Progress in Lipid Research 62 (2016) 25–40

Contents lists available at ScienceDirect

Progress in Lipid Research

journal homepage: www.elsevier.com/locate/plipres

Review

Long-chain polyunsaturated fatty acid biosynthesis in chordates: Insights into the evolution of Fads and Elovl gene repertoire

L. Filipe C. Castro [a,b], Douglas R. Tocher [c], Oscar Monroig [c,*]

[a] CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal
[b] Biology Department, FCUP, University of Porto, Portugal
[c] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 28 December 2015
Accepted 1 January 2016
Available online 6 January 2016

Keywords:
Chordates
Elongation of very long-chain fatty acid protein
Evolution
Front-end desaturase
Long-chain polyunsaturated fatty acids

ABSTRACT

Long-chain polyunsaturated fatty acids (LC-PUFA) are major components of complex lipid molecules and are also involved in numerous critical biological processes. Studies conducted mainly in vertebrates have demonstrated that LC-PUFA can be biosynthesized through the concerted action of two sets of enzymes, namely fatty acyl desaturases (Fads) and elongation of very long-chain fatty acid (Elovl) proteins. While LC-PUFA research is a thriving field, mainly focused on human health, an integrated view regarding the evolution of LC-PUFA biosynthetic genetic machinery in chordates is yet to be produced. Particularly important is to understand whether lineage specific life history trajectories, as well as major biological transitions, or particular genomic processes such as genome duplications have impacted the evolution of LC-PUFA biosynthetic pathways. Here we review the gene repertoire of Fads and Elovl in chordate genomes and the diversity of substrate specificities acquired during evolution. We take advantage of the magnitude of genomic and functional data to show that combination duplication processes and functional plasticity have generated a wide diversity of physiological capacities in extant lineages. A clear evolutionary framework is provided, which will be instrumental for the full clarification of functional capacities between the various vertebrate groups.

Contents

Abbreviations: aa, amino acid; ACP, acyl carrier protein; ALA, α-linolenic acid (18:3n-3); ARA, arachidonic acid (20:4n-6); BHT, butylated hydroxytoluene; cDNA, complementary DNA; CoA, Coenzyme A; DHA, docosahexaenoic acid (22:6n-3); ELOVL, elongation of very long-chain fatty acid protein; EPA, eicosapentaenoic acid (20:5n-3); ER, endoplasmic reticulum; FACES, fatty acid chain elongation system; FADS, fatty acyl desaturase; FAE1, fatty acid elongase 1; FAS, fatty acid synthase; HADC, β-hydroxyacyl-CoA dehydrase; KAR, β-ketoacyl-CoA reductase; KCS, β-ketoacyl-CoA synthase; LA, linoleic acid (18:2n-6); LC-PUFA, long-chain (C_{20-24}) polyunsaturated fatty acids; ORF, open reading frame; PKS, polyketide synthase; PUFA, Polyunsaturated fatty acid; SCD, stearoyl-CoA desaturase; TER, trans-2-enoyl-CoA reductase; VLC-PUFA, very long-chain (>C_{24}) polyunsaturated fatty acid; WGD, whole genome duplication.

* Corresponding author at: Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK.
E-mail address: oscar.monroig@stir.ac.uk (O. Monroig).

http://dx.doi.org/10.1016/j.plipres.2016.01.001
0163-7827/

1. Introduction

Unlike proteins and carbohydrates, that generally have structures based on long chains (polymers) of amino acid (polypeptides) or sugar (polysaccharides) residues, lipids are a much more diverse range of compounds with considerable variations in structure. However, although lipids do not have "building blocks" as such, fatty acids come closest, being components of many lipid classes including acylglycerols (glycerides and phosphoglycerides) and sphingolipids. In complex lipids, fatty acids are esterified to alcohol or amino groups and, as these lipid classes are the predominant forms, fatty acids constitute the bulk of lipid. All fatty acids play important roles in key biological processes including energy supply, structure and functions of biological membranes. Some fatty acids, particularly polyunsaturated fatty acids (PUFA) and their derivatives, are highly biologically active and involved in signalling and the regulation of lipid metabolism, inflammatory response and cell division [1]. Saturated and monounsaturated fatty acids can be biosynthesized by all organisms whereas PUFA generally have to be obtained in the diet of animals although they can be converted to long-chain ($C_{20–24}$) PUFA (LC-PUFA) in some species. This review will focus on two sets of enzymes, fatty acyl desaturases and elongases that participate in the biosynthesis of LC-PUFA in chordates. In particular, we will review the currently available data on the desaturase and elongase gene[1] repertoire present in chordate genomes and the diversity of substrate specificities that have been acquired by the encoded enzyme proteins during evolution. For clarity purposes, we will first provide a description of the fatty acid nomenclature system used in this paper, as well as a definition of the groups of organisms that compose the chordate phylum.

