TiPS

Screening *Salvia macrosiphon* transcriptome for 4-coumarate CoA ligase enzyme coding genes

Rasool Asnafi[1], Shiva Hemmati[1,2,*]

[1]Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.
[2]Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

Abstract

4-Coumarate CoA ligase (4CL) is one of the key enzymes in the phenylpropanoid pathway, which converts cinnamic acid derivatives to active thioesters. Active thioesters are precursors of a vast number of secondary metabolites. *Salvia macrosiphon* as a valuable medicinal plant grows in Iran, Turkey, and Afghanistan. This plant produces pharmacologically active metabolites, including rosmarinic acid, flavonoids (apigenin, luteolin), sesquiterpenes, and coumarins. Identification of genes encoding 4cl provides the opportunity to manipulate the biosynthetic pathways and mediate the carbon flux toward the corresponding metabolites. Mature wild type *S. macrosiphon* plants were collected in flowering and seed-bearing stages. *In vitro* cultures of *S. macrosiphon* were established on Murashige & Skoog (1/2 MS) medium. To increase the possibility of finding 4cl isoforms in the transcriptome, cultures were elicited by chitosan. Since the genomic sequence of *S. macrosiphon* was not available, degenerate and CODEHOP primers were designed based on the identified 4CL protein sequences. Using the CODEHOP primers two isoforms of putative 4cl genes were identified in all organs of the wild type plant. Degenerate primers could only amplify the same genes from roots and seed-containing capsules. This might be due to a higher expression level of the genes in these organs. No 4cl isoform were detected from cultures, which might be due to the lower abundance of the transcript at early stages of *in vitro* plantlets. Phylogenetic analysis showed that the two isoforms of 4cl genes from *S. macrosiphon* and *Salvia miltiorrhiza* have been evolved from a common ancestor.

*Keywords*: *Salvia macrosiphon*, 4-coumarate CoA ligase, CODEHOP primers, Degenerate primers, Phylogeny.

1. Introduction

The phenylpropanoid pathway in plants provides precursors for a vast number of secondary metabolites. 4-coumarate Co:A Ligase (4CL) is an Enzyme in this pathway, which plays a pivotal role in branching the carbon flux toward a special secondary metabolite biosynthetic pathway. To enhance the reactivity of cinnamic acid and its hydroxylated derivatives, 4CL activates the carboxyl group of these compounds by

replacing -OH (a poor leaving group) with Coenzyme A and produces CoA thioesters in a two-step reaction (Figure 1) (1). The thioester products of 4CL are the precursors of a vast number of secondary metabolites, including lignins, flavonoids, stilbens, coumarins, and chalcones. Following the formation of *p*-coumaroyl CoA, a series of reactions are then initiated. Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT), couples *p*-coumaroyl CoA with shikimate and produces *p*-coumaroyl shikimic acid (2). This compound is then hydroxylated by p-coumaroyl shikimate/quinate 3'-hydroxylase (C3H) and produces caffeoyl shikimic acid,

Corresponding Author: Shiva Hemmati, Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.
Email: hemmatish@sums.ac.ir
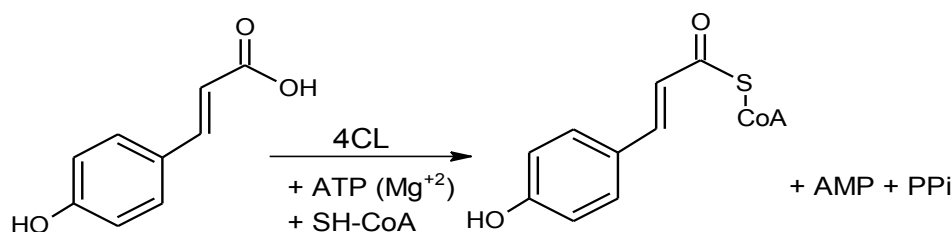
Figure 1. The reaction catalyzed by 4CL enzyme. A cinnamic acid derivative is activated by adenylation. The adenylate intermediate is attacked by a SH-CoA molecule to form the thioester product.

which in turn is converted by different enzymes to coniferyl alcohol and sinapyl alcohol. Together with p-coumaryl alcohol, these compounds are the precursors of lignin biosynthesis. Lignin is one of the most abundant biopolymers in nature, which protects plants against pathogen attacks and provides structural strength (2). p-coumaroyl CoA, the product of 4CL catalysis, might be used as a substrate for another reaction. It is condensed with 3 molecules of malonyl CoA by stilbene synthase and produces resveratrol. Resveratrol is a phytoalexin that protects plants from fungal attacks (3, 4). A number of studies have shown the beneficial effects of resveratrol for human health, including anti-carcinogenic and lipid-lowering effects (5). p-coumaroyl CoA is also able to condense with three molecules of malonyl CoA under the effect of chalcone synthase and aurone synthase, which results in the formation of hispidol and apigenin (6). These compounds are aurone flavonoids and serve as plant pigments. p-coumaroyl CoA, feruloyl CoA, and caffeoyl CoA are the precursors of coumarin biosynthesis. The carcinogenic effect of some carcinogens is modulated in the presence of coumarins, which might be due to their antioxidant activity (7). Coumarins are also important as antiretroviral compounds (8). Rosemarinic acid as a potent antioxidant, anti-inflammatory, antibacterial, and antiviral compound is produced from p-coumaroyl CoA and 4-hydroxyphenyllactate (9).

In most of the plant species studied so far, 4CL has more than one isoform. In parsley, for example, 4CL family has two members. There are 4 isoforms of 4CL in *Arabidopsis thaliana* (10), 4 isoforms in *Physcomitrella patens* (moss) (11), 5 isoforms in *Oryza sativa* (rice) (12), and 2 isoforms in *Petroselinum crispum* (parsley) (13). These isoforms show different affinities toward

different cinnamate derivatives. At4CL1 from *A. thaliana* converts 4-coumarate efficiently and has almost no affinity for ferulate and 5-hydroxyferulate. At4CL2 converts cinnamate and 4-coumarate and shows less affinity toward caffeate. At4CL3 converts cinnamate and caffeate more efficiently (14). At4CL4 has a rare ability to convert sinapate to sinapoyl CoA, while there is no such ability in the other 4CL isoforms of this plant (10). Although *in vitro* studies show that cinnamate, 4-coumarate, caffeate, ferulate, and sinapate all can be substrates of 4CL, *in vivo* studies have demonstrated that in the phenylpropanoid pathway only 4-coumarate is converted by 4CL. Whether the other products are actually produced in plants and what their roles might be is still under question.

