



Chapter 5

Isolation and characterization of cinnamate-4-hydroxylase (C4H) in *Helichrysum*. *aureonitens*

5.1 Abstract

Cinnamate 4-hydroxylase (C4H) is the second key enzyme of the phenylpropanoid pathway, and synthesizes numerous secondary metabolites that participate in development adaptations. Its function is the hydroxylation at the C-4' position on the B ring. One C4H copy was isolated and cloned from *Helichrysum aureonitens*. It has 1518-base pairs (including the stop codon, TAA) the open reading frame encoding a 506-amino-acid polypeptide. It shows the highest homology to *Echinacea angustifolia* from the Asteraceae's C4H, with 83.6 % identity on the nucleotide level but 93 % identity and 97 % positives on the protein level. The genomic DNA sequence of isolated C4H from *H. aureonitens* indicates the presence of three introns with a longer size compared to the *Arabidopsis thaliana* C4H gene structure. The presence of the first intron has not been reported in the C4H gene from any other species and it is be therefore a new finding from the isolated C4H of *H. aureonitens*.

5.2 Introduction

The second enzyme in the phenylpropanoid pathway, cinnamate 4-hydroxylase (C4H), produces a large number of biologically important secondary metabolites. (Chen et al., 2007). C4H converts *trans*-4-cinnamate to *trans-p*-coumarate by catalyzing the hydroxylation at the C-4' position of the B ring. C4H belongs to the P450s family (Russel and Conn, 1967). The cDNA sequence of C4H has been isolated from different plant species and is called CYP73. According to the P450 enzyme database (<http://dranderson.utmem.edu/public/73.htm1>), 20 different CYP73 members have been identified so far and include alfalfa, *Arabidopsis* and artichoke (Fahrendorf and Dixon, 1993; Mizutami et al., 1997; Teutsch et al., 1993). In cDNA libraries 16 ESTs have been found but pea, *Arabidopsis* and parsley seemingly have only a single gene encoding C4H (Frank et al., 1996; Bell-Lelong et al., 1997; Koopmann et al., 1999).

Galangin is one of the flavonoids produced in *H. aureonitens*. The main difference between the chemical structure of galangin and other members of flavonols is in the existence of the OH group at the C-4' position of the aromatic ring B. According to the results in the previous chapter, one of the possibilities for producing galangin in *H. aureonitens* is the presence of an alternative biosynthetic pathway in which cinnamate activates the 4CL enzyme as a substrate acid to produce cinnamoyl-CoA and not *p*-coumaric via C4H. This means that in the production of galangin in *H. aureonitens*, the C4H enzyme is not involved in this step (Scheme 3.2). Since kaempferol was also detected in the leaf ethanolic extracts of *H. aureonitens* in which the C-4' on ring B is hydroxylated (Fig. 4.2), it is proposed that C4H could be responsible for the hydroxylation at the 4' position to convert galangin to kaempferol. Based on the results obtained in the previous chapters, *p*-coumaric acid, the product of the activity of C4H in plants due to the

4'-OH biosynthetic pathway for flavonols, was absent in the leaf ethanolic extract of *H. aureonitens*.

It was also indicated that some 4'-OH intermediates in this pathway were not detected while some non 4'-OH intermediates were detected which belong to the non 4'-OH biosynthetic pathway (without C4H). The first step in this part of the investigation is the isolation of C4H from *H. aureonitens* leaf samples. The following objectives were set for this part of the study:

- Is C4H present in *H. aureonitens*?
- Characterization of C4H in *H. aureonitens*.
- What is the relationship between C4H in *H. aureonitens* and other plants?



5.3 Materials and methods

5.3.1 Plant materials

Fully expanded mature leaves of intact plants grown in the greenhouse and cells from the cell suspension cultures of *H. aureonitens* were used as plant materials for the gene isolation experiment in this part of the project.

5.3.2 Genomic DNA extraction

Harvested cells from a cell suspension culture of *H. aureonitens* (100 mg) were ground in liquid nitrogen using a mortar and pestle. Genomic DNA was extracted using a modified CTAB-based method (cetyl-trimethyl-ammonium-bromide) according to Gawel and Jarret (1991). To remove any RNA contamination, the extracted DNA was treated with 100 µg/ml DNase free RNase and the DNA was recovered with 96 % ethanol (2-times volume) with sodium-acetate (NaAC) (3M, pH 6.8) (1/10 volume) added. The DNA was finally washed with 70 % ethanol. Genomic DNA was dissolved in sterile distilled water and stored at -20 °C until used. The quality and concentration of DNA was determined by running the DNA on a 1 % agarose gel to visually test for the purity, and a spectrophotometer (Nanodrop) analysis (A260/280 ratio) was used to determine the quantity and purity.

5.3.3 PCR amplification of an internal fragment of C4H using degenerated primers

To amplify an internal fragment of the C4H gene from *H. aureonitens* two degenerated primers were designed. The upstream primer [5'-AGAAGGGCGAGATCAACGA (A/G) GA (C/T) AA (C/T) GT-3'] and the downstream primer [5'-GGCGTCGTGCAGGTTCAT (A/G) TG (A/C/G/T) GG-3'] were designed using CODEHOP software

(<http://blocks.fhrc.org/codehop.html>) based on the conserved area of C4H homologues from some closely related plant species from the Asteraceae family (Table 5.1) with amino acid sequences deposited in NCBI database according to Table 5.1. Primer pairs were designed to amplify approximately 500 bp fragment of the C4H gene. A PCR amplification of the internal fragment of C4H was carried out with the genomic DNA. The PCR reaction was performed in a 50 µl reaction mixture containing 100-150 ng genomic DNA, 1.5 µl mM MgCl₂ (50 mM), 5 µl 10x *Taq* DNA polymerase buffer, 4 µl dNTP (2.5 mM of each), 1 µl of 20 µ M of each primer and 0.2 µl *Taq* polymerase (5 U/µl, Bio line, Germany). DNA samples were initially denatured for 4 minutes at 94 °C and amplified in two steps with 10 and 25 cycles. DNA denaturation was carried out at 94 °C for 30 seconds, the primer annealing for 30 seconds at 50 °C, and the extension for 1:30 minutes at 72 °C. The second step was done using the same program and increasing the extension period with 10 seconds in each cycle. Cycling was completed with a final extension for 10 minutes at 72 °C in a thermocycler (BioRad icycler-USA). Amplified products were separated on a 1 % (w/v) agarose gel containing 0.5 µg/ml ethidium bromide. The ethidium-bromide stained amplification products were visualized and imaged on a UV transilluminator.

Table 5. 1. List of plants for designing the degenerated primers to amplify the internal fragment in C4H.

Scientific name	Family name	Gene bank accession no.
<i>Helianthus tuberosum</i>	Asteraceae	Z17369
<i>Zinnia violacea</i>	Asteraceae	U19922
<i>Gossypium arboreum</i>	Malvaceae	AF286648
<i>Nicotiana tobacco</i>	Solanaceae	DQ350353
<i>Solanum tuberosum</i>	Solanaceae	DQ341174
<i>Petroselinum crispum</i>	Apiaceae	L38898.1
<i>Arabidopsis thaliana</i>	Brassicaceae	NM_128601

5.3.4 Gel recovery and TA cloning

The DNA of the PCR product on the agarose gel was recovered, purified and eluted using the GFX™ PCR DNA and gel band purification kit (GE Healthcare, UK). The quality and concentration of the DNA was determined on a 1 % agarose gel and a UV/VIS spectrophotometer (Nanodrop-ND1000-USA) was used to determine the A₂₆₀/A₂₈₀ ratio. Gel-eluted DNA was cloned into the TOPO TA vector (Invitrogen, USA) and *E. coli* JM 109 competent cells were transformed with the vector DNA (Invitrogen, USA) by a heat shock treatment based on the manufacturer's instructions. Distinct kanamycin-resistant colonies derived from a Luria-Bertani medium containing plate (LB, 1 % tryptone, 0.5 % yeast extract and 0.5 % NaCl) were inoculated in a liquid LB medium overnight and the plasmids were purified from transformed cells using a commercial plasmid isolation kit (Fermentas, Europe). The plasmids were digested with the restriction enzyme *EcoRI* in a reaction containing 1 µl, 2 µl *EcoRI* and 10x reaction buffer in a total volume of 20 µl. Digestion was performed at 37 °C for 1 hour, and the digested DNA was run on a 1 % (w/v) agarose gel to visualize.

