Chapter 5: Molecular cloning and characterization of 4 coumarate: coenzyme A ligase (4CL) and chalcone synthase (CHS) from Coleus forskohlii

5.1 Introduction
Flavonoids are an important class of secondary metabolites that are involved in wide array of processes like floral pigmentation, pollination, nitrogen fixation, response to stress, UV, pathogens, insects etc. (Taylor and Grotewold 2005; Lillo, Lea, and Ruoff 2008; Fini et al. 2011; Winkel-Shirley 2002; Päsold et al. 2010). Several flavonoids have been shown to possess anti-oxidant, anti-inflammatory and antitumor activities (Cazarolli et al. 2008). 4-coumarate: CoA ligase (4CL) and chalcone synthase (CHS) are important key enzymes in the biosynthesis of flavonoids. 4CL channelizes the carbon flow towards flavonoid biosynthesis through general phenylpropanoid pathway whereas, CHS commits the channelization of carbon flux from phenylpropanoid pathway towards biosynthesis of flavonoids (Figure 5.1).

In plants, phenylpropanoid pathway involves the sequential action of three enzymes: phenylalanine ammonia-lyase (PAL; EC 4.3.1.5), cinnamate 4-hydroxylase (C4H; EC 1.14.13.11) and 4-coumarate: CoA ligase (4CL; EC 6.2.1.12). Firstly, PAL transforms L-phenylalanine to cinnamic acid and then C4H catalyzes the transfer of a hydroxyl group at the para position of the phenyl ring of cinnamic acid, producing p-coumaric acid. Finally, 4CL activates the conversion of 4-coumaric acid into coumaroyl-CoA and its related derivatives into their respective thioesters such as feruloyl-CoA, caffeoyl-CoA, and cinnamoyl-CoA (Hahlbrock & Scheel, 1989). In this process, ATP is utilized. These esters serve as intermediates for the synthesis of a wide array of compounds such as flavonoids, isoflavonoids, lignins, lignans, anthocyanins, coumarins, stilbenoids, etc. (Douglas et al. 1992; Vogt 2010). 4CL has been cloned from many plant species like Arabidopsis thaliana (L.) Heynh. (Brassicaceae), Pinus taeda L. (Pinaceae), Salvia miltiorrhiza Bunge (Lamiaceae), Rubus idaeus L. (Rosaceae), etc. (Ehlting et al., 1999; Kumar & Ellis, 2003; Zhang & Chiang, 1997; Zhao, Hu, Liu, & Leung, 2006). 4CL is encoded by a small gene family consisting of two to four members (Douglas et al. 1992; Ehlting et al. 1999; Zhao et al. 2006) and is grouped into the AMP-binding superfamily (Cukovic, Ehlting, VanZiffle, & Douglas, 2001). Differential expression pattern of 4CL in different tissues, development stages
and in response to various elicitors/phytohormones, herbivory attack, pathogens, UV, wounding, etc. has been studied in various plants (Hu et al., 2010; Lee, Ellard, Wanner, Davis, & Douglas, 1995; Lindermayr et al., 2002; Porth, Hamberger, White, & Ritland, 2011; Voo, Whetten, O’Malley, & Sederoff, 1995).

Chalcone synthase (CHS) is an important enzyme that channelizes the carbon flux of phenylpropanoid pathway towards biosynthesis of flavonoids (Figure 5.1). Phenylpropanoid pathway also provides precursors for biosynthesis of other important metabolites, including rosmarinic acid, which is a known antimicrobial agent and has been shown to be effective against Japanese encephalitis (Swarup et al. 2007). CHS belongs to Type III polyketide synthase (PKS) superfamily (Schröder 1997), catalyzing the condensation of three molecules of malonyl-CoA and one molecule of CoA ester of cinnamic acid or its derivatives like coumaric acid yielding naringenin chalcone (Martens and Mithöfer 2005). Chalcone isomerase (CHI) then reversibly converts it to flavanone (Moustafa 1967). These intermediates mark the entry point from where pathway diverges into several branches, each resulting in a different class of flavonoids like anthocyanins, phytoalexins, phytoanticipins, etc. (Jez et al. 2001; Dao, Linthorst, and Verpoorte 2011). Its expression may be induced in response to herbivory attack, plant pathogens and various abiotic stresses like UV, wounding or exogenous phytohormone/elicitor treatment (Richard et al. 2000; Schenk et al. 2000).

