Molecular cloning and 3D model of a fatty-acid elongase in a carnivorous freshwater teleost, the European perch (*Perca fluviatilis*)

Emmanuel Tinti^{1,2} · Florian Geay³ · Maximilien Lopes Rodrigues^{1,5} · Patrick Kestemont^{2,4} · Eric A. Perpète^{1,2,5} · Catherine Michaux^{1,5,6}

Received: 14 March 2019 / Accepted: 20 May 2019 / Published online: 30 May 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

The European perch (*Perca fluviatilis*) is a carnivorous freshwater fish able to metabolise polyunsaturated fatty acids (PUFA) into highly unsaturated fatty acids (HUFA). This makes it a potential candidate for sustainable aquaculture development. In this study, special attention is given to the fatty-acid elongase (ELOVL) family, one of the two enzymatic systems implied in the HUFA biosynthesis. Structural information on European perch enzyme converting PUFA into HUFA is obtained by both molecular cloning and in silico characterization of an ELOVL5-like elongase from *P. fluviatilis* (*pfELOVL*). The full-length cDNA sequence consists of a 885-base pair Open Reading Frame coding for a 294-amino acid protein. Phylogenetic analysis and sequence alignment with fish elongases predict the *pfELOVL* clusters within the ELOVL5 sub-group. The amino-acid sequence displays the typical ELOVL features: several transmembrane α helices (TMH), an endoplasmic reticulum (ER) retention signal, and four “conserved boxes” involved in the catalytic site. In addition, the topology analysis predicts a 7-TMH structure addressed in the ER membrane. A 3D model of the protein embedded in an ER-like membrane environment is also provided using de novo modelling and molecular dynamics. From docking studies, two putative enzyme–substrate-binding modes, including H bonds and CH– π interactions, emphasize the role of specific residues in the “conserved boxes”.

Keywords Cloning · Fatty-acid elongase · HUFA · Modelling · *Perca fluviatilis*

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1773-x>) contains supplementary material, which is available to authorized users.

✉ Catherine Michaux
catherine.michaux@unamur.be

¹ Laboratoire de Chimie Physique des Biomolécules, UCPTS, University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium

² Institute of Life-Earth-Environment, University of Namur, Namur, Belgium

³ BioMar AS, 7 Havnegata, 7010 Trondheim, Norway

⁴ Research Unit in Environmental and Evolutionary Biology, University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium

⁵ Namur Institute of Structures Matter, University of Namur, Namur, Belgium

⁶ Namur Research Institute for Life Sciences, University of Namur, Namur, Belgium

Introduction

One of the major challenges of future society will be the population supplying of food resources. In this context, aquaculture has been developed to offset the increase in water products demand (D’Odorico et al. 2014), supplying almost half of halieutic products intended for human diet (Godfray et al. 2010).

As the principal products of aquaculture, fish are the major dietary sources of highly unsaturated fatty acids (HUFA) for humans, especially for the n-3 HUFA series, such as eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3) (Strobel et al. 2012). These FA, with the n-6 series arachidonic acid (ARA, 20:4n-6), are essential fatty acids (EFA) for fish as well as for humans and mammals in general (Tocher 2015). EFA also have particularly important roles in cellular and physiological processes (Walls et al. 2016; Yeagle 2014). Moreover, ARA, EPA, and DHA are eicosanoids precursors, implicated in practically every organ, tissue, and cell, which explains

the very diversified physiological roles of EFA (Calder 2014). Particularly, essential HUFA are involved in immune response (Russell and Bürgin-Maunders 2012), and balance between them can improve or impair resistance against pathogens, cause cardiac diseases, inflammatory, or autoimmune disorders (Nayak et al. 2017; Geay et al. 2015a).

Up to now, the main n-3 HUFA source in fish food is fish oil (FO), coming from small marine species such as anchovies and sardines (Graziano da Silva 2014). This practice greatly forces the continuing growth of aquaculture activities (Tacon et al. 2012). However, alternatives do exist, such as vegetable oils (VO) (Tocher 2015). Unlike FO, VO lack HUFA but are rich in their C18 polyunsaturated fatty-acid (PUFA) precursors: linoleic acid (LA, 18:2n-6) and α -linolenic acid (ALA, 18:3n-3), respectively, for the n-6 and n-3 series (Lenihan-Geels et al. 2013).

The capacity of fish to synthesize HUFA from PUFA varies among species. Indeed, freshwater fish, mainly herbivorous or omnivorous, can produce DHA from ALA, and ARA from LA. In contrast, the mostly carnivorous marine species are unable to convert C18 into C20 and C22 at a physiologically significant rate, either because of a repression of enzymatic stakeholders or because of their absence, depending on regarded species (Tocher 2015; Geay et al. 2015b). The enzymes converting PUFA into HUFA are members of fatty-acid desaturases and very long-chain fatty-acid elongases (ELOVL) families. In fish, two desaturases and three elongases are involved in the PUFA/HUFA metabolism (Monroig et al. 2013a; Xue et al. 2014).

European perch (*Perca fluviatilis* L., 1758) is a carnivorous freshwater species, widely distributed in Eurasia. Although belonging to a nearly exclusively marine order (Perciformes), its capacity to synthesize HUFA from PUFA, especially n-3 series (Henrotte et al. 2011), points it out as a potential candidate for a sustainable aquaculture development and diversification, regarding the consumer dietary preference for carnivorous species (Steenfeldt et al. 2015). However, few is known about the molecular fundamentals of the HUFA synthesis pathway in this fish species. While we have already partially characterized its $\Delta 6$ fatty-acid desaturase (Geay et al. 2016), no investigation on elongase enzymes was performed. In that context, we here have focused on the molecular cloning and in silico characterization of an ELOVL from *P. fluviatilis* (*pfELOVL*) to have the first structural information on a European perch enzyme converting PUFA into HUFA. Currently, the Protein Data Bank (PDB) does not contain any 3D structure of ELOVL and very few studies were dedicated to their structural features (Denic and Weissman 2007; Hernandez-Buquer and Blacklock 2013; Vrinten et al. 2010; Logan et al. 2014; Nayak et al. 2018).

