

[Click here to view linked References](#)

This is a post-peer-review, pre-copyedit version of an article published in *Planta* (2019). The final authenticated version is available online at: <https://doi.org/10.1007/s00425-019-03126-1>

1
2 **Functional characterization and structural modeling of *Helianthus annuus***
3
4 **(sunflower) ketoacyl-CoA synthases and role in seed oil composition**
5
6

7
8
9 González-Mellado, Damián¹; Salas, Joaquín J.¹; Venegas-Calderón, Mónica²; Moreno-
10 Pérez, Antonio J.¹; Garcés, Rafael¹ and Martínez-Force, Enrique¹
11

12
13 ¹Instituto de la Grasa (CSIC), Edificio 46, Campus Universitario Pablo de Olavide, Ctra.
14 de Utrera Km 1, 41013-Seville, Spain.
15

16
17 ²Departamento de Genética, Facultad de Biología, Universidad de Sevilla, 41012
18 Seville, Spain.
19

20
21
22
23
24
25
26
27
28
29 Running title: Sunflower ketoacyl-CoA synthases
30
31
32
33
34
35

36 Corresponding author: Joaquín J. Salas, Instituto de la Grasa (CSIC), Edificio 46,
37 Campus Universitario Pablo de Olavide, Ctra. de Utrera Km 1, 41013-Seville, Spain.
38
39 Email: jjsalas@ig.csic.es, Tlf: +34 954611550, Fax: +34 954616790
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Abstract

1
2 **Main conclusion:** The enzymes *HaKCS1* and *HaKCS2* are expressed in sunflower seeds
3
4 and contribute to elongation of C18 fatty acids, resulting in the C20-C24 fatty acids in
5
6 sunflower oil.
7

8
9
10
11 Most plant fatty acids are produced by plastidial soluble fatty acid synthases that
12
13 produce fatty acids of up to 18 carbon atoms. However, further acyl chain elongations
14
15 can take place in the endoplasmic reticulum, catalysed by membrane-bound synthases
16
17 that act on acyl-CoAs. The condensing enzymes of these complexes are the ketoacyl-
18
19 CoA synthase (KCSs), responsible for the synthesis of very long chain fatty acids
20
21 (VLCFAs) and their derivatives in plants, these including waxes and cuticle
22
23 hydrocarbons, as well as fatty aldehydes. Sunflower seeds accumulate oil that contains
24
25 around 2-3% of VLCFAs and studies of the fatty acid elongase activity in developing
26
27 sunflower embryos indicate that two different KCS isoforms drive the synthesis of these
28
29 fatty acids. Here, two cDNAs encoding distinct KCSs were amplified from RNAs
30
31 extracted from developing sunflower embryos and named *HaKCS1* and *HaKCS2*. These
32
33 genes are expressed in developing seeds during the period of oil accumulation and they
34
35 are clear candidates to condition sunflower oil synthesis. These two KCS cDNAs
36
37 complement a yeast elongase null mutant and when expressed in yeast, they alter the
38
39 host's fatty acid profile, proving the encoded KCSs are functional. The structure of these
40
41 enzymes was modelled and their contribution to the presence of VLCFAs in sunflower
42
43 oil is discussed based on the results obtained.
44
45
46
47
48
49
50

51 **Keywords:** *Helianthus annuus*, very long chain fatty acids, fatty acid elongase,
52
53 ketoacyl-CoA synthase, sunflower seeds.
54
55
56
57
58
59
60
61
62
63
64
65

Introduction

1
2 Plant fatty acids are synthesized by successive elongation of acyl-ACP
3 derivatives in plastids, events catalysed by the fatty acid synthase (FAS) soluble type 2
4 multienzyme system that brings together ketoacyl-ACP synthase (KAS), ketoacyl-ACP
5 reductase, hydroxyacyl-ACP dehydratase and enoyl reductase activities. Several
6 condensing enzymes have been described and characterized in the soluble system,
7 differing in the length of the substrate they condense. Thus, KAS III mainly catalyses an
8 initial reaction that uses malonyl-ACP and acetyl-ACP as substrates, whereas
9 condensations from C4 to C16 are undertaken by the KAS I enzyme and KAS II catalyses
10 the elongation of palmitoyl-ACP to stearoyl-ACP. The final products of plastidial fatty acid
11 synthase are C16 and C18 acyl chains, the precursor of most fatty acids present in plant
12 glycerolipids (Ohlrogge and Jaworski 1997). Nevertheless, fatty acids longer than 18C
13 are often present in plant tissues and these fatty acids are synthesized by elongation of
14 different acyl-CoAs in the endoplasmic reticulum (ER) by a membrane bound enzyme
15 complex with characteristics similar to FAS, known as fatty acid elongase or FAE
16 (Leonard et al. 2004; Haslam and Kunst 2013). FAE elongates acyl-CoA derivatives
17 rather than plastidial acyl-ACP derivatives, and this involves the same reactions of
18 condensation, dehydration and reduction, malonyl-CoA representing the substrate that
19 provides the carbon for the chain elongations. Moreover, the condensing activities within
20 this complex are the ketoacyl-CoA synthases (KCSs). In this process, the condensation
21 reaction is widely acknowledged to be the rate limiting and more determinant step (Millar
22 and Kunst 1997).

23
24 Biochemically, the active site and reaction mechanism of plant KAS and KCS
25 enzymes is identical, and quite similar to that of enzymes like plant chalcone synthase
26 or the KAS forms in *E. coli* (Jiang et al. 2008) which have been crystallized and
27 characterized structurally (Huang et al. 1998). Structural and site directed mutagenesis
28 demonstrated that a catalytic triad of residues is responsible for catalysis, formed by a
29 cysteine residue, one histidine and one His/Asn. The acyl moiety is transferred to the
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Cys residue of this triad, permitting the ensuing nucleophilic attack of the malonyl-CoA
2 substrate to form the ketoacyl product. The main structural difference evident between
3 KCS and KAS is the presence of two domains in the N-terminal region that anchor the
4 enzyme to the ER membranes.
5
6

7
8 Arabidopsis plants contain 21 copies of different KCS genes (Beisson et al.
9 2003), among which *AtKCS1* has been functionally characterized (Todd et al. 1999) and
10 is known to play an important role in wax synthesis. As such, knocking out the *KCS1*
11 gene produces an 80% decrease in C26 to C30 wax alcohols and aldehydes, having a
12 weaker impact on hydrocarbons and ketones. However, since the total amount of waxes
13 remains unaltered, this gene would appear to be redundant to some extent in
14 Arabidopsis, yet knockout plants have thinner stems and they are susceptible to
15 desiccation. The involvement of KCSs in the production of surface lipids and their
16 protective functions have been confirmed in Arabidopsis and other plant species (Vogg
17 et al. 2004; Lee et al. 2009). In addition, KCSs also play a role in determining the fatty
18 acid composition of the seed oil in some species, for example regulating the large
19 amounts of erucic acid (22:1) that accumulates in *Brassica napus* oil. In this regard,
20 BnFAE1.1 and BnFAE1.2 genes (Barret et al. 1998) encoding KCS proteins were shown
21 to be responsible, around 90.6% of the trait (Jourden et al. 1996), for the presence of
22 erucic acid in rapeseed oil, and mutations in the active site of these enzymes produce a
23 plant with a low erucic phenotype (Han et al. 2001; Roscoe et al. 2001). Similar KCS
24 isoforms have been defined in other Brassicaceae species like Arabidopsis or Camelina
25 and these enzymes elongate oleoyl-CoA or linoleoyl-CoA, yielding the unsaturated very
26 long chain fatty acids (VLCFAs) present in these species. This situation contrasts with
27 that reported in sunflower, as sunflower oil lacks unsaturated VLCFAs and only contains
28 small amounts of saturated ones: 0.5-1% of arachidic acid, 20:0; 1-2% behenic acid,
29 22:0; and trace amounts of lignoceric acid, 24:0 (Salas et al. 2005). However, the
30 accumulation of these fatty acids increases in sunflower mutants deficient in stearate
31 desaturase, with a higher content of stearic acid and VLCFAs (Fernández-Moya et al.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

2005). 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 *HaKCS1* and *HaKCS2*. The proteins encoded by these genes displayed similarities to *AtKCS2*, *AtKCS20* and *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.

