



Classification and substrate head-group specificity of membrane fatty acid desaturases

Dongdi Li^a, Ruth Moorman^a, Thomas Vanhercke^b, James Petrie^b, Surinder Singh^b, Colin J. Jackson^{a,*}

^a Research School of Chemistry, Australian National University, Canberra, Australia

^b CSIRO Agriculture and Food, Canberra, Australia

ARTICLE INFO

Article history:

Received 30 June 2016

Received in revised form 29 August 2016

Accepted 30 August 2016

Available online 12 September 2016

Keywords:

Membrane-bound fatty acid desaturases

Lipid-substrate specificity

Sequence similarity networks (SSNs)

ABSTRACT

Membrane fatty acid desaturases are a diverse superfamily of enzymes that catalyze the introduction of double bonds into fatty acids. They are essential in a range of metabolic processes, such as the production of omega-3 fatty acids. However, our structure–function understanding of this superfamily is still developing and their range of activities and substrate specificities are broad, and often overlapping, which has made their systematic characterization challenging. A central issue with characterizing these proteins has been the lack of a structural model, which has been overcome with the recent publication of the crystal structures of two mammalian fatty acid desaturases. In this work, we have used sequence similarity networks to investigate the similarity among over 5000 related membrane fatty acid desaturase sequences, leading to a detailed classification of the superfamily, families and subfamilies with regard to their function and substrate head-group specificity. This work will facilitate rapid prediction of the function and specificity of new and existing sequences, as well as forming a basis for future efforts to manipulate the substrate specificity of these proteins for biotechnology applications.

Crown Copyright © 2016 Li et al.. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In contrast to soluble fatty acid desaturases (FADs), which are acyl-carrier protein (acyl-ACP) specific [1,2], membrane FADs are a diverse family of proteins that display a range of lipid substrate preferences including acyl-CoAs, sphingolipids (SP), phospholipids (PL) and galactolipids (GL) [3,4]. The acyl chains of these substrates can, in general, be quite similar, but the “head-groups” of these lipids differ and contribute to the different physiological roles (Fig. 1). For example, acyl-ACPs are important intermediates for a number of different metabolic pathways, including lipid biosynthesis [5], whereas PL and GL are not only important structural lipids, but also play essential roles in different cellular signalling pathways [6]. Monogalactosyl diacylglycerol (MGDG) is a type of non-phosphorous structural glycerolipid that is abundant in the cellular membranes of photosynthetic organisms [7–9]. Similarly, SP is also a class of structural lipids, although its roles in the regulation of cellular processes, such as apoptosis, cell migration and cold tolerance, makes this class of lipid particularly important [10–13]. Acyl-CoA substrates are frequently acted upon by lipid-modifying enzymes such as desaturases, acyltransferases and elongases [14–19]. Therefore, acyl-CoA substrates are especially important for

the production of unusual lipids such as docosahexaenoic acid and eicosapentaenoic acid [14,20].

The substrate specificity and regioselectivity (double bond positioning) of membrane FADs is largely determined by the interaction between the enzyme and the lipid head-group. The recently published human and mouse stearyl-CoA desaturase structures reveal the interaction between the lipid substrate and the membrane FADs [21,22]. The hydrophilic CoA head-group of the substrate forms electrostatic interactions and hydrogen bonds with residues in the cytoplasmic domain and transmembrane helix (TM) 1 of the desaturase, and orients the acyl group into the long hydrophobic tunnel with the target carbon presented at di-metal active site [22]. The acyl group is surrounded by hydrophobic residues in the substrate binding cavity including W262 on TM4, which holds the substrate in place for $\Delta 9$ desaturation [22]. However, the substrate binding mechanisms of the desaturases that are not specific to acyl-CoA are still poorly understood. A domain-swapping study between the acyl-phosphatidylcholine (acyl-PC) specific $\Delta 6$ desaturase and the sphingolipid specific $\Delta 8$ desaturase from *Borage officinalis* discovered that TM helices, as well as the cytosolic loops, of the desaturases play essential roles in differentiating between head-groups [23]. However, there is not enough structural data to either identify specific roles for the functionally important residues in substrate binding, or to describe an accurate mechanism for the substrate binding of the acyl-PC-specific and sphingolipid-specific FADs. Moreover, there is currently no established sequence-based classification available that

* Corresponding author.

E-mail address: colin.jackson@anu.edu.au (C.J. Jackson).

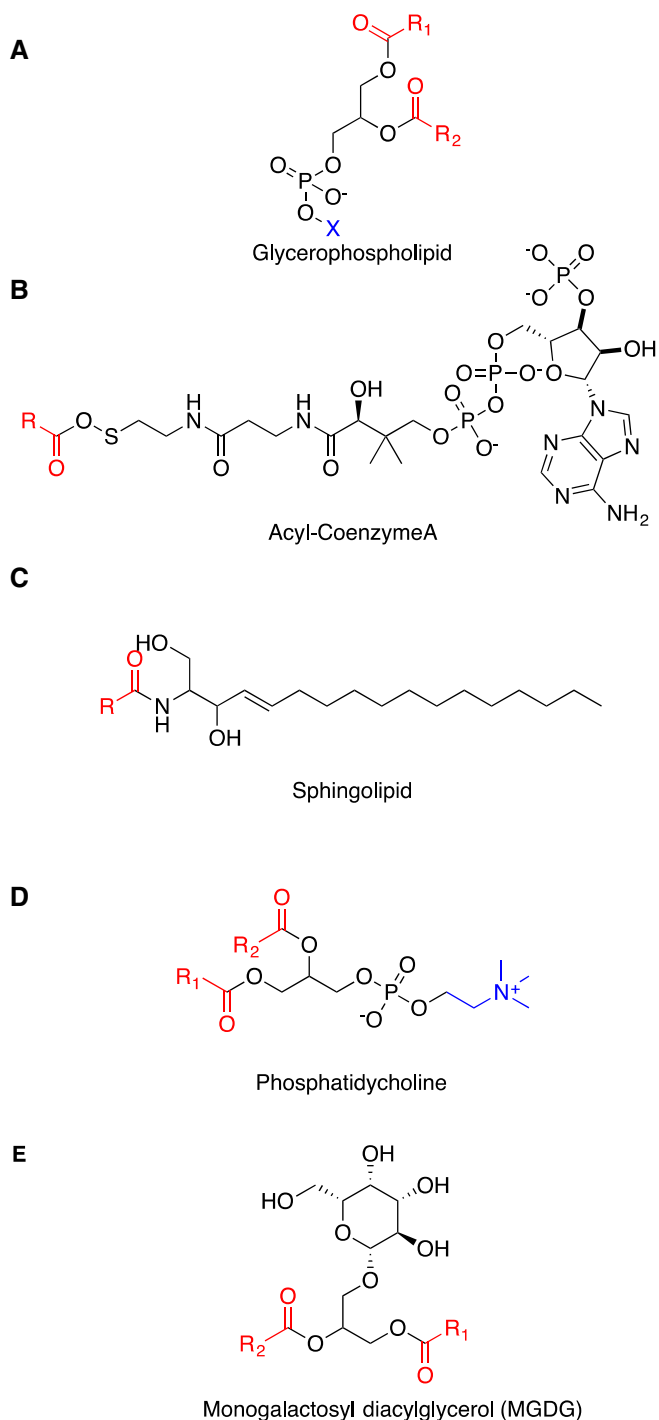


Fig. 1. The chemical structures of glycerophospholipid (A), acyl-Coenzyme A (B), sphingolipid (SP), phosphatidylcholine as a major phospholipid (PC) and monogalactosyldiacylglycerol (MGDG). The acyl groups are coloured red. The X head-group on the glycerophospholipid backbone varies between different classes with the structure of PC given as an example.

accurately predicts head-group specificity and regioselectivity of uncharacterized sequences.