Over the past decade, ELOVL2, -4 and -5, membrane enzymes of the endoplasmic reticulum (ER) have been

sequenced in various freshwater (Kuah et al. 2015; Fonseca-Madrugal et al. 2014; Ren et al. 2012; Monroig et al. 2010; Tan et al. 2010; Gregory and James 2014), marine (Monroig et al. 2011a; Morais et al. 2011, 2012), anadromous (Monroig et al. 2013b; Gregory and James 2014; Carmona-Antoñanzas et al. 2011, 2013; Morais et al. 2009), and catadromous species (Wang et al. 2014; Mohd-Yusof et al. 2010; Tu et al. 2012a). Alignments of these sequences allow to highlight their general features and conserved motifs putatively involved in the catalytic site: several (from 5 to 7) transmembrane α helices (TMH), one ER retention signal in the C-terminal region, and four conserved boxes (KEDT-, His-, Tyr-, Gln-box) (Monroig et al. 2016). The importance of these amino acids is supported by several site-directed mutagenesis studies and chimeric protein construction experiments in unicellular eukaryotes (Denic and Weissman 2007; Hernandez-Buquer and Blacklock 2013; Vrinten et al. 2010; Logan et al. 2014; Kihara 2012; Blacklock et al. 2008). In vertebrates, ELOVL shows a weak substrate specificity without discrimination based on length or double-bond number (Kihara 2012). ELOVL2 shows a preference towards C20n-3 and C22n-3, while ELOVL5 prefers C18n-3 and C20n-3. ELOVL4 is present in a limited number of tissues and shows an exclusive activity towards C22n-3 HUFA (Gregory and James 2014).

In this contribution, we perform (1) a consensus analysis of the topology, (2) a sequence and phylogenetic analysis for *pfELOVL5*-like elongase, and we predict (3) its 3D structure embedded in a native-like membrane environment, highlighting the conserved motifs, a potential active site, and two putative substrate-binding modes.

Materials and methods

Isolation and cloning of the ELOVL cDNA

All animal-handling procedures were approved by the local ethics committee for animal research of the University of Namur. Total RNA was extracted from adult individual liver using the InnuSPEED Tissue RNA© kit (Analytik Jena) and reverse-transcribed with the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas). Degenerated and specific primers were designed from available fish elongase sequences on NCBI database (NLM–NIH) and used to perform degenerated, 5'/3'-RACE and full-length PCR. Amplicons were obtained using Advantage® cDNA PCR kit (Clontech). Each cDNA product was purified by the High-Pure PCR Product Purification kit (Roche Applied Science) and cloned in *E. coli* NovaBlue competent cell strain (Clonables™ Kit, Novagen) with the pGEM®-T Easy Vector System I (Promega).

Purification of plasmid vector was performed using Gen-Elute™ Five-Minute Plasmid Miniprep Kit® (Sigma). Sequencing was achieved by MacroGen Inc.

Sequence analysis and 2D-structural patterns

The *p/ELOVL* primary structure was translated with the ExPASy Translate tool (Swiss Institute of Bioinformatics, Swiss). For comparative analyses, amino-acid sequences were aligned by MAFFT (CBRC-AIST, Japan) using BLOSUM matrix to assess identity and similarity, while phylogeny was predicted with the neighbour-joining method (Saitou and Nei 1987) and constructed by MEGA5 (Arizona State University, USA).

The proportion of secondary structures was estimated by the NPS@ secondary structure prediction tool (Université Lyon I—Claude Bernard, France). The putative presence of TMH was checked by hydropathy profiles produced by 15 programs: DAS (Research Institute of Molecular Pathology, Austria), ExPASy TMPred tool, HMMTOP (Hungarian Academy of Sciences, Hungary), OCTOPUS (Stockholm University, Sweden), NPS@ PHDhtm tool, Philius (University of Washington, USA), Phobius (Stockholm Bioinformatics Center, Sweden), PolyPhobius (Stockholm Bioinformatics Center), SCAMPI (Stockholm Bioinformatics Center), SOSUI (Nagahama Institute of Bio-Science and Technology, Japan), SPLIT 4.0 (University of Split, Croatia), SPOCTOPUS (Stockholm Bioinformatics Center), TM-Finder (Toronto University, Canada), TMHMM (Technical University of Denmark, Denmark), TopPred (Université Paris VII—Diderot, France). Hydropathy profiles were also exploited to predict the membrane orientation of N- and C-terminal extremities and soluble segments.

3D-structure prediction

Two strategies were selected to predict the 3D structure of the *p/ELOVL*: the fold recognition and the de novo methods. RaptorX (University of Chicago, USA) managed the fold recognition, whereas the Rosetta@home suite (University of Washington, USA) was operated to realise de novo modelling. The required inputs were obtained from experience (protein sequence), Robetta (fragment library), OCTOPUS (hydropathy profile), and Rosetta@home (lipophilicity profile). The 3D models were generated with the “Membrane-initio2” protocol. Ramachandran plots were drawn with RAMPAGE (University of Cambridge, UK). All output files were viewed with MacPyMOL (Schrödinger Inc., USA) or Visual Molecular Dynamics (VMD) (University of Illinois, USA).

