Front. Agric. China 2008, 2(4): 456–462
DOI 10.1007/s11703-008-0047-x

RESEARCH ARTICLE

Anhe CHEN, Jiana LI, Yourong CHAI, Rui WANG, Jun LU

Cloning and sequence analysis of a mutation-type cinnamate 4-hydroxylase gene from *Brassica oleracea* L. var. *acephala* DC.

Abstract A 2431-bp full-length cinnamate 4-hydroxylase gene, *BoC4H*, was cloned from *Brassica oleracea* L. var. *acephala* DC.. It contains 2 introns. Its mRNA is 1715 bp, encoding a deduced 481-amino-acid polypeptide with wide homologies to C4Hs from other plants. It possesses cytochrome P450 conserved domains and motifs such as the haem-iron binding motif, the E-R-R triad, the T-containing binding pocket motif and the hinge motif necessary for optimal orientation of the enzyme. It also has most of the canonical C4H/CYP73A5-featured substrate-recognition sites (SRSs) and active site residues. However, owing to a single-base deletion at C_{2242} and subsequent frame shift within the 3′ coding region as compared with *C4H* genes from *Arabidopsis thaliana* and other plants, BoC4H shows a 36-aa deletion/variation at its C-terminus and the SRS6 motif together with active site residues therein are absent. Thus BoC4H may be of no function or low activity. BoC4H is a membrane protein and is probably associated with the endoplasmic reticulum. Its secondary structure is dominated by alpha helices and random coils. The Swiss-Model could not predict its tertiary structure. *B. oleracea* contains a *C4H* gene family with at least 5 members.

Keywords *Brassica oleracea* L. var. *acephala* DC., cloning, cinnamate 4-hydroxylase (C4H), mutation, sequence analysis

Translated from *Acta Horticulturae Sinica*, 2007, 34(4): 915–922 [译自: 园艺学报]

Anhe CHEN, Jiana LI (✉), Yourong CHAI (✉), Rui WANG, Jun LU
Chongqing Rapeseed Engineering Research Center, Key Laboratory of Biotechnology and Crop Quality Improvement of Ministry of Agriculture, College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China
E-mail: ljn1950@swu.edu.cn; chaiyourongl@163.com

Anhe CHEN
College of Bio-information, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

1 Introduction

Phenylpropanoid pathway produces a large number of biologically important secondary metabolites such as lignin and various flavonoid compounds from L-phenylalanine through several important branch pathways located downstream of the common phenylpropanoid pathway (Barber and Mitchell, 1997). Lignin, after cellulose, is the second most abundant terrestrial organic polymer, accounting for up to 30% of all vascular plant tissues. By reinforcing plant cell walls, lignin provides plant tissues with mechanical rigidity, water impermeability and protection against pathogen invasion. It plays a fundamental role in mechanical support, solute conductance and disease resistance in higher plants (Li et al., 2006). Flavonoid compounds, such as flavones, flavonols, phlobaphenes, aurones, anthocyanins and proanthocyanidins etc., play important roles in plant development, resistance, quality and nodule symbiosis (Winkel-Shirley, 2001; Harakava, 2005).

Cinnamate 4-hydroxylase (C4H, EC 1.14.13.11) belongs to the CYP73A group of the cytochrome P450-dependent monooxygenase superfamily. As a membrane-associated protein, it forms a tight regulatory unit with other core enzymes of phenylpropanoid metabolism, including phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL) and NADPH:cytochrome P450 oxidoreductase (CPR). C4H is the second key enzyme of the common phenylpropanoid pathway and plays a key action in those enzymatic steps. C4H catalyzes the hydroxylation of cinnamate to *p*-coumaric acid. *p*-coumaric acid is then transformed into *p*-coumaroyl-coenzyme A (CoA) by 4-coumaroyl:CoA ligase. *p*-coumaroyl-CoA goes into the downstream pathways such as the flavonoid pathway and the lignin pathway and so on. *P*-coumaric acid can also directly go into the lignin pathway, through further hydroxylation, yielding caffeic acid. Furthermore, caffeic acid can be transformed into ferulic acid by the caffeic acid *O*-methyltransferase. Finally, ferulic acid is transformed into the sinapic acid through further hydroxylation and the methylation. Those phenolic acids are used as the carbonic

bones for the biosynthesis of lignin and phytoalexin (Baber and Mitchell, 1997; Li et al., 2006).

