Indian Journal of Agricultural Sciences **84** (8): 920–4, August 2014/Article

Diversity in KCS2 (Ketoacyl-CoA Synthase) of selected plants and its molecular implications: A computational analysis

MONENDRA GROVER[1]

National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012

Received: 16 February 2013; Revised accepted: 30 June 2014

ABSTRACT

The majority of calories in human food are derived from plant fatty acids. Besides,plant fatty acids are also a major component of a variety of products useful to human beings such as paints, cosmetics, biofuels, lubricants, detergents and soaps. Ketoacyl-CoA synthase is a key enzyme involved in the fatty acid elongation in plants In this study, we have analyzed the diversity in the KCS2 proteins of a selected plant species. We conclude that though there are extensive similarities in the KCS2 proteins studied with respect to total number of negatively charged residues, total number of positively charged residues, and domain organization, there are notable differences for other features such as extinction coefficients, protein stability, kinase specific phosphorylation sites, number of O-GlcNAc sites, predicted sumoylation sites, residues contributing to nuclear export signal and transmemebrane helices. These differences may have repercussions for the quantitative efficiency of the 3-Ketoacyl-CoA synthase enzyme which catalyzes the condensation of c2 units to acyl coA during the fatty acid elongation process, and its regulation. This paper showcases molecular implications of diversity in KCS2 , which can be used to create a diverse genetic base for engineering KCS 2 genes.

Key words : Acyl-CoA, Fatty acids, KCS2, Proteins

The majority of human caloric intake is derived from plant fatty acids (Broun *et al.* 1999). In addition to this the plant fatty acids are also a major component of paints, cosmetics, biofuels, lubricants, detergents and soaps (Ohlrogge 1994). There are two major reasons for mainstream attention to genetic engineering of vegetable oils. First, storage form of fatty acids is a good target for genetic engineering and efforts in this direction are unlikely to have repercussions on the physiology of the plant. Secondly, a vast array of different fatty acid structures with potentially useful properties occur in plants (Voelker and Kinny 2001). Besides, the plant oils can provide viable bio based alternatives to petroleum in the long run.

It is important to understand and analyze genetic diversity which is available to human beings, since modern agricultural practices have led to the erosion of genetic base in the crop plants. In this study, we have analyzed the genetic diversity of 3-Ketoacyl-CoA synthase (encoded by *KCS2*) from the selected plants in respect to various structural and psysico-chemical parameter and its molecular implications. This will be important in creating a library of KCS2 genes (with known characteristics of the proteins coded by them) which may be used for genetic transformation and thus increase the genetic variability of KCS2 genes.

[1]e mail: monendra_grover@yahoo.com, Present Address: Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Pusa Campus, New Delhi 110 012

MATERIALS AND METHODS

The Arabidopsis KCS2 protein (NM_100303) (AT1G04220) was retrieved from NCBI. This protein was subjected to sequence homology analysis using BLASTP against the nr-Protein database. Some of the hits were selected for further analysis (Table 1). The tools used in this study have been listed in Table 2.

Table 1 KCS2 proteins included in this study

Protein id	Plant
ABV60087.1	*Gossypium hirsutum*
XP_002283935.1	*Vitis vinifera*
EAY99126.1	*Oryza sativa Indica*
XP_002437234.1	*Sorghum bicolour*
ACG36525.1	*Zea mays*
AAT65207.1	*Brassica napus*

RESULTS AND DISCUSSION

Fatty acids with chains of twenty or more carbons are termed as VLCFAs (very long chain fatty acids). VLCFAs are precursors of various lipids in plants such as sphingolipids and the phospholipids in the membranes (Schneiter *et al*. 1996, Devaiah *et al.* 2006, Dickson *et al.* 2006), triacylglycerols accumulating in the seeds (Stefansson

