Functional characterization and structural modeling of *Helianthus annuus* (sunflower) ketoacyl-CoA synthases and role in seed oil composition

González-Mellado, Damián; Salas, Joaquín J.; Venegas-Calerón, Mónica; Moreno-Pérez, Antonio J.; García, Rafael and Martínez-Force, Enrique

1Instituto de la Grasa (CSIC), Edificio 46, Campus Universitario Pablo de Olavide, Ctra. de Utrera Km 1, 41013-Seville, Spain.

2Departamento de Genética, Facultad de Biología, Universidad de Sevilla, 41012 Seville, Spain.

Running title: Sunflower ketoacyl-CoA synthases

Corresponding author: Joaquín J. Salas, Instituto de la Grasa (CSIC), Edificio 46, Campus Universitario Pablo de Olavide, Ctra. de Utrera Km 1, 41013-Seville, Spain.

Email: jisalas@ig.csic.es, Tlf: +34 954611550, Fax: +34 954616790
Abstract

Main conclusion: The enzymes HaKCS1 and HaKCS2 are expressed in sunflower seeds and contribute to elongation of C18 fatty acids, resulting in the C20-C24 fatty acids in sunflower oil.

Most plant fatty acids are produced by plastidial soluble fatty acid synthases that produce fatty acids of up to 18 carbon atoms. However, further acyl chain elongations can take place in the endoplasmic reticulum, catalysed by membrane-bound synthases that act on acyl-CoAs. The condensing enzymes of these complexes are the ketoacyl-CoA synthase (KCSs), responsible for the synthesis of very long chain fatty acids (VLCFAs) and their derivatives in plants, these including waxes and cuticle hydrocarbons, as well as fatty aldehydes. Sunflower seeds accumulate oil that contains around 2-3% of VLCFAs and studies of the fatty acid elongase activity in developing sunflower embryos indicate that two different KCS isoforms drive the synthesis of these fatty acids. Here, two cDNAs encoding distinct KCSs were amplified from RNAs extracted from developing sunflower embryos and named HaKCS1 and HaKCS2. These genes are expressed in developing seeds during the period of oil accumulation and they are clear candidates to condition sunflower oil synthesis. These two KCS cDNAs complement a yeast elongase null mutant and when expressed in yeast, they alter the host’s fatty acid profile, proving the encoded KCSs are functional. The structure of these enzymes was modelled and their contribution to the presence of VLCFAs in sunflower oil is discussed based on the results obtained.

Keywords: Helianthus annuus, very long chain fatty acids, fatty acid elongase, ketoacyl-CoA synthase, sunflower seeds.
Introduction

Plant fatty acids are synthesized by successive elongation of acyl-ACP derivatives in plastids, events catalysed by the fatty acid synthase (FAS) soluble type 2 multienzyme system that brings together ketoacyl-ACP synthase (KAS), ketoacyl-ACP reductase, hydroxyacyl-ACP dehydratase and enoyl reductase activities. Several condensing enzymes have been described and characterized in the soluble system, differing in the length of the substrate they condense. Thus, KAS III mainly catalyses an initial reaction that uses malonyl-ACP and acetyl-ACP as substrates, whereas condensations from C4 to C16 are undertaken by the KAS I enzyme and KAS II catalyses the elongation of palmitoyl-ACP to stearoyl-ACP. The final products of plastidial fatty acid synthase are C16 and C18 acyl chains, the precursor of most fatty acids present in plant glycerolipids (Ohlrogge and Jaworski 1997). Nevertheless, fatty acids longer than 18C are often present in plant tissues and these fatty acids are synthesized by elongation of different acyl-CoAs in the endoplasmic reticulum (ER) by a membrane bound enzyme complex with characteristics similar to FAS, known as fatty acid elongase or FAE (Leonard et al. 2004; Haslam and Kunst 2013). FAE elongates acyl-CoA derivatives rather than plastidial acyl-ACP derivatives, and this involves the same reactions of condensation, dehydration and reduction, malonyl-CoA representing the substrate that provides the carbon for the chain elongations. Moreover, the condensing activities within this complex are the ketoacyl-CoA synthases (KCSs). In this process, the condensation reaction is widely acknowledged to be the rate limiting and more determinant step (Millar and Kunst 1997).

Biochemically, the active site and reaction mechanism of plant KAS and KCS enzymes is identical, and quite similar to that of enzymes like plant chalcone synthase or the KAS forms in E. coli (Jiang et al. 2008) which have been crystallized and characterized structurally (Huang et al. 1998). Structural and site directed mutagenesis demonstrated that a catalytic triad of residues is responsible for catalysis, formed by a cysteine residue, one histidine and one His/Asn. The acyl moiety is transferred to the
Cys residue of this triad, permitting the ensuing nucleophilic attack of the malonyl-CoA substrate to form the ketoacyl product. The main structural difference evident between KCS and KAS is the presence of two domains in the N-terminal region that anchor the enzyme to the ER membranes.