C4H is one of the most studied enzymes in the cytochrome P450-dependent monooxygenase superfamily. According to Nelson's P450 Web (http://drnelson.utmem.edu/biblioD.html#73A), 54 kinds of C4H from various plants have been cloned to date. Except for the full-length *C4H* gene from the model plant *Arabidopsis thaliana* and some short *C4H* tags from family Brassicaceae, no other full-length cruciferous *C4H* gene have been reported or submitted to GenBank, though many important oilseed and vegetable crops are from Brassicaceae.

Brassica oleracea L. var. *acephala* DC. is an important ornamental plant. A study on *C4H* helps to understand the mechanism of diseases resistance, stress resistance, growth and development and quality improvement of *B. oleracea*. It is also helpful to clarify the molecular evolution of *C4H*. No full-length *C4H* gene was cloned from this species, except for four short *C4H* tags (GenBank accession Nos. AF230674–AF230677). Here, we report the cloning and systematical sequence analysis of a full-length *C4H* gene from *B. oleracea*.

2 Materials and methods

2.1 Materials

The leaves and roots of *B. oleracea* var. *acephala* were sampled from the experimental field of the Chongqing Rapeseed Engineering Research Center, Southwest University, China. *Escherichia coli* strain DH5α was preserved in our laboratory. GeneRacer Kit was the product of Invitrogen Co., USA. Plant Total RNA Purification Mini Kit and DNA Gel Extraction Mini Kit were the products of Watson Biotechnologies (Shanghai, China). T-vector pMD18-T and Taq DNA polymerase were the products of TaKaRa Biotechnology Co. (Dalian, China). Other major bio-molecular reagents such as dNTPs, agarose, CTAB, peptone, yeast extract, ampicillin (Amp), X-gal, IPTG and so on, were the products of the Shanghai Biotechnology Co.. Primer synthesis and DNA sequencing were commercially completed by the Invitrogen Co. (Shanghai, China).

2.2 Methods

2.2.1 Extraction of total genomic DNA and RNA

Total genomic DNA of *B. oleracea* was extracted by a CTAB method (Rechards, 1995). Total RNA was extracted, using a Mini Plant Total RNA Purification Kit, from the elongated roots of *B. oleracea* plants at the budding stage because *C4H* genes were almost constitutively expressed and the roots were highly lignified.

2.2.2 PCR amplification of the *C4H* conserved region

Primers FBNC4A (5′-AGAGCTCCTTCGGAAACTGGCTCCAAG-3′) and RBNC4A (5′-AGGATCCCTCAAAGCTCTGAGCCAACC-3′) were designed based on the multi-alignment on Vector NTI Advance 9.0 of *C4H*s from *A. thaliana* and other plants. PCR amplification of the *C4H* conserved region was carried out in a 50-μL PCR system containing 5 μL of 10 × PCR buffer, 3 μL of 25 mmol·L^{-1} $MgCl_2$, 1 μL of 10 mmol·L^{-1} dNTPs, 1 μL of 10 μmol·L^{-1} FBNC4A, 1 μL of 10 μmol·L^{-1} of RBNC4A, 1 μL (0.4 μg) of genomic DNA and 1 μL of 5 U·μL^{-1} Taq DNA polymerase. PCR process was as follows: 94°C for 1 min, followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min and followed by 72°C for 10 min. The PCR product was separated by using 1% agarose gel electrophoresis. The special band was recovered with DNA Gel Extraction Mini Kit after dyeing with GoldView and photographing.

2.2.3 Subcloning and sequencing

The recovered DNA fragment was linked with pMD18-T vector. The cloned DNA was transformed into DH5α competent cells by the $CaCl_2$ method (Seidman et al., 1995) and screened on the agar plate with X-gal + IPTG + Amp. Positive white colonies were PCR-identified and sequenced with general primers M13F/M13R.

