INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
15–503/2016/18–2–412–418
DOI: 10.17957/IJAB/15.0104
http://www.fspublishers.org

Full Length Article

A Novel *Hydroxymethyldihydropterin Pyrophosphokinase-dihydropteroate Synthase* (*HPPK-DHPS*) Gene from a Nutraceutical Plant Seabuckthorn, Involved in Folate Pathway is Predominantly Expressed in Fruit Tissue

Shazia Arif[1], M. Ramzan Khan[2*], Syed Dilnawaz Ahmed Gardezi[1], Zaheer Abbas[2] and Ghulam Muhammad Ali[2*]
[1]*Faculty of Agriculture Rawalakot Poonch, Azad Jammu and Kashmir University, Muzaffarabad, Azad Kashmir*
[2]*National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Center, Park Road, Islamabad, Pakistan*
[*]For correspondence: drmrkhan_nigab@yahoo.com; drgmali@yahoo.ca

Abstract

Seabuckthorn (*Hippophae rhamnoides* L.) is a hardy fruit-producing plant. With medicinal and nutraceutical properties, its fruits, seeds and leaves are enriched with astonishing array of nutrients. Especially, its bark contains vitamins, essential oils and minerals, for which most of the staple crops are deficient. This study was envisaged to isolate the folate pathway *HPPK-DHPS* ortholog from seabuckthorn. The target gene Hydroxymethyldihydropterin pyrophosphokinase–dihydropteroate synthase linked in the pathway of folate biosynthesis was successfully isolated and cloned. The sequences analysis revealed that this novel genomic locus is 2354 bp in size. The coding region is interrupted by a single large intron. The coding sequence is 1539 bp long which is similar to its ortholog in tomato. The newly isolated gene has prominent nucleotide mutations randomly distributed over the entire length of the sequence. Mutations are only of substitution type, and no deletions or insertions are detectable. All the mutations were of non-synonymous type. Expression profile of *HrHPPK-DHPS* with semi-quantitative RT-PCR revealed the higher accumulation of transcripts in leaf and fruits tissues. Phylogenetic reconstruction revealed gene duplication in soybean and Arabidopsis orthologs. A considerable conservation in 3-dimensional structure of folate protein HPPK-DHPS was observed; however, notable differences in substrate binding pockets were also visible. Hence, *HrHPPK-DHPS* is a novel ortholog isolated from medicinal plant seabuckthorn.

Keywords: Folate; HPPK-DHPS; Seabuckthorn; Gene expression; Biofortification

Introduction

Seabuckthorn (*Hippophae rhamnoides* L.) is a hardy fruit producing plant of the family *Elaeganaceae*, naturally distributed in Asia and Europe. The fruit of seabuckthorn is the store house of vitamins and other important bioactive substances (Xurong *et al*., 2001). Seabuckthorn berries contain appreciable levels of vitamin B1, P and K (Lu, 1992) and folate vitamers (Gutzeit *et al*., 2008). Micronutrient deficiency is preventing an estimated one third of the world population from reaching their physical and intellectual potentials.

Folate or folic acid, a water-soluble vitamin B is identified as advantageous to human health. Seabuckthorn fruit is considered as high source of folate found in the range of 29 µg/100 g fresh weight (Marcus, 2005; Virendra, 2006). It has been reported previously that total folate contents in seabuckthorn berries is found in the range of about 39 µg/100 g, and 5-methyltetrohydrofolate was illustrated as the principal folate vitamers found in seabuckthorn berry. Tracings of tetrahydrofolate were noticed, although distantly lower than the detecting point (Stralsjo *et al*., 2003). Total folate substances of seabuckthorn fruit berries and juice analyzed ranged from 29 up to 81 µg/100 g. Comparing seabuckthorn folate contents with other berries, reveals that the maximum folate contents of 96 µg/100 g are investigated amongst these berries (Stralsjo *et al*., 2003; Gutzeit *et al*., 2008).

Folate is a B-group vitamin critical for normal cellular function and division. Its deficiency in humans leads to a number of serious diseases, most notably anemia and neural tube defects in newborns (Lucock, 2000), and mental disorders such as psychiatric syndromes, and decreased cognitive performance among the elders (Calvaresi and Bryan, 2001; Hultberg *et al*., 2001). In addition, higher risk of cardio -vascular diseases seems to correlate with high Hcy levels in blood (Quinlivan *et al*., 2002) caused by the inefficient re-methylation of Hcy upon folate deficiency (Stover, 2004). Increased plasma-Hcy content may also be a risk factor for different forms of dementia including Alzheimer's disease (Seshadri *et al*., 2002).