Molecular dynamics

The lipid membrane environment was built with the CHARMM-GUI program (Lee et al. 2016). Molecular dynamics (MD) simulations were performed thanks to GROMACS 5.0.2 program (Abraham et al. 2015; Berendsen et al. 1995). The simulation systems consist of cubic boxes of $10 \times 12 \times 12$ nm. Water thickness was set to 1.7 nm on both sides, which makes approximately 40,000 water molecules in total. Bilayers consist of 172 1-stearoyl-2-linoleoyl-sn-glycero-phosphocholine (SLPC). The three-site water model TIP3P was chosen to describe the solvent, and periodic boundary conditions were applied in all directions. Eleven chloride anions were randomly placed to neutralize the positive charge of the protein. The selected force field is charmm36 (Huang and Mackerell 2013) and additional parameters for lipids were taken from CHARMM-GUI. H bonds were constrained using the linear constraint solver (LINCS) algorithm (Hess et al. 1997). The Verlet cut-off scheme was set to 1.2 nm for pair interactions and the cut-off distance for Van der Waals and electrostatic interactions was also set to 1.2 nm. The electrostatic interactions were described via the particle-mesh Ewald (PME) method with a grid spacing of 1.0 Å (Essmann et al. 1995). The energy minimization step used the steepest descent method. The systems were equilibrated to relax the uncorrelated initial configurations. A first equilibration was performed in the NVT ensemble for 0.5 ns with protein coordinates fixed and a second one in the NPAT ensemble for 0.325 ns without restraints. Production simulations ran for 210 ns in the NPT ensemble with a time step of 2 fs, with protein, lipids, ions, and water independently coupled with a Nosé–Hoover thermostat. Temperature was set at 298 K. Pressure was kept constant to 1 atm by the semi-isotropic Parrinello–Rahman pressure coupling algorithm (Parrinello and Rahman 1981). The results were analyzed by integrated GROMACS functions and by VMD (Humphrey et al. 1996). The secondary structure analysis was done using the DSSP algorithm (Kabsch and Sander 1983).

Molecular docking

The docking between our model and its substrate was performed with the GOLD suite of programs (Cambridge Crystallographic Data Centre, UK). The binding region was defined as a 10 Å radius sphere centered on Lys120. The following amino acids were defined as being part of the catalytic site (belonging to the 4 conserved boxes): K¹²⁰, E¹²³, T¹²⁷ (from KEDT-box), H¹⁴⁶, H¹⁴⁷ (from His-box), N¹⁷³, H¹⁷⁷, M¹⁸⁰, Y¹⁸¹, Y¹⁸³ (from Tyr-box), T²⁰³ (from Gln-box). γ -linolenic acid (GLA) and ARA were chosen as substrates. For the 20-genetic algorithm runs, a total of 100,000 genetic operations were carried out on five islands, each containing

100 individuals. The niche size was set to 2, and the value for the selection pressure was set to 1.1. Genetic operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. The scoring function used to rank the dockings was Goldscore. The outputs were analyzed with MacPyMOL and VMD.

Results and discussion

Cloning and sequence analysis

We cloned the full-length cDNA of an elongase from the European perch. PCR with specific primers (Table 1) resulted in the amplification of a 885-base pair Open Reading Frame coding for a 294-amino acid protein (GenBank accession no. KR360724), with a calculated MW and a pI of 35 kD and 9, respectively. Phylogenetic analysis (Fig. 1) compares fish elongases and sets the perch sequence close to other Perciformes (*Sparus aurata* and *Epinephelus coioides*) and within the marine teleost ELOVL (*Nibea mitsukurii*, *Argyrosomus regius*, *Dicentrarchus labrax*, *Siganus canaliculatus*, *Scophthalmus maximus*, etc.). More generally, *pfELOVL* clusters within the ELOVL5 sub-group of the ELOVL2/5 group, distinct from its ELOVL4 and ELOVL1/6/7 counterparts. A BLASTp analysis on the translated amino-acid sequence indicates that *pfELOVL* shows 90%, 89%, 78%, and 73% identity with *Lates calcarifer*, *Epinephelus coioides*, *Salmo salar*, and *Danio rerio* ELOVL5, respectively, and 77% with *Oncorhynchus mykiss* ELOVL2. Regarding the independent evolution of the different ELOVL elongase sub-families amongst biological lineage (Xue et al. 2014), and considering *pfELOVL* clustered within ELOVL5 sub-group, it is most likely our protein belongs to this one. A functional study in heterologous system would definitely confirm this hypothesis.

Moreover, amino-acid sequence alignment with fish elongases (ELOVL2/5 and ELOVL4 groups) from freshwater (*Oreochromis niloticus*, *O. mykiss*, and *D. rerio*),

marine (*S. aurata* and *Rachycentron canadum*), and anadromous species (*S. salar*) confirms that *pfELOVL* includes all the characteristics of elongases: several TMH (N³¹...K⁵¹, L⁶⁵...T⁸⁵, I¹⁰⁹...F¹²⁹, F¹⁴¹...V¹⁶¹, G¹⁶⁹...I¹⁸⁹, Y²⁰¹...A²²¹ and K²²⁹...F²⁴⁹), an ER retention signal (K²⁸⁹KLR²⁹²) and the four “conserved boxes” (KEDT-box, K¹²⁰LIEFMDT¹²⁷; His-box, Q¹³⁸MTFLHIYHH¹⁴⁷; Tyr-box, N¹⁷³SFVHVVMYSY¹⁸⁴; and Gln-box, T²⁰³QLQLIQ²⁰⁹) (Fig. 2).

2D structural patterns

In our *pfELOVL*-5 model, 113 residues are involved in α helix (38.44%), 52 in β sheet (17.69%), 120 in random coil (40.82%), and 9 (3.06%) are not assigned. Globally, this result corroborates a higher proportion of α secondary structure than a β one, in accordance with the literature for the elongase family (Leonard et al. 2004; Tvrdik et al. 2000).

Furthermore, we found an endoplasmic reticulum retention signal (K²⁸⁹Kxx²⁹²) and the protein is predicted to be addressed in the ER, data confirmed by the literature (Jakobsson et al. 2006). However, we could not find any signal peptide. Nonetheless, for proteins with multiple α helices, it is known membrane addressing can be directed by an N-terminal “hydrophobic signal” corresponding to one of the two first α helices inserted in the membrane (Monné et al. 2005). We predict this is the case for the ELOVL family.