2.2.4 RACE handling

An aliquot of 5 μg total root RNA was used as template to synthesize the first strand total cDNA using a GeneRacer Kit. The first strand total cDNA was gained through dephosphorylation of the incomplete RNA 5′ end, decapping of the complete mRNA, ligation of the RNA Oligo to the decapped mRNA, reverse transcription, etc., in terms of manual instruction.

2.2.4.1 3′-RACE amplification

Based on the sequences of the conserved region of *C4H* genes from *B. oleracea* var. *acephala*, forward primers FC4H5-31 (5′-ACTTACGGATGGGTCAACGAG-3′) and FC4H5-32 (5′-GATCGTGGTGAGGAAGCGT-3′) were designed. FC4H5-31 and GeneRacer 3′ Primer (5′-GCTGTCAACGATACGCTACGTAACG-3′) were used in the primary PCR using 2 μL of total root cDNA as template and were annealed at 62°C and extended for 1.5 min in the cycling. FC4H5-32 and GeneRacer 3′ Nested Primer (5′-CGCTACGTAACGGCATGACAGTG-3′) were used for nested PCR, in which 0.1 μL of the primary PCR product was used as template. Other PCR conditions were the same as the above mentioned. Gel recovery, TA-cloning and sequencing of the PCR product were performed.

2.2.4.2 5′-RACE amplification

Based on the sequence of the *C4H* conserved region, reverse primers RC4H5-51 (5′-CAACTGCAAACGC-TTCCTCAC-3′) and RC4H5-52 (5′-CTCGTTGACCC-ATCCGTAAGT-3′) were designed. RC4H5-51 and GeneRacer 5′ Primer (5′-CGACTGGAGCACGAGG-ACACTGA-3′) were used for primary PCR using 2 μL of total root cDNA as template. 0.1 μL of the primary PCR product was used as template for nested PCR using RC4H5-52 and GeneRacer 5′ Nested Primer (5′-GGACACTGACATGGACTGAAGGAGTA-3′). Other PCR ingredients, cycling procedures, subcloning, etc. were the same as the 3′-RACE handling.

2.2.5 Amplification of the full-length cDNA and genomic DNA sequences

Based on the sequencing results of the 3′ and 5′ RACE products, sense primer FP (5′-GAACTTGAGTTACTT-ATATTAAATAAATAATAAC-3′) and antisense primer RP (5′-AAAGATTTTACACAATATGGCTTATTC-3′) were designed corresponding to the utmost 5′ and 3′ cDNA ends to amplify the full-length cDNA and corresponding genomic sequence. The templates for the amplification of the full-length cDNA and corresponding genomic sequence were 1 μL first strand total cDNA and 1 μL (0.4 μg) genomic DNA, respectively. Other PCR conditions were the same as described in the conserved region amplification. The cycling conditions were: pre-denaturation at 94°C for 4 min, followed by 35 cycles of amplification (94°C for 1 min, 52°C for 1 min, and 72°C for 2.5 min), then followed by 72°C for 10 min. Gel recovery, TA-cloning and sequencing of the PCR products were performed.

2.2.6 Bioinformatic analysis

Sequence alignment, open reading frame (ORF) translation and molecular weight calculation of the predicted protein were carried out with Vector NTI Advance 9.0. BLAST was done at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/), while structural analysis of deduced protein was carried out on the website (http://cn.expasy.org/tools/).