Based on their regiospecificity and sequence homology, the membrane desaturase superfamily has been further divided into families [3,24,25]. Front-end (FE) desaturases are capable of introducing carbon-carbon double bonds into unsaturated lipid substrates at positions between an existing double bond and the terminal carboxyl group [4]. It was found that the varied substrate preference of the front-end desaturases is the root cause of the substrate dichotomy bottleneck in

the ω 3-LCPUFAs biosynthesis pathway [20,63,64]. This is because some Δ 6 desaturases utilize acyl-PC substrates, whereas the elongases at the next step utilize acyl-CoA substrates [20]. Thus, for acyl-PC-specific Δ 6 desaturases, the product of the Δ 6 desaturation has to be converted into an acyl-CoA molecule by an acyltransferase, which limits the metabolic flux in the recombinant ω 3-LCPUFA biosynthesis pathway in plants [20]. Second, first desaturases (FDs), typically catalyze the formation of the first C=C bond at the Δ 9 position of an acyl group. It has been postulated that the Δ 9 desaturases are the most ancient desaturases among the three groups because of their universal distribution in organisms [24]. The FDs utilize acyl-CoA as substrates except for the FDs of plant plastids [8]. Third, the methyl-end (ME) desaturases are responsible for introducing a C=C between a pre-existing C=C and the methyl end of the acyl group [26]. ME desaturases are known to utilize PL substrates [27,28]. Finally, the Δ 4 sphingolipid desaturases (Δ 4-SPs) are particularly important for cellular signalling [29]. This classification was based on a previous limited phylogenetic analysis of eukaryotic membrane-bound desaturases [25].

For a very diverse protein family, such as the membrane FADs, constructing a high-quality multiple sequence alignment can be challenging [30], which limits our ability to obtain evolutionary information from phylogenetic analysis and to annotate the possible functions for identified genes with confidence [30]. Protein sequence similarity networks (SSNs) were developed in 2009 by Babbitt and co-workers to facilitate functional annotation based on known sequence data [31]. SSNs can illustrate the global sequence-structure-function diversity of protein superfamilies [32–34], because they are based on many pairwise alignments of proteins instead of the entire alignment of a large protein datasets, so that the inaccuracy of a large protein alignment is largely eliminated. A SSN presents the level of similarity between members in a protein family in a graphical way in which proteins sharing high sequence identities are clustered. In the SSNs, the edges represent all-vs-all BLAST *E*-values and each node represents a protein or a group of highly similar proteins. The information about their organism of origin and structural understanding can be mapped and visualized in the network manually or automatically [35]. This method can be combined with other computational or biochemical techniques to annotate and characterize the function-structure relationships of protein subfamilies. Two example applications of the use of SSNs are the re-classifications of the cytosolic glutathione transferases superfamily [36], and an oxidoreductase superfamily [30], where biochemical characterization is incorporated to provide a more in-depth analysis of the functional and structural diversity of protein subfamilies. Thus, SSNs can serve as the foundation for the characterization of protein superfamilies by providing information about the sequence-structure-function divergence [37].

In this work, membrane FADs were analysed using SSNs to determine the distinctions in sequence and structure to provide a clearer picture of the differences in the possible mechanisms of substrate binding. Because of the high commercial value of the FE desaturases in the production of poly-unsaturated fatty acids, such as ω 3-LCPUFAs [14], a more in-depth analysis of this family was performed. These results provide new directions for the future engineering membrane FADs.

2. Results and discussion

2.1. The sequence-structure-function relationships of all known membrane-bound desaturases

In order to characterize the sequence and functional diversity in the membrane FADs, particularly the substrate head-group specificity, 5245 sequences were collected using the PFAM fatty acid desaturase family PF00487 as seed sequence clusters within the length range of 350 aa to 550 aa. This length range was chosen to limit the search to the single domain desaturases and the desaturases with a fused cytochrome b5 domain. The free cytochrome b5 proteins and the unrelated long

cytochrome b5 domain containing fusion proteins are eliminated with this filter. The collected sequences were analysed by generating a SSN using EFI-EST [35], where an initial network containing 2878 representative nodes (clusters of protein sequences with >60% amino acid identity) was produced. The edges represent all-vs-all BLAST *E*-values between the clusters. The cytochrome b5 domain was only included in the *E*-value calculations when all of the protein members in the connecting nodes also contained cytochrome b5 domains. Otherwise,

only the fatty acid desaturase sequences were used to calculate the *E*-values. A literature review was performed to collect the known lipid head-group preferences of the characterized desaturases and their organism distributions. This data was mapped onto the networks (Fig. 2).

The SSNs in Fig. 2 are presented with different colour coding of the nodes to identify the kingdom in which the respective membrane FADs are found and the substrate head-group preference. It is apparent that there are a large number of uncharacterized prokaryotic

