specific *cis*-regulatory elements present in non-TNL chickpea genes.

No	Seq ID	WBOX	DRE	CBF	GCC
1	Ca_00258	1	0	0	0
2	Ca_00260	2	0	0	0
3	Ca_00263	2	1	0	0
4	Ca_00340	6	0	0	0
5	Ca_00576	6	0	0	0
6	Ca_02007	1	0	0	0
7	Ca_02119	9	0	0	0
8	Ca_05181	9	1	0	0
9	Ca_07224	4	0	2	0
10	Ca_07225	2	0	0	0
11	Ca_07226	2	0	0	1
12	Ca_08122	3	0	0	0
13	Ca_08196	3	0	0	0
14	Ca_08361	7	0	0	0
15	Ca_08364	2	0	0	0
16	Ca_08947	4	0	0	2
17	Ca_08948	2	0	0	0
18	Ca_08949	0	0	0	0
19	Ca_08950	4	0	0	0
20	Ca_09168	1	0	0	0

21	Ca_09241	2	3	2	0
22	Ca_09242	4	0	0	0
23	Ca_09773	3	0	1	0
24	Ca_10860	2	1	0	2
25	Ca_10861	2	0	0	0
26	Ca_11041	2	0	0	0
27	Ca_11443	6	0	0	0
28	Ca_11733	4	0	0	0
29	Ca_12872	2	0	0	0
30	Ca_13008	2	0	0	0
31	Ca_13839	0	0	0	0
32	Ca_13840	0	0	0	0
33	Ca_13859	3	0	0	0
34	Ca_13939	1	0	0	0
35	Ca_14250	5	0	0	0
36	Ca_14301	2	0	0	0
37	Ca_15117	5	0	0	0
38	Ca_15733	7	0	0	0
39	Ca_17101	2	0	0	0
40	Ca_17995	2	1	0	0
41	Ca_18721	7	1	0	0
42	Ca_19906	4	0	0	0
43	Ca_20025	5	0	0	0
44	Ca_20316	3	0	0	0
45	Ca_20318	4	0	0	0
46	Ca_20639	4	0	2	0
47	Ca_20667	2	0	0	0
48	Ca_21455	4	0	0	0
49	Ca_23751	6	0	0	0
50	Ca_24830	4	2	0	0
51	Ca_23754	5	0	0	0
52	Ca_23758	1	0	0	0
53	Ca_23345	2	0	0	0
54	Ca_23360	5	1	0	0
55	Ca_17099	2	0	0	0
56	Ca_15666	5	1	0	0
57	Ca_11732	8	0	0	0
58	Ca_10908	3	0	0	0
59	Ca_10905	4	0	0	0
60	Ca_10186	4	0	0	0
61	Ca_09772	6	1	0	0
62	Ca_00351	4	0	2	1
63	Ca_00577	4	0	0	1
64	Ca_08360	3	0	0	0

Table 6.10: List of disease resistance specific *cis*-regulatory elements present in TNL chickpea genes.

No.	Seq id	W-BOX	DRE	CBF	GCC
1	Ca_07569	3	0	0	0
2	Ca_08101	8	0	0	0
3	Ca_09385	4	0	0	0
4	Ca_09445	3	1	0	1
5	Ca_10020	8	0	0	0
6	Ca_10029	9	0	0	0
7	Ca_10030	3	0	0	0
8	Ca_10064	5	0	0	0
9	Ca_10065	4	0	0	0
10	Ca_10445	4	0	0	0
11	Ca_10449	2	0	0	1
12	Ca_13726	1	1	0	0
13	Ca_16325	3	0	0	0
14	Ca_16351	2	0	0	0
15	Ca_19533	4	1	0	0
16	Ca_23839	5	1	0	0
17	Ca_24609	8	1	0	0
18	Ca_16352	6	0	0	1
19	Ca_24238	4	0	0	0

6.14 Conclusion

The NBS disease resistance proteins constitute around 0.35% of the total chickpea proteome. Other sequenced genomes like *M. truncatula*, *P. trichocarpa*, and *A. thaliana* genomes encode quite a high number of NBS resistance genes i.e. 333, 402 and 207 each. On comparison chickpea seems to encode lesser number of NBS resistance genes. This might happen due to reduced clustered arrangement of these genes. Similar to chickpea, plants such as *B. rapa* (92), *Z. mays* (109), cucumber (57), and *Carica papaya* (78) also encode lesser number NBS resistance genes. According to Wan et al. (2013), absence of whole-genome duplication, small number of tandem gene duplication and a few segmental duplications might be the reason behind such a low number of NBS resistance genes in cucumber. Likewise fewer number of duplication events might be the reason for the reduced number of NBS resistance genes in chickpea too.

Although the number of NBS resistance genes in chickpea is quite less as compare to some of the other sequenced genomes, it has maintained both the families

TNL and non-TNL, which suggests that chickpea has few but diverse set of resistance genes. The NBS motifs also showed diversity in the two families. Six motifs (P-loop, Kinase-2, RNBS-B, RNBS-C, GLPL, and MHDL) are common to both families, whereas RNBS-A-TIR and RNBS-D-TIR are exclusive to TIR-NBS family. Similarly, RNBS-A-nonTIR and RNBS-D-nonTIR are specific to non-TIR-NBS family. Out of the two new motifs identified the first one is located in 9 protein sequences of the non-TNL family and second in 24 (not shown). Three new motifs CNBS1, CNBS2, and TNBS1 were also found in the NBS domain of chickpea NBS resistance proteins.

The primary objective of crop breeding is the development of improved disease resistant and stress resistant varieties of plants with increased nutrient values. Therefore, in an attempt to help developing stress tolerant and pathogen resistant varieties, we have identified and studied 100 putative NBS resistance genes in chickpea genome in the current research. The results reported here will provide insights into NBS resistance gene evolution leading to functional diversification of the non-TIR- and TIR-NBS-LRR R-genes in chickpea. The findings accounted in this report have direct agricultural significance and will provide a strong background for the isolation of candidate stress tolerant or pathogen resistance genes. This will aid in the development of more efficient, stress and disease resistant varieties of chickpea through molecular breeding approaches and help in improving their quality, yield and nutritional value.

Chapter 7

Identification, characterization, and tissue specific gene expression studies of more stress genes from chickpea genome

Abiotic stresses usually cause protein denaturation and dysfunction. Maintaining proteins in their functional conformations and preventing the aggregation of protein molecules with non-native structure are particularly important for cell survival under stress. As already mentioned, abiotic stress includes several environmental disasters like extreme temperature, excessive light, oxidative stress, drought, flood, salinity, wind, anaerobic condition, metal toxicity, nutrient deprivation etc. In this chapter, other stress proteins not considered previously and involved in salinity, water deficit conditions, chilling, heat stress, wounding, and pathogenesis responses were identified in chickpea genome and analyzed.

7.1 Proteins involved in other stress conditions

7.1.1 Pathogenesis related (PR) proteins

Chitinases, considered as pathogenesis related, are generally found in organisms that either needs to reshape their own chitin or dissolve and digest the chitin of fungi or animals (Sivaji *et al*, 2014). The cell walls of various fungi including plant pathogens consists of chitin and beta-1,3-, 1-6-glucan. Many chitinases and beta-1,3-glucanases are produced by plants to inhibit fungal growth by dissolving the tips of germ tubes and hyphae. Previous studies have shown that a combined action of chitinase and beta-1,3-glucanase is stronger towards a wider range of fungi than their individual action (Mauch *et al*, 1988). Thaumatin belongs to another class of PR-proteins in plants that gets induced in katemfe in response to an attack by viroid pathogens. Several members of the thaumatin protein family display significant *in vitro* inhibition of hyphal growth and sporulation by various fungi (Crammer, 2008). In cell, the primary role of lipid transfer proteins (LTPs) is to shuttle phospholipids and other fatty acid groups between the cell membranes (Kader, 1996). In addition to this, they play a major role in plant defense mechanism too (Yeats & Rose JKC, 2008).

7.1.2 Proteins in heat stress and desiccation

Another environmental stress is heat shock, which is countered by an important protein family namely heat shock proteins (HSPs). HSPs are like chaperons that prevent the stress-induced unfolding and denaturation of functional proteins (Forreiter & Nover, 1998). The late embryogenesis abundant proteins (LEA) are hydrophilic in

nature and their accumulation is tightly correlated with acquisition of desiccation tolerance. They stabilize other proteins and membranes during drying, especially in the presence of sugars like trehalose (Hand *et al*, 2011).

7.1.3 Proteins in alleviating oxidative stress and healing wound

Peroxidase, a major peroxide scavenging enzyme comes into action during oxidative burst and wounding conditions. The reactive oxygen species generated from peroxide are harmful for cell contents which are thought to also contribute in aging and pathogenesis of a variety of human diseases. Peroxidases break the peroxide molecule into harmless substances by adding hydrogen, obtained from another molecule with the result that the peroxide is reduced to form water molecule, and the molecule donating hydrogen getting oxidized.

7.2 Gene identification

Various abiotic and biotic factors play major role in chickpea productivity loss. Some of them are salinity, drought, heat, oxidative stress, wounding, and pathogenesis related protein. The proteins were identified in the similar manner using HMM profile of individual family as described above (Table 7.1).

Table 7.1: The HMM profiles of various classes of stress genes employed for the gene identification study.

Serial no.	Protein	HMM profile
1	**Chitinase**	
	Chitinase 1	PF00182
	Chitinase 2	PF00704
2	**Glucanase**	PF00332
3	**Thaumatin**	
4	**Heat shock proteins (HSPs)**	
	HSP20	PF00011
	HSP70	PF00012
	HSP90	PF00183
5	**Late embryogenesis abundant (LEA)**	
	LEA1	PF03760

	LEA2	PF03168
	LEA3	PF03242
	LEA4	PF02987
	LEA5	PF00477
	LEA6	PF10714
6	**LTPs**	PF00234
7	**Peroxidases**	PF00141
8	**Glutathione peroxidase**	PF00255

A hidden Markov model profile search resulted in the identification of total 329 stress proteins in chickpea playing role in several abiotic environmental stress conditions (Table 7.2, Figure 7.1). Out of these 329 genes, locations of 300 were known and their names were assigned based on the order of their location on chromosomes. The genomic locations of remaining 29 sequences were unknown, therefore assigned to scaffolds. The GenBank accession numbers were assigned for 64 sequences [GenBank: KJ808768-KJ808774 (Chitinase); KM052180-KM052195 (Glucanase); KM052196-KM052200 (Thaumatin); KJ808803-KJ808810 (HSPs); KJ808775-KJ808781 (LEA); KJ808782-KJ808787 (LTPs); KJ808788-KJ808802 (Peroxidase)]. The remaining sequences either had the REFSEQ accessions predicted and assigned by automated computational analysis, present on scaffolds or had ambiguous nucleotide 'N' in the gene sequence (CD-Table S-7.1).

Table 7.2: Number of stress genes identified by the HMM search in chickpea genome

No.	Protein class	Number of hits
1	Chitinase (I and II)	7, 20
2	Glucanase	51
3	Thaumatin	24
4	LTP	37
5	Heat shock protein (HSP20, HSP70, and HSP90)	27, 21, 5
6	Late embryogenesis abundant protein (LEA1, 2, 3, 4, 5, 6)	5, 45, 3, 4, 2, 2
7	Peroxidase	70, 6

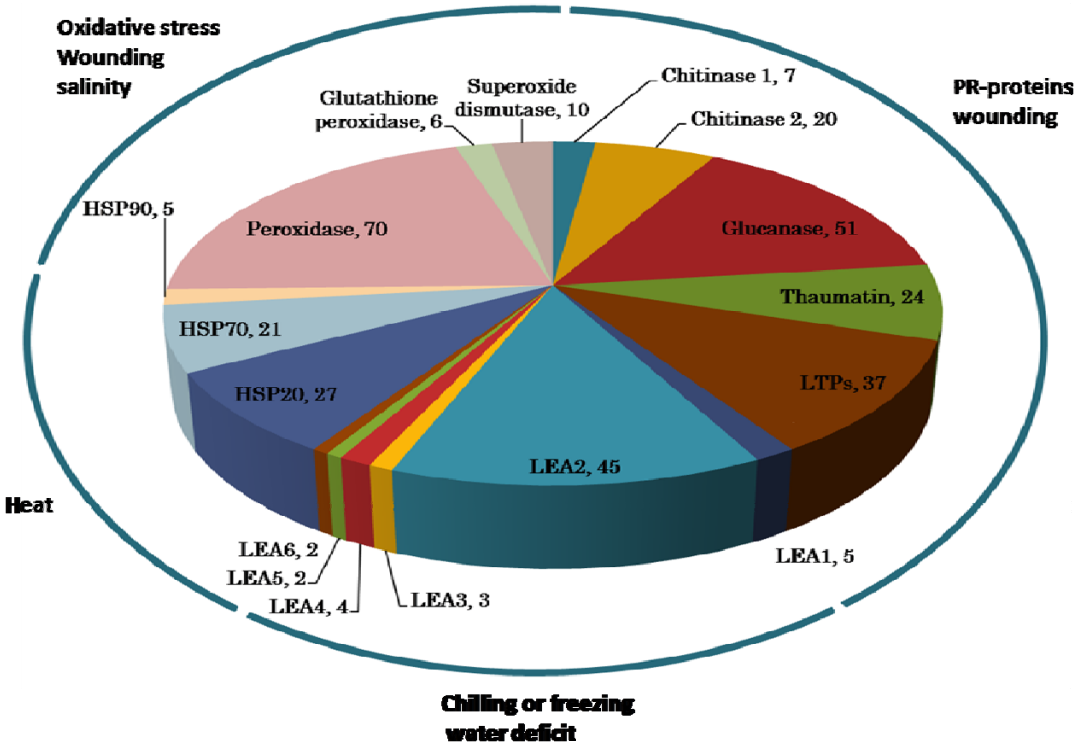


Figure 7.1: Pie chart depicting the distribution of members of different classes of additional stress genes identified in chickpea.

7.3 Identification of orthologs and divergence of genes

The six homologous dicot plant genomes of *M. truncatula*, *G. max*, *L. Japonicus*, *P. vulgaris*, *V. vinifera*, and *A. thaliana* were explored for identifying close orthologous relationship with the stress proteins considered here. Maximum number of orthologs was detected in *G. max* (496) while least number was found in *A. thaliana* (36). Out of the total 329 identified stress proteins, 281 had orthologs in at least one of the six dicot plants considered in the analysis, however, 48 of them showed large divergence (CD-Table S-7.2).

7.4 Recent gene duplication events

The identified genes were further explored to check if any recent gene duplication events have taken place by comparing sequence similarity. We could observe four recent gene duplication events in the seven different stress gene classes (Table 7.3).