3 Results

3.1 Sequence cloning of *BoC4H*

Electrophoresis detection and optical density (OD) value measurements showed that the concentration and quality of the genomic DNA and root total RNA isolated from *B. oleracea* were suitable for the following handling. A 550-bp band was detected in the amplification of the conserved region from the total genomic DNA, which was in accordance with the length predicted based on *C4H* genes from *A. thaliana* and other plants. The sequenced exact length was 538 bp. Agarose gel electrophoresis of the PCR-amplified reverse-transcribed total cDNA were 100 bp−8 kb, concentrated at 1.5 kb, indicating that the reverse transcription was successful. In the 3′-RACE, the primary PCR product was a wide band larger than 1 kb and the nested PCR product was a specific band of about 1.2 kb with a sequenced exact length of 1079 bp (not including the poly(A) tail). In the 5′-RACE, the primary PCR product and the nested PCR product were a 650-bp wide band and a 400-bp specific band, respectively. The sequenced exact length was 353 bp (RNA Oligo sequence removed). A 1.7-kb band and a 2.5-kb band were amplified in full-length cDNA and corresponding genomic DNA amplifications with exact sequenced lengths of 1715 bp and 2431 bp, respectively. They are identical to each other at the exon regions, indicating that they are from the same gene which is named as *BoC4H* here.

3.2 Analysis of *BoC4H*

3.2.1 Molecular characterization of nucleotide sequence of *BoC4H*

The 2431-bp genomic sequence of *BoC4H* contains 3 exons and 2 introns (910–984 bp and 1113–1753 bp). The length of mRNA is 1715 bp (not counting poly(A) tail). The 5′ untranslated region (5′ UTR) and the 3′ UTR are 124 bp and 131 bp, respectively. From 125 to 1570 bp, the open reading frame (ORF) is 1446 bp, coding a deduced protein of 481 amino acids (Fig. 1). G+C contents of 5′ UTR, 3′ UTR, intron 1 and intron 2 are 35.5%, 35.2%, 33.3% and 31.2%, respectively, while 49.1% for the coding region which is distinctly higher than those of the non-coding regions.

3.2.2 Homologies of *BoC4H*

The National Center for Biotechnology Information (NCBI) BLASTn indicated that *BoC4H* shows convincing identities to other *C4H*s, e.g. 84%–92% to *AtC4H* (ATU71080) in local alignment. Their calculated full-length genomic identity by Vector NTI is only 54.1%, since their intron regions drastically diverged. *BoC4H* shows high BLASTn local identities to the known *C4H* tags from *Brassica* species, e.g. 87%–90% to *B. rapa* cultivar R500 *C4H–BR–1* (AF230678) and *C4H-BR-3* (AF230680), and 86%–89% to *B. oleracea C4H-BO-1–C4H-BO-4* (AF230674–AF230677). *BoC4H* also shows extensive homologies to non-cruciferous *C4H*s, e.g. 77%–78% BLASTn local identities to Asian cotton (*Gossypium arboretum*) *C4H* mRNA (AF286648).

Fig. 1 The full-length *BoC4H* gene and its deduced amino acid sequence

Note: ATG represents the start codon; TAG represents the stop codon; introns are dash-underlined. Solid-underlined amino acids from P_{34} to R_{481} correspond to conserved pfam00067 (p450) domain. Between boxed G_{2241} and T_{2242}, a single-base deletion has occurred on C_{2242} as compared with *C4H*s from other plants. The C-terminal 14 residues in gray background are encoded by the frame-shifted bases after G_{2241}.

3.2.3 Analysis of the deduced BoC4H protein

3.2.3.1 Basic parameters, subcellular localization, and possible post-transcriptional modifications of BoC4H

The 481-aa BoC4H has a molecular weight of 55.38 kD. It is a basic protein with an isoelectric point of 9.06. NetNGlyc 1.0 (Blom et al., 2004) and PROSITE predicted that BoC4H is not N-glycosylated. NetPhos 2.0 predicted that BoC4H has 11, 5 and 3 potential phosphorylation sites of Ser, Thr and Tyr, respectively, implying that phosphorylation might take part in regulating the activity of BoC4H. SignalP 3.0 (Bendtsen et al., 2004) predicted that BoC4H has a probability of 0.473 for a signal peptide and the probability for a signal anchor is 0.518 and the probability of cutting site at G_{28}-K_{29} is 0.362. TargetP 1.1 (Tusnády and Simon, 2001) predicted that the probability for BoC4H to have a signal peptide is 0.859, but no cutting site was predicted. Therefore, BoC4H may have a signal anchor other than a signal peptide. Softberry-ProtComp 6.1 predicted that BoC4H is a membrane protein and is localized in the endoplasmic reticulum with a score of 3.3. TMpred (Hofmann and Stoffel, 1993) predicted that BoC4H has 2 significant transmembrane regions: L_9–P_{26} and S_{444}–V_{463}.