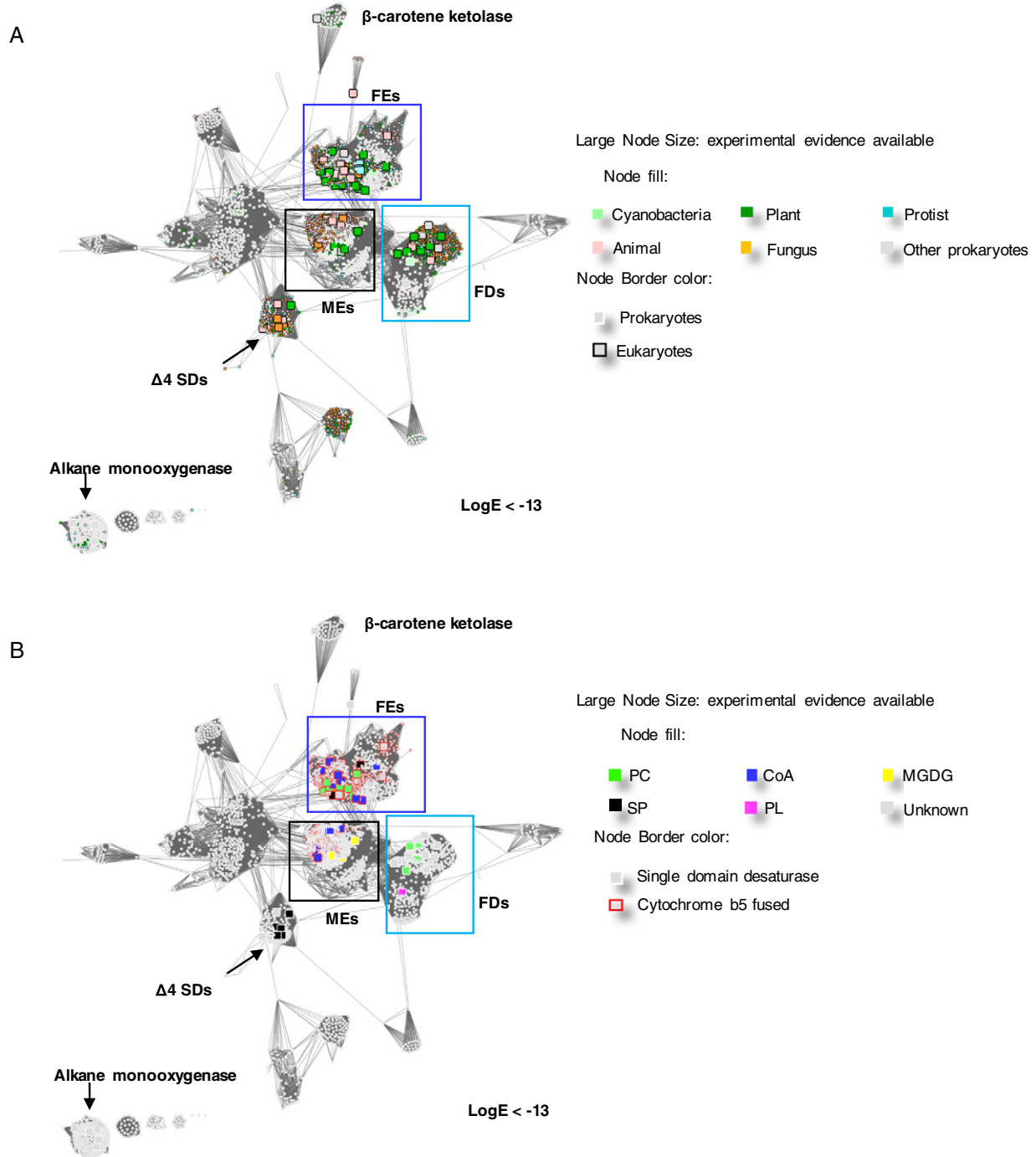


Fig. 2. Overview of the sequence similarity relationships in the membrane-bound desaturase family. These representative networks show 2878 nodes representing the 5245 proteins in the membrane-bound desaturase family, in the Pfam database (pfam.xfam.org) (the list of the UniProt IDs of the sequences included in this network are provided in Supplementary Table S1). The clusters with functionally characterized member sequences are labelled, including the functionally dissimilar alkane monooxygenases cluster and beta-carotene ketolases clusters. There are three major fatty acid desaturase families (first desaturases, FDs; methyl-end desaturases, MEs; front-end desaturases, FEs). Δ4 sphingolipid desaturases (SDs) are also shown. Larger squares represent nodes with at least one functionally characterized member. Edges or lines connecting the nodes are shown if the pairwise similarity score between the sequences of representative nodes is lower than the threshold of Log BLAST *E*-value of -13 . The 366,864 edges in the overview network have a median sequence identity of 30% over 275 residues. A and B are identical networks with different colour coding as explained in the figure keys. A. The nodes in this overview network are coloured by the kingdom of the organism of origin. B. The nodes in this overview network are coloured by the substrate head-group specificity of the characterized members.

desaturase-like proteins in the database. The functionally distinct beta-carotene ketolases and alkane monooxygenases are also present in the network, because they all carry the three conserved and functionally important histidine-rich motifs, which coordinate the catalytic di-iron centres of these enzymes [38–40]. However, they are sufficiently different in terms of sequence identity and function to be excluded from this membrane FAD-focused study. Four major clusters of desaturases with characterized members are visualized in the SSNs. By examining the characterized members in each major cluster, the four clusters are identified as the first desaturases (FDs) [41,42], methyl-end desaturases (MEs) [43], front-end desaturases (FEs) [4] and $\Delta 4$ sphingolipid desaturases ($\Delta 4$ SPs) [44]. The different clusters include multiple sequences from different evolutionary kingdoms as well as a range of different substrate preferences (Fig. 2). This coarse separation of the clusters is consistent with the previously proposed membrane FAD classification, which was based on the phylogenetic analysis of eukaryotic membrane-bound desaturases [25]. The $\Delta 4$ sphingolipid desaturases catalyze the $\Delta 4$ desaturation of important signalling molecules such as (E)-sphing-4-enine-1-phosphate, which is the messenger for the epidermal growth factor (EGF) receptor family [29]. As $\Delta 4$ s are not known to be specific for substrates other than sphingolipids, they were not subjected to further detailed analysis in this study.

2.2. Substrate specificity in the first desaturase family is determined by the presence of charged residues within the substrate binding site

The three main families within the membrane FAD superfamily (FD, ME, FE) can be further resolved by reducing the LogE filter to remove weak associations. The FD family contains the structurally characterized mammalian $\Delta 9$ stearoyl-CoA desaturases (SCD1s) [41], which share approximately 32% amino acid identity with other members of the FD subfamily. They are responsible for introducing the first C=C bond into a saturated hydrocarbon chain [21]. At a logE value of -56 , eight major

clusters are formed (Fig. 3). The largest cluster consists of acyl-CoA specific $\Delta 9$ FDs (FD-A). The only other cluster that contains experimentally characterized desaturases is the FD-C cluster, which contain the characterized bi-functional acyl-lipid-specific *Arabidopsis thaliana* $\Delta 7/\Delta 9$ desaturases (ADS1) [8].

The FD-A cluster consists of the acyl-CoA $\Delta 9$ single domain FADs, predominantly from animals, and fusion desaturases that have a cytochrome b5 domain fused at the C-terminus of the desaturase, predominantly from fungi (Fig. 3). Two crystal structures of single-domain animal desaturases have been published and have defined the geometry of the acyl-CoA as bound in the desaturase [21,22]. However, there is no crystal structure of the fused fungal desaturases to date for structural comparison. Given that there is significant amino acid sequence identity (higher than 30%) in the desaturase domain between the fused and unfused proteins, the former most likely exhibit similar binding characteristics as the single domain desaturases.

The FD-C cluster is composed of plant plastid FADs and prokaryotic FADs, including the bi-functional ADS1 protein, which revealed an intriguing subcellular-specific substrate-specificity and regioselectivity [8]. This desaturase is capable of catalysing $\Delta 9$ desaturation of acyl-PC, when targeted to endoplasmic reticulum (ER) membrane [8]. It can also catalyze $\Delta 7$ desaturation of acyl-groups presented on MGDG when targeted to the plastid membrane [8]. However, because most ADS enzymes (with the exception of the plastid-localized ADS3) are localized to the ER, their native function is likely to involve desaturation of acyl-PC. As the structures of MGDG and acyl-PC are only different by the hydrophilic functional head-groups at the sn3 position of the glycerol backbone (Fig. 1), overlapping substrate specificity is not entirely surprising.

Although both FD-A and FD-C clusters belong to the FD family, they have diverged in terms of head-group specificity. When the sequence differences between the FD-A human SCD1 and the FD-C ADS1 and are compared, a number of non-conservative substitutions are evident

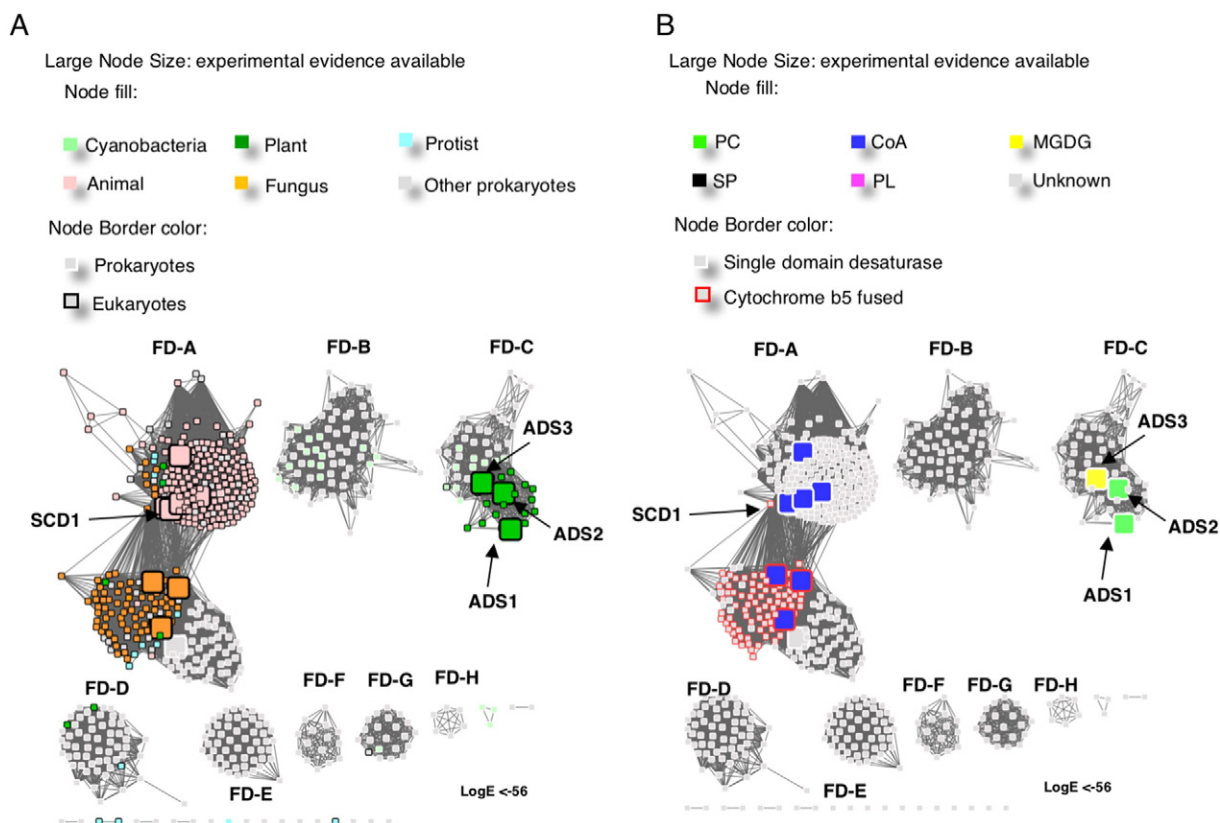


Fig. 3. The representative networks of the FDs show more detailed subgroupings. A and B are the same networks generated from the FDs cluster in Fig. 2 at a higher stringency, LogE < -56 . Eight major clusters (clusters with more than 50 member sequences in each) formed. The nodes are coloured by organism kingdom information (A) or their substrate specificities (B).