5.3.5 Sequencing and alignment

To confirm the presence of the right insertion and determine the sequence of fragments, the plasmids were sequenced using a M13 universal primer. Sequencing of the insert was carried out by using the BigDye Terminator Cycle Sequencing Ready Reaction Kit, v 4.1 (Perkin Elmer, Applied Biosystems, USA). A 10 µl sequencing PCR mixture reaction containing 2 µl BigDye ready reaction mix, 1 µl 5X dilution buffer, 10 pmol M13 forward or reverse primer (Table 5.2) and 200-300 ng DNA was used for this. The PCR sequencing reactions were performed using denaturation at 96 °C for 10 seconds, primer annealing at 50 °C for 5 seconds and DNA

extension at 60 °C for 4 minutes. The sequencing PCR product was then cleaned by using a Sephadex column (6.65 %) and the DNA was dried before submitting it to the DNA sequencer (3130/xl genetic analyzer, Applied biosystems, Foster City, USA). The NCBI programs, Blastn, blastx and Blastp were used for sequence alignment.

Table 5. 2. List of primers for the identification and amplification of the C4H gene in *Helichrysum aureonitens*.

Primer	Sequence	Comments
Adaptor 1	5'-GGATCCTAATACGACTCACTATAGGGC-3'	Adaptor primer 1
Adaptor 2	5'-AATAGGGCTCGAGCGGC-3'	Adaptor primer 2
CH1-F	5'-AACCTTATGGTCTATCGAAT-3'	Primary first walking primer
CH1-R	5'-AGATTTTATCCGTGTACTAA-3'	
CH1-NF	5'-ATGGCTATTCCTCTTTTAGT-3'	Nested first walking primer-FW
CH1-NR	5'- TTCTCAACAATGTAAAGAAC-3'	
CH2-NR	5'-CATCAAATGATCATGCAAGCCTAC-3'	
CH3-NR	5'-CGCTGCTGCCGATTCTTCTT-3'	
CH4-NR	5'-GACCGAAATTTTTGGCGTAATCGGT-3'	
CH5-NR	5'-CCCGAAGATGGGGACGGGGATAA-3'	
CH6-NR	5'-ACGGGGATTGGACCCGGTGGAA-3'	
C4H-F	5'-ATGGATCTACTCCTTTTGGAGAAA-3'	} To amplify full length fragment
C4H-R	5'-TTAAAGAGATCTTGGTTTGGCAA-3'	

5.3.6 Genome walking

5.3.6.1 Genomic DNA restriction digestion

Isolated genomic DNA (2-2.5 µg) was digested with four blunt-end cutting restriction enzymes (*EcoRV*, *PvuII*, *ScaI* and *StuI*) at 37 °C over night in a 100 µl reaction mixture using the modified procedure of Sibert et al. (1995). Inactivation of the enzymes was done by heat treatment. DNA was then extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) followed by DNA precipitation by adding 1/10 volume of 3 M NaOAc (pH 5.2). Then EtOH 100 % (2.5-times volume) and 1 µl glycogene (20 µg from a 20 mg/ml stock) was added. After incubation at -70 °C for 1 hour, the mixture was centrifuged at 13000 rpm for 15 minutes at 4 °C and the resulting pellet was washed with 70 % EtOH. After washing the pellet was air dried and then dissolved in 20 µl sterile water. The dried DNA was stored at -20 °C until used.

5.3.6.2 Ligation

Twenty microliter of the restriction enzyme-digested DNA was heated at 50 °C for 2 minutes on a heat block and then placed on ice. Adaptor 1 (5'- CTA ATA CGA CTC ACT ATA GGG CTC GAG CGG CCG CCC GGG CAG GT-3') and adaptor 2 (5'- ACCTGCCC-3') were used at a concentration of 150 pmol each and were heated at 95 °C for 3 min, then chilled slowly. Ligation of the adaptors was accomplished at 20 °C overnight in a 30 µl reaction mixture containing 20 µl heated, digested DNA, 3 µl ligase buffer (10X), 2 µl T4 Ligase (10 U), 1.5 µl of each adaptor (adaptor 1 and 2), and 2 µl sterile water. Ligation was terminated by incubating the mixture at 65°C for 10 minutes then placed on ice. Finally, the mixture was diluted to 100 µl by adding 70 µl sterile water and then stored at -20 °C until used.

5.3.6.3 PCR amplification

The primary PCR reaction was carried with adaptor primer pair AP1 in a total volume of 25 μ l containing 50 mM $MgCl_2$ (1 μ l), 0.25 μ M each of dATP, dTTP, dCTP, dGTP, 20 pmol of each primer (adaptor-specific and gene-specific primers), 2.5 μ l DNA polymerase reaction buffer (10X), 0.5 μ l of *Taq* DNA polymerase (5 U/ μ l, Bio line, Germany). The PCR reaction was run at 94°C for 4 minutes, followed by 35 amplification cycles (94°C for 30 seconds, primer annealing temperature based on T_m of the lower primer for 30 seconds, 72 °C for 2 minute and 72°C for 10 minutes for final DNA extension using a thermocycler (BioRad, icycler-USA). The PCR product from the first run was diluted 100- times with sterile the distilled water and 1 μ l of diluted PCR product was used as a template for the second (nested) PCR. The reaction product was analyzed by 1 % (w/v) agarose gel electrophoresis.

The nested PCR reaction was performed in a total volume of 50 μ l reaction mixture containing $MgCl_2$ (50 mM) 2 μ l, 0.2 μ M each of dATP, dTTP, dCTP, dGTP, 20 pmol of each primer (adaptor-specific and gene-specific primers), 5 μ l DNA polymerase reaction buffer (x10), 1 μ l of *Taq* DNA polymerase (5 U/ μ l, Bio line) and 1 μ l of the 100-times diluted PCR product as a template. The nested PCR reaction was performed with the same conditions as outlined above for the primary PCR reaction. The nested PCR product was run on 1 % (w/v) agarose gel containing 0.5 μ g/ml ethidium bromide. DNA band/s were visualized and imaged on a UV transluminator. .

5.3.7 Phylogenetic analysis

The phylogenetic relation of the C4H sequence in *H. aureonitens* with the C4H sequences in other plants was created by retrieving the C4H sequences from *Helianthus tuberosum* (Z17368), *Zinnia elegans* (U19922), *Echinacea angustifolia* (EU676019) from Asteraceae family, *Gossypium arboreum* (AF286648), *Nicotiana tobacco* (DQ350353), *Solanum tuberosum* (DQ341174), *Ammi majus* (AY 219918), *Petroselinum crispum* (L38898), *Arabidopsis thaliana* (cDNA, NM128601), *Arabidopsis thaliana* (gDNA, NC_003071) and *Ammy majous* (AY219918). The alignment was performed using phylogenetic and molecular evolutionary analyses, which were conducted using the Bootstrap consensus for neighbour joining, MEGA version 4.0 software (Kumar et al., 2004).

5. 3. 8 Secondary structure of C4H from *H. aureonitens*

Amino acid sequence and open reading frame (ORF) translation were carried out with MEGA software version 4.0 as well as expasy web site (au.expasy.org/tools/dna/html). Secondary structure of C4H from *H. aureonitens* was predicted on line with Psi-pred (bioinf.cs.ucl.ac.uk/psipred). To find the homology for *H. aureonitens* C4H as a template for modeling of C4H from *H. aureonitens*, the C4H amino acid sequence was submitted to HHPRED web site. The creation of secondary structure model and characterization of C4H from *H. aureonitens* was carried out with SWISS PDB viewer and RASMOL.

Briefly the expasy translated amino acid sequence of C4H was submitted to Psi-pred (to predict secondary structure of C4H) as well as HHPRED web site (to find a closest homology for modeling). Homology modeling was created by HHPRED and the characterization was performed by SWISS PDB viewer. The structure figure was prepared by RASMOL software.