3.2.3.2 Homology analysis of BoC4H

NCBI BLASTp showed that BoC4H has very wide similarities to C4Hs from other plants, with the highest to short C4H tags from *B. rapa* (AAK14957–AAK14959), *B. napus* (AAK14951–AAK14952) and *B. oleracea*

(AAK14953–AAK14955). Besides *Brassica* C4H tags, AtC4H has the highest homology to BoC4H. In pairwise-alignment on Vector NTI, the identities/positives of BoC4H to the C4H proteins from *A. thaliana, C. acuminate*, and Asian cotton were 84.0%/88.9%, 77.6%/84% and 76.4%/83.4%, respectively. In the phylogenetic tree of full-length BoC4H and C4Hs from representative plant taxological species built on Vector NTI Advance 9.0, BoC4H was tightly grouped with AtC4H, then with C4H from *Ginkgo biloba*, a gymnosperm and C4H from *Mesembryanthemum crystallinum*, a dicotyledon, to form a large group. Thereafter, it was grouped with the C4Hs from two monocotyledonous species, *Allium cepa* and *Helianthus tuberosus*, and finally with other dicotyledonous C4Hs (Fig. 2).

3.2.3.3 Prediction of the conserved domains/motifs and active site residues in BoC4H

In NCBI Conserved Domain (CD) search (Marchler-Bauer and Bryant, 2004), two cytochrome P450 conserved domains were detected: pfam00067 (p450) and COG2124 (CypX), located at the regions of P_{34}–R_{481} and L_{31}–R_{481}. BoC4H possesses the specific motifs of cytochrome P450, such as the haem-iron binding domain P_{437}-FGVGRRSCPG$_{447}$, the T-containing binding pocket motif A_{304}AIETT$_{309}$, E_{361}-R_{364}-R_{418} triad, and the hinge motif P_{34}PGP(M/I)PIP$_{41}$ (Fig. 3) necessary for optimal orientation of the enzyme (Rupasinghe et al., 2003). BoC4H also possesses 4 motifs of the 5 specific substrate-binding sites (SRSs) of C4H/CYP73A5: SRS1, SRS2, SRS4 and SRS5, but lacks the C-terminal SRS6 motif KGGQFSLHI because of the frame shift caused by C-terminal single-base deletion (Fig. 3). Furthermore, BoC4H possesses the typical conserved amino acids accounting for activity sites of C4H/CYP73A5 (Hasemann et al., 1995; Schoch et al., 2003), such as I_{109}, K_{113}, V_{118}, F_{220}, N_{302}, R_{366} and so on, but the amino acids of the enzyme activity sites located within SRS6 and C-terminus are lost because of the lack of the C-terminal region including the SRS6.

3.2.4 Secondary and tertiary structures of BoC4H

Predicted by SOPMA (Geourjon and Deléage, 1995), the secondary structure of BoC4H is mainly composed of α-helices (51.35%) and random coils (35.14%), as well as extended strands (9.98%) and β-turns (3.53%). α-helices are mainly distributed at the middle region and the N-terminus (Fig. 4). Prediction of tertiary structure of BoC4H by the Swiss-Model did not yield a result.

4 Discussion

BoC4H was cloned using the cDNA, the product of the reverse transcription, as a template based on the RACE-anchoring amplification technique, implying that *BoC4H* is normally transcribed *in vivo*. The carboxyl terminals of SRS5, SRS6 and SRS4 are important sites to bind the aromatic ring of the substrate, while the amino terminals of SRS1, SRS2 and SRS4 take a very important part in binding the fatty chains of the substrates (Hasemann et al., 1995; Schoch et al., 2003). The lack of SRS6 means that BoC4H loses the ability of binding the substrate or has a very low enzymatic activity. The results of BLASTn of *BoC4H* to *C4H* genes from many other plants show that the reason leading to the loss of SRS6 is the deletion of the cytosine after G_{2241}, thereafter leading to frame shift of ORF of *BoC4H*. Compared with AtC4H, BoC4H loses C-terminal 36 amino acids. The substituting sequence F_{468}–R_{481} has no homology to the known C4H proteins. The Swiss Model could not predict the tertiary structure of BoC4H, therefore, BoC4H protein is likely to have no or a very low enzymatic activity. Repeating the cloning and sequencing of this gene proves that this frame shift mutation is authentic in the *B. oleracea* plants used in our study.

