Comparative Biochemistry and Physiology, Part D, Vol. 5, pp. 45-54 (2010)

# COMPARATIVE GENOMICS AND PROTEOMICS OF VERTEBRATE DIACYLGLYCEROL ACYLTRANSFERASE (*DGAT*), ACYL CoA WAX ALCOHOL ACYLTRANSFERASE (*AWAT*) AND MONOACYLGLYCEROL ACYLTRANSFERASE (*MGAT*)

**Roger S Holmes**

School of Biomolecular and Physical Sciences, Griffith University, Nathan 4111 Brisbane Queensland Australia

Email: r.holmes@griffith.edu.au

Keywords: Diacylglycerol acyltransferase-Monoacylglycerol transferase-Human-Mouse-Opossum-Zebrafish-Genetics-Evolution-X chromosome

Running Head: **Genomics and proteomics of vertebrate acylglycerol acyltransferases**

**ABSTRACT**

BLAT (BLAST-Like Alignment Tool) analyses of the opossum (*Monodelphis domestica*) and zebrafish (*Danio rerio*) genomes were undertaken using amino acid sequences of the acylglycerol acyltransferase (AGAT) superfamily. Evidence is reported for 8 opossum monoacylglycerol acyltransferase-like (*MGAT*) (E.C. 2.3.1.22) and diacylglycerol acyltransferase-like (*DGAT*) (E.C. 2.3.1.20) genes and proteins, including *DGAT1*, *DGAT2*, *DGAT2L6* (DGAT2-like protein 6), *AWAT1* (acyl-CoA wax alcohol acyltransferase 1), *AWAT2*, *MGAT1*, *MGAT2* and *MGAT3*. Three of these genes (*AWAT1*, *AWAT2* and *DGAT2L6*) are closely

localized on the opossum X chromosome. Evidence is also reported for six zebrafish *MGAT*- and *DGAT*-like genes, including two *DGAT1*-like genes, as well as *DGAT2*-, *MGAT1*-, *MGAT2*- and *MGAT3*-like genes and proteins. Predicted primary, secondary and transmembrane structures for the opossum and zebrafish MGAT-, AWAT- and DGAT-like subunits and the intron-exon boundaries for genes encoding these enzymes showed a high degree of similarity with other members of the AGAT superfamily, which play major roles in triacylglyceride (DGAT), diacylglyceride (MGAT) and wax ester (AWAT) biosynthesis. Alignments of predicted opossum, zebrafish and other vertebrate DGAT1, DGAT2, other DGAT2-like and MGAT-like amino acid sequences with known human and mouse enzymes demonstrated conservation of residues which are likely to play key roles in catalysis, lipid binding or in maintaining structure. Phylogeny studies of the human, mouse, opossum, zebrafish and pufferfish MGAT- and DGAT-like enzymes indicated that the common ancestors for these genes predated the appearance of bony fish during vertebrate evolution whereas the *AWAT*- and *DGAT2L6*-like genes may have appeared more recently prior to the appearance of marsupial and eutherian mammals.

## 1. INTRODUCTION

Acylglycerol acyltransferases (AGATs) are predominantly responsible for triglyceride synthesis in the body, via two major pathways: the glycerol phosphate (GP) pathway (Kennedy, 1957) and the monoacylglycerol (MG) pathway (see Coleman and Lee, 2004; Yen et al, 2008). The final step of both pathways involves diacylglycerol and fatty acyl CoA being catalytically converted into triglyceride via two distinct DGAT (diacylglycerol O-acyltransferase, E.C.2.3.1.20) families, DGAT1 and DGAT2 (Cases et al., 1998; Oelkers et al, 1998). The two pathways differ in the mode of synthesis of diacylglycerols prior to catalysis by DGAT1 and/or DGAT2, with the GP pathway involving *de*

*novo* synthesis from glycerol-3-phosphate and fatty acyl CoA, whereas the MG pathway uses partially hydrolyzed monoacylglycerols and fatty acyl CoA. This penultimate step in the MG pathway is catalyzed by enzymes encoded by the monoacylglycerol acyltransferase (*MGAT*; 2-acylglycerol O-acyltransferase; EC 2.3.1.22) gene sub-family, for which three *MGAT* (also designated as *MGOT* and *MOGAT*) genes have been reported in humans (Yen et al., 2002; 2003; Cheng et al., 2003). The human *DGAT2*-like gene family comprises at least seven members, including *DGAT2* (Cases et al., 2001), *MGAT1* (Yen et al., 2002), *MGAT2* (Yen and Farese, 2003), *MGAT3* (Cheng et al., 2003), and three genes located together on the X chromosome (*DGAT2L6*, *AWAT1* and *AWAT2*) (Turkish et al., 2005). The latter two genes encode enzymes with acyl CoA wax alcohol acyltransferase (AWAT) activity, and are specifically expressed in sebocytes (skin), and play a role in preventing surface desiccation by the formation of wax esters.

The triglyceride forming pathways are differentially distributed in the body with the GP pathway being widely distributed and responsible for triglyceride synthesis in most tissues of the body. The MG pathway however is predominantly localized in specific cell types, including enterocytes (intestine), hepatocytes (liver) and adipocytes (adipose tissue), where large amounts of triglycerides are synthesized or stored. In particular, MG is the major pathway of the small intestine, where partially hydrolyzed fats are used to synthesize triglycerides following lipid ingestion (see reviews in Coleman and Lee, 2004; Yen et al, 2008).

This study describes the predicted sequences, structures and phylogeny of diacylglycerol acyltransferase-like (*DGAT*) and monoacylglycerol acyltransferase-like (*MGAT*) genes and enzymes from eutherian (human and mouse) and marsupial (opossum) mammals and from a bony fish (zebrafish)

species. Computational methods were used to predict the primary, secondary and transmembrane structures for these enzymes, as well as gene locations, exonic structures and sequences for *MGAT*- and *DGAT*-like genes, using published data from genome sequences. Predictions of MGAT- and DGAT-like enzyme sequences from a wider range of vertebrates were also used to examine the phylogeny and evolution of these genes and to identify conserved amino acid residues, including likely candidates for the active sites and substrate binding regions for these enzymes.

## 2. MATERIALS AND METHODS

### 2.1 *DGAT* and *MGAT* Gene and Enzyme Identification.

BLAT (BLAST-Like Alignment Tool) studies were undertaken using the UC Santa Cruz web browser [http://genome.ucsc.edu/cgi-bin/hgBlat] (Altschul et al., 1990; Kent et al., 2002) with the default settings. UniProtKB/Swiss-Prot database derived amino acid sequences for human, mouse and zebrafish DGAT-like and MGAT-like enzymes [http://au.expasy.org], GenBank derived amino acid sequences for zebra fish DGAT-like and MGAT-like enzymes [http://www.ncbi.nlm.nih.gov/Genbank/], and predicted NCBI sequences for opossum genes and enzymes [http://www.ncbi.nlm.nih.gov/] (see Table 1) were used to interrogate vertebrate genome sequences, particularly human (*Homo sapiens*) and mouse (*Mus musculus*) (MGC Project Team, 2004), opossum (*Monodelphis domestica*) (Mikkelson et al., 2007) and zebrafish (*Danio rerio*) genomes (http://www.ensembl.org/Danio_rerio/Info/Index). Several other vertebrate genomes were also interrogated to examine DGAT-like and MGAT-like amino acid sequences for conserved residues that may play key roles in catalysis, substrate binding and maintaining enzyme structure and function (see Supplementary Table). Gene locations, predicted gene structures and

protein subunit sequences were observed for each of the predicted *DGAT*-like and *MGAT*-like genes and proteins. Structures for the major human DGAT, AWAT and MGAT isoforms (splicing variants) were obtained using the AceView website to interrogate the database of human mRNA sequences (Thierry-Mieg and Thierry-Mieg, 2006):

(http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human)

**2.2 Predicted Structures and Properties for Vertebrate DGAT-like and MGAT-like Subunits.**

Predicted secondary and transmembrane structures for mammalian DGAT-like and MGAT-like subunits were obtained using the SWISS MODEL web tools (Schwede et al., 2003) (http://swissmodel.expasy.org/workspace) and the transmembrane prediction server (http://www.cbs.dtu.dk/services/TMHMM-2.0/) provided by the Center for Biological Sequence Analysis of the Technical University of Denmark (von Heijne, 1992; Tusnady and Simon, 2001). Theoretical isoelectric points and molecular weights for MGAT and DGAT subunits were obtained using Expasy web tools: (http://au.expasy.org/tools/pi_tool.html).

**2.3 Alignment of Conserved Vertebrate DGAT-like and MGAT-like Residues.**

Alignments of vertebrate DGAT-like and MGAT-like sequences were undertaken using a ClustalW-technique (http://www.ebi.ac.uk/clustalw/) (Chenna et al., 2004) and previously reported sequences for the human, mouse, zebrafish and other sequences obtained from BLAT and BLAST analyses following analyses of NCBI databases (http://www.ncbi.nlm.nih.gov/)(Table 1; see Supplementary Table). Predictions of key DGAT-like and MGAT-like amino acid residues were obtained from previous biochemical analyses of mammalian DGAT1

(Cases et al., 1998; Guo et al., 2001) and from the conserved sequences observed for the human, mouse, opossum, zebrafish and other vertebrate enzymes.

### 2.4 Phylogenetic Studies and Sequence Divergence.

Phylogenetic trees were constructed using a ClustalW-derived amino acid alignment of DGAT-like and MGAT-like protein sequences, obtained with default settings and corrected for multiple substitutions (http://www.ebi.ac.uk/clustalw/) (Chenna et al., 2003). Percentage sequence identities for the DGAT1 and DGAT2-like family subunits examined were determined using the SIM-Alignment tool for Protein Sequences (http://au.expasy.org/tools/simprot.html) (Schwede et al., 2003).

## 3. RESULTS AND DISCUSSION

### 3.1 Alignments of opossum and zebrafish DGAT1 amino acid sequences with human and mouse DGAT1. Evidence for two zebrafish *DGAT1*-like genes.