Table 2 Tools used in this study

Analysis	Tool, Web address
Domain analysis	Interproscan (http://www.ebi.ac.uk/Tools/InterProScan/)
Prediction of O-GlcNAc sites	Yin-O-Yang, http://www.cbs.dtu.dk/services/YinOYang/
Analysis of physico-chemical properties	Protparam, http://expasy.org/tools/protparam.html
Prediction of Phosphorylation sites	NetPhos, http://www.cbs.dtu.dk/services/NetPhos/
Prediction of kinase specific phosphorylation sites	NetPhosK, http://www.cbs.dtu.dk/services/NetPhosK/
Chloroplast targeting	ChloroP, http://www.cbs.dtu.dk/services/ChloroP/
Mitochondrial targeting	Mito P, http://ihg.gsf.de/ihg/mitoprot.html
Peroxisomal targeting	PTS1 predictor, http://mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp

et al. 1961; Lassner *et al.* 1996; Barret *et al.* 1998), the aliphatic suberin embedded in the cell walls of the root endodermis and the periderm of shoots and roots (Kolattukudy 2001, Bernards 2002, Franke and Schreiber 2007, Pollard *et al.* 2008). At the plant level VLCFA derivatives act as signaling molecule in the membranes and also provide energy storage in the seeds. Complex mixtures of VLCFAs and their derivatives such as ketones, wax esters, primary alcohols, secondary alcohols, aldehydes and alkanes constitute the cuticular waxes (Post-Beittenmiller 1996, Kunst and Samuels 2003, Samuels *et al.* 2008*)*. The cuticular waxes ensure protection from UV light (Reicosky and Hanover 1978) and provide safety against lipophilic pathogenic spores and dust (Barthlott and Neinhuis 1997*)*. The fatty acid elongation in plants proceeds through a series of four reactions: condensation, reduction, dehydration followed by reduction again. In the first condensation step, the C2 carbon moiety is condensed to acyl coA 3-Ketoacyl-CoA synthase, the 3-Ketoacyl coA is then reduced by the action of 3-ketoacyl-CoA reductase (KCR), which in turn is followed by dehydration of 3-hydroxyacyl- CoA by 3-hydroxyacyl CoA dehydratase and finally there is reduction of trans-2,3-enoyl CoA by trans-2-enoyl CoA reductase. The VLCFAs generated during the above process are then transformed into aldehydes ketones, alkanes and secondary alcohols by decarbonylation pathway or into primary alcohols and wax esters by primary reduction pathway. (Kunst and Samuels 2003, Samuels *et al*. 2008). During the fatty acid elongation process, the first step is condensation of C2 units to acyl CoA by KCS. In the Arabidopsis genome a total of 22 *KCS* genes have been identified (Beisson *et al.* 2003). When Arabidopsis *KCS* genes were expressed in yeast, KCS2/DAISY catalyzed the two-carbon elongation of the docosanoic acid (C20 fatty acid) (Franke *et al.* 2009). Studies by Lee *et al.* (2009) indicate that KCS20 and KCS2/

Table 3 Physico-chemical properties of the KCS2 proteins used in this study

Protein id	Plant	MW	Theoretical pI	Total number of negatively charged residues	Total number of positively charged residues	Extinction coefficients	Instability index
NM_100303	*Arabidopsis thaliana*	518	9.39	44	62	58955	39.51 stable
ABV60087.1	*Gossypium hirsutum*	59707.5	9.39	48	68	68925	36.56 stable
XP_002283935.1	*Vitis vinifera*	58751.8	9.17	48	63	66070	35.71 stable
EAY99126.1	*Oryza sativa Indica Group*	58061.2	9.28	45	61	67435	37.91 This classifies the protein as stable
XP_002437234.1	*Sorghum bicolor*	57952.42.4	9.15	50	64	64580	44.64 unstable
ACG36525.1	*Zea mays*	57914.34.3	9.09	51	64	64580	44.52 unstable

Table 4 Analysis of the secretory/non-secretory nature of the proteins studied

Accession number	Plant	Nature
NM_100303	*Arabidopsis thaliana*	Non secretory
ABV60087.1	*Gossypium hirsutum*	Non secretory
XP_002283935.1	*Vitis vinifera*	Non secretory
EAY99126.1	*Oryza sativa indica*	Non secretory
XP_002437234.1	*Sorghum bicolor*	Signal anchor
ACG36525.1	*Zea mays*	Signal anchor
AAT65207.1	*Brassica napus*	Signal anchor