The gene duplication event was considered by setting percent identity cut off to ≥ 90%.

Table 7.3: The gene pair involved in recent gene duplication events are enlisted below

Serial no.	Gene pair	% identity
1	Chitinase 2 & 3	90.2
2	Glucanase 12 & 14	90.5
3	HSP20_19 & HSP20_21	92.8
4	LEA2_5 & LEA2_6	94.1

7.5 Gene expression analysis

7.5.1 RNA-seq data

Out of the total 326 genes identified 176 were reported to show medium to high expression (FPKM >= 5) in one of the 5 tissues selected (flower, flower bud (FB), shoot apical meristem (SAM), young leaves (YL), and germinating seedling (G. seed)) whereas 44 showed low expression (5 > FPKM > 0) and 106 showed no expression (FPKM=0) in any of the five tissues under study (Figure 7.2-7.6, Table 7.4, CD-Table S-7.3). Most of the chitinases were found to express in germinating seeds and young leaves. Expression was also seen in flower tissues. Characterization and gene expression studies in *G. max*, *H. vulgare*, *Z. mays*, *T. aestivum*, *L. esculentum* and *C. melo* (Gijzen *et al*, 2001, Ramos *et al*, 1998; Krishnaveni *et al*, 1999, Swegle *et al*, 1992, Cordero *et al*, 1994; Caruso *et al*, 1999, Wu *et al*, 2001, Witmer *et al*, 2003) revealed high gene expression level of chitinases in germinating seeds. Since the germinating seeds are more prone to attack by soil pathogens, accumulation of chitinases in seeds of several species is a part of their developmental program, while others can be induced in response to microbial attack (reviewed by Gomez *et al*, 2002). Seed chitinases may protect against chitin-containing pathogenic fungi, because substrates for chitinase are found in some of the fungal cell walls, but not in plants themselves (Powning & Irzykiewicz, 1965; Graham & Sticklen, 1994;

Gomez *et al*, 2002). Interestingly the five thaumatin genes (Thaumatin 8, 12, 13, 15, and 16) express in multiple plant tissues with a significantly high FPKM values. The heat map of HSPs showed no expression for a single member of HSP20 though the members of remaining two classes differentially expressed in more than one plant tissue. The unexpressed genes are more in number for HSPs, LEA, and peroxidase but that doesn't attribute non-functionality to the gene. These genes might be expressing in other plant tissues not included in the analysis or gets activated when attacked by pests and pathogens. Our analysis on two other variety of chickpea genotype showed up-regulation as well as down-regulation of some of the stress genes under drought condition as compared to control conditions (CD-Figure S-7.1, CD-Table S-7.4).

Table 7.4: Information about the number of expressed and unexpressed genes in the five plant tissues in chickpea is given. The third and fourth column represents the number of genes with FPKM value ≥ 5 and FPKM value < 5. The last column enlists the total number of unexpressed genes in each protein class.

Serial number	Protein class	Number of gene with high to medium expression	Number of gene with low expression	Number of unexpressed gene
1	Chitinase	15	12	0
2	Glucanase	36	12	0
3	Thaumatin	20	4	0
4	HSPs	19	4	30
5	LEA	21	1	39
6	LTP	35	2	0
7	Peroxidase	30	9	37

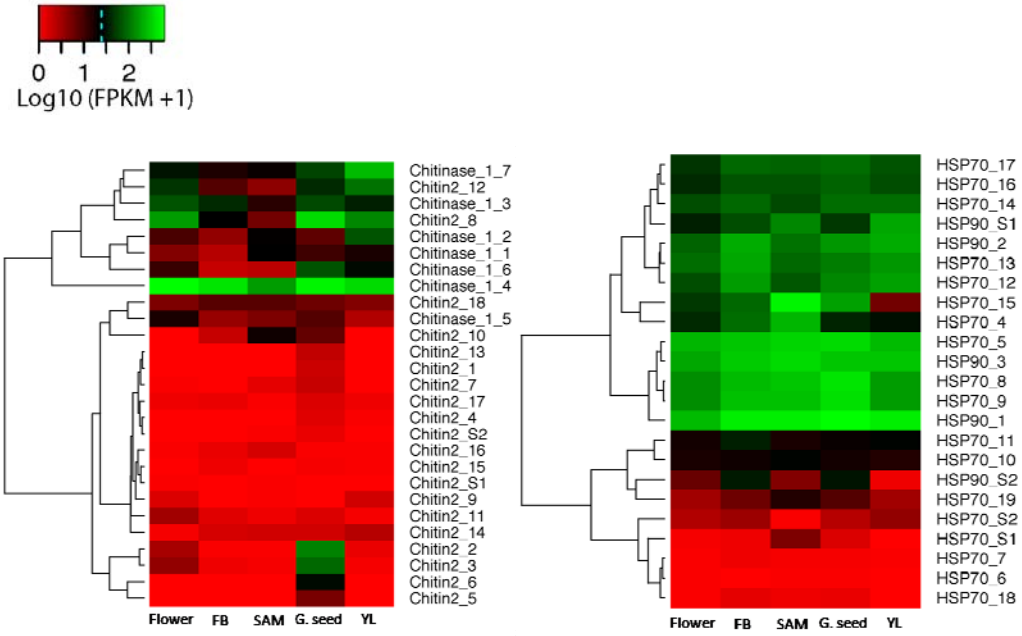


Figure 7.2: Heatmap shows relative gene expression of chickpea chitinase and HSPs genes in various tissues samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for identified genes in different tissues. The genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

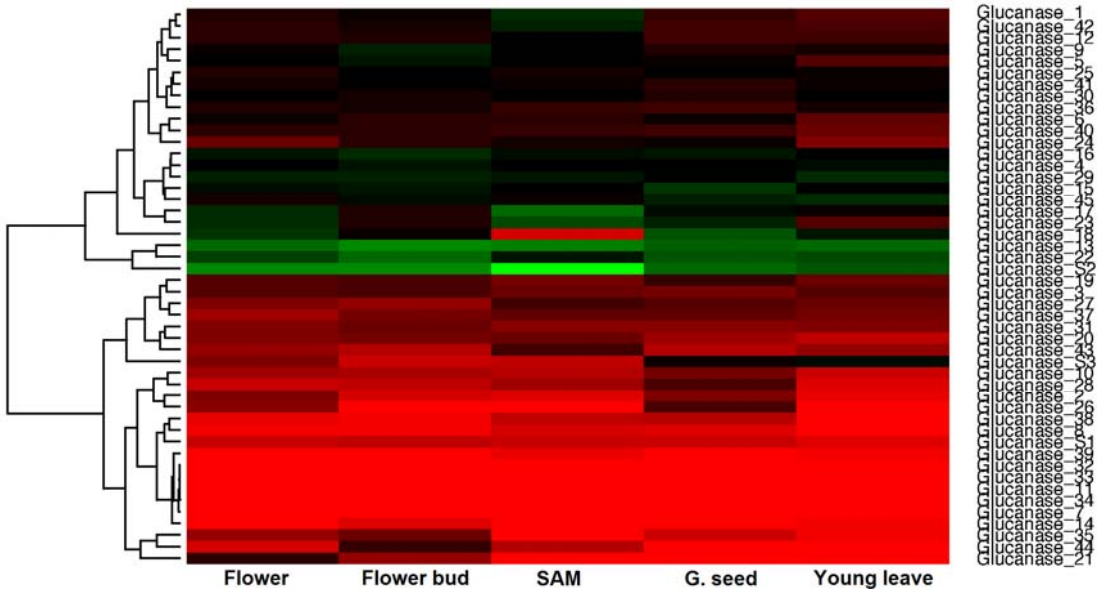


Figure 7.3: Heatmap shows relative gene expression of chickpea glucanase gene in various tissues samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for identified genes in different tissues. The genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

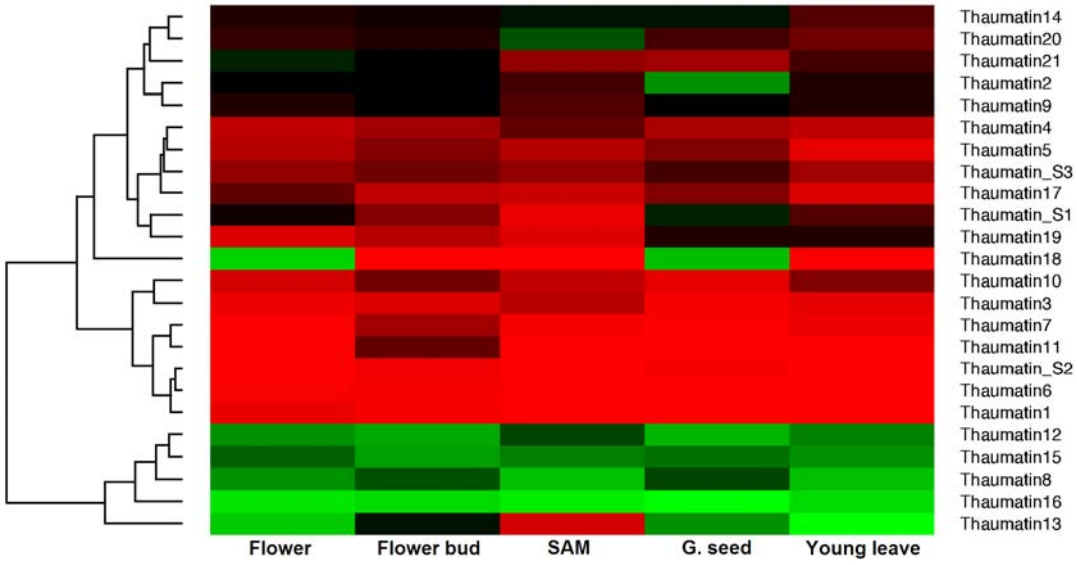


Figure 7.4: Heatmap show relative genes expression of chickpea thaumatin genes in various tissues samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for identified genes in different tissues. The genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

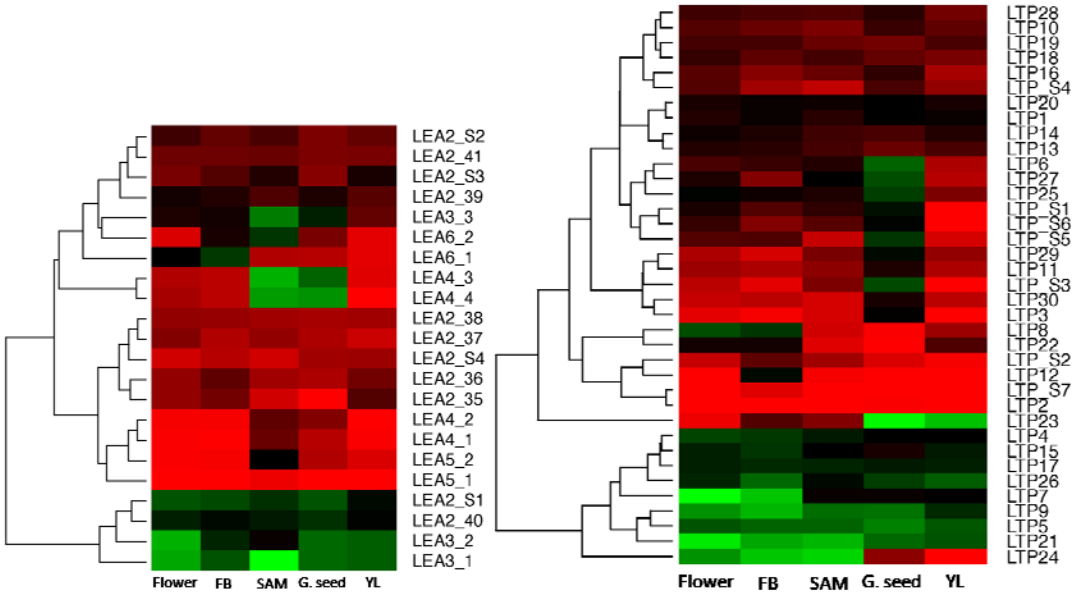


Figure 7.5: Heatmap shows relative gene expression of chickpea LEA and LTP genes in various tissues samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for identified genes in different tissues. The genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

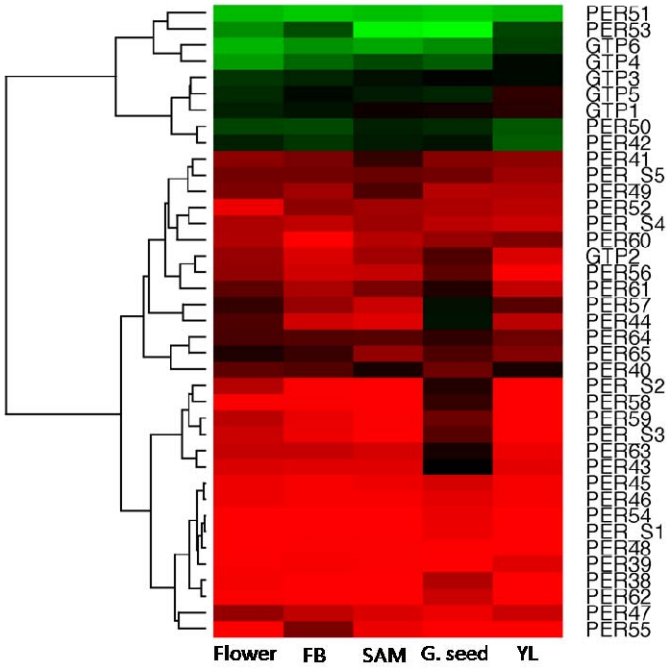


Figure 7.6: Heatmap shows relative gene expression of chickpea peroxidase genes in various tissues samples by RNA-seq data analysis. The color scale represents log (base 10) transformed FPKM values, calculated by comparing fragment kilo base transcript per million (FPKM) value for identified genes in different tissues. The genes with FPKM > 0 are included in the analysis. Dendrogram on the left side of the heatmap shows hierarchical clustering of genes using complete linkage approach.

7.5.2 EST data

The coding sequence of identified stress genes were queried against the expressed sequence tag (EST) libraries of chickpea available in NCBI to get supporting evidence on expression of predicted genes. The sequence identity constraint was kept ≥ 90% between the EST and the coding sequence. Due to lack of sufficient expression data, we could retrieve expression data for only 11 chitinase, 13 glucanase, 6 thaumatin, 23 LEA, 22 LTPs, and 26 peroxidase genes which may express in root, shoot, immature seed, root & collar, and leaf tissues (Table 7.5). However, these genes also showed expression in the RNA-seq analysis though most of the plant tissues tested was different. The gene expression was matching for specific genes when checked in the same tissues (leaf) by using both the methods. The candidate genes of LEA5 and 6 groups don't have any EST support.