Materials and Methods

Biological material

The wild-type CAS-6 Sunflower line (*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 $\mu\text{mol m}^{-2}\text{s}^{-1}$. 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,

1 root and seedling cotyledons) were collected from plants 20 days after germination. All
2 the samples were frozen and stored at -80 °C.
3

4 The *Escherichia coli* strain XL1-Blue (Stratagene, La Jolla, CA, USA) was used
5 as the plasmid host for *HaKCS1* and *HaKCS2* cloning. The bacteria were grown at 37
6 °C with shaking in LB medium (1% bacto tryptone, 0.5% bacto yeast extract, 1% NaCl
7 [pH 7]). Plasmid selection was based on ampicillin resistance (50 µg mL⁻¹).
8
9

10 The *Saccharomyces cerevisiae* strain W301A (MAT α *ura3-52 leu2-3, 112 trp1-*
11 *289a his3-1*) was used for heterologous gene expression and the TDY7005 *elo2* Δ -
12 *elo3* Δ /pELO3 strain was used for complementation assays (MAT α *lys2 ura3-52 trp1 leu2*
13 *elo2* Δ TRP *elo3* Δ TRP/pADHura-ELO3: kindly provided by Fred Beaudoin, Rothamsted
14 Research, UK).
15
16
17
18
19
20
21
22
23
24
25
26

27 **Cloning of the sunflower HaKCS1 and HaKCS2 genes**

28
29 Approximately 0.1 g of sunflower developing seed embryos at 18 DAF was
30 ground in liquid nitrogen in a precooled sterile mortar and pestle. Total RNA was
31 extracted using a Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO, USA) and
32 mRNA was isolated from this total RNA using the GenElute mRNA Miniprep kit (Sigma-
33 Aldrich). The mRNA pellet was resuspended in 33 µL of RNAase free TE buffer (10 mM
34 Tris-HCl, 1 mM EDTA [pH 8]) and the corresponding cDNA was synthesized using a
35 Ready-To-Go T-Primed First Strand Kit (GE Healthcare Life Science, Buckinghamshire,
36 UK).
37
38
39
40
41
42
43
44
45
46

47 Plant KCS protein sequences from public databases were aligned to identify
48 regions of homology using the ClustalX 2.0 program (Thompson et al. 1997: Figure S1).
49 From that data, three degenerate primers were designed to clone sunflower KCS genes
50 expressed during seed development: KCSEST-F1 (48 times degenerate), KCSEST-F2
51 (four times degenerate) and KCSEST-R1 (eight times degenerate: Table S1). A PCR
52 fragment corresponding to the 3'-end of *HaKCS1* (374 bp long) was amplified using the
53
54
55
56
57
58
59
60
61
62
63
64
65

1 oligonucleotide KCSEST-F2 in combination with the primer FA2Z (Table S1),
2 complementary to the sequence incorporated during the initial cDNA synthesis. The 5'-
3 end was amplified using the Smart™-RACE cDNA amplification kit (BD Bioscience
4 Clontech Company, Palo Alto, CA, USA) and the specific primer KCS1B-R4 (Table S1).
5
6 Similarly, an internal PCR fragment of *HaKCS2* (703 bp long) was amplified using
7 degenerate primers KCSEST-F1 and KCSEST-R1 (Table S1). The full sequence was
8 obtained at the 3'-end using the primer FA2Z and the specific primer KCS2A-F1 (Table
9 S1), while the 5'-end was amplified using the Smart™-RACE cDNA amplification kit and
10 the specific primer KCS2A-R2 (Table S1).
11
12
13
14
15
16
17
18
19

20 Primers were synthesized by MWG Biotech AG (Ebersberg, Germany) and all
21 the PCR fragments were cloned into the pGEM-T-Easy® vector (Promega, Madison, WI,
22 USA), transformed into XL1-Blue, sequenced by Secugen SL (Madrid, Spain), and
23 assembled to obtain DNA sequences of about 1720 bp and 1572 bp. Once their identities
24 were confirmed using the Blast software (Altschul et al. 1990), the complete cDNA
25 sequences of the sunflower ketoacyl-CoA synthases KCS1 and KCS2 were deposited in
26 GenBank with the accession numbers EU442581 and EU496864, respectively.
27
28
29
30
31
32
33
34
35
36
37

38 **Modelling of the HaKCS three-dimensional structures**

39
40 Homology modelling of the putative *HaKCS1* and *HaKCS2* protein structures was
41 carried out with the Deepview and Swiss Model Workspace software (Guex and Peitsch,
42 1997; <http://www.expasy.org/spdbv/>), using their protein sequences and the chalcone
43 synthase crystal structure from *Medicago sativa* as a template (Protein Data Bank
44 accession 1i86: Jez et al. 2001). The *HaKCS1* (85-467) and *HaKCS2* (104-485) residues
45 were modelled against this template, with sequence identities of 19.6% and 20.0%,
46 respectively. Molecular docking experiments were performed using I-TASSER
47 (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), and with malonyl-CoA and arachidic
48 acid molecules as substrates (Yang et al. 2015). Critical residue mapping and
49 visualization was performed using the UCSF Chimera package (Pettersen et al. 2004),
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 and the transmembrane tendency hydrophobicity profiles of these sunflower KCS
2 enzymes were determined using the ProtScale software (Gasteiger et al. 2005).
3
4
5

6 **Bioinformatics analysis**

7
8 Sequences homologous to the predicted protein sequences of sunflower KCSs
9 were retrieved using the BLASTP program (www.ncbi.nlm.nih.gov) and an alignment of
10 the amino acid sequences for the KCS proteins deposited at GENBANK was performed
11 with the Clustal X v.2.0 program using the default settings (Thompson et al. 1997). These
12 alignments were used to generate a phylogenetic tree based on the neighbor-joining
13 algorithm (Saitou and Nei 1987), and the resulting phenogram was drawn using the
14 MEGA 4.0 program (Tamura et al. 2007). Other programs used to identify putative
15 transmembrane anchorage regions or the transmembrane tendency were TMHMM
16 (Krogh et al. 2001), OCTOPUS (Viklund and Elofsson 2008) and ProtScale (Gasteiger
17 et al. 2005).
18
19
20
21
22
23
24
25
26
27
28
29
30

31 **RT-QPCR studies of gene expression**

32
33 The cDNAs from different sunflower tissues (leaf, stem, roots, developing seed
34 **embryos** and cotyledons) were subjected to real-time quantitative PCR (RT-QPCR) on
35 a CFX96™ Real-Time PCR Detection System (Bio-Rad, CA, USA) using SYBR Green I
36 (SsoAdvanced™ SYBR® Green Supermix, Bio-Rad, CA, USA) and specific primer pairs:
37 HaKCS1_QPCR-F and HaKCS1_QPCR-R for the *HaKCS1* gene; and HaKCS2_QPCR-
38 F and HaKCS2_QPCR-R for the *HaKCS2* gene (Table S1). The PCR products obtained
39 were 108 bp and 142 bp long, respectively. Polymerase activation and DNA denaturation
40 were carried out at 95 °C for 30 s before performing 40 PCR cycles of 95 °C for 15 s and
41 60 °C for 30 s, during which the resulting fluorescence was monitored. The calibration
42 curves were drawn up using sequential dilutions of the cDNAs and the Livak method was
43 employed to calculate the comparative expression of the samples (Livak and Schmittgen
44 2001). The sunflower *HaACT1* actin gene (GenBank Accession number FJ487620) was
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2 used as an internal reference to normalize the readings to the relative amount of cDNA
3 in each sample, amplified with the specific HaActin-F1 and HaActin-R1 primers (Table
4 S1).
5
6
7

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

9

10 The *HaKCS1* and *HaKCS2* genes were expressed heterologously in yeast using
11 the pYES2 and pYES3 expression vectors, which carry the genes required for uracil
12 (*URA3*) and tryptophan (*TRP1*) biosynthesis under the control of a *GAL1* promoter.
13 *HaKCS1* was amplified with the primers KCS1-F pYES2 and KCS1-R pYES2 and
14 *HaKCS2* with KCS2-F pYES3 and KCS2-R pYES3 (Table S1). These primers included
15 restriction sites for *KpnI* and *BamHI*, which were used to clone these cDNAs into the
16 vectors. Recombinant plasmids were sequenced and introduced into the yeast strain
17 W301A, transforming the yeast by the PLATE method (Becker and Lundblad, 1994).
18
19
20
21
22
23
24
25
26
27
28
29
30