Arabidopsis plants contain 21 copies of different KCS genes (Beisson et al. 2003), among which AtKCS1 has been functionally characterized (Todd et al. 1999) and is known to play an important role in wax synthesis. As such, knocking out the KCS1 gene produces an 80% decrease in C26 to C30 wax alcohols and aldehydes, having a weaker impact on hydrocarbons and ketones. However, since the total amount of waxes remains unaltered, this gene would appear to be redundant to some extent in Arabidopsis, yet knockout plants have thinner stems and they are susceptible to desiccation. The involvement of KCSs in the production of surface lipids and their protective functions have been confirmed in Arabidopsis and other plant species (Vogg et al. 2004; Lee et al. 2009). In addition, KCSs also play a role in determining the fatty acid composition of the seed oil in some species, for example regulating the large amounts of erucic acid (22:1) that accumulates in Brassica napus oil. In this regard, BnFAE1.1 and BnFAE1.2 genes (Barret et al. 1998) encoding KCS proteins were shown to be responsible, around 90.6% of the trait (Jourden et al. 1996), for the presence of erucic acid in rapeseed oil, and mutations in the active site of these enzymes produce a plant with a low erucic phenotype (Han et al. 2001; Roscoe et al. 2001). Similar KCS isoforms have been defined in other Brassicaceae species like Arabidopsis or Camelina and these enzymes elongate oleoyl-CoA or linoleoyl-CoA, yielding the unsaturated very long chain fatty acids (VLCFAs) present in these species. This situation contrasts with that reported in sunflower, as sunflower oil lacks unsaturated VLCFAs and only contains small amounts of saturated ones: 0.5-1% of arachidic acid, 20:0; 1-2% behenic acid, 22:0; and trace amounts of lignoceric acid, 24:0 (Salas et al. 2005). However, the accumulation of these fatty acids increases in sunflower mutants deficient in stearate desaturase, with a higher content of stearic acid and VLCFAs (Fernández-Moya et al. 2003).
Biochemical studies on the KCSs present in microsomes isolated from developing sunflower embryos pointed to the presence of two main KCS isoforms with different substrate preferences (Salas et al. 2005): KCS-I displayed stronger activity towards arachidoyl-CoA, whereas KCS-II was more active in elongating palmitoyl-CoA and stearoyl-CoA. Neither displayed significant activity towards unsaturated acyl-CoAs. The characterization of the genes encoding these KCS is of interest from the point of view of plant biochemistry, as it could provide information about the structural features that determine the substrate specificity of these enzymes, as well as allowing us to better understand the synthesis of VLCFAs and their derivatives: waxes, cutin and suberin.

In the present work, we identified two KCS genes expressed strongly in developing sunflower embryos, named \textit{HaKCS1} and \textit{HaKCS2}. The proteins encoded by these genes displayed similarities to \textit{AtKCS2}, \textit{AtKCS20} and \textit{AtKCS11}, and their expression profiles fit well with a role in the synthesis of fatty acids present in sunflower oil. These genes were expressed in yeast, altering the fatty acid composition of the transformed cells, which displayed important increases in saturated VLCFAs. The contribution of these enzymes to the quality of sunflower oil is discussed in light of the results.

\textbf{Materials and Methods}

\textbf{Biological material}

The wild-type CAS-6 Sunflower line \textit{(Helianthus annuus} L.: Sunflower Collection of Instituto de la Grasa, CSIC, Sevilla, Spain) was grown as described elsewhere (Álvarez-Ortega et al. 1997). Plants were cultivated in growth chambers at 25/15 °C (day/night cycles), with a 16 h photoperiod and a photon flux density of 200 μmol m⁻²s⁻¹. Seeds used for the synthesis of cDNA were collected 15 days after flowering (DAF), whereas to study quantitative expression by RT-QPCR, seed embryo samples were collected every other day from 12 to 22 DAF. Samples of vegetative tissues (stem, leaf,
root and seedling cotyledons) were collected from plants 20 days after germination. All
the samples were frozen and stored at -80 °C.

The Escherichia coli strain XL1-Blue (Stratagene, La Jolla, CA, USA) was used
as the plasmid host for HaKCS1 and HaKCS2 cloning. The bacteria were grown at 37
°C with shaking in LB medium (1% bacto tryptone, 0.5% bacto yeast extract, 1% NaCl
[pH 7]). Plasmid selection was based on ampicillin resistance (50 μg mL⁻¹).

The Saccharomyces cerevisiae strain W301A (MATα ura3-52 leu2-3, 112 trp1-
289a his3-1) was used for heterologous gene expression and the TDY7005 elo2Δ-
elo3Δ/pELO3 strain was used for complementation assays (MATα lys2 ura3-52 trp1 leu2
elo2ΔTRP elo3ΔTRP/pADHura-ELO3: kindly provided by Fred Beaudoin, Rothamsted
Research, UK).

Cloning of the sunflower HaKCS1 and HaKCS2 genes

Approximately 0.1 g of sunflower developing seed embryos at 18 DAF was
ground in liquid nitrogen in a precooled sterile mortar and pestle. Total RNA was
extracted using a Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO, USA) and
mRNA was isolated from this total RNA using the GenElute mRNA Miniprep kit (Sigma-
Aldrich). The mRNA pellet was resuspended in 33 μL of RNAase free TE buffer (10 mM
Tris–HCl, 1 mM EDTA [pH 8]) and the corresponding cDNA was synthesized using a
Ready-To-Go T-Primed First Strand Kit (GE Healthcare Life Science, Buckinghamshire,
UK).

Plant KCS protein sequences from public databases were aligned to identify
regions of homology using the ClustalX 2.0 program (Thompson et al. 1997: Figure S1).
From that data, three degenerate primers were designed to clone sunflower KCS genes
expressed during seed development: KCSEST-F1 (48 times degenerate), KCSEST-F2
(four times degenerate) and KCSEST-R1 (eight times degenerate: Table S1). A PCR
fragment corresponding to the 3’-end of HaKCS1 (374 bp long) was amplified using the
oligonucleotide KCSEST-F2 in combination with the primer FA2Z (Table S1), complementary to the sequence incorporated during the initial cDNA synthesis. The 5´-end was amplified using the Smart™-RACE cDNA amplification kit (BD Bioscience Clontech Company, Palo Alto, CA, USA) and the specific primer KCS1B-R4 (Table S1). Similarly, an internal PCR fragment of HaKCS2 (703 bp long) was amplified using degenerate primers KCSEST-F1 and KCSEST-R1 (Table S1). The full sequence was obtained at the 3´-end using the primer FA2Z and the specific primer KCS2A-F1 (Table S1), while the 5´-end was amplified using the Smart™-RACE cDNA amplification kit and the specific primer KCS2A-R2 (Table S1).

Primers were synthesized by MWG Biotech AG (Ebersberg, Germany) and all the PCR fragments were cloned into the pGEM-T-Easy® vector (Promega, Madison, WI, USA), transformed into XL1-Blue, sequenced by Secugen SL (Madrid, Spain), and assembled to obtain DNA sequences of about 1720 bp and 1572 bp. Once their identities were confirmed using the Blast software (Altschul et al. 1990), the complete cDNA sequences of the sunflower ketoacyl-CoA synthases KCS1 and KCS2 were deposited in GenBank with the accession numbers EU442581 and EU496864, respectively.