Table 7.5: Result of EST search against chickpea chromosome assembly. The gene IDs with EST support and respective tissues in which the genes may express are given in the below table.

No.	Protein class	Gene ID	Tissues
1	Chitinase	Chitinase1_1, Chitinase1_4, Chitinase1_6, Chitinase1_7, Chitinase2_1, Chitinase2_2, Chitinase2_3, Chitinase2_6, Chitinase2_8, Chitinase2_12, Chitinase2_13	Leave, Root, Immature seed, Root and collar, Shoot
2	Glucanase	Glucanase 4, 5, 6, 15, 16, 21, 23, 34, 37, 39, S1, S3, S4	Leave, Root, Root and collar, Shoot
3	Thaumatin	Thaumatin 13, 14, 15, 17, 18, S1	Leave, Root, Root and collar, Shoot
4	HSPs	HSP20_2, HSP20_3, HSP20_4, HSP20_10, HSP20_12, HSP20_14, HSP20_16, HSP20_17, HSP20_18, HSP20_19, HSP20_20, HSP20_21, HSP20_22, HSP20_24, HSP20_26, HSP70_1, HSP70_2, HSP70_4, HSP70_5, HSP70_8, HSP70_9, HSP70_10, HSP70_11, HSP70_12, HSP70_13, HSP70_14, HSP70_15, HSP70_16, HSP70_17, HSP70_19, HSP90_1, HSP90_2, HSP90_3, HSP90_S1, HSP90_S2	Leave, Root, Immature seed, Root and collar, Shoot
5	LEA	LEA1_1, LEA1_2, LEA1_3, LEA2_3, LEA2_9, LEA2_11, LEA2_12, LEA2_13, LEA2_14, LEA2_21, LEA2_22, LEA2_23, LEA2_27, LEA2_28, LEA2_40, LEA2_S1, LEA2_S2, LEA3_1, LEA3_2, LEA3_3, LEA4_2, LEA4_3, LEA4_4	Leave, Root, Root and collar, Shoot
6	LTPs	LTP1, 3, 5, 6, 7, 8, 9, 14, 15, 20, 21, 24, 25, 26, 27, 28, 29, S1, S3, S4, S5, S6	Leave, Root, Root and collar, Shoot
7	Peroxidase	PER2, 3, 4, 5, 6, 7, 12, 13, 22, 27, 36, 37, 40, 42, 50, 51, 53, 54, 56, 57, 61, 62, 63, GTP4, 5, 6	Leave, Root, Root and collar, Shoot

7.6 Exon-intron position in stress genes

The exon-intron architecture of the above mentioned stress genes was explored further to analyze the intron position and number, intron phases, and untranslated regions (Table 7.6-7.12). Most of the chitinases genes were intronless, however gene structure of some of the chitinase1 family members posses one or two introns with an exception of five introns in Chitin2_18 (Table 7.6). The gene structure of glucanases family members consists of 0 to 3 introns with an exception of maximum 9 introns in Glucanase24. Six glucanase genes were intronless and two genes possess 5' and 3' UTR regions (Glucanase 14 & 34) (Table 7.7). Similar pattern of only few introns were observed in thaumatin genes also (Table 7.8). Genes of chickpea HSP20 had

very less number of introns as compared to HSP70 and HSP90 genes of chickpea (Table 7.9). Most of the members of LEA and LTP gene family were intronless although a few genes had one or two introns present (Table 7.10 & 7.11). Almost half the peroxidases genes possessed 3 introns, and only PER36 and PER42 had as many as 11 introns (Table 7.12).

Table 7.6: Exon-intron arrangement of chitinase genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
Chitinase1	1	1	Chitin2_7		
Chitinase2	0		Chitin2_8		
Chitinase3	2	1, 2	Chitin2_9		
Chitinase4	2	1, 2	Chitin2_10		
Chitinase5	1	2	Chitin2_11	2	1, 0
Chitinase6	0		Chitin2_12		
Chitinase7	2		Chitin2_13		
Chitin2_1			Chitin2_14		
Chitin2_2	1	0	Chitin2_15		
Chitin2_3			Chitin2_16		
Chitin2_4			Chitin2_17	1	2
Chitin2_5			Chitin2_18	5	0,2,0,2,0, 1,1
Chitin2_6					

Table 7.7: Exon-intron arrangement of glucanase genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
Glucanase1	1	1	Glucanase24	9	2, 0, 0, 1, 2, 0, 2, 0, 1
Glucanase2	3	1, 2, 2	Glucanase25	3	1, 1, 2
Glucanase3	1	1	Glucanase26	1	1
Glucanase4	2	0, 2	Glucanase27	2	1, 2
Glucanase5	0		Glucanase28	2	1, 1
Glucanase6	1	1	Glucanase29	2	1, 1
Glucanase7	3	1, 2, 2	Glucanase30	1	1
Glucanase8	0		Glucanase31	3	1, 1, 2
Glucanase9	2	2, 2	Glucanase32	2	1, 1
Glucanase10	2	1, 1	Glucanase33	0	
Glucanase11	0		Glucanase34	2	2, 0
Glucanase12	1	1	Glucanase35	1	2
Glucanase13	1	1	Glucanase36	0	
Glucanase14	0		Glucanase37	1	2
Glucanase15	1	0	Glucanase38	1	0
Glucanase16	1	2	Glucanase39	3	0, 0, 2
Glucanase17	3	0, 0, 2	Glucanase40	2	1, 2
Glucanase18	2	1, 1	Glucanase41	2	2, 1
Glucanase19	4	1, 0, 1, 2	Glucanase42	2	0, 2
Glucanase20	1	2	Glucanase43	1	1

Glucanase21	4	2, 0, 0, 2	Glucanase44	1	1
Glucanase22	1	2	Glucanase45	3	1, 1, 2
Glucanase23	3	1, 0, 0			

Table 7.8: Exon-intron arrangement of thaumatin genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
Thaumatin1	0		Thaumatin12	1	1
Thaumatin2	1	1	Thaumatin13	0	
Thaumatin3	1	1	Thaumatin14	2	1, 2
Thaumatin4	3	2, 1, 2	Thaumatin15	2	1, 2
Thaumatin5	1	2	Thaumatin16	0	
Thaumatin6	0		Thaumatin17	3	1, 1, 2
Thaumatin7	0		Thaumatin18	0	
Thaumatin8	2	1, 2	Thaumatin19	0	
Thaumatin9	1	1	Thaumatin20	2	1, 2
Thaumatin10	1	2	Thaumatin21	2	1, 2
Thaumatin11	1	2			

Table 7.9: Exon-intron arrangement of HSP genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
HSP20_1	1	2	HSP20_25	1	1
HSP20_2	0		HSP20_26	1	0
HSP20_3	1	2	HSP70_1	7	0, 0, 1, 1, 0, 2, 0
HSP20_4	0		HSP70_2	5	0, 2, 0, 0, 0
HSP20_5	1	1	HSP70_3	8	0, 0, 0, 0, 0, 0, 0, 2
HSP20_6	1	1	HSP70_4	1	1
HSP20_7	5	1, 1, 2, 0, 0	HSP70_5	2	2, 2
HSP20_8	0		HSP70_6	2	0, 2
HSP20_9	1	1	HSP70_7	4	1, 0, 0, 0
HSP20_10	0		HSP70_8	1	2
HSP20_11	0		HSP70_9	1	1
HSP20_12	1	1	HSP70_10	7	0, 0, 0, 0, 0, 0, 0
HSP20_13	0		HSP70_11	13	0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0
HSP20_14	0		HSP70_12	7	1, 0, 2, 2, 2, 1, 2
HSP20_15	1	1	HSP70_13	7	1, 0, 2, 2, 2, 1, 2
HSP20_16	1	1	HSP70_14	5	0, 2, 0, 0, 0
HSP20_17	0		HSP70_15	1	2
HSP20_18	0		HSP70_16	8	0, 0, 0, 0, 0, 0, 0, 2
HSP20_19	0		HSP70_17	8	0, 0, 0, 0, 0, 0, 0, 2
HSP20_20	0		HSP70_18	0	
HSP20_21	0		HSP70_19	0	
HSP20_22	1	1	HSP90_1	3	2, 2, 0
HSP20_23	0		HSP90_2	15	0, 0, 1, 2, 0, 0, 2, 2, 0, 1, 2, 0, 0, 1, 2

HSP20_24	0	HSP90_3	2	2, 0

Table 7.10: Exon-intron arrangement of LEA genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
LEA1_1	1	0	LEA2_26	0	
LEA1_2	1	0	LEA2_27	0	
LEA1_3	1	0	LEA2_28	2	0, 0
LEA2_1	0		LEA2_29	0	
LEA2_2	2	0, 0	LEA2_30	0	
LEA2_3	2	0, 0	LEA2_31	0	
LEA2_4	0		LEA2_32	2	0, 0
LEA2_5	0		LEA2_33	0	
LEA2_6	0		LEA2_34	0	
LEA2_7	0		LEA2_35	0	
LEA2_8	0		LEA2_36	2	0, 0
LEA2_9	0		LEA2_37	1	1
LEA2_10	0		LEA2_38	0	
LEA2_11	1	2	LEA2_39	2	0, 0
LEA2_12	0		LEA2_40	2	0, 0
LEA2_13	0		LEA2_41	2	0, 0
LEA2_14	0		LEA3_1	1	2
LEA2_15	0		LEA3_2	1	1
LEA2_16	0		LEA3_3	1	1
LEA2_17	0		LEA4_1	1	0
LEA2_18	0		LEA4_2	1	0
LEA2_19	0		LEA4_3	1	0
LEA2_20	0		LEA4_4	2	1, 0
LEA2_21	0		LEA5_1	1	2
LEA2_22	0		LEA5_2	1	2
LEA2_23	0		LEA6_1	0	
LEA2_24	0		LEA6_2	0	
LEA2_25	0				

Table 7.11: Exon-intron arrangement of LTP genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
LTP1	2	2, 2	LTP16	2	2, 2
LTP2	0		LTP17	1	2
LTP3	0		LTP18	2	2, 2
LTP4	0		LTP19	2	2, 2
LTP5	0		LTP20	2	2, 2
LTP6	0		LTP21	1	0
LTP7	0		LTP22	0	
LTP8	0		LTP23	0	
LTP9	0		LTP24	0	
LTP10	1	2	LTP25	0	

LTP11	2	1, 1	LTP26	0	
LTP12	0		LTP27	0	
LTP13	2	1, 1	LTP28	2	1, 1
LTP14	0		LTP29	2	1, 1
LTP15	3	2, 2, 2	LTP30	3	2, 2, 2

Table 7.12: Exon-intron arrangement of peroxidase genes in chickpea.

Protein	Introns no.	Intron phase	Protein	Introns no.	Intron phase
PER1	3	0, 0, 1	PER37	3	0, 0, 1
PER2	2	2, 0	PER38	3	0, 0, 1
PER3	3	0, 0, 1	PER39	2	0, 0
PER4	3	0, 0, 1	PER40	9	1, 2, 1, 1, 2, 0, 2, 1, 2
PER5	3	0, 0, 1	PER41	3	0, 0, 1
PER6	5	0, 0, 1, 0, 1	PER42	11	0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0
PER7	3	0, 0, 1	PER43	3	2, 0, 0
PER8	3	0, 0, 1	PER44	3	0, 0, 1
PER9	3	0, 0, 1	PER45	3	2, 0, 0
PER10	3	0, 0, 1	PER46	3	0, 0, 1
PER11	3	2, 0, 0	PER47	2	0, 0
PER12	2	0, 0	PER48	3	2, 0, 0
PER13	3	2, 0, 0	PER49	2	0, 0
PER14	2	2, 0	PER50	8	0, 1, 0, 2, 0, 0, 2, 1
PER15	1	0	PER51	7	2, 0, 0, 1, 0, 2, 0
PER16	3	0, 0, 1	PER52	2	2, 0
PER17	1	1	PER53	3	0, 0, 1
PER18	1	0	PER54	3	2, 0, 0
PER19	3	2, 0, 0	PER55	3	2, 0, 0
PER20	2	0, 0	PER56	3	0, 0, 1
PER21	2	0, 0	PER57	3	2, 0, 0
PER22	3	0, 0, 1	PER58	3	0, 0, 1
PER23	3	0, 0, 1	PER59	3	2, 0, 0
PER24	3	0, 0, 1	PER60	2	0, 0
PER25	1	0	PER61	3	2, 0, 0
PER26	3	2, 0, 0	PER62	3	2, 0, 0
PER27	1	0	PER63	3	0, 0, 1
PER28	4	0, 0, 1, 0	PER64	8	0, 1, 0, 2, 0, 0, 2, 1
PER29	2	0, 0	PER65	3	2, 0, 0
PER30	3	0, 0, 1	GTP1	4	0, 2, 1, 0
PER31	4	2, 0, 0, 2	GTP2	5	0, 0, 2, 1, 0
PER32	3	2, 0, 0	GTP3	5	0, 0, 2, 1, 0
PER33	3	2, 0, 0	GTP4	5	0, 2, 1, 0, 0
PER34	1	0	GTP5	5	0, 0, 2, 1, 0
PER35	3	0, 0, 1	GTP6	4	0, 0, 2, 1
PER36	11	2, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0			

7.7 Promoter analysis

A 2 kb upstream region of the identified stress genes was extracted and explored in detail to identify the regulatory elements. The following *cis*-regulatory elements enlisted in Table 7.13 related to stress conditions and pathogen attack were found overrepresented in the 2 kb region upstream of the stress genes. Dong *et al*, 2003 reported the importance of WBOX cassettes associated with WRKY transcription factor being overrepresented in the 2 kb region upstream of the stress and defense genes. The importance of WBOXNTERF3 *cis* elements in wounding response, MYBCORE in water stress condition, and GT1GMSCAM4 in pathogen attack and salt stress condition have been well studied in *Arabidopsis* (Li *et al*, 2012). Although another *cis*-element namely AGCBOXNPGLB is known to be conserved in most PR-protein genes we could find only one chitinase and glucanase gene associated with this box (Yamamoto *et al*, 2007).

The heat shock acts through a highly conserved *cis*-regulatory promoter element, the heat shock element (HSE) located in the TATA-box-proximal 5′-flanking regions of heat-shock genes. The occurrence of multiple HSEs within a few hundred base pairs is a signature of most eukaryotic heat-shock genes. The eukaryotic HSE consensus sequence has been ultimately defined as alternating units of 5′-nGAAn-3′. In plants the optimal HSE core consensus was shown to be 5′-aGAAg-3′ (Barros *et al*, 1992). HSEs are the binding sites for the *trans*-active HSF, and efficient binding requires at least three units, resulting in 5′-nGAAnnTTCnnGAAn-3′. The role of CCAATBOX1 in heat shock is also reported in literature (Sharma *et al*, 2011).