31 **S. cerevisiae fatty acid analysis**

32

33 The yeast cells were harvested by centrifugation at 1500g for 5 min at 4 °C and
34 they were washed with distilled water before determining the total fatty acid composition
35 using the one-step method proposed previously (Garcés and Mancha 1993). A volume
36 of 3.3 mL methanol/toluene/dimethoxypropane/sulphuric acid (39:20:5:2) and 1.7 mL
37 heptane was added to the cell pellet, and the mixture was heated at 80 °C for 1 h. After
38 cooling, the upper phase containing the fatty acid methyl esters was transferred to a
39 fresh tube, washed with 6.7% sodium sulphate and evaporated to dryness with nitrogen.
40 The methyl esters were dissolved in an appropriate volume of heptane and analysed by
41 GLC (Hewlett–Packard 6890 gas chromatography apparatus; Palo Alto, CA, USA) using
42 a Supelco SP-2380 fused-silica capillary column (30 m length, 0.25 mm i.d., 0.20 µm film
43 thickness: Supelco, Bellefonte, PA, USA). Hydrogen was used as the carrier gas at 28
44 cm s⁻¹, the temperature of the flame ionization detector and injector was 200 °C, the oven
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 *elo2*Δ *elo3*Δ/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 pI of 9.07. The predicted *HaKCS2* protein was 514 amino acids long, it was slightly larger at 57.60 kDa and with a pI 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 comunis*, 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

1 online server Swiss Model (<http://swissmodel.expasy.org/>), based on the known
2 structures of homologous proteins (see the secondary and tertiary structures in Figures
3 3 and 4). These models did not include the N-terminal ends, corresponding to the
4 transmembrane regions, and they involved residues 83-467 and 101-485 of *HaKCS1*
5 and *HaKCS2*, respectively. The models predicted these enzymes as homodimers, with
6 both coupled monomers displaying triangle forms (Figure 4a and b), similar to polyketide
7 synthases and type III ketoacyl-ACP synthases. These homodimers may be associated
8 with the presence of a putative leucine zipper motif in C-terminal region analogous to
9 that present in BnFAE1 (Barret et al, 1998). The body of the two enzymes is made up of
10 an α - β - α - β - α structure at the N-terminal end and an α - β - α - α structure at the C-terminal
11 end. These structures embraced the catalytic residues formed by the Cys223, His390
12 and Asn423 triad in *HaKCS1*, and Cys241, His408 and Asn441 in *HaKCS2*, and they
13 were grouped close to each other to form a hydrophobic cavity across the protein that
14 allows substrate entry (Figure 5a, b and c). Docking analysis with substrates like malonyl-
15 CoA or arachidic acid showed that the hydrophobic pocket can accommodate both
16 substrates at the same time, allowing us to predict the amino acid residues involved in
17 substrate-enzyme binding (Figure 5 and Table S2).
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **Analysis of gene expression in leaves and vegetative tissues**

41
42 The expression of the *HaKCS1* and *HaKCS2* genes was studied by RT-QPCR in
43 different plant organs: leaves, root, stem, cotyledons from seedlings and developing
44 seeds. *HaKCS2* was expressed more strongly than *HaKCS1* in all the tissues except in
45 the developing seed. These differences were more significant in roots, stems and
46 cotyledons, where the increase reached approximately one order of magnitude (Figure
47 6). In the developing seeds, *HaKCS1* expression increased from DAF 12 to 18 and it
48 reduced thereafter. The expression of *HaKCS2* was more stable than that of *HaKCS1*
49 during seed development, peaking in seeds harvested at 18 DAF.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

1
2 The *Saccharomyces elo2ΔTrp elo3ΔTrp* strain lacks the ELO2 and ELO3
3 endogenous elongases, which provokes lethality and thus, this mutant was used to
4 confirm that *HaKCS1* and *HaKCS2* act as functional elongases. The experiment
5 consisted in growing mutants that harbour the *HaKCS* genes cloned into the pADH-LEU
6 plasmid and pADH-URA::elo3 to complement the mutation. Yeast transformed with
7 pADH-LEU::HaKCS1, pADH-LEU::HaKCS2 and pADH-LEU::AtKCS9 were plated on
8 minimal medium containing 5'-fluoroacetic acid. The yeast containing pADH-URA::elo3
9 and the empty pADH-LEU plasmid could not grow on this medium, whereas the lines
10 transformed with the plasmids containing the KCS genes were able to grow,
11 demonstrating that the genes identified in this study encoded active elongases (Figure
12 7).

Heterologous expression of *HaKCS* genes in *S. cerevisiae*

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 The *HaKCS1* and *HaKCS2* genes were cloned into the pYES2 and pYES3
32 vectors, and the resulting constructs were used to transform the wild type W301A yeast
33 strain in order to investigate their effect on the yeast's fatty acid composition (Figure 8).
34 In the case of *HaKCS1*, cultures displayed an important increase in VLCFAs upon the
35 induction of gene expression. This increase predominantly affected the proportion of 22:0
36 and 24:0 fatty acids, which increased 3-4 fold, and to a lesser extent to the levels of 26:0
37 fatty acids, the content of which doubled (Figure 8a). At the same time, the level of stearic
38 and oleic acids in the cultured yeast decreased, with a significant increase in the levels
39 of palmitoleic acid. The expression of *HaKCS2* had a weaker impact on the VLCFA
40 content than its homologue, mainly affecting the 22:0 and 24:0 content that doubled in
41 the induced cultures (Figure 8b). There were no significant differences in 26:0 content,
42 while the impact of *HaKCS2* induction on the other fatty acids was similar to that
43 observed for *HaKCS1*, producing a decrease in stearic and oleic acid and an increase in
44 palmitoleic acid.

Discussion

Sequences and phylogenetic studies

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. *Brasica napus*, *Sorghum bicolor*, *Oryza sativa*, etc.) and subsequently RACE. The sequences obtained corresponded to two similar proteins with an alkaline pI, 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 *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 *HaKCSs*. 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 *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). *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 (α , β , γ , δ , ζ , ϵ , η and θ), arising from an ancestral genome duplication that can be inferred from their chromosomal location (Blanc et al. 2003; Joubès et al. 2008). The *HaKCS1*

1 and *HaKCS2* proteins cluster within a group in the ζ subclass that includes
2 dicotyledonous, monocotyledonous plants and briophyta: *HaKCS1* displays 75 and 76%
3 homology with the *AtKCS2/DAISY* and *AtKCS20* proteins from *Arabidopsis* respectively,
4 which are genes related with the synthesis of surface lipids (Franke et al., 2009; Lee et
5 al., 2009). Moreover, *HaKCS2* displays 81% homology with *KCS11* from the same plant.
6
7 Of all the different forms included in this study, the most similar KCSs to the sunflower
8 proteins found were those from *Vitis vinifera* and *Ricinus communis*.
9

10
11
12
13
14
15 The genes coding for *HaKCS1* and *HaKCS2* corresponded to
16 *HanXRQChr16g0514651* and *HanXRQChr13g0402491* locus tags from the sunflower
17 genome recently disclosed (Badouin et al 2017; [https://www.heliagene.org/HanXRQ-](https://www.heliagene.org/HanXRQ-SUNRISE/)
18 *SUNRISE/*). Compared to their homologs in *Arabidopsis* (*AtKCS2* and
19 *AtKCS20/HaKCS1* and *AtKCS11/HaKCS2*; see Figure 1), the structures of *HaKCS1* and
20 *HaKCS2* genes are pretty conserved, displaying one and zero introns each one
21 respectively.
22
23
24
25
26
27
28
29
30

31 **Structural model of sunflower KCSs**

32
33 Unlike plastidial condensing enzymes, KCSs are membrane-bound enzymes
34 located in the ER (Joubès et al. 2008), which requires the presence of transmembrane
35 domains in at least some cases. The hydrophobicity profiles of *HaKCS1* and *HaKCS2*
36 indicated the presence of two high hydrophobic regions close to the N-terminus that
37 would bind the enzyme tightly to the ER membrane. These transmembrane domains
38 were similar to those described in *AtKCS1*, which also has two hydrophobic regions in
39 the N-terminus (Joubès et al. 2008), the KCS from *Tropaeolum majus* (Mietkiewska et
40 al. 2004) and that from *Lunaria annua* (Guo et al. 2009).
41
42
43
44
45
46
47
48
49
50