The degeneration mutation of *BoC4H* can not affect the growth and development of the plants of *B. oleracea* because the plants we used are the typical type *B. oleracea* with black seeds and they can normally grow and develop. Fourmann et al. (2002) cloned 4 short tags of *C4H* genes

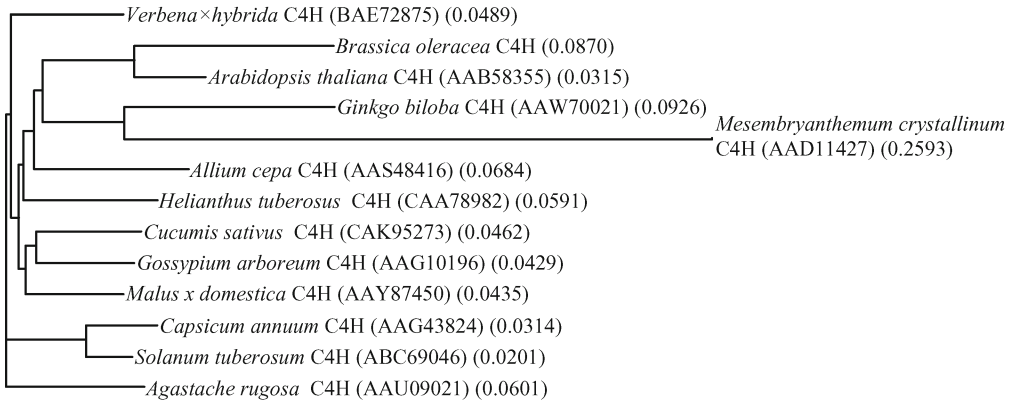

Fig. 2 Phylogenetic tree of BoC4H and C4Hs from other plants

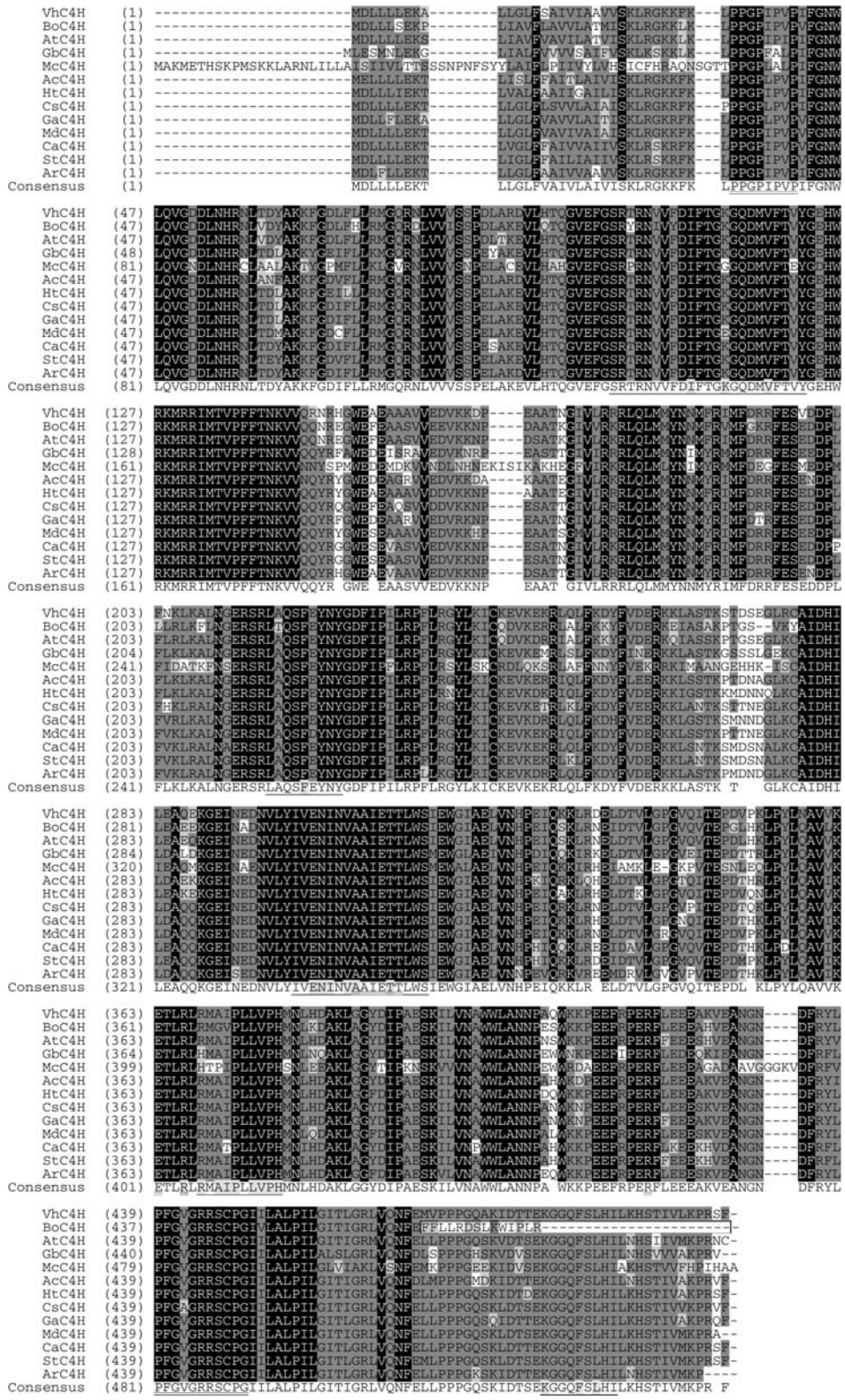

Fig. 3 Multi-alignment of BoC4H with C4Hs from other plants

Note: C4H proteins aligned are the same as those in Fig. 2. Completely identical residues are reverse-displayed, while residues with dark gray, light gray and white backgrounds are conserved, weakly similar and non-similar residues. In the consensus, double-underlined regions/residues indicate P450-featured motifs, i.e. the hinge region, the T-containing binding pocket motif, the E_{361}-R_{364}-R_{418} triad and the haem domain, while single-underlines represent the 5 SRS regions in which the possible active site residues are in gray background. The missing/mutated 36-residue region of the C-terminus of BoC4H is boxed.

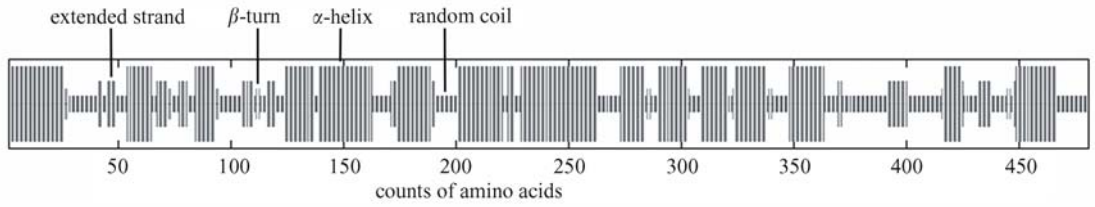

Fig. 4 Secondary structures of BoC4H

from *B. oleracea* during developing *Brassica* molecular markers with the functional genes of *A. thaliana*. The results of BLASTn show that *BoC4H* does not correspond to any of the 4 short tags, and there are enormous sequence differences between *BoC4H* and the 4 short tags, implying that there is a *C4H* gene family with more than 5 members in *B. oleracea*. The disagreement between the subgroups and the taxonomic distances in Fig. 2 also shows that there are several kinds of *C4H* genes in the plant kingdom. Many questions, such as the members containing the standard *C4H* structure, the members possessing the deletion mutation like *BoC4H*, and the differences in function among the members, will be answered through cloning the full-length sequences, molecular characterization and functional identification of all members of the whole *C4H* gene family of *B. oleracea*. The cloning of full-length *BoC4H* will promote the cloning of the full-length sequences of the whole *BoC4H* gene family and the study of the functional effects of the carboxyl terminal deletion of C4H.