Most of the LEA genes have ABA responsive element (ABRE) and/or low temperature response (LTRE) elements in their promoters and gets induced by ABA, cold or drought (Hundertmark & Hincha, 2008). Other *cis*- regulatory elements namely MYB1AT, RAV1AAT, MYBCORE, DRE, and CBFHV also contribute towards transcriptional regulation of LEA genes (Yamamoto *et al*, 2007). Few peroxidases genes possess G-box element in their promoter region which is known to play an important role in oxidative stress (Smykowski *et al*, 2010). The role of peroxidases genes during wounding response is well known (Mohan *et al*, 1993; Kawaoka *et al*, 1994). As expected, WBOXNTERF3 *cis*- element was found to be

overrepresented in the 2kb upstream region of all the chickpea peroxidase genes, minimum being 1 and maximum 10, except *PER*49.

Table 7.13: List of important *cis*-regulatory elements present in the different classes of stress genes identified here.

Promoter	Promoter sequence	Response	Gene involved
WBOX	TGAC(C/T)	Pathogenesis	Chitinase, LTP, Peroxidase
WBOXNTERF3	TGACY	Wounding	Peroxidase
ASF1MOTIFCAMV	TGACG	Pathogenesis (Methyl jasmonic acid)	Chitinase, LTP
GT1GMSCAM4	GAAAAA	Pathogenesis, salt	Chitinase, LTP
GCC BOX	GCCGCC	Pathogenesis	Chitinase
AGCBOXNPGLB	AGCCGCC	Pathogenesis	Chitinase, Glucanase
CCAATBOX1	CCAAT	Heat response	HSPs
HSE	CNNGAANNTTGNNG	Heat response	HSPs
ABRE		Water deficit, salinity	LEA
MYB1AT	WAACCA	Dehydration	LEA
RAV1AAT	CAACA	Cold response	LEA
MYBCORE	CNGTTR	Water stress	LEA
LTRE	ACCGACA CCGAC CCGAAA	Low temperature	LEA
DRE	ACCGAC RCCGAC	Drought	LEA
CBFHV	RYCGAC	Dehydration	LEA
G-BOX	CACGTG	Oxidative stress	Peroxidase

7.8 Conclusion

A large number of genes involved in biotic and abiotic stress conditions were identified in chickpea genome. The stress conditions studied in this chapter are oxidative stress, heat, chilling, drought, salinity, pathogenesis etc.

Mostly the close orthologs of the identified proteins were detected in *G. max*. The exon-intron arrangement of these stress genes showed variation in the patterns of occurrence. A differential pattern of gene expression was observed in more than one plant tissue studied. The expression values showed that most of the genes expressed at basal level in normal condition though they might express when encountered by

adverse environmental conditions which was validated by analyzing RNA-seq data of the drought stressed root tissues. The expression information reported here will be useful for further investigation of the functional characterization of these genes under various stress conditions. The promoter regions of the identified stress genes were analyzed to identify the important *cis*-regulatory elements. We could identify unique *cis*-elements present in each individual class of the stress genes studied here. These studies could increase our understanding of the roles of these genes in chickpea, but further functional analysis of stress-responsive genes is required to confirm their role in stress tolerance.

Chapter 8

Conclusion

Plants, being sessile in nature, are continuously exposed to different stress factors in combination. Such unfavorable conditions affect the survival of plants adversely. Therefore they have developed specific mechanisms to detect precise environmental changes and respond in such a manner so as to minimize damage and simultaneously conserve valuable resources required for growth and reproduction. Environmental factors can be categorized into two classes: abiotic and biotic factors. Biotic factors are caused due to interactions with other organisms causing infection, mechanical damage by herbivory or trampling, as well as symbiosis or parasitism. Abiotic factors include temperature, humidity, light intensity, metal toxicity, supply of water, and many more. These factors determine the overall growth and development of plant.

Recently, in January 2013, a draft genome of chickpea was released that provided ample genomic data for the researchers and plant breeders to use with valuable tools to improve this vital crop's yield in different environments. The major constraints limiting chickpea production include abiotic and biotic stresses, out of which drought accounts for about 50% reduction in yield globally. In addition to that, heat stress and soil salinity have become other major constraints to chickpea production. *Fusarium* wilt (FW), caused by *Fusarium oxysporum* f. sp. ciceri, dry root rot caused by *Rhizoctonia bataticola*, and collar rot, caused by *Sclerotium rolfsi*, are important root diseases of chickpea which cause plant mortality. *Ascochyta* blight (AB), caused by *Ascochyta rabiei* (Pass.) Labr., and *Botrytis* grey mold (BGM) caused by *Botrytis cineria* Pres., are the important foliar diseases of chickpea. Pod borer (*Helicoverpa armigera* Hubner) is a highly polyphagous and important pest of chickpea worldwide. It can feed on various plant parts, such as leaves, tender shoots, flower buds, and immature or tender seeds. The viral diseases like rust (*Uromyces ciceris-arietini*), root nematodes (*Meloidogyne* sp.), *Phytophthora* root rot (*Phytophthora medicaginis*), cutworm (*Agrotis* sp.) and leaf miner (*Liriomyza cicerina*) are also some of the important rate limiting factors in chickpea production. Therefore there is an urgent demand for a change of focus in plant stress research, in order to understand the behavior of multiple stress responses and to create avenues for the development of plants that are resistant to multiple stresses yet maintain high

yields. Thus it is worth studying the genes involved in the plant stress and defense mechanism in the case of chickpea.

In the research reported here the high copy number genes encoding stress proteins present in chickpea genome were identified. The study involves detailed analysis of stress genes in chickpea such as glycosyltransferases, proteases, protease inhibitors, NBS-LRR genes, chitinases, glucanases, thaumatin, peroxidase, LTPs, HSPs, and LEA. Identification, genomic location, evolutionary relatedness, close orthologs in other genomes and protein as well as gene features of different classes of stress proteins were evaluated.

Chickpea being the world's third most important food legume is affected by environmental stress factors as mentioned earlier. One of the important stress gene class involved in glycosylation of secondary metabolites produced under stress condition, i.e. UGTs, were identified in chickpea and their functional annotation was performed based on their relatedness with other experimentally validated proteins. Role of glycosyltransferases in pathogen attack and abiotic stress is well studied in several plant species like *Arabidopsis* (Meurinne-Langlois *et al*, 2005), *Populus* (Babst *et al*, 2014), and Tobacco (Chong *et al*, 2002). Since, UGTs apart from being plant stress-related protein, have many applications such as in increasing the solubility of drug molecules etc. a detailed study of their structure-function in general have been carried out. As mentioned in the third and fourth chapters the glycosyltransferase family-1 members display remarkable diversity in their donor, acceptor and product specificity and thereby can generate an infinite number of glycoconjugates, oligo- and polysaccharides. Sequence and phylogenetic analysis of a large dataset of plant UGTs revealed the presence of a group of UGTs specific towards glycosylation of 3-OH group of flavonoid (F3GT), one of the important plant secondary metabolites. A considerably high degree of conservation at the N- and C-terminal domain was found through sequence and structural studies. Interaction studies with various flavonoids and UDP-sugar substrates revealed the presence of eight conserved regions exposed at the vicinity of acceptor binding. These regions are very crucial in deciding the site of glycosylation and the nature of the glycosylated product formed. The above findings were supported by molecular dynamics simulations and by performing *in-silico* experiments on experimentally validated F3GT through exploiting positive and

negative acceptor ligands as controls. The eight regions identified by us in this study will be beneficial in order to assign functions to UGTs based on their acceptor specificity.

The conserved eight regions of chickpea UGTs and other experimentally validated proteins were compared that resulted in functional assignment of 74 chickpea UGTs into the following classes: Hydroquinone glucosyltransferases, Limnoid UDP-glucosyltransferase, Scopoletin glucosyltransferases, Anthocyanidin 5,3-O-glucosyltransferase, Anthocyanidin 3-O glucoside 2"-O-glucosyltransferase, Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase, Abscisate beta-glucosyltransferase, UDP-glucose flavonoid 3-O-glucosyltransferase, Anthocyanidin 3-O-glucosyltransferase, Zeatin O-glucosyltransferase, Soyasapogenol B glucuronide galactosyltransferase, Soyasaponin III rhamnosyltransferase, and UDP-glycosyltransferase 73XX, 85XX, 76XX, 71XX, & 89XX. Similar gene architecture of chickpea UGTs and their respective orthologs revealed close relationship that also shows their probable origin from a common ancestral gene. Heat maps of gene expression values shows high degree of expression in rapidly dividing cells that reflect an indispensible role played by this important gene family in normal conditions. RNA-seq analysis of drought challenged root tissues revealed the over expression of 30 UGTs which further support the fact that they are involved in countering stress situations.

A significantly high number of protease and protease inhibitors with involvement in countering stress were identified in chickpea genome. The functional classification of proteases was carried out based on the phylogenetic clustering. Preference for specific codon of the catalytic residues was observed. Analysis of structure of genes and the domains present in the aspartate proteases shows a strong correlation with the phylogenetic analysis. Structural and molecular dynamics studies revealed the formation of more stable complex after 10 ns time scale. Kunitz-type inhibitors identified in chickpea exhibit a high degree of variation in amino acid sequence, mainly in the reactive site loop bearing the inhibitory residues. The variations in these residues and the conformation of the loop largely confer inhibitory specificity towards cognate enzymes. RNA-seq analysis of drought challenged root tissues revealed over expression of some of these proteases and PIs. NBS-LRR gene

family, one of the important *R*-gene family, was also identified in chickpea genome. The promoter regions of these genes were analyzed in detail that showed high occurrence of disease resistance-specific *cis*-regulatory elements. Such an abundant presence of *cis*-regulatory elements revealed the importance of these gene family members in chickpea survival. RNA-seq analysis showed gene expression at basal level in normal condition that might get activated when encountered an attack by pathogen or under stress condition. This was further proved upon analyzing RNA-seq data of drought challenged root tissues that revealed over expression of 24 non-TNL & 10 TNL genes. Similar analysis on other major classes of stress genes showed almost similar results.

The points to highlight on the observations recorded and conclusions drawn from this study are:

(1) UGTs have been studied as a stress-related protein in chickpea as well its function as an important enzyme with several applications.

(2) Modeling and docking studies of UGTs helped in the characterization of eight conserved regions in the acceptor binding vicinity that decides the flavonoid glycosylation preference.

(3) Molecular dynamics simulation has demonstrated and has further supported the importance of above eight regions identified in enzyme function.

(4) Structural studies of stress proteins showed high binding affinity for substrates and formation of more stable control complexes.

(5) Comparative genomics showed close orthologous relationship of chickpea stress proteins with corresponding ones in *M. truncatula* and *G. max*.

(6) Promoter analysis of stress genes showed over-representation of stress-specific *cis*-elements.

(7) Gene expression studies have shown that some of the genes identified are expressed at basal level in normal conditions and get activated on encountering stress such as drought.

(8) The gene expression profile of drought stressed root samples revealed induced expression of some of the members, which we have identified as belonging to stress gene families.

(9) The results obtained in this study have helped us to conclude that the importance of stress genes in chickpea is revealed by the presence of an ample number of copies of such genes in genome.

(10) It may be noted that the number of copies of genes coding protease inhibitors is few, which also may be its bane by succumbing to attack by certain biotic agents.

(11) The findings will help in selecting genes, understanding expression profiles, and their effective manipulation to meet the stress conditions by chickpea.

Developing disease resistant and stress tolerant varieties of plants is an important objective of breeding crops. The studies reported here, thus, take us closer towards identifying and understanding the role of the stress and defense genes in order to help developing stress tolerant and pathogen resistant varieties. A large fraction of chickpea genome encodes such vital stress-responsive genes which provide resistance against various environmental stress factors. A differential pattern of gene expression was observed in more than one plant tissue under study although at basal level in normal condition. However, they may express upon encountered by adverse environmental conditions. The expression information reported here will be useful for further investigation of the functional characterization of these genes under various stress conditions. The findings reported here could increase our understanding about the roles of these genes in chickpea, but their further expression and functional analysis is required to confirm their role in stress tolerance.

BIBLIOGRAPHY

1. Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE (1998) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. Proc Natl Acad Sci USA 95:10306-10311.
2. Abadía J, Vázquez S, Rellán-Álvarez R, El Jendoubi H, Abadía A, Álvarez-Fernandez, López-Millán AF (2011) Towards a knowledge-based correction of iron chlorosis. Plant Physiol Biochem 49: 471-482.
3. Abiko M, Akibayashi K, Sakata T, Kimura M, Kihara M, Itoh K, Asamizu E, Sato S, Takahashi H, Higashitani A (2005) High-temperature induction of male sterility during barley (*Hordeum vulgare* L.) anther development is mediated by transcriptional inhibition. Sex Plant Reprod 18: 91-100.
4. Acharjee S, Sarmah BK (2013) Biotechnologically generating 'super chickpea' for food and nutritional security. Plant Science 207: 108-116.
5. Acharya BR, Assmann SM (2009) Hormone interactions in stomatal function. Plant Mol Biol 69: 451-462.
6. Agarwal S, Grover A (2006) Molecular biology, biotechnology and genomics of flooding-associated low O2 stress response in plants. Critical Reviews in Plant Sciences 25: 1-21.
7. Ahuja I, de Vos RC, Bones AM, Hall RD (2010) Plant molecular stress responses face climate change. Trends Plant Sci. 15: 664-674.
8. Al-Khatib K, Paulsen GM (1990) Photosynthesis and productivity during high temperature stress of wheat genotypes from major world regions. Crop Sci 30: 1127-1132.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.
10. Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu Hongyan, Roe B, Young ND, Cannon SB (2008) Identification and Characterization of Nucleotide-Binding Site-Leucine-Rich Repeat Genes in the Model Plant *Medicago truncatula*. Plant Physiol 146: 5-21.
11. Amtmann A, Troufflard S, Armengaud P (2008) The effect of potassium nutrition on pest and disease resistance in plants. Physiol Plant 133: 682-691.

12. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol 60: 685-699.
13. Arend J, Warzecha H, Stoeckigt J (2000) Hydroquinone:O-glucosyltransferase from cultivated Rauvolfia cells: enrichment and partial amino acid sequences Phytochemistry 53: 187-193
14. Artlip TS, Funkhouser EA (1995) Protein synthetic responses to environmental stresses. In: M. Pessarakli, ed. Handbook of Plant and Crop Physiology. New York: Marcel Dekker, 627-644.
15. Askari H, Edqvist J, Hajheidari M, Kafi M, Salekdeh GH (2006) Effects of salinity levels on proteome of *Suaeda aegyptiaca* leaves. Proteomics 6: 2542-2554.
16. Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63: 3523-3543.
17. Ausubel FM (2005) Are innate immune signalling pathways in plants and animals conserved? Nat Immunol 6: 973-979.
18. Babst BA, Chen HY, Wang HQ, Payyavula RS, Thomas TP, Harding SA, Tsai CJ (2014) Stress-responsive hydroxycinnamate Glycosyltransferase modulates phenylpropanoid metabolism in Populus. J Exp Bot 65: 4191-200.
19. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202-W208.
20. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28-36.
21. Bailey TL, Gribskov M (1998) Combining evidence using p-value: application to sequence homology searches. Bioinformatics 14: 48-54.
22. Bailey-Serres J, Freeling M (1990) Hypoxic stress induced changes in ribosomes of maize seedling roots. Plant Physiology 94: 1237-1243.
23. Bailey-Serres J, Voesenek LA (2008) Flooding stress: acclimations and genetic diversity. Annu Rev Plant Biol 59: 313-339.
24. Bale JS, Masters GJ, Hodkinson ID, Awmack C, Bezemer TM, Brown VK, Butterfield J, Buse A, Coulson JC, Farrar J, Good JEG, Harrington R, Hartley

S, Jones TH, Lindroth RL, Press MC, Symrnioudis I, Watt AD, Whittaker JB (2002) Herbivory in global climate change research: direct effects of rising temperatures on insect herbivores. Global Change Biology 8: 1-16.

25. Barros MD, Czarnecka E, Gurley WB (1992) Mutational analysis of a plant heat shock element. Plant Mol Biol 19: 665-675.
26. Bartels D, Hussain SS. Current status and implications of engineering drought tolerance in plants using transgenic approaches (2008) CAB Reviews: Perspect Agri, Vet Sci, Nutri Nat Resour 3: 020.
27. Bartoli CG, CasalonguéCA, Simontacchi M (2012) Interactions between hormone and redox signalling pathways in the control of growth and cross tolerance to stress. Environment and Experimental Botany. 52: 139-147.
28. Benedict C, Geisler M, Trygg J, Huner N, Hurry V (2006) Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signaling pathway in Arabidopsis. Plant Physiol 141: 1219-1232.
29. Berendsen HJC, Postma JPM, Gunsteren van WF, DiNola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81:3684-3690.
30. Berendsen HJC, Spoel D Van der, Drunen R Van (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Comp Phys Comm 91:43-56.
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235-242.
32. Berry J, Bjorkman O (1980) Photosynthesis response and adaptation to temperature in higher plants. Annu Rev Plant Physiol 31: 491-543.
33. Bethune MT, Strop P, Tang Y, Sollid LM, Khosla C (2006) Heterologous expression, purification, refolding, and structural-functional characterization of EP-B2, a self-activating barley cysteine endoprotease. Chem Biol 13: 637-647.
34. Birker D, Heidrich K, Takahara H, Narusaka M, Deslandes L, Narusaka Y, Reymond M, Parker JE, O'Connell R (2009) A locus conferring resistance

to *Colletotrichum higginsianum* is shared by four geographically distinct Arabidopsis accessions Plant J 60: 602–613.

35. Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: Integrated platform of small molecules and biological activities. Annual Reports in Computational Chemistry 4: 217-241.
36. Boston RS, Viitanen PV, Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32: 191-222.
37. Bowles DJ (1990) Defense-related proteins in higher plants. Annu Rev Biochem 59: 873-907.
38. Breton C, Bettler E, Joziasse DH, Geremia RA, Imberty A (1998) Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. J Biochem 123: 1000-1009.
39. Buchner J (1999) Hsp90 & Co. – a holding for folding. Trends Biochem Sci 24: 136-141.
40. Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92: 351-366.
41. Cabello F, Jorrin JV, Tena M (1994) Chitinase and beta-1, 3-glucanase activities in chickpea (*Cicer arietinum*)- Induction of different isoenzymes in response to wounding and ethephon. Physiol Plant 92: 654-660.
42. Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationship within the TIR and non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol 54: 548-562.
43. Cao PJ, Bartley LE, Jung KH, Ronald PC (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. Mol Plant 1: 858-77.
44. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaption of plants to life on land. Plant J 69: 1030-42.
45. Caruso G, Cavaliere C, Guarino C, Gubbiotti R, Foglia P, Lagana A (2008) Identification of changes in Triticum. durum L. leaf proteome in response to

salt stress by two-dimensional electrophoresis and MALDI-TOF mass spectrometry. Anal Bioanal Chem 391: 381-390.

46. Chaturvedi P, Mishra M, Akhtar N, Gupta P, Mishra P, Tuli R (2012) Sterol glycosyltransferases-identification of members of gene family and their role in stress in *Withania somnifera*. Mol Biol Rep 39: 9755-64.
47. Chaves MM, Oliveira MM (2004) Mechanisms underlying plant resilience to water deficits: prospects for water-saving agriculture. J Exp Bot. 55: 2365-2384.
48. Cheng Y, Li X, Jiang H, Ma Wei, Miao W, Yamada T, Zhang M (2012) Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize. FEBS Journal 279:2431-2443
49. Chérif M, Arfaoui A, Rhaiem A (2007) Phenolic Compounds and their Role in Bio-control and Resistance of chickpea to Fungal Pathogenic Attacks. Tunisian Journal of Plant Protection 2: 7-21
50. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev 17: 1043-1054.
51. Chong J, Baltz R, Schmitt C, Beffa R, Fritig B, Saindrenan P (2002) Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferases reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. Plant Cell 14: 1093-107.
52. Clarke S, Mur LAJ, Wood JE, Scott IM (2004) Salicylic acid dependent signalling promotes basal thermotolerance but is not essential for acquired thermotolerance in *Arabidopsis thaliana*. Plant J 38: 432-447.
53. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2:1511-1519.
54. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.
55. Crammer B (2008) Selected topics in the chemistry of natural products. Chapter 7: Recent trends of some natural sweet substances from plants. World Scientific 189-208.

56. Cui J, Zhou Y, Ding JG (2011) Role of nitric oxide in hydrogen peroxide-dependent induction of abiotic stress tolerance by brassinosteroids in cucumber. Plant Cell Environ. 34: 347-358.
57. Dalal P, Knickelbein J, Haymet ADJ, Sönnichsen FD, Madura JD (2001) Hydrogen bond analysis of Type 1 antifreeze protein in water and the ice-water interface. Phys Chem Comm 7: 1-5.
58. Darden T, York D, Pedersen LG (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J Chem Phys 98:10089-10092.
59. Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu Jk, Bohnert HJ, Cheeseman JM (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nat. Genet 43: 913-918.
60. Delaney TP (1997) Genetic dissection of acquired resistance to disease. 113: 5-12.
61. Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617-625.
62. DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7:1243-1249.
63. Didierjean L, Frendo P, Nasser W, Genot G, Marivet J, Burkard G (1996) Heavy-metal-responsive genes in maize: identification and comparison of their expression upon various forms of abiotic stress. Planta 199: 1-8.
64. Divi UK, Rahman T, Krishna P (2010) Brassinosteroid-mediated stress tolerance in Arabidopsis shows interactions with abscisic acid, ethylene and salicylic acid pathways. BMC Plant Biol 10: 151.
65. Dong J, Chen C, Chen Z (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol 51:21-37.
66. Dong X (1998) SA, JA, ethylene, and disease resistance in plants. Curr Opin Plant Biol 1: 316-323.
67. Dubey RS (1999) Protein synthetic by Plants Under Stressful Conditions. In: M. Pessarakli, ed. Handbook of Plant and Crop Stress. New York: Marcel Dekker, 365-398.

68. Dupuis I, Dumas C (1990) Influence of temperature stress on maize (*in vitro*) fertilization and heat shock protein synthesis in maize (*Zea mays* L.) reproductive tissues. Plant Physiol 94: 665-670
69. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755-763.
70. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792:97.
71. Eitas TK, Dangl JL (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr. Opin. Plant Biol. 13: 472–477.
72. Flandez-Galvez H, Ades PK, Ford R, Pang ECK, Tayler PWJ (2003) QTL analysis for *Ascochyta* blight resistance in an intraspecific population of chickpea (*Cicer arietinum* L.). Theor Appl Genet 107:1257–1265.
73. Florian J, Leighton P, Graham JE, Katrin M, Peter JAC, Frank W, Sanjeev KS, Dan B, Glenn B, Jonathan DGJ, Ingo H (2012) Identification and localization of the NB-LRR gene family within the potato genome. BMC Genomics 13:75.
74. Flowers TJ, Gaur PM, Gowda CLL, Krishnamurthy l, Samineni S, Siddique KHM, Turner NC, Vadez V, Varshney RK, Colmer TD (2010) Salt sensitivity in chickpea. 33: 490-509.
75. 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 glucosylate anthocyanidins. in vivo J Biol Chem 273: 9224-9233.
76. Forreiter C, Nover L (1998) Heat induced stress proteins and the concept of molecular chaperones. J Biosci 23: 287-302.
77. Freeman M (2003) Rhomboids. Curr Biol 13: R586.
78. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shaw DE, Shelley M, Perry JK, Francis P, Shenkin PS (2004) "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy" J Med Chem 47:1739–1749.
79. Frydman J (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem 70: 603-647.

80. Fukuda H (1996) Xylogenesis: initiation, progression, and cell death. Annu Rev Plant Physiol. Plant Mol Biol 47: 299-325.
81. Fülöp V, Böcskei Z, Polgár L (1998) Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. Cell 94: 161–170.
82. Furtek D, Schiefelbein JW, Johnston F, Nelson OE Jr. (1988) Sequence comparisons of 3 wild-type bronze-1 alleles from *Zea mays*. Plant Mol Biol 11: 473-481.
83. Gachon C, Baltz R, Saindrenan P (2004) Over-expression of a scopoletin glucosyltransferase in *Nicotiana tabacum* leads to precocious lesion formation during the hypersensitive response to tobacco mosaic virus but does not affect virus resistance Plant Mol Biol 54: 137-146
84. Galleni M, Lamotte-Brasseur J, Raquet X, Dubus A, Monnaie D, Knox JR, Frere JM (1995) The enigmatic catalytic mechanism of active-site serine beta-lactamases. Biochem Pharmacol 49: 1171-1178.
85. Garcia-Olmedo F, Salcedo G, Sanchez-Monge R, Gomez L, Royo J, Carbonero P (1987) Plant proteinaceous inhibitors of proteinases and α-amylases. Oxf Surv Plant Mol. Cell Biol 4: 275–334.
86. Ghanem ME, Hichri I, Smigocki AC, Albacete A (2011) Root-targeted biotechnology to mediate hormonal signalling and improve crop stress tolerance. Plant Cell Rep 30: 807-23.
87. Gillespie KM, Rogers A, Ainsworth EA (2011) Growth at elevated ozone or elevated carbon dioxide concentration alters antioxidant capacity and response to acute oxidative stress in soybean (*Glycine max*) 62: 2667-2678.
88. Glick RE, Schlagnhaufer CD, Arteca RN, Pell EJ (1995) Ozone-induced ethylene emission accelerates the loss of ribulose-1,5-Bisphosphate Carboxylase/Oxygenase and nuclear-encoded mRNAs in senescing potato leaves. Plant Physiol 109: 891-898.
89. Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, Dangl JL, Grant SR (2008) The Pseudomonas syringae type III effector HopAM1 enhances virulence on water-stressed plants. Mol Plant Microbe Interact 21: 361-370.
90. Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC, Bouwmeester H, Becard G,

Beveridge CA, Rameau C, Rochange SF (2008) Strigolactone inhibition of shoot branching. Nature 455: 189-194.

91. Gong Q, Li P, Ma S, Rupassara SI, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile Thellungiella halophila in comparison with its relative Arabidopsis thaliana. Plant J 44: 826-839.
92. Gong Z, Yamazaki M, Sugiyama M, Tanaka Y, Saito K (1997) Cloning and molecular analysis of structural genes involved in anthocyanin biosynthesis and expressed in a forma-specific manner in *Perilla frutescens*. Plant Mol Biol 35: 915-927.
93. Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol 313: 903-19.
94. Greenberg JT (1996) Programmed cell death: A way of life for plants. Proc Natl Acad Sci USA 93: 12094–12097.
95. Gregersen L, Christensen AB, Sommer-Knudsen J, Collinge DB (1994) A putative O-methyltransferase from barley is induced by fungal pathogens and UV light. Plant Mol Biol 26: 1797-806.
96. Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J, Schwab W (2008) Multi-substrate flavonol O-glucosyltransferases from strawberry (Fragaria x ananassa) achene and receptacle. J Exp Bot 59: 2611-2625.
97. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithm and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.
98. Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. Yi Chuan 29: 1023–1026. [http://gsds.cbi.pku.edu.cn]
99. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G (2010) Probing the alpha-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. Chem Biol Drug Des 75: 348-359.
100. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW (2012) Plant disease resistance genes: Current status and future directions. Physiol Mol Plant P 78:51-65.

101. Habib H, Fazili KM (2007) Plant protease inhibitors: a defense strategy in plants. Biotechnology and Molecular Biology 2: 068-085.
102. Hand SC, Menze MA, Toner M, Boswell L, Moore D (2011) LEA proteins during water stress: not just for plants anymore. 73: 115-34.
103. Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381: 571-580.
104. Hasanuzzaman M, Nahar K, Alam MM, Roychowdhury R, Fujita M (2013) Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. Int J Mol Sci 14: 9643-9684.
105. Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. Annu Rev Plant Physiol Plant Mol Biol 51: 463-499.
106. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comp Chem 18:1463-1472.
107. Higo K, Ugawa Y, Iwamoto M, Higo H (1998) PLACE: A database of plant cis-acting regulatory DNA elements. Nucleic Acids Res 26: 358-359.
108. Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58: 2369-2387.
109. Hiromoto T, Honjo E, Tamada T, Noda N, Kazuma K, Suzuki M, Kuroki R (2013) Crystal structure of UDP-glucose: anthocyanidin 3-O-glucosyltransfease from Clitoria ternatea. J Synchrotron Radiat 20: 894-898.
110. Hirotani M, Kuroda R, Suzuki H, Yoshikawa T (2000) Cloning and expression of UDP-glucose:flavonoid 7-O-glucosyltransferase from hairy root cultures of *Scutellaria baicalensis*. Planta 210: 1006-13.
111. Hodel AE, Hodel MR, Griffis ER, Hennig KA, Ratner GA, Xu S, Powers MA (2002) The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. Mol Cell 10: 347-358.
112. Hossain MA, Piyatida P, Teixeira da Silva JA, Fujita M (2011) Molecular mechanism of heavy metal toxicity and tolerance in plants: central role of glutathione in detoxification of reactive oxygen species and methylglyoxal and in heavy metal chelation. Journal of Botany doi: 10.1155/2012/872875.