In a study, it was found that Ipomoea purpurea (L.) plants with nonfunctional copies of CHS suffered almost twice the intensity of infection from Rhizoctonia solani and about 25% greater damage from herbivorous attack than the wild type, proving the role of CHS in pathogen resistance (Dao et al. 2011).

C. forskohlii (Willd.) Briq. (Lamiaceae) is an important herb having medicinal value and a diverse metabolic profile (Kavitha, Rajamani, & Vadivel, 2010; Paul, Radha, & Kumar, 2013). Here, in this study, we carried out cloning and characterization two key enzymes of flavonoids biosynthetic pathway from C. forskohlii.
5.2 Results

5.2.1 Cloning and sequence analysis of *Cf4CL* and *CfCHS*

5.2.1.1 Cloning and sequence analysis of *Cf4CL*
Degenerate primers were used to obtain core fragment of 464 base pairs (bps). 5’ and 3’ RACE primers were designed using the sequence information of the core amplicon. RACE-PCR was carried out to obtain the 5’ and 3’ ends of the cDNA, giving amplicons of 740 bps and 882 bps respectively (Figure 5.2A). The full-length clone of 1900 bps containing an open reading frame (ORF) of 1626 bps, starting with an ATG start codon at position 70 and ending with a TGA stop codon at position 1696 (Figure 5.3), was sequenced and designated as *Cf4CL* (NCBI GenBank accession no. KF643242). Its theoretical translation encoded a protein of 542 amino acids having molecular weight 58.77 kDa and pI 5.55. The 5’ and 3’ untranslated regions (UTR) were 69 bps & 202 bps long respectively. We could identify the conserved AMP-binding motif SSGTTGLPKGV in the *Cf4CL* sequence. Another motif, present in *Cf4CL*, that is highly conserved in all the 4CLs is GEICIRG which is associated with the stability and catalytic activity of 4CL (Schmelz and Naismith 2009) (Figure 5.3).

Closest orthologs of *Cf4CL* were identified using BlastP (NCBI) and were aligned. *Cf4CL* shared 87% identity and 90.5% similarity with the 4CL from *M. officinalis* (Figure 5.4A and 5.4B). Phylogenetic analysis was carried out using 4CL protein sequences from different organisms, and as expected, the *Cf4CL* falls in the clade of dicotyledonous plants, showing close similarity with *M. officinalis* (Figure 5.5A). With respect to class, *Cf4CL* clustered closely with *Petroselinum crispum* (Mill.) Mansf. (Apiaceae) 4CL (class I 4CL) (Figure 5.5B).

5.2.1.2 Cloning and sequence analysis of *CfCHS*
Degenerate primers were used to obtain a core fragment of 569 bp. Using the sequence information of core amplicon, 5’ and 3’ RACE primers were designed. RACE-PCR was carried out to obtain the 5’ and 3’ ends of the cDNA, giving an amplicon size of 842 bp and 362 bp respectively (Figure 5.2B). The full-length clone of 1598 bp contained an open reading frame (ORF) of 1176 bp, starting with an ATG start codon at position 245 and ending with a TGA stop codon at position 1420 (Figure 5.6). The clone was sequenced and designated as *CfCHS* (NCBI GenBank accession no. KF643243). Theoretical translation of *CfCHS* nucleotide sequence
Figure 5.1 Schematic representation of flavonoid biosynthetic pathway. 4CL and CHS (gene under study) are key enzymes involved in channelizing the carbon flux toward the biosynthesis of flavonoids.

Figure 5.2 Cloning of Cf4CL and CfCHS genes from C. forskohlii.