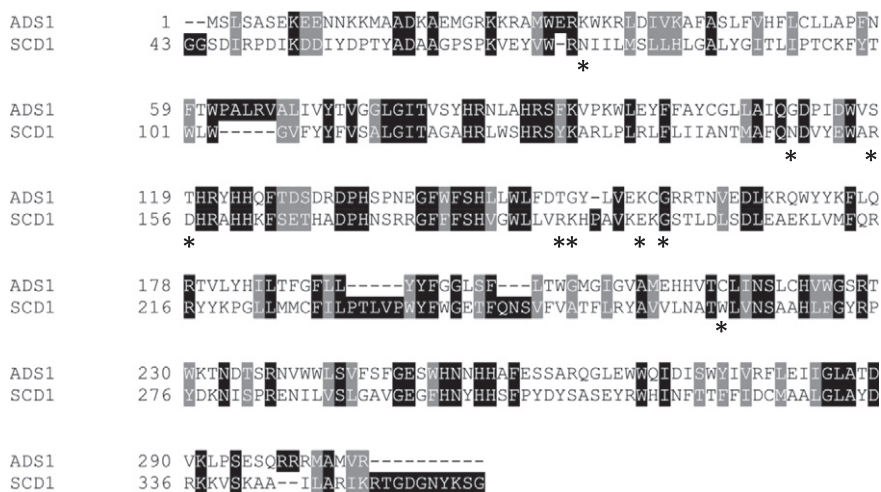


Fig. 4. The alignment of ADS1 and SCD1. The substrate head-group binding residues of the crystal structure of SCD1 [22] are denoted by asterisks.

(Fig. 4). Specifically, a number of charged or polar residues that are responsible for CoA binding in SCD1 are replaced by uncharged amino acids in ADS1, which result in significant changes in the surface charge distribution in the substrate binding cavity (Fig. 5). The loss of positive charge, which is complementary to the extensive negative charge on coenzyme-A but is not required for binding of the glycerol-sugar moiety of MGDG, is consistent with the change in specificity.

2.3. Methyl-end desaturases specific for phospholipid substrates

The methyl-end desaturase subfamily can be separated into two groups at the LogE filter of -30 : a prokaryotic $\Delta 5$ phospholipid FAD group ($\Delta 5$ PLs) and the canonical MEs including $\Delta 12$ -specific [45], $\Delta 15$ -specific [46] and bifunctional ($\Delta 12$ and $\Delta 15$) FADs [47] (Fig. 6). The $\Delta 5$ PL desaturase from *Bacillus subtilis* (UniProt ID: O34653) has been experimentally proven to have a six-transmembrane (TM) helix topology [27,28]. Functionally, the $\Delta 5$ PL cluster is similar to the first desaturases, although it shares higher sequence similarity with the ME family. Given the high amino acid sequence identity between these clusters (higher than 21%), the 6-TM-helix topology is likely to be the common topology of the MEs cluster and makes this subfamily structurally distinct from the FDs. Even though the bacterial *B. subtilis* $\Delta 5$ desaturase is suggested to be specific for phospholipid, this study did not specify the class of phospholipid that is the primary substrate [27, 28]. Several enzymes from the ME cluster have been shown to catalyze

the desaturation of phosphatidylcholine including the plant FAD2 FAD3, FAD6, FAD7 and FAD8 desaturases (Fig. 6) [48–51]. Thus, it is likely that both clusters utilize phospholipid substrates. It is notable that the $\Delta 5$ phospholipid FAD group consists almost entirely of prokaryotic and cyanobacterial sequences (with a handful of sequences from plants and protists), whereas the canonical ME cluster consists of sequences from fungi, animals, plants, cyanobacteria and other prokaryotes.

2.4. Front-end desaturases have diverse substrate specificity

The FE desaturases are responsible for introducing carbon–carbon double bonds into unsaturated acyl chains between the pre-existing double bonds and the carboxyl group of lipid substrates [52]. There is currently no crystal structure of any member of the FE desaturase family. Thus, our understanding of their mechanism and regioselectivity is primarily based on mutagenesis studies, which have been reviewed by Meesapyodsuk and Qiu [4]. A range of FEs with $\Delta 4$, $\Delta 5$, $\Delta 6$ desaturases activity and specificity for acyl-CoA or acyl-PC substrates have been identified [4,20,53–57]. This level of diversity is not seen in the other FAD families, making FE desaturases distinct from MEs and FDs. With a LogE < -20 filter, the broad FE desaturase family separated into four clusters. For ease of explanation, the clades are named as FE1–4 (Fig. 7). The fusion-FADs, which have a cytochrome b5 domain fused at the N-terminus of the desaturase domain, are only found in the FE1–3 clusters.

Clusters 1 and 2 of the FE desaturases are the largest and both include sequences from prokaryotes and eukaryotes, although only the eukaryotic proteins have fused cytochrome b5 domains at the N-termini. This observation suggests that the gene fusion event likely took place after the evolution of eukaryotes. The FE1 cluster includes the $\Delta 6$ desaturases from *M. pusilla* and *O. tauri*, which have been shown to function as desaturases with omega-3 and omega-6 fatty acids [14,58]. A single prokaryotic fusion desaturase from the cyanobacterium *Leptolyngbya* sp. PCC 7375 (UniProt ID: K9ES17) falls within those found in the FE2 cluster. It is unclear if K9ES17 is the result of a horizontal gene transfer event between eukaryotes and prokaryotes, or if it is the descendent of an ancient gene fusion that was the evolutionary origin of the eukaryotic fusion proteins. The smallest cluster (FE3) includes proteins that are predominantly found in insects, with the exception of a few proteins from algae (*Nannochloropsis gaditana*) and simple eukaryotes and are believed to be housekeeping genes involved in lipid metabolism [59]. Finally, the FE4 cluster is primarily composed of single domain (non-fused) desaturases from prokaryotic species, but also includes some genes from eukaryotes that encode single domain proteins. The only characterized protein in this family is the NADPH

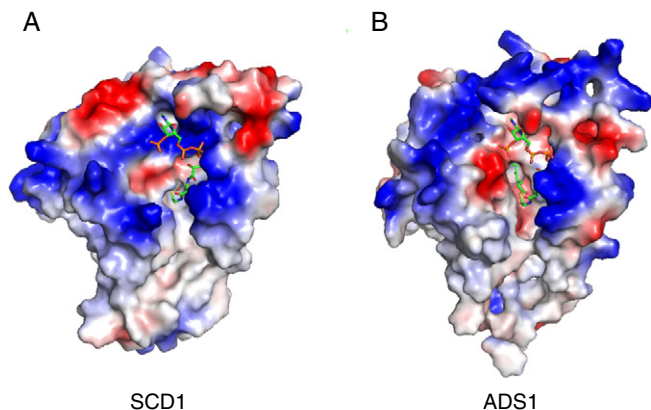


Fig. 5. The surface charge distribution comparison between an acyl-CoA-specific FD-A and an MGDG-specific FD-C. The difference in surface charge distributions visualized on the surface of the crystal structure of human stearyl-CoA $\Delta 9$ desaturase (SCD1, PDB ID: 4ZYO) and the homology model of ADS1 inside and around the substrate-binding cavity.