There are a number of genes coding for vitamins particularly folate in this nutraceautical plant. The plant bi-functional *HPPK-DHPS* is an important enzyme in

To cite this paper: Arif, S., M.R. Khan, S.D.A. Gardezi, Z. Abbas and G.M. Ali, 2016. A novel *hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase* (*HPPK-DHPS*) gene from a nutraceutical plant seabuckthorn, involved in folate pathway is predominantly expressed in fruit tissue. *Int. J. Agric. Biol*., 18: 412–418

the folate biosynthetic pathway. Tetrahydrofolate and its one-carbon derivatives, collectively termed folates, are essential cofactors for one-carbon transfer reactions in the biosynthesis of certain amino acids, pantothenate, purines, and thymidylate (Hanson *et al*., 2000; Ravanel *et al*., 2004). During organ differentiation, folate is synthesized-preferentially in highly dividing tissues and in photosynthetic leaves. This is associated with high levels of the *HPPK-DHPS* protein and a pool of folate, 3- to 5-fold higher than in the rest of the plant.

In plant *HPPK-DHPS*, the *DHPS* reaction is feedback inhibited by dihydropteroate dihydrofolate and THF-Glu suggesting that this domain could be a potential regulatory point of the mitochondrial branch of folate pathway (Mouillon *et al*., 2002). Arabidopsis is unique among higher plants with sequenced genome in having two genes coding for *HPPK-DHPS* (Storozhenko *et al*., 2007). The first one encodes the mitochondrial isoform involved in *de novo* synthesis of THF, and the second is highly expressed in developing seeds and encodes a cytosolic enzyme whose function remains to be established. The third mitochondrial step is the ATP dependent attachment of glumate to the carboxyl moiety of pABA to form dihydrofolate. It is catalyzed by a mono-functional dihydrofolate synthase (Ravanel *et al*., 2001).

Seabuckthorn has a great nutraceautical potential for genetic improvement of cereal crops through genetic engineering, hitherto little efforts has been made in isolating folate genes from this plant. But before this application it is important to isolate and analyze these genes and to study their expression patterns using molecular techniques. Therefore we isolated and clone *HPPK-DHPS* gene, to analyze the sequence of this gene using bioinformatics tools, to study the expression patterns of this genes in various tissues at different developmental stages and comparative homology modeling for deciphering the 3D structure of HPPK-DHPS protein.

Materials and Methods

Seabuckthorn (*Hippophae rhamnoides* ssp. *Sinensis*) plants were grown in glass house and the fields of National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Centre, Islamabad Pakistan. For nucleic acid extraction, leaf, berry, bud and seed samples were harvested from plants at different bud, leaf, fruit and seed development stages. Fruit was harvested once or twice a week during the ripening period from seabuckthorn nursery, frozen in liquid nitrogen and stored at -80°C until processed for nucleic acid extraction and expression analysis.

Designing of Primers

Nucleotide sequences of *HPPK-DHPS* genes were retrieved from National Center for Biotechnology Information (NCBI) database. A pair of gene specific primers designed from the conserved region of *HPPK-DHPS* for the amplification of full length coding region of *HrHPPK-DHPS* gene. Primers for expression analysis through RT-PCR were designed from newly isolated sequence of *HrHPPK-DHPS* gene. The sequences of the primers used in the study are listed in Table 1.

DNA Extraction

Advancements in functional genomics has revealed that there is positive correlation between number of introns found and gene expression, furthermore, introns containing transgenes showed enhancements of gene expression than the same gene lacking introns (Brinster *et al.*, 1988). Therefore we set out to isolate the *HrHPPK-DHPS* gene from DNA instead of cDNA. Genomic DNA was isolated from the leaf samples of seabuckthorn using a standard CTAB protocol (Doyle and Doyle, 1987).

PCR Amplification and Molecular Cloning

Diluted DNA was used as template for standard PCR amplification for *HrHPPK-DHPS* gene amplification using the Takara Ex Taq™ polymerase (Cat# RR001). PCR amplification was performed in a total reaction volume of 50 µL using the gene specific primers with initial denaturation step of 94°C for 3 min, amplification with 35 cycles of 94°C for 60s, annealing at 57°C for 75s, elongation at 72°C for 145s, followed by final extension at 72°C for 10 min. The amplified product of *HPPK-DHPS* was gel purified using the Gene JET™ Gel extraction kit (Thermo Fisher Scientific USA) and cloned into the pTZ57R/T vector (Fermentas USA) for sequencing from MACROGEN (Korea).

Total RNA Isolation and Gene expression Analysis

Total RNA from different tissues of seabuckthorn was extracted using the TRIZOL® Reagent (Invitrogen USA). It was quantified with Thermo Scientific's NanoDropTM Lite (USA). Quality of extracted RNA was assessed on 1.5% agarose gel. The first strand cDNA was synthesized using AMV-RT reverse transcriptase (Fermentas USA #EP0641). Multiplex semi-quantitative RT-PCR was carried out to discern the differential expression patterns of newly isolated *HrHPPK-DHPS* gene transcripts in leaf, fruit bud and seed of seabuckthorn (Khan *et al*., 2013). The *18SrRNA* gene was used as an internal control. The reaction was repeated thrice for both the biological as well as technical replicates. The amplified products were resolved on 2% agarose gel and photographed.