5.4 Results

5.4.1 Amplification of an internal fragment of C4H using degenerated primers

Genomic DNA of *H. aureonitens* was used for the isolation of a full length gene. Based on the database for the C4H sequence of *Arabidopsis thaliana*, degenerated primers were designed for the amplification of an internal fragment, including the second intron (Fig. 5.1). According to the *A. thaliana* C4H structure, the expected size of the internal fragment should be about 500 bp. PCR amplification of the internal fragment for C4H with two degenerated primers resulted in two distinct bands, one about 500 and the other about 750 bp (Fig. 5.2). To find the right fragment, both amplicons were cloned in TA cloning vector and sequenced. NCBI blastn and blastx analysis indicated that the 500 bp band was not identical to the *A. thaliana* C4H, (data not shown). 11 kanamycin-resistant *E-coli* colonies of the upper band (750 bp) were selected and sequenced. Eight clones showed wide homologies to known C4H sequences with highest identities to *A. thaliana* (NC_003071) (Fig.5. 3). Since the gene structure of C4H from *A. thaliana* (NC_003071) is only the reference for the C4H structure map on the database, the major observed difference between *H. aureonitens* and *A. thaliana* in the obtained internal fragment was due to the size of the intron which was found at 472 and 220 bp respectively without regarding the minor differences in nucleotide sequence.

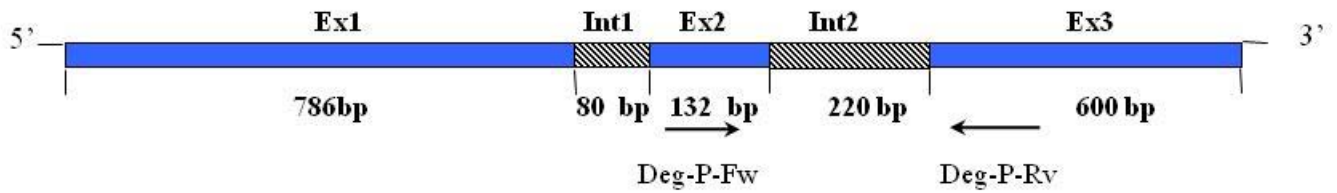


Fig. 5. 1. C4H gene map in *Arabidopsis thaliana* (NC_003071.3). The blue boxes represent the coding region and the grey box an untranslated region. Arrows show the position of the designed degenerated primers used to amplify the internal fragment.

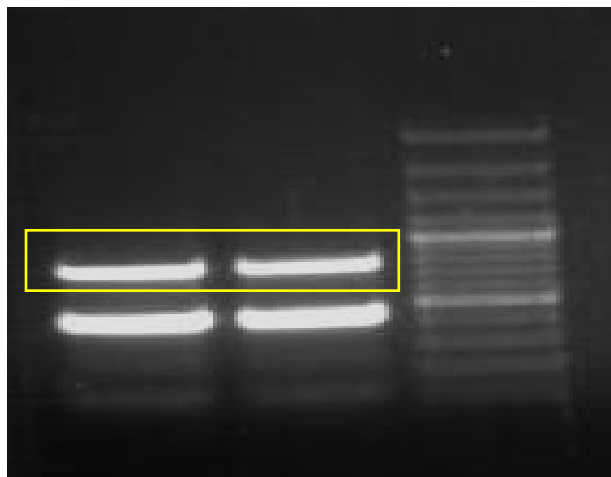


Fig. 5. 2. Amplification of the internal partial fragment of C4H from the genomic DNA of *Helichrysum aureonitens* by degenerated primers. Boxed fragments were cloned and sequenced.

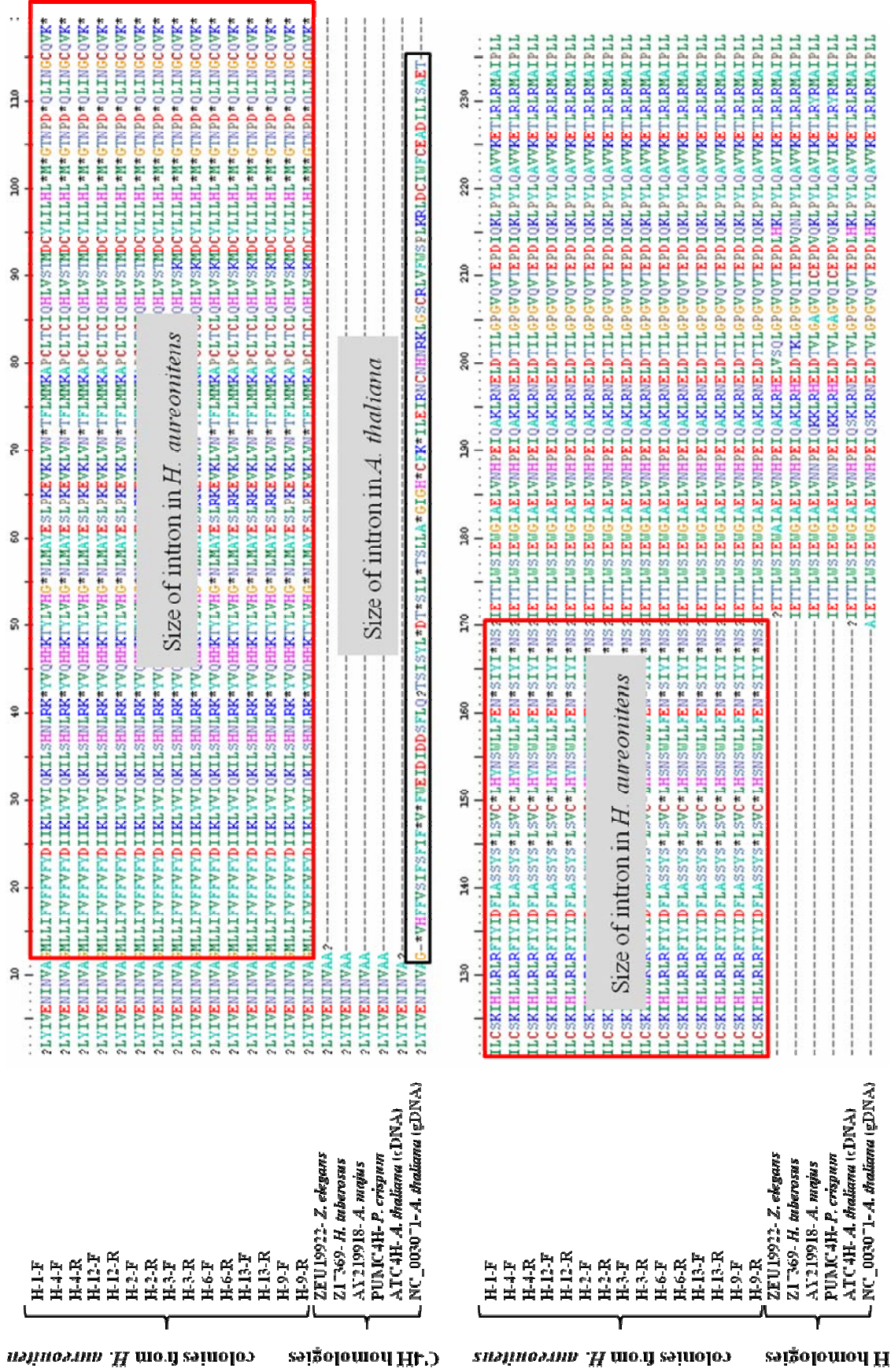


Fig. 5. 3. Multi-alignment of the sequences of the internal fragment of C4H in *Helichrysum aureoniens* from eight colonies with C4H homologues of other plants. The boxed region represents the position of intron.

5.4.2 Genome walking

5.4.2.1 Genomic DNA restriction digestion

The genomic DNA was digested with four blunt-end restriction enzymes, *EcoRV*, *PvuII*, *ScaI* and *StuI*. As shown in figure 5.4, the complete digestion of the genomic DNA with blunt-end restriction enzymes revealed a smear when using gel electrophoresis, which shows a wide range of different fragment sizes of the genomic DNA. The blunt-end digested fragments, which were ligated with adaptors, were used as a DNA template for the primary PCR reaction of the first genome walking to find the upstream and downstream regions of the known area. Two gene specific primers, CH1-F and CH1-R and adaptor primer 1 (AP1) (Table 5.2) were used. Figure 5.5-A shows the PCR product of the primary reaction with AP1 (Adaptor primer 1), CH1-F and CH1-R (Gene specific primers) resulted in a smear in both directions, up and downstream to the known internal fragment. A second PCR (nested PCR) was performed with two other gene specific primers, CH1-NF and CH1-NR, in combination with a adaptor primer 2 (AP2) (Table 5.2). The nested amplification used to identify the downstream region resulted in a bright band of about 750 and 600 bp in *EcoRV* and *ScaI* digested DNA respectively (Fig. 5.5-B) The nested PCR upstream of the known area also resulted in a bright band of about 500 bp only in *ScaI* digested DNA, but no amplification was observed in the *EcoRV* digested DNA. All bands for both directions were recovered from the gel and subcloned into the TOPO-TA vector. The plasmids from kanamycin resistant clones of up and downstream were sequenced from the *ScaI* and *EcoRV* digested DNA experiments. The blastn and blastx analysis for the downstream region indicated that all the clones were highly conserved to known C4H sequences in the database and showed the 3' end sequence of C4H in *H. aureonitens*.