In addition to substrate specificity, there is another interesting characteristic to this enzyme family, which is the distinct compartmental expression of each isoform. In *Ruta graveolens* (common rue), 4CL1 is mostly expressed in the flowers, while 4CL2 has been expressed in the shoots. In *Populus tremuloides* (aspen), there are two isoforms. Ptre4CL1 has been shown to be expressed in the developing xylem tissues, which correlate with lignin biosynthesis, while Ptre4CL2 has higher expression in non-lignin tissues (like epidermal cells) and is clearly involved in the biosynthesis of other phenolic compounds like flavonoids (15). In Rubus idaeus (raspberry), the transcription level of the 3 isoforms has been analyzed. While Ri4CL1 was expressed in leaves, Ri4CL2 was mainly expressed in shoots and Ri4CL3 transcripts had high abundance in different developmental stages of flowers and fruits (16).

4CL is a member of ANL superfamily. This superfamily consists of Acyl-CoA synthetases, adenylate forming domain of non-ribosomal
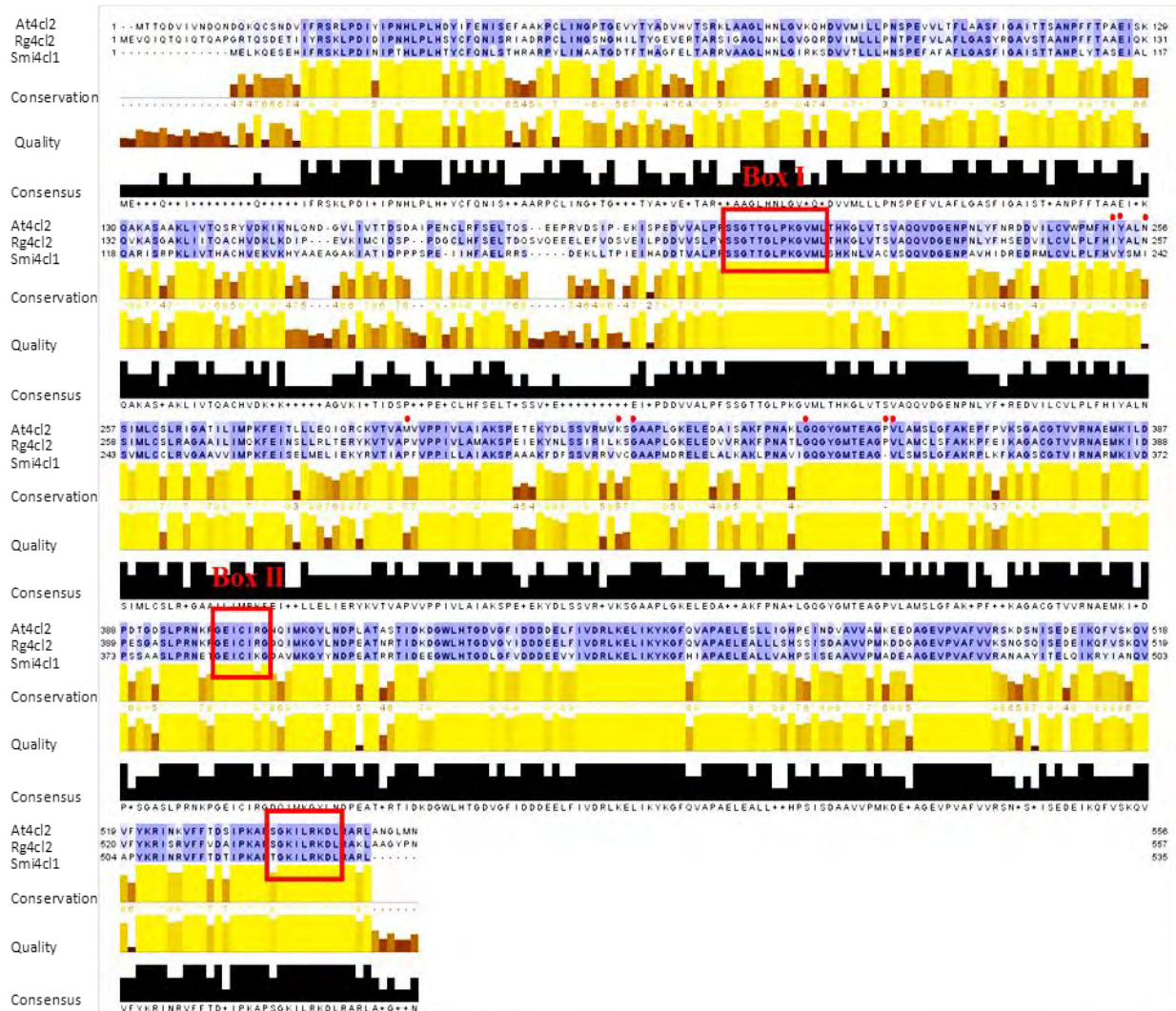
Figure 2. Multiple sequence alignment of *A. thaliana* 4CL2, *R. graveolens* 4CL2 and *S. miltiorrhiza* 4CL1. The conserved motifs (Box I and Box II) are shown. The third box containing GKILRKD sequence is the basis of designing a third pair of primers, SH102 and SH103. The residues marked by red bullets at the top of each row are the amino acids that define substrate specificity of 4CL (48).

peptide synthetases and luciferases. The members of this superfamily all form an adenylate intermediate to activate their substrate (17). The motif box I (Figure 2) with a peptide sequence consisting of "SSGTTGLPKGV" is a potential nucleotide-binding motif and is highly conserved not only in all 4CLs, but also in other members of ANL superfamily (17). The motif box II, usually consisting of GEICIRG, is more specific to the 4CL family (Figure 2). Some studies believe that the central cystein is involved in the catalysis, while there are evidences proving the otherwise. These motifs have been used to prime PCR amplification of 4cl gene from species that their genome has not been sequenced yet (16, 18).

Salvia macrosiphon a species of the Lamiaceae family grows mostly in Iran. It is a perennial plant, with a height of 60 cm and a distinct strong aroma (19). It flowers in May, and the seeds are ready to harvest in June. The seeds have always been of great importance in Persian traditional medicine as demulcent and antitussive agent. Analysis of the aerial parts has revealed the presence of different sesquiterpenes, flavonoids, and rosmarinic acid (20, 21). This suggests that different secondary metabolite biosynthetic pathways are active in this plant, so studying the characteristics of *S. macrosiphon* 4CL as the enzyme, which controls the carbon flux to a specific pathway is of special interest for genetic engineering purposes.