Inhibiting the enzyme activity of C4H, the contents of lignin in the transgenic plants will decrease. Regulating the expression of *C4H* gene can also change the compositions of lignin (Sewalt et al., 1997; Blee et al., 2001). Increasing the expression of *C4H* and accelerating the biosynthesis of lignin and phytoalexin can increase the disease resistance, lodging resistance and stress resistance. Theoretically, regulating the expression of *C4H* genes, upwards or downwards, may increase or decrease the flow into the pathway of anthocyanin synthesis, influencing the color of flowers. For the same reason, regulating or controlling the expression of *C4H* genes in the seed coat may regulate and influence the synthesis of the pigments and lignin in the seed coats and improve the properties of seed coat. Cloning *BoC4H* lays a base for commencing on these researches and for studying the evolution of *C4H* genes from the relatives such as *B. rapa* and *B. napus*.

Acknowledgements This research was financially supported by the National Natural Science Foundation of China (Grant No. 30330400) and the Chongqing Municipal Natural Science Foundation, China (No. 8446).

References

Barber M S, Mitchell H J (1997). Regulation of phenylpropanoid metabolism in relation to lignin biosynthesis in plants. Int Rev Cytol, 172: 243–293

Bendtsen J D, Nielsen H, von Heijne G, Brunak S (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Biol, 340: 783–795

Blee K, Choi J W, O'Connell A P, Jupe S C, Schuch W, Lewis N G, Bolwell G P (2001). Antisense and sense expression of cDNA coding for CYP73A15, a class II cinnamate 4-hydroxylase, leads to a delayed and reduced production of lignin in tobacco. Phytochemistry, 57: 1159–1166

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics, 4: 1633–1649

Fourmann M, Barret P, Froger N, Baron C, Charlot F, Delourme R, Brunel D (2002). From *Arabidopsis thaliana* to *Brassica napus*: development of amplified consensus genetic markers (ACGM) for construction of a gene map. Theor Appl Genet, 105: 1196–1206

Harakava R (2005). Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*. Genet Mol Biol, 28(Suppl3): 601–607

Hasemann C A, Kurumbail R G, Boddupalli S S, Peterson J A, Deisenhofer J (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure, 3: 41–62

Hofmann K, Stoffel W (1993). TMbase: A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler, 374: 166

Li B, Liang Y, Chai Y R (2006). Achievements in research on plant cinnamoyl-*CoA* reductase (CCR) genes. Mol Plant Breed, 4(Suppl3): 55–65 (in Chinese)

Marchler-Bauer A, Bryant S H (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res, 32: W327–W331

Rechards E J (1995). Preparation and analysis of DNA. In: Short Protocols in Molecular Biology. Ausubel F M, Brent R, Kingston R E, Moore D D, Seidman J G, Smith J A, Struhl K, eds. New York: John Wiley and Sons

Rupasinghe S, Baudry J, Schuler M A (2003). Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. Protein Eng, 16: 721–731

Schoch G A, Attias R, Le Ret M, Werck-Reichhart D (2003). Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1. Homology guided site-directed mutagenesis. Eur J Biochem, 270: 3684–3695

Seidman C E, Struhl K, Sheen J (1995). *Escherichia coli*, plasmids, and bacteriophages. In: Short Protocols in Molecular Biology. Ausubel F M, Brent R, Kingston R E, Moore D D, Seidman J G, Smith J A, Struhl K, eds. New York: John Wiley and Sons

Sewalt V J H, Ni W, Blount J W, Jung H G, Masoud S A, Howles P A, Lamb C, Dixon R A (1997). Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonialyase or cinnamate 4-hydroxylase. Plant Physiol, 115: 41–50

Tusnády G E, Simon I (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics, 17: 849–850

Winkel-Shirley B (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol, 126: 485–493