The deduced amino acid sequences of a predicted opossum DGAT1 subunit and two predicted forms of zebrafish DGAT1 subunits (designated as DGAT1A and DGAT1B) are shown in Figure 1 together with the previously reported sequences for human and mouse DGAT1 (Oelkers et al., 1998; Buhman et al., 2002) (see Table 1). The enzymes contained 488-505 amino acid residues in sequence and exhibited predicted MWs of ~ 55-58 kDa's and high theoretical isoelectric points (pI values of 9.4-9.6) (Table 1). The alignments showed high levels of sequence identities for human and mouse (85%) and for human and opossum (78%) DGAT1 but lower levels of identity for the mammalian sequences with zebrafish DGAT1A and DGAT1B sequences (59-61%) (Supplementary Table 2). Regions of high

sequence identity were observed for the mammalian and zebrafish DGAT1 sequences, with the exception of the N- and C-termini and residues 229-240 for the human DGAT1 sequence (Figure 1) where more diverse sequences were observed. All of the DGAT1 sequences examined were distinct from the DGAT2-like and MGAT-like sequences with <14% identity. This suggests that these enzymes are products of the same *DGAT* gene family (*DGAT1*) which is separate from the *DGAT2* family. The amino acid sequences observed for zebrafish DGAT1A and DGAT1B were ~63% identical (Figure 1; Supplementary Table 2), and the genes encoding these enzymes (*DGAT1A* and *DGAT1B*) were localized on different zebrafish chromosomes (16 and 19 respectively) (see Table 1). This supports the presence of two *DGAT1*-like genes in this species.

### 3.2 Alignments of opossum and zebrafish MGAT-like and DGAT2-like amino acid sequences with the corresponding human and mouse enzymes.

Predicted opossum DGAT2-like subunits were designated DGAT2, AWAT1, AWAT2, DGAT2L6, MGAT1, MGAT2 and MGAT3 following alignment of deduced amino acid sequences with the corresponding human enzymes (Figure 2). High levels of sequence identity with the human enzymes were observed in each case: DGAT2 (92%); AWAT1 (71%); AWAT2 (71%); DGAT2L6 (63%); MGAT1 (69%); MGAT2 (73%); and MGAT3 (72%)(Supplementary Table 2). A significant level of sequence identity was also observed between these opossum enzymes (≥42%), and it is likely that they form part of a single gene family, previously designated as *DGAT2*-like (*DGAT2L*) proteins for the human enzymes (Turkish et al., 2005). Human and opossum DGAT2 sequences were longer (388 and 366 amino acid residues, respectively) than those for human and opossum AWAT1 (328 and 332), AWAT2 (333 and 335), DGAT2L6 (337 and 328), MGAT1 (334 and 335), MGAT2 (both 334 residues) and MGAT3 (341 and 327) sequences, as a result of extended N-termini for the human and opossum DGAT2 sequences (Figure 2). All of the

opossum DGAT2-like subunits exhibited high theoretical isoelectric points (pI values of 8.8-9.8) (see Table 1).

Predicted DGAT2-like subunit sequences are also reported for zebrafish, although only four enzymes were deduced from the BLAT analyses of the zebrafish genome: DGAT2 (69% identity with human DGAT2); MGAT1 (56% identity with human MGAT1); MGAT2 (63% identity with human MGAT2); and MGAT3 (50% identity with human MGAT3) (Table 1; Figure 2; Supplementary Table 2). Zebrafish DGAT2 also exhibited a longer amino acid sequence (361 residues) with an extended N-terminus, in comparison with the other zebrafish DGAT2-like enzymes: MGAT1, MGAT2 and MGAT3 (334, 330 and 336 residues, respectively). All of the zebrafish DGAT2-like subunits exhibited high isoelectric points (pI values of 8.5-9.6), which are similar to those reported for the human, mouse and opossum enzymes (Table 1).

### 3.3 Predicted secondary and transmembrane structures for human, mouse, opossum and zebrafish DGAT1 and DGAT2-like subunits.

Predicted secondary and transmembrane (TrM) structures for human, mouse, opossum and zebrafish DGAT1 subunits are compared in Figures 1 and 3. Similar α-helix and predicted TrM structures were observed for all of the vertebrate DGAT1 subunits examined, which were almost devoid of β-sheet structures. For each vertebrate DGAT1 subunit examined, nine high probability regions for TrM sequences were observed, each located in similar positions with those previously predicted for human DGAT1 (Oelkers et al., 1998). This supports the location of vertebrate DGAT1 subunits in the endoplasmic reticulum (ER), which has been previously reported for human DGAT1 (Stone et al., 2006). Figures 1 and 3 also compare the predicted topology of the DGAT1 TrM structures, showing an extended N-terminus located within the ER lumen (amino acid residues 1-103 for human DGAT1), as were residues 151-165

(between TrMs 2 and 3); an extended sequence located in the center of the subunit (residues 211-281) (between TrMs 4 and 5); and residues 353-406 between TrMs 6 and 7. This topology was retained for all of the vertebrate DGAT1 subunits examined and suggests major functional significance for the microlocalization of this enzyme within vertebrate ER membranes.

DGAT2-like subunits are also strongly membrane bound due to their highly hydrophobic character and have been difficult to purify and characterize (see Turkish et al., 2005; Yen et al, 2008). Transmembrane (TrM) predictions for the human, opossum and zebrafish DGAT2-like subunits (Figures 2 and 3) revealed distinct structures in comparison with those observed for the vertebrate DGAT1 subunits. All of the DGAT2-like subunits showed one or two (in proximal locations) predicted TrM(s) near the N-terminus, commencing within the lumen of the ER for the N-terminus and then extending into the cytosol following the transmembrane bound structure (human MGAT3 was an exception). These included the AWAT-like subunits (AWAT1 and AWAT2), for which human and opossum AWAT2 exhibited a second predicted TrM structure near the center of the subunit (residues 130-150 for human AWAT2). This second predicted TrM AWAT2 structure allows for an intermediate segment to extend into the cytosol (residues 58-129 for human AWAT2) prior to the location in the lumen following the second TrM structure. These differences in TrM structures may play major roles in contributing to the separate roles for these enzymes *in vivo*.

### 3.4 Identification of Key Residues for Vertebrate DGAT1 and DGAT2-like Enzyme Subunits

Comparisons of predicted mouse, opossum and zebrafish DGAT1 sequences with the human DGAT1 sequence enabled identification of key residues which

may contribute to catalysis and function (Figure 1; Table 2). Likely active site residues for human DGAT1, previously proposed by Oelkers and coworkers (1998), included Asn378 and His415 (catalytic residues) and a heptapeptide sequence (Phe360-Tyr361-Arg362-Asp363-Trp364-Trp365-Asn366), recognized as a potential fatty acyl CoA binding site. Evidence for identifying these residues was based on similarities with the active site residues for human acyl CoA cholesterol acyltransferase (ACAT) and other members of the membrane bound O-acyltransferase (MBOAT) superfamily (see Hoffmann, 2000; Guo et al., 2001; Chang et al., 2001). The putative catalytic His415 for vertebrate DGAT1 was also predicted within a transmembrane sequence (TrM 7), which is similar to the active site positioning for ACAT, which led Chang and coworkers (2001) to propose that the catalytic transfer of the acyl group takes place within the lipid bilayer of the ER. These 'putative' active site residues have been retained for all mammalian, amphibian and zebrafish DGAT1 subunits examined (Figure 1; Supplementary Table 1).

Comparisons were also conducted of mammalian, opossum, amphibian (Xenopus) and zebrafish DGAT2-like sequences with the previously reported sequences for human and mouse DGAT2 and related family members, AWAT1, AWAT2 and DGAT2L6 (Cases et al., 1998; 2001; Oelkers et al., 1998; Turkish et al., 2005), together with MGAT1 (Yen et al., 2002), MGAT2 (Lardizibal et al., 2001; Yen and Farese, 2003; Lockwood et al., 2003; Cao et al., 2003) and MGAT3 (Cheng et al., 2003) (Figure 2; Table 2). A likely neutral lipid octapeptide binding domain was identified at residues 80-87 (human DGAT2 numbers used) [Phe-Leu-X-Leu-X-X-X-n], where n is a nonpolar amino acid (Table 4). This is consistent with a lipid binding domain previously described for human DGAT2 by Yen and coworkers (2008) and also reported for several enzymes which bind or metabolize neutral lipids, such as hormone-sensitive lipase, cholesterol ester transfer protein, lecithin:cholesterol

acyltransferase, cholesterol 7α-hydroxylase, cholesterol esterase and triglyceride hydrolase (Au-Young and Fielding, 1992; Alam et al., 2006). This sequence is also similar to the consensus sequences described here for vertebrate DGAT2, mammalian DGAT2L6 and for two members of the MGAT-subfamily, MGAT1 and MGAT2 (Table 2). Mammalian AWAT1 and AWAT2 and vertebrate MGAT3 aligned octapeptide consensus sequences, however, do not match the neutral binding domain sequence, and these may reflect differences in lipid binding specificities for the AWAT and MGAT3 enzymes. A tetrapeptide sequence (His-Pro-His-Gly; residues 161-164 for human DGAT2) was conserved for all vertebrate DGAT2-like sequences examined (Figure 2; Table 2), which may serve as a catalytic site for this family of enzymes, as site-directed mutations for this region reduce DGAT2 activity (Stone et al., 2006). Several other vertebrate DGAT2-like amino acid residues have also been conserved for all family members, although their specific roles remain to be identified (Figure 2).

### 3.5 Predicted gene locations and exonic structures for human, mouse, opossum and zebrafish *DGAT1*-like and *DGAT2*-like genes.

Table 1 summarizes the predicted locations for human, mouse, opossum and zebrafish *DGAT1* and *DGAT2*-like genes examined, including three closely located mammalian *DGAT2-like* genes (*AWAT1*, *AWAT2* and *DGAT2L6*) on the X-chromosome; and two closely localized autosomal *DGAT2*-like genes (*DGAT2* and *MGAT2*) in humans (chromosome 11), mice (chromosome 7), opossum (chromosome 4) and zebrafish (chromosome 10). *DGAT1*, *MGAT1* and *MGAT3* (*MGAT3* not found in mice and rats) were also identified and localized elsewhere on the genomes examined (Table 1; see Supplementary Table 1). Two *DGAT1*-like zebrafish genes (designated *DGAT1A* and *DGAT1B*) were located on separate chromosomes (16 and 19, respectively) (Table 1). These BLAT interrogations of the respective

genomes used reported sequences for human DGAT1 (Oelkers et al., 1998), AWAT1, AWAT2 and DGAT2L6 (Turkish et al., 2005) DGAT2 (Cases et al., 2001), MGAT1 (Yen et al., 2002), MGAT2 (Yen and Farese, 2003) and MGAT3 (Cheng et al., 2003) and the UC Santa Cruz Web Browser (Kent et al., 2002) ([http://genome.ucsc.edu/cgi-bin/hgBlat). The human, mouse, opossum and zebrafish *DGAT1* and *MGAT3* genes, human and mouse *AWAT2* genes, mouse, opossum and zebrafish *DGAT2* genes, mouse, opossum and zebrafish *MGAT2* genes, as well as opossum *AWAT1* and *DGAT2L6* genes were all predicted for transcription on the negative strand, whereas others were transcribed on the positive strand (Table 1).