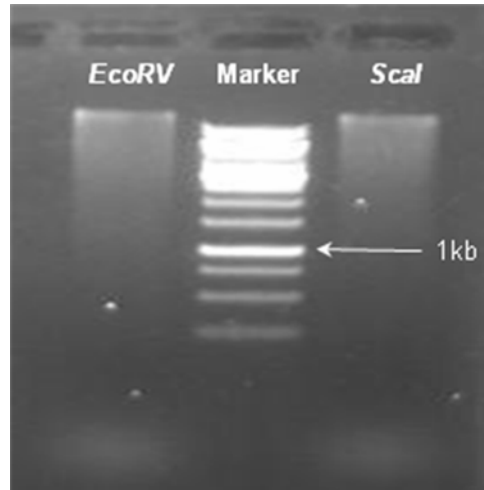


Fig. 5. 4. Complete digestion of the genomic DNA using blunt-end restriction enzymes, *PvuII* and *StuI*. The second lane shows the DNA ladder and the arrow represents the size of 1 kb in the ladder.

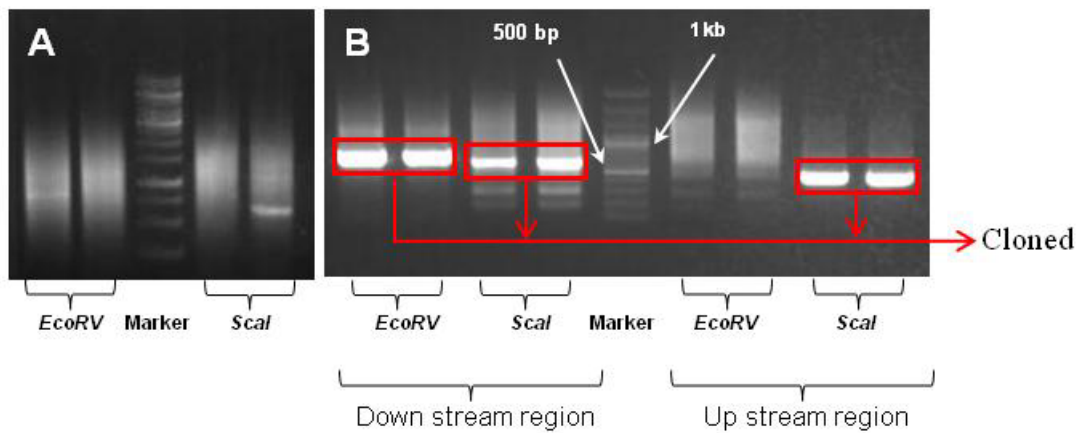


Fig. 5. 5. PCR amplification for the first genome walking with *EcoRV* and *SacI* DNA digestion. **A:** Primary PCR. **B:** Second PCR (nested PCR); the red boxed fragments were cloned and sequenced to identify the up and down stream regions of the known internal fragment.

The results of blastn and blastx for the upstream direction showed that there is a wide similarity among all clones and good homology between clones and known C4H sequences from the other plants on the database. This is true until 119 bp from CH1-NR and after that another intron could start. It means that the sequence of the exon in that partial fragment of *H. aureonitens* is identical to the sequence of genomic DNA in *A. thaliana* (NC_003071.3) (Fig. 5.1).

The second genome walking was performed by using *PvuII* and *StuI* digested DNA as template. Primers CH1-NR and CH2-NR were used for primary and nested PCR respectively. Primer CH2-NR was designed based on the sequences obtained from the first genome walking experiment allowing the identification of the upstream part of the result of first walking. Results of the primary and nested PCR showed a fragment of about 1 kb only in *PvuII* digested DNA and no amplification was observed in *StuI* digested DNA (Fig. 5.6). The blastn and blastx results for the 1 kb fragment showed the presence of a second intron in the genomic DNA of *H. aureonitens*. It further demonstrated that the second intron in the genomic DNA of *H. aureonitens* is longer than the second intron in the genomic DNA of *A. thaliana* (NC_003071.3).

The third walking was carried out based on the result of sequencing in a second genome walking. In this step the CH3-NR primer was used to amplify a 300 bp fragment from *StuI* digested DNA. After sequencing, the blastn and blastx analysis showed that a further intron is present in the genomic DNA of *H. aureonitens*. The size of this intron was determined to be 630 bp. This intron is not present in the genomic DNA of *A. thaliana* (NC_003071.3).

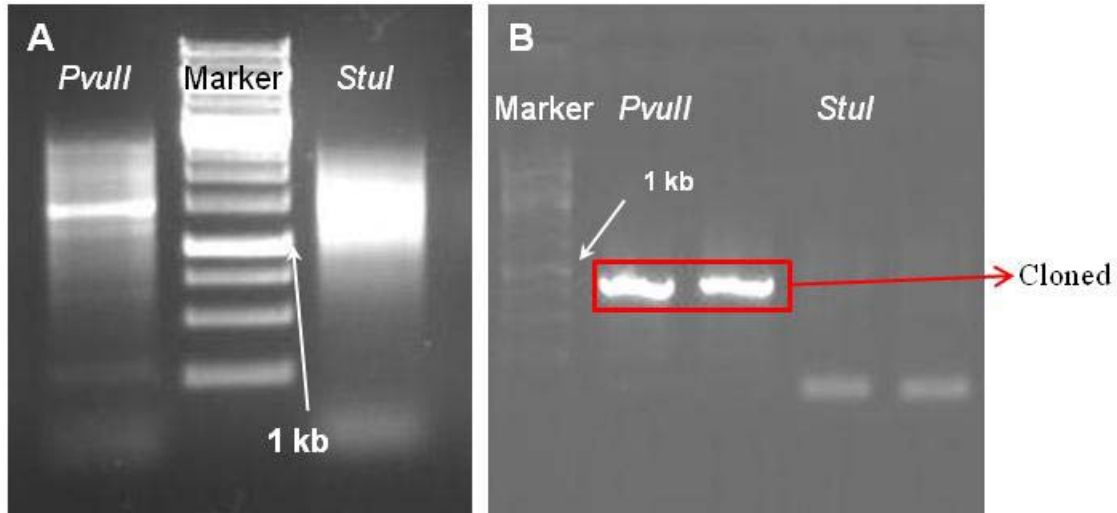


Fig. 5. 6. PCR amplification for the first genome walking with *PvuII* and *StuI* digested DNA for the second genome walking. A: Primary PCR; B: Second PCR (nested PCR); the red boxed fragments were cloned and sequenced to identify the upstream region of known the fragment from the first genome walking.

5.4.3 Amplification of the full length cDNA and genomic C4H from *H. aureonitens*

Amplification of a full length of cDNA and genomic DNA of C4H in *H. aureonitens* by using primers C4H-F and C4H-R yielded a specific band of about 1500 bp in full length C4H amplification of cDNA and another band of about 3400 bp in a full length C4H amplification genomic DNA. These primers were designed based on a obtained sequence of 5' and 3' ends of the genome walking experiments, including a start codon (ATG) and stop codon (TAA). The sequencing results of the full length of cDNA were identical to the genomic DNA sequence except the intron regions.

5.4.4 Molecular characterization of the nucleotide sequence of C4H from *H. aureonitens*

The full length sequence of the genomic DNA of C4H from *H. aureonitens* is 3399 bp. Sequencing results of genomic DNA of C4H from *H. aureonitens* indicated the presence of three introns. Two of them are also present in the genomic DNA of *A. thaliana* (NC_003071.3) and the third one is a new finding in C4H from *H. aureonitens*. The size of the three introns (I, II, III) from the 5' to 3' side is 630 (445-1074), 852 (1420-2196) and 472 bp (2326-2799) respectively (Fig. 5.7 and 5.8).