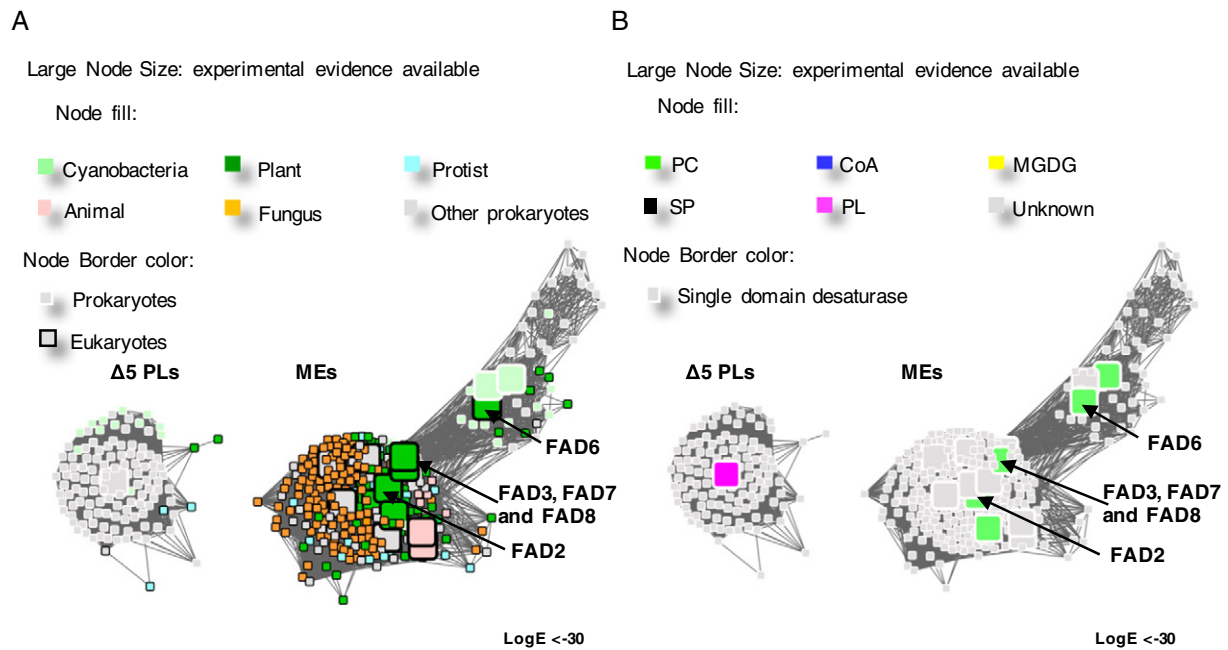


Fig. 6. The representative network of the MEs cluster in Fig. 2 show more detailed subgroupings. The plant FAD2, FAD3, FAD6, FAD7 and FAD8 desaturases are denoted. A and B are the same networks generated from the MEs cluster in Fig. 2 at a higher stringency, $\text{LogE} < -56$. The prokaryotic $\Delta 5$ PLs formed a separate cluster. The nodes are coloured by organism kingdom information (A) and their substrate specificities reported in the literature (B).

oxidoreductase-dependent C16:0/C18:0-CoA $\Delta 9$ desaturase from *Mycobacterium* (UniProt ID: P9WNZ3) [60]. Its function mimics the role of FDs as this enzyme is responsible for introducing the first double bond into fatty acid chains.

Increasing the LogE filter to < -65 can further differentiate the FE1 cluster, revealing the presence of clusters with distinct lipid head-

group preferences including a sphingolipids/acyl-PC-specific cluster (FE1-S), a predominantly acyl-PC-specific cluster (FE1-PC) and acyl-CoA-specific proteins (FE1-AC) (Fig. 8). The separation of sequences within the FE1 sub-cluster is supported by a phylogenetic analysis of the functionally characterized FE desaturases (Fig. 9). The FE1-AC sub-cluster includes the animal $\Delta 5$ and $\Delta 6$ acyl-CoA desaturases, as well as

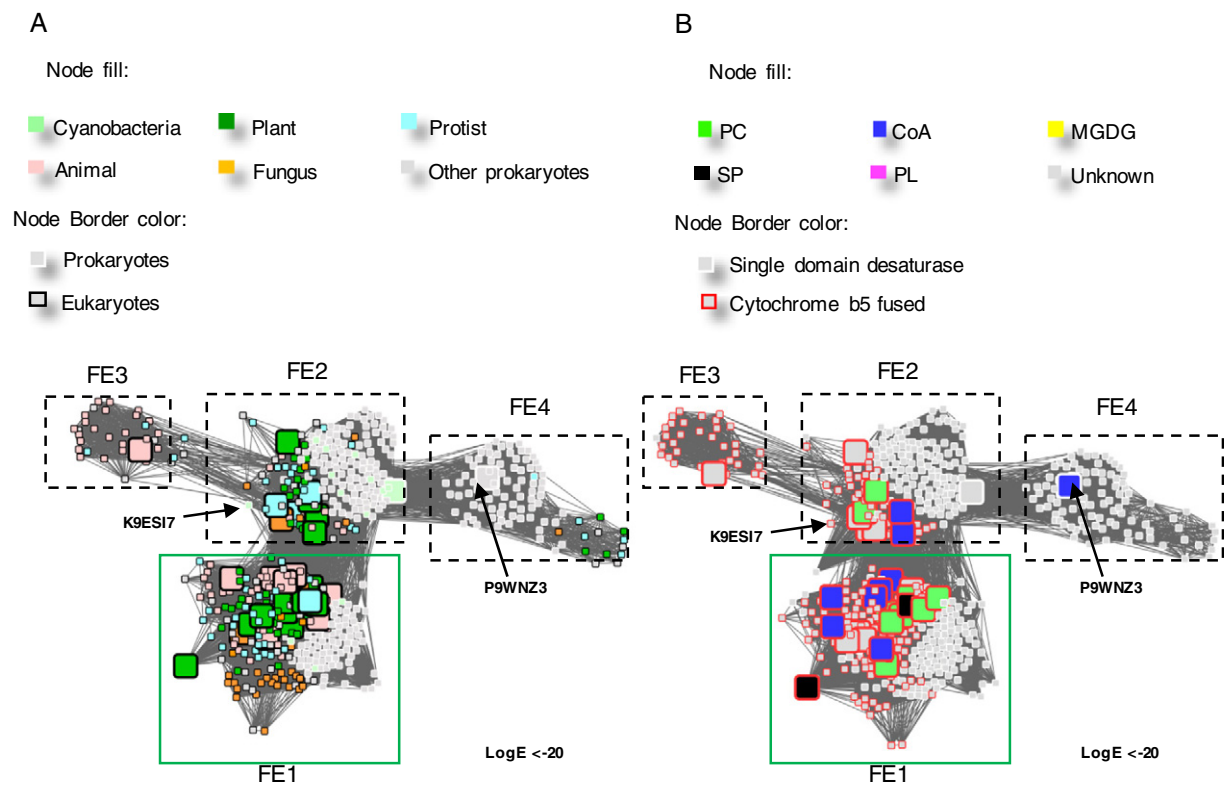


Fig. 7. Separation of the FE desaturase family into smaller groups. A and B are the same networks generated from the FEs cluster in Fig. 2 at a higher stringency, $\text{LogE} < -20$. The nodes are coloured by organism kingdom information (A) and their substrate specificities reported in the literature (B).