Table 1. GenBank accession numbers of plant species which have been used for phylogenetic analysis in this study.

Plant species	Number of 4CL isoforms	GenBank/EMBL accession number	Plant species	Number of 4CL isoforms	GenBank/EMBL accession number
Allium sativum	1	AEM44785.1	*Panicum virgatum*	2	ADZ96250.1
Arabidopsis thaliana	1	AAD47191.1	*Petroselinum crispum*	1	CAA31696.1
Arabidopsis thaliana	2	AAD47193.1	*Petroselinum crispum*	2	CAA31697.1
Arabidopsis thaliana	3	AAD47194.1	*Petunia x hybrida*	1	AEO52694.1
Arabidopsis thaliana	4	AAM19949.1	*Physcomitrella patens*	1	ABV60447.1
Bambusa emeiensis	1	ACA09448.1	*Physcomitrella patens*	2	ABV60448.1
Camellia sinensis	1	ABA40922.1	*Physcomitrella patens*	3	ABV60449.1
Cannabis sativa	1	AHA42444.1	*Physcomitrella patens*	4	ABV60450.1
Eucalyptus camaldulensis	1	AAZ79469.1	*Pinus radiata*	1	ACF35279.1
Galega orientalis	1	ACZ64784.1	*Pinus taeda*	1	AAB42383.1
Glycine max	1	AAL98709.1	*Pinus taeda*	2	AAB42382.1
Glycine max	2	AAC97600.1	*Plagiochasma appendiculatum*	1	AJT43268.1
Glycine max	3	AAC97599.1	*Populus tomentosa*	1	AAL02145.1
Glycine max	4	CAC36095.1	*Populus tremuloides*	1	AAC24503.1
Hibiscus cannabinus	1	ADK24217.1	*Populus tremuloides*	2	AAC24504.1
Hypericum calycinum	1	AFS60176.1	*Populustrichocarpa x Populus-deltoides*	1	AAC39366.1
Leucaena leucocephala	1	ACI23348.1	*Populustrichocarpa x Populus-deltoides*	2	AAC39365.1
Lithospermum erythrorhizon	1	BAA08365.1	*Prunella vulgaris*	1	AHY94891.1
Lithospermum erythrorhizon	2	BAA08366.2	*Prunus avium*	1	ADZ54779.1
Lolium perenne	1	AAF37732.1	*Pueraria montana var. lobata*	1	AGW16013.1
Lolium perenne	2	AAF37733.1	*Pueraria montana var. lobata*	2	AGW16014.1
Lolium perenne	3	AAF37734.1	*Rubus idaeus*	1	AAF91310.1
Melissa officinalis	1	CBJ23825.1	*Rubus idaeus*	2	AAF91309.1
Nicotiana tabacum	1	BAA07828.1	*Rubus idaeus*	3	AAF91308.1
Nicotiana tabacum	1	AAB18637.1	*Ruta graveolens*	1	ABY81910.1
Nicotiana tabacum	2	AAB18638.1	*Ruta graveolens*	2	ABY81911.1
Ocimum basilicum	1	AGP02119.1	*Salvia miltiorrhiza*	1	AAP68990.1
Ocimum tenuiflorum	1	ADO16242.1	*Salvia miltiorrhiza*	2	AAP68991.1
Oryza sativa	1	NP_001061353.1	*Solanum tuberosum*	1	AAD40664.1
Oryza sativa	2	NP_001047819.1	*Sorbus aucuparia*	1	ADF30254.1
Oryza sativa	3	NP_001046069.1	*Sorbus aucuparia*	2	ADE96996.1
Oryza sativa	4	NP_001058252.1	*Sorbus aucuparia*	3	ADE96997.1
Oryza sativa	5	NP_001061935.1	*Vanilla planifolia*	1	O24540.1
Panicum virgatum	1	ACD02135.1	*Zea mays*	1	NP_001105258.2

To our knowledge, the only species from the *Salvia genus* with the known 4CL enzyme is *Salvia miltiorrhiza* (Chinese sage) (22).

In this research, we tried to amplify 4cl copies from the transcriptome of different organs of the wild type *S. macrosiphon*. To compare the level of 4cl expression in *in vitro* cultures with the wild type plants qualitatively, shoot cultures of *S.*

Table 2. Primers for amplification of *S. macrosiphon* 4cls.

Name of the primer	Nucleotide sequence	Type of primer	Conserved amino acid motifs used for primer designing	Fold of degeneracy
SH100	TWY TCN TCN GGN ACN ACN GGN	Degenerate Forward	Y(F)SSGTTG*	16386
SH101	TAC TCC TCC GGG ACN ACN GGN	CODEHOP Forward	Y(F)SSGTTG	64
SH102	RTC YTT YCT NAR DAT NAT NTT NCC	Degenerate Reverse	GKILRKD**	12288
SH103	CAG ATC TTT TCT CAA DAT NTT NCC	CODEHOP Reverse	GKILRKD	64
SH104	NCC YCK DAT RCA DAT YTC NCC	Degenerate Reverse	GEICIRG***	2304
SH105	TCC YCT AAT GCA DAT YTC NCC	CODEHOP Reverse	GEICIRG	24

* The motif Box I. **This region was conserved in the alignment of 4CLs. *** The motif box II.

macrosiphon has been established. In addition, to induce the expression of the 4cl gene *in vitro*, plantlets were exposed to chitosan as an elicitor. A comparative approach has been used to evaluate the efficacy of degenerate and CODEHOP primer (Consensus-Degenerate Hybrid Oligonucleotide Primer) designing to amplify 4cl genes (23). We have conducted a phylogenetic analysis based on the peptide sequences of the plant 4CLs that have been characterized by enzymatic assays.

2. Materials and Methods
2.1. Chemicals

RNeasy® Plant Mini Kit was from Qiagen (Hilden/Germany). RevertAid first strand cDNA Synthesis Kit and PCR Master Kit were from Thermo Fisher Scientific (USA) and Cinnagen (Iran), respectively. GEL/PCR Purification Minikit was from Yekta Tajhiz (Iran). Synthesis of oligonucleotides was performed at MWG-Biotech (Martinsried, Germany). Murashige & Skoog (MS) media including vitamins were purchased from Duchefa Biochemie B.V (Harleem, The Netherlands). Benzylaminopurine (BAP) and Indole Acetic Acid (IAA) were gifts from A. Mohagheghzadeh, Shiraz, Iran. Chitosan was donated by Z. Amoozgar, Boston, USA.