Sequence Analysis and Phylogenetic Reconstruction

The new gene sequence was edited using Mac Vector™ 7.2.3 software (Accerlrys Inc.) gcg/Wisconsin Package,

Table 1: Detail of primers used for *Hr-HPPK-DHPS* gene amplification and expression studies

Gene	Primer name	Primer sequence (5'-3')	Application
HPPK-DHPS	HPPK-F	ATGAATGTCTTCAGGCGTTTTCTGCCC	Cloning
HPPK-DHPS	HPPK-R.	TTAACTGTTTTGTCTCTGTTTTCTTT	Cloning
HPPK-DHPS	HPPK-F	ATGTCTGCCTTGTGTATCTA	Expression analysis
HPPK-DHPS	HPPK-R	TTATACATCCATCAAGATTC	Expression analysis
18sRNA	18s RNA-F	GTAACCCGTTGAACCCCATT	Internal control
18sRNA	18s RNA-R	CCATCCAATCGGTAGTAGCG	Internal control
M13	M13 Seq -F	GTAAAACGACGGCCAGTG	Colony PCR & plasmid PCR
M13	M13 Seq- R	GGAAACAGCTATGACCATG	Colony PCR & plasmid PCR

university of Wisconsin USA). After the sequences were retrieved for each gene family, the sequences and phylogenetic analyses were performed using Mac Vector version 7.2.3, and MEGA 5 (Kumar *et al*., 2008). Clustal W (Thompson *et al*., 1994) in Mac Vector 7.2 was used as a multiple alignment tool for aligning the amino acids. The nucleotide sequences were translated in the Mac Vector program. The default parameters were used. Sequences too diverged disturbing the alignments were removed manually while performing the multiple alignment procedure. Using the alignment file phylogenetic tree was constructed using the Neighbor Joining method (Saitou and Nei, 1987; Russo *et al*., 1996) implemented in MEGA5. The complete deletion option was used to exclude any site which postulated a gap in the sequence. As amino acid substitution models, uncorrected proportion (p) of amino acid differences and distance were used. The bootstrap method (Felsenstein, 1985) was used to test the authenticity of the generated tree topologies. The calculation at 1000 pseudo replicates gave the bootstrap probability of each interior branch in the tree.

Homology Modeling and Protein Structure Prediction

The three dimensional (3D) structural modeling of folate protein (HrHPPK-DHPS) was performed by MODELLAR 9V10 homology modeling program. The amino acid sequence of HrHPPK-DHPS (512aa) retrieved in FASTA format was subjected to protein-protein blast (Blastp) search against the protein data bank (PDB) to identify a appropriate template structure for comparative modeling. The 2BMB_A was preferred as a proper template with query sequence having 33% identity, 89% query coverage and E-value of 7e-70. The modeler quality was assessed based on sequence alignment by Blast and template structure used. The structure was predicted by MODELLAR with alpha helices and beta-pleated sheets visualized by Chimera 1.6.

Results

A New *HPPK-DHPS* Sequence Identified from Seabuckthorn

In order to isolate *HPPK-DHPS* ortholog, gene specific primers were used to amplify dihydropterin gene fragment

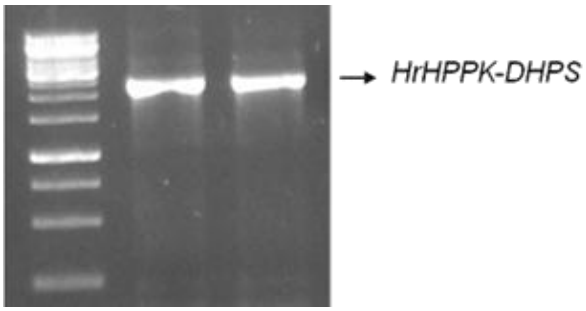

Fig. 1: Amplification and accumulation of *HPPK-DHPS* gene. PCR amplification of full length sequence of folate gene (*HPPK-DHPS*) containing intron from seabuckthorn genomic DNA

using genomic DNA as template. A full length 2354 bp target genomic fragment was amplified. The distinct high molecular weight band of *HPPK-DHPS* observed on agarose gel (Fig. 1) was purified, cloned and sequenced.