Predicted exonic start sites for the vertebrate *DGAT1 and DGAT2-like* genes were examined with vertebrate *DGAT1* genes containing 17 exons, which are located in similar or identical positions to the 17 exons identified for human *DAGT1* (Oelkers et al., 1998) (Figure 1). Vertebrate *DGAT2*-like genes contained 8 exons for the vertebrate *DGAT2* gene; 6 exons for vertebrate *MGAT1* and *MGAT2* genes; and 7 exons for vertebrate *MGAT3* genes (Table 1; Figure 2). Mammalian *AWAT1*, *AWAT2* and *DGAT2L6* genes also exhibited 7 predicted exonic start sites in similar positions to the human, mouse and opossum genes examined (Table 1; Figure 2). Figure 4 summarizes the comparative exonic and intronic structures for the major human DGAT1 and DGAT2-like isoforms, illustrating the distinctive nature and sizes for these genes and their transcripts for the human genome.

### 3.6 Phylogeny and divergence of vertebrate DGAT1 and DGAT2-like sequences.

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of human, mouse, opossum and zebrafish DGAT1 and DGAT2-like amino acid sequences. The phylogram clustered into two major groups overall,

corresponding to DGAT1 and DGAT2-like proteins and genes. The latter group was further divided into 7 groups, including mammalian AWAT1, AWAT2 and DGAT2L6, and vertebrate DGAT2, MGAT1, MGAT2 and MGAT3, which is consistent with the nomenclature proposed for these enzymes. The tandem locations on the X-chromosome for the three mammalian *DGAT2*-like genes (*AWAT1*, *AWAT2* and *DGAT2L6*) and their apparent absence from chicken, frog and zebrafish genomes (Table 1; Supplementary Table 1) indicate that these genes may have arisen from successive duplications and translocations of an ancestral *DGAT2*-like gene to the X chromosome, prior to the common ancestor for marsupial and eutherian mammals. In addition, the tandem locations and levels of sequence identities observed for the vertebrate *DGAT2* and *MGAT2* genes (~45% identical) (Supplementary Table 2) is suggestive of an earlier common origin for these genes by tandem duplication of an ancestral *DGAT2*-like gene prior to the appearance of bony fish. The reasons for the retention of proximal locations of *DGAT2* and *MGAT2* genes over > 500 million years of vertebrate evolution remain to be established.

These studies of the comparative sequences of *DGAT1* and *DGAT2*-like genes and proteins derived from mammalian and bony fish genomes strongly suggest that these two gene families have evolved separately during vertebrate evolution. This conclusion is based on the low levels of sequence identity (<14%) and the distinct phylogenetic relationships observed (Table 2; Figure 5). A single DGAT1-like gene and protein was observed for most vertebrates, including representative mammals, birds (chicken) and amphibians (frog), whereas two distinct zebrafish and pufferfish *DGAT1*-like genes (with similar sequences (63% identical for zebrafish DGAT1-like proteins) with the genes located on separate chromosomes) (Table 1; Supplementary Table 1). This is suggestive of a gene duplication event giving rise to these genes within the ancestor for these bony fish species while other vertebrates have

apparently retained only one *DGAT1*-like gene (Table 1). The vertebrate *DGAT2*-like gene family, however, exists in least 4 separate genes in both bony fish genomes examined, which have been provisionally designated as *DGAT2*, *MGAT1*, *MGAT2* and *MGAT3*, based on the clustering and higher percentages of identity of these sequences with the corresponding mammalian sequences (Figure 5; Supplementary Table 2). It is likely that these genes have appeared early in vertebrate evolution, certainly prior to the appearance of bonyfish, which occurred > 500 million years ago (Kumar and Hedges, 1998). The three mammalian X-linked genes (*AWAT1*; *AWAT2* and *DGAT2L6*), however, were observed only in mammals (Table 1; Supplementary Table 1) and these may have resulted from further gene duplication events prior to the appearance of the common ancestor for marsupial and eutherian mammals (~ 180 million years ago) (Woodburne et al., 2003).

### 3.7 Functions of DGAT1 and DGAT2-like Enzymes

Human and mouse DGAT1 and DGAT2 catalyze the final step in triglyceride biosynthesis by catalyzing the joining of diacylglycerol with fatty acyl CoA. The enzymes are differentially distributed in mammalian tissues with DGAT1 occurring in most tissues, whereas DGAT2 exhibits highest activity in the intestine, liver and adipose tissue, where triglycerides are required for major lipid biosynthesis for lipid digestion and transport, storage and lipoprotein synthesis (see Yen et al., 2008). Knock-out mice for both enzymes have shown that *DGAT1*-deficient mice are viable but show major reductions in tissue triglyceride content but have normal plasma triglceride levels, whereas *DGAT2*-deficient mice die within hours of birth, and have severely reduced levels of triglycerides (Smith et al., 2000; Stone et al., 2004). In addition to the catalysis of triacylglycerol biosynthesis, DGAT1 is also capable of synthesizing diacylglycerols, waxes and retinyl esters (Yen

et al., 2005). It is concluded however that DGAT2 has a more significant role in triglyceride biosynthesis in the body than DGAT1 although significant side effects were also observed for DGAT1-deficient mice, including an increased sensitivity to insulin and leptin, and atrophy of the skin sebaceous glands, resulting in hair loss and reduced formation of wax esters. The high levels of sequence identities for opossum and zebrafish DGAT1 and DGAT2 with the human and mouse sequences, (Figure 1; Supplementary Table 2), indicate that these enzymes serve similar metabolic functions reported for the human and mouse enzymes.

While there are no reports of knockout mice for the other DGAT2-like genes and enzymes, biochemical analyses and tissue expression studies provide clues as to their metabolic roles in the body. The human *AWAT1* and *AWAT2* genes, for example, are differentially expressed within the sebaceous glands of the skin and exhibit kinetic properties which are consistent with specific roles in the formation of wax esters (Turkish et al., 2005). AWAT1 shows a preference for saturated fatty acyl CoAs, as compared with AWAT2, which prefers unsaturated fatty acyl CoAs, reflecting distinct but overlapping roles in the biosynthesis of wax esters, which protect the external surfaces of the body from desiccation.

The *MGAT* gene and enzyme subfamily have also been investigated in terms of their tissue expression profiles and substrate specificities. MGATs preferentially catalyze reactions involving monoacylglycerols and fatty acyl CoAs, forming diacylglycerols, which participate in triglyceride and phospholipid synthesis, which store energy, contribute to lipoprotein and membrane formation and serve as an intracellular signal for protein kinase activation (see Yen et al., 2008). MGAT2 is the major enzyme involved in fat digestion and absorption in the gut, where products of pancreatic lipase hydrolysis of triacylglycerols in the diet, *sn*-2-monacylglycerols and fatty

acids, are used in the resynthesis of triacylglycerols by the intestine (Yen and Farese, 2003). MGAT2 is widely distributed in human and mouse tissues and exhibits high levels of expression within intestinal enterocytes and a preference for monoacylglycerols containing unsaturated fatty acyl CoAs. *MGAT1* is expressed at much lower levels in human tissues, but shows higher expression levels in mouse tissues. Mouse MGAT1 is specific for catalyzing diacylglycerol synthesis and exhibits a preference towards monoacylglycerols containing unsaturated fatty acyl groups, and may play a role in recycling essential fatty acids in the body (Yen et al., 2002). Human *MGAT3* is specifically expressed in the digestive system, with high expression levels in the small intestine and colon, and may serve specific roles in lipid absorption by the digestive tract (Cheng et al., 2003).

## 4. Conclusions

BLAT analyses of the opossum and zebrafish genomes using the amino acid sequences reported for human and mouse DGAT1 and DGAT2-like protein subunits were undertaken to interrogate these genomes. Evidence is reported for an opossum *DGAT1* gene and six *DGAT2*-like genes, including *DGAT2*, *AWAT1*, *AWAT2*, *DGAT2L6*, *MGAT1*, *MGAT2* and *MGAT3* genes. Evidence for two zebrafish and pufferfish *DGAT1-like* genes and four *DGAT2*-like genes (*DGAT2*, *MGAT1*, *MGAT2* and *MGAT3)* is also described. Three of the opossum genes (*AWAT1*, *AWAT2* and *DGAT2L6*) were closely localized on the X chromosome, which is comparable to the locations of these genes in human, mouse and several other mammalian genomes examined (see Supplementary Table). Two other opossum *DGAT2*-like genes (*DGAT2* and *MGAT2*) were also closely located on chromosome 4, for which tandem locations for these genes was observed for other mammalian, amphibian and zebrafish genomes examined (see Supplementary Table 1). The predicted amino acid sequences, secondary and transmembrane structures for the opossum

and zebrafish DGAT1 and DGAT2-like subunits showed a high degree of similarity with the corresponding human and mouse enzymes. Phylogenetic analyses undertaken with human, mouse, opossum, zebrafish and pufferfish DGAT1 and DGAT2-like proteins supported the designation of these enzymes with relevant families and sub-families previously described. It is likely that opossum and zebrafish enzymes perform similar functions to those reported for human and mouse in triacylglycerol (DGAT1, DGAT2 and DGAT2L6 subunits), wax (AWAT1 and AWAT2 for the opossum) and diacylglycerol (MGAT1, MGAT2 and MGAT3) biosynthesis.

**ACKNOWLEDGEMENTS**

I am grateful to Dr Laura Cox and Dr John VandeBerg of the Southwest Foundation for Biomedical Research in San Antonio Texas USA for helpful discussions and advice.

**REFERENCES**

Alam, M., Gilham, D., Vance, D.E., Lehner, R., 2006. Mutation of F417 but not L418 or L420 in the lipid binding domain decreases the activity of triacylglycerol hydrolase, *J. Lipid Res.* **47:** 375-383.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., 1990. Basic local alignment search tool, *J. Mol. Biol*. **215:** 403-410.

Au-Young, J., Fielding, J., 1992. Synthesis and secretion of wild-type and mutant human plasma cholesteryl ester transfer protein in baculovirus-transfected insect cells: the carboxyl-terminal region is required for both lipoprotein binding and catalysis of transfer, *Proc. Natl. Acad. Sci. USA* **89:** 4094-4098.