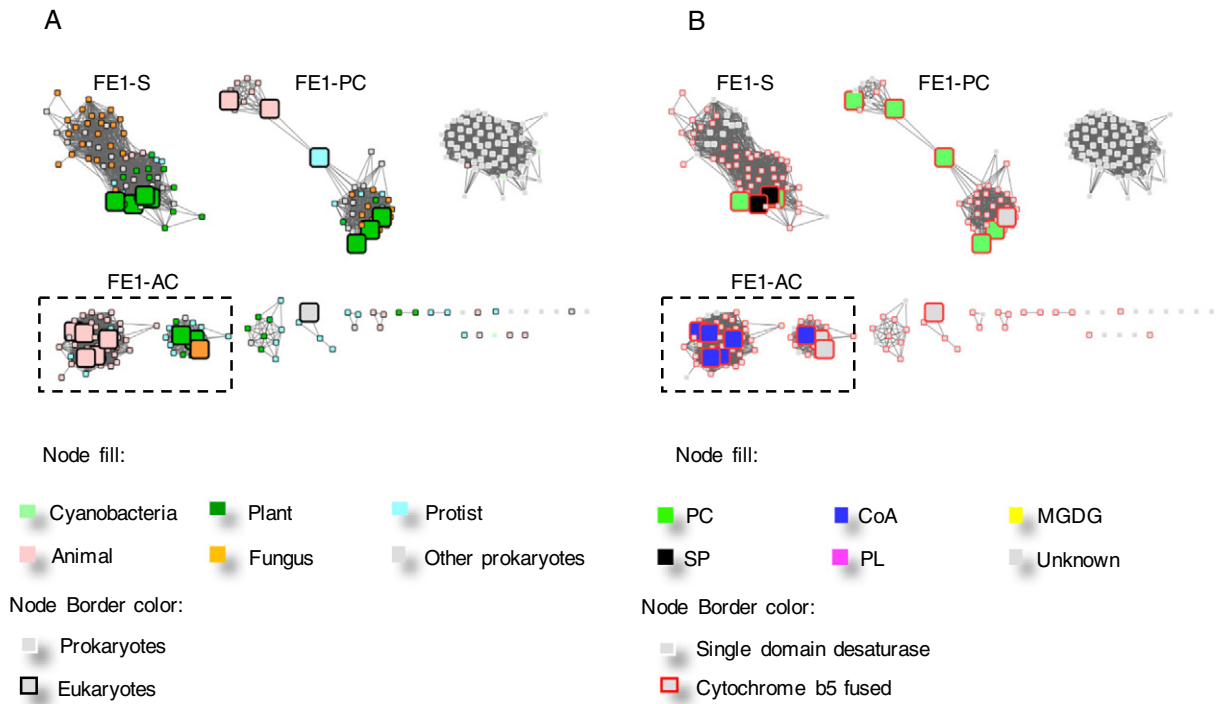


Fig. 8. The representative sub-clusters within the FE1 cluster. A and B are the same networks generated from the FE1 cluster in Fig. 2 at a higher stringency, $\text{LogE} < -65$. The nodes are coloured by organism kingdom information (A) and their substrate specificities reported in the literature (B).

the algal and fungal acyl-CoA desaturases. The FE1-PC sub-cluster includes genes from *C. elegans*, moss and liverworts. Within this sub-cluster, *Marchantia polymorpha* $\Delta 6$ desaturase (UniProt ID: Q696V8) has been shown to have some promiscuous activity with acyl-CoA

substrates in addition to acyl-PC substrates [61]. The FE1-S sub-cluster is complex and difficult to resolve at a $\text{LogE} < -65$ filter, and includes both acyl-PC and acyl-SP desaturases. Within the FE1-S sub-cluster, higher plant FE desaturases cluster with a group of fungal proteins.

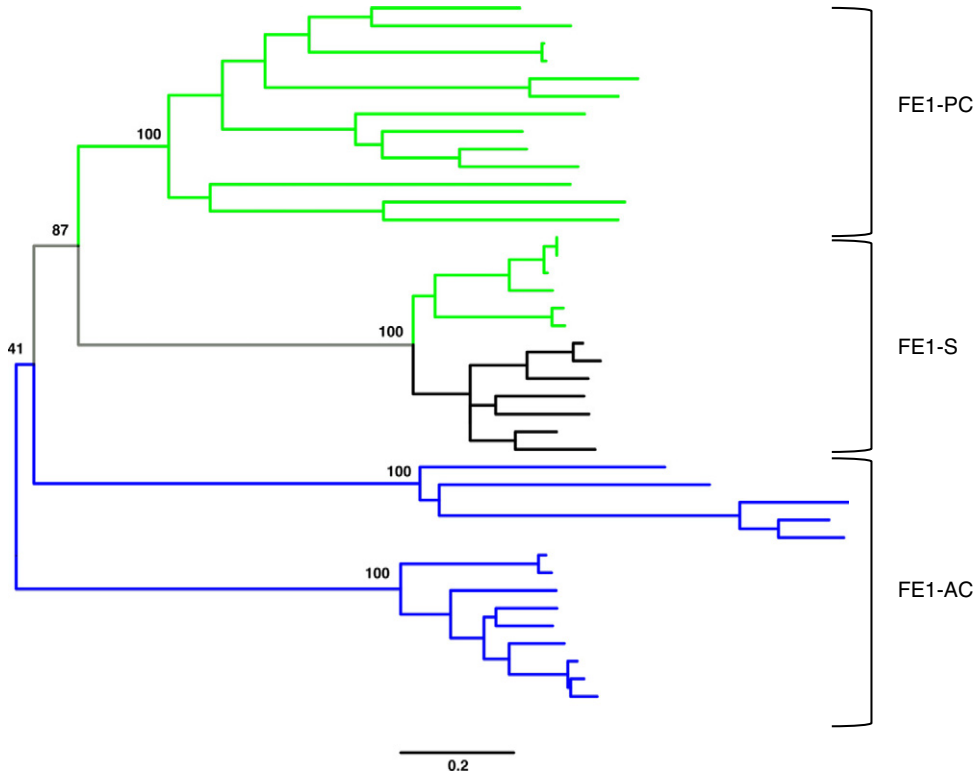


Fig. 9. Phylogenetic analysis of the FE1 sub-cluster. The tree was constructed using MEGA v.7. A list of the included desaturase proteins is detailed in the Supplementary Materials. The bootstrap values of 100 replicates are denoted at the major nodes. The green branches indicate acyl-PC specificity. The black branches indicate sphingolipid specificity. The blue branches indicate acyl-CoA specificity.

The topology of the FE cluster has not been experimentally confirmed, which makes it difficult to determine the specific substrate-binding mode of these desaturases. It is possible that the modes of interaction between the proteins of the FE cluster and the different types of lipid substrates could also be distinctive from the observed binding mechanism in SCD1 [22]. The recently published structure of the yeast integral membrane fatty acid α -hydroxylase (scScs7p) [62] could indicate an alternative lipid–substrate binding mechanism that might be relevant to these enzymes. However, given that it only shares very low (~14%) amino acid sequence identity with members of the FE cluster, we think that the structures are likely to be too different to allow the inference of alternative substrate-binding modes.

3. Conclusion

This broad sequence–structure survey of the membrane FAD superfamily provides several notable observations. First, there are a large number of uncharacterized FADs that are essentially distinct from all known clusters and these proteins are abundant in prokaryotes. Further work to characterize these abundant proteins will be necessary to identify their evidently important physiological roles. Second, the membrane FAD superfamily is diverse, with the divergence of FE, ME, FD families, as well as a number of smaller and more specialized families, such as the $\Delta 4$ sphingolipid desaturases. This evolutionary separation of proteins on the basis of their function and regioselectivity (first, methyl end, or front end) produces a number of sub-clusters in the FD and FE families that have specificity for fatty acids with different head-groups. The FE desaturases, for example, have evolved to become specific for acyl-CoA, acyl-PC and sphingolipid substrates. In this work, we have classified this diverse superfamily for the first time in detail. This figure also highlights the gaps in our current understanding—for instance, what are the functions of the FD2 and FD4 groups?