Modelling of the HaKCS three-dimensional structures

Homology modelling of the putative HaKCS1 and HaKCS2 protein structures was carried out with the Deepview and Swiss Model Workspace software (Guex and Peitsch, 1997; http://www.expasy.org/spdbv/), using their protein sequences and the chalcone synthase crystal structure from Medicago sativa as a template (Protein Data Bank accession 1i86: Jez et al. 2001). The HaKCS1 (85-467) and HaKCS2 (104-485) residues were modelled against this template, with sequence identities of 19.6% and 20.0%, respectively. Molecular docking experiments were performed using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), and with malonyl-CoA and arachidic acid molecules as substrates (Yang et al. 2015. Critical residue mapping and visualization was performed using the UCSF Chimera package (Pettersen et al. 2004),
and the transmembrane tendency hydrophobicity profiles of these sunflower KCS enzymes were determined using the ProtScale software (Gasteiger et al. 2005).

**Bioinformatics analysis**

Sequences homologous to the predicted protein sequences of sunflower KCSs were retrieved using the BLASTP program ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and an alignment of the amino acid sequences for the KCS proteins deposited at GENBANK was performed with the Clustal X v.2.0 program using the default settings (Thompson et al. 1997). These alignments were used to generate a phylogenetic tree based on the neighbor-joining algorithm (Saitou and Nei 1987), and the resulting phenogram was drawn using the MEGA 4.0 program (Tamura et al. 2007). Other programs used to identify putative transmembrane anchorage regions or the transmembrane tendency were TMHMM (Krogh et al. 2001), OCTOPUS (Viklund and Elofsson 2008) and ProtScale (Gasteiger et al. 2005).

**RT-QPCR studies of gene expression**

The cDNAs from different sunflower tissues (leaf, stem, roots, developing seed embryos and cotyledons) were subjected to real-time quantitative PCR (RT-QPCR) on a CFX96™ Real-Time PCR Detection System (Bio-Rad, CA, USA) using SYBR Green I (SsoAdvanced™ SYBR® Green Supermix, Bio-Rad, CA, USA) and specific primer pairs: HaKCS1_QPCR-F and HaKCS1_QPCR-R for the *HaKCS1* gene; and HaKCS2_QPCR-F and HaKCS2_QPCR-R for the *HaKCS2* gene (Table S1). The PCR products obtained were 108 bp and 142 bp long, respectively. Polymerase activation and DNA denaturation were carried out at 95 °C for 30 s before performing 40 PCR cycles of 95 °C for 15 s and 60 °C for 30 s, during which the resulting fluorescence was monitored. The calibration curves were drawn up using sequential dilutions of the cDNAs and the Livak method was employed to calculate the comparative expression of the samples (Livak and Schmittgen 2001). The sunflower *HaACT1* actin gene (GenBank Accession number FJ487620) was
used as an internal reference to normalize the readings to the relative amount of cDNA in each sample, amplified with the specific HaActin-F1 and HaActin-R1 primers (Table S1).

**Heterologous expression of HaKCS1 and HaKCS2 in yeast**

The HaKCS1 and HaKCS2 genes were expressed heterologously in yeast using the pYES2 and pYES3 expression vectors, which carry the genes required for uracil (URA3) and tryptophan (TRP1) biosynthesis under the control of a GAL1 promoter. HaKCS1 was amplified with the primers KCS1-F pYES2 and KCS1-R pYES2 and HaKCS2 with KCS2-F pYES3 and KCS2-R pYES3 (Table S1). These primers included restriction sites for KpnI and BamHI, which were used to clone these cDNAs into the vectors. Recombinant plasmids were sequenced and introduced into the yeast strain W301A, transforming the yeast by the PLATE method (Becker and Lundblad, 1994).

**S. cerevisiae fatty acid analysis**

The yeast cells were harvested by centrifugation at 1500g for 5 min at 4 °C and they were washed with distilled water before determining the total fatty acid composition using the one-step method proposed previously (Garcés and Mancha 1993). A volume of 3.3 mL methanol/toluene/dimethoxypropane/sulphuric acid (39:20:5:2) and 1.7 mL heptane was added to the cell pellet, and the mixture was heated at 80 °C for 1 h. After cooling, the upper phase containing the fatty acid methyl esters was transferred to a fresh tube, washed with 6.7% sodium sulphate and evaporated to dryness with nitrogen. The methyl esters were dissolved in an appropriate volume of heptane and analysed by GLC (Hewlett-Packard 6890 gas chromatography apparatus; Palo Alto, CA, USA) using a Supelco SP-2380 fused-silica capillary column (30 m length, 0.25 mm i.d., 0.20 μm film thickness: Supelco, Bellefonte, PA, USA). Hydrogen was used as the carrier gas at 28 cm s⁻¹, the temperature of the flame ionization detector and injector was 200 °C, the oven
temperature was 170 °C, and the split ratio was 1:50. Peaks were identified by comparing their retention times with those of the corresponding commercial standards.

Complementation of S. cerevisiae elo double mutants

Complementation studies of the cloned sunflower KCS genes in Saccharomyces cerevisiae were based on the use of the pADH-LEU plasmid, derived from the constitutively expressed, commercial pGAD424 (GenBank number U07647). Cloning of HaKCS2 into the plasmid was carried out using the KpnI and HindIII sites, and HaKCS1 was cloned using KpnI and a blunt end as it has an internal HindIII site. To obtain these constructs the genes were amplified with the primer pairs KCS2-F pYES2/KCS2-R pADH-LEU and KCS1-F pADH/KCS1-R pADH (Table S1).

The Arabidopsis thaliana AtKCS9 gene (At2g16280) was used as the control for the complementation of Saccharomyces TDY7005 elo2Δ elo3Δ/pELO3 double mutant, which is not viable due to the lack of long chain fatty acids (Oh et al. 1997; Paul et al. 2006). Recombinant plasmids were firstly cloned into E. coli XL1Blue, their sequences were then confirmed and used to transform the TDY7005 yeast strain.
Results

HaKCS1 and HaKCS2 genes and phylogenetic tree

Two complete cDNA sequences corresponding to plant elongases were cloned from developing sunflower embryos using degenerate oligomer primers designed from conserved protein regions. These genes were named HaKCS1 and HaKCS2 and their sequences were deposited in GenBank: EU442581 and EU496864, respectively. The predicted HaKCS1 protein was 498 amino acids long, with a molecular weight of 56.26 kDa and a pl of 9.07. The predicted HaKCS2 protein was 514 amino acids long, it was slightly larger at 57.60 kDa and with a pl value of 9.19. There was 77% identity between the two proteins.