(A) Gel picture shows the PCR amplicon of core fragment, 5’ and 3’ RACE PCR fragment and full length cDNA of Cf4CL. (B) Gel picture shows the PCR amplicon of core fragment, 5’ and 3’ RACE PCR fragment and full length cDNA of CfCHS.
Figure 5.3 Nucleotide and deduced amino acid sequence of Cf4CL from *C. forskohlii*. The ATG start codon at position 70, the TAA stop codon at position 1696 and conserved amino acid motif-SSGTTLPKGV and GEICIRC are also shown.
Figure 5.4 Multiple sequence alignment of Cf4CL sequence and its homologs from other plant species. A: Multiple sequence alignment of Cf4CL protein sequence, with the homologous protein from other species S. baicalensis (GenBank Acc. No. BAD90937.1), M. officinalis (GenBank Acc. No. CBJ23825.1), P. fortunei (GenBank Acc. No. ACL31667.1) and S. miltiorrhiza (GenBank Acc. No. APP68991.1). B: Represents the percentage similarity and identity of Cf4CL with other homologous 4CL protein sequence of different plant species, calculated using MatGAT tool.
Figure 5.5 Phylogenetic tree analysis of Cf4CL. (A) Clustering of sequences from different taxonomic groups of plants, animals, fungi and bacteria. Cf4CL grouped with related sequences from dicotyledonous plants. (B) Clustering of Cf4CL two known classes of 4CL: class I and class II. Cf4CL falls in class I category.
Figure 5.6 Nucleotide and deduced amino acid sequence of CfCHS from *C. forskohlii*. The ATG start codon is positioned at 245, the TAA stop codon at position 1420 and conserved region are highlighted. Letter in bold-face indicate strictly conserved amino acid.
Figure 5.7 Multiple sequence alignment of CfCHS sequence and its homologs from other plant species.

A: Multiple sequence alignment of CfCHS protein sequence, with the homologous proteins from other plant species A. rugosa (GenBank Acc. No. AFL72079.1), S. scutellaroides (GenBank Acc. No. ABP57071.1), P. frutescens (GenBank Acc. No. BAA19548.1) and M. orontium (GenBank Acc. No. CAJ44127.1).

B: Represents the percentage similarity and identity of CfCHS with other homologous CHS protein sequence of different plant species, calculated using MatGAT tool.
Figure 5.8 Phylogentic tree analysis of CfCHS. Phylogentic tree of CfCHS showing clustering of CHS sequences from different taxonomic groups of plants, fungi and bacteria. CfCHS clusters with related sequences from dicotyledonous plants.
encoded a protein of 392 amino acids having a molecular weight 45.28 kDa and pI 5.36 (Figure 5.3). The 5’ and 3’ un-translated regions (UTR) were 244 bp & 178 bp long respectively. Closest homologs of CfCHS protein sequence were identified using BlastP (NCBI) and multiple sequence alignment was carried out (Figure 5.5A). CfCHS sequence sharing 90.1% identity and 94.1% similarity with the CHS from A. rugosa (Figure 5.7). A phylogenetic tree was constructed, using CHS protein sequences from different organisms, and as expected, CfCHS falls in the clade of dicotyledonous plants, showing close phylogenetic similarity with CHS from S. scutellarioides, belonging to the same family (Figure 5.8).

5.2.2 Expression analysis

5.2.2.1 Expression analysis of 4CL
Spatial expression pattern of Cf4CL was determined in different tissues: flowers, young leaves, mature leaves, stems, roots and root tips using qRT-PCR. Mature leaves showed the highest expression followed by young leaves while stems, roots, and root tips showed negligible expression (Figure 5.9A). Similar results have been reported in Rubus idaeus where three 4CL genes were cloned (Ri4CL1, Ri4CL2, and Ri4CL3) and Ri4CL1, clustering with class I 4CLs, showed the highest expression in leaves (Kumar and Ellis 2003).
Differential expression of Cf4CL on treatment with various phytohormones/elicitors like MeJA, SA, ABA and 2, 4-D, UV and wounding was also determined. Cf4CL expression was strongly induced in response to ABA and 2, 4-D treatment while treatment with MeJA and SA, at the tested conditions, did not affect its expression (Figure 5.9B).