4. Materials and methods

4.1. Sequence similarity networks (SSNs)

SSNs were generated by EFI-EFT using sequences belonging to the PF00478 fatty acid desaturase superfamily in PFAM database, in which the BLAST *all-vs-all* LogE values were used as the edges and a LogE cutoff of -5 was applied to the initial network generation [35,63]. The network consists of the nodes representing protein clusters with 60% sequence identity which were visualized using Cytoscape v.3.2.1 [64]. The sequences shorter than 250 aa or longer than 550 aa were excluded to avoid partial proteins or proteins with more than two domains. The network was curated to remove unrelated sequences. The subgroupings in each major cluster were visualized by gradually increasing the stringency of the LogE filter of the networks. The published functional data was used to determine the consensus function and substrate preference of each subfamily.

4.2. Phylogenetic analysis

The FE desaturases with experimentally proven functional data were collected from the literature. A total of 38 sequences (Supplementary Table S2) belonging to FE1 cluster were aligned by Molecular Evolutionary Genetics Analysis program Version 7 (MEGA7) using MUSCLE algorithm [65,66]. A phylogenetic tree was constructed using the maximum-likelihood method and LG matrix in (MEGA7) [65].

4.3. Structural modelling

The protein sequence of the *Arabidopsis thaliana* $\Delta 9$ desaturase (ADS1, UniProt ID: O65797) was submitted to the Phyre2 modelling

server [67] using the default settings of the "normal" modelling mode, with human stearoyl-CoA desaturase (PDB: 4ZY0) as template.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.csbj.2016.08.003>.

References

- [1] Fox BG, Shanklin J, Somerville C, Munck E. Stearoyl-acyl carrier protein delta 9 desaturase from *Ricinus communis* is a diiron-oxo protein. *Proc Natl Acad Sci U S A* 1993;90:2486–90.
- [2] Cahoon EB, Lindqvist Y, Schneider G, Shanklin J. Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proc Natl Acad Sci U S A* 1997;94:4872–7.
- [3] Sperling P, Ternes P, Zank TK, Heinz E. The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids* 2003;68:73–95.
- [4] Meesapyodsuk D, Qiu X. The front-end desaturase: structure, function, evolution and biotechnological use. *Lipids* 2012;47:227–37.
- [5] Byers DM, Gong H. Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family. *Biochem Cell Biol* 2007;85:649–62.
- [6] Ghosh S, Strum JC, Bell RM. Lipid biochemistry: functions of glycerolipids and sphingolipids in cellular signaling. *FASEB J* 1997;11:45–50.
- [7] Zauner S, Jochum W, Bigorowski T, Benning C. A cytochrome b5-containing plastid-located fatty acid desaturase from *Chlamydomonas reinhardtii*. *Eukaryot Cell* 2012;11:856–63.
- [8] Heilmann I, Pidkovich MS, Girke T, Shanklin J. Switching desaturase enzyme specificity by alternate subcellular targeting. *Proc Natl Acad Sci U S A* 2004;101:10266–71.
- [9] Awai K, Kakimoto T, Awai C, Kaneko T, Nakamura Y, Takamiya K, Wada H, Ohta H. Comparative genomic analysis revealed a gene for monoglucosyldiacylglycerol synthase, an enzyme for photosynthetic membrane lipid synthesis in cyanobacteria. *Plant Physiol* 2006;141:1120–7.
- [10] Chen M, Markham JE, Cahoon EB. Sphingolipid Delta8 unsaturation is important for glucosylceramide biosynthesis and low-temperature performance in *Arabidopsis*. *Plant J* 2012;69:769–81.
- [11] Tian Y, Zhao GY, Fang W, Xu Q, Tan RX. Delta10(E)-sphingolipid desaturase involved in fusaricide mycosynthesis and stress adaptation in *Fusarium graminearum*. *Sci Rep* 2015;5:10486.
- [12] Alden KP, Dhondt-Cordelier S, McDonald KL, Reape TJ, Ng CK, McCabe PF, Leaver CJ. Sphingolipid long chain base phosphates can regulate apoptotic-like programmed cell death in plants. *Biochem Biophys Res Commun* 2011;410:574–80.
- [13] Rivera IG, Ordonez M, Presa N, Gangoi P, Gomez-Larrauri A, Trueba M, Fox T, Kester M, Gomez-Munoz A. Ceramide 1-phosphate regulates cell migration and invasion of human pancreatic cancer cells. *Biochem Pharmacol* 2016;102:107–19.
- [14] Petrie JR, Shrestha P, Mansour MP, Nichols PD, Liu Q, Singh SP. Metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA Delta6-desaturase with omega 3-preference from the marine microalga *Micromonas pusilla*. *Metab Eng* 2010;12:233–40.
- [15] Domergue F, Abbadi A, Ott C, Zank TK, Zahringer U, Heinz E. Acyl carriers used as substrates by the desaturases and elongases involved in very long-chain polyunsaturated fatty acids biosynthesis reconstituted in yeast. *J Biol Chem* 2003;278:35115–26.
- [16] Cao J, Burn P, Shi Y. Properties of the mouse intestinal acyl-CoA:monoacylglycerol acyltransferase, MGAT2. *J Biol Chem* 2003;278:25657–63.
- [17] McFie PJ, Stone SL, Banman SL, Stone SJ. Topological orientation of acyl-CoA:diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the N terminus in dimer/tetramer formation. *J Biol Chem* 2010;285:37377–87.
- [18] Liu Q, Siloto RM, Snyder CL, Weselake RJ. Functional and topological analysis of yeast acyl-CoA:diacylglycerol acyltransferase 2, an endoplasmic reticulum enzyme essential for triacylglycerol biosynthesis. *J Biol Chem* 2011;286:13115–26.
- [19] Petrie JR. Recruiting a new substrate for triacylglycerol synthesis in plants: the monoacylglycerol acyltransferase pathway. *PLoS One* 2012;7.
- [20] Hoffmann M, Wagner M, Abbadi A, Fulda M, Feussner I. Metabolic engineering of omega3-very long chain polyunsaturated fatty acid production by an exclusively acyl-CoA-dependent pathway. *J Biol Chem* 2008;283:22352–62.
- [21] Bai Y, McCoy JG, Levin EJ, Sobrado P, Rajashankar KR, Fox BG, Zhou M. X-ray structure of a mammalian stearoyl-CoA desaturase. *Nature* 2015;524:252–6.
- [22] Wang H, Klein MG, Zou H, Lane W, Snell G, Levin I, Li K, Sang BC. Crystal structure of human stearoyl-coenzyme A desaturase in complex with substrate. *Nat Struct Mol Biol* 2015;22:581–5.
- [23] Song LY, Zhang Y, Li SF, Hu J, Yin WB, Chen YH, Hao ST, Wang BL, Wang RR, Hu ZM. Identification of the substrate recognition region in the delta-fatty acid and delta-sphingolipid desaturase by fusion mutagenesis. *Planta* 2013;239:753–63.
- [24] Alonso DL, Garcia-Maroto F, Rodriguez-Ruiz J, Garrido JA, Vilches MA. Evolution of the membrane-bound fatty acid desaturases. *Biochem Syst Ecol* 2003;31:1111–24.
- [25] Hashimoto K, Yoshizawa AC, Okuda S, Kuma K, Goto S, Kanehisa M. The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J Lipid Res* 2008;49:183–91.
- [26] Uttaro AD. Biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB Life* 2006;58:563–71.
- [27] Aguilar PS, Cronan JE, de Mendoza D. A *Bacillus subtilis* gene induced by cold shock encodes a membrane phospholipid desaturase. *J Bacteriol* 1998;180:2194–200.
- [28] Diaz AR, Mansilla MC, Vila AJ, de Mendoza D. Membrane topology of the acyl-lipid desaturase from *Bacillus subtilis*. *J Biol Chem* 2002;277:48099–106.