In addition, the majority (12 of 15) of the secondary structure prediction tools (Table 1) have also shown seven TMH, also in agreement with the literature, which commonly admits between 5 and 7 α helices (Fig. 3). Furthermore, the orientation of the loop elements is supported by experimental data regarding the localisation of the putative catalytic site: precursor synthesis in the cytosol, cofactors localisation on the ER cytoplasmic face, and proof of an enzyme activity inhibition outside the ER (Monné et al. 1999, 2005; DePierre and Ernster 1977).

Table 1 Sequence and melting temperature of degenerate and gene-specific primers

PCR	Primers	Primer sequences	T_m (°C)
Degenerated PCR	Elovl-deg-F1	5'-YVCTDTGGTGGTAYTAYTTCTC-3'	46.0
	Elovl-deg-R1	5'-TRTANGTCTKWATGTAGAAGTT-3'	42.3
5'-RACE-PCR	Elo-5'-R1	5'-TAGAAGGACAAGACTGTGAGGCCAGATTGT-3'	58.0
	Elo-5'-R2	5'-TACGGCTGCCTGTGTTTCATGTACTTG-3'	54.6
	Elo-5'-R3	5'-TAGTTGTCGAGCAGCAGCCATCCCTG-3'	57.6
3'-RACE-PCR	Elo-3'-F1	5'-CAGCTTCGTCCACGTTGTGATGTATTC-3'	54.6
	Elo-3'-F2	5'-GAAGAAGTACATCACACAGTTACAGCTGA-3'	53.6
	Elo-3'-F3	5'-AAGGGATGGCTGTACTCCCAACAAG-3'	54.4
FL-PCR	Elvol5-dig-F	5'-CCCAAGCTTATGGAGACCTTTAATCATAA-3'	52.2
	Elovl5-dig-R	5'-GCGGGATCCTCAATCCACCCTCAGTTTCT-3'	59.3

Fig. 1 Phylogenetic relationship of *Perca fluviatilis* ELOVL with elongase proteins from other organisms

3D structure model embedded in an artificial membrane

As no ELOVL homologues structure is available in the PDB to perform homology modelling, a 3D model was set up using the de novo method Rosetta, which is well

known for efficiently modelling membrane proteins (Das and Baker 2008). Based on (1) the determined amino-acid sequence, (2) the 7 TMH topology, (3) the calculated lipophilicity profile, and (4) a template identified by a fold-recognition method (RaptorX), i.e., the NADH-CoQ reductase/NADH dehydrogenase from *Escherichia coli*