2.2. Plant materials

Wild type *S. macrosiphon* plants and seeds were collected from the natural location, Roknabad, Shiraz, (4900 feet altitude) in May and June 2014, respectively. Plant samples were identified by S. Khademian and voucher specimens were deposited at the herbarium of School of Pharmacy, Shiraz University of Medical Sciences. To overcome the dormancy of *S. macrosiphon* seeds, they were scarified as mentioned earlier (24). Seeds were sterilized by detergent, ethanol, and sodium hypochlorite (25). To induce shoot cultures *in vitro*, sterilized seeds were planted on ½ MS medium (pH 5.6) without hormone and solidified with 1% Agar. Cultures were incubated in diurnal cycles of light/dark (16 hr light/8 hr dark) at 25 °C (26). The plantlets were sub-cultured to a new medium every three weeks and were harvested after three rounds of sub-cultivation. To evaluate the effect of growth regulators on the organogenesis, BAP, and IAA were used in a separate group of *in vitro* cultures (27). The final concentration of BAP and IAA was 2 and 0.75 μM, respectively. In the elicitor treated group, chitosan solution was added to the medium to a final concentration of 0.004% (w/v). Plantlets were then transferred to this medium, which after 72 hours they were harvested like the control group for RNA extraction.

Figure 3. Plantlets of *S. macrosiphon* on the ½ MS medium.

2.3. Primer design to amplify S. macrosiphon 4cl genes (Sma4cl)

Two isoforms of *S. miltiorrhiza* 4CL amino acid sequences were used as query to identify 4CL sequences with over 60% similarity by BLAST tools (GenBank accession numbers are listed in Table 1). Selected sequences were then aligned by Clustal program using blosum protein weight matrix with an open gap penalty of 10 and extended gap penalty of 0.05 (28). Three conserved areas (including box I and II) (Figure 2) were used to design degenerate and CODEHOP primers (23). The efficacy of both methods was then compared. Six sets of primers were designed to amplify *S. macrosiphon* 4cl (Sma4cl) genes (Table 2).

2.4. Isolation of RNA, cDNA synthesis, and amplification of Sma4cl genes

In vitro shoot cultures with and without chitosan treatment, along with young leaves, aerial parts (in flowering and seed-bearing stages), and roots of the wild type plants were separately frozen in liquid nitrogen. Total RNA was extracted from 100 mg of frozen homogenized tissue. cDNA was synthesized using 0.3 μg of the extracted RNA according to the manufacturer's protocol. One μL of the synthesized cDNA and primers with the final concentration of 0.7 μM were added to the master mix for a 25 μL reaction. Different combinations of primers were used. Degenerate forward primers were used with the two sets of degenerate reverse primers, and CODEHOP primers were used with each other (Table 2) (SH100+SH102, SH100+SH104, SH101+SH103, SH101+SH105). PCRs were performed in a gradient mode to find the best annealing temperature. The reaction was initiated at 95 °C for 2 min, followed by 35 cycles

Figure 4. Agarose gel electrophoresis of the PCR products from the root cDNA with degenerate primers SH100 and SH104. Lane 1: 1Kb DNA ladder, lane 2: Amplified Sma4cl1 and Sma4cl2 genes at Ta=46 °C.

of 35 sec at 93 °C, 30 sec at 53±7 °C ,and 1 min at 72 °C. Final extension was 72 °C for 3 min. The PCR product was separated by gel electrophoresis. Amplified fragments smaller than 1000 bp were isolated, purified, and sequenced. The result of nucleotide sequencing was translated to corresponding peptide sequence using ExPASy translate tool (29). A BLAST search was conducted to identify the most similar sequences to the isolated fragments from *S. macrosiphon*.

2.5. Phylogenetic analysis of 4CLs

The deduced amino acid sequences of 71 4CL enzymes, which have been proved by experimental data to have the ability to activate cinnamic acid derivatives were retrieved from protein databases. Using MEGA6 package a phylogenetic tree was constructed with neighbor-joining (NJ) method and 1000 bootstrap replicates (30).

3. Results
3.1. Establishment of S. macrosiphon in vitro cultures

Seeds of *S. macrosiphon* that were cultivated on ½ MS medium, without any growth regulator started to germinate in 10 days. After germination of seeds and formation of seedlings, plantlets with roots, shoots, and leaves have been formed under described conditions in methods (Figure 3). Molecular experiments have been conducted on this group of plantlets. Exposure of seedlings to BAP and IAA did not promote organogenesis. Nine days after addition of the hormones undifferentiated callus masses were observed, which has not been reported in other *Salvia* species (31, 32). When these callus cultures were transferred to a hormone-free ½ MS medium, they were differentiated.

3.2. Treatment of S. macrosiphon seedlings with chitosan

Plantlets that were treated with 0.004% w/v chitosan experienced senescence 72 hours following the exposure. The roots became brown atrophied, and the leaves turned yellow.

3.3. Amplification of Sma4cl genes

A total number of 4 pairs of primers tar-geting three regions in the 4cl gene were used for amplification of 4cl from *S. macrosiphon* cDNA (Table 2). In the roots and seed-bearing capsules of the wild type plant, combination of degenerate primers SH100 and SH104 corresponding to box I and box II motifs, respectively, could amplify two fragments with a size in the range of 500-750 bp, at 46 °C and 48.5 °C as annealing temperature (Figure 4). This set of degenerate primers did not result in gene amplification in other plant tissues. Combination of the corresponding CODEHOP primers, SH101 and SH105, amplified the same two segments in roots, seed-bearing capsules, and leaves with similar annealing temperatures. The CODEHOP primers SH101 and SH103 at an annealing temperature of 56.2 °C resulted in the amplification of a gene about 750 bp. Using different sets of primers, no DNA fragment could be amplified from untreated or treated in vitro cultures with chitosan.

Amplified products were separated by electrophoresis, which after purification, both bands smaller than 1000 bp was sequenced. The larger amplified band (Sma4cl1) was 95% similar to *S. miltiorrhiza* 4CL1 by blast analysis. The shorter fragment (Sma4cl2) was 88% similar to *S. miltiorrhiza* 4CL2. The rest of 100 top peptides lining up in the BLASTsearch were at least 65% similar to the query all from the 4CL family.