The amino acid sequences of these enzymes were compared with homologous proteins from species in different phylogenetic groups, such as Arabidopsis thaliana, Oryza sativa and Physcomitrella patens (Figures 2S and 3S). The resulting alignments highlighted a strong variability in the N-terminal domain and broadly conserved regions around the catalytic residues (KCS1: C223, H390 and N423; KCS2: C241, H408 and N441). A BLAST search produced more homologous proteins, allowing a selection of sequences from different phylogenetic clades to be aligned using Clustal X (including all the forms from Arabidopsis) and used to construct a phylogenetic tree using the MEGA 4.0 program (Thompson et al. 1997: Figure 1). The HaKCS1 protein clustered close to the enzymes from Vitis vinifera and Ricinus communis, lying very close to KCS2 and KCS20 from Arabidopsis thaliana. On the other hand, HaKCS2 displayed strong homology to KCS11 from the Arabidopsis. These three KCS from Arabidopsis (KCS2, KCS11 and KCS20) were included in the KCS subclass ζ (Joubès et al. 2008).

Transmembrane domains and structural model of HaKCS1 and HaKCS2

Both KCS proteins had two predicted transmembrane regions close to the N-terminus that involved residues 23-43 and 62-82 in HaKCS1, and 41-61 and 80-100 in HaKCS2 (Figure 2). The structural models of these enzymes were constructed using the
online server Swiss Model (http://swissmodel.expasy.org/), based on the known
structures of homologous proteins (see the secondary and tertiary structures in Figures
3 and 4). These models did not include the N-terminal ends, corresponding to the
transmembrane regions, and they involved residues 83-467 and 101-485 of HaKCS1
and HaKCS2, respectively. The models predicted these enzymes as homodimers, with
both coupled monomers displaying triangle forms (Figure 4a and b), similar to polyketide
synthases and type III ketoacyl-ACP synthases. These homodimers may be associated
with the presence of a putative leucine zipper motif in C-terminal region analogous to
that present in BnFAE1 (Barret et al, 1998). The body of the two enzymes is made up of
an α-β-α-β-α structure at the N-terminal end and an α–β–α–α structure at the C-terminal
end. These structures embraced the catalytic residues formed by the Cys223, His390
and Asn423 triad in HaKCS1, and Cys241, His408 and Asn441 in HaKCS2, and they
were grouped close to each other to form a hydrophobic cavity across the protein that
allows substrate entry (Figure 5a, b and c). Docking analysis with substrates like malonyl-
CoA or arachidic acid showed that the hydrophobic pocket can accommodate both
substrates at the same time, allowing us to predict the amino acid residues involved in
substrate-enzyme binding (Figure 5 and Table S2).

Analysis of gene expression in leaves and vegetative tissues

The expression of the HaKCS1 and HaKCS2 genes was studied by RT-QPCR in
different plant organs: leaves, root, stem, cotyledons from seedlings and developing
seeds. HaKCS2 was expressed more strongly than HaKCS1 in all the tissues except in
the developing seed. These differences were more significant in roots, stems and
cotyledons, where the increase reached approximately one order of magnitude (Figure
6). In the developing seeds, HaKCS1 expression increased from DAF 12 to 18 and it
reduced thereafter. The expression of HaKCS2 was more stable than that of HaKCS1
during seed development, peaking in seeds harvested at 18 DAF.
Complementation of the S. cerevisiae elo2ΔTrp elo3ΔTrp mutant

The Saccharomyces elo2ΔTrp elo3ΔTrp strain lacks the ELO2 and ELO3 endogenous elongases, which provokes lethality and thus, this mutant was used to confirm that HaKCS1 and HaKCS2 act as functional elongases. The experiment consisted in growing mutants that harbour the HaKCS genes cloned into the pADH-LEU plasmid and pADH-URA::elo3 to complement the mutation. Yeast transformed with pADH-LEU::HaKCS1, pADH-LEU::HaKCS2 and pADH-LEU::AtKCS9 were plated on minimal medium containing 5'-fluoroacetic acid. The yeast containing pADH-URA::elo3 and the empty pADH-LEU plasmid could not grow on this medium, whereas the lines transformed with the plasmids containing the KCS genes were able to grow, demonstrating that the genes identified in this study encoded active elongases (Figure 7).

Heterologous expression of HaKCS genes in S. cerevisiae

The HaKCS1 and HaKCS2 genes were cloned into the pYES2 and pYES3 vectors, and the resulting constructs were used to transform the wild type W301A yeast strain in order to investigate their effect on the yeast’s fatty acid composition (Figure 8). In the case of HaKCS1, cultures displayed an important increase in VLCFAs upon the induction of gene expression. This increase predominantly affected the proportion of 22:0 and 24:0 fatty acids, which increased 3-4 fold, and to a lesser extent to the levels of 26:0 fatty acids, the content of which doubled (Figure 8a). At the same time, the level of stearic and oleic acids in the cultured yeast decreased, with a significant increase in the levels of palmitoleic acid. The expression of HaKCS2 had a weaker impact on the VLCFA content than its homologue, mainly affecting the 22:0 and 24:0 content that doubled in the induced cultures (Figure 8b). There were no significant differences in 26:0 content, while the impact of HaKCS2 induction on the other fatty acids was similar to that observed for HaKCS1, producing a decrease in stearic and oleic acid and an increase in palmitoleic acid.
Discussion

Sequences and phylogenetic studies

\textit{HaKCS} cDNAs have been cloned by PCR amplification from developing sunflower seed cDNA, initially using degenerate primers designed from consensus KCS protein sequences from other species (e.g. \textit{Brasica napus}, \textit{Sorghum bicolor}, \textit{Oryza sativa}, etc.) and subsequently RACE. The sequences obtained corresponded to two similar proteins with an alkaline $p$I, and alignment with KCS proteins from other plant species displayed high levels of conservation of the sequences around the residues involved in the enzyme’s catalytic triad: cysteine, histidine and asparagine (supplemental material Figures S2 and S3). The cysteine residue is the substrate acceptor in the first step of the condensation (Lassner et al. 1996; Ghanevati and Jaworski 2001) and it is conserved in all known KCSs (Joubès et al. 2008), just like the other two residues of the triad (Ghanevati and Jaworski 2001). Interestingly, this is the same catalytic triad found in the FabH enzyme from \textit{E. coli} (Heath and Rock 2002), which suggests bacterial KAS III enzymes may be the phylogenetic ancestor of plant KCSs.