- [29] Pyne S, Pyne NJ. Sphingosine 1-phosphate signalling in mammalian cells. *Biochem J* 2000;349:385–402.
- [30] Ahmed FH, Carr PD, Lee BM, Afriat-Jurnou L, Mohamed AE, Hong NS, Flanagan J, Taylor MC, Greening C, Jackson CJ. Sequence–structure–function classification of a catalytically diverse oxidoreductase superfamily in Mycobacteria. *J Mol Biol* 2015; 427:3554–71.
- [31] Devos D, Valencia A. Intrinsic errors in genome annotation. *Trends Genet* 2001;17: 429–31.
- [32] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. *Nucleic Acids Res* 2014;42:D222–30.
- [33] Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger H, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* 2006;34:D187–91.
- [34] Gerlt JA, Babbitt PC. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 2001;70:209–46.
- [35] Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. Enzyme Function Initiative–Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *BBA Proteins Proteomics* 2015;1854:1019–37.
- [36] Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, Jagessar K, Kim J, Patskovsky Y, Seidel RD, Stead M, Toro R, Vetting MW, Almo SC, Armstrong RN, Babbitt PC. Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS Biol* 2014; 12, e1001843.
- [37] Wichelecki DJ, Balthazor BM, Chau AC, Vetting MW, Fedorov AA, Fedorov EV, Lukk T, Patskovsky YV, Stead MB, Hillerich BS, Seidel RD, Almo SC, Gerlt JA. Discovery of function in the enolase superfamily: D-mannonate and d-gluconate dehydratases in the D-mannonate dehydratase subgroup. *Biochemistry* 2014;53:2722–31.
- [38] Shanklin J, Whittle E, Fox BG. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* 1994;33:12787–94.
- [39] Shanklin J, Cahoon EB. Desaturation and related modifications of fatty acids. *Annu Rev Plant Physiol* 1998;49:611–41.
- [40] Ye RW, Stead KJ, Yao H, He H. Mutational and functional analysis of the beta-carotene ketolase involved in the production of canthaxanthin and astaxanthin. *Appl Environ Microbiol* 2006;72:5829–37.
- [41] Wang J, Yu L, Schmidt RE, Su C, Huang X, Gould K, Cao G. Characterization of HSCD5, a novel human stearyl-CoA desaturase unique to primates. *Biochem Biophys Res Commun* 2005;332:735–42.
- [42] Hashimoto K, Yoshizawa AC, Saito K, Yamada T, Kanehisa M. The repertoire of desaturases for unsaturated fatty acid synthesis in 397 genomes. *Genome informatics International Conference on Genome Informatics*, 17; 2006 173–83.
- [43] Alloatti A, Uttaro AD. Highly specific methyl-end fatty-acid desaturases of trypanosomatids. *Mol Biochem Parasitol* 2011;175:126–32.
- [44] Michaelson LV, Zauner S, Markham JE, Haslam RP, Desikan R, Mugford S, Albrecht S, Warnecke D, Sperling P, Heinz E, Napier JA. Functional characterization of a higher plant sphingolipid Delta4-desaturase: defining the role of sphingosine and sphingosine-1-phosphate in Arabidopsis. *Plant Physiol* 2009;149:487–98.
- [45] Covello PS, Reed DW. Functional expression of the extraplasmidial *Arabidopsis thaliana* oleate desaturase gene (FAD2) in *Saccharomyces cerevisiae*. *Plant Physiol* 1996;111:223–6.
- [46] Pan Z, Rimando AM, Baerson SR, Fishbein M, Duke SO. Functional characterization of desaturases involved in the formation of the terminal double bond of an unusual 16:3Delta(9,12,150) fatty acid isolated from *Sorghum bicolor* root hairs. *J Biol Chem* 2007;282:4326–35.
- [47] Sayanova O, Haslam R, Guschina I, Lloyd D, Christie WW, Harwood JL, Napier JA. A bifunctional Delta12,Delta15-desaturase from *Acanthamoeba castellanii* directs the synthesis of highly unusual n-1 series unsaturated fatty acids. *J Biol Chem* 2006; 281:36533–41.
- [48] Iwabuchi M, Kohno-Murase J, Imamura J. Delta12-oleate desaturase-related enzymes associated with formation of conjugated trans-Delta11, cis-Delta13 double bonds. *J Biol Chem* 2003;278:4603–10.
- [49] Jones AL, Lloyd D, Harwood JL. Rapid induction of microsomal delta 12(omega 6)-desaturase activity in chilled *Acanthamoeba castellanii*. *Biochem J* 1993;296(Pt 1): 183–8.
- [50] Miquel M, Browse J. Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. Biochemical and genetic characterization of a plant oleoyl-phosphatidylcholine desaturase. *J Biol Chem* 1992;267:1502–9.
- [51] Gibson S, Arondel V, Iba K, Somerville C. Cloning of a temperature-regulated gene encoding a chloroplast omega-3 desaturase from *Arabidopsis thaliana*. *Plant Physiol* 1994;106:1615–21.
- [52] Napier JA, Sayanova O, Stobart AK, Shewry PR. A new class of cytochrome b5 fusion proteins. *Biochem J* 1997;328(Pt 2):717–8.
- [53] Domergue F, Lerchl J, Zahringer U, Heinz E. Cloning and functional characterization of *Phaeodactylum tricoratum* front-end desaturases involved in eicosapentaenoic acid biosynthesis. *Eur J Biochem* 2002;269:4105–13.
- [54] Sayanova OV, Beaudoin F, Michaelson LV, Shewry PR, Napier JA. Identification of primula fatty acid Delta(6)-desaturases with n-3 substrate preferences. *FEBS Lett* 2003;542:100–4.
- [55] Sayanova O, Haslam R, Caleron MV, Napier JA. Cloning and characterization of unusual fatty acid desaturases from *Anemone leveillei*: identification of an acyl-coenzyme A C-20 Delta(5)-desaturase responsible for the synthesis of sciadonic acid. *Plant Physiol* 2007;144:455–67.
- [56] Zhou XR, Robert SS, Petrie JR, Frampton DM, Mansour MP, Blackburn SI, Nichols PD, Green AG, Singh SP. Isolation and characterization of genes from the marine microalga *Pavlova salina* encoding three front-end desaturases involved in docosahexaenoic acid biosynthesis. *Phytochemistry* 2007;68:785–96.
- [57] Pereira SL, Leonard AE, Huang YS, Chuang LT, Mukerji P. Identification of two novel microalgal enzymes involved in the conversion of the omega3-fatty acid, eicosapentaenoic acid, into docosahexaenoic acid. *Biochem J* 2004;384:357–66.
- [58] Domergue F, Abbadi A, Zahringer U, Moreau H, Heinz E. In vivo characterization of the first acyl-CoA Delta6-desaturase from a member of the plant kingdom, the microalga *Ostreococcus tauri*. *Biochem J* 2005;389:483–90.
- [59] Helmkampf M, Cash E, Gadau J. Evolution of the insect desaturase gene family with an emphasis on social Hymenoptera. *Mol Biol Evol* 2015;32:456–71.
- [60] Chang Y, Fox BG. Identification of Rv3230c as the NADPH oxidoreductase of a two-protein DesA3 acyl-CoA desaturase in *Mycobacterium tuberculosis* H37Rv. *Biochemistry* 2006;45:13476–86.
- [61] Kajikawa M, Yamato KT, Kohzu Y, Nojiri M, Sakuradani E, Shimizu S, Sakai Y, Fukuzawa H, Ohyama K. Isolation and characterization of delta(6)-desaturase, an ELO-like enzyme and delta(5)-desaturase from the liverwort *Marchantia polymorpha* and production of arachidonic and eicosapentaenoic acids in the methylotrophic yeast *Pichia pastoris*. *Plant Mol Biol* 2004;54:335–52.
- [62] Zhu G, Koszelak-Rosenblum M, Connelly SM, Dumont ME, Malkowski MG. The crystal structure of an integral membrane fatty acid alpha-hydroxylase. *J Biol Chem* 2015;290:29820–33.
- [63] Zhang J, Madden TL. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res* 1997;7:649–56.
- [64] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- [65] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–4.
- [66] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–7.
- [67] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10:845–58.