The synthesized C4H cDNA from total RNA of *H. aureonitens* showed the full length size of 1518 bp (Fig. 5.9). G + C content of the open reading frame (ORF) is 44.4 %, while the introns have a total of 30.27 %.

NCBI Blastn analysis indicated that the open reading frame of isolated C4H from *H. aureonitens* shows high identities to known C4H tags from Asteraceae family with the highest homologies to *Echinacea angustifolia* (EU676019) with a 83 % similarity. The other high similarity was to *Zinnia violacea* (U19922), *Helianthus tuberosus* (Z17369) 81 % as inter family and to AtC4H (NM_128601) 74 %. They also show moderate identities to many C4H/CYP73A genes from different families, such as those from *Nicotiana tabacum* (DQ350352), 78 %, *Catharanthus roseus* (Z32563) 77 %, *Gossypium arboreum* (AF286648), 77 %. NCBI blastp indicated that the C4H shows very many similarities to the C4H genes from other plants. When pairwise-aligned on the whole molecule scale, the C4H shows identities / positives of 94 % / 97.0 % to intra family *E. angustifolia* (ACF74449); *H. tuberosus* (CAAB78982), 93.0 % / 97.0% to *Z. violacea* (AAB42024), 92.0 % / 96.0 %; to C4H from *N. tabacum* (ABC69412), 88.0 % / 94.0 %, *Solanum tuberosum* 87.0 % / 95.0 %; *G. arboreum* (AAG10197) 88.0 % / 95.0 %; *C. roseus* (CAA83552) 89.0 % / 94.0 % *Arabidopsis thaliana* monocot (AAB58355) 85.0 % / 93.0 % and



to C4H from *Sorghum bicolor* (AAK54447) 75.0 % / 87.0 % and to the gymnosperm C4H of *Ginkgo biloba* (CAA70596) 81.0 % / 90.0 % (Fig. 5. 10).

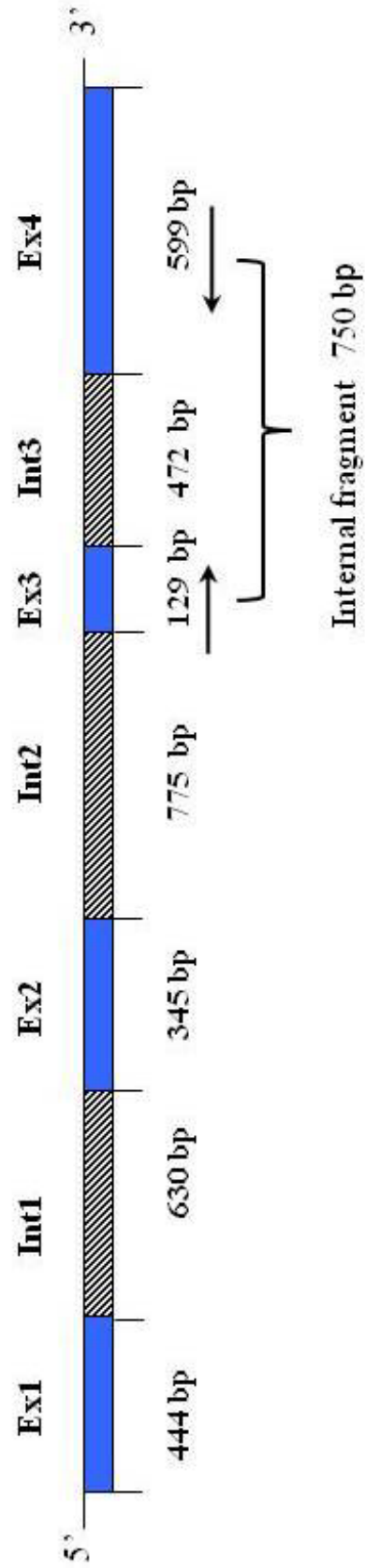


Fig. 5. 8. C4H gene structure in *Helichrysum aureonitens*. Arrows show the position of the designed degenerated primers used to amplify the internal fragment.


```
ATGGATCTACTCCTTTTGGAGAAA →  
ATGGATCTACTCCTTTTGGAGAAAACACTCTTAGCCCTGTTTGCAACCATTTTAGGTGCAATC  
TTCATCTCCAAGCTACGTGGCAAACGTTTCAAGCTGCCACCTGGACCAATCCCAGTTCCCAT  
TTTTGGAAACTGGCTCCAAGTTGGTGATGATCTTAATCATCGTAATTTAACCGATCTCGCCAA  
GAAATTTGGCGAGATCTTTTTATTACGAATGGGACAACGCAACCTTGTTGTTGTGTCGTC  
CGGATCTTGCCAAAGATGTGCTTCATACACAAGGTGTGCGAGTTCGGATCACGAACTCGAAA  
CGTTGTGTTTGATATCTTCACGGGAAAAGGACAAGATATGGTTTTACGGTTTATGGGGAA  
ATTGGCGTAAAGTGCGGAGGATCATGACTGTTCCGTTTTTTACTAACAAAGTCGTCCAACAA  
TACAGGTATGGATGGGAGGCAGAGGCTGCAGCTGTTGTGGAAGATGTGAAGAAGAATCCG  
GCAGCAGCGACAGAAGGTGTGGTGATCCGAAGACGATTACAATTGATGATGTACAACAACA  
TGTTCAGGATTATGTTTGATAGAAGGTTTGAGAGTGAGGATGATCCTTTGTTCTTGAACTA  
AAGGCATTAATGGAGAAAGGAGTCGATTGGCACAAGCTTCGAGTATAACTATGGCGATT  
TCATCCCGATTTTGAGGCCGTTTTTGAGAAGTTACTTGAAGTTATGTAAGGAAGTTAAAGAG  
AAGAGGTTGCAGCTATTCAAGGATTACTTTGTTGATGAAAGGAGGGAATTGGGAAGCACA  
AAAAGCATGGACAACAACCAACTGAAATGTGCAATTGATCATATTCTTGAAGCCAAGGACA  
AGGGAGAGATCAATGAGGACAATGTTCTTACATTGTTGAGAACATCAACGTTGCTGCAATC  
GAGACAACCTTATGGTCTATTGAATGGGGAATCGCTGAGCTTGTGAACCACCCCGAGATCC  
AAGCCAAACTCAGGAACGAGCTTGACACCATACTTGACCAGGAGTACAAGTCACGGAGC  
CTGACATCCAAAACTTCCGTACCTCCAAGCTGTGGTTAAGGAAACTCTTCGTCTTCGTATG  
GCTATTCCTCTTTTAGTCCCACACATGAACCTTCACGATGCTAAGCTAGGTGGCTTTGACATC  
CCAGCCGAGAGCAAGATCTTGGTCAACGCTTGGTGGCTCGCTAATAATCCTGAGCAATGGA  
AGAAACCCGAGGAGTTTAGGCCCGAGAGGTTCTTTGAAGAAGAAAGTAATGTTGAAGCCA  
ATGGTAATGACTTTAGGTAAGTGCCTTTCGGTGTGGAAGGAGGAGTTGTCCTGGGATTATT  
CTTGCCTTGGCAGTCTAGGAATCACGATCGGGCGTTTGGTGCAGAATTCGAGTTGTTGCC  
GCCACCCGGAGTGTCAAAGATTGACACGAGTGAAAAAGGTGGACAGTTTAGTCTTCATATC  
TTGAAGCATTCTACTGTTGTTGCCAAACCAAGATCTCTTTAA  
← TTGCCAAACCAAGATCTCTTAA
```

Fig. 5. 9. cDNA full length sequence of the C4H in *Helichrysum aureonitens* (1518 bp including stop codon, TAA). Arrows show the position of the attaching place of the specific primers (C4H-F and C4H-R) used to amplify the full length C4H gene. The under-lined nucleotides represent the start and stop codons for the isolated C4H gene from *Helichrysum aureonitens*.