3.4. Phylogenetic analysis of Sma4CLs in relation to other 4CL enzymes

To analyze the relationship between the peptides proved to be capable of catalyzing cinnamic acid derivatives, the full lengths of those peptides that have been reported to possess the catalytic ability of converting cinnamic acid derivatives to subsequent CoA thioesters were retrieved. A total number of 71 peptide sequences were collected (Table 1). Together with the partial sequences that were obtained from *S. macrosiphon* (Sma-4CL1 and Sma4CL2), the phylogenetic tree was constructed. Three main clades were observed in the Hierarchichal tree (Figure 5). Clade I includes the cinnamate CoA ligase of the bacteria *Streptomyces coelicolor* and the 4CL of a liverwort. The only member of clade II is *A. thaliana* 4CL4, and clade III consists of the rest of the peptides includ-
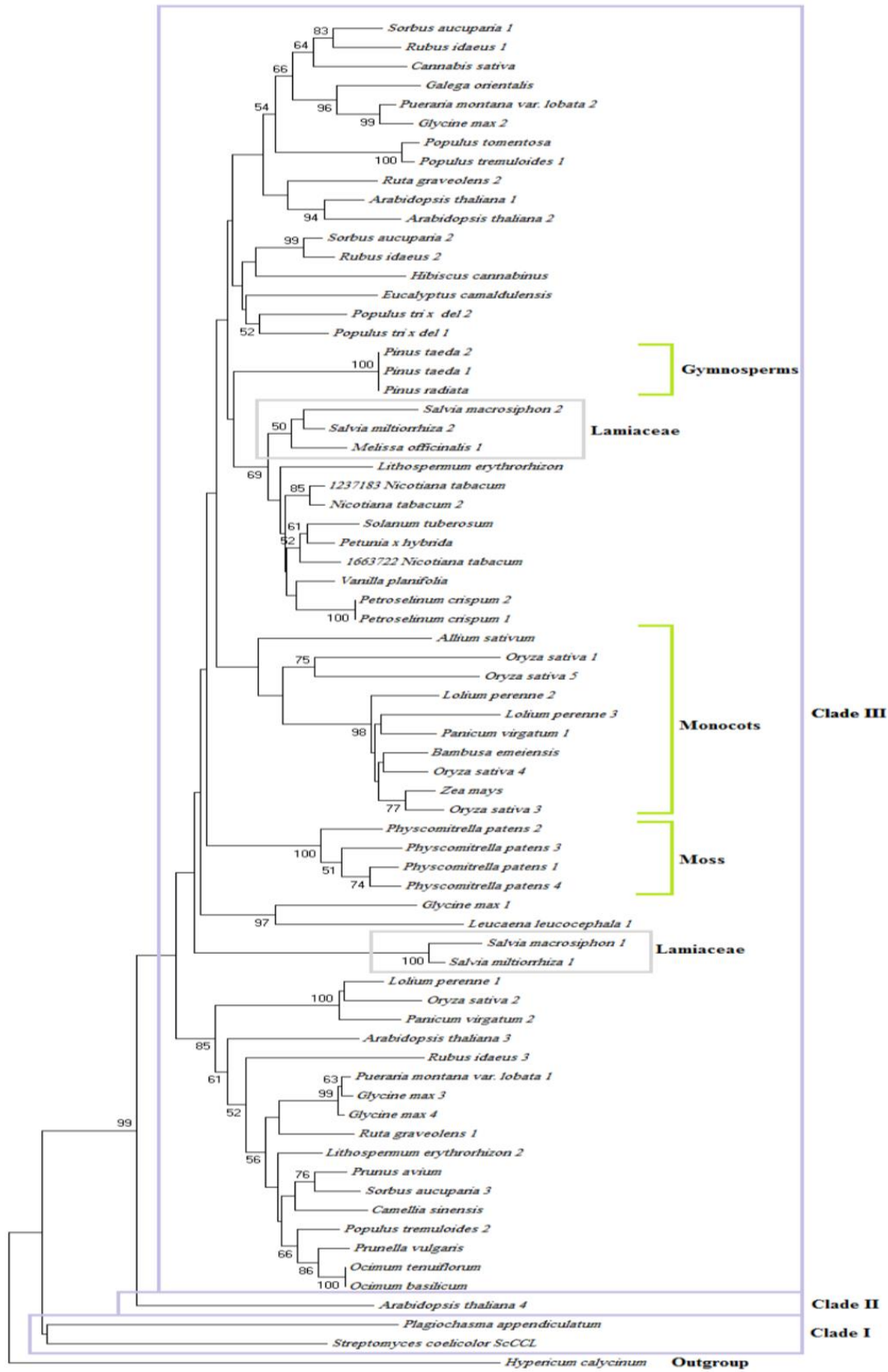
Figure 5. Phylogenetic tree of 4CL proteins. Three main clades are observed in the tree. In the clade III mosses, gymnosperms and monocots are shown by brackets. The two partial sequences of putative *S. macrosiphon* 4CLs determined in this study are gray highlighted in boxes named "Lamiaceae".

ing those obtained from a moss, gymnosperms, and angiosperms, all forming subclades of their own. Sma4CLs have been grouped with 4CLs of *S. miltiorrhiza* in two different clades.

4. Discussion

S. macrosiphon is a herbaceous perennial plant endemic to Iran, in which the activity of several secondary metabolite biosynthetic pathways have been observed (19). The identification of the role of Sma4CL isoenzymes, which channel the carbon flux into a specific pathway, is of particular interest. This provides the opportunity to manipulate each isoenzyme for the desired purposes.