51 While there are currently no crystallographic data for KCS enzymes, the
52 structures of related enzymes have been elucidated, such as bacterial polyketide
53 synthases and KAS III. On the basis of these structures, and more specifically on the C-
54 terminal domain of chalcone synthase of *Medicago sativa* (PDB 1i86: Jez et al. 2001), it
55 is possible to obtain three-dimensional models of KCSs by applying modelling programs
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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'

1 biosynthetic machinery in order to synthesise the long chain fatty acids necessary for
2 cell viability, demonstrating the functionality of these genes.
3
4
5

6 **Effects on fatty acid composition in yeast**

7
8 The activity of the sunflower *HaKCS1* and *HaKCS2* was evaluated by heterologous
9 expression in yeast, the system of choice to evaluate membrane enzymes expressed in
10 the ER as they undergo correct organ compartmentation. Indeed, this approach has
11 already been used to study the substrate specificity of these enzymes in *Arabidopsis*
12 *thaliana* (Blacklock and Jaworski 2006) and *Brassica napus* (Katavic et al. 2002). The
13 system used for this approach was based on the pYES plasmids and the W301A cell
14 line, driving protein expression through the activity of the galactose operon.
15
16
17
18
19
20
21
22
23

24 To characterize these enzymes they were expressed separately in the microbial
25 host, and the fatty acid composition of the microorganism was analysed before and after
26 enzyme induction to examine the impact of the sunflower elongases. *HaKCS1*
27 expression produced a clear effect on fatty acid production, increasing the VLCFA
28 content of the yeast between 7 to 3.5 fold, from 1-2% (Welch and Burlingame 1973) up
29 to 7.2%. This increase is predominantly due to the accumulation of saturated 22:0, 24:0
30 and 26:0 fatty acids, in the absence of any additional unsaturated VLCFAs like erucic or
31 gondoic fatty acids. The increase in these fatty acids was paralleled by a decrease in the
32 proportion of stearic and oleic acid in the culture. There was also an increase in
33 palmitoleic acid (16:1), which suggested that the yeast cells try to compensate for the
34 deficit in oleic acid by increasing their desaturase activity, which in this case desaturates
35 palmitic acid to palmitoleic. This effect could be a mechanism to maintain the normal
36 fluidity of the cell membranes. By contrast, the impact of *HaKCS2* on lipid composition
37 was less pronounced, mainly affecting 20:0, 22:0 and 24:0 but not 26:0 fatty acids, which
38 remained at the same proportion as in uninduced cells. Again, the effect on unsaturated
39 VLCFAs was limited and the effect on endogenous fatty acids from yeast was similar to
40 that caused by *HaKCS1*, producing a decrease in stearic and oleic acids, and enhancing
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 the desaturation of palmitic to palmitoleic acid. Furthermore, these results suggest that
2 *HaKCS2* has a lower specificity towards longer acyl-CoAs, in good agreement with
3 structural models given the narrower and shallower hydrophobic pocket to accommodate
4 the substrate.
5
6

7
8 Both sunflower KCSs influence the VLCFAs present in sunflower oil, essentially
9 arachidic and behenic acids with only traces of lignoceric acid. All of them are saturated
10 VLCFAs, not accompanied by any unsaturated ones, meaning that unlike the enzymes
11 in other species (e.g. *Brassica napus*, *Camelina sativa* or *Arabidopsis*), sunflower KCSs
12 are very specific for saturated derivatives. Earlier biochemical studies indicated that
13 sunflower seed microsomes contain at least two forms of KCS that can be separated by
14 conventional purification methods (Salas et al. 2005). These forms could elongate
15 several acyl-CoA substrates, with highest activities towards stearoyl-CoA and
16 arachidoyl-CoA, whereas no such elongation was observed with oleoyl-CoA as the
17 substrate. These specificity profiles share similarities with those of the *HaKCS1* and
18 *HaKCS2* enzymes, which were expressed in seeds and were only active towards
19 saturated substrates. Important amounts of VLCFAs longer than 22:0 are not detected
20 in sunflower oil, even though the sunflower KCSs studied here could elongate up to C26
21 fatty acids. This may be due to discrimination of these fatty acids by the enzymes
22 responsible for TAG synthesis or because these fatty acids were quickly metabolized in
23 other pathways active in sunflower seeds, such as those involved in the synthesis of
24 waxes or other surface lipids (Franke et al., 2009; Lee et al., 2009; Broughton et al.
25 2018).
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 **Conclusions**

52 The synthesis of VLCFAs requires the action of membrane bound elongase
53 complexes on acyl-CoA derivatives, and the condensing enzymes of these complexes
54 or KCSs are limiting and determinant. In the present work, two KCS genes were cloned
55 from developing sunflower seeds, *HaKCS1* and *HaKCS2*, and they correspond to
56
57
58
59
60
61
62
63
64
65

1 enzymes with two transmembrane domains that **may** anchor them to ER membranes.
2 Their structures were modelled, identifying hydrophobic pockets of different sizes that
3 suggest different substrate **preferences**. The expression profile of these genes differs
4 from that of other KCSs expressed in seeds, like brassica *FAE1*, as they are ubiquitously
5 expressed in different plant organs, especially *HaKCS2*. Nevertheless, their expression
6 in developing seeds increased during the period of oil accumulation. When
7 heterologously expressed in yeast, these genes were able to complement the lethality of
8 the *elo2elo3* mutant that is deficient in fatty acid elongation, proving they encoded active
9 enzymes. Their expression in yeast modified the fatty acid composition of the cells,
10 increasing the concentrations of VLCFAs. *HaKCS1* induced saturated VLCFA
11 accumulation up to C26, whereas *HaKCS2* only affected C20-C24 fatty acids, again
12 suggesting differences in substrate specificity that fitted well with the protein structure.
13 The activity of the proteins was similar to that of KCS from sunflower microsomes
14 previously characterized, such that they could be involved in the synthesis of the VLCFAs
15 present in sunflower oil and contribute to the synthesis of surface lipids present in
16 sunflower **embryos**.

37 **Acknowledgements**

38 We acknowledge Fred Beaudoin (Rothamsted Research, UK) for kindly providing the
39 *Saccharomyces cerevisiae* double mutant *elo2* and *elo3* strain.
40
41
42
43
44
45

46 **Availability of data and material**

47 The dataset and genes produced in the present study are available on reasonable
48 request.
49
50
51
52
53
54
55

56 **Ethics approval and consent to participate**

57
58
59
60
61
62
63
64
65

1 The experimental research in this manuscript was complied with the institutional, national
2 and international guidelines.
3
4
5

6 **Consent for publication**

7
8 Not applicable
9

10 **Competing interest**

11
12 The authors declare they have not competing interests
13
14
15

16 **Funding**

17
18 This work was funded by the MINECO and FEDER Projects AGL2014-53537-R and
19 AGL2017-83449-R. They are free competition projects funded for the Spanish
20 government for the development of scientific research.
21
22
23
24
25
26
27
28
29
30

31 **Authors' contribution**

32
33 DG-M performed most of the experimental work. AM-P participated in plasmid
34 construction and rtPCR. MVC, EM-F, RG and JJS participated in the work direction and
35 experimental design. JJS and EM-F wrote and revised the manuscript.
36
37
38
39
40
41
42

43 **References**

44
45 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment
46 search tool. *J Mol Biol* 215:403-410. [https://doi.org/10.1016/S0022-](https://doi.org/10.1016/S0022-2836(05)80360-2)
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Berard A, Berges H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadiou C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Vares D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Munos S, Vincourt P, Rieseberg LH, Langlade NB (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546:148-152. <https://doi.org/10.1038/nature22380>

Barret P, Delourme R, Renard M, Domergue F, Lessire R, Delseny M, Roscoe TJ (1998) A rapeseed *FAE1* gene is linked to the E1 locus associated with variation in the content of erucic acid. *Theor Appl Genet* 96:177-186. <https://doi.org/10.1007/s001220050725>

Beisson F, Koo AJ, Ruuska S, Schwender J, Pollard M, Thelen JJ, Paddock T, Salas JJ, Savage L, Milcamps A, Mhaske VB, Ohlrogge JB (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 Physiol* 132:681-697. <https://doi.org/10.1104/pp.103.022988>