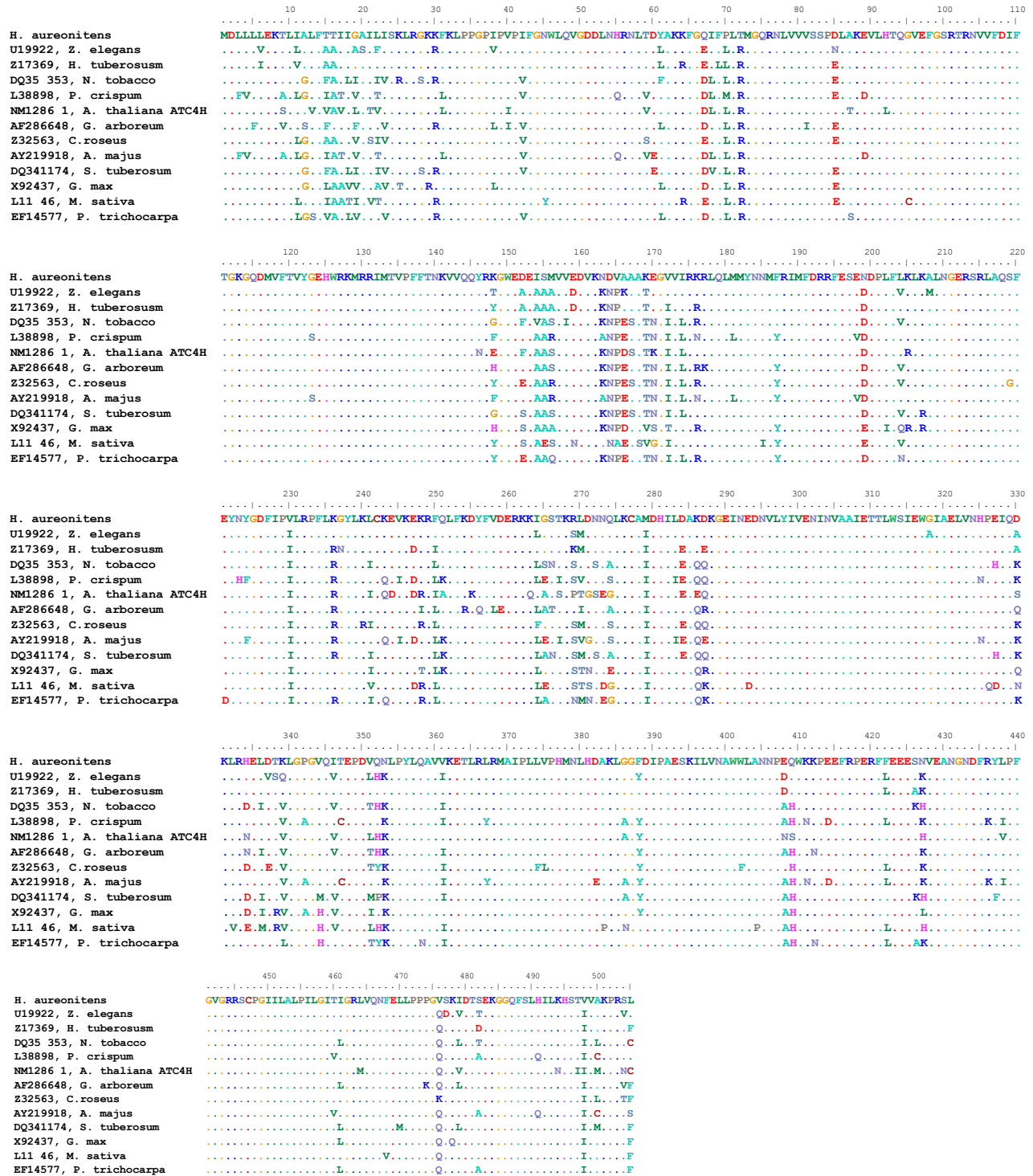


Fig. 5. 10. Multi-alignment of C4H amino acid from *Helichrysum aureonitens* with C4H homologies from other plants. The spots represent the similar amino acid sequences to the C4H from *Helichrysum aureonitens*.

5.4.5 Phylogenetic tree

A preliminary phylogenetic analysis carried out to determine the relationship between C4H from *H. aureonitens* with other C4H genes in other plants revealed that the C4H from *H. aureonitens* can be categorized in the same group of C4H as found in plants belonging to the Asteraceae family including *Z. eleganse* (Ye, et al., 1996), *H. tuberosus* (Teutsch, et al., 1993), and *E. angustifolia* (Feng, et al., 2008), but the C4H in *H. aureonitens* is closest to *Z. violacea* compared to the other inter family tested plants (Fig.5.11). It also indicated that there is a significant difference between C4H from Asteraceae and C4H from others (Fig.5.11).

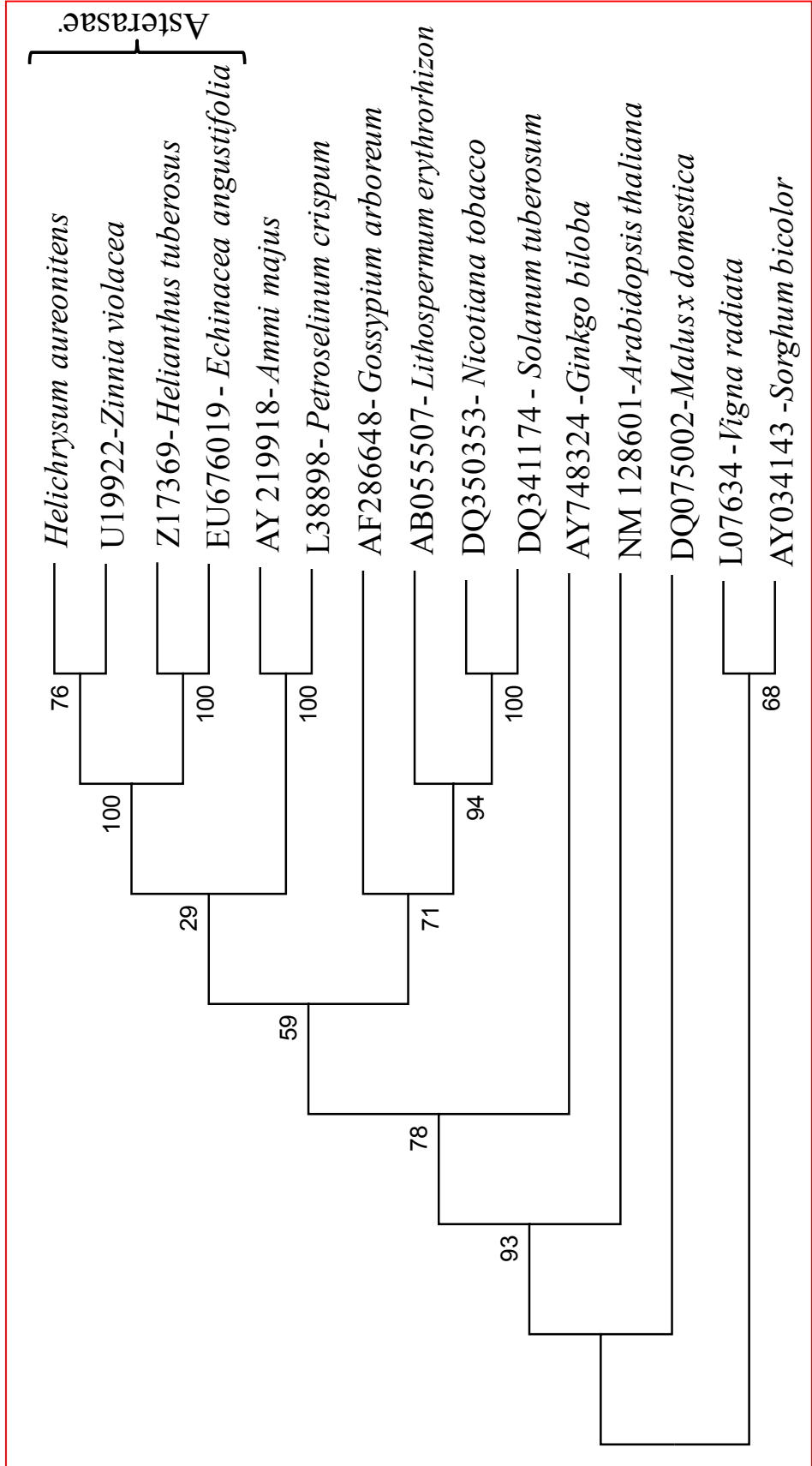


Fig. 5. 11. Phylogenetic tree of the C4H in *Helichrysum aureonitens*.