Higher plants usually have several KCS genes, so a large number of sequences were homologous to \textit{HaKCS}s. Thus, a selection of sequences from different phylogenetic clades was necessary to produce an intelligible phylogenetic tree. This selection was based on the availability of sequenced genomes of these species, yet the phylogenetic tree obtained highlighted the divergences between plant KCS protein subclasses, as defined previously (Joubès et al. 2008). The different clades have been associated with different KCS forms from \textit{Arabidopsis thaliana}, and each clade contains homologous sequences from dicotyledonous and monocotyledonous plants, suggesting duplications prior to the evolutionary divergence between these plant groups (Guo et al. 2016). \textit{Arabidopsis thaliana} contains 21 genes encoding KCSs in its genome (Costaglioli et al. 2005). The analysis of these proteins indicates the existence of eight subclasses ($\alpha$, $\beta$, $\gamma$, $\delta$, $\zeta$, $\epsilon$, $\eta$ and $\theta$), arising from an ancestral genome duplication that can be inferred from their chromosomal location (Blanc et al. 2003; Joubès et al. 2008). The \textit{HaKCS}1
and HaKCS2 proteins cluster within a group in the ζ subclass that includes dicotyledonous, monocotyledonous plants and briophyta: HaKCS1 displays 75 and 76% homology with the AtKCS2/DAISY and AtKCS20 proteins from Arabidopsis respectively, which are genes related with the synthesis of surface lipids (Franke et al., 2009; Lee et al., 2009). Moreover, HaKCS2 displays 81% homology with KCS11 from the same plant.

Of all the different forms included in this study, the most similar KCSs to the sunflower proteins found were those from Vitis vinifera and Ricinus communis.

The genes coding for HaKCS1 and HaKCS2 corresponded to HanXRQChr16g0514651 and HanXRQChr13g0402491 locus tags from the sunflower genome recently disclosed (Badouin et al. 2017; https://www.heliagene.org/HanXRQ-SUNRISE/). Compared to their homologs in Arabidopsis (AtKCS2 and AtKCS20/HaKCS1 and AtKCS11/HaKCS2; see Figure 1), the structures of HaKCS1 and HaKCS2 genes are pretty conserved, displaying one and zero introns each one respectively.

**Structural model of sunflower KCSs**

Unlike plastidial condensing enzymes, KCSs are membrane-bound enzymes located in the ER (Joubès et al. 2008), which requires the presence of transmembrane domains in at least some cases. The hydrophobicity profiles of HaKCS1 and HaKCS2 indicated the presence of two high hydrophobic regions close to the N-terminus that would bind the enzyme tightly to the ER membrane. These transmembrane domains were similar to those described in AtKCS1, which also has two hydrophobic regions in the N-terminus (Joubès et al. 2008), the KCS from Tropaeolum majus (Mietkiewska et al. 2004) and that from Lunaria annua (Guo et al. 2009).

While there are currently no crystallographic data for KCS enzymes, the structures of related enzymes have been elucidated, such as bacterial polyketide synthases and KAS III. On the basis of these structures, and more specifically on the C-terminal domain of chalcone synthase of Medicago sativa (PDB 1i86: Jez et al. 2001), it is possible to obtain three-dimensional models of KCSs by applying modelling programs.
like the Swiss Model server. In the structural models for HaKCS1 and HaKCS2 obtained, the regions modelled did not include transmembrane domains or N-terminal ends, embracing residues 85 to 467 for HaKCS1 and 104 to 485 for HaKCS2. These models defined a triangular form in which one of the arms was made up of two α-helices, corresponding to the transmembrane domains that anchor the protein to the ER membrane. The main body of the secondary structure of both proteins displayed two differentiated conformations. On the one hand an α-β-α-β-α structure at the N-terminal end, also seen when modelling thiolase proteins (Mathieu et al. 1994), and on the other, an α–β–α–α structure at C-terminal end in the modelled secondary structure of AtKCS1 (Joubès et al. 2008). Residues involved in the enzyme’s catalytic activity are found in both domains. The catalytic triad for HaKCS1 is probably made up by the amino acids Cys223, His390 and Asn423, whereas in HaKCS2, it corresponds to residues Cys241, His408 and Asn441, all them accessible to the acyl moiety of the substrate that would be set into a cavity that crossed by the protein from one side to the other. These residues are analogous to those previously described for catalytic activity of FabH proteins (Qiu et al. 1999; 2001), supporting a phylogenetic relationship between KCSs and bacterial KASIII from a cyanobacterial ancestor.

Previous studies on Arabidopsis KCSs set out to find a relationship between the active site topology and substrate preference (Joubès et al. 2008). The conformations of the active site of sunflower KCSs give rise to a narrower hydrophobic cavity in HaKCS2, which would produce differences in substrate specificity. As such, the small hydrophobic pocket in the AtKCS18 enzyme is associated with a preference for C20 acyl-CoAs, whereas AtKCS5 and AtKCS6 have wider cavities and a preference towards C26 and C24 substrates, respectively (Blacklock and Jaworski 2006). Hence, there appears to be a relationship between the size of the predicted hydrophobic cavity and the substrate specificity of these enzymes. Furthermore, when domain swaps and site-directed mutagenesis was carried out to identify additional features contributing to KCSs substrate specificity, a 99 amino acid N-terminal region immediately downstream of the
transmembrane helices was seen to be sufficient to change substrate specificity (Blacklock and Jaworski 2002). In the sunflower KCS tertiary structure and docking models, a smaller protein domain that included this previously identified region (Blacklock and Jaworski 2002) corresponds to the bottom of the substrate pocket, and its amino acid sequence varies considerably between both sunflower KCSs, evidence of its implication in substrate length selectivity. Hence, HaKCS1 and HaKCS2 are likely to show different substrate preferences.