Becker DM, Lundblad V (1994) Introduction of DNA into yeast cells, *Curr Protoc Mol Biol* 27:13-7. <https://doi.org/10.1002/0471142727.mb1307s27>

Blacklock BJ, Jaworski, JG (2002) Studies into factors contributing to substrate specificity of membrane-bound 3-ketoacyl-CoA synthases. *Eur J Biochem* 269:4789-4798. <https://doi.org/10.1046/j.1432-1033.2002.03176.x>

Blacklock BJ, Jaworski JG (2006) Substrate specificity of Arabidopsis 3-ketoacyl-CoA synthases. *Biochem Biophys Res Comm* 346:583-590. <https://doi.org/10.1016/j.bbrc.2006.05.162>

1
2 Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-
3 scale duplications in the Arabidopsis genome. *Genome Res.* 13:137-144.
4 <https://doi.org/10.1101/gr.751803>
5

6 Broughton R, Ruíz-Lopez N, Hassall KL, Martínez-Force E, Garcés R, Salas JJ,
7 Beaudoin F (2018). New insights in the composition of wax and sterol esters in
8 common and mutant sunflower oils revealed by ESI-MS/MS. *Food Chem* 269:70-
9 79. <https://doi.org/10.1016/j.foodchem.2018.06.135>
10
11
12
13
14

15 Chiron H, Wilmer J, Lucas MO, Nesi N, Delseny M, Devic M, Roscoe TJ (2015)
16 [Regulation of FATTY ACID ELONGATION1 expression in embryonic and vascular](https://doi.org/10.1007/s11103-015-0309-y)
17 [tissues of *Brassica napus*. *Plant Mol Biol* 88:65-83. \[https://doi.org/10.1007/s11103-\]\(https://doi.org/10.1007/s11103-015-0309-y\)](https://doi.org/10.1007/s11103-015-0309-y)
18 [015-0309-y](https://doi.org/10.1007/s11103-015-0309-y)
19
20
21
22
23

24 Costaglioli P, Joubès J, García C, Stef M, Arveile, B, Lessiere R, Garbay B (2005).
25 Profiling candidate genes involved in wax biosynthesis in *Arabidopsis thaliana* by
26 microarray analysis. *Biochim Biophys Acta* 1734:247-258.
27 <https://doi.org/10.1016/j.bbalip.2005.04.002>
28
29
30
31
32

33 Fernández-Moya V, Martínez-Force E, Garcés R (2005). Oils from improved high stearic
34 acid sunflower seeds. *J Agric Food Chem* 53:5326-5330.
35 <https://doi.org/10.1021/jf0503412>
36
37
38
39

40 Franke R, Höfer R, Briesen I, Emsermann M, Efremova N, Yephremov A, Schreiber L
41 (2009). The DAISY gene from *Arabidopsis* encodes a fatty acid elongase
42 condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and
43 the chalaza-micropyle region of seeds. *Plant J.* 57:80-95.
44 <https://doi.org/10.1111/j.1365-313X.2008.03674.x>
45
46
47
48
49
50

51 Garcés R, Mancha M (1993) One-step lipid extraction and fatty acid methyl esters
52 preparation from fresh plant tissues. *Anal Biochem* 211:139-143.
53 <https://doi.org/10.1006/abio.1993.1244>
54
55
56
57

58 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005)
59 Protein identification and analysis tools on the ExpASy Server. In: *The Proteomics*
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Protocols Handbook, John M Walker (ed), Humana Press (2005), pp. 571-607.

<https://doi.org/10.1385/1-59259-890-0:571>

Ghanevati M, Jaworski JG (2001) Active-site residues of a plant membrane-bound fatty acid elongase β -ketoacyl-CoA synthase, FAE1 KCS. *Biochem Biophys Acta* 1530:77-85. [https://doi.org/10.1016/S1388-1981\(00\)00168-2](https://doi.org/10.1016/S1388-1981(00)00168-2)

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723. <https://doi.org/10.1002/elps.1150181505>

Guo Y, Mietkiewska E, Francis T, Katavic V, Brost JM, Giblin M, Barton DL, Taylor, DC (2009) Increase in nervonic acid content in transformed yeast and transgenic plants by introduction of a *Lunaria annua* L. 3-ketoacyl-CoA synthase (KCS) gene. *Plant Mol Biol* 69: 565-575. <https://doi.org/10.1007/s11103-008-9439-9>

Guo HS, Zhang YM, Sun XQ, Li MM, Hang YY, Xue JY (2016) Evolution of the KCS gene family in plants: the history of gene duplication, sub/neofunctionalization and redundancy. *Mol Genet Genomics* 291:739-752. <https://doi.org/10.1007/s00438-015-1142-3>

Han J, Lühs W, Sonntag K, Zähringer U, Borchardt DS, Wolter FP, Heinz E, Frentzen M (2001) Functional characterization of β -ketoacyl-CoA synthase genes from *Brassica napus* L. *Plant Mol Biol* 46:229-239. <https://doi.org/10.1023/A:101066512>

Haslam TM, Kunst L (2013) Extending the story of very-long-chain fatty acid elongation. *Plant Sci* 210:93-107. <https://doi.org/10.1016/j.plantsci.2013.05.008>

Heath RJ, Rock CO (2002) The claisen condensation in biology. *Natural Product Reports* 19:581-596. <https://doi.org/10.1039/B110221B>

Huang W, Jia J, Edwards P, Dehesh K, Schneider G, Lindqvist Y (1998). Crystal structure of β -ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J* 17:1183-1191. <https://doi.org/10.1093/emboj/17.5.1183>