Buhman, K.K., Smith, S.J., Stone, S.J., Repa, J.J., Knapp, F.F., Burri, B.J., Hamilton, R.L., Abumrad, N.A., Farese, R.V., 2002. DGAT is not essential for

intestinal triacylglycerol absorption or chylomicron synthesis, *J. Biol. Chem.* **277:** 25474-25479.

Cao, J., Lockwood, J., Burn, P., Shi, Y., 2003. Cloning and functional characterization of a mouse intestinal acyl CoA: monoacylglycerol acyltransferase, *J. Biol. Chem.* **278:** 13860-13866.

Cases, S., Smith, S.J., Zheng, Y.W., Myers, H.M., Lear, S.R., Sande, E., Novak, S., Collins, C., Welch, C.B., Lusis, A.J., 1998. Identification of a gene encoding acyl CoA:diacylglycerol transferase, a key enzyme in triacylglycerol synthesis, *Proc. Natl. Acad. Sci. USA* **95:** 13018-13023.

Cases, S., Stone, S.J., Zhou, P., Yen, E., Tow, K.D., Lardizabal, K.D., Voelker, T., Farese, R.V., 2001. Cloning of DGAT2, a second mammalian diacylglycerol acyl transferase, and related family members, *J. Biol. Chem.* **276:** 38870-38876.

Chang, T.Y., Chang, C.C.Y., Lu, X., Lin, S., 2001. Catalysis of ACAT may be completed within the plane of the membrane: a working hypothesis*, J. Lipid Res.* **42:** 1933-1938.

Cheng, D., Nelson, T.C., Chen, J., Walker, S.J., Wardwell-Swanson, J., Meegalla, R., Taub, R., Billheimer, J.T., Ramaker, M., Feder, J.N., 2003. Identification of acyl coenzyme A:monoacylglycerol acyltransferase 3, an intestine specific enzyme implicated in dietary fat absorption, *J. Biol. Chem.* **278:** 13611-13514.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., J.D. Thompson, J.D. Multiple sequence alignment with the Clustal series of programs., 2003. *Nucleic Acids Res.* **31:** 3497-3500.

Coleman, R.A., Lee, D.P., 2004. Enzymes of triacylglycerol synthesis and their regulation. *Prog. Lipid Res.* **43:** 134-176.

Guo, Z., Cromley, D., Billheomer, J.T., Sturley, S.L., 2001. Identification of potential substrate-binding sites in yeast and human acyl-CoA sterol

acyltransferases by mutagenesis of conserved sequences, *J. Lipid. Res.* **42:** 1282-1291.

Hofmann, K., 2000. A superfamily of membrane-bound O-acyl-transferases with implications for Wnt signaling, *Trends Biochem. Sci.* **25:** 111-112.

Kennedy, E., 1957. Metabolism of lipides. *Ann. Rev. Biochem*. **26**: 119-148.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D. The human genome browser at UCSC, (2002). *Genome Res.* **12:** 994-1006.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. (1998). *Nature* **392:** 917-920.

Lardizabal, K.D., Mai, J.T., Wagner, N.W., Wyrick, A., Voelker, T., Hawkins, D.J., 2001. DGAT2 is a new diacylglycerol acyltransferase gene family. Purification, cloning and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacyl-glycerol acyltransferase activity, *J. Biol. Chem.* **276:** 38862-38869.

Lockwood, J.F., Cao, J., Burn, P., Shim, Y., 2003. Human intestinal monoacylglycerol acyltransferase: differential features in tissue expression and activity, *Am J. Physiol.* **285:** E297-E937.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server, *Bioinformatics* **16**: 4044-5.

Mikkelsen, T.S., Wakefield, M.J, Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., Jurka, J., Kamal, M., Mauceli, E., Searle, S.M.J., Sharpe, T., Baker, M.L., Batzer, M.A., Benos, P.V., Belov, K., Clamp, M., Cook, A., Cuff, J., Das, R., Davidow, L., Deakin, J.E.,Fazzari, M.J., Glass, J.L., Grabherr, M., Greally, J.M., Gu, W., Hore, T.A., Huttley, G.A., Kleber, M., Jirtle, R.L., Koin, E., Lee, J.T., Mahony, S., Marra, M.A., Miller, R.D., Nicholls, R.D., Oda, M., Papenfuss, A.T., Parra, Z.E., Pollock, D.D., Ray, D.A., Schein, J.E., Speed, T.P.,

Thompson, K., VandeBerg, J.L., Wade, C.M., Walker, J.A., Waters, P.D., Webber, C., Weidman, J.R., Xie, X., Zody, M.C., Marshall Graves, J.A., Ponting, C.P., Breen, M., Samollow, P.B., ELander, E.S., Lindblad-Toh, J., 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**: 167-177.

Nilsson, M.A., Arnason, U., Spencer, P.B., Janke, A., 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana, *Gene* **340**: 189-196.

Oelkers, P., Behari, A., Cromley, D., Billheimer, J.T., Sturley, S.L., 1998. Characterization of two human genes encoding acyl CoA cholesterol acyltransferase-related enzymes, *J. Biol. Chem.* **273**: 26765-26771.

Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology modeling server., *Nucleic Acids Res*. **31**: 3383-3385.

Smith, S.J., Cases, S., Jensen, D.R., Chen, H.C., Sande, E., Tow, B., Sanan, D.A., Raber, J., Eckel, R.H., Farese, R.V., 2000. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking DGAT, *Nature Genet.* **25**: 87-90.

Stone, S.J., Myers, H., Brown, B.E., Watkins, S.M., Feingold, K.R., Elias, P.M., Farese, R.V., 2004. Lipopenia and skin barrier abnormalities in DGAT2-deficient mice, *J. Biol. Chem.* **279**: 11767-11776.

Stone, S.J., Levin, M., Farese, R.V., 2006. Membrane topology and identification of key functional amino acid residues of murine acyl-CoA: diacylglycerol acyltransferase-2, *J. Biol. Chem.* **281**: 40273-42082.

The MGC Project Team, 2004. The status, quality and expansion of the NIH cDNA project: the mammalian gene collection, *Genome Res.* **14**: 2121-2127.

Thierry-Mieg, D., Thierry-Mieg, J., 2006. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **7**: S12.

Turkish, A.R., Henneberry, A.L., Cromley, D., Padamsee, M., Oelkers, P., Bazzi, H., Christian, A.M., Billheimer, J.T., Sturley, S.L., 2005. Identification of two novel human acyl CoA wax alcohol acyltransferases, *J. Biol. Chem.* **280:** 14755-14764.

Tusnady, G.E., Simon, I., 2001. The HMMTOP transmembrane topology prediction server, *Bioinform.* **17:** 849-850.

von Heijne,G., 1992. Membrane protein structure prediction. *J. Mol. Biol.* **255:** 487–494.

Woodburne, M.O., Rich, T.H., Springer, M.S., 2003. The evolution of tribospheny and the antiquity of mammalian clades, *Mol. Phylogenet. Evol.* **28:** 360-385.

Yen, C-L.E., Farese, R.V., 2003. MGAT2, a monoacylglycerol acyltransferase expressed in the small intestine, *J. Biol. Chem.* **278:** 18532-18537.

Yen, C-L.E., Stone, S.J., Cases, S., Zhou, P., Farese, R.V., 2002. Identification of a gene encoding MGAT1, a monoacylglycerol acyltransferase, *Proc. Natl. Acad. Sci. USA* **99:** 8512-8517.

Yen, C-L.E., Monetti, M., Burri, B.J., Farese, R.V., 2005. The triacylglycerol synthesis enzyme DGAT1 also catalyzes the synthesis of diacyl glycerols, waxes and retinyl esters. *J. Lipid. Res.* **46:** 1502-1511.

Yen, C-L.E., Stone, S.J., Koliwad, S., Harris, C., Farese, R.V., 2008. Glycerolipids. DGAT enzymes and triacylglycerol biosynthesis, *J. Lipid. Res.* **49:** 2283-2301.