Finally, 3D and docking models provide valuable information about the amino acid residues involved in substrate binding. The entrance to the substrate binding site is surrounded with a positively charged region that will interact with the 3'-phosphoadenosine of CoA. Interestingly, the models predict that one residue from the other monomer, Thr196 in HaKCS1 and Thr214 in HaKCS2, protrudes into the active site, suggesting that it will participate in the interaction with the substrate or its catalysis. As such, KCS will be similar to type III ketoacyl-ACP synthases and polyketide synthases, where a single phenylalanine residue from each monomer protrudes into the active site of the other, suggesting that dimer formation plays an important role in catalysis (Haslam and Kunst 2013).

**Tissue specific expression of the HaKCS genes**

Gene expression studies show that both genes, HaKCS1 and HaKCS2, are widely expressed in the examined tissues being HaKCS1 more strongly expressed in developing seeds. KCSs are enzymes necessary to produce the VLCFAs needed to generate the surface lipids that prevent water loss in vegetative tissues and that protect against pest attack: waxes, cutin or suberin. The genes encoding HaKCS1 and HaKCS2 were cloned from seed transcripts and not surprisingly, these enzymes were expressed strongly in seeds, in a profile that paralleled that of oil accumulation but also the enlargement of the seed shell. The expression of HaKCS1 in developing seeds increases continuously until seeds reach 18 DAF, the point of maximal oil accumulation, suggesting...
it play a role in the synthesis of the VLCFAs present in sunflower oil. The expression of these genes resembles that reported previously for *Brassica napus* FAE1 elongase, which is responsible for the synthesis of erucic acid in this species, and that is mainly expressed in seeds but also in vegetative tissues (Chiron et al. 2015). Indeed, the 20:0 and 22:0 fatty acids from sunflower oil were synthesized by enzymes involved in surface lipid production that were not specific for oil synthesis and accumulation. Thus, the ubiquity of the expression of those enzymes, and the high level of homology with the genes AtKCS2/DAISY (Franke et al, 2009) and AtKCS20 (Lee et al., 2009), point to a role of them in the production of surface lipids or polyesters present in cuticle covering the aerial part of the plants, acting as a protective barrier in these organs. This fact seemed to point that genes related with surface lipid synthesis can also contribute to the synthesis of fatty acids present in triacylglycerols (TAGs) accumulated in oilseeds.

**Complementation of the *S. cerevisiae elo2ΔTrp elo3ΔTrp* mutant**

The *S. cerevisiae elo2elo3* mutant lacks the ELO2 and ELO3 elongase enzymes, and it is not viable as it cannot synthesize enough fatty acids for its membranes (Oh et al. 1997). This mutant line was used for complementation studies to confirm that the *HaKCS1* and *HaKCS2* genes encode functional condensing enzymes. The growth and propagation of this mutant strain is usually complemented by the pADH-URA::elo3 plasmid that expresses the yeast ELO3 gene. In these studies, the double mutant containing the pADH-URA::elo3 plasmid was transformed with plasmids pADH-LEU, pADH-LEU::HaKCS1, pADH-LEU::HaKCS2 and pADH-LEU::AtKCS9, expressing the elongase genes from sunflower or the KCS9 from Arabidopsis as a positive control. When the four yeast transformants were plated on minimal medium containing 5´-FOA, they lose the pADH-URA::elo3 and in these conditions, the elo2elo3 mutant is not viable unless it expresses another functional elongase. Both *HaKCS1* and *HaKCS2* complement the elo2elo3 mutation, as did the positive *AtKCS9* control. Hence, the sunflower genes encode a fully functional KCS capable of integration into the yeasts’
biosynthetic machinery in order to synthesise the long chain fatty acids necessary for cell viability, demonstrating the functionality of these genes.

Effects on fatty acid composition in yeast

The activity of the sunflower HaKCS1 and HaKCS2 was evaluated by heterologous expression in yeast, the system of choice to evaluate membrane enzymes expressed in the ER as they undergo correct organ compartmentation. Indeed, this approach has already been used to study the substrate specificity of these enzymes in Arabidopsis thaliana (Blacklock and Jaworski 2006) and Brassica napus (Katavic et al. 2002). The system used for this approach was based on the pYES plasmids and the W301A cell line, driving protein expression through the activity of the galactose operon.

To characterize these enzymes they were expressed separately in the microbial host, and the fatty acid composition of the microorganism was analysed before and after enzyme induction to examine the impact of the sunflower elongases. HaKCS1 expression produced a clear effect on fatty acid production, increasing the VLCFA content of the yeast between 7 to 3.5 fold, from 1-2% (Welch and Burlingame 1973) up to 7.2%. This increase is predominantly due to the accumulation of saturated 22:0, 24:0 and 26:0 fatty acids, in the absence of any additional unsaturated VLCFAs like erucic or gondoic fatty acids. The increase in these fatty acids was paralleled by a decrease in the proportion of stearic and oleic acid in the culture. There was also an increase in palmitoleic acid (16:1), which suggested that the yeast cells try to compensate for the deficit in oleic acid by increasing their desaturase activity, which in this case desaturates palmitic acid to palmitoleic. This effect could be a mechanism to maintain the normal fluidity of the cell membranes. By contrast, the impact of HaKCS2 on lipid composition was less pronounced, mainly affecting 20:0, 22:0 and 24:0 but not 26:0 fatty acids, which remained at the same proportion as in uninduced cells. Again, the effect on unsaturated VLCFAs was limited and the effect on endogenous fatty acids from yeast was similar to that caused by HaKCS1, producing a decrease in stearic and oleic acids, and enhancing
the desaturation of palmitic to palmitoleic acid. Furthermore, these results suggest that
*HaKCS2* has a lower specificity towards longer acyl-CoAs, in good agreement with
structural models given the narrower and shallower hydrophobic pocket to accommodate
the substrate.