- 1
2 Jez JM, Bowman ME, Noel JP (2001) Structure-guided programming of polyketide chain-
3 length determination in chalcone synthase. *Biochem* 40:14829-14838.
4 <https://doi.org/10.1021/bi015621z>
5
- 6 Jiang C, Kim SY, Suh DY (2008). Divergent evolution of the thiolase superfamily and
7 chalcone synthase family. *Mol Phylogenetics Evol* 49:691-701.
8 <https://doi.org/10.1016/j.ympev.2008.09.002>
9
- 10 Joubès J, Raffaele S, Bourdenx B, García C, Laronche-Traineau J, Moreau P, Domergue
11 F, Lessire R (2008) The VLCFA elongase gene family in *Arabidopsis thaliana*:
12 phylogenetic analysis, 3D modeling and expression profiling. *Plant Mol. Biol.*
13 67:547-566. <https://doi.org/10.1007/s11103-008-9339-z>
14
- 15 **Jourdren C, Barret P, Horvais R, Foisset N, Delourme R, Renard M (1996) Identification**
16 **of RAPD markers linked to the loci controlling erucic acid level in rapeseed. *Mol***
17 **Breed 2: 61-71. <https://doi.org/10.1007/BF00171352>**
18
- 19 Katavic V, Mietkiewska E, Barton DL, Giblin EM, Reed DW, Taylor DC (2002) Restoring
20 enzyme activity in nonfunctional low erucic acid *Brassica napus* fatty acid elongase
21 1 by a single amino acid substitution. *Eur J Biochem* 269:5625-5631.
22 <https://doi.org/10.1046/j.1432-1033.2002.03270.x>
23
- 24 Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane
25 protein topology with a hidden Markov model: application to complete genomes. *J*
26 *Mol Biol* 305:567-580. <https://doi.org/10.1006/jmbi.2000.4315>
27
- 28 Lassner MW, Lardizabal K, Metz JG (1996) A jojoba β -ketoacyl-CoA synthase cDNA
29 complements the canola fatty acid elongation mutation in transgenic plants. *Plant*
30 *Cell* 8:281-292. <https://doi.org/10.1105/tpc.8.2.281>
31
- 32 Lee SB, Jung SJ, Go YS, Kim HU, Kim JK, Cho HJ, Ohkmae KP, Suh MC (2009). Two
33 *Arabidopsis* 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are
34 functionally redundant in cuticular wax and root suberin biosynthesis, but
35 differentially controlled by osmotic stress. *Plant J* 60:462-475.
36 <https://doi.org/10.1111/j.1365-313X.2009.03973.x>
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1 Leonard AE, Pereira SL, Sprecher H, Huang YS (2004). Elongation of long-chain fatty
2 acids. *Prog. Lipid Res.* 43:36-54. [https://doi.org/10.1016/S0163-7827\(03\)00040-7](https://doi.org/10.1016/S0163-7827(03)00040-7)
3
- 4 Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-
5 time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 25:402-408.
6
7 <https://doi.org/10.1006/meth.2001.1262>
8
9
- 10 Mathieu M, Zeelen J, Pauptit RA, Erdmann R, Kunau WH, Wierenga RK (1994) The 2.8Å
11 Crystal Structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces*
12 *cerevisiae*: a five-layered $\alpha\beta\alpha\beta\alpha$ structure constructed from two core domains of
13 identical topology. *Structure* 2:797-808. <https://doi.org/10.1016/S0969->
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Mietkiewska E, Giblin EM, Wang S, Barton DL, Dirpaul J, Brost JM, Katavic V, Taylor
DC (2004) Seed-specific heterologous expression of a nasturtium FAE gene in
Arabidopsis results in a dramatic increase in the proportion of erucic acid. *Plant*
Physiol 136:2665-2675. <https://doi.org/10.1104/pp.104.046839>
- Millar AA, Kunst L (1997) Very-long-chain fatty acid biosynthesis is controlled through
the expression and specificity of the condensing enzyme. *Plant J.* 12:121-131.
<https://doi.org/10.1046/j.1365-313X.1997.12010121.x>
- Oh CS, Toke DA, Mandala S, Martin CE (1997) *ELO2* and *ELO3*, homologues of the
Saccharomyces cerevisiae ELO1 gene, function in fatty acid elongation and are
required for sphingolipid formation. *J Biol Chem* 272:17376–17384.
<https://doi.org/10.1074/jbc.272.28.17376>
- Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. *Annu. Rev Plant*
Physiol Plant Mol Biol 48:109-136.
<https://doi.org/10.1146/annurev.arplant.48.1.109>
- Paul S, Gable K, Beaudoin F, Cahoon E, Jaworski J, Napier JA, Dunn TM (2006)
Members of the *Arabidopsis* FAE1-like 3-ketoacyl-CoA synthase gene family
substitute for the elop proteins of *Saccharomyces cerevisiae*. *J Biol Chem*
281:9018-9029. <https://doi.org/10.1074/jbc.M507723200>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin, TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612. <https://doi.org/10.1002/jcc.20084>
- Qiu X, Janson CA, Konstantinidis AK, Nwagwu S, Silverman C, Smith WW, Khandekar, S, Lonsdale J, Abdel-Meguid SS (1999) Crystal structure of β - ketoacyl-acyl carrier protein synthase III: a key condensing enzyme in bacterial fatty acid biosynthesis. *J Biol Chem* 274:36465-36471. <https://doi.org/10.1074/jbc.274.51.36465>
- Qiu X, Jason CA, Smith WW, Head M, Lonsdale J, Konstantinidis AK (2001) Refined structures of beta-ketoacyl-acyl carrier protein synthase III. *J Mol Biol* 307:341-356. <https://doi.org/10.1006/jmbi.2000.4457>
- Roscoe T, Lessire R, Puyaubert J, Renard M, Delseny M (2001) Mutations in the fatty acid elongation 1 gene are associated with a loss of β -ketoacyl-CoA synthase activity in low erucic acid rapeseed. *FEBS Lett* 492:107-111. [https://doi.org/10.1016/S0014-5793\(01\)02243-8](https://doi.org/10.1016/S0014-5793(01)02243-8)
- Salas JJ, Martínez-Force E, Garcés R (2005). Very long chain fatty acid synthesis in sunflower kernels. *J Agric Food Chem* 53:2710-2716. <https://doi.org/10.1021/jf047939e>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599. <https://doi.org/10.1093/molbev/msm092>
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882. <https://doi.org/10.1093/nar/25.24.4876>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
- Todd J, Post-Beittenmiller D, Jaworski JG (1999) KCS1 encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. *Plant J* 17:119-130. <https://doi.org/10.1046/j.1365-313X.1999.00352.x>
- Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*. 24:1662-1668. <https://doi.org/10.1093/bioinformatics/btn221>
- Vogg G, Fischer S, Leide J, Emmanuel E, Jetter R, Levy AA, Riederer M (2004) Tomato fruit cuticular waxes and their effects on transpiration barrier properties: functional characterization of a mutant deficient in a very-long-chain fatty acid β -ketoacyl-CoA synthase. *J Exp Bot* 55:1401-1410. <https://doi.org/10.1093/jxb/erh149>
- Welch JW, Burlingame AL (1973) Very long-chain fatty acids in yeast. *J Bacteriol* 115:464-466. ISSN: 1098-5530
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* 12:7-8. <https://doi.org/10.1038/nmeth.3213>

38 **Figure Legends**

39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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, *Brasica juncea*; Bn, *Brasica 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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.

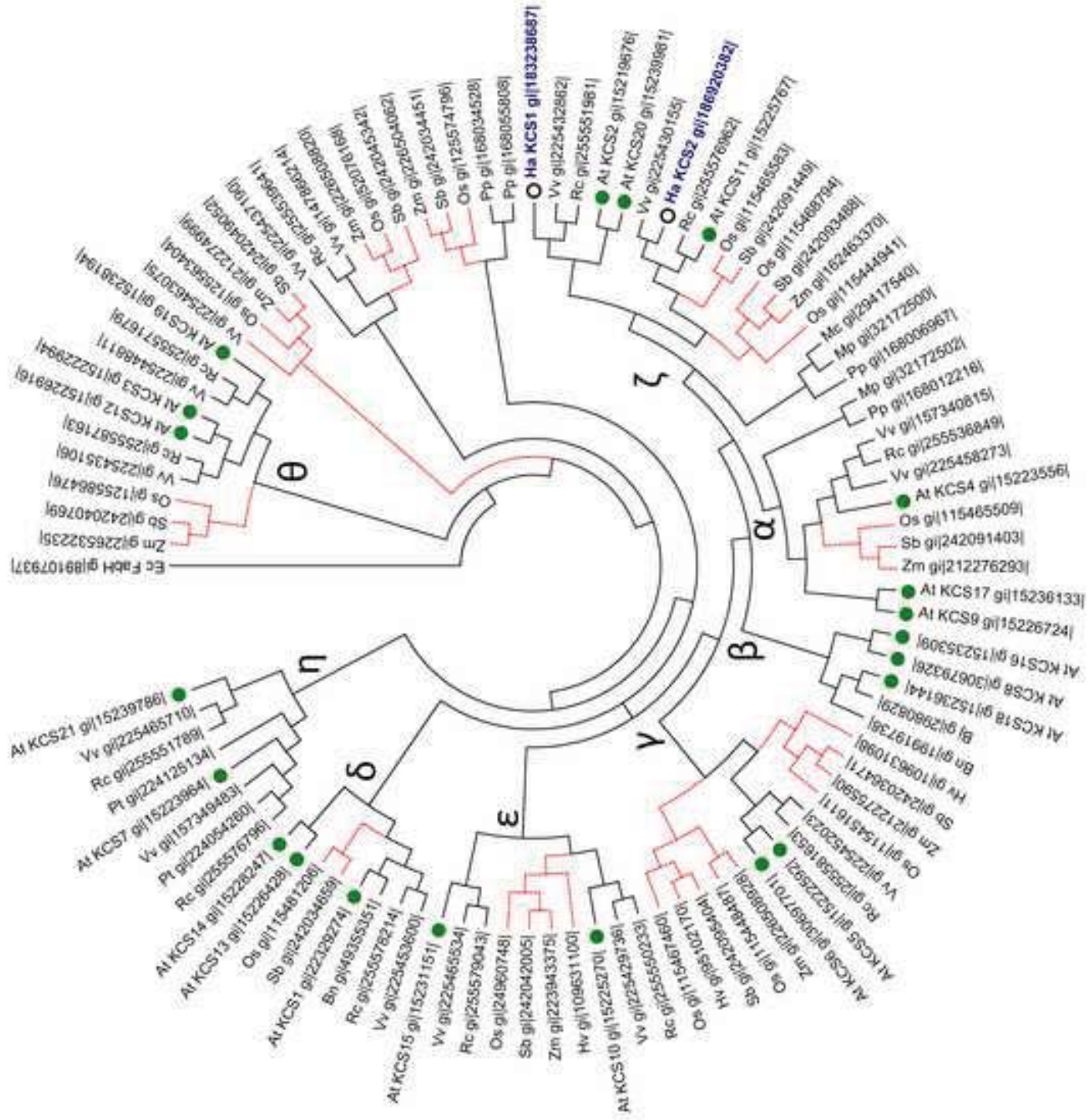
Figure 4. Proposed structural models for sunflower *HaKCS1* (a, c and e) and *HaKCS2* (b, d and f) KCS homodimers based on that of *Medicago sativa* chalcone synthase (1i86.1A: Jez et al. 2001). Panels a and b are ribbon diagrams in which the catalytic triad residues are in green. Panels c, d, e and 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).