113. Hou B, Lim E-K, Higgins GS, Bowles DJ (2004) N-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana* J Biol Chem 279: 47822-47832

114. Hsiao TC (1973) Plant responses to water stress. Annu Rev Plant Physiol 24: 519-570.

115. Huang GT, Ma SL, Bai LP, Zhang L, Ma H, Jia P, Liu J, Zhong M, Guo ZF (2012) Signal transduction during cold, salt, and drought stresses. 39: 969-987.

116. 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.

117. Humphrey W, Dalke A, Schulten K (1996) "VMD – Visual Molecular Dynamics". J Mol Graph 14:33-38.

118. Hundertmark M, Hincha DK (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9: 118.

119. Hwang SY, VanToai TT (1991) Abscisic acid induces anaerobiosis tolerance in corn. Plant Physiol 97: 593-597.

120. Ikegami A, Akagi T, Potter D, Yamada M, Sato A, Yonemori K, Kitajima A, Inoue K (2009) Molecular identification of 1-Cys peroxiredoxin and anthocyanidin/flavonol 3-*O*-galactosyltransferase from proanthocyanidin-rich young fruits of persimmon (*Diospyros kaki* Thunb.). Planta 230: 841-855.

121. Impact, version 5.5, Schrödinger, LLC, New York, NY, 2005.

122. Inan G, Zhang Q, Li P, Wang Z, Cao Z, Zhang H, Zhang C, Quist TM, Goodwin SM, Zhu J, Shi H, Damsz B, Charbaji T, Gong Q, Ma S, Fredricksen M, Galbraith DW, Jenks MA, Rhodes D, Hasegawa PM, Bohnert HJ, Joly RJ, Bressan RA, Zhu JK (2004) Salt cress. A halophyte and cryophyte Arabidopsis relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. 135: 1718-1737.

123. IPCC. 2007. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the

Intergovernmental Panel on Climate Change. Cambridge, UK & New York, NY, USA: Cambridge University Press.

124. Irie K, Hosoyama H, Takeuchi T, Iwabuchi K, Watanabe H, Abe M, Abe K, Arai S (1996) Transgenic rice established to express corn cystatin exhibits strong inhibitory activity against insect gut proteinases. Plant Mol Biol 30: 149–157.
125. Izuhara K, Kanaji S, Arima K, Ohta S, Shiraishi H (2008) Involvement of cysteine protease inhibitors in the defense mechanism against parasites. *Med Chem* 4: 322-7.
126. James C, Global Status of Commercialized Biotech/GM Crops, 2010 http://www.isaaa.org/resource/publications/briefs/ [accessed 19.03.11].
127. Jellouli N, Ben Jouira H, Skouri H, Ghorbel A, Gourgouri A, Mliki A (2008) Proteomic analysis of Tunisian grapevine cultivar Razegui under stress. J Plant Physiol 165: 471-481.
128. Jian R, Longping W, Xinjiao G, Changjiang J, Yu X and Xuebiao Y (2009) DOG 1.0: Illustrator of Protein Domain Structures. Cell Research 19: 271–273.
129. Jones JDG and Dangl JL (2006) The plant immune system. Nature 444: 323-329.
130. Jones P, Vogt T (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213:164-174.
131. Jörg S, Frank M, Peer B, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci U S A.* 95: 5857-64.
132. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110:1657-1666.
133. Jukanti AK, Gaur PM, Gowda CLL, Chibbar RN (2012) Nutritional quality and health benefits in chickpea (Cicer arietinum L.): a review. Br J Nutr 108: S11-S26.
134. Jung JL, Maurel S, Fritig B, Hahne G (1995) Different pathogenesis related proteins are expressed in sunflower (*Helianthus annus* L.) in response to physical, chemical and stress factors. J Plant Physiol 145: 153-160.

135. Kader JC (1996) Lipid-transfer proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 628-646.

136. Kale SM, Pardeshi VC, Barvkar VT, Gupta VS, Kadoo NY (2013) Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure. Genome 56:91-99

137. Kanadaswami C, Lee LT, Lee PPH, Hwang JJ, Ke FC, Huang YT, Lee MT (2005) The Antitumor Activities of Flavonoids. in vivo 19: 895-910.

138. Kant S, Kant P, Raveh E, Barak S (2006) Evidence that differential gene expression between the halophyte, *Thellungiella halophila*, and *Arabidopsis thaliana* is responsible for higher levels of the compatible osmolyte proline and tight control of Na^+ uptake in *T. halophila*. Plant Cell Environ 29: 1220-1234.

139. Katz A, Waride P, Shevchenko A, Pick U (2007) Salt-induced changes in the plasma membrane proteome of the halotolerant alga Dunaliella. Salina as revealed by blue native gel electrophoresis and nano-LC-MS/MS analysis. Mol Cell Proteomics 6: 1459-1472.

140. Kawaoka A, Kawamoto T, Ohta H, Sekine M, Takano M, Shinmyo A (1994) Wound-induced expression of horseradish peroxidase. Plant Cell Reports 13: 149-154.

141. Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. 18: 3947-55.

142. Kita M, Hirata Y, Moriguchi T, Endo-Inagaki T, Matsumoto R, Hasegawa S, Suhayda CG, Omura M (2000) Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in Citrus FEBS Lett 469: 173-178

143. Kobe B, Deisenhofer J (1995) A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374:183-186.

144. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F (2008) Genome-wide identification of NBS resistance genes in *Populas trichocarpa*. Plant Mol Biol 66:619-636.

145. Komatsu S, Yang G, Khan M, Onodera H, Toki S, Yamaguchi M (2007) Over-expression of calcium-dependent protein kinase 13 and calreticulin interacting protein 1 confers cold tolerance on rice plants. Mol Genet Genomics 277: 713-723.
146. Koo AJK, Howe GA (2009) The wound hormone jasmonate. Phyochemistry 70: 1571-1580.
147. Kosová K, Vítámvás P, Prášil IT, Renaut J (2011) Plant proteome changes under abiotic stress-contribution of proteomics studies to understanding plant stress response. J. Proteomics 74: 1301-1322.
148. Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf KD (2007) Complexity of the heat stress response in plants. 10: 310-316.
149. Kramer PJ (1980) Drought, stress, and the origin of adaptations. Adaptations of plants to water and high temperature stress. pp. 7-20. John-Wiley & Sons, New York.
150. Kroon J, Souer E, de Graaff A, Xue Y, Mol J (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.
151. (A) Larkindale J, Mishkind M, Vierling E (2005) Plant responses to high temperature. In Plant Abiotic Stress. Blackwell Publishing 100-144.
152. (B) Larkindale J, Hall JD, Knight MR, Vierling E (2005) Heat stress phenotypes of Arabidopsis mutants implicate multiple signalling pathways in the acquisition of thermotolerance. Plant Physiol 138: 882-897.
153. Larkindale J, Huang B (2004) Thermotolerance and antioxidant systems in *Agrostis stolonifera*: involvement of salicylic acid, abscisic acid, calcium, hydrogen peroxide, and ethylene. J Plant Physiol 161: 405-413.
154. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291.
155. Law RD, Crafts-Brandner SJ (1999) Inhibition and acclimation of photosynthesis to heat stress is closely correlated with activation of ribulose-1, 5-bisphosphate carboxylase/oxygenase. Plant Physiol 120: 173-181.
156. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol 25: 1307-20.

157. Lee BH, Lee H, Xiong L, Zhu JK (2002) A mitochondrial complex I defect impairs cold-regulated nuclear gene expression. Plant Cell 14: 1235-1251.
158. Leister RT, Katagiri F (2000) A resistance gene product of the nucleotide binding site-leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant J 22: 345-354.
159. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40: D302-D305.
160. Levitt J (1972) Responses of plants to environmental stresses. Academic Press, New York.
161. Li J, Farmer AD, Lindquist IE, Dukowic-Schulze S, Mudge J, Li T, Retzel EF, Chen C (2012) Characterization of a set of novel meiotically-active promoters in *Arabidopsis.* BMC Plant Biology 12: 104.
162. Li J, Lin X, Chen A, Peterson T, Ma K, Bertzky M, Ciais P, Kapos V, Peng C, Poulter B (2013) Global priority conservation areas in the face of 21st century climate change. PLoS One 8: e54839.
163. Li Y, Baldauf S, Lim EK, Bowles DJ (2000) Phylogenetic Analysis of the UDP-glycosyltransferase Multigene Family of *Arabidopsis thaliana*. J Biol Chem 276: 4338-4343.
164. Licausi F (2010) Regulation of the molecular response to oxygen limitations in plants. New Phytol 190: 550-555.
165. Liska AJ, Shevchenko A, Pick U, Katz A (2004) Enhanced photosynthesis and redox energy production contribute to salinity tolerance in Dunaliella as revealed by homology-based proteomics. Plant Physiol 136: 2806-2817.
166. Liu HT, Liu YY, Pan QH, Yang HR, Zhan JC, Huang WD (2006) Novel interrelationship between salicylic acid, abscisic acid, and PIP2-specific phospholipase C in heat acclimation-induced thermotolerance in pea leaves. J Exp Bot 57: 3337-3347.
167. Liu HT, Sun DY, Zhou RG (2005) Ca2+ and AtCaM3 are involved in the expression of heat shock protein gene in *Arabidopsis*. Plant Cell Environ 28: 1276-1284.
168. Llorente F, Oliveros JC, Martínez-Zapater JM, Salinas J (2000) A freezing-sensitive mutant of Arabidopsis, frs1, is a new aba3 allele. Planta 211: 648-655.

169. Louveau T, Leitao C, Green S, Hamiaux C, Rest B et al. (2011) Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit. FEBS J 278: 390-400.

170. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G (2012) Genome-Wide Identification and Mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. PLoS ONE 7(4):e34775.doi10.1371/journal.pone.0034775

171. Luck J, Spackman M, Freeman A, Tre¸bicki P, Griffiths W, Finlay K, Chakraborty S (2011) Climate change and disease of food crops. Plant Pathology 60: 113-121.

172. Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG (2000) Regions outside of the leucine-rich repeats of Flax rust resistance proteins play a role in specificity determination. Plant Cell 12:1367-1377.

173. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85.

174. Mackerness SAH, Liu L, Thomas B, Thompson WF, Jordan BR, White MJ (1998) Individual members of the light-harvesting complex II chlorophyll a/b-binding protein gene family in pea (*Pisum sativum*) show differential responses to ultraviolet-B radiation. Physiologia Plantarum 103: 377-384.

175. Madgwick J, West J, White R, Semenov M, Townsend J, Turner J, Fitt BL (2011) Impacts of climate change on wheat anthesis and fusarium ear blight in the UK. Eur J Plant Pathol 130: 117-131.

176. Maekawa T, Cheng W, Spiridon LN, Toller A, Lukasik E, Saijo Y, Liu P, Shen QH, Micluta MA, Somssich IE, Takken FL, Petrescu AJ, Chai J () Coiled-coil domain-dependent homodimerzation of intracellular MLA immune receptors defines a minimal functional module for triggering cell death. Cell Host Microbe 9: 187-199.

177. Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys 444: 139-158.

178. Marino M, Braun L, Cossart P, Ghosh P (1999) Structure of the InIB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen L. moocytogenes. Mol Cell 4: 1063-72.

179. Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23-61.

180. Martin RC, Mok MC, Habben JE, Mok DWS (2001) A maize cytokinin gene encoding an O-glucosyltransferase specific to cis-zeatin Proc Natl Acad Sci U.S.A. 98: 5922-5926

181. Martin RC, Mok MC, Mok DWS (1999) Isolation of a cytokinin gene, ZOG1, encoding zeatin O-glucosyltransferase from *Phaseolus lunatus* Proc Natl Acad Sci U.S.A. 96: 284-289

182. Martonàk R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello-Rahman method revisited. Physiol Rev Lett 90:075503 .

183. Masada S, Terasaka K, Mizukami H (2007) A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glycosyltransferase), a Group D Family I glucosyltransferase from Catharanthus roseus. FEBS Lett 581: 2605-10.

184. 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 of UDP-Galactose: Flavonoid 3-O-Galactosyltransferase (UF3GaT) Expressed in *Vigna mungo* Seedlings. Plant Cell Physiol 39: 1145-1155.

185. Mauch F, Mauch-Mani B, Boller T (1988) Antifungal hydrolases in pea tissue II. Inhibition of fungal growth by combinations of chitinase and beta-1,3-glucanase. Plant Physiol 88: 936-942

186. McKersie BD, Leshem Y (1994) Stress and stress coping in cultivated plants. Klumer Academic Publishers, Netherland.

187. Mehta PA, Rebala KC, Venkataraman G, Parida A (2009) A diurnally regulated dehydrin from Avicennia marina that shows nucleo-cytoplasmic localization and is phosphorylated by Casein kinase II in vitro. Plant Physiol Biochem 47: 701-709.

188. Mehta RA, Parsons BL, Mehta AM, Nakhasi HL, Mattoo AK (1991) Differential protein metabolism and gene expression in tomato fruit during wounding stress. Plant Cell Physiol 32: 1057-1065.

189. Meurinne-Langlois M, Gachon CMM, Saindrenan P (2005) Pathogen-responsive expression of glycosyltransferase genes UGT73B3 and UGT73B5

is necessary for resistance to *Pseudomonas syringae* pv tomato in Arabidopsis. Plant Physiol 139: 1890-901.

190. Meyers BC, Dickerman AW, Michelmore RW (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20: 317-332.
191. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis. Plant Cell 15:809-834.
192. Meyers BC, Morgante M, Michelore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. Plant J 32:77-92.
193. Miller G, Mittler R (2006) Could heat shock transcription factors function as hydrogen peroxide sensors in plants? Ann Bot 98: 279-288.
194. Miller JD, Arteca RN, Pell EJ (1999) Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. Plant Physiol 120: 1015-24.
195. Miller KD, Guyon V, Evans JNS, Shuttleworth WA, Taylor LP (1999) Purification, cloning, and heterologous expression of a catalytically efficient flavonol 3-O-galactosyltransferase expressed in the male gametophyte of *Petunia hybrid*. J Biol Chem 274: 34011-34019.
196. Minina EA, Filonova LH, Daniel G, Bozhkov PV (2013) Detection and Measurement of Necrosis in Plants. Necrosis, Methods in Molecular Biology 1004: 229-248.
197. Mittler R, Blumwald E (2010) Genetic engineering for modern agriculture: challenges and perspectives. Annu Rev Plant Biol 61: 443-462.
198. Mittler R, Finka A, Goloubinoff P (2012) How do plants feel the heat? Trends Biochem Sci. 37: 118-125.
199. Modolo LV, Blount JW, Achnine L, Naoumkina MA, Wang X (2007) A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. Plant Mol Biol 64: 499-518.
200. Mohan R, Bajar AM, Kolattukudy PE (1993) Induction of a tomato anionic peroxidase gene (tap1) by wounding in transgenic tobacco and activation of

tap1/ GUS and tap2/GUS chimeric gene fusions in transgenic tobacco by wounding and pathogen attack. Plant Mol Biol 21: 341-54.

201. Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109:1434-1447.
202. Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC (2011) Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferase responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). Plant J 65: 106-18.
203. Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N, Nitasaka E, Noguchi H, lida S (2005) Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2"-O-glucosyltransferase, due to 4-bp insertions in the gene. Plant J 42: 353-363.
204. Mun JH, Yu HJ, Park S, Park BS (2009) Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. Mol Genet Genomics 282:617-631.
205. Munns R (2005) Bringing them together. Tansley Rev New Phytol 167: 645-663.
206. Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59: 651-681.
207. Nakamoto H, Vigh L (2007) The small heat shock proteins and their clients. Cell Mol Life Sci 64: 294-306.
208. Nakashima K, Yamaguchi-Shinozaki K (2006) Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. Physiologia Plantarum 126: 62-71.
209. Nakatsuka T, Sato K, Takahashi H, Yamamura S, Nishihara M (2008) Cloning and characterization of the UDP-glucose:anthocyanin 5-O-glucosyltransferase gene from blue-flowered gentian. J Exp Bot 59: 1241-1252.
210. Nicol JM, Turner SJ, Coyne DL, Nijs Ld, Hockland S, Maafi ZT (2011) Current nematode threats to world agriculture. In: Jones J, Gheysen G, Fenoll

C, eds. Genomics and molecular genetics of plant-nematode interactions. Amsterdam, the Netherlands: Springer, 21-43.

211. Nissen MS, Kumar GNM, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowlws NR, Kang CH (2009) Characteization of *Solanum tuberosum*multicystatin and its structural comparison with other cystatins. Plant Cell 21: 861–875
212. Ogata J, Itoh Y, Ishida Y, Yoshida H, Ozeki Y (2004) Cloning and heterologous expression of cDNAs encoding flavonoid glucosyltransferase from *Dianthus caryophyllus*. Plant Biotechnol 21: 367-375.
213. Ogata J, Kanno Y, Itoh Y, Tsugawa H, Suzuki M (2005) Plant biochemistry: anthocyanin biosynthesis in roses. Nature 435: 757-758.
214. Ohme-Takagi M, Suzuki K, Shinshi H (2000) Regulation of ethylene-induced transcription of defense genes. Plant Cell Physiol 41:1187-1192.
215. Orvar BL, Sangwan V, Omann F, Dhindsa RS (2000) Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. Plant J 23: 785-794.
216. Osmani SA, Bak S, Moller BL (2009) Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochem 70: 325-347.
217. Paetzel M, Karla A, Strynadka NC, Dalbey RE (2002) Signal peptidases. Chem Rev 102: 4549-4580.
218. Pang Q, Chen S, Dai S, Chen Y, Wang Y, Yan X (2010) Comparative proteomics of salt tolerance in *Arabidopsis thaliana* and *Thellungiella halophila*. J Proteome Res 9: 2584-2599.
219. Patil DN, Chaudhary A, Sharma AK, Tomar S, Kumar P (2012) Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. FEBS J 279: 4547-4564.
220. Perata P, Alpi A (1993) Plant responses to anaerobiosis. Plant Science 93: 1-17.
221. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.

222. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z (2014) ZDOCK Server: Interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 30: 1771-1773.

223. Pittaway JK, Robertson IK, Ball MJ (2008) Chickpeas may influence fatty acid and fiber intake in an ad libitum diet, leading to small improvements in serum lipid profile and glycemic control. J Am Diet Assoc 108:1009-13

224. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Mol Genet Genomics 281:609-626.

225. Prasad PVV, Boote KJ, Allen LH Jr (2006) Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain sorghum [*Sorghum bicolor* (L.) Moench] are more severe at elevated carbon dioxide due to higher tissue temperatures. Agric For Meteorol 139: 237-251.

226. Pratt WB, Krishna P, Olsen LJ (2001) Hsp90-binding immunophilins in plants: the protein movers. Trends Plant Sci 6: 54-58.

227. Przymusiński R, Rucińska R, Gwóźdź EA (1995) The stress-stimulated 16 kDa polypeptide from lupin roots has properties of cytosolic Cu:Zn-superoxide dismutase. Environ Exp Bot 35: 485-495.

228. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290-D301.

229. Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417: 618-624.

230. Rachwat D, Nebesny E, Budryn G (2013) Chickpea – composition, nutritional value, health benefits, application to bread and snacks: A review. Crit Rev Food Sci Nutr.

231. Rambaut A (2009) FigTree, ver. 1.3.1 [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/ [2009, December 21].

232. Ranson NA, White HE, Saibil HR (1998) Chaperonins. Biochem J 333: 233-242.

233. Rao MV, Paliyath C, Ormrod DP (1996) Ultraviolet-B and ozone-induced biochemical changes in antioxidant enzymes of *Arabidopsis thaliana*. Plant Physiol 110: 125-136.

234. Rawlings ND, Morton FR, Barrett AJ (2006) MEROPS: the peptidase database. Nucleic Acids Res 34: D270-2.

235. Ricard B, Rivoal J, Spiteri A, Pradet A (1991) Anaerobic stress induces the transcription and translation of sucrose synthase in rice. Plant Physiol 95: 669-674.

236. Richter K, Buchner J (2001) Hsp90: chaperoning signal transduction. J Cell Physiol 188: 281-290.

237. Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When Defense Pathways Collide: The Response of Arabidopsis to a Combination of Drought and Heat Stress. Plant Physiol. 134: 1683-1696.

238. Rosenblum JS, Kozarich JW (2003) Prolyl peptidases: a serine protease subfamily with high potential for drug discovery. Curr Opin Chem Biol 7: 496-504.

239. Rotanova TV, Melnikov EE, Khalatova AG, Makhovskaya OV, Botos I, Wlodawer A, Gustchina A (2004) Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. Eur J Biochem 271: 4865-4871.

240. Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396: 336-342.

241. Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner HY, Hunt MD (1996) Systemic acquired resistance. Plant Cell 8: 1809-1819.

242. Sabbavarapu MM, Sharma M, Chamarthi SK, Swapna N, Rathore A, Thudi M, Gaur PM, Pande S, Singh S, Kaur L, Varshney RK (2013) Molecular mapping of QTLs for resistance to *Fusarium* wilt (race 1) and *Ascochyta* blight in chickpea (*Cicer arietinum* L.). Euphytica 193: 121-133.

243. Sachs MM, Subbaiah CC, Saab IN (1996) Anaerobic gene expression and flooding tolerance in maize. J Exp Bot 47: 1-15.

244. Sakata T, Takahashi H, Nishiyama I, Higashitani A (2000) Effects of high temperature on the development of pollen mother cells and microspores in barley (*Hordeum vulgare* L.) J Plant Res 113: 395-402.

245. Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi-Shinozaki K (2006) Dual function of an Arabidopsis transcription factor

DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proc Natl Acad Sci USA 103:18822-18827.

246. Sangwan V, Foulds I, Singh J, Dhindsa RS (2001) Cold-activation of Brassica napus BN115 promoter is mediated by structural changes in membranes and cytoskeleton, and requires Ca2+ influx. Plant J 27: 1-12.
247. Santner A, Estelle M (2010) The ubiquitin-proteasome system regulates plant hormone signaling. Plant J 61: 1029-1040.
248. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27: 221-234.
249. Sato S, Kamiyama M, Iwata T, Makita N, Furukawa H, Ikeda H (2006) Moderate increase of mean daily temperature adversely affects fruit set of *Lycopersicon esculentum* L. by disrupting specific physiological processes in male reproductive development. Ann Bot 97: 731-738.
250. Sauter A, Davies WJ, Hartung W (2001) The long-distance abscisic acid signal in the droughted plant: the fate of the hormone on its way from root to shoot. J Exp Bot 52: 1991-1997.
251. Savchenko GE, Klyuchareva EA, Abramck LM, Serdyuchenko EV (2002) Effect of periodic heat shock on the inner membrane system of etioplasts. Russ J Plant Physiol 49: 349-359.
252. Schaller A and Ryan CA (1996) Molecular cloning of a tomato leaf cDNA encoding an aspartic protease, a systemic wound response protein. Plant Mol Biol 31: 1073-1077.
253. Schrödinger Release 2013-1: Desmond Molecular Dynamics System, version 3.4, D. E. Shaw Research, New York, NY, 2013. Maestro-Desmond Interoperability Tools, version 3.4, Schrödinger, New York, NY, 2013.
254. Schrödinger Release 2013-2: Prime, version 3.3, Schrödinger, LLC, New York, NY, 2013.
255. Schroeder JI, Kwak JM, Allen GJ (2001) Guard cell abscisic acid signalling and engineering drought hardiness in plants. Nature 410: 327-330.
256. Schüttelkopf AW, Aalten van DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Cryst D60:1355-1363.

257. Shah K, Dubey RS (1995) Cadmium induced changes on germination, RNA level and ribonuclease activity in rice seeds. Plant Physiol Biochem (India) 22: 101-107.

258. Shah K, Dubey RS (1998) A 18 kDa cadmium inducible protein complex: its isolation and characterization from rice (*Oryza sativa* L.) seedlings. J Plant Physiol 152: 448-454.

259. Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosltransferase from *Medicago truncatula*. Plant Cell 17: 3141-3154.

260. Shao HB, Guo QJ, Chu LY et al (2007) Understanding molecular mechanism of higher plant plasticity under abiotic stress. Colloids Surf B Biointerfaces 54: 37-45.

261. Sharma N, Russell S, Bhalla PL, Singh MB (2011) Putative *cis*-regulatory elements in genes highly expressed in rice sperm cells. 4: 319.

262. Sharma R, Panigrahi P, Suresh C.G. (2014) *In-Silico* analysis of binding site features and substrate selectivity in plant flavonoid-3-O glycosyltransferases (F3GT) through molecular modeling, docking and dynamics simulation studies. PLoS ONE 9(3): e92636. doi:10.1371/journal.pone.0092636.

263. Sharma S, Yadav N, Singh A, Kumar R (2011) Nutrition and antinutrition profile of newly developed chickpea (Cicer arietinum L) varieties. Int Food Res J 20: 805-810.

264. Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett 584: 2258-2264

265. Siezen RJ, Leunissen JA (1997) Subtilases: the superfamily of subtilisinlike serine proteases. Protein Sci 6: 501-523.

266. Singh VK, Garg R, Jain M (2013) A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. Plant Biotechnology Journal 11: 691-701.

267. Sivaji M, Sadasivam V, Narayanasamy J, Samuel S, Fan C (2014) Detection, Characterization and Evolution of Internal Repeats in Chitinases of Known 3-D Structure. PLOS ONE 10.1371/journal.pone.0091915

268. Smirnoff N (1993) Role of active oxygen in the response of plants to water deficit and desiccation. New Phytol 125: 27-58.

269. Smykowski A, Zimmermann P, Zentgraf U (2010) G-Box Binding Factor1 Reduces CATALASE2 Expression and Regulates the Onset of Leaf Senescence in Arabidopsis. Plant Physiol 153: 1321-1331.

270. Sobhanian H, Motamed N, Jazii FR, Nakamura T, Komatsu S (2010) Salt stress induced differential proteome and metabolome response in the shoots of Aeluropus lagopoides (Poaceae), a halophyte C_4 plant. J proteome Res 9: 2882-2897.

271. Sohn KH, Zhang Y, Jones JDG (2009) The Pseudomonas syringae effector protein, AvrRPS4, requires in planta processing and the KRVY domain to function. Plant J 57: 1079–1091.

272. Spiers A, Lamb HK, Cocklin S, Wheeler KA, Budworth J, Dodds AL, Pallen MJ, Maskell DJ, Charles IG, Hawkins AR (2002) PDZ domains facilitate binding of high temperature requirement protease A (HtrA) and tail-specific protease (Tsp) to heterologous substrates through recognition of the small stable RNA A (ssrA)-encoded peptide. J Biol Chem 277: 39443-39449.

273. Spinelli F, Cellini A, Marchetti L, Nagesh KM, Piovene C (2011) Emission and function of volatile organic compounds in response to abiotic stress. Agricultural and Biological Sciences "Abiotic Stress in Plants-Mechanism and Adaptations" Chapter 16 DOI: 10.5772/24155.

274. Spoel D Van der, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: Fast, Flexible and Free. J Comp Chem 26:1701-1719.

275. Steffens B, Sauter M (2010) G proteins as regulators in ethylene-mediated hypoxia signalling. Plant Signal Behav 5: 375-378.

276. Strid A, Chow WS, Anderson JM (1994) UV-B damage and protection at the molecular level in plants. Photosynth Res 39: 475-489.

277. Subbaiah CC, Sachs MM (2003) Molecular and cellular adaptations of maize to flooding stress. Ann Bot 90: 119-127.

278. Suite 2012: LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2012.

279. Sun W, Gao F, Fan H, Shan X, Sun R, Liu L, Gong W (2013) The structures of Arabidopsis Deg5 and Deg8 reveal new insights into HtrA proteases. Acta Crystallogr D Biol Crystallogr 69: 830-7.

280. Sun YG, Wang B, Jin SH, Qu XX, Li YJ, Hou BK (2013) Ectopic expression of Arabidopsis glycosyltransferase UGT85A5 enhances salt stress tolerance in Tobacco. Plos One 10.1371/journal.pone.0059924

281. Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R (2014) Abiotic and biotic stress combinations. New Phytol 203: 32-43.

282. Tada Y, Kashimura T (2009) Proteomic analysis of salt-responsive proteins in the mangrove plant, Bruguiera gymnorhiza. Plant Cell Physiol 50: 439-446.

283. Taguchi G, Yazawa T, Hayashida N, Okazaki M (2001) Molecular cloning and heterologous expression of novel glucosyltransferase from tobacco cultured cells that have broad substrate specificity and are induced by salicylic acid and auxin. Eur J Biochem 268: 4086-4094.

284. Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu JK, Shinozaki K (2004) Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray. Plant Physiol 135: 1697-1709.

285. Takeda K, Akira S (2005) Toll-like receptors in innate immunity. Int. Immunol. 17: 1–14.

286. Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL, Haring MA, Cornelissen BJ (2002) The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell 14: 2929-2939.

287. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2713-9.

288. Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, Bent AF, Michelmore RW (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol 7:56.