Both sunflower KCSs influence the VLCFAs present in sunflower oil, essentially
arachidic and behenic acids with only traces of lignoceric acid. All of them are saturated
VLCFAs, not accompanied by any unsaturated ones, meaning that unlike the enzymes
in other species (e.g. *Brassica napus*, *Camelina sativa* or *Arabidopsis*), sunflower KCSs
are very specific for saturated derivatives. Earlier biochemical studies indicated that
sunflower seed microsomes contain at least two forms of KCS that can be separated by
conventional purification methods (Salas et al. 2005). These forms could elongate
several acyl-CoA substrates, with highest activities towards stearoyl-CoA and
arachidoyl-CoA, whereas no such elongation was observed with oleoyl-CoA as the
substrate. These specificity profiles share similarities with those of the *HaKCS1* and
*HaKCS2* enzymes, which were expressed in seeds and were only active towards
saturated substrates. Important amounts of VLCFAs longer than 22:0 are not detected
in sunflower oil, even though the sunflower KCSs studied here could elongate up to C26
fatty acids. This may be due to discrimination of these fatty acids by the enzymes
responsible for TAG synthesis or because these fatty acids were quickly metabolized in
other pathways active in sunflower seeds, such as those involved in the synthesis of
waxes or other surface lipids (Franke et al., 2009; Lee et al., 2009; Broughton et al.
2018).

**Conclusions**

The synthesis of VLCFAs requires the action of membrane bound elongase
complexes on acyl-CoA derivatives, and the condensing enzymes of these complexes
or KCSs are limiting and determinant. In the present work, two KCS genes were cloned
from developing sunflower seeds, *HaKCS1* and *HaKCS2*, and they correspond to
enzymes with two transmembrane domains that may anchor them to ER membranes. Their structures were modelled, identifying hydrophobic pockets of different sizes that suggest different substrate preferences. The expression profile of these genes differs from that of other KCSs expressed in seeds, like brassica \textit{FAE1}, as they are ubiquitously expressed in different plant organs, especially \textit{HaKCS2}. Nevertheless, their expression in developing seeds increased during the period of oil accumulation. When heterologously expressed in yeast, these genes were able to complement the lethality of the \textit{elo2elo3} mutant that is deficient in fatty acid elongation, proving they encoded active enzymes. Their expression in yeast modified the fatty acid composition of the cells, increasing the concentrations of VLCFAs. \textit{HaKCS1} induced saturated VLCFA accumulation up to C26, whereas \textit{HaKCS2} only affected C20-C24 fatty acids, again suggesting differences in substrate specificity that fitted well with the protein structure. The activity of the proteins was similar to that of KCS from sunflower microsomes previously characterized, such that they could be involved in the synthesis of the VLCFAs present in sunflower oil and contribute to the synthesis of surface lipids present in sunflower embryos.

\textbf{Acknowledgements}

We acknowledge Fred Beaudoin (Rothamsted Research, UK) for kindly providing the \textit{Saccharomyces cerevisiae} double mutant elo2 and elo3 strain.

\textbf{Availability of data and material}

The dataset and genes produced in the present study are available on reasonable request.

\textbf{Ethics approval and consent to participate}
The experimental research in this manuscript was complied with the institutional, national and international guidelines.

Consent for publication

Not applicable

Competing interest

The authors declare they have not competing interests

Funding

This work was funded by the MINECO and FEDER Projects AGL2014-53537-R and AGL2017-83449-R. They are free competition projects funded for the Spanish government for the development of scientific research.

Authors´ contribution

DG-M performed most of the experimental work. AM-P participated in plasmid construction and rtPCR. MVC, EM-F, RG and JJS participated in the work direction and experimental design. JJS and EM-F wrote and revised the manuscript.

References


Figure Legends

Figure 1. Phylogenetic tree of different plant KCS proteins, including sunflower KCS1 and KCS2. The tree was rooted at the FabH protein from Escherichia coli. The groups and species included embraced Dicots (At, Arabidopsis thaliana; Bj, Brassica juncea; Bn, Brassica napus; Ha, Helianthus annuus; Pt, Populus trichocarpa; Rc, Ricinus communis; Vv, Vitis vinifera), Monocots (Zm, Zea mays; Sb, Sorghum bicolor; Os, Oryza sativa; Hv, Hordeum vulgare) and Bryophytes (Mp, Marchantia polymorpha; Mc, Marchantia chenopoda; Pp, Physcomitrella patens). The KCS enzymes from sunflower are in blue and the green dots indicate KCSs from Arabidopsis. The red lines represent groups of
Monocots and the eight distinct KCS subclasses are labelled as described by Joubès et al. (2008).

**Figure 2.** Prediction of transmembrane domains in the HaKCS1 (a) and HaKCS2 (b) proteins, achieved using the ProtScale program (Gasteiger et al. 2005).

**Figure 3.** Comparison of the deduced amino acid sequences and the predicted secondary structures of sunflower KCS1 and KCS2, with the sequence and secondary structure of *Medicago sativa* chalcone synthase (1i86.1A: Jez et al. 2001). The structural α-helices and β-sheets in the N- and C-terminal domains are labelled with * or **, respectively: h, α-helix; s, β-sheet. Residues from the triad involved in substrate catalysis are marked with a black arrowhead and the residues conserved in the three proteins are highlighted in black.

**Figure 4.** Proposed structural models for sunflower HaKCS1 (a, c and e) and HaKCS2 (b, d and f) KCS homodimers based on that of *Medicago sativa* chalcone synthase (1i86.1A: Jez et al. 2001). Panels a and b are ribbon diagrams in which the catalytic triad residues are in green. Panels c, d, e and f, frontal (c and d) and lateral (e and f) views of the molecular surfaces showing the electrostatic potentials according to Coulomb’s laws with a positively charged patch pointing to the entrance of the substrate binding pocket (shown in green).

**Figure 5.** Front (a), lateral (b) and rear (c) views of malonyl-CoA (yellow) and arachidic acid (green) docking in the HaKCS2 substrate binding pocket, with the key residues (triad C, H and N) at the catalytic centre in red. HaKCS2 substrate docking was modelled using the iterative threading assembly refinement (I-TASSER) method (Yang et al. 2015).
Figure 6. Expression of genes encoding HaKCS1 and HaKCS2 in developing seeds and vegetative tissues of the common CAS-6 sunflower line. Expression was determined by RT-QPCR using the actin gene from *H. annuus* (GenBank FJ487620) as the reference gene. The data correspond to the mean ± SD from three independent measurements.