4.1. Screening 4cl transcripts in the wild type and in vitro cultures of S. macrosiphon

We have established *in vitro* cultures of *S. macrosiphon* on ½ MS medium with 3% sucrose as a suitable medium for *S. macrosiphon* organogenesis. No additional growth regulators were needed for shoot culture. Chitosan as an elicitor caused senescence to *S. macrosiphon* plantlets. Chitosan is a deacetylated derivative of chitin, which has been used abundantly in plant cell culture (33). There are several factors which impact chitosan cytotoxicity, including the degree of acetylation, pH, solvent, concentration, type of plant cell, and the duration of exposure (33). The toxic concentration depends on the plant species and varies from 0.006% for *Pinus elliottii* to 0.01% for *Mimosa pudica* (33, 34). Duration of exposure varied from 3 hrs for *M. pudica* to 9 days for *Nicotiana tabacum* and *Eschscholtzia californica* (33, 35). The cytotoxicity mechanisms of chitosan might be due to the positively charged nature of chitosan that binds to negatively charged DNA (36) as well as ROS formation and apoptosis (37), pore formation in plasma membrane of plant cells and electrolyte leakage (38). The appearance of brown spots on the leaves of 14-days-old seedlings has been associated with the use of chitosan in *O. sativa* seedlings (39). In our analysis, chitosan concentration and exposure time affected cell viability more than it could provoke a transient rise in 4cl transcription as a defense mechanism of the plant. Although, the toxic concentration of chitosan for most plant cultures was higher than the concentration used in

this study (36, 39, 40), lower concentrations and shorter exposure times of chitosan treatment might react differently.

Plants grown in *in vitro* cultures are under imposed conditions, which in turn might act as a stress factor. In the wild type plants that were collected from nature, several defense mechanisms were needed to cope with the environmental stresses. Since the controlled condition of the culture may have reduced the need of the plantlets to secondary metabolites, the expression level of 4cl genes was undetectable. It seems that in *in vitro* culture plantlets, the level of 4cl transcripts may have been very low, since other studies have reported similar observations for other cell cultures (41, 42). It is shown that under different conditions of culture, and depending on the age of culture the level of 4cl mRNA varies significantly. In addition, using a pool of nonspecific primers reduces the optimal concentration of the two oligonucleotides in the primer pool. This reduces the ability to bind to the target gene specifically. When the low primer concentration is exposed to a PCR mix that contains low concentration of target transcript, the chance of successful gene amplification would be lower. Elicitation with chitosan was not different from control group of *in vitro* plantlets in the ability to amplify 4cl copies. It has been reported that chitosan as an elicitor results in a rise in the level of gene expression transiently (35, 43). However, appropriate sampling time is of crucial importance. In *Sorbus aucuparia* cell suspension culture, the level of all three 4cl transcripts were raised 6 hours after addition of the chitosan as an elicitor, while the level of transcripts of all Sa4cl isoforms were undetectable in the control group. On the other hand, elicitors might not have been effective, since not every copy of 4cl is inducible by elicitor, and not just any elicitors can provoke a rise in the gene expression level. When cell suspension cultures of *S. aucuparia* were exposed to light, only the level of Sa4cl3 transcript was increased and the other two isoforms remained undetectable. At4cl3 in *A. thaliana* was induced by UV irradiation more strongly than At4cl1 and At4cl2. The treatment with *Phythophthora parasitica* spores induced At4cl1 and At4cl2 mRNA accumulation and had no effect on At4cl3 (14).

Using the two conserved motifs (box I & II), we could amplify two copies of 4cl genes from *S. macrosiphon* leaves, root and seed-containing capsules. The plant was collected from a rather dry soil. Drought, pathogen, sunlight, exposure to insects, and probable wounding are some natural elicitors, which induce the phenylpropanoid pathway. In addition to natural stresses, the stage of the plant growth has a crucial impact on 4cl expression. Sma4cls display a compartmental expression. Although the two isoforms were detected in the leaves, root, and seed-bearing capsules of the plant, the same two isoforms were not identified in flowering aerial parts. This observation strongly suggests a distinct pattern of expression of Sma-4cls in different plant organs and stages of growth. Similar observations have been reported for isoenzymes of other species (44). The two putative Sma4cls do not show selectivity of expression in different organs over one another. Although, this needs further examination.

4.2. Comparison of degenerate vs. CODEHOP primers to amplify Sma4cl genes

Designing primers for genes from species whose genome sequencing projects have not been conducted yet, might be rather complicated. Specific protein families with conserved motifs provide possibilities to design specific primers accordingly. To target the nucleotide sequence that encodes a certain motif in a peptide, the concept "Degeneracy" is proposed so all the probable nucleotides that may be placed in a certain residue are determined. Therefore, a number of oligonucleotides based on this degeneracy will be designed. This is helpful and challenging at the same time (23). The result is not a single oligonucleotide, but a pool of oligonucleotides in which only one of them can match the target gene specifically. Unspecific oligonucleotides as primers may start replication of unwanted genes that might belong to the same gene family. In addition, when the number of primers increases, the concentration of a single primer will be reduced. Using a CODEHOP primer, where the 3' end of the primers are degenerate and the 5' end is the most probable sequence, it is possible to keep the degeneracy to a certain level, and still have a chance not to repli-

cate some unwanted genes (23). In our study, this class of primers proved to be efficacious. In all the plant organs that degenerate primers resulted in gene amplification, CODEHOP primers worked as well. In some organs that the transcription of the gene was low, like in the leaves of the mature plant, degenerate primers could not amplify any genes, where CODEHOP primers resulted in amplification. For example degeneracy of the degenerate primer SH100 was 16386 while degeneracy of the CODEHOP equivalent of the same primer (SH101) was 64. In all cases a final concentration of 0.7 μM of primers were used. So the final concentration of each oligonucleotide in the degenerate primer pool of SH100 was 0.043×10^{-3} μM, while a single oligonucleotide's concentration in CODEHOP primer pool of SH101 was 0.011 μM, which is 256 times higher. Although the consensus 5' ends of the CODEHOP primers were probably not completely matching the gene sequence that did not lessen the efficacy of the CODEHOP primers. The degenerate 3' end had it covered. In general, we found CODEHOP primer designing method more valuable, since it was as effective as degenerate primers with lower financial costs.

4.3. Phylogenetic analysis of 4CL enzymes

Determining a certain criterion to identify true 4CLs from databanks is critical. Some articles have defined the presence of the conserved motifs (box I and box II) as their main criterion, some have added the segment QGYGMTEA to the previous motifs, and some have performed the experiment to find the true peptide sequence (45). To exclude pseudo-positive and pseudo-negative results and to include all the 4CL sequences that have been proved by the experiment as functional enzymes, 71 peptides truly registered as 4CL by enzyme assays in literature were mined out for phylogenetic analysis (Table 1).