| Species | Gene | GenBank ID | UNIPROT ID | NCBI RefSeq ID[1] | Chromosome location (strand) | Gene Size (bps) | Coding Exons | Amino Acids | pI | Subunit MW | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Human | AWAT1 | BC153034 | Q58HT5 | NM_001013579 | X:69,371,271-69,376,865 (+ve) | 5,595 | 7 | 328 | 9.1 | 37,759 | DGT |
| | AWAT2 | BC153070 | Q6E213 | NM_001002254 | X:69,177,117-69,186,513 (-ve) | 9,397 | 7 | 333 | 9.4 | 38,094 | DGT |
| | DGT2L6 | BC172336 | Q6ZPD8 | NM_198512 | X:69,314,061-69,342,278 (+ve) | 28,218 | 7 | 337 | 9.9 | 38,593 | DG |
| | DGAT1 | BC015762 | O75907 | NM_012079 | 8:145,511,028-145,521,107 (-ve) | 10,080 | 17 | 488 | 9.4 | 55,278 | |
| | DGAT2 | BC015234 | Q96PD7 | NM_032564 | 11:75,157,685-75,189,198 (+ve) | 31,514 | 8 | 388 | 9.5 | 43,831 | |
| | MGAT1 | BC146518 | Q96PD6 | NM_058165 | 2:223,244,749-223,282,850 (+ve) | 38,102 | 6 | 334 | 9.5 | 38,630 | MO |
| | MGAT2 | BC103876 | Q3SYC2 | NM_025098 | 11:75,106,582-75,119,976 (+ve) | 13,395 | 6 | 334 | 9.5 | 38,196 | MO |
| | MGAT3 | BC100953 | Q86VF5 | NM_178176 | 7:100,625,950-100,630,855 (-ve) | 4,906 | 7 | 341 | 8.9 | 38,730 | MO |
| Mouse | AWAT1 | BC147680 | A2ADU9 | NP_001074605 | X:97,767,586-97,773,545 (+ve) | 5,549 | 7 | 328 | 8.9 | 37,573 | Dgt |
| | AWAT2 | BC138085 | Q6E1M8 | NP_808414 | X:97,598,128-97,607,874 (-ve) | 9,747 | 7 | 333 | 9.5 | 38,145 | Dgt |
| | DGT2L6 | AK079438 | [2] | XM_001000486 | X:97,720,262-97,740,936 (+ve) | 20,675 | 7 | 337 | 9.8 | 38,362 | Dgt |
| | DGAT1 | BC003717 | Q9Z2A7 | NM_010046 | 15:76,332,669-76,342,211 (-ve) | 9,543 | 17 | 498 | 9.5 | 56,790 | DGT |
| | DGAT2 | BC043447 | Q9DCV3 | NM_026384 | 7:106,303,059-106,331,022 (-ve) | 27,964 | 8 | 388 | 9.5 | 43,770 | DGT |
| | MGAT1 | BC106135 | Q91ZV4 | NM_026713 | 1:78,508,537-78,534,670 (+ve) | 26,134 | 6 | 335 | 8.9 | 38,803 | Mo |
| | MGAT2 | BC052831 | Q80W94 | NM_177448 | 7:106,368,318-106,387,075 (-ve) | 18,758 | 6 | 334 | 9.7 | 38,591 | Mo |
| Opossum | AWAT1 | [2] | [2] | XP_001368551 | X:59,097,964-59,104,042 (-ve) | 6,079 | 7 | 332 | 9 | 38,599 | DGT |
| | AWAT2 | [2] | [2] | ENSMODP5771[1] | X:59,181,618-59,199,570 (+ve) | 17,953 | 7 | 335 | 9.8 | 38,536 | DGT |
| | DGT2L6 | [2] | [2] | XP_001368584 | X:59,116,360-59,125,987 (-ve) | 9,628 | 7 | 328 | 9.3 | 37,141 | DG |
| | DGAT1 | [2] | [2] | XP_001371565 | 3:435,794,358-435,825,910 (-ve) | 31,553 | 17 | 505 | 9.4 | 56,590 | |
| | DGAT2 | [2] | [2] | XP_001364751 | 4:339,142,799-339,188,735 (-ve) | 45,937 | 8 | 366 | 9.4 | 41,552 | |
| | MGAT1 | [2] | [2] | XP_001365685 | 7:228,853,668-228,886,321 (-ve) | 32,654 | 6 | 335 | 8.9 | 39,134 | MO |
| | MGAT2 | [2] | [2] | XP_001364890 | 4:339,284,620-339,315,264 (-ve) | 30,645 | 6 | 334 | 9.6 | 38,391 | MO |
| | MGAT3 | [2] | [2] | XP_001371248 | 2:279,175,446-279,179,682 (-ve) | 4,237 | 7 | 327 | 8.8 | 35,316 | MO |
| Zebrafish | DGAT1A | BC076012 | [2] | NM_001002458 | 16:27,596,419-27,647,463 (-ve) | 51,045 | 17 | 501 | 9.6 | 58,058 | |
| | DGAT1B | BC063070 | Q6P3J0 | NP_956024 | 19:45,373,531-45,395,506 (-ve) | 21,976 | 17 | 499 | 9.5 | 57,152 | |
| | DGAT2 | BC096927 | [2] | NM_001030196 | 10:28,025,496-28,047,075 (-ve) | 21,580 | 8 | 361 | 9.3 | 40,865 | |
| | MGAT1 | BC159110 | [2] | NM_001122623 | 23:40,648,506-40,659,082 (+ve) | 10,577 | 6 | 334 | 9.3 | 37,708 | MO |
| | MGAT2 | BC083237 | [2] | NM_001006083 | 10:28,058,463-28,081,932 (-ve) | 23,470 | 6 | 330 | 9 | 38,013 | MO |
| | MGAT3 | BC129046 | [2] | NM_001008626 | 5:29,806,778-29,812,172 (-ve) | 5,395 | 7 | 336 | 8.5 | 37,801 | MO |

**Table 1: Diacylglycerol acyltransferase (DGAT) and monoacylglycerol acyltransferase (MGAT) genes and enzymes examined**

GenBank mRNA (or cDNA) IDs identify previously reported sequences or predicted sequences (see http://www.ncbi.nlm.nih.gov/Genbank/); UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual DGAT and MGAT enzymes (see http://kr.expasy.org); DGAT and MGAT sequences were provided by the above sources; opossum AWAT, DGAT2L6, DGAT1, DGAT2, MGAT1, MGAT2 and MGAT3 protein sequences were obtained from a blast of the opossum genome using web tools of the National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the corresponding human sequences. Predicted exon/intron locations and gene sizes were obtained by BLAT interrogations of the opossum genome using the predicted DGAT and MGAT sequences and the UC Santa Cruz web tools ( http://genome.ucsc.edu); NCBI refers to database sources from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sites/entrez); [1] refers to the predicted Ensembl sequence observed from the BLAT analysis of opossum AWAT2; [2] not observed; bps: base pairs; pI: theoretical isoelectric point.

| DGAT2-like (Consensus)[5] | Putative Active Site[1] | Neutral Lipid Binding Domain[4] F L X L X X X n[6] |
|---|---|---|
| DGAT2[2] | H161-P162-H163-G164 | **F**80-**L**81-V82-**L**83-G84-V85-A86-**C**87 |
| DGAT2L6[2] | H107-P108-H109-G110 | **F**27-**L**28-V29-**L**30-G31-V32-A33-**C**34 |
| MGAT1[2] | H107-P108-H109-G110 | F27-L28-L29-L30-A31-Q32-V33-C34 |
| MGAT2[2] | H108-P109-H110-G111 | **F**27-**L**28-A29-**L**30-A31-Q32-I33-**C**34 |
| MGAT3[2] | H114-P115-H116-G117 | **F**33-**L**34-F35-M36-G37-P38-F39-F40 |
| AWAT1[2] | H103-P104-H105-G106 | Y22-**L**23-A24-I25-F26-W27-I28-L29 |
| AWAT2[2] | H104-P105-H106-G107 | L26-**L**27-I28-V29-T30-T31-V32-I33 |

| DGAT1 | Putative Active Site[1] | Fatty Acyl CoA Binding Site[3] |
|---|---|---|
| Human | N378 H415 | F360-Y361-R362-E363-W364-W365-N366 |
| Mouse | N389 H426 | F371-Y372-R373-E374-W375-W376-N377 |
| Opossum | N390 H427 | F372-Y373-R374-E375-W376-W377-N378 |
| Zebrafish (1A) | N388 H425 | F370-Y371-R372-E373-W374-W375-N376 |
| Zebrafish (1B) | N387 H424 | F359-Y360-R361-E362-W363-W364-N365 |

**Table 2: Key DGAT1 and DGAT2-like amino acid residues for vertebrate enzymes.**

[1]Identification of predicted active site residues for vertebrate DGAT1 is based on previous studies of Oelkers et al., 1998; prediction of active site residues for DGAT2-like sequences is based on a region containing a conserved tetrapeptide sequence with histidine residues (the active site residue for human DGAT1); [2]DGAT2-like consensus sequences were obtained following alignment of vertebrate sequences for the particular sub-family of DGAT2-like enzymes (sequences used included human, mouse, opossum, zebrafish (see Figure 2) and 9 other vertebrate sequences (see supplementary list of species examined) using the ClustalW technique (see methods); [3]identification of the fatty acyl CoA binding site sequences for vertebrate DGAT1 is based on a previous report (Oelkers et al., 1998); [4] identification of a potential neutral lipid binding domain for vertebrate DGAT2-like sub-families is based on a report from Yen and coworkers (2008).

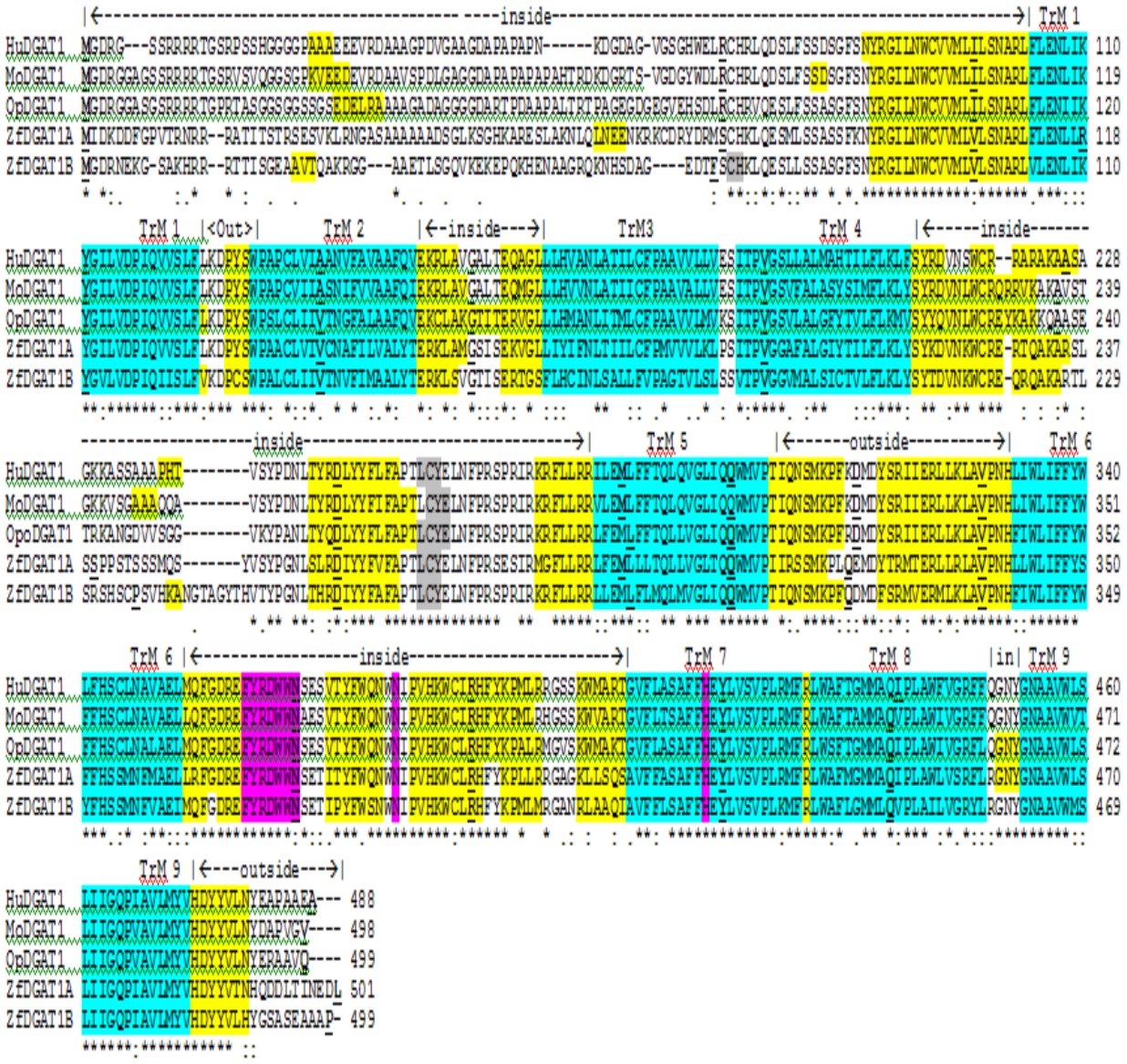


**Figure 1: Amino acid sequence alignments for human, mouse, opossum and zebrafish DGAT1-like sequences.**

See Table 1 for sources of DGAT1 sequences: HuDGAT1: human DGAT1; MoDGAT1: mouse DGAT1; OpDGAT1: opossum DGAT1; ZfDGAT1A: zebrafish DGAT1A; ZfDGAT1B: zebrafish DGAT1B; * shows identical residues; : shows one conservative amino acid substitution; . shows two conservative amino acid substitutions; **underlined bold font** shows known or predicted exon junctions; predicted β-sheet and α-helix secondary structures are shown; key residues (active site H within TrM 7) and fatty acyl CoA binding site (FYRDWWN) are shown; predicted transmembrane structures are shown and numbered (TrM1-TrM9); the predicted topology for the polypeptide chains are represented as inside the endoplasmic reticulum (ER) (the lumen); and outside the ER (cytosol).