5.4.6 Secondary and tertiary structures of C4H from *H. aureonitens*

Figure 5. 12 shows the secondary structure of C4H from *H. aureonitens*, which was predicted online by Psi-pred. In this predicted secondary structure of C4H, it was demonstrated that there are some helices with different sizes containing 3 to 29 residues. The longest helix located in the middle between two strand regions. The first helix at N-terminus can be attributed to the signal peptide/anchor. To find the homology for the C4H protein from *H. aureonitens*, the translated amino acid sequence of C4H from *H. aureonitens* was applied to HHPRED web site. Figure 5. 13 shows the alignment between amino acid sequences of C4H from *H. aureonitens* and a cytochrome P450, 2C9 monooxygenase, drug metabolizing enzyme, oxidoreductase (IR90-A,) as one of the closest homologies to *H. aureonitens* C4H. As there is no known structure for plant-derived C4H on the database, the closest homologies were selected from human cells. The first 23 amino acids are not included in this alignment because of no homology at the N terminus of C4H from *H. aureonitens* to the IR90-A (Fig. 5. 13). According to this result the secondary structure of *H. aureonitens* C4H modeling was created in the H H PRED online web site (Fig 5. 14). Based on the C4H created model the secondary structure from *H. aureonitens* is composed of 16 alpha helices (45.35 %), random coils (46.4 %) and strands (8.22 %). Alpha helices were mainly distributed throughout the protein, containing 2 to 31 amino acids. The longest helix is placed around the middle of C4H protein. Extended strands mainly distributed in two regions: one is the ~100-residue N-terminal region, and another is the ~120-residue region at the C-terminal helix and the central helices with 31 residues. In these 2 regions, extended strands distribute in an interlaced manner with random coils and small alpha helices (Fig. 5.14).

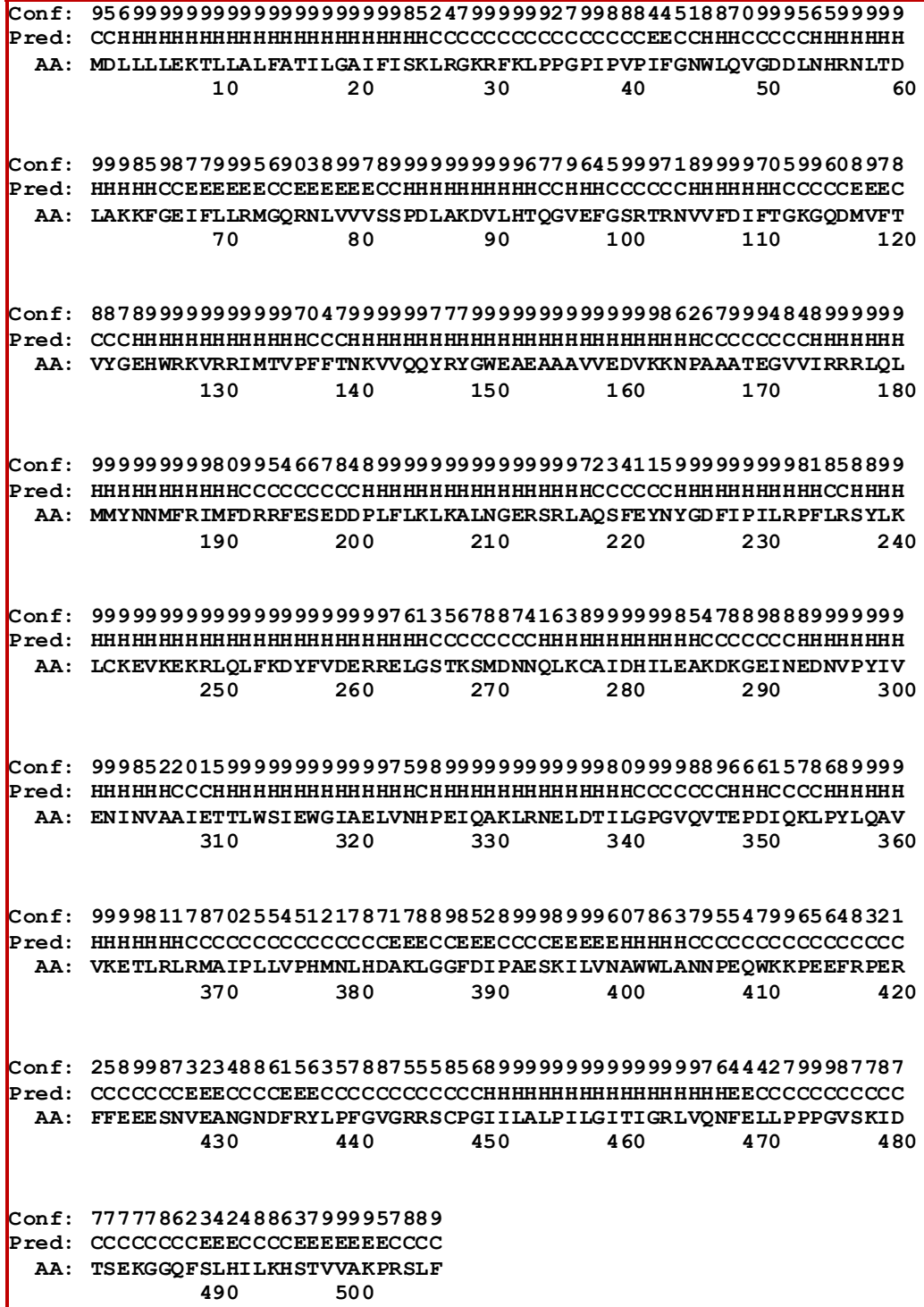


Fig. 5. 12. Secondary structure of C4H from *Helichrysum aureonitens* was predicted in Psi-pred (Psipred V2.6 by David Jones) (bioinf.cs.ucl.ac.uk/psipred). Conf: Confidence (0=low, 9=high); Pred: Predicted secondary structure (H=helix, E=strand, C=coil); AA: Target sequence.

