Molecular cloning, characterization and expression of two orthologs of *Arabidopsis thaliana phenylalanine ammonia-lyase 1* (*AtPAL1*) from rapeseed (*Brassica napus* L.)

NI Yu, JIANG Honglin, LI Jiana, CHAI Yourong

Chongqing Rapeseed Engineering & Technology Research Center, Southwest University, Chongqing, 400716, China
Agronomy and Biotechnology College, Southwest University, Chongqing, 400716, China
Email: ljn1950@swu.edu.cn

Abstract

Phenylalanine ammonia-lyase (PAL; E.C. 4.3.1.5) plays important roles in determine flavonoid- and lignin-related traits in plants including rapeseed crops. Two *PAL* full-length cDNAs, *BnPAL1-1* and *BnPAL1-2*, were isolated from *Brassica napus* using rapid amplification of cDNA ends (RACE) method. *BnPAL1-1* is 2460 bp with an open reading fame (ORF) of 2169 bp, while *BnPAL1-2* is 2396 bp with an ORF of 2160 bp. They share only 86.8% identities on nucleotide level but as high as 94.9% identities and 97.0% positives on protein level to each other. They are most homologous to *Arabidopsis thaliana PAL1* (*AtPAL1*), then to *AtPAL2*, highly homologous to known dicot *PAL*s, moderately homologous to *AtPAL4*, *AtPAL3* and *PAL*s from monocots, gymnosperms, ferns and mosses, while least homologous to fungal *PAL*s. The deduced 722 aa BnPAL1-1 protein is 78.32 kDa with a pI of 5.90, while the 719 aa BnPAL1-2 is 78.05 kDa with a pI of 5.93. Pair-wise alignment and preferential amino acid analysis indicated that *BnPAL1-1* and *BnPAL1-2* are orthologs of *AtPAL1*, suggesting that the divergence of *AtPAL1* and *AtPAL2* was earlier than the divergence of genera *Arabidopsis* and *Brassica*. Conserved domain search, active sites analysis as well as secondary and tertiary structure predictions proved that *BnPAL1-1* and *BnPAL1-2* encode typical PAL proteins. Southern blot hybridization revealed that in *B. napus* PAL is encoded by a multi-gene family, and there may be 2 to 6 orthologs of *AtPAL1*. Transcripts of *BnPAL1-1* and *BnPAL1-2* are both most abundant in flowers, but *BnPAL1-2* shows stronger tissue-specificity than *BnPAL1-1*. *BnPAL1-1* transcription shows sharp decrease in late stage seeds, while *BnPAL1-2* transcription in stems, leaves and most stages of seeds are obviously lower than in other organs. *Escherichia coli* expressed 6×His-tagged proteins of the two genes were both bioactive, but BnPAL1-2 was much higher in tested catalytic activity than BnPAL1-1.

Key words: *Arabidopsis thaliana*, *Brassica napus* L., Cloning; Expression, Phenylalanine ammonia-lyase (PAL), RACE

Introduction

Phenylalanine ammonia-lyase (PAL; E.C. 4.3.1.5) catalyzes the first committed step in the phenylpropanoid pathway and plays important roles in the accumulation of polyphenol pigments and cell wall lignins. Fluctuations of PAL levels are the key element in regulating phenolic compounds synthesis in plants. PAL has been extensively studied since its first description in 1961. *PAL* genes have been isolated and sequenced from a number of plant species, including gymnosperms, monocots, and herbaceous and woody dicots. To date, however, there has been no report on isolation and characterization of the structure and function of full-length *PAL* gene from *B. napus* though it is a very important oil crop. The 490 bp short sequence (GenBank accession number AF229428) putatively encoding partial codons of *B. napus* PAL was just a genomic DNA fragment derived from an amplified consensus genetic marker research, and in other *Brassica* species there are also only very limited short tags of this kind (AY055752, AF229430 and AF229429). Enzymatic test has shown that PAL activity showed decreased levels in yellow-seeded rape seed coat than in black-seeded lines , indicating possible relatedness of *PAL* gene regulation to seed coat color formation. Here we report the cloning, molecular characterization and expression of two full-length cDNAs (*BnPAL1-1* and *BnPAL1-2*), orthologous to *A. thaliana PAL1*, of *B. napus PAL* gene family.

Material and Methods

Seeds of black-seeded rapeseed (*B. napus*) stock line 5B were kept by Chongqing Rapeseed Technology Research Center and planted in field under normal cultivation. The roots, hypocotyls, cotyledons, stems, leaves, flower buds, flowers, 10 DAF (days after flowering) seeds, 20 DAF seeds and 30 DAF seeds were collected and immediately frozen using liquid nitrogen and stored at –80 °C until genomic DNA and total RNA isolation.