Table 5 O-ß-GlcNAc attachment sites in the KCS2proteins

Protein id	Plant	No. of sites
NM_100303	*Arabidopsis thaliana*	14
ABV60087.1	*Gossypium hirsutum*	9
XP_002283935.1	*Vitis vinifera*	7
EAY99126.1	*Oryza sativa Indica Group*	10
XP_002437234.1	*Sorghum bicolor*	8
ACG36525.1	*Zea mays*	10
AAT65207.1	*Brassica napus*	7

Table 6 Phosphorylation sites in the KCS2 proteins

Protein id	Plant	Phosphorylation sites
NM_100303	*Arabidopsis thaliana*	Ser: 17 Thr: 6 Tyr: 2
ABV60087.1	*Gossypium hirsutum*	Ser: 11 Thr: 6 Tyr: 2
XP_002283935.1	*Vitis vinifera*	Ser: 6 Thr: 6 Tyr: 2
EAY99126.1	*Oryza sativa indica*	Ser: 9 Thr: 7 Tyr: 3
XP_002437234.1	*Sorghum bicolor*	Ser: 9 Thr: 4 Tyr: 4
ACG36525.1	*Zea mays*	Ser: 11 Thr: 2 Tyr: 5
AAT65207.1	*Brassica napus*	Ser: 16 Thr: 8 Tyr: 6

Table 7 Kinase specific phosphorylation sites in KCS2 proteins

Protein id	Plant	Kinases for which sites are present
NM_100303	*Arabidopsis thaliana*	PKC PKA cdc2 CKI DNAPK ATM CKII PKG SRC p38MAPK GSK3 cdk5 (12)
ABV60087.1	*Gossypium hirsutum*	CKII cdc2 PKA PKC p38MAPK DNAPK RSK SRC GSK3
XP_002283935.1	*Vitis vinifera*	CKI PKA PKC PKG CKII cdc2 DNAPK RSK SRC EGFR (10)
EAY99126.1	*Oryza sativa Indica*	PKC PKA CKI cdc2 RSK PKG CKII SRC DNAPK p38MAPK DNAPK INSR (12)
XP_002437234.1	*Sorghum bicolor*	cdc2 PKC PKA p38MAPK PKG CKII DNAPK INSR PKG CKII ATM GSK3 (12)
ACG36525.1	*Zea mays*	cdc2 PKC PKA p38MAPK CKII DNAPK INSR RSK PKG ATM CKI SRC EGFR GSK3 (14)
AAT65207.1	*Brassica napus*	PKA cdc2 PKC CKII CKI RSK p38MAPK GSK3 cdk5 DNAPK INSR PKG (12)

DAISY are functionally redundant in the two-carbon elongation to C22 VLCFA that is required for root suberin and cuticular wax biosynthesis. However, their expression is differentially controlled under osmotic stress conditions.

In this study we computationally analyze the properties of the proteins ecoded by *KCS 2* gene in Arabidopsis and their homologues in selected plants. Protparam was used to analyze the the physico-chemical properties of the KCS2 proteins. It was found that there is slight variation in the number of negatively charged amino acids (48–51). the

Table 8 Sumoylation sites in the KCS2 proteins

Protein id	Plant	Number of sumoylation sites
NM_100303	*Arabidopsis thaliana*	0
ABV60087.1	*Gossypium hirsutum*	2
XP_002283935.1	*Vitis vinifera*	0
EAY99126.1	*Oryza sativa Indica*	1
XP_002437234.1	*Sorghum bicolor*	0
ACG36525.1	*Zea mays*	0
AAT65207.1	*Brassica napus*	0

Table 9 NES (Nuclear export signals) in KCS2 proteins

NM_100303	*Arabidopsis thaliana*	Predicted NES signals: 1
ABV60087.1	*Gossypium hirsutum*	10
XP_002283935.1	*Vitis vinifera*	10
EAY99126.1	*Oryza sativa Indica*	10
XP_002437234.1	*Sorghum bicolor*	10
ACG36525.1	*Zea mays*	10
AAT65207.1	*Brassica napus*	2

positively charged residues (62–68) in plants studied. The extinction coefficient of all the proteins varied from 58955 to 68925. The sorghum and maize proteins were found to be unstable in contrast to other proteins studied. The theoretical pI of the proteins studied was found to be broadly similar (Table 3).