1.1. Nomenclature and structure of fatty acids

A fatty acid is essentially an organic molecule with a carboxylic acid group at the end of an aliphatic chain containing four or more carbons, usually an even number up to 24, although odd-numbered and longer carbon chains are also found (Fig. 1). The aliphatic chain can be "saturated", where all carbon–carbon linkages are single bonds and all other carbon bonds are taken by hydrogen, or "unsaturated", where some carbons are linked by double bonds. Several systems have been used historically for fatty acid nomenclature, but the most commonly used and internationally accepted is that defined by the International Union of Pure and Applied Chemistry (IUPAC) in the Compendium of Chemical Terminology (IUPAC, 1997). In the n-x (or "omega x") system of nomenclature, fatty acids are described by the general formula, C:Dn-x, where C = chain length, D = number of ethylenic/double bonds, and n-x (or ωx) indicates the position of the first double bond relative to the methyl end of the chain. Therefore, in this nomenclature 18:0 represents a saturated fatty acid containing an 18-carbon aliphatic chain with no double bonds, and 18:1n-9 (18:1 ω9) denotes a monounsaturated fatty acid with an 18-carbon aliphatic chain with a single, *cis* double bond 9 carbons from the methyl group. Polyunsaturated fatty acids (PUFA), any fatty acid containing two or more double bonds that are most commonly separated by methylene (CH_2) groups, are represented as in the following example, 20:5n-3 (20:5 ω3), which denotes a 20-carbon aliphatic chain containing five double bonds with the first situated three carbons from the methyl group (Fig. 1). However, in the present review, the main alternative nomenclature, the Δx (delta-x) system, is equally important as this is the one commonly used for specifying activities of the fatty acyl desaturase (Fads) enzymes studied herein. In this nomenclature the double bonds are numbered from the carboxyl end of the molecule and so 20:5n-3 is written as $20:5^{\Delta5,8,11,14,17}$. Thus, a Fads that introduces an ethylenic (double) bond five carbons from the carboxyl end of the aliphatic chain is described as having Δ5 activity. Additionally, fatty acids are often still described using trivial names, often reflecting their main sources, such as palmitic acid (16:0) and oleic acid (18:1n-9) from palm and olive oils, respectively. Semi-systematic names, such as eicosapentaenoic acid (EPA; 20:5n-3) and docosahexaenoic acid (DHA; 22:6n-3), are more useful as they at least indicate the numbers of carbons (e.g. eicosa-, 20) and double bonds (e.g. pentaenoic, 5). It is important to clarify that, while PUFA applies to any fatty acid with at least two double bonds and generally, with chain lengths from 18 carbons or more, we will use the term long-chain PUFA (LC-PUFA) for fatty acids with an aliphatic chain length from C_{20} to C_{24}, and two or more double bonds. Similarly, very long-chain PUFA (VLC-PUFA) refer to PUFA with two or more double bonds, with fatty acyl chains $>C_{24}$.

1.2. Classification of chordates

Metazoans, a classification that refers to animals, are characterized by being multicellular and heterotrophic, possessing epithelial cells and having the ability to produce sperm and eggs. Their overwhelming diversity can be organized into defined groups on the basis of phylogenetic relationships. The first major division separates two clades on the basis of their body plan symmetry: the pre-Bilateria (e.g. sponges, coral and jellyfish) and the Bilateria. Within the latter, two other groups can be distinguished based on early embryonic development features: the protostomes (e.g. molluscs) and the deuterostomes (e.g. sea urchins, mammals). In this review we will focus primarily on the phylum chordates that, together with hemichordates and echinoderms, comprise the deuterostomes. Chordates are subdivided into three subphyla, namely cephalochordates, tunicates and vertebrates (Fig. 2). Cephalochordates and tunicates are usually referred to as "invertebrate chordates" as they lack vertebrae (Fig. 2). Chordates share a number of characteristics including a dorsal hollow nerve cord, pharyngeal slits, postanal tail and a prominent axial notochord. Over the years a consensus over the phylogenetic relationships between these three lineages emerged, with tunicates now considered the sister clade of the

[1] Gene/protein nomenclature: The standard vertebrate gene symbol formatting determines that different conventions apply to name gene/protein in different model organisms including human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*), Carolina anole (*Anolis carolinensis*), frog (*Xenopus laevis* or *X. tropicalis*) and zebrafish (*Danio rerio*). Using as example "Elovl5", the human gene is referred to as "*ELOVL5*" and the predicted protein as "ELOVL5"; for mouse and rat, gene will be named as "*Elovl5*", whereas protein will be "ELOVL5"; for chicken and other birds, gene will be termed as "*ELOVL5*", whereas protein will be "ELOVL5"; for anole and other reptiles, gene will be termed as "*elovl5*", whereas protein will be "ELOVL5"; for frog and other amphibians, gene will be named as "*elovl5*", whereas protein will be "Elovl5"; similarly, for zebrafish and other fish, gene will be named as "*elovl5*", whereas protein will be "Elovl5". For non-vertebrate organisms and agnathans, we have used the same symbols as described for fish ("*elovl5*" for genes and "Elovl5" for proteins).