P. fluviatilis ELOVL	---METFNHKLNT---YFETWMGPRDQVRVQGWLLLD	<u>NYPPTFALTVMYLLIVWMGPK</u> YMK
S. aurata ELOVL5	---METFNHKLNV---YFETWMGPRDQVRVQGWLLLD	<u>NYPPTFALTVMYLLIVWMGPK</u> YMK
O. niloticus ELOVL5	---METFNHKLNA---YIDSWMGPRDQVRVQGWLLLD	<u>NYPPTFALTVMYLLIVWMGPK</u> YMK
S. salar ELOVL5	---MEAFNHKLNT---YIDSWMGPRDERVQGWLLLD	<u>NYPPTFALTVMYLLIVWLGP</u> YMR
O. mykiss ELOVL2	MNHLQSLDERLNA---LFYFLFEDRDSRVQGWLLMD	<u>SYLPTLSLTILYLLTVYLGSK</u> YMR
D. rerio ELOVL2/5	---METFSHRVNS---YIDSWMGPRDLRVGTWFLLD	<u>DYIPTFIETVMYLLIVWMGPK</u> YMK
R. canadum ELOVL4	---MEVVFTHFVNDTVEFYKWSLTIAKRVENWPMMA	<u>SPLPTLAISCLYLLFLWVGPR</u> YMQ
	: : . . : *	: * ** . : : . *** : : : ** : : *
P. fluviatilis ELOVL	HRQPYSRCRGL	<u>LVLVNLGLTVLSFYMFYELVT</u> AVWYGGYNFYCQDTHSAKE-ADNK
S. aurata ELOVL5	HRQPYPYPCRG	<u>LVLVNLGLTVLSFYMFYELVT</u> AVWYGGYNFYCQDTHSAQE-VDNK
O. niloticus ELOVL5	HRQPYSRCRAV	<u>MVFYNLGLTVLSFYMFYELVS</u> AAWHGGYNFYCQDTHSAEE-ADIK
S. salar ELOVL5	HRQPVSCQGL	<u>LVLVNLGLTVLSFYMFYELVS</u> AVWQGGYNFYCQDTHSAGE-TDTK
O. mykiss ELOVL2	NRPAYSLKGV	<u>LOVYNFVSMTLSLYMLVELVS</u> ATLSAGYRLQCQGLHEAGE-ADLR
D. rerio ELOVL2/5	NRQAYSCRAL	<u>LVPYNLCLTVLSLYMFYELVM</u> SVYQGGYNFYCQDTHSGGD-ADNR
R. canadum ELOVL4	DRQPYTLRRT	<u>LIVYNFVSMTVNFYIAKELLI</u> ATRAAGYSYLCQPVNYSNDVNEVR
	. *	: : ** : : . * : : : * : : . ** ** : : : : . *
P. fluviatilis ELOVL		<u>WWYFYSKLIEFMDTEFF</u> FILRKNNH <u>QMTFLHIYHHATMLNIWWFVNVW</u> PCGHTYF <u>GPTIN</u>
S. aurata ELOVL5		<u>WWYFYSKLIEFMDTEFF</u> FILRKNNH <u>QITFLHTYHHASMLNIWWFVNVW</u> PCGHSYF <u>GASLN</u>
O. niloticus ELOVL5		<u>WWYFYSKLIEFMDTEFF</u> FILRKNNH <u>QITFLHTYHHASMLNIWWFVNVW</u> PCGHSYF <u>GASLN</u>
S. salar ELOVL5		<u>WWYFYSKVIEFMDTEFF</u> FILRKNNH <u>QITFLHIYHHASMLNIWWFVNVW</u> PCGHSYF <u>GASLN</u>
O. mykiss ELOVL2		<u>WWYFYSKVIEFMDTEFF</u> FILRKNNH <u>QITFLHVIYHHASMLNIWWFVNVW</u> PCGHSYF <u>GPTLN</u>
D. rerio ELOVL2/5		<u>WWYFYSKLIEFMDTEFF</u> FILRKNNH <u>QITFLHVIYHHATMLNIWWFVNVW</u> PCGHSYF <u>GATFN</u>
R. canadum ELOVL4		<u>WWYFYSKVIEFMDTEFF</u> FILRKNNH <u>QITFLHVIYHHCTMFLWVIGIKWV</u> PGGQAF <u>GATIN</u>
	**** : : ** : : . * : : : * : : : * : : : * : : : *	**** : : ** : : . * : : : * : : : * : : : * : : : *
P. fluviatilis ELOVL		<u>SFVHVVMYSYGLSAI</u> -PAMRPYLWKKY <u>YITQLOLIOFFLTITVQTMCAA</u> IWPCGF <u>PKGWL</u>
S. aurata ELOVL5		<u>SFVHVVMYSYGLSAI</u> -PAMRPYLWKKY <u>YITQLOLIOFFLTMSQITFAV</u> IWPCGF <u>PDGWL</u>
O. niloticus ELOVL5		<u>SFIHVVMYSYGLSAI</u> -PAIRPYLWKKY <u>YITQLOLIOFFLTVTQTTLAV</u> IWPCGF <u>PIGWL</u>
S. salar ELOVL5		<u>SFVHVLMYSYGLSAV</u> -PAIRPYLWKKY <u>YITQLOLIOFFLTMSQITCAV</u> IWPCGF <u>PRGWL</u>
O. mykiss ELOVL2		<u>SFIHVCMSYGLSTI</u> -PSMQYLLWKKY <u>YITQLOLIOFFLTITITHTLSAI</u> VVPCGF <u>PVGCL</u>
D. rerio ELOVL2/5		<u>SFIHVLMYSYGLSAV</u> -PALRPYLWKKY <u>YITQLOLIOFFLTITITHTLSAI</u> VVPCGF <u>PMGWL</u>
R. canadum ELOVL4		<u>SSIHVLMYGYGLAAL</u> GPQMQLWKKY <u>YITQLOLIOFFLTITITHTLSAI</u> VTGCPF <u>PCWMO</u>
	* : ** ** : : ** : : : * : : : ** : : : * : : : *	* : ** ** : : ** : : : * : : : ** : : : * : : : *
P. fluviatilis ELOVL		<u>YSQTSYMTLILFLFNF</u> YIQTYYKHSKS---LKKEHHNGSPVSTNGNANATPSLERTAH-
S. aurata ELOVL5		<u>YFQIGYMTLILFLFNF</u> YIQTYYKHSKS---LRKEHQNGSPLSTNGHANGTPSMEHTAH-
O. niloticus ELOVL5		<u>YFQISYMTLILFLFNF</u> YIQTYYKHSKS---QRKHPQNSLLSTNGHANGTPSTDTFTAP-
S. salar ELOVL5		<u>YFQIFYMASLIAFFSNE</u> YIQTYYKHSKS---QKEYHQNGSVDSLNGHANGTPSTDTFTITH-
O. mykiss ELOVL2		<u>YFQISYMTLILFLFNF</u> YIQTYYKHSKS---IKSSRPNGHTVSTNGTSF-----
D. rerio ELOVL2/5		<u>YFQISYMTLILFLFNF</u> YIQTYYKHSKS---RKSDYPNGS---VNGHTNGVMSSEKIKH-
R. canadum ELOVL4		<u>YFQISYMTLILFLFNF</u> YIQTYYKHSKS---YHAYRGKPSSSQKGGKPIANGTSVVTNGHSHKVEEDNGKRO
	* : : : * ** : : *	* : : : * ** : : *
P. fluviatilis ELOVL	KKLRVD--	
S. aurata ELOVL5	KKLRVD--	
O. niloticus ELOVL5	KKLRVD--	
S. salar ELOVL5	RKVRVD--	
O. mykiss ELOVL2	KK-RK---	
D. rerio ELOVL2/5	RKARAD--	
R. canadum ELOVL4	KKGRAKRE	
	: * *	

Fig. 2 Alignment of the elongase amino-acid sequence from European Perch with fish orthologs. The GenBank accession numbers of the aligned amino-acid sequence areas follows: *Sparus aurata* (ADD50001), *Oreochromis niloticus* (NP_001266389), *Salmo salar* (NP_001130024), *Oncorhynchus mykiss* (NM_001124636), *Danio*

rerio (NP_956747), *Rachycentron canadum* (ADG59898.1). The ER retention motif is coloured in grey, the TMH in blue (underlined when overlapping with conserved boxes), and the conserved boxes in red. * sequence identity, : conservative substitution, . semi-conservative substitution

(3RKO, 3.0 Å resolution), we have obtained a valid model of the structure. Indeed, Ramachandran plots support the quality of the simulated model (more than 99% of favoured or allowed region) (Fig. 4). Geometrical constraints were

applied to the structure to conserve a minimal distance between residues in the four conserved boxes (with 5 Å).