289. Tanaka Y, Yonekura K, Fukuchi-Mizutani M, Fukui Y, Fujiwara H (1996) Molecular and biochemical characterization of three anthocyanin synthetic enzymes from *Gentiana triflora*. Plant Cell Physiol 37: 711-716.

290. Taniguchi N, Honke K, Fukuda M (2003) Handbook of Glycosyltransferases and Related Genes. Springer 68: 707-708.

291. Teixeira MT, Fabre E, Dujon B (1999) Self-catalyzed cleavage of the yeast nucleoporin Nup145p precursor. J Biol Chem 274: 32439-32444.

292. Tekeoglu M (2004) QTL analysis of *Ascochyta* blight resistance in chickpea. Turk J Agric For 28:183–187.

293. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

294. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882.

295. Tripathi LP, Sowdhamini R (2008) Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes. 9: 549.

296. Trivedi DK, Ansari MW, Tuteha N (2013) Multiple abiotic stress responsive rice cyclophilin (OsCYP-25) mediates a wide range of cellular responses. Communication & Intergrative Biology 6: 5, e25260.

297. Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K, Kyozuka J, Yamaguchi S (2008) Inhibition of shoot branching by new terpenoid plant hormones. Nature 455: 195-200.

298. Urban S, Lee JR, Freeman M (2001) Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. Cell 107: 173-182.

299. Vacca RA, de Pinto MC, Valenti D, Passarella S, Marra E, De Gara L (2004) Production of reactive oxygen species, alteration of cytosolic ascorbate peroxidase, and impairment of mitochondrial metabolism are early events in heat shock-indiced programmed cell death in tobacco bright-yellow 2 cells. Plant Physiol 134: 1100-1112.

300. Van der Hoorn RAL, Jones JDG (2004) The plant proteolytic machinery and its role in defence. Curr Opin Plant Biol 7: 400-407.

301. Van der Hoorn RAL, Jones JDG (2004) The plant proteolytic machinery and its role in defence. *Curr Opin Plant Biol* 7: 400-407.

302. van Loon LC (1999) Occurrence and properties of plant pathogenesis related proteins. In: Dutta SK, Muthukrishnan S (eds.) Pathogenesis related proteins in plants. CRC Press, Boca Raton

303. van Loon LC, Pierpoint WS, Boller T, Conejero V (1994) Recommendations for naming plant pathogenesis-related proteins. Plant Molecular Biology Reporter 12: 245-264.

304. van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 44: 1-28.

305. Van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol, 45:43-72.

306. Varshney RK, Song C, Saxena RK, Azam S, Yu Sheng, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol* 31:240-246.

307. Vergnolle C, Vaultier MN, Taconnat L, Renou JP, Kader JC, Zachowski A, Ruelland E (2005) The cold-induced early activation of phospholipase C and D pathways determines the response of two distinct clusters of genes in Arabidopsis cell suspensions. Plant Physiol 139: 1217-1233.

308. Volkov RA, Panchuk II, Mullineaux PM, Schöffl F (2006) Heat stress-induced H_2O_2 is required for effective expression of heat shock genes in Arabidopsis. Plant Mol Biol 61: 733-746.

309. von Moltke J, Ayres JS, Kofoed EM, Chavarría-Smith J Vance RE (2013) Recognition of bacteria by inflammasomes. Annu. Rev. Immunol. 31: 73–106.

310. Wakeel A, Asif AR, Pitann B, Schubert S (2011) Proteome analysis of sugar beet (Beta vulgaris L.) elucidates constitutive adaptation during the first phase of salt stress. J Plant Physiol 168: 519-526.

311. Walsh T, Strickland J (1993) Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. Plant Physiol 103: 1227–1234.

312. Wan H, Yuan Wei, Bo K, Shen J, Pang X, Chen J (2013) Genome wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genomics 14:109.

313. Wang MC, Peng ZY, Li CL, Li F, Liu C, Xia GM (2008) Proteomics analysis on a high salt tolerance introgression strain of *Triticum aesticum*/*Thinopyrum ponticum*. Proteomics 8: 1470-1489.

314. Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta. 218: 1-14.

315. Wang X (2009) Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Letters 583: 3303-3309.

316. Wang X, Fan P, Song H, Chen X, Li X, Li Y (2009) Comparative proteomics analysis of differentially expressed proteins in shoots of *Salicornia europaea* under different salinity. J Proteome Res 8: 3331-3345.

317. Wei H, Li W, Sun X, Zhu S, Zhu J (2013) Systematic Analysis and Comparison of Nucleotide-Binding Site Disease Resistance Genes in a Diploid Cotton *Gossypium raimondii*. PLoS ONE 8(8):e68435.doi10.1371/journal.pone.0068435

318. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407-W410.

319. Williams ME, Torabinejad J, Cohick E, Parker K, Drake EJ, Thompson JE, Hortter M, Dewald DB (2005) Mutations in the Arabidopsis phosphoinositide gene SAC9 lead to overaccumulation of Ptdlns (4, 5)P2 and constitutive expression of the stress-response pathway. Plant Physiol 138: 686-700.

320. Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, Ve T, Ma Y, Saucet SB, Ericsson DJ, Casey LW, Lonhienne T, Winzor DJ, Zhang X, Coerdt A, Parker JE, Dodds PN, Kobe B, Jones JDG (2014) Structural basis

for assembly and function of a heterodimeric plant immune receptor. Science 344: 299-303.

321. Wilson KE, Ivanov AG, Öquist G, Grodzinski B, Sarhan F, Hüner NPA (2006) Energy balance, organellar redox status, and acclimation to environmental stress. Can J Bot 84: 1355-1370.
322. Wise RP, Rohde W, Salamini F (1990) Nucleotide sequence of the Bronze-1 homologous gene from *Hordeum vulgare*. Plant Mol Biol 14: 277-279.
323. Woo HH, Jeong BR, Hirsch AM, Hawes MC (2007) Characterization of *Arabidopsis* AtUGT85A and AtGUS gene families and their expression in rapidly dividing tissues. Genomics 90: 143-153.
324. Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, Dixon RA, Lamb C (2004) An extracellular aspartic protease functions in Arabidopsis disease resistance signalling. *The EMBO Journal* 23: 980-988.
325. Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291:118-120.
326. Xiong L, Ishitani M, Lee H, Zhu JK (2001) The Arabidopsis LOS5/ABA3 locus encodes a molybdenum cofactor sulfurase and modulates cold stress- and osmotic stress-responsive gene expression. Plant Cell 13: 2063-2083.
327. Xiong LM, Schumaker KS, Zhu JK (2002) Cell signalling during cold, drought, and salt stress. Plant Cell 14: S165-S183.
328. Xu ZJ, Nakajima M, Suzuki Y, Yamaguchi I (2002) Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. Plant Physiol 129: 1285-1295.
329. Yadav SS, Redden R, Chen W, Sharma B (2007) Chickpea Breeding and Management. CAB International 538–554.
330. Yadeta KA, Thomma BPHJ (2013) The xylem as battleground for plant hosts and vascular wilt pathogens. Front Plant Sci 4: 97.
331. Yaish MW, Saenz de Miera LE, Perez de la Vega M (2004) Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (NBS) type from Lens species. Genome 47:650-659.

332. Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57: 781-803.

333. Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics 8: 67.

334. Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Saito K (1999) Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin J Biol Chem 274: 7405-7411.

335. Yan HB, Lou ZZ, Li L, Brindley PJ, Zheng Y, Luo X, Hou J, Guo A, Ji WZ, Cai X (2014) Genome-wide analysis of regulatory proteases sequences identified through bioinformatics data mining in *Taenia solium*. BMC Genomics 15: 428.

336. Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics 280:187-198.

337. Yasuda E, Ebinuma H, Wabiko H (1997) A novel glycine-rich/hydrophobic 16 kDa polypeptide gene from tobacco: similarity to proline-rich protein genes and its wound-inducible and developmentally regulated expression. Plant Mol Biol 33: 667-678.

338. Yeats TH, Rose JKC (2008) The biochemistry and bioogy of extracellular plant lipid-transfer proteins (LTPs). Protein Sci 17: 191-198.

339. Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J 66: 182-93.

340. Yoshihara N, Imayama T, Mizutani-Fukuchi M, Okuhara H, Tanaka Y, Ino Ikuo, Yabuya T (2005) cDNA cloning and characterization of UDP-glucose: Anthocyanidin 3-O-glucosyltransferase in *Iris hollandica*. Plant Sci 169: 496-501.

341. Young JC, Moarefi I, Hartl FU (2001) HSP90: a specialized but essential protein-folding tool. J Cell Biol 154: 267-73.

342. Yu J, Chen S, Zhao Q, Wang T, Yang C, Diaz C, Sun G, Dai S (2011) Physiological and proteomic analysis of salinity tolerance in *Puccinellia tenuiflora*. J Proteome Res 10: 3852-3870.

343. Zhou T, Wang Y, Chen J-Q, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expression of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271:402-415.

344. Zhu H, Cannon SB, Young ND, Cook DR (2002) Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*. Mol Plant Microbe Interact 15: 529-539.

345. Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53: 247-273.

List of Publications

1. Genome-wide identification and structure-function studies of proteases and protease inhibitors in chickpea genome.
   (doi: http://dx.doi.org/10.1016/j.compbiomed.2014.10.019: **Computers in Biology and Medicine**)
   Ranu Sharma, C.G. Suresh

2. Genome-wide identification and tissue specific expression analysis of UDP-glycosyltransferases genes in *Cicer arietinum* (chickpea) genome
   (doi:10.1371/journal.pone.0109715: **PLOS ONE**)
   Ranu Sharma, Vimal Rawat, C.G. Suresh

3. *In-Silico* Analysis of Binding Site Features and Substrate Selectivity in Plant Flavonoid-3-O Glycosyltransferases (F3GT) through Molecular Modeling, Docking and Dynamics Simulation Studies
   (doi:10.1371/journal.pone.0092636: **PLOS ONE**)
   Ranu Sharma, Priyabrata Panigrahi, C.G. Suresh

4. An improved method for specificity annotation shows a distinct evolutionary divergence among the microbial enzymes of the Cholylglycine hydrolase family (doi: 10.1099/mic.0.077586-0: **Microbiology**).
   Priyabrata Panigrahi, Manas S Sule, **Ranu Sharma**, Sureshkumar Ramasamy, Suresh C.G.

5. Structural effects of Leigh syndrome mutation on the function of human mitochondrial Complex-1 Q module,
   Jaokar TM, **Sharma R**, Suresh C.G. (2013). ***Biochem Physiol*** S2. doi:10.4172/2168-9652.S2-004

6. Cloning, expression and in silico studies of a serine protease from a marine actinomycete (Nocardiopsis sp. NCIM 5124) (under review in **Process Biochemistry**) Sonali Rohamare, Sushama Gaikwad, Dafydd Jones, Varsha Bhavnani, Jayanta Pal, **Ranu Sharma**, Prathit Chatterjee.

7. *In Silico* substrate specificity in bmgt1 and bmgt2 genes of *Bacopa monniera* glycosyltransferases 12(2):413-430 **Online J Bioinform.**
   Sharma R, Ruby Zargar, Khan BM, Suresh CG (2011).

8. Genome-wide identification of nucleotide-binding site (NBS) - leucine-rich repeat (LRR) gene family in *Cicer arietinum* (chickpea) (communicated),
   Ranu Sharma, Vimal Rawat, C.G. Suresh

9. Leigh syndrome mutations affected the structural stability and ubiquinone binding affinity of mitochondrial Complex-I subunit NDUFS7, evidence from biophysical and computational studies (under preparation).
   Tulika Jaokar, Deepak Patil, **Ranu Sharma**, Shouche Yogesh, Sushama Gaikwad, Suresh C.G.

Contents of CD

Sr. No	Description
	Chapter 3
CD-Figure S-3.1	MSA between alignment between query and template
CD-Figure S-3.2	MSA of 30 F3GT protein sequences
CD-Table S-3.1	Dataset of 101 GT sequences
CD-Table S-3.2	Percent identity matrix of 30 F3GT protein sequences
CD-Video S1	Video showing firm binding KMP and UPG in the binding pocket of F3GT
CD-Video S2	Video showing firm binding KMG and UDP in the binding pocket of F3GT
	Chapter 4
CD-Figure S-4.1	MSA of 96 chickpea UGTs
CD-Figure S-4.2	MSA of 4 chickpea UGTs identified by HMM
CD-Figure S-4.3	MSA of chickpea UGTs with experimentally validated UGT proteins
CD-Figure S-4.4	MSA between target chickpea UGTs and templates
CD-Figure S-4.5	Heatmap showing expression of *Ca*UGTs under drought
CD-Table S-4.1	Summary of 96 chickpea UGTs
CD-Table S-4.2	Orthologs of chickpea UGTs in four selected dicot plants
CD-Table S-4.3	Standard deviation of protein lengths of *Ca*UGTs
CD-Table S-4.4	Expression value for CaUGTs under drought stress
	Chapter 5
CD-Figure S-5.1	MSA of proteases and PIs
CD-Figure S-5.2	Dendrogram of serine proteases showing bootstrap values
CD-Figure S-5.3	Domain arrangement of chickpea MPs
CD-Figure S-5.4	Heatmap showing relative gene expression of

	proteases and PIs under drought condition
CD-Figure S-5.5	MSA between query and templates
CD-Table S-5.1	Details of proteases and PIs
CD-Table S-5.2	Complete list of identified proteases and PIs
CD-Table S-5.3	Orthologs of chickpea proteases and PIs
CD-Table S-5.4	Description of *Cicer arietinum* EST BLAST hits against the chickpea dbEST in NCBI
CD-Table S-5.5	Expression values of proteases and PI genes of chickpea
CD-Table S-5.6	Expression value for proteases and PI under drought stress
	Chapter 6
CD-Figure S-6.1	Motif sequence logos of the non-TNL family members of chickpea
CD-Figure S-6.2	Motif sequence logos of the TNL family members of chickpea
CD-Figure S-6.3	Heatmap showing relative gene expression of non-TNL and TNL genes in drought condition
CD-Table S-6.1	Details of disease resistance NBS encoding genes/proteins of chickpea
CD-Table S-6.2	Orthologs identification of chickpea NBS proteins
CD-Table S-6.3	The motif patterns of NBS gene family of chickpea
CD-Table S-6.4	Expression value for NBS genes under drought stress
	Chapter 7
CD-Figure S-7.1	Heatmap showing relative gene expression of of identified stress genes in drought condition
CD-Table S-7.1	Details of other stress genes of chickpea
CD-Table S-7.2	Othologs identification in other stress genes of chickpea
CD-Table S-7.3	Expression values of other stress genes
CD-Table S-7.4	Expression values of other stress genes under drought stress