5.2.2.2 Expression analysis of CfCHS
To understand the spatial pattern of distribution, expression profiling of CfCHS was carried out in flower, young leaf, mature leaf, stem, root, and root tip, using qRT-PCR. Young leaf showed highest expression of CfCHS, followed by flower, while root and root tip showed negligible expression (Figure 5.10A).
Treatments with chemical elicitors are known to mimic the effects of environmental stresses (Tuteja and Sopory 2008). The changes in expression pattern of CfCHS, on treatment with various phytohormones/elicitors like MeJA, SA, ABA and 2, 4-D,
were also determined. *CfCHS* expression was strongly induced in plants treated with MeJA (Figure 5.10B) whereas, ABA was found to marginally induces expression of *CfCHS* at the concentration of ABA and time-point we tested (Figure 5.10B). In our results, we also observed that 2, 4-D had a negligible effect on expression of *CfCHS* (Figure 5.10B). However, *CfCHS* appeared to be downregulated in response to SA treatment (Figure 5.10B).

### 5.2.3 miRNA prediction

#### 5.2.3.1 miRNA prediction study for *Cf4CL*

Our computational analysis revealed that expression of *Cf4CL* may be post-transcriptionally regulated by miRNAs, whose target sequence is conserved in at least three orthologs of 4CL from four different plant species: *S. miltiorhiza*, *S. baicalensis*, *P. fortunei* and *M. officinalis*. miR6249 was predicted to regulate *Cf4CL* RNA levels through cleavage, while miR5643, miR5648 and miR5661 may act by translational repression. miR1886 may act through cleavage and translational repression both (Figure 5.11A). However, this is only a preliminary analysis and further studies are needed to validate the predictions.

#### 5.2.3.2 miRNA prediction study for *CfCHS*

miRNAs are small non-coding RNAs that play a fundamental role in various biological processes (Bartel 2004; Sunkar et al. 2007). Our *in silico* search for microRNAs that may regulate *CfCHS* expression identified four miRNAs targeting *CfCHS*, whose target sequence is conserved in at least three other homologs of CHS from four different plant species (*S. scutellarioides*, *A. rugosa*, *P. frutescens* and *M. orontium*) that were tested. miR34, miR166, and miR395 were predicted to regulate *CfCHS* RNA levels through cleavage, while miR477 may act by translational repression and cleavage both (Figure 5.11B).

### 5.3 Discussion

Flavonoids are one of the extensively studied metabolites in the plant kingdom. Several flavonoids have been shown to possess anti-oxidant, anti-inflammatory and antitumor activities (Cazarolli et al. 2008). 4-coumarate: CoA ligase (4CL) and
**Figure 5.9 Expression profile study of Cf4CL**

(A) Tissue expression study of *Cf4CL* in flower, young leaf, mature leaf, stem, root and root tip of *C. forskohlii* was determined by quantitative real time RT-PCR (qPCR). Actin was used as housekeeping control and expression of *Cf4CL* in stem was used as baseline for calculating fold change in different tissues. (B) Expression study of Cf4CL under the effect of elicitor: MeJA, ABA, SA, 2,4-D, UV A and wounding; was determined in *C. forskohlii* (leaves) by quantitative real time RT-PCR (qPCR).

**Figure 5.10 Expression profile study of CfCHS.**

(A) Tissue expression study of *CfCHS* in flower, young leaf, mature leaf, stem, root and root tip of *C. forskohlii* was determined by quantitative real time RT-PCR (qPCR). Actin was used as housekeeping control and expression of *CfCHS* in stem was used as baseline for calculating fold change in different tissues. (B) Expression study of *CfCHS* under the effect of elicitor: MeJA, ABA, SA, 2,4-D; was determined in *C. forskohlii* (leaves) by quantitative real time RT-PCR (qPCR).
Figure 5.11 Prediction of miRNA targeting Cf4CL and CfCHS.

(A) Pictures shows conservation of target sites of predicted miRNAs in Cf4CL and its homologs from other plant species, and their possible modes of action (cleavage or inhibition of translation).