As *pELOVL5* is a membrane protein, the modelled structure was embedded in an artificial lipid bilayer mimicking

Fig. 3 2D theoretical model of *pELOVL5* (<http://www.sacs.ucsf.edu/cgi-bin/open-topo2.py/>) inserted in ER membrane with the cytosol above and lumen side below. The four conserved boxes are coloured in red and the ER retention signal in purple

the natural environment of the protein. The overall structure was then refined using MD (Fig. 5a, b). A 1-stearoyl-2-linoleoyl-sn-glycero-phosphocholine (SLPC, C18:0/C18:2) lipid bilayer was selected, because phosphatidylcholine represents more than 50% of the ER lipid content in eukaryotes (OPM database). In addition, its linoleic acid residue (C18:2 Δ 9,12) is closely related to GLA (C18:3 Δ 6,9,12), an ELOVL5 substrate. Moreover, linoleic acid is a direct substrate of ELOVL5 in the Δ 8 desaturase alternative pathway (Fonseca-Madriral et al. 2014; Monroig et al. 2011b, 2013b; Tu et al.

2012b). The time evolution of Root Mean Square Deviation (RMSD) of the protein backbone atoms is analyzed to check the stability of the system. At the end of the simulation, the RMSD is almost steady, attesting the system's stabilization (Fig. 6a). Focusing on the flexible regions of the protein, the Root Mean Square Fluctuation (RMSF) plot was generated with respect to individual residues (Fig. 6b). The residues involved in loops and, therefore, more exposed to the solvent, show greater fluctuations, the C-terminal end not bound to the membrane being the most flexible. Regarding

Number of residues in favoured region (~98.0% expected) : 239 (93.4%)
 Number of residues in allowed region (~2.0% expected) : 11 (4.3%)
 Number of residues in outlier region : 6 (2.3%)

RAMPAGE by Paul de Bakker and Simon Lovell available at <http://www-cryst.bioc.cam.ac.uk/rampage/>

Please cite: S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson & D.C. Richardson (2002) Structure validation by C α geometry: ϕ/ψ and C β deviation. *Proteins: Structure, Function & Genetics*. **50**: 437-450

Fig. 4 Ramachandran plot of the model of *pELOVL5*

Fig. 5 Our *pf*ELOVL5 model in SLPC bilayer. **a** Side view; **b** top view

the secondary structural content, each motif remains stable during the simulation, with a content of 60–70% α -helix and no β -sheet/bridge/bulge (Fig. 7a). Finally, the radius of gyration (R_g) reflecting the compactness of *pf*ELOVL5 is stable over the time, and ranges from 2.25 to 2.45 nm (Fig. 7b). All these analyses demonstrate the robustness of our model.

Identification of a potential catalytic site

From our reliable 3D model embedded in membrane, a molecular docking analysis with two key substrates (GLA and ARA) was performed to identify essential residues involved in the formation of the enzyme–substrate complex,

Fig. 6 **a** Evolution of C α RMSD during the MD simulation, **b** RMSF of C α averaged over all subunits from the last 1 ns

an essential step to understand the mechanism of action of elongation.

Two binding modes of GLA have been identified (Fig. 8a, b). In the first one, called “I”, three types of molecular interactions are involved: a strong electrostatic interaction between K¹²⁰ and –COO[–], two H bonds between –COO[–], and the K¹²⁰ and H¹⁷⁷, and a CH– π interaction (Takahashi et al. 2010) between the ligand alkyl chain and the indole group of W¹¹⁴ (Fig. 8a). K¹²⁰ and H¹⁷⁷ are part of the KEDT- and Tyr-box, respectively. The alkyl chain is close to hydrophobic and aromatic amino acids: L²⁸, W²⁶, L¹¹³, Y¹¹⁷, V¹⁷⁶, and L²¹².

In the second one, called “II”, K¹²⁰ is also involved in an electrostatic interaction with –COO[–], while three H bonds between the ligand carboxylate group and Y¹¹⁷, Y¹⁸⁴, and Q²⁰⁹ are also observed (Fig. 8b). Q²⁰⁹ and Y¹⁸⁴ are in the Gln- and Tyr-box, respectively. The alkyl chain is surrounded by several hydrophobic residues such as L²⁸, A¹⁰⁵, I¹¹⁰, L¹¹³, V¹⁷⁶, and W¹¹⁴. For ARA, the mode II seems the preferential-binding mode with two H bonds observed

Fig. 7 **a** Rg plot of *pfELOVL5* during MD simulation, **b** secondary structure % as a function of time

between the carboxylate moiety and Y¹¹⁷ and H¹⁷⁷ (data not shown).

At this point, it is difficult to decide for one or other mode but these studies already highlight potential essential amino acids present in both binding modes, i.e., K¹²⁰, H¹⁷⁷, and W¹¹⁴. Interestingly, amongst them, K¹²⁰ and H¹⁷⁷ are absolutely conserved along fatty-acid elongases (Carmona-Antoñanzas et al. 2013). This contribution, presenting the model of an ELOVL from the European perch, a carnivorous freshwater fish and prime candidate for the diversification of European aquaculture, paves the way to better understanding of the substrate specificities of this enzyme family and the potential mechanisms of fatty-acid elongation.

Conclusion

In this study, we isolated and cloned the coding sequence of a very long-chain fatty-acid ELOVL5-like elongase from a freshwater fish, the European perch (*P. fluviatilis*).

Fig. 8 Molecular docking of GLA on *p/ELOVL5*. **a** Mode I, **b** mode II

Sequencing results from genetics experiments allowed to predict a 3D model structure embedded in an ER-like artificial membrane, used to provide two putative substrate-binding modes and highlight essential amino acids. Assume to belong to ELOVL5 sub-group of the ELOVL2/5 sub-family, it remains necessary to perform a functional study in heterologous system to confirm this hypothesis. Otherwise, our modelling studies give key structural data on a carnivorous teleost elongase belonging to an important protein family for human and animal nutrition.