Total genomic DNA was isolated from 5B leaf tissue following the protocols described previously. A CTAB-based method was used to isolate total RNA for RACE and analysis of gene expression in various tissues.

Results

2.1. Cloning and analysis of BnPAL1-1 and BnPAL1-2 full-length cDNAs

The 3′ RACE nested PCR yielded a DNA band of about 300 bp, but it was somewhat wide and obscure. Intended to clone all the possible isoforms, we recovered the whole wide band and *E. coli* colony identification showed polymorphism in

the amplified fragment length of the inserts. Sequencing results yielded two unique 3′ cDNA ends, 244 bp and 187 bp respectively not including poly dA tail, that were most homologous to *AtPAL1* revealed by NCBI blastn. Here the genes corresponding to them are denoted as *BnPAL1-1* and *BnPAL1-2* respectively. Both primer pairs FPALC / RPAL1-1 and FPALC / RPAL1-2 successfully amplified an anticipated band of about 2 kb using root total cDNA as template, sequencing of which turned out to be 2065 bp and 2008 bp respectively. Based on these 2 fragments, isoform-specific 5′ RACE primers were designed and conducted. The 5′ RACE nested PCR amplifications for *BnPAL1-1* and *BnPAL1-2* both resulted in a DNA band of about 500 bp and their sequenced length was 506 bp and 499 bp respectively after excision of the kit artificial fragment corresponding to RNA oligo. They both had an 84 bp completely overlapping region with previously amplified cDNA main regions with only several variant bases in the FPLAC primer region, implying that the 5′ RACE products were the exact 5′ cDNA ends of *BnPAL1-1* and *BnPAL1-2*. Primer pairs FPAL1-1 / RPAL1-1 and FPAL1-2 / RPAL1-2 both successfully amplified a specific cDNA band of about 2.4 kb with an accurate length of 2460 bp and 2396 bp respectively. They were completely consistent with the assembled putative full-length cDNAs of *BnPAL1-1* and *BnPAL1-2*. Conclusively, we have successfully isolated the full-length cDNAs of *BnPAL1-1* and *BnPAL1-2*.

The sequence of *BnPAL1-1* contains a 98 bp 5′ untranslated region (5′ UTR), a 2169 bp open reading frame (ORF) and a 193 bp 3′ untranslated region (3′ UTR). *BnPAL1-2* contains a 100 bp 5′ UTR, a 2160 bp ORF and a 136 bp 3′ UTR. These two full-length cDNAs show 86.8% identities to each other with the UTRs more variable, but NCBI blastn showed that they both are most homologous to *AtPAL1*, then to *AtPAL2*, *PAL* genes from non-cruciferous species, *AtPAL4* and *AtPAL3*. Pair-wise alignment on full-length cDNA level indicated that *BnPAL1-1* is 82.2% and 75.1% identical to *AtPAL1* (NM_129260) and *AtPAL2* (NM_115186), and *BnPAL1-2* is 79.6% and 74.6% identical to *AtPAL1* and *AtPAL2*, respectively. On ORF level, *BnPAL1-1* shows 84.9%, 79.3%, 70.2% and 67.3% identities to *AtPAL1*, *AtPAL2*, *AtPAL4* (AY303130) and *AtPAL3* (NM_120505) respectively; while *BnPAL1-2* shows 84.6%, 79.3%, 71.0% and 66.9% identities to *AtPAL1*, *AtPAL2*, *AtPAL4* and *AtPAL3* respectively. Nucleotide analysis indicated that *BnPAL1-1* and *BnPAL1-2* are both orthologs of *AtPAL1*.

2.2. Characterization of the deduced BnPAL1-1 and BnPAL1-2 proteins

The deduced BnPAL1-1 protein is 722 amino acid residues (aa) in length with a predicted molecular mass (Mw) of 78.32 kDa and a theoretical isoelectric point (pI) of 5.90, while the BnPAL1-2 is 719 aa in length with a Mw of 78.05 kDa and a pI of 5.93. The calculated molecular masses of the two deduced BnPAL1 proteins are in agreement with the estimation of enzyme purified from *A. thaliana* AtPAL1 (78.7 kDa). BnPAL1-1 and BnPAL1-2 are 94.9% identical and 97.0% positive to each other, which are much higher than the identities on nucleotide level. As has been noted in other PAL protein sequences, the most divergence was found in the N-terminal region. The N-terminal region of BnPAL1-2 is 3 aa less than that of BnPAL1-1.