The search for domains in the proteins using interproscan revealed that domain organization of all the proteins studied is broadly similar with domains IPR012392, IPR013601, IPR016038, IPR013747, IPR016039 present in all the proteins. The results also show that the proteins are not targeted to mitochondria, chloroplast or peroxisomes. The Signal P analysis showed that none of the proteins are secretory in nature (Table 4). However, a signal anchor (uncleaved signal peptide) is present in *Sorghum bicolor*, *Zea mays* and *Brassica napus*. These observations are consistent with our results which show that these proteins are membrane proteins (Table 9).

O-GlcNAc sites in the proteins were predicted using YinOYang. Predictions for O-ß-GlcNAc attachment sites in the studied proteins are shown in Table 5. The addition of a carbohydrate moeity to the side chain of a residue in a protein chain results in modification of the physico-chemical properties of the protein. Glycosylation is known to result in changes in protein solubility, stability, proteolytic resistance, and local structure (Lis and Sharon 1993, Hounsell *et al*. 1996). The number of potential O-glcNAc attachment sites in all the proteins studied is highly variable.

Protein phosphorylation is an important post-translational modification of proteins in which a threonine tyrosine or serine residue is phosphorylated by a protein kinase by the addition of a phosphate group. Protein phosphorylation at serine, threonine or tyrosine residues affects a multitude of cellular signaling processes (Blom *et al.* 1999). Net Phos analysis (Table 5) shows that all the studied proteins are potentially phosphorylated. However the sites of phosphorylation, however, vary among the studied proteins (Table 7).

Table 10 Trans-membrane helices in KCS2 proteins

Protein id	Plant	Predicted TMH	Number of amino acids in the TMHs
NM_100303	*Arabidopsis thaliana*	3	66
ABV60087.1	*Gossypium hirsutum*	3	73
XP_002283935.1	*Vitis vinifera*	2	62
EAY99126.1	*Oryza sativa indica*	2	75
XP_002437234.1	*Sorghum bicolor*	2	54
ACG36525.1	*Zea mays*	2	58
AAT65207.1	*Brassica napus*	3	70

NetphosK was used to predict kinase specific phosphorylation sites (Table 8). All the plants studied showed potential presence of kinase specific phosphorylation sites. The number of sites was largely similar in the plants studied.

The prediction of Sumoylation sites was done in the selected proteins Small Ubiquitin-like Modifier or SUMO proteins are a family of small proteins that are detached from and attached to other proteins to modify their function. SUMOylation could modify the sub-cellular localization, activity or stability etc of proteins (Fernandez-Loris *et al.* 2006). Protein sumoylation can play an important roles in a variety of biological processes, such as, cell cycle progression, transcriptional regulation, signaling transduction and differentiation etc. Only the rice and cotton proteins showed the presence of sumoylation sites (Table 8).

The studies on the presence of nuclear export signals show that all the studied proteins have nuclear export signal. All the proteins except the *B. napus* protein have 10 NESs. The *B. napus* protein has 2 NESs (Table 10).

The studies on the transmembrane helices show that all these proteins are membrane proteins. All the proteins show the presence of transmembrane helices, with a significant number of amino acids present in the helices (Table 11).

In the present study, we have analyzed the diversity in the KCS2 protein involved in the fatty acid elongation pathway and its molecular implications. Fatty acids are crucial in many human activities. Our study showed that there are a few similarities in the biochemical and biophysical properties of the proteins encoded by the KCS2 genes, there are significant differences. These differences could be related with the fatty acid quantity and quality among plant species and the potential KCS2 genes may be used for genetic genetic engineering of plants for improving oil quality.

REFERENCES

Broun P, Gettner S and Somerville C. 1999. Genetic engineering of plant lipids. *Annual Review Nutrition.* **19**: 197–216.

Ohlrogge J B. 1994. Design of new plant products: Engineering of fatty acid metabolism. *Plant Physiology* **104** : 821–6.