(B) Picture shows conservation of target sites of predicted miRNAs in CfCHS and its homologs from other plant species, and their possible modes of action (cleavage or inhibition of translation).
chalcone synthase (CHS) are important key enzymes in biosynthesis of flavonoids. We have cloned these two genes from *C. forskohlii*. Cf4CL gene, cloned in the study was found to cluster with class I 4CLs on phylogenetic tree analysis, thereby indicating that Cf4CL might be a class I 4CL (Figure 5.5B). Class I 4CLs are reported to be involved in lignin accumulation (Cao, Hu, Huang, Ren, & Lu, 2012; Chowdhury et al., 2013; Hamberger & Hahlbrock, 2004). Downregulation of 4CL1 gene family has been shown to reduce lignin content in poplar, thereby improving feedstock quality for paper and bioethanol production (Jia et al., 2004; Voelker et al., 2010). Cf4CL may also be involved in lignin biosynthesis. qPCR study has shown that Cf4CL expression is dominant in leaf. Similar results have been reported in *Rubus idaeus* where three 4CL genes were cloned (Ri4CL1, Ri4CL2, and Ri4CL3) and Ri4CL1, clustering with class I 4CLs, showed the highest expression in leaves (Kumar & Ellis, 2003). Differential expression of Cf4CL on treatment with various phytohormones/elicitors like MeJA, SA, ABA, 2, 4-D, UV and wounding was also determined. Application of chemical elicitors can mimic the effects of various abiotic stresses (Tuteja & Sopory, 2008). ABA is known to be involved in plant’s response to drought and osmotic (salt) stress (Cutler et al. 2010; Tuteja, 2007; Zhu, 2002) while 2, 4-D is a synthetic plant hormone (auxin), a herbicide that is absorbed through leaves and results in toxicity by accumulating reactive oxygen species (ROS) and nitric oxide (NO) (Egan, Maxwell, Mortensen, Ryan, & Smith, 2011; Song, 2014). Cf4CL expression was strongly induced in response to ABA and 2, 4-D treatment while treatment with MeJA and SA, at the tested conditions, did not affect its expression (Figure 5.9B). This is similar to earlier reported results where Class I 4CL from *Hibiscus cannabinus* L. (Malvaceae) was also unresponsive to MeJA and SA treatment (Chowdhury et al., 2013). SA levels have been reported to be inversely proportional to the amount of lignin (Gallego-Girald, Escamilla-Trevino, Jackson, & Dixon, 2011) and class I 4CLs are associated with lignin accumulation. Interestingly, Cf4CL was unresponsive to SA. UV An exposure and wounding also increased Cf4CL transcript levels manifold (Figure 5.9C). Previous studies have also reported similar results (Chowdhury et al., 2013; Ehlting et al., 1999; Kao, Harding, & Tsai, 2002). Since 4CL is an important enzyme for the biosynthesis of a wide number of secondary metabolites like flavonoids, stilbenoids, suberins, lignins, etc. (Douglas et al. 1992; Vogt 2010). Upregulation of
Cf4CL transcripts in response to abiotic stresses is indicative of its crucial role in plant’s defense machinery. Besides, C. forskohlii has a diverse metabolic profile (Kavitha et al., 2010; Paul et al., 2013). Phenolic compounds such as caffeic acid, coleside B and colexanthone have been isolated from C. forskohlii (Ahmed and Vishwakarma 1988; Alasbahi and Melzig 2010). Flavonoid- genkwanin (7-O-methylapigenin) and a phenylpropanoid- guaiacol glycerin ether, have also been reported from C. forskohlii (Alasbahi and Melzig 2010). 4CL provides intermediates for the synthesis of compounds like flavonoids, isoflavonoids, lignins, lignans, anthocyanins, coumarins, stilbenoids, etc. (Douglas et al. 1992; Vogt 2010).

miRNA are known to regulate plant genes post-transcriptionally. miR6249, miR5648 and miR5661 were predicted to regulate Cf4CL post-transcriptionally. However, this is only a preliminary analysis and further studies are needed to validate the predictions. Earlier, Nta-miR399a was predicted to target 4CL in response to topping in flue-cured Nicotiana tabacum L. (Solanaceae) roots (Guo, Kan, and Liu 2011). Cloning of 4CL gene from C. forskohlii provides an opportunity to understand and engineer the biosynthetic machinery for over production of these compounds as well as to block their synthesis in order to understand their roles in plant physiology.

CfCHS gene cloned in the study is an important key enzyme involved in the biosynthesis of flavonoids. CfCHS expression was found to be dominant in leaves and flower (Figure 5.10A). The data is consistent with the previous studies where highest expression is observed in aerial parts of the plant (Thain et al. 2002; Fritze et al. 1991). Since CHS plays an important role in flower pigmentation (van der Krol et al. 1990; Napoli, Lemieux, and Jorgensen 1990), maximum accumulation of CHS was observed in the corolla and young leaves of Antirrhinum majus L. (Plantaginaceae), where pigmentation is intense (Fritze et al. 1991).