The newly isolated sequence was edited using Mac Vector™ 7.2.3, software (Accerlrys Inc.) gcg/Wisconsin Package university of Wisconsin USA). Blast results exhibited maximum hits with *HPPK-DHPS* gene of *S. lycopersicom* followed by *P. sativum, M. truncatula, C. arietinum, G. max, A. thaliana, A. lyrata, T. aestivum, B. distachyo* and *O. sativa*, thereby confirming the presence of new sequence of *HPPK-DHPS* gene in seabuckthorn. All positions of codon containing gaps and missing data were eliminated. The length of the sequenced clone was 2.3 kb due to the presence of a single large intron. The 1539 bp coding sequence was extracted and found to be exactly equal in length to tomato gene that codes for 513aa. Thus, *HrHPPK-DHPS* gene sequence revealed no variation in the length of the nucleotide sequence. Clustal W multiple alignments of *HPPK-DHPS* gene with the original gene sequence were generated and considerable variability was observed in the sequence of this novel gene. A total of 35 nucleotide mutations were detectable in the final dataset. Mutations in the sequence were only of substitution type, and no deletion and insertion were found. The mutations found were randomly distributed in the entire sequence without specific domain and as these mutations resulted in changes in the protein structure (non-synonymous mutations). The amino acid sequence deduced from the nucleotide sequence indicates that the product of

translation is a precursor of about 512 amino acid residues. The sequence identity was 97% at amino acid level. *HPPK-DHPS* protein sequences were retrieved from NCBI protein database of different species. Multiple sequence alignment has shown the high sequence similarity of *HrHPPK-DHPS* translated amino acid sequence with different proteins from number of plant species including tomato, *Arabidopsis*, soybean and wheat. Sequence similarities and difference among selected protein homologue sequences are shown in Fig. 2.

Phylogenetic Relationship of Seabuckthorn *HPPK-DHPS* Gene with other Plants Species

In order to probe the evolutionary relationship between newly isolated sequence and those in the database, phylogenetic reconstruction was carried out. For this purpose the coding sequences of *HPPK-DHPS* homologue from different plant species such as *S. lycopersicum, T. aestivum B. distachyon, C. arietinum, O. sativa, A. thaliana, P. sativum, M. truncatula, G. max* and *A. lyrata* were employed. A neighbour-joining tree in Fig. 3 demonstrates that sequences are clearly differentiated into different clades. Four groups comprising brassicaeceous *HPPK-DHPS*, cereals, pea family and tomato as a member of Solanaceae are observed. The un-rooted tree clearly separates out *HPPK-DHPS* of seabuckthorn and establishes its relationship with tomato. This indicates divergence in sequences. Interestingly two gene duplication events were observed. *A. thaliana* and *G. max* showed two sequences of *HPPK-DHPS* on different chromosomes; therefore this is gene duplication as species are diploid. This is very interesting finding as duplications are important for evolutionary natural selection. The monocots *i.e. T. aestivum, B. distachyon, O. sativa,* seems to be the progenitors as present at the base. Higher bootstrap values are indicative of increased reliability of the tree. These results demonstrated that *HPPK-DHPS* genes are well differentiated in different clusters. The newly isolated *HrHPPK-DHPS* sequence though shows clustering with tomato genes, however, divergence do exist.

The comparison of cumulative data from cDNA sequence, amino acid sequence, phylogenetic reconstruction and expression pattern demonstrated that there is a considerable variability in the gene structure and evolutionary relationship of *HrHPPK-DHPS* with other homologs. To what extent mutations in *HrHPPK-DHPS* contribute to functional divergence, characterization through site-directed mutagenesis and transformation is necessary. But before transformation, it is paramount to do the expression analysis of *HrHPPK-DHPS* in different tissues of the seabuckthorn.

***HPPK-DHPS* is Differentially Expressed in Seabuckthorn Tissues**

The relative abundance of the total RNA coding for

Fig. 2: Clustal W nucleotide alignment of newly isolated *HPPK-DHPS* amino acid sequence with Mac Vector^{TM}7.2.3. (Accerlrys Inc.) gcg/Wisconsin pakage university of Wisconsin) from seabuckthorn and its ortholog from *S. lycopersicum, G. max, A. thaliana* and *T. aestivum.* Amino acids in blocks highlight similarity

HrHPPK-DHPS was estimated by quantitative RT-PCR. In order to discern the expression patterns of *HPPK-DHPS* gene in different tissues of seabuckthorn, total RNA from four different tissues like leaf, fruit bud and seed were extracted. Normalization was done using constitutive expression *18SrRNA* gene. cDNA was synthesized from equal amount of total RNA. The transcript signals of *HrHPPK-DHPS* were strongly detectable in the fruit and leaf tissues (Fig. 4) and RNA signals barely detectable in fruit bud. Nevertheless, weak expression of this gene was found in seed. Hence, differences in expression patterns in different tissues are quite obvious and *HrHPPK-DHPS* gene is strongly expressed in metabolically active part of the plant tissues.