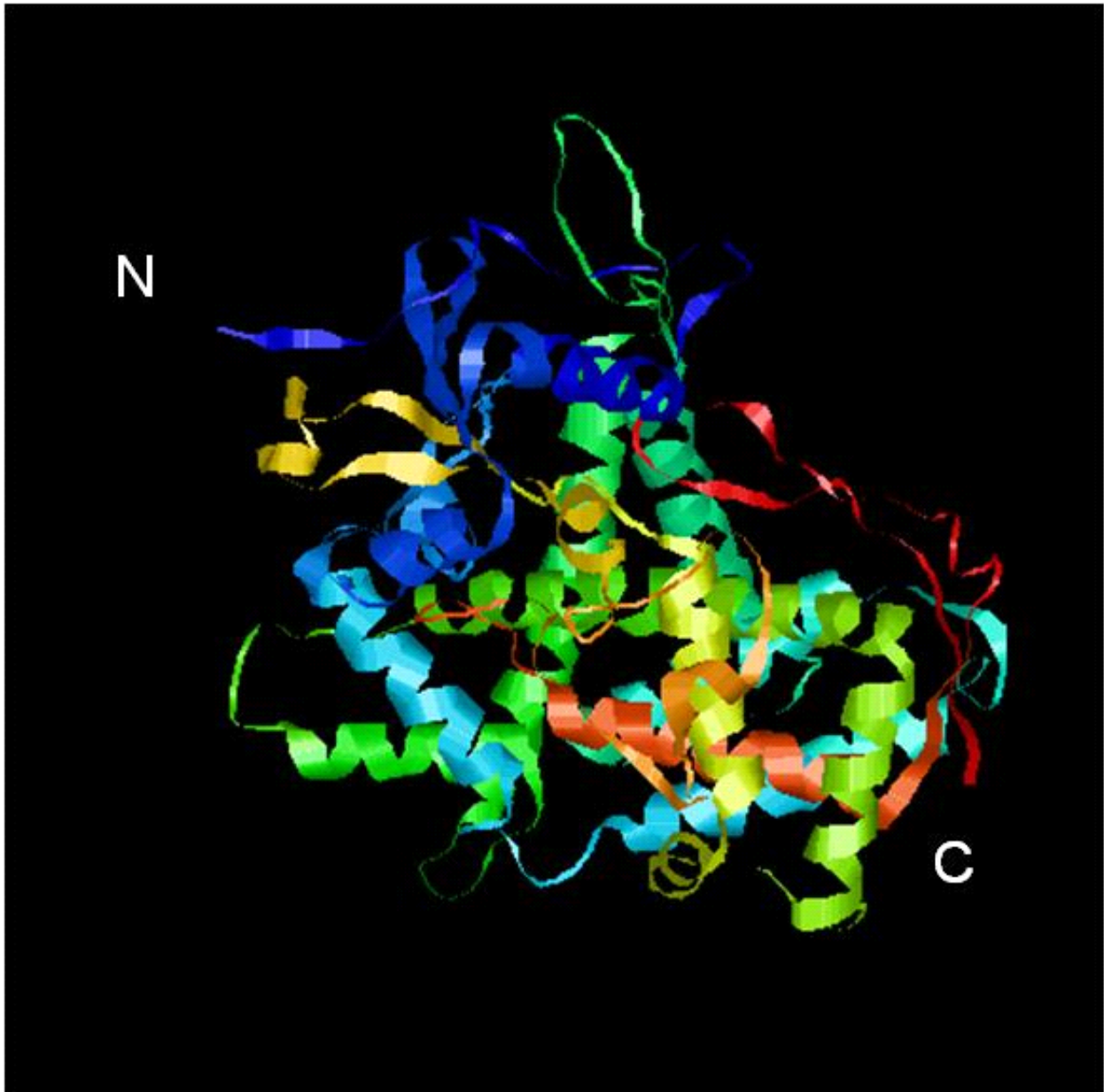


Fig. 5. 14. RASMOL secondary structure model for C4H from *Helichrysum aureonitens*. N: N terminus; C: C terminus.

5.5 Discussion

In this study a copy of C4H from *H. aureonitens* was isolated and sequenced. Although all the transformed colonies from the internal fragment were sequenced and all of them showed the same sequences, there is no clear evidence to conclude that *H. aureonitens* has only one copy of C4H. According to the database there is no report to show that members of the Asteraceae such as *Z. violacea* (Ye, et al., 1993), *H. tuberosus* (Teutsch, et al., 1993), and *E. Angustifolia* (Feng, et al., 2008) have more than one C4H gene. Pea, *Arabidopsis* and parsley are thought to contain only one copy of C4H (Frank et al., 1996; Bell-Lelong et al., 1997; Mizutani et al., 1997; Koopmann et al., 1999), but in members of the Brassicaceae family, there are some with as much as 6 copies of C4H present in *Brassica napus* (Chen et al., 2007). However to confirm the number of C4H copies in *H. aureonitens* further experiments need to be carried out.

According to the genomic DNA sequencing, the obtained gene structure for *H. aureonitens* C4H is different from all the present data on the database. Based on the C4H gene structure in *A. thaliana* (NC_003071.3) (Fig. 5.1) there are three exons (Ex1, 786 bp; Ex2, 132 bp; Ex3, 600 bp) and two introns (In1, 80 bp and In2, 220 bp), but the C4H gene map obtained for *H. aureonitens* indicates four exons (Ex1, 444 bp; Ex2, 345 bp; Ex3, 129 bp; Ex4, 599) and three introns (In1, 630 bp; In2, 775 bp and In3, 472 bp). The length of the open reading is 1518 bp which is similar to the coding region of C4H in species of the Asteraceae family such *Z. violacea* (U19922) (Ye, et al., 1993), *H. tuberosus* (Z17369) (Teutsch, et al., 1993), and *E. angustifolia* (EU676019) (Feng, et al., 2008) as well as other plants from different families such as *A. thaliana* (AtC4H, NM_128601) (Mizutani et al., 1997). The existence of different sizes of introns of a specific gene in different plants has already been reported. Two copies (BnC4H-1 and BnC4H-2) of C4H were isolated from *Brassica napus*. The genomic DNA sequence of these indicated that there are

two introns in each one, while the size of introns in BnC4H-1 (In1, 70; In2, 378) and BnC4H-2 (In1, 64; In2, 326) are different (Chen et al., 2007). According to the results of this study regarding the gene structure of C4H, it is the first time that the presence of a new intron (Intron 1, Fig. 5.1) for C4H has been reported.

The G+C content of C4H in *H. aureonitens* in the open reading frame and introns showed 44.4 %, and 30.27 % respectively. These results were confirmed by Chen et al. (2007), whose experiment showed higher G+C content in the coding regions than the non-coding regions in the C4H genes isolated from *B. napus*.

The results of the C4H secondary structure model from *H. aureonitens*, in the case of number of helixes, seems different to isolated C4Hs from *Brassica napus* by Chen et al. (2006). They reported 10 to 11 helixes in BnC4H-1 and BnC4H-2 respectively while in this study we reported 16 helixes. The reason for the different number of helixes can be contributed to the different consideration of counted helixes. In this study all helixes with distinct position were considered as separate helixes, but Chen et al. (2006) for example, the longest helix has two separate parts, which joined together with some random coils. The size of longest helix in C4H from *H. aureonitens* in this study was reported 31 residues, while Chen et al. (2006) reported 62 in BnC4H-1 and in BnC4H-2 (two pieces). In C4H from *H. aureonitens*, there are 4 helixes close to each other with a total size of 62 residues (31, 14, 4 and 13) that can consider as a big helix, but it needs more study.

Predictions on the signal peptide/anchor and subcellular localization of C4H from *H. aureonitens* by different software gave inconsistent results. C4H and other microsomal P450s have been extracted from microsomes of various organisms, and it has been presumed that microsomal P450s have a N-terminal hydrophobic helix which serves to anchor the enzyme to the ER

membrane (Winkel-Shirley, 1999; Werck- Reichhart et al., 2002). In C4H the N-terminal hydrophobic helix has been predicted with multiple identities, i.e. signal peptide with cutting site, signal anchor, and transmembrane helix, by different software in this species. A second strong transmembrane helix was also predicted at the C-terminus. Another fact is that the predicted tertiary structures of the C4H protein from *H. aureonitens* do not contain the first 23 residues, and no N-terminal helix can be seen in Fig. 5. 13, although figure 5. 12 shows a helix region at N-terminal with low confidence. Considering P450s in general, there is more than one report indicating effective or possible plastidic localization. These make it difficult to draw a definite conclusion on the properties of the N-terminal helix and subcellular localization of C4H from *H. aureonitens*. But it is obvious that both the N-terminal and the C-terminal sequences of C4H from *H. aureonitens* do not show essential difference from those of AtC4H and most other C4Hs, so the location and topology of C4H from *H. aureonitens* should be similar to those of typical C4H proteins.

5.6 Conclusions

The results in the previous chapter showed that *p*-coumaric acid, the product of C4H activity, was not detected in *H. aureonitens* samples, and based on the known phenylpropanoid biosynthetic pathway C4H may not be active at that position in this species. However, in the leaf samples of *H. aureonitens*, kaempferol, the hydroxylated compound at the 4' position, was detected. Based on these results the initial question that arose was whether C4H is present or expressed in *H. aureonitens*. The results of this part of the study showed that C4H is present in *H. aureonitens* and its characteristics show the highest homologies to *Echinacea angustifolia* from the Asteraceae C4H with 83.6 % identity on nucleotide level but 93 % identity and 97 % positives on the protein level. Although there are some differences in the case of the gene structure, the full length coding region sequence of C4H from *H. aureonitens* is mostly similar to that of other plants (1518 bp including stop codon, TAA). It was observed that the similarity of isolated C4H from *H. aureonitens* was quite similar to that of other Asteraceae species present in the database. The phylogenetic tree result indicated that there is a close relationship among C4H nucleotide sequences from Asteraceae and that they can be grouped in one category, but there is a big difference between C4H sequences of this group and C4Hs from other plants. The predicted secondary structure of C4H from *H. aureonitens* is almost similar to C4Hs from other plants, but the characterization of C4H from *H. aureonitens* still needs more experiments.

In proving the hypothesis for the presence of an alternative biosynthetic pathway for kaempferol production in *H. aureonitens* and the possibility of some link/s between an non 4'-OH pathway and the 4'-OH pathway, the expression of isolated C4H from *H. aureonitens* will be discussed in the next chapter.