NCBI blasp indicated that both BnPAL1-1 and BnPAL1-2 are most homologous to AtPAL1 (NP_181241), followed by AtPAL2 (NP_190894), AtPAL4 (AAP59440), AtPAL3 (NP_196043), other plant PALs and fungal PALs. They also show certain homologies to histidine ammonia-lyase (HAL) from various microbes and animals with local identities below 36% and positives below 54%. Pair-wise alignment on full protein level indicated that BnPAL1-1 has 94.6%/97.2%, 88.8%/92.8%, 78.4%/86.2% and 69.9%/79.0% identities/positives to AtPAL1, AtPAL2, AtPAL4 and AtPAL3 respectively; and BnPAL1-2 shows 94.2%/96.3%, 88.7%/92.7%, 78.9%/85.8% and 70.2%/78.7% identities/positives to AtPAL1, AtPAL2, AtPAL4 and AtPAL3 respectively. Multi-alignment and phylogenetic analysis indicated that BnPAL1-1 and BnPAL1-2 are most similar to AtPAL1, with a little less similar to AtPAL2. These 4 sequences form a tight subgroup; they are further grouped with PALs from other dicot species, then with AtPAL4 and AtPAL3, to form a dicot large group. PALs from certain monocot species, gymnosperms, ferns, mosses and fungi are grouped together to form another large group. From our analysis, it can be seen that PALs from dicots and fungi are at two polars of evolution, while PALs from monocot, gymnosperms, ferns and mosses lie between them, with their bootstrap values in consensus with their taxonomic relationships. Though BnPAL1-1 and BnPAL1-2 show high homologies to both AtPAL1 and AtPAL2, it is obvious that at many amino acid positions they are preferentially similar to AtPAL1 than to AtPAL2, e.g. the underlined residues A/L, N/G, A/S, V/T, K/T, S/N, H/Qs, I/V, F/L, T/S, N/D, A/S, E/Ks, K/E, S/T, LS/QA, Y/Q, R/S, V/I, Y/F, A/V, I/M, S/T, D/G, A/P, T/V, E/D in multi-alignment panel in.

NCBI Conserved Domain search detected both Phenylalanine ammonia-lyase (PAL, cd01442) and Histidine ammonia-lyase (HAL, cd01441) conserved domains within E_{67}-K_{598} of BnPAL1-1 and E_{64}-K_{595} of BnPAL1-2. But the scores of having a PAL conserved domain (709 and 720 bits respectively) are much higher than those of having a HAL domain (295 and 302 bits respectively). The highly conserved residues especially within catalytic center of PAL were scrutinized through multi-alignment of BnPAL1-1 and BnPAL1-2 with PALs from wide origin including the well characterized model PAL proteins from parsley (*Petroselinum crispum*) and *Rhodosporidium toruloides*. In BnPAL1-1 and BnPAL1-2, all the highly conserved residues are as identical as in AtPAL1 and other plant PALs. All the revealed highly conserved residues especially within the catalytic center of plant PALs are not changed in these two proteins, e.g. the residues corresponding to A_{202}-S_{203}-G_{204} which autocatalytically form the cofactor 4-methylidene-imidazole-5-one (MIO), N_{384}, Q_{348}, R_{354}, N_{260}, F_{400}, L_{138}, L_{206}, Y_{351}, N_{487}, Q_{488}, E_{484}, Y_{110}, and R_{174} from *P. crispum* PAL. Even when compared with the far-relational fungal PAL from *R. toruloides* , most of the conserved residues are also identical, and the non-identical residues in BnPAL1-1 and BnPAL1-2 are identical to AtPAL1 and most plant PALs .

SignalP 3.0 did not predict any signal peptide in BnPAL1-1 and BnPAL1-2. PSORT prediction showed a certainty of 0.7 of locating the two proteins at plasma membrane, but PENCE PA-SUB Server 2.5 absolutely predicted that they are located in