Figure 7. Complementation of the *elo2ΔTrp elo3ΔTrp* Saccharomyces cerevisiae mutant with *HaKCS1* and *HaKCS2* from sunflower, using the *KCS9* (At2g16280) gene from Arabidopsis as a positive control.

Figure 8. Changes in the fatty acid composition of the W301A yeast strain by the induction of *HaKCS1* (a) and *HaKCS2* (b). These genes were expressed from the pYES2 vector and their expression was induced by supplying galactose as the carbon source in the medium. Non-induced cultures were grown using glucose as the carbon source. The data correspond to the mean ± SD of three independent measurements.
Figure 2

(a) Graph showing transmembrane tendency versus position.

(b) Graph showing transmembrane tendency versus position.
<table>
<thead>
<tr>
<th>HaKS1</th>
<th>83</th>
<th>RPKKVLVLPACAYKPNDTSNTVESRETFSQHSLGATGDNDALFQTQLRLSSGQLGQKTYPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>101</td>
<td>RPRVYLVFSCSYPKEREEKTRCTQFMERTAGTSTSGSESGFQREKQLRSLGGEYLPL</td>
</tr>
<tr>
<td>1186.1A</td>
<td>30</td>
<td>CVEQSTYDPAYFKITNSEQHKTLEKQTFQRMCDSKMRRYMILYTERELKENDH---PNVEYMM</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>sssss sss sssss hhhhh hhhhhhhhhhh sssss</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>ss hhhhhhh hh sssss hh sssss hhhhhhhhhhh sssss</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssss sssss hhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>HaKS1</td>
<td>143</td>
<td>DAVLQVPNPFCAERAKEAMPMVFAGAADDLRLRTKGVKAKIDGILVNCNSLFQPEPHELSSMM</td>
</tr>
<tr>
<td>HaKS2</td>
<td>161</td>
<td>EAFLQVPVFPCARKEAEHLMVFAGAADDLRKTSVPKIDGILVNCNSLFQPEPHELASMM</td>
</tr>
<tr>
<td>1186.1A</td>
<td>88</td>
<td>APSLARQDMVVEEPRLGKEAVAKIEKGQPK---SKRTRLVCTTSVGDMEQAGYQB</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>hhhhhhhhhhhhhhh hssss hhh</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>hhhhhhhhhhhhhhh hssss sssss sssss hhh</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>hhhhhhhhhhhhhhh hssss sssss sssss hhh</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \]

<table>
<thead>
<tr>
<th>HaKS1</th>
<th>203</th>
<th>VVNYKGFRGNISSYNLGSGGQCAgLISIDLRAQLPTQVNFNSYALWMS---IQHWNGYFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>221</td>
<td>IVNVYKGFRGNISSYNLGSGGQCAgLISIDLRAQLPTQVNFNSYALWMS---IQHWNGYFG</td>
</tr>
<tr>
<td>1186.1A</td>
<td>144</td>
<td>LTKLGLRVPVKRYMYQGCGACTGTLRLRAQAILAENRKGARLVCSEVTAVFRGPD</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>hhhhh sssss sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \quad \alpha^* \]

<table>
<thead>
<tr>
<th>HaKS1</th>
<th>261</th>
<th>NNRSNLSSCQLRFGRGAAALLLS-NRSSDRRRSKQYQLYTVRHKGADDKCYSCVFQEDN</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>279</td>
<td>NERSNLSSCQLRFGRGAAALLLS-NKTSGRRRSKQYQLYTVRHKGADDKSCVTQEDP</td>
</tr>
<tr>
<td>1186.1A</td>
<td>204</td>
<td>THLDSLVQALFPGDAIALIGVPMQFPEIKPFEMATQIAPDSEGAIADHLREAG-</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>hhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \quad \alpha^* \quad \alpha^* \]

<table>
<thead>
<tr>
<th>HaKS1</th>
<th>320</th>
<th>EKKIGVALSKDLMADVAG-EALNTNITTLGELVLPSEQLLFFVTLVARVKVFMKIKPYIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>338</td>
<td>VGKIGVALSKDLMADVAG-DATNDITTLGELVLPSEQLLFFATLVGKKFRMKIKPYIP</td>
</tr>
<tr>
<td>1186.1A</td>
<td>263</td>
<td>---LTFLNKDVPGIVLSKNITGALVEAEFLBIGISDYNSIFWIA---------</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>hhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \quad \alpha^* \quad \alpha^* \]

<table>
<thead>
<tr>
<th>HaKS1</th>
<th>379</th>
<th>DFLKAFEHFCINAGGRGAVLDEIEKKNLSEWHRFPMDNFYRTGNTSSSLWYDAAXCEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>397</td>
<td>DFLKAFEHFCINAGGRGAVLDEIEKKNLSEWHRFPMDNFYRTGNTSSSLWYDAAXSE</td>
</tr>
<tr>
<td>1186.1A</td>
<td>303</td>
<td>---HFGPDPLDDQVEQMLAKPEKNMCLIVSEYFEGSSACVLFIDLEMRK</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>hhhhh sssss hhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssss hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh hhhhhhhhhhh</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \quad \alpha^* \quad \alpha^* \]

<table>
<thead>
<tr>
<th>HaKS1</th>
<th>439</th>
<th>RGRKKGDRSWQIAFSGFKCNSAVMRALERTIDBTKEKTNPWTSEIHEFPVVPKIAQIDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HaKS2</td>
<td>457</td>
<td>RGRKKGDRSWQIAFSGFKCNSAVMRALERTIDBTKEKTNPWTSEIHEFPVVPKIAQIDL</td>
</tr>
<tr>
<td>1186.1A</td>
<td>352</td>
<td>GSTQ---EMLKQTGEGLEGLVFLFGFY---LTIETVVLRSVAI</td>
</tr>
<tr>
<td>HaKS1</td>
<td></td>
<td>sssssss sssssss sssssss</td>
</tr>
<tr>
<td>HaKS2</td>
<td></td>
<td>sssssss sssssss</td>
</tr>
<tr>
<td>1186.1A</td>
<td></td>
<td>sssssss sssssss</td>
</tr>
</tbody>
</table>

\[ \alpha^* \quad \beta^* \quad \alpha^* \quad \alpha^* \]
Figure 7