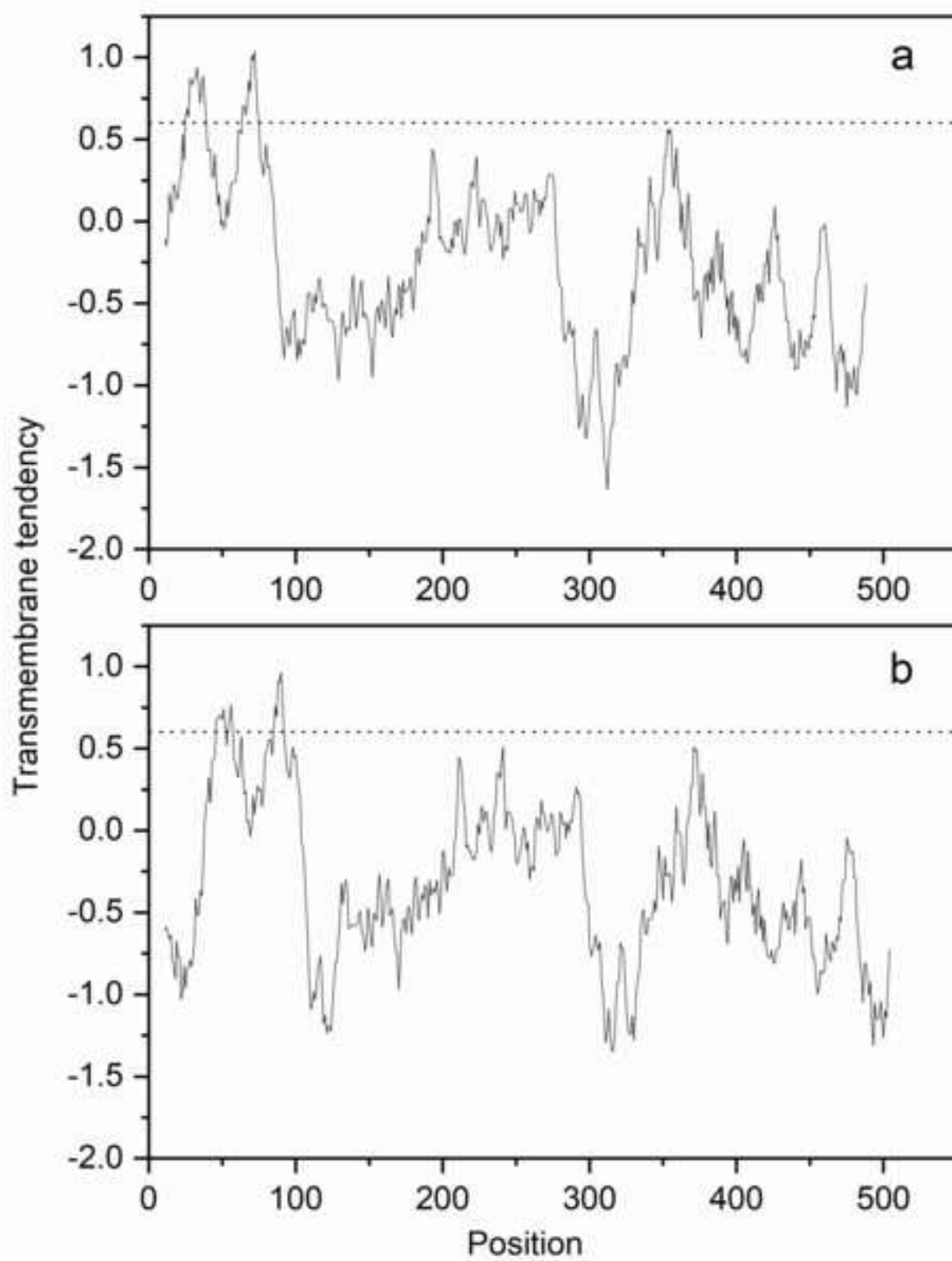
Figure 5. Front (a), lateral (b) and rear (c) views of malonyl-CoA (yellow) and arachidic acid (green) docking in the *HaKCS2* substrate binding pocket, with the key residues (triad C, H and N) at the catalytic centre in red. *HaKCS2* substrate docking was modelled using the iterative threading assembly refinement (I-TASSER) method (Yang et al. 2015).

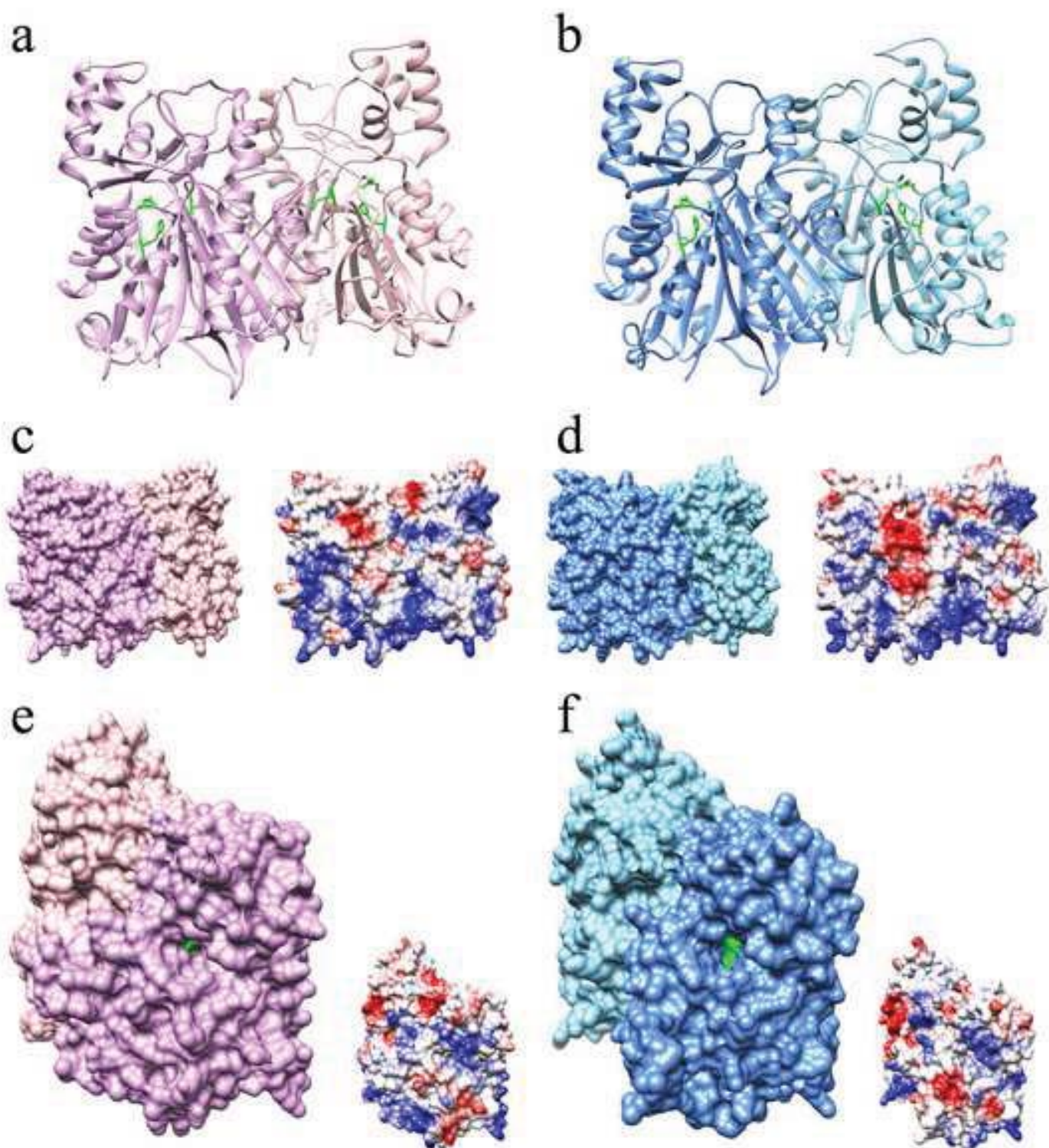
1
2
3
4
5
6
7
8
9
10
Figure 6. Expression of genes encoding *HaKCS1* and *HaKCS2* in developing seeds and vegetative tissues of the common CAS-6 sunflower line. Expression was determined by RT-QPCR using the actin gene from *H. annuus* (GenBank FJ487620) as the reference gene. The data correspond to the mean \pm SD from three independent measurements.