Tertiary Structure of Folate Protein

The 3D structure of selected original *Saccharomyces cerevisiae* gene templates (Lawrence *et al*., 2005) were used as starting point for homology modeling (Fig. 5A). Modeled tertiary protein structure of different plant species including seabuckthorn *HPPK-DHPS* (B), *S. lycopersicum (*C), *A. thaliana (*D), *G. max* (E) and *T. aestivum* (F) were compared and shown in Fig. 5, respectively. The domain architecture consisting of Pterin domain from 230–498

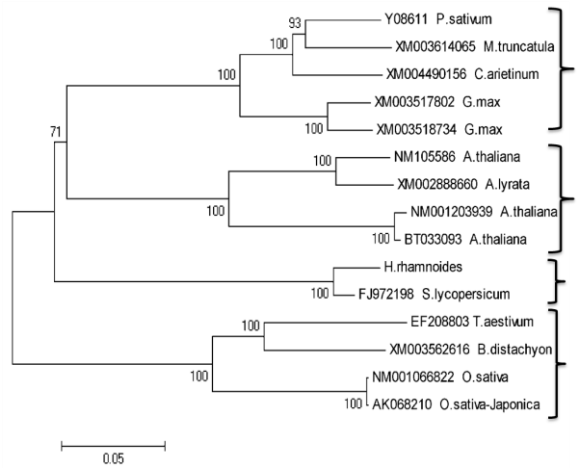

Fig. 3: Phylogenetic reconstruction of *HPPK-DHPS* gene. Neighbor Joining tree is constructed using MEGA5.0 software. Values on the nodes indicate the bootstrap replication of 1000. Phylogenetic inference of enzymes containing the two domains *HPPK-DHPS* was made using 15 complete coding sequences *H. rhamnoides, S. lycopersicum* (GenBank accession No. FJ972198) *T. aestivum* (EF208803), *B. distachyon* (XM003562616), *O. sativa* (NM001066822), *O. sativa* (AK068210), *P. sativum* (Y08611), *M. truncatula* (XM003614065), *C. arietinum (*XM004490156) *G. max* (XM003518734), *G. max (*XM003517802), *A. thaliana* (NM105586), *A. thaliana* (NM001203939), *A. lyrata* (XM002888660) and *A. thaliana* (BT033093) were included in the alignment. Four clusters are enclosed by brackets. A scale is given at the bottom

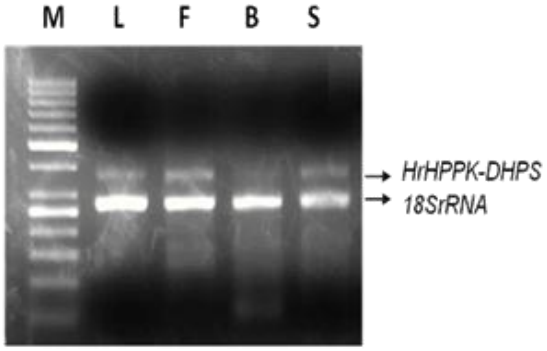

Fig. 4: *HPPK-DHPS* gene expression analysis using RT-PCR. RT-PCR analysis of *HPPK-DHPS* gene transcripts in different tissues of seabuckthorn is shown. M stands for 50bp ladder; L, leaf; F, Fruit; B, Bud; S, Seed. Expression was normalized with *18SrRNA* gene control

residues and *HPPK* domain from 47–172 amino acid residues was evaluated using the Chimera program for seabuckthorn protein. *HPPK* is the upstream and the adjacent enzyme to *DHPS* in the folate biosynthetic pathway. All the 3D structures of 5 different protein fold

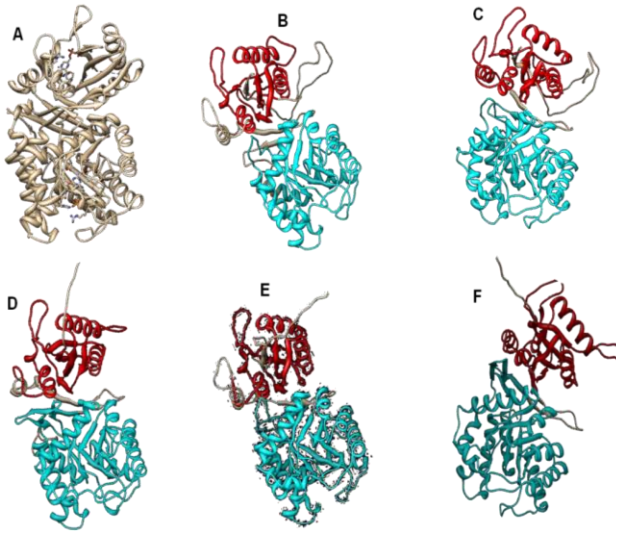

Fig. 5: Predicted 3D structure of HPPK-DHPS proteins from different plant species shown at the same scale. **(A)** Crystal structure of the template bifunctional 6-hydroxymethyl-7, 8- dihydroxypterin pyrophosphokinase dihydropteroate synthase from *Saccharomyces cerevisiae* (Lawrence *et al.*, 2005). (**B)** 3D structure of seabuckthorn HPPK-DHPS produced using MODELER 9v10. (**C)** 3D model of *S. lycopersicum.* **(D)** 3D model of *A. thaliana.* (**E)** 3D Model of *G. max.* (**F)** Model of *T. aestivum.* Models in different protein structure were colored according to different domains. Pterin binding domain DHPS was given blue color, HPPK domain was shown by red color. Inter-domain linker was given grey color