Four main clades are observed in the phylogenetic tree (Figure 5). In Clade I, *S. coelicolor cinnamate* CoA Ligase (ScCCL) is grouped with the 4CL of liverwort *Plagiochasma appendiculatum*. The bacterial enzyme has unknown role and has more affinity towards cinnamate than 4-coumarate, which is not common within plant 4CLs (46). Since bacteria evolved earlier than other spe-

cies, it is therefore deduced that Scccl gene has been transferred to the genome of the liverwort by a horizontal gene transfer and then has evolved functionally. In comparison to ScCCL, Pl4CL is capable of converting caffeate and 4-coumarate more efficiently, since the substrate binding pocket has been enlarged (47). The only member of the clade II in the tree is *A. thaliana* 4CL4. This enzyme is distinct for converting sinapate in addition to other substrates.

In the clade III, P. patens 4CLs and 4CLs from gymnosperms have been separated from those of angiosperms (Figure 5). In angiosperms, two classes were observed in our tree. Hamberger and Hahlbrock have concluded that this pattern is the result of three gene duplications (10). At4CL1 and At4CL2 separated as a result of recent gene duplications, At4CL3 is the oldest functional copy, and At4CL4 has evolved recently. There are two subclades in clade III that consist of 4CLs of the Lamiaceae family. In one subclade, *S. miltiorrhiza* 4CL1, Sma4CL1 (partial fragment from this study) and *Melissa officinalis* 4CL1 have grouped together. In another subclade, Sma4CL2 (the other partial fragment from this study) have been grouped together. This interesting pattern strongly suggests that there has been two isoforms of 4CL in the ancestor of the two species that probably have emerged by a gene duplication and further sub-functionalization of the second isoform. When the two species diverged, they both inherited the two copies of the gene. *S. miltiorrhiza* 4CL2 has gained more functions during evolution and is the main 4CL of *S. miltiorrhiza* (22). Since the two species have appeared so close to each other, we therefore predict that the similar pattern of function might exist in *S. macrosiphon*. Gene duplication has occurred before the divergence of monocots and dicots (10). Although monocots and dicots do not always appear in completely different clades, the presence of a clade consisting of only monocots is remarkable. The clade consists of 4CLs of *O. sativa*, *Allium sativum*, *Zea mays*,

Lollium perenne, *Panicum virgatum* and *Bambusa emeiensis*. *Petroselinum crispum* 4CL1 and 4CL2 are very similar. Since their substrate specificity profile is also very similar to each other. It has been concluded that they have evolved by recent gene duplication (13). Another interesting pattern in the tree is the relationship between *R. idaeus* and *S. aucuparia*, from the Rosaceae family. They both have three isoforms. It is speculated that they have diverged from a common ancestor, when after two rounds of gene duplications three isoforms of 4CL emerged. The three isoforms then were passed to the two species *R. idaeus* and *S. aucuparia*. The situation of Ri4CL1 and Sa4CL1, Ri4CL2 and Sa4CL2, and Ri4CL3 and Sa4CL3 in the clade III proves this theory

5. Conclusion

4CL is a pivotal enzyme in the phenylpropanoid pathway, whose products channel to a vast number of specialized metabolite biosynthetic pathways. Two partial sequences of putative 4cl genes were mined out from wild type *S. macrosiphon* (Sma4cl1 and Sma4cl2) with degenerate and CODEHOP primers. The phylogenetic analysis of characterized plant 4CLs showed that *S. macrosiphon* and *S. miltiorrhiza* have inherited their copies of 4cl genes from a common ancestor. The identification of two copies of 4cl genes in the genome of *S. macrosiphon* provides the opportunity to genetically manipulate the plant for the production of the desired metabolite both *in vitro* and *in vivo*.

Conflict of interest

None declared.

6. References

1.	Pietrowska-Borek M, Stuible H-P, Kombrink E, Guranowski A. 4-Coumarate: coenzyme A ligase has the catalytic capacity to synthesize and reuse various (di) adenosine polyphosphates. *Plant Physiol.* 2003;131:1401-10.
2.	Weng J-K, Li X, Stout J, Chapple C. Independent origins of syringyl lignin in vascular

plants. *Proc Natl Acad Sci USA.* 2008;105:7887-92.

3. Reinprecht Y, Yadegari Z, Perry GE, Siddiqua M, Wright LC, McClean PE, *et al. In silico* comparison of genomic regions containing genes coding for enzymes and transcription factors for the phenylpropanoid pathway in *Phaseolus vulgaris* L. and *Glycine max* L. *Front Plant Sci.* 2013;4:317.

4. Hain R, Bieseler B, Kindl H, Schröder G, Stöcker R. Expression of a stilbene synthase gene in *Nicotiana tabacum* results in synthesis of the phytoalexin resveratrol. *Plant Mol Biol.* 1990;15:325-35.

5. Beekwilder J, Wolswinkel R, Jonker H, Hall R, de Vos CR, Bovy A. Production of resveratrol in recombinant microorganisms. *J Appl Environ Microbiol.* 2006;72:5670-72.

6. Farag MA, Deavours BE, de Fátima Â, Naoumkina M, Dixon RA, Sumner LW. Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in *Medicago truncatula* cell cultures. *Plant physiol.* 2009;151:1096-1113.

7. Scheel LD. The biological action of the coumarins. In Kadis S, Ciegler A, Ajl SJ, Editor: Microbial Toxins. Vol VIII. New York-London; Academic Press; 1972. P. 47-66.

8. Garro HA, Pungitore CR. Coumarins as Potential Inhibitors of DNA Polymerases and Reverse Transcriptases. Searching New Antiretroviral and Antitumoral Drugs. *Curr Drug Discov Technol.* 2015;12:66-79.

9. Petersen M, Simmonds MS. Rosmarinic acid. *Phytochemistry.* 2003;62:121-5.

10. Hamberger B, Hahlbrock K. The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc Natl Acad Sci USA*. 2004;101:2209-14.

11. Silber MV, Meimberg H, Ebel J. Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. *Phytochemistry.* 2008;69:2449-56.

12. Hamberger B, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the Populus lignin toolbox and conservation and diversification of angiosperm

gene families. *Can J Bot*. 2007;85:1182-1201.

13. Lozoya E, Hoffmann H, Douglas C, Schulz W, Scheel D, Hahlbrock K. Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate: CoA ligase genes in parsley. *Eur J Biochem.* 1988;176:661-7.

14. Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* 1999;19:9-20.

15. Hu W-J, Kawaoka A, Tsai C-J, Lung J, Osakabe K, Ebinuma H, *et al.* Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). *Proc Natl Acad Sci USA.* 1998;95:5407-12.