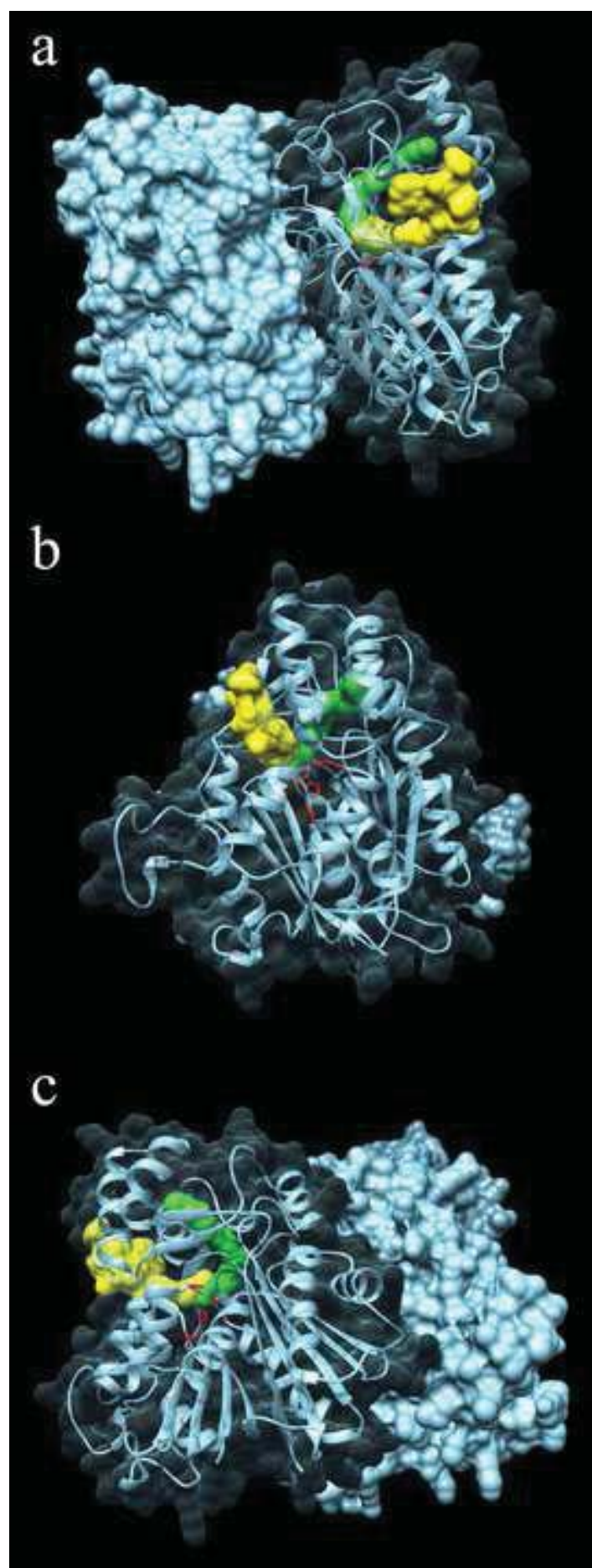
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Figure 7. Complementation of the *elo2 Δ Trp elo3 Δ Trp* *Saccharomyces cerevisiae* mutant with *HaKCS1* and *HaKCS2* from sunflower, using the *KCS9* (**At2g16280**) gene from *Arabidopsis* as a positive control.

66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
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 \pm SD of three independent measurements.









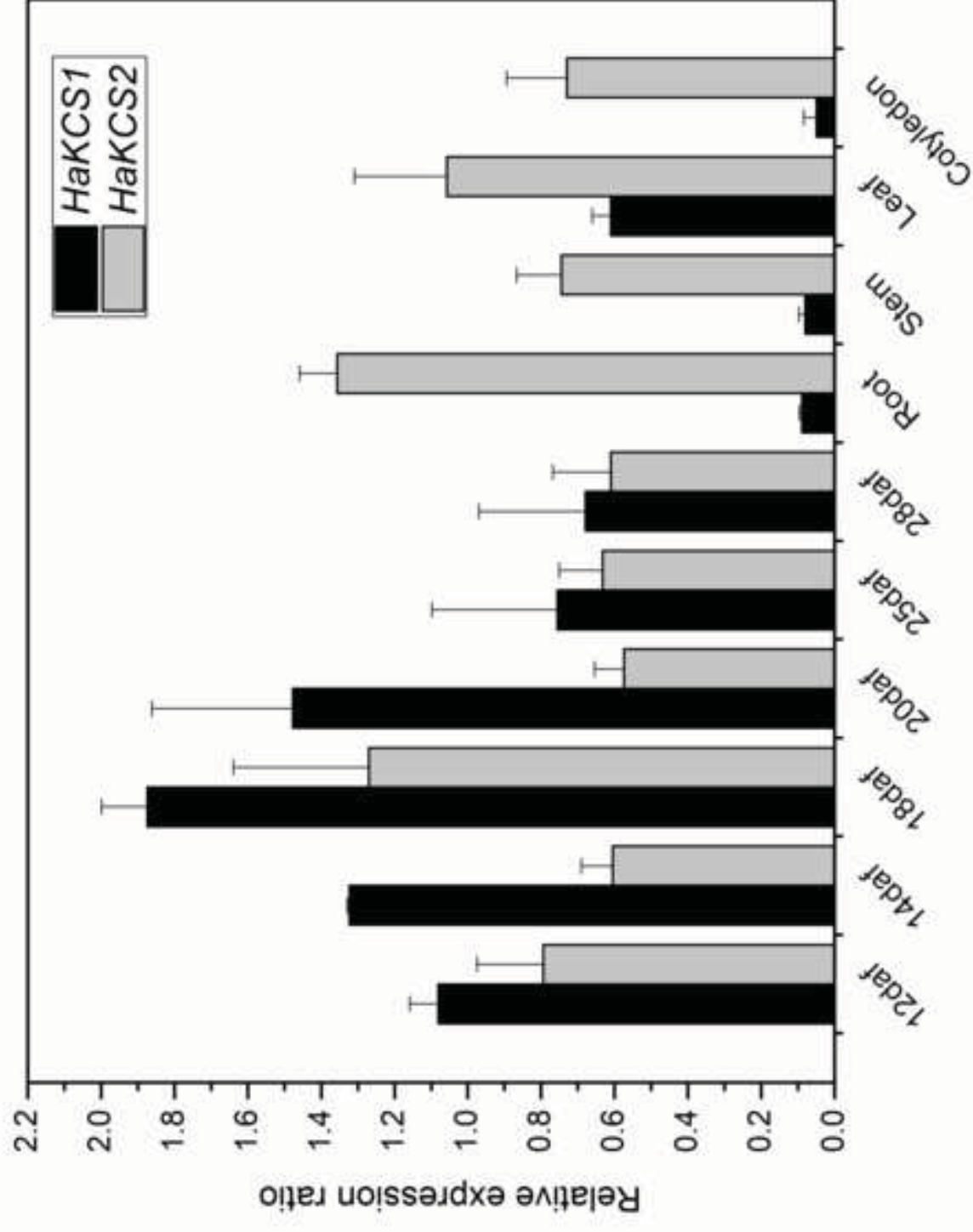


Figure 6

pADH-LEU



**pADH-LEU
::HaKCS2**

**pADH-LEU
::AtKCS9**

SC+FOA+Ura

pADH-LEU::HaKCS1