classes ranged in size from 512–561 amino acid residues. The ~ 250-residue pterin-binding *(DHPS)* domains were found to have (beta/alpha) 8 barrel fold, which adopted shape of a distorted cylinder. It has eight alpha-helices gathered around the outside of an inner cylinder of parallel beta-strands. It is partially buried within the (beta/alpha) 8 barrel. The pterin binding residues are highly conserved among five species. The 3D structure of the target protein was also searched from protein data bank (PDB) but no 3D structure of folate protein from plant has been reported to date. The 3-dimentional structure can better predict the function of the protein. The work revealed the significant structural variation in petering binding domain among the various orthologus proteins. We detect a slight but noticeable difference with respect to the substrate binding pockets opening that might be responsible for structure and function of this protein.

Discussion

Seabuckthorn is a super vitamin rich in antioxidants wild plant. Its berries contain a large variety of substances which possess a strong biological activity. Since inadequate folate

intake is a worldwide problem in human nutrition, hence, raising folate levels in grain is a target for metabolic engineering (Scott *et al*., 2000; Bouis, 2002). To solve the problem of micronutrient deficiency, the fortification of staple food has proven to be very efficient for certain micronutrients. A practicable approach is genetic engineering to create a cereal crop plant producing higher levels of folate. In this scenario, availability of complete plant genomes, the growing number of ESTs, and the advances made in HPLC analysis of folates hold great promise (Hanson and Gregory, 2002).

The isolation and cloning of a single gene (*HPPK-DHPS*) in folate biosynthesis pathway was successfully carried out for the first time from seabuckthorn in this study. A full length fragment of *HPPK-DHPS* gene containing large single intron was obtained using standard PCR amplification. The *HPPK-DHPS* gene has already been cloned from several plants including *Arabidopsis*, other dicots and monocots. As there were no deletions and insertions in the extracted coding sequence, the size of newly isolated sequence remained conserved. The length of coding sequence was restricted to 1539 bp for this gene. The amino acid sequence deduced from the nucleotide sequence indicates that the product of translation is a precursor of about 512 amino acid residues. The homology analysis using Clustal W alignments showed that *HrHPPK-DHPS* sequence was 97% identical to sequence of *S. lycopersicom* cultivar (*SlHPPK-DHPS*). The *HrHPPK-DHPS* fragment cloned in this study contained a large single intron. This will be further cloned in expression vector for maximum expression in transgenic plant. It is well established that introns results in increased stability and expression of genes (Buchman and Berg, 1988; Okkema *et al*., 1993). Introns can significantly affect gene expression in plants and many other eukaryotes in a variety of ways. Some introns contain enhancer elements or alternative promoters, while many others elevate mRNA accumulation by a different process.

It is obvious from expression study that there was a remarkable difference in the accumulation pattern of *HPPK-DHPS* between different stages of development. The expression was not detectable in bud while its accumulation gradually lowered in seed but in the fruit and leaf tissues transcript expression was rather similar and considerably higher than other tissues. Expression of *HPPK-DHPS* gene in tomato, *Arabidopsis* and pea in dissimilar tissues and at different developmental stages has been reported (Fabrice *et al*., 1997).

The proteins involved in the folate synthesis are different from one species to another. The phylogenetic analysis differentiated the *HPPK-DHPS* genes from different species into 4 different groups. *HPPK-DHPS* from seabuckthorn conglomerated with *S. lycopersicum* gene. It can be inferred that these two proteins might have the same properties. But functional characterization of these two proteins through ectopic expression or knocking down is required to validate this hypothesis. Nevertheless, the divergence among different species exists due to their clustering. Interestingly duplication of *HPPK-DHPS* gene in some plants is also observed. This might provide the evolutionary advantage to this family during selection processes.