Acknowledgements This work was supported by funding from the Belgian National Fund for Scientific Research (F.R.S.-FNRS) through a postdoctoral fellowship to F. Geay (FRFC project no. 6.8073.70) and a PhD fellowship to E. Tinti (FRIA project no. FC89770), and from the Agència de Gestió d'Ajuts Universitaris i de Recerca de Catalunya through a PhD fellowship to M. Lopes Rodrigues. C. Michaux

and E. A. Perpète thank the F.R.S.-FNRS for their Research Associate and Senior Research Associate position, respectively. This research used resources of the Platform High Performance Computing (PTCI) located at the University of Namur, Belgium, which is supported by the F.R.S.-FNRS (project no. 2.5020.11). The PTCI is member of the “Consortium des Équipements de Calcul Intensif (CÉCI)”.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abraham MJ et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25
- Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91:43–56
- Blacklock BJ, Kelley D, Patel S (2008) A fatty acid elongase ELO with novel activity from *Dictyostelium discoideum*. *Biochem Biophys Res Commun* 374:226–230
- Calder PC (2014) Very long chain omega-3 (n-3) fatty acids and human health. *Eur J Lipid Sci Technol* 116:1280–1300
- Carmona-Antoñanzas G et al (2011) Biosynthesis of very long-chain fatty acids (C > 24) in Atlantic salmon: cloning, functional characterisation, and tissue distribution of an Elovl4 elongase. *Comp Biochem Physiol Part B Biochem Mol Biol* 159:122–129
- Carmona-Antoñanzas G et al (2013) An evolutionary perspective on Elovl5 fatty acid elongase: comparison of Northern pike and duplicated paralogs from Atlantic salmon. *BMC Evol Biol* 13:85–97
- Das R, Baker D (2008) Macromolecular modeling with Rosetta. *Annu Rev Biochem* 77:363–382
- Denic V, Weissman JS (2007) A molecular caliper mechanism for determining very long-chain fatty acid length. *Cell* 130:663–677
- DePierre JW, Ernster L (1977) Enzyme topology of intracellular membranes. *Annu Rev Biochem* 46:201–262
- D’Odorico P et al (2014) Feeding humanity through global food trade. *Earth’s Future* 2:458–469
- Essmann U et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
- Fonseca-Madrigal J et al (2014) Diversification of substrate specificities in teleostei Fads2: characterization of D4 and D6D5 desaturases of *Chirostoma estor*. *J Lipid Res* 55:1408–1419
- Geay F et al (2015a) Effects of dietary linseed oil on innate immune system of Eurasian perch and disease resistance after exposure to *Aeromonas salmonicida* achromogen. *Fish Shellfish Immunol* 47:782–796
- Geay F et al (2015b) Dietary linseed oil reduces growth while differentially impacting LC-PUFA synthesis and accretion into tissues in Eurasian Perch (*Perca fluviatilis*). *Lipids* 50:1219–1232
- Geay F et al (2016) Cloning and functional characterization of $\Delta 6$ fatty acid desaturase (FADS2) in Eurasian perch (*Perca fluviatilis*). *Comp Biochem Physiol Part B Biochem Mol Biol* 191:112–125
- Godfray HCJ et al (2010) Food security: the challenge of feeding 9 billion people. *Science* 327:812–818
- Graziano da Silva J (2014) The State of World Fisheries and Aquaculture 2014. FAO, Rome
- Gregory MK, James MJ (2014) Rainbow trout (*Oncorhynchus mykiss*) Elovl5 and Elovl2 differ in selectivity for elongation