the cytoplasm. TMpred predicted two strong transmembrane helices in both proteins, i.e. L_{69}-S_{89} and A_{282}-M_{300} of BnPAL1-1 while L_{66}-L_{86} and L_{261}-F_{277} of BnPAL1-2. The prediction results of subcellular localization and transmembrane topology of these two proteins are consistent with the characteristics of potato wound-induced PAL. NetPhos 2.0 search predicted 38 significant phosphorylation sites in BnPAL1-1 (21 for S, 9 for T and 8 for Y respectively) and 36 significant phosphorylation sites in BnPAL1-2 (19 for S, 10 for T and 7 for Y respectively). Phosphorylation of PAL has been suggested as a ubiquitous regulatory mechanism in higher plants. It is involved in marking PAL subunits for turnover. The phosphorylation site in French bean PAL is most likely T_{545} in the sequence V_{539}AKRTLTT$_{546}$. The corresponding residues in BnPAL1-1 and BnPAL1-2 are T_{555} and T_{553} respectively. But these two T residues were predicted as non-significant weak sites for phosphorylation, suggesting that in these two proteins the phosphorylation sites may have been shifted to other residues. The relationship between PAL and glycosylation has not been deeply characterized. Enzyme purification study has shown that the fungal PAL from *Ustilago maydis* did not show any evidence of glycosylation, but in potato the wound-induced PAL showed absolutely necessity for glycosylation and inhibition of the N-linked glycosylation by tunicamycin resulted in imperfect protein folding and enzyme inactivation. NetNGlyc 1.0 predicted 4 significant N-glycosylation sites in BnPAL1-1 (N_{24}, N_{266}, N_{448} and N_{649}) and 5 significant N-glycosylation sites in BnPAL1-2 (N_4, N_{21}, N_{263}, N_{458} and N_{646}), suggesting that, as in potato PAL, glycosylation may also be a kind of post-translational modification in the two rapeseed PAL proteins.

SOPMA predicted that in BnPAL1-1 and BnPAL1-2 alpha helices (50.42% and 51.18% respectively) and random coils (30.33% and 29.90% respectively) constitute the most portion of the secondary structure, while extended strands (10.66% and 11.13% respectively) and beta turns (8.59% and 7.79% respectively) contribute a little portion. In most parts of the proteins, the 4 types of the predicted secondary structures distribute in an interlaced manner, but a distinct feature is that there are several large helices, e.g. T_{507}-K_{552} of BnPAL1-1 and T_{504}-K_{549} of BnPAL1-2. The feature of secondary structure of BnPAL1-1 and BnPAL1-2 is quite similar to that of the parsley PAL1.

Swiss-Model prediction indicated that the tertiary structures of BnPAL1-1 and BnPAL1-2 proteins are almost the same. BnPAL1-1 shows 84.55% of identities to both models 1w27A.pdb and 1w27B.pdb, while BnPAL1-2 shows 83.85% of identities to both models 1w27A.pdb and 1w27B.pdb. 1w27A.pdb and 1w27B.pdb are subunit models of suggested homotetrameric quaternary structure of parsley PAL yielded by X-Ray crystal diffraction at a resolution of 1.7 Å .

From above analyses, it is concluded that the *BnPAL1-1* and *BnPAL1-2* genes we cloned are typical *PAL* genes encoding bioactive PAL enzymes and are orthologs of *A. thaliana PAL1* (*AtPAL1*).

Discussion

Previous study has reported that the *A. thaliana PAL1* (*AtPAL1*) expressed most abundantly in roots, so we choose total RNA from *B. napus* roots for RACE. Our results proved the feasibility of this strategy. But RT-PCR analysis of the expression of the two genes cloned here showed that in *B. napus* the root is not the most abundantly expressed tissue for *PAL* transcription. We detected much higher level of transcripts of *BnPAL1-1* and *BnPAL1-2* in opened flowers than in other tissues. Apparently, the brightly yellow petals and pollens is a strong driven force leading to large amount of biosynthesis of flavonoid pigments. It seems that most of the flavonoid pigments in rapeseed mature flowers are synthesized de novo at or just prior to flower opening stage, not by transformation of pre-stored flavonoid pathway intermediates. Our results also indicated that *BnPAL1-1* is more ubiquitous in transcription than *BnPAL1-2* among various organs/tissues, whereas *BnPAL1-2* keeps certain degree of tissue-specificity. This is to say, the promoters of *BnPAL1-1* and *BnPAL1-2* may have occurred certain degrees of divergence from each other in cis regulatory elements responsible for tissue-specificity. However, unlike reported *A. thaliana PAL1* which is most abundantly transcribed in roots, transcription of both *BnPAL1-1* and *BnPAL1-2* in roots is not much more than many other organs such as leaves and cotyledons etc..

In our research, catalytic activity of *E. coli* expressed 6×His-tagged BnPAL1-1 was much lower than that of BnPAL1-2. This may be a real reflection of in planta status of these two enzymes, but also may be resulted from different behaviors of these two proteins under our testing conditions, the fact should be further identified.

Dissection of the molecular mechanism of PAL in determining bio-traits together with genetic engineering of these traits is a focus in PAL research, and hopeful achievements have been made in modifying lignin content or disease resistance of plants. But to date no full-length *PAL* gene has been cloned from any species of genus *Brassica*, and little study has been taken on *Brassica PAL* genes on molecular level. The two *PAL* genes clones here will undoubtedly speed up our further investigation of the relationship of *PAL* genes to some important traits of rapeseed on molecular level, also will enable us to modify *PAL*-related traits of rapeseed through genetic engineering.

References

Omitted.