The evolutionary trajectory of a protein through sequence analysis is constrained by its function. *HPPK* is the upstream and adjacent enzyme to *DHPs* in the folate biosynthetic pathway. The polypeptide consist of two structural domains which is its respective mono-functional *HPPK* and *DHPs* domains (Sandeep *et al*., 2012), this allowed us to construct a 3D protein structure of *H. rhamnoides* by applying homology modeling approach. The rationale behind this method is the observation that proteins with similar sequence usually share the same overall 3D folding patterns. Therefore, experimentally determined protein structures were usually used as templates to predict the confirmation of another protein with a similar amino acid sequence (Havel and Snow, 1991; David, 1997). This work also revealed the significant structural variation in petering binding domain among the various orthologus proteins. We detect a slight but noticeable difference with respect to the substrate binding pockets opening. There might be a clue about functional attributes regarding 3D structure of different binding proteins that is triggering the function. The eminence of homology model performed through MODELLER and Chimera program depends on the quality of the sequence alignment by Blast and template protein structure used. The structural analysis and figures generated with MODELLER software (Swiss Institute of Bioinformatics Switzerland) were presented in Fig. 5. These structures provide a good foundation for functional analysis of genes.

Based on structural and experimental analysis we posit that there exists a novel *HPPK-DHPS* gene in seabuckthorn assumed to be recruited in folate synthesis pathway. No deletion or insertions are detectable, however, non-synonymous substitutions are found in the entire sequence. Expression profile of *HrHPPK-DHPS* with quantitative RT-PCR flaunted the higher accumulation of transcripts in leaf and fruits tissues. Phylogenetic reconstruction revealed that *HPPK-DHPS* from seabuckthorn is sister to tomato ortholog and there seems to be gene duplication in *Arabidopsis* and soybean orthologs. In spite of considerable similarities in 3D structures, there exist variations in substrates binding sites that may confer different functions.

Conclusion

The current investigation present evidences for varying expression levels of the bi-functional *HPPK-DHPS* on the folate status in plants. Based on sequence similarity, clustering and phylogenetic differentiation of genes, it is concluded that *HPPK-DHPS* isolated from seabuckthorn is a novel gene isolated from this plant. In this study we predicted and compared three dimensional structure of

HPPK-DHPS protein, which was not available in plant species. This study has future implications in exploring the potential of wild medicinal plants for genetic improvement of staple food crops through biofortification to meet the folate deficiency in human diet.

Acknowledgment

This research was supported by National Institute for Genomics and Advanced Biotechnology project and partially by ingenuous Ph.D. scholarship as well as six month IRSIP program awarded to SA by Higher Education Commission of Pakistan. We are grateful to Mr. Noman Iftikhar (NIGAB) for help in generating 3-dimensional structure of HrHPPK-DHPS protein and two anonymous reviewers for valuable comments and suggestions.

References

Brinster, R.L., J.M. Allen, R.R. Behringer, R.E. Gelinas and R.D. Palmiter, 1988. Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. USA*, 85: 836–840

Bouis, H.E., 2002. Plant breeding: a new tool for fighting micronutrient malnutrition. *J. Nutr*., 132: 491S–494S

Buchman, A.R. and P. Berg, 1988. Comparison of intron-dependent and intron-independent gene expression. *Mol. Cell Biol.,* 8: 4395–4405

Calvaresi, E. and J. Bryan, 2001. B vitamins, cognition, and aging: a review. *J. Gerontol. B. Psychol. Sci.*, 56: 327–339

David, T.J., 1997. Progression in protein structure prediction. *Curr. Opin. Str. Biol*., 7: 377–387

Doyle, J.J. and J.L. Doyle, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*., 19: 11–15

Fabrice, R., D. Macherel, J.M. Mouillon, J.R. Garin and D. Roland, 1997. Folate biosynthesis in higher plants: purification and molecular cloning of a bifunctional 6-hydroxymethyl 7,8- dihydropterin pyrophosphokinase/7,8-dihydropteroate synthase Localized in mitochondria. *EMBO J*., 16: 947–957

Felsenstein, J., 1985. Confidence limits on Phylogenies: An approach using the Bootstrap. *Evolution*, 39: 783–791

Gutzeit, D., S. Monch, G. Jerz, P. Winterhalter and M. Rychlik, 2008. Folate content in seabuckthorn berries and related products (*Hippophae rhamnoides* L. ssp. *rhamnoides*): LC-MS/MS determination of folate vitamers stability influenced by processing and storage assessed by stable isotope dilution assay. *Anal Bioanal. Chem*., 391: 211–219

Hanson, A.D., D.A. Gage and Y. Shachar-Hill, 2000. Plant one carbon metabolism and its engineering. *Trends Plant Sci*., 5: 206–213

Hanson, A.D. and J.F. Gregory III, 2002. Synthesis and turnover of folates in plants. *Curr. Opin. Plant Biol*., 5: 244–249

Havel, T.F. and M.E. Snow, 1991. A new method for building protein conformations from sequence alignments with homologous of known structures. *J. Mol. Biol*., 217: 1–7

Hultberg, B., A. Isaksson, K. Nilsson and L. Gustafson, 2001. Markers for the functional availability of cobalamin/folate and their association with neuropsychiatric symptoms in the elderly. *Int. J. Geriatr. Psychiatr*., 16: 873–878