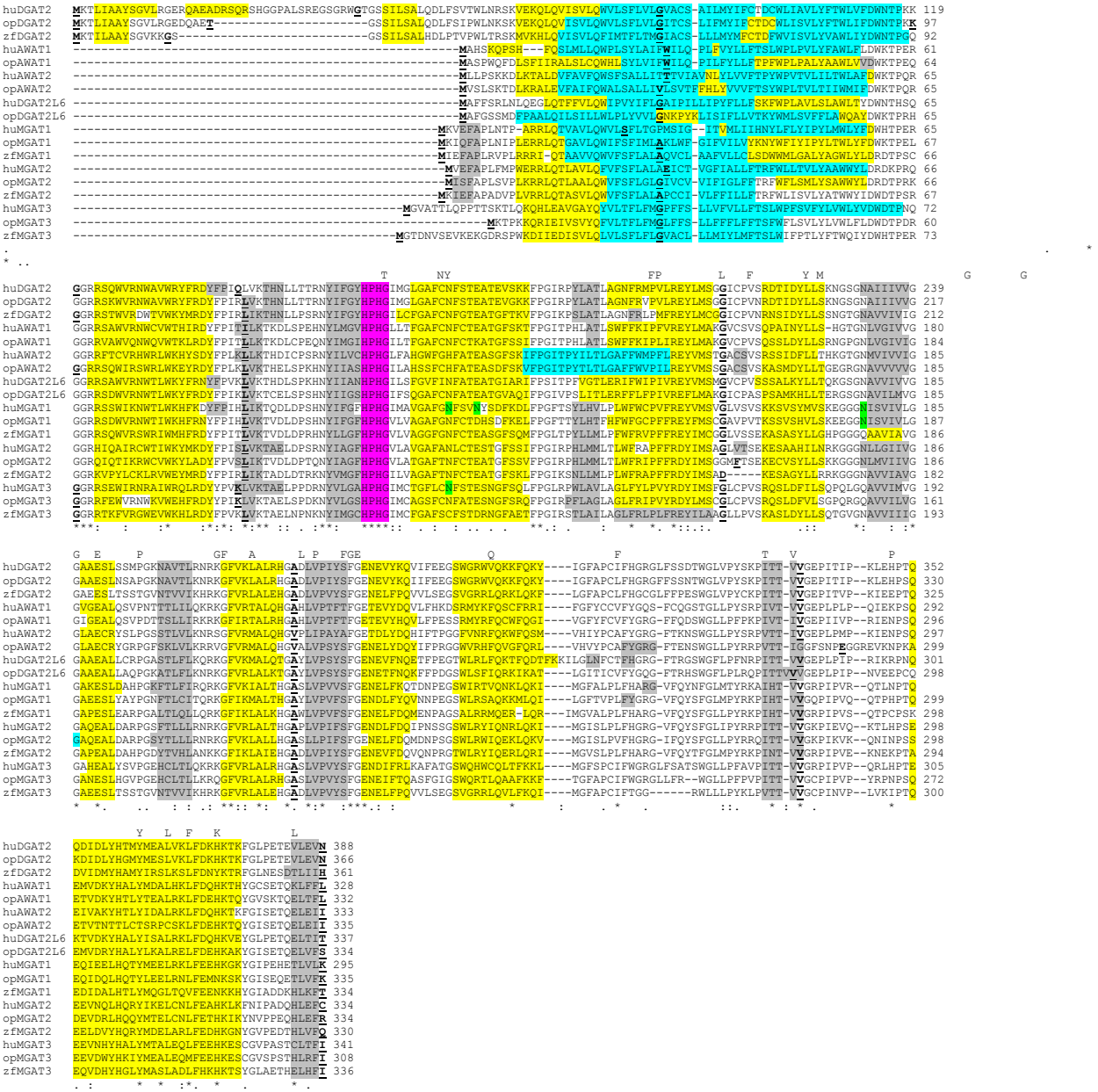

**Figure 2: Amino acid sequence alignments for human, mouse, opossum and zebrafish DGAT2-like sequences.**

See Table 1 for sources of DGAT2-like sequences: huDGAT2: human DGAT2; opDGAT2: opossum DGAT2; zfDGAT2: zebrafish DGAT2; huAWAT1: human AWAT1; opAWAT1: opossum AWAT1; huAWAT2: human AWAT2; opAWAT2: opossum AWAT2; huDGAT2L6: human DGAT2L6; opDGAT2L6: opossum DGAT2L6; huMGAT1: human MGAT1; opMGAT1: opossum MGAT1; zfMGAT1: zebrafish MGAT1; huMGAT2: human MGAT2; opMGAT2: opossum MGAT2; zfMGAT2: zebrafish MGAT2; huMGAT3: human MGAT3; opMGAT3: opossum MGAT3; zfMGAT3: zebrafish MGAT3; * shows identical residues; : shows one conservative amino acid substitution; . shows two conservative amino acid substitutions; **underlined bold font** shows known or predicted exon junctions; predicted β-sheet and α-helix secondary structures are shown; potential active site and neutral lipid binding domain residues are shown eg HPHG (see Table 3); other fully conserved residues observed for all of the DGAT2-like sequences are also shown; predicted transmembrane structures are shown, note the additional transmembrane structure observed for the human and opossum sequences.

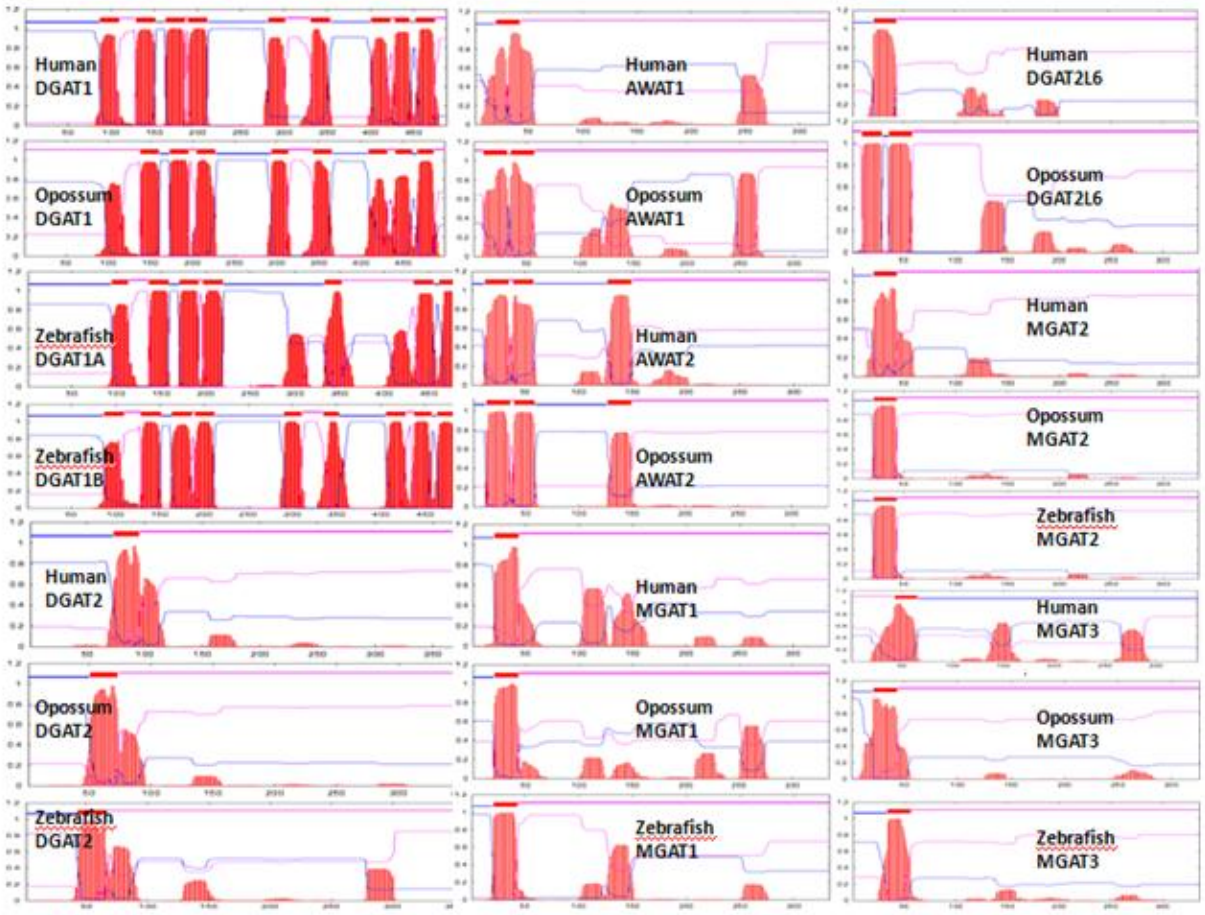


**Figure 3: Predicted transmembrane helices for human, opossum and zebrafish DGAT1-like and DGAT2-like sequences.**

See Table 1 for sources of DGAT1 and DGAT2-like sequences. The TMHMM web tools of the Center for Biological Sequence Analysis, Technical University of Denmark TMHMM Server plots the probability of the ALDH sequence forming a transmembrane helix (0-1) (shown in red for the relevant amino acid sequences) (http://www.cbs.dtu.dk/services/TMHMM/). Nine predicted transmembrane helices are identified for each of the DGAT1-like sequences; in contrast, only one major transmembrane helix (or helices) were observed for DGAT2-like sequences, with the exception of human and opossum AWAT2 sequences, which show a second major transmembrane region; regions of DGAT1 or DGAT2-like sequences predicted to be located inside or outside the membrane are shown in blue and pink, respectively.