- of omega-3 docosapentaenoic acid. *BBA Mol Cell Biol Lipids* 1841:1656–1660
- Henrotte E et al (2011) n-3 and n-6 fatty acid bioconversion abilities in Eurasian perch (*Perca fluviatilis*) at two developmental stages. *Aquac Nutr* 17:e216–e225
- Hernandez-Buquer S, Blacklock BJ (2013) Site-directed mutagenesis of a fatty acid elongase ELO-like condensing enzyme. *FEBS Lett* 587:3837–3842
- Hess B et al (1997) LINCOS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
- Huang J, Mackerell AD Jr (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34:2135–2145
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38
- Jakobsson A, Westerberg R, Jacobsson A (2006) Fatty acid elongases in mammals: their regulation and roles in metabolism. *Prog Lipid Res* 45:237–249
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637
- Kihara A (2012) Very long-chain fatty acids: elongation, physiology and related disorders. *J Biochem* 152:387–395
- Kuah MK, Jaya-Ram A, Shu-Chien AC (2015) The capacity for long-chain polyunsaturated fatty acid synthesis in a carnivorous vertebrate: functional characterisation and nutritional regulation of a Fads2 fatty acyl desaturase with $\Delta 4$ activity and an Elovl5 elongase in striped snakehead (*Channa striata*). *Biochim Biophys Acta* 1851:248–260
- Lee J et al (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput* 12:405–413
- Lenihan-Geels G, Bishop KS, Ferguson LR (2013) Alternative sources of omega-3 fats: can we find a sustainable substitute for fish? *Nutrients* 5:1301–1315
- Leonard AE et al (2004) Elongation of long-chain fatty acids. *Prog Lipid Res* 43:36–54
- Logan S et al (2014) Endoplasmic reticulum microenvironment and conserved histidines govern ELOVL4 fatty acid elongase activity. *J Lipid Res* 55:698–708
- Mohd-Yusof NY et al (2010) Investigation of highly unsaturated fatty acid metabolism in the Asian sea bass, *Lates calcarifer*. *Fish Physiol Biochem* 36:827–843
- Monné M et al (1999) N-Tail translocation in a eukaryotic polytopic membrane protein. Synergy between neighboring transmembrane segments. *Eur J Biochem* 263:264–269
- Monné M et al (2005) Competition between neighboring topogenic signals during membrane protein insertion into the ER. *FEBS J* 272:28–36
- Monroig O et al (2010) Expression and role of Elovl4 elongases in biosynthesis of very long-chain fatty acids during zebrafish *Danio rerio* early embryonic development. *Biochim Biophys Acta* 1801:1145–1154
- Monroig Ó et al (2011a) Biosynthesis of long-chain polyunsaturated fatty acids in marine fish: characterization of an Elovl4-like elongase from cobia *Rachycentron canadum* and activation of the pathway during early life stages. *Aquaculture* 312:145–153
- Monroig O, Li Y, Tocher DR (2011b) Delta-8 desaturation activity varies among fatty acyl desaturases of teleost fish: high activity in delta-6 desaturases of marine species. *Comp Biochem Physiol Part B Biochem Mol Biol* 159:206–213
- Monroig Ó et al (2013a) Functional characterisation of a Fads2 fatty acyl desaturase with $\Delta 6/\Delta 8$ activity and an Elovl5 with C16, C18 and C20 elongase activity in the anadromous teleost meagre (*Argyrosomus regius*). *Aquaculture* 412:14–22
- Monroig Ó, Tocher DR, Navarro JC (2013b) Biosynthesis of polyunsaturated fatty acids in marine invertebrates: recent advances in molecular mechanisms. *Mar Drugs* 11:3998–4018
- Monroig Ó et al (2016) Evolutionary functional elaboration of the Elovl2/5 gene family in chordates. *Sci Rep* 6:1–10
- Morais S et al (2009) Highly unsaturated fatty acid synthesis in Atlantic salmon: characterization of ELOVL5- and ELOVL2-like elongases. *Mar Biotechnol* 11:627–639
- Morais S et al (2011) Expression of fatty acyl desaturase and elongase genes, and evolution of DHA:EPA ratio during development of unfed larvae of Atlantic bluefin tuna (*Thunnus thynnus* L.). *Aquaculture* 313:129–139
- Morais S et al (2012) Long chain polyunsaturated fatty acid synthesis in a marine vertebrate: ontogenetic and nutritional regulation of a fatty acyl desaturase with $\Delta 4$ activity. *Biochim Biophys Acta* 1821:660–671
- Nayak S et al (2017) Dietary arachidonic acid affects immune function and fatty acid composition in cultured rabbitfish *Siganus rivulatus*. *Fish Shellfish Immunol* 68:46–53
- Nayak M et al (2018) Molecular characterization, tissue distribution and differential nutritional regulation of putative Elovl5 elongase in silver barb (*Puntius gonionotus*). *Comp Biochem Physiol Part B Biochem Mol Biol* 217:27–39
- Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52:7182–7190
- Ren H, Yu J, Xu P, Tang Y (2012) Influence of dietary fatty acids on muscle fatty acid composition and expression levels of D6 desaturase-like and Elovl5-like elongase in common carp (*Cyprinus carpio* var. Jian). *Comp Biochem Physiol Part B Biochem Mol Biol* 163:184–192
- Russell FD, Bürgin-Maunders CS (2012) Distinguishing health benefits of eicosapentaenoic and docosahexaenoic acids. *Mar Drugs* 10:2535–2559
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Steenfeldt S et al (2015) Current status of Eurasian percid fishes aquaculture. In: Kestemont P, Dabrowski K, Summerfelt RC (eds) *Biology and culture of percid fishes*. Springer, Dordrecht, pp 817–841
- Strobel C, Jahreis G, Kuhnt K (2012) Survey of n-3 and n-6 polyunsaturated fatty acids in fish and fish products. *Lipids Health Dis* 11:144
- Tacon AGJ et al (2012) Aquaculture feeds: addressing the long-term sustainability of the sector. In: *Farming waters people food. Proceedings of the global conference on aquaculture 2010*, FAO/NACA, Thailand & Rome, pp 193–231
- Takahashi O, Kohno Y, Nishio M (2010) Relevance of weak hydrogen bonds in the conformation of organic compounds and bioconjugates: evidence from recent experimental data and high-level ab initio MO calculations. *Chem Rev* 110:6049–6076
- Tan SH, Chung HH, Shu-Chien AC (2010) Distinct developmental expression of two elongase family members in zebrafish. *Biochem Biophys Res Commun* 393:397–403
- Tocher DR (2015) Omega-3 long-chain polyunsaturated fatty acids and aquaculture in perspective. *Aquaculture* 449:94–107
- Tu WC et al (2012a) An alternative n-3 fatty acid elongation pathway utilising 18:3n-3 in barramundi (*Lates calcarifer*). *Biochem Biophys Res Commun* 423:176–182
- Tu WC et al (2012b) Barramundi (*Lates calcarifer*) desaturase with $\Delta 6/\Delta 8$ dual activities. *Biotechnol Lett* 34:1283–1296
- Tvrđik P et al (2000) Role of a new mammalian gene family in the biosynthesis of very long chain fatty acids and sphingolipids. *J Cell Biol* 149:707–717
- Vrinten PL, Hoffman T, Bauer J, Qiu X (2010) Specific protein regions influence substrate specificity and product length in

- polyunsaturated fatty acid condensing enzymes. *Biochemistry* 49:3879–3886
- Walls J, Sinclair L, Finlay D (2016) Nutrient sensing, signal transduction and immune responses. *Semin Immunol* 28:396–407
- Wang S et al (2014) Investigating long-chain polyunsaturated fatty acid biosynthesis in teleost fish: functional characterization of fatty acyl desaturase (Fads2) and Elovl5 elongase in the catadromous species, Japanese eel *Anguilla japonica*. *Aquaculture* 434:57–65
- Xue X et al (2014) Characterization of the fatty acyl elongase (elovl) gene family, and hepatic elovl and delta-6 fatty acyl desaturase transcript expression and fatty acid responses to diets containing camelina oil in Atlantic cod (*Gadus morhua*). *Comp Biochem Physiol Part B Biochem Mol Biol* 175:9–22
- Yeagle PL (2014) *Lipids*. Wiley, New York, pp 1–9