Voelker T and Kinney A. 2001. Variations in the biosynthesis of seed storage lipids. *Annual Review Plant Physiology and Plant Molecular Biology* **52**: 335–361

Schneiter R, Hitomi, M, Ivessa A S, Fasch E V, Kohlwein S P and Tartakoff M. 1996. A yeast acetyl coenzyme A carboxylase mutant links very-long-chain fatty acid synthesis to the structure and function of the nuclear membrane-pore complex. *Molecular Cell Biology* **16**: 7 161–72.

Devaiah S P, Roth M R, Baughman E, Li M, Tamura P, Jeannotte R, Welti R. and Wang X. 2006. Quantitative profiling of polar glycerolipid species from organs of wild-type Arabidopsis and a phospholipase D 1 knockout mutant. *Phytochemistry* **67**: 1 907–24.

Dickson R C, Sumanasekera C and Lester R L. 2006. Functions and metabolism of sphingolipids in *Saccharomyces cerevisiae. Prog. Lipid Research* **45**: 447–65.

Stefansson B R, Hougen E W and Downey R K. 1961. Note on the isolation of rape plants with seed oil free from erucic acid. *Canadian Journal of Plant Science* **41**: 218–9.

Lassner M W. Lardizabal K and Metz J G. 1996. A jojoba β-ketoacyl-CoA synthase cDNA complements the canola fatty acid elongation mutation in transgenic plants. *Plant Cell* **8**: 281–92.

Barret P, Delourne R, Renard M, Domergue F, Lessire R, Delseny M and Roscoe T J. 1998. A rapeseed FAE1 gene is linked to the E1 locus associated with variation in the content of erucic acid. *Theoretical Applied Genetics*. **96**: 177–86.

Kolattukudy P E. 2001. Polyesters in higher plants. *Advances Biochemical Engineering Biotechnology*. **71**: 1–49.

Bernards M A. 2002. Demystifying suberin. *Canadian Journal of Botany* **80**: 227–40.

Franke R and Schreiber L. 2007. Suberin – a biopolyester forming apoplastic plant interfaces. *Current Opinion in Plant Biology* **10**: 252–9.

Pollard M, Beisson F, Li Y and Ohlrogge J B. 2008. Building lipid barriers: biosynthesis of cutin and suberin. *Trends in Plant Science* **13**: 236–46.

Post-Beittenmiller D. 1996. Biochemistry and molecular biology of wax production in plants. *Annual Review Plant Physiology and Plant Molecular Biology*. **47**: 405–30.

Kunst L and Samuels A L. 2003. Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Research*. **42**: 51–80.

Samuels L, Kunst L and Jetter R. 2008. Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annual Review of Plant Biology*. **59**: 683–707.

Reicosky D A and Hanover J W. 1978. Physiological effects of surface waxes. I. Light reflectance for glaucous and non-glaucous Picea pungens. *Plant Physiology* **62**: 101–4.

Barthlott W and Neinhuis C. 1997. Purity of the sacred lotus, or escape from contamination in biological surfaces. *Planta* **202**: 1–8.

Kunst L and Samuels A L. 2003. Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Research* **42**: 51–80.

Beisson F, Koo A J K, Ruuska S *et al.* 2003. Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiology* **132**: 681–97.

Franke R, Höfer R, Briesen I, Emsermann M, Efremova N, Yephremov A and Schreiber L. 2009. The DAISY gene from Arabidopsis encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza–micropyle region of seeds. *Plant Journal* **57**: 80–95.

Lee S B, Jung S J, Go Y S, Kim H U, Kim J K, Cho H J, Park O K and Suh M C. 2009. Two Arabidopsis 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2*/*DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *Plant Journal* **60**: 462–75.

Lis H and Sharon N. 1993. Protein glycosylation: Structural and functional aspects. C*urrent Journal of Biochemistry* **218**: 1–27.

Hounsell E F, Davies M J and Renouf D V. 1996. Olinked protein glycosylation structure and function. *Glycoconjugate Journal* **13**: 19–26.

Blom N, Gammeltoft S and Brunak S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* **294**(5): 1 351–62.

Fernández-Lloris R, Osses N, Jaffray E, Shen LN, Vaughan OA, Girwood D, Bartrons R, Rosa J L, Hay R T, Ventura F. 2006. Repression of SOX6 transcriptional activity by SUMO modification. *FEBS Letters* **580**(5): 1 215–21.