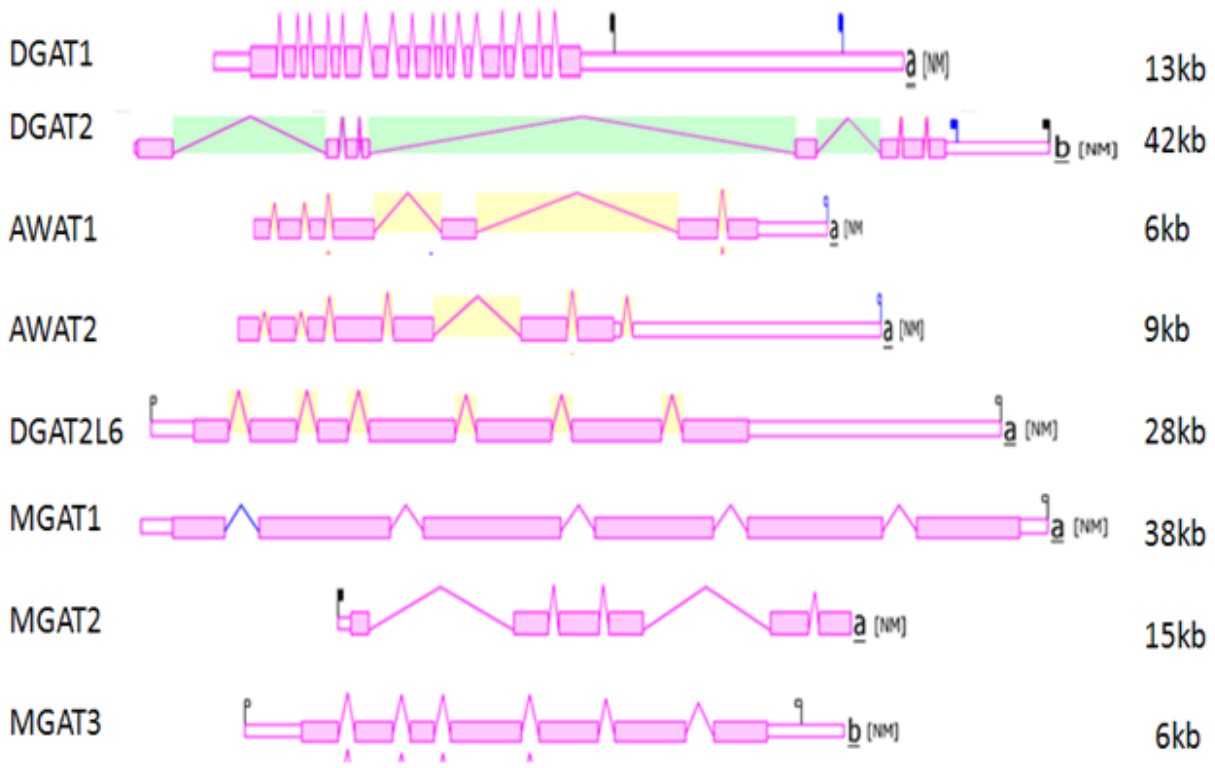


**Figure 4: Gene structures and major transcript isoforms for human DGAT1 and DGAT2-like genes.**

Derived from the AceView website (Thierry-Mieg and Thierry-Mieg, 2006); coding exons are shown as shaded boxes; untranslated exon regions are shown as blank boxes; introns are shown by the lines joining the exons; capped 5'- and validated 3'-ends are also shown. The genes are not drawn to scale but gene sizes are shown as kilobases of DNA. NM refers to validated sequences from NCBI database sources (National Center for Biotechnology Information).

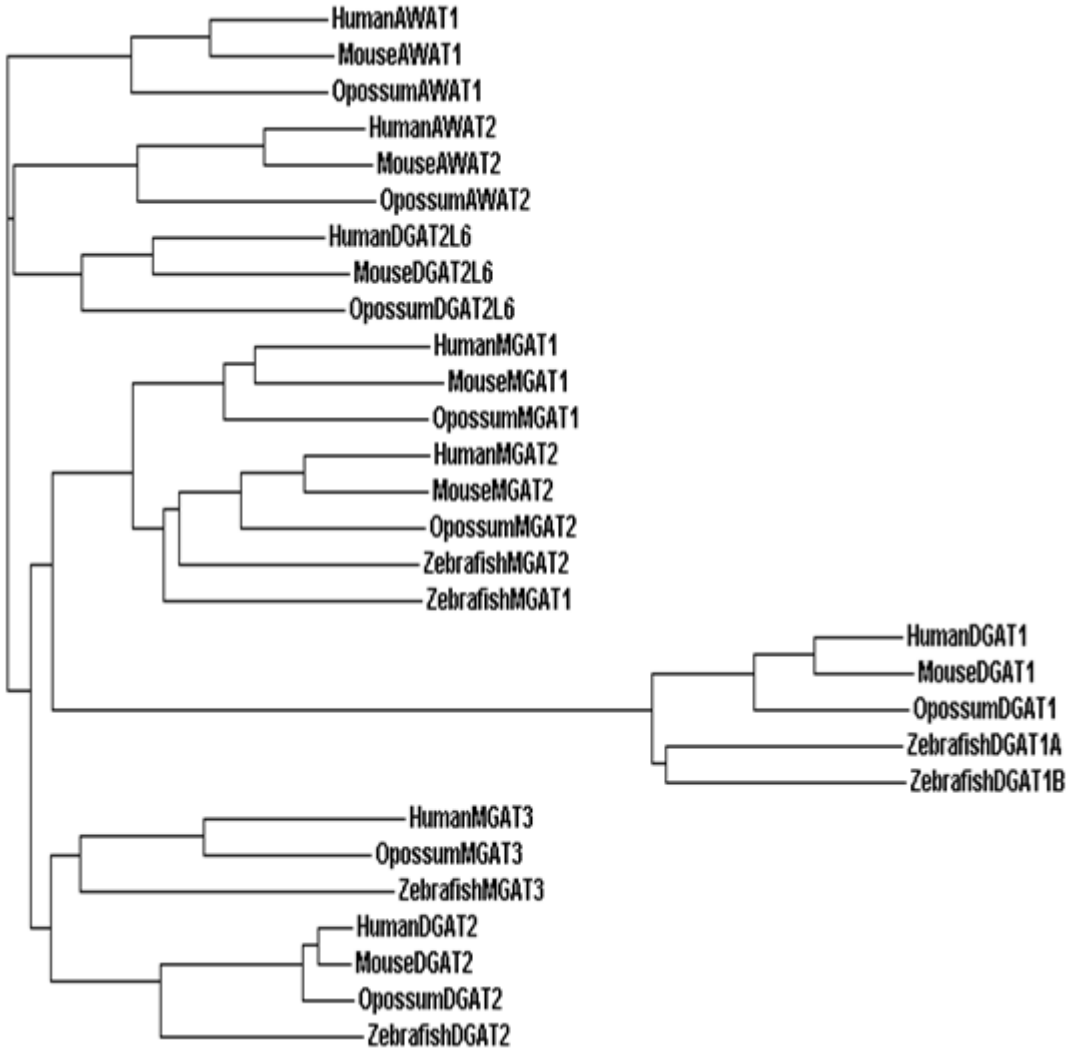


**Figure 5: Phylogenetic tree of human, mouse, opossum, zebrafish and pufferfish DGAT1 and DGAT2-like sequences.**

Each branch of the tree is labeled with the name of the vertebrate (human, mouse, opossum, zebrafish and pufferfish [Tetraodon]) followed by an abbreviation for the gene name: DGAT1; DGAT2; AWAT1; AWAT2; DGAT2L6; MGAT1; MGAT2; and MGAT3). Note the clustering into two groups: DGAT1 and DGAT2-like, and additional clustering of 7 groups for the DGAT2-like enzymes: DGAT2; AWAT1; AWAT2; DGAT2L6; MGAT1; MGAT2; and MGAT3.