Khan, M.R., I. Khan and G.M. Ali, 2013. MPF2-like MADS-box genes affecting SOC1 and MAF1 expression are implicated in flowering time control. *Mol. Biotechnol*., 54: 25–36

Kumar, S., J. Dudley, M. Nei and K. Tamura, 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*., 9: 299–306

Lawrence, M.C., P. Iliades, R.T. Fernley, J. Berglez, P.A. Pilling and I.G. Macreadie, 2005. The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of *Saccharomyces cerevisiae*. *J. Mol. Biol.*, 348: 655–670

Lucock, M., 2000. Folic acid: nutritional biochemistry, molecular biology, and role in disease processes. *Mol. Genet. Metab*., 71: 121–138

Lu, R., 1992. *Seabuckthorn-A Multipurpose Plant Ostrobothnia in Finland*. COST action 836, Meeting species for fragile mountains, ICIMOD Occasional WG 6 Cane and bush fruits. Institute of Food Paper No. 20. Katmandu, Nepal, 62p. *Chemistry and Technology*, Graz, Austria, p: 19

Marcus, C.C.W.E., 2005. Seabuckthorn - *Hippophae rhamnoides* L. *A Whole Food Answer to Better Nutrition*. riyasilver.trdinfotech.com/seabuckthorn.htm.

Mouillon, J.M., S. Ravanel, R. Douce and F. Rébeillé, 2002. Folate synthesis in higher plant mitochondria: coupling between the dihydropterin pyrophosphokinase and the the dihydropteroate synthase activities. *Biochem. J.*, 363: 313–319

Okkema, P.G., S.W. Harrison, V. Plunger, A. Aryana and A. Fire, 1993. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics*, 135: 385–404

Quinlivan, E.P., J. McPartlin, H. McNulty, M. Ward, J.J. Strain, D.G. Weir and J.M. Scott, 2002. Importance of both folic acid and vitamin B12 in reduction of risk of vascular disease. *Lancet*, 359: 227–228

Ravanel, S., H. Cherest, S. Jabrin, D. Grunwald, Y. Surdin-Kerjan, R. Douce and F. Rebeille, 2001. Tetrahydrofolate biosynthesis in plants; molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, 98: 15360–15365

Ravanel, S., R. Douce and F. Rébeillé, 2004. The uniqueness of tetrahydrofolate synthesis and one-carbon metabolism in plants. Advances in Photosynthesis and Respiration. *In*: *Plant Mitochondria: From Genome to Function*, Vol. 17, pp: 277–292. Day D.A., Millar H. and Whelan J. (eds.) Great Britain: Kluwer, The Netherlands

Russo, C.A.M., N. Takezaki and M. Nei, 1996. Efficiencies of different genes and tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol*., 13: 525–536

Saitou, N. and M. Nei, 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., 4: 406–425

Sandeep, C., D. Olan, M.C. Brett, N. Janet, S.S. Jamie, G.M. Ian, F. Ross, S.P. Thomas and D.S. James, 2012. Structure of *S. aureus HPPK* and the Discovery of a New Substrate Site Inhibitor Inhibitor. *PLoS ONE*., 7: e29444

Scott, J.M., F. Rbeill and J. Fletcher, 2000. Folic acid and folates the feasibility of nutritional enhancement in plant foods. *J. Sci. Food Agric*., 80: 795–824

Seshadri, S., A. Beiser, J. Selhub, P.F. Jacques, I.H. Rosenberg, R.B. D'Agostino, P.W. Wilson and P.A. Wolf, 2002. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *N. Engl. J. Med*., 346: 476–483

Stralsjo, L., H. Ahlin, C.M. Witthoft and J. Jastrebova, 2003. Folate determination in Swedish berries by radioprotein-binding assay (RPBA) and high performance liquid chromatography (HPLC). *Eur. Food Res. Technol*., 216: 264–269

Storozhenko, S., V. De Brouwer, M. Volckaert, O. Navarrete, D. Blancquaert, G.F. Zang, W. Lambert and D. Van der Straeten, 2007a. Folate fortification of rice by metabolic engineering. *Nat. Biotech*., 25: 1277–1279

Stover, P.J., 2004. Physiology of Folate and Vitamin B12 in Health and Disease. *Nutr Rev*., 62: S3–12

Thompson, J.D.I., D.G. Higgins and T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., 22: 4673–4680

Virendra, S., 2006. Seabuckthorn (*Hippophae* L.) A Multipurpose Wonder Plant Vol. II: Biochemistry and Pharmacology, ISBN:9788170354154 8170354153

Xurong, T., M. Peter and A. Tigerstedt, 2001. Variation of physical and chemical characters within and elite Seabuckthorn (*Hyppophae rhamnoides* L.) breeding population. *Sci. Hortic.*, 88: 203–214

(Received 11 May 2015; Accepted 28 September 2015)