Treatments with chemical elicitors are known to mimic the effects of environmental stresses (Tuteja and Sopory 2008). The changes in expression pattern of CfCHS, on treatment with various phytohormones/elicitors like MeJA, SA, ABA and 2, 4-D, were also determined. CfCHS expression was strongly induced in plants treated with MeJA (Figure 5.10B), which mimics herbivore attack or mechanical wounding (Creelman and Mullet 1997; Kessler and Baldwin 2002). MeJA can diffuse to distal parts of the plant (Ruiz-Medrano, Xoconostle-Cázares, and Lucas 2001; Karban et al.
Molecular cloning………..Coleus forskohlii

CHAPTER: 5

2000), thereby, mediating systemic signaling. The enhanced expression of CfCHS in response to MeJA would probably result in accumulation of plant secondary metabolites (Sánchez-Sampedro, Fernández-Tárrago, and Corchete 2005; Lei et al. 2010; Tamari et al. 1995). It has been reported that methyl jasmonate promoted silymarin production and enhanced CHS activity in cell cultures of Silybum marianum (L.) Gaertn. (Asteraceae) (Sánchez-Sampedro, Fernández-Tárrago, and Corchete 2005). MeJA also induced CHS expression in Petunia corollas (Tamari et al. 1995).

Abscisic acid is another important phytohormone that serves myriad functions ranging from developmental to adaptive stress responses (Cutler et al. 2010). ABA has been shown to promote synthesis of flavonoids and CHS expression in Ginkgo biloba (L.) (Ginkgoaceae) (Yan 2002; Shuiyuan et al. 2004; Li et al. 2014). In our studies, we found that ABA treatment marginally induces expression of CfCHS at the concentration of ABA and time-point we tested (Figure 5.10B). 2, 4-D is another phytohormone which suppresses anthocyanin (class of flavonoids) synthesis (Ban et al. 2003; Takeda et al. 1993), however, it has also been reported that anthocyanin synthesis, once induced by light, could not be suppressed by further addition of 2, 4-D (Takeda 1990). In our results also, we observed that 2, 4-D had a negligible effect on expression of CfCHS (Figure 5.10B). However, CfCHS appeared to be downregulated in response to SA treatment (Figure 5.10B). Our results are in concurrence with the previous studies performed in constitutive salicylic acid producing transgenic Nicotiana tabacum cv. Samsun NN (Solanaceae) plants, where CHS was suppressed, relative to the wild type plants (Nugroho, Verberne, and Verpoorte 2002).

CHS catalyzes the first committed step of flavonoid biosynthesis, which is an important regulatory branch point channeling flux of phenylpropanoid pathway towards biosynthesis of flavonoids. Several species of Coleus, including C. forskohlii are known to produce rosmarinic acid. Rosmarinic acid also utilizes p-Coumaroyl-CoA as a precursor. Downregulating/silencing CfCHS could help in diverting the metabolic flux towards the synthesis of alternate compounds of commercial value such as rosmarinic acid or other phenylpropene phytochemicals. For instance, downregulating the strawberry CHS resulted in enhanced production of phenylpropene aroma compounds like eugenol (Hoffmann et al. 2011). In view of this, an in silico search to find the microRNAs (miRNAs) that may target CfCHS was carried out. miR34, miR166, miR395, and miR477 were found to regulate CfCHS
expression post-transcriptionally (Figure 5.10A). In earlier studies of *Phelipanche ramosa* (L). Pomel (Orobanchaceae), it is reported that miR166 was down regulated in response to viroid infection (Ivanova et al. 2014). Flavonoids are already known to be involved in defense responses against pathogen infection (Jasiński et al. 2009; Päsold et al. 2010) and in a study, it was found that *CHS* expression was induced in Norway spruce after pathogen infection (Nina Elisabeth Nagy 2004). So, there could be some correlation between down regulation of microRNAs and up regulation of *CHS* gene during pathogen infection. However, further studies are needed to validate the predictions.