16. Kumar A, Ellis BE. 4-coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. *Plant Mol Biol.* 2003;51:327-40.

17. Gulick AM. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol*. 2009;4:811-27.

18. Endler A, Martens S, Wellmann F, Matern U. Unusually divergent 4-coumarate: CoA-ligases from *Ruta graveolens* L. *Plant Mol Biol.* 2008;67:335-46.

19. Gohari AR, Ebrahimi H, Saeidnia S, Foruzani M, Ebrahimi P, Ajani Y. Flavones and flavone glycosides from *Salvia macrosiphon* Boiss. *Iran J Pharm Res.* 2011;10:247-51.

20. Moghddam FM, Farimani MM, Taheri S, Tafazoli M, Amin G. Chemical constituents from *Salvia macrosiphon. Chem Nat Comp.* 2008;44:518-9.

21. Shekarchi M, Hajimehdipoor H, Saeidnia S, Gohari AR, Hamedani MP. Comparative study of rosmarinic acid content in some plants of Labiatae family. *Phcog Mag.* 2012;8:37-41

22. Zhao SJ, Hu ZB, Liu D, Leung FC. Two Divergent Members of 4-Coumarate: Coenzyme A Ligase from *Salvia miltiorrhiza* Bunge: cDNA Cloning and Functional Study. *J Integr Plant Biol.* 2006;48:1355-64.

23. Staheli J, Boyce R, Kovarik D, Rose T. CODEHOP PCR and CODEHOP PCR Primer Design. In Park DJ, editor: PCR Protocols. Methods in Molecular Biology. Totowa, New Jersey; Hu-

mana Press; 2011. p. 57-73.

24. Mohagheghzadeh A, Hemmati S, Alfermann AW. Quantification of aryltetralin lignans in *Linum album organs* and *in vitro* cultures. *Iranian Journal of Pharmaceutical Sciences.* 2006;2:47-56.

25. Mohagheghzadeh A, Hemmati S, Mehregan I, Alfermann AW. *Linum persicum*: Lignans and placement in Linaceae. *Phytochem Rev.* 2003;2:363-9.

26. Hemmati S, Schmidt TJ, Fuss E. (+)-Pinoresinol/(-)-lariciresinol reductase from *Linum perenne* Himmelszelt involved in the biosynthesis of justicidin B. *Febs Lett.* 2007;581:603-10.

27. Mohagheghzadeh A, Dehshahri S, Hemmati S. Accumulation of lignans by *in vitro* cultures of three *Linum* species. *Z Naturforsch C.* 2009;64:73-6.

28. Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947-8.

29. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012;40:597-603.

30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725-9.

31. Avato P, Fortunato IM, Ruta C, D'Elia R. Glandular hairs and essential oils in micropropagated plants of *Salvia officinalis* L. *Plant Sci J.* 2005;169:29-36.

32. Misic D, Grubisic D, Konjevic R. Micropropagation of *Salvia brachyodon* through nodal explants. *Biol Plant.* 2006;50:473-6.

33. Amborabé B-E, Bonmort J, Fleurat-Lessard P, Roblin G. Early events induced by chitosan on plant cells. *J Exp Bot.* 2008;59:2317-24.

34. Lesney MS. Growth responses and lignin production in cell suspensions of *Pinus elliottii* 'elicited'by chitin, chitosan or mycelium of *Cronartium quercum* f. sp. fusiforme. *Plant Cell Tissue Organ Cult.* 1989;19:23-31.

35. Brodelius P, Funk C, Häner A, Villegas M. A procedure for the determination of optimal chitosan concentrations for elicitation of cultured plant cells. *Phytochemistry.* 1989;28:2651-4.

36. Hadwiger LA, Beckman JM. Chitosan as a component of pea-*Fusarium solani* interactions. *Plant Physiol.* 1980;66:205-11.

37. Vasil'ev L, Dzyubinskaya E, Zinovkin R, Kiselevsky D, Lobysheva N, Samuilov V. Chitosan-induced programmed cell death in plants. *Biochemistry (Mosc).* 2009;74:1035-43.

38. Young DH, Kauss H. Release of calcium from suspension-cultured Glycine max cells by chitosan, other polycations, and polyamines in relation to effects on membrane permeability. *Plant Physiol.* 1983;73:698-702.

39. Agrawal GK, Rakwal R, Tamogami S, Yonekura M, Kubo A, Saji H. Chitosan activates defense/stress response (s) in the leaves of *Oryza sativa* seedlings. *Plant Physiol Biochem.* 2002;40:1061-9.

40. Kendra DF, Hadwiger LA. Characterization of the smallest chitosan oligomer that is maximally antifungal to *Fusarium solani* and elicits pisatin formation in *Pisum sativum*. *Exp Mycol.* 1984;8:276-81.

41. Lindermayr C, Möllers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, *et al.* Divergent members of a soybean (*Glycine max* L.) 4-coumarate: coenzyme A ligase gene family. *Eur J Biochem.* 2002;269:1304-15.

42. Gaid MM, Scharnhop H, Ramadan H, Beuerle T, Beerhues L. 4-Coumarate: CoA ligase family members from elicitor-treated *Sorbus aucuparia* cell cultures. *J Plant Physiol.* 2011;168:944-51.

43. Popp MP, Lesney MS, Davis JM. Defense responses elicited in pine cell suspension cultures. *Plant Cell Tissue Organ Cult.* 1997;47:199-205.

44. Hemmati S, von Heimendahl CB, Klaes M, Alfermann AW, Schmidt TJ, Fuss E. Pinoresinol-lariciresinol reductases with opposite enantiospecificity determine the enantiomeric composition of lignans in the different organs of *Linum usitatissimum* L. *Planta Med.* 2010;76:928-34.

45. Gao H, Guo DM, Liu WJ, Ran JH, Wang XQ. Evolution of the 4-coumarate: coenzyme A ligase (4CL) gene family: Conserved evolutionary pattern and two new gene classes in gymnosperms. *J Systemat Evol.* 2012;50:195-205.

46. Kaneko M, Ohnishi Y, Horinouchi S. Cinnamate: coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3 (2). *J Bacteriol.* 2003;185:20-7.

47. Gao S, Yu HN, Xu RX, Cheng AX, Lou H-X. Cloning and functional characterization of

Rasool Asnafi *et al.*

a 4-coumarate CoA ligase from liverwort *Pla-giochasma appendiculatum*. *Phytochemistry.* 2015;111:48-58.

48.   Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, *et al.* The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proc Natl Acad Sci USA.* 2003;100:8601-6.