| Species | Gene | NCBI RefSeq ID[1] | Chromosome location (strand) | Gene Size (bps) | Coding Exons | Amino Acids |
|---|---|---|---|---|---|---|
| Orangutan | DGAT1 | chr8.780.1[2] | 8:152,680,375-152,690,672 (-ve) | 10,298 | 17 | 476 |
| | AWAT1[1] | chrX.373.1[2] | X:67,880,885-67,886,778 (+ve) | 6,133 | 6[1] | 328 |
| | AWAT2 | chrX.369.1[2] | X:67,699,769-67,708,297 (-ve) | 8,529 | 7 | 333 |
| | DGAT2L6 | chrX.372.1[2] | X:67,820,883-67,851,887 (+ve) | 31,005 | 7 | 336 |
| | DGAT2 | chr11.880.1[2] | 11:71,216,741-71,248,907 (+ve) | 32,167 | 8 | 388 |
| | MGAT2 | chr11.879.1[2] | 11:71,168,993-71,183,490 (+ve) | 14,498 | 6 | 334 |
| | MGAT3[1] | chr7.161.1[2] | 7:10,267,219-10,270,493 (-ve) | 3,275 | 4[1] | 220[1] |
| Rhesus | DGAT1 | XP_001090134 | 8:146,942,899-146,952,955 (-ve) | 10,157 | 17 | 491 |
| | AWAT1 | XP_001083656 | X:69,201,424-69,207,156 (+ve) | 5,733 | 7 | 328 |
| | AWAT2 | XP_001085075 | X:69,009,116-69,017,097 (-ve) | 7,982 | 7 | 333 |
| | DGAT2L6 | XP_001083431 | X:69,150,939-69,177,795 (+ve) | 26,857 | 7 | 337 |
| | DGAT2 | chr14.74.017[2] | 14:73,939,668-73,971,731 (+ve) | 32,064 | 8 | 388 |
| | MGAT1 | XP_001107993 | 12:86,549,294-86,590,162 (+ve) | 40,869 | 6 | 335 |
| | MGAT2 | XP_001086603 | 14:73,891,210-73,906,244 (+ve) | 15,035 | 6 | 334 |
| | MGAT3 | XP_001107764 | 3:48,480,648-48,485,922 (-ve) | 5,275 | 7 | 341 |
| Marmoset | DGAT1 | Contig6937.7[2] | Contig6937:123602-134361 (+ve) | 10,760 | 16 | 515 |
| | AWAT1 | Contig7065.001[2] | Contig7065:4097-10,235 (-ve) | 6,139 | 7 | 328 |
| | AWAT2[1] | contig4884.005[2] | Contig4884:169,128-171,732 (-ve) | 2,605 | 5[1] | 303[1] |
| | DGAT2L6[1] | contig7065.001[2] | Contig7065:29,847-57,530 (-ve) | 27,684 | 6[1] | 315[1] |
| | DGAT2 | contig1148.002[2] | Contig1148:150,614-182,017 (-ve) | 31,404 | 8 | 388 |
| | MGAT1 | contig1051.003[2] | Contig1051:357,042-398,741 (+ve) | 41,700 | 6 | 322 |
| | MGAT2 | contig1148.003[2] | Contig1148:219,348-230,968 (-ve) | 11,621 | 6 | 334 |
| | MGAT3[3] | contig4929.008[2] | Contig4929:227,880-230,020 (+ve) | 21,141[3] | 4[1] | 198[3] |
| | MGAT3[3] | contig10131.5[2] | Contig10131:75,520-75,975 (-ve) | 456[3] | 2[1] | 118[3] |
| Dog | DGAT1 | XP_849176 | 13:40,777,205-40,786,043 (-ve) | 8,839 | 17 | 498 |
| | AWAT1 | XP_549058 | X:57,606,359-57,611,805 (+ve) | 5,447 | 7 | 328 |
| | AWAT2 | XP_549056 | X:57,445,045-57,459,716 (-ve) | 14,672 | 7 | 333 |
| | DGAT2L6 | XP_849355 | X:57,550,407-57,584,831 (+ve) | 34,425 | 7 | 337 |
| | DGAT2 | XP_542303 | 21:25,839,827-25,904,558 (-ve) | 64,732 | 8 | 369 |
| | MGAT1 | XP_545667 | 37:31,731,937-31,763,738 (+ve) | 31,802 | 6 | 335 |
| | MGAT2 | XP_542304 | 21:25,906,094-25,923,425 (-ve) | 17,332 | 6 | 334 |
| | MGAT3 | XP_850305 | 6:11680127-11,683,533 (+ve) | 3,407 | 7 | 345 |
| Rat | DGAT1 | BC081742[4] | 7:114,552,269-114,562,391 (-ve) | 10,123 | 17 | 500 |
| | AWAT1 | NP_001102841 | X:88,570,733-88,576,631 (+ve) | 5,899 | 7 | 328 |
| | AWAT2 | XP_228583 | X:88,395,932-88,404,189 (-ve) | 8,258 | 7 | 333 |
| | DGAT2L6 | XP_001053251 | X:88,516,501-88,539,678 (+ve) | 31,178 | 7 | 337 |
| | DGAT2 | BC089846[4] | 1:156,448,541-156,478,563 (-ve) | 30,023 | 8 | 388 |
| | MGAT1 | XP_001059515 | 9:77,875,371-77,908,608 (+ve) | 33,328 | 6 | 340 |
| | MGAT2 | BC166995[4] | 1:156,517,248-156,540,870 (-ve) | 23,623 | 6 | 334 |
| Horse | DGAT1 | XP_001917097 | 9:82,848,978-82,858,030 (-ve) | 9,053 | 17 | 490 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | AWAT1 | XP_001490162 | X:52,206,564-52,212,113 (+ve) | 5,550 | 7 | 328 |
| | AWAT2 | XP_001496780 | X:51,957,754-51,965,116 (-ve) | 3,263 | 7 | 327 |
| | DGAT2L6 | XP_001490853 | X:52,164,937-52,183,186 (+ve) | 18,350 | 7 | 337 |
| | DGAT2 | XP_001495352 | 7:68,414,190-68,443,579 (-ve) | 29,390 | 8 | 383 |
| | MGAT1 | XP_001495519 | 6:11,748,405-11,768,285 (+ve) | 19,981 | 6 | 335 |
| | MGAT2 | XP_001917405 | 7:68,477,526-68,493,887 (-ve) | 16,362 | 7 | 355 |
| | MGAT3[1] | chr13.155.1[2] | 13:9,131,363-9,133,483 (-ve) | 2,121 | 7 | 266[1] |
| Cow | DGAT1 | BC118146[4] | 14:444,097-446,531[1] (+ve) | 2435[1] | 17 | 489 |
| | AWAT1 | XP_608355 | X:50,032,716-50,038,781 (-ve) | 6,066 | 7 | 328 |
| | AWAT2 | XP_599558 | Un.004.31: 456,157-465,814 (-ve) | 9,658 | 7 | 331 |
| | DGAT2L6 | NP_001095331 | X:50,087,116-50,107,900 (-ve) | 20,785 | 7 | 337 |
| | DGAT2 | NP_991362 | 15:54,717,587-54,748,692 (+ve) | 31,106 | 8 | 361 |
| | MGAT1 | NP_001001153 | 2:115,337,770-115,367,035 (+ve) | 29,266 | 6 | 335 |
| | MGAT2 | XP_001253431 | Un.004.554:9,151-22,088 (+ve) | 12,938 | 6 | 334 |
| | MGAT3 | XP_875499 | 25:37,644,824-37,647,942 (+ve) | 3,119 | 7 | 395 |
| Chicken | DGAT2 | XP_419374[1] | Un:45202148-45,211,195 (+ve) | 9,036 | 8 | 352 |
| | MGAT1 | chr9.162.1[2] | 9:9,089,059-9,109,902 (+ve) | 20,844 | 6 | 326 |
| | MGAT2 | XP_424082 | 1:199,685,261-199,697,403 (-ve) | 12,143 | 7 | 351 |
| | MGAT3 | XP_426251 | 4:1,199,365-1,202,827 (+ve) | 3,463 | 8 | 361 |
| Xenopus | DGAT1 | scaffold_1631.3[2] | scaffold1631:25,368-41,663 (+ve) | 16,296 | 14 | 413 |
| | DGAT2 | BC064191[4] | 725:199,049-214,323 (+ve) | 15,275 | 7 | 361 |
| | MGAT1 | CR942398[4] | 224:1,013,476-1,018,678 (+ve) | 5,203 | 6 | 335 |
| | MGAT2 | BC158509[4] | 725:156,829-176,037 (+ve) | 19,029 | 5 | 331 |
| | MGAT3 | BC088019[4] | 738:466,628-473,346 (+ve) | 6,719 | 6 | 335 |
| Melanogaster | DGAT1 | AF468650[4] | 2L:16,820,210-16,826,713 (+ve) | 6,504 | 9 | 536 |

**Supplementary Table 1: Diacylglycerol acyltransferase (DGAT) and monoacylglycerol acyltransferase (MGAT) genes and enzymes examined**

**Orangutan, rhesus, marmoset, dog, rat, horse and cow AWAT, DGAT2L6, DGAT1, DGAT2, MGAT1, MGAT2 and MGAT3 protein sequences; chicken DGAT2, MGAT1, MGAT2 and MGAT3 sequences; frog DGAT1, DGAT2, MGAT1, MGAT2 and MGAT3 sequences; pufferfish DGAT1, DGAT2, MGAT1, MGAT2 and MGAT3 sequences; and the fruit fly DGAT1 sequence were obtained either from BLATs of the respective genomes using UC Santa Cruz web tools (http://genome.ucsc.edu) or from Blasts using the National Center for Biotechnology Information website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the corresponding human sequences. Predicted chromosomal locations, gene sizes, numbers of coding exons and amino acids for the encoded proteins were obtained by BLAT interrogations of the respective genomes using the predicted human DGAT and MGAT sequences and the UC Santa Cruz web tools ( http://genome.ucsc.edu); [1] refers to NCBI database sources from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sites/entrez); [2] predictions of sequence using the N-SCAN gene structure prediction software provided by the Computational Genomics Lab at Washington University in St. Louis, MO, USA; [3] two partial sequences were observed; [4] GenBank IDs identified previously reported sequences (see http://www.ncbi.nlm.nih.gov/Genbank/); [5]predictions of protein sequences according to the GAZE program: Howe KL, Chothia T and Durbin R (2002), "GAZE: A Generic Framework for the Integration of Gene-Prediction Data by Dynamic Programming", *Genome Res.* 12:1418-1427; [6]unknown chromosomal location.**

| Gene | huDGAT1 | opDGAT1 | huDGAT2 | opDGAT2 | huAWAT1 | opAWAT1 | huAWAT2 | opAWAT2 | huDGAT2L6 | opDGAT2L6 | huMGAT1 | opMGA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| huDGAT1 | 100 | **78** | 5 | 6 | 6 | 8 | 8 | 6 | 4 | 13 | 9 | 6 |
| opDGAT1 | **78** | 100 | 7 | 6 | 8 | 6 | 7 | 5 | 4 | 7 | 6 | 6 |
| huDGAT2 | 5 | 7 | 100 | **92** | 50 | 51 | 48 | 45 | 50 | 49 | 48 | 46 |
| opDGAT2 | 6 | 6 | **92** | 100 | 50 | 49 | 48 | 45 | 50 | 50 | 48 | 46 |
| huAWAT1 | 6 | 8 | 50 | 50 | 100 | **71** | 51 | 46 | 52 | 51 | 40 | 43 |
| opAWAT1 | 8 | 6 | 51 | 49 | **71** | 100 | 50 | 46 | 52 | 49 | 42 | 44 |
| huAWAT2 | 8 | 7 | 48 | 48 | 51 | 50 | 100 | **66** | 51 | 47 | 40 | 41 |
| opAWAT2 | 6 | 5 | 45 | 45 | 46 | 46 | **66** | 100 | 51 | 47 | 41 | 39 |
| huDGAT2L6 | 4 | 4 | 50 | 50 | 52 | 52 | 51 | 51 | 100 | **63** | 44 | 45 |
| opDGAT2L6 | 13 | 7 | 49 | 50 | 51 | 49 | 47 | 47 | **63** | 100 | 41 | 44 |
| huMGAT1 | 9 | 6 | 48 | 48 | 40 | 42 | 40 | 41 | 44 | 41 | 100 | **69** |
| opMGAT1 | 6 | 6 | 46 | 46 | 43 | 44 | 41 | 39 | 45 | 44 | **69** | 100 |
| huMGAT2 | 7 | 8 | 45 | 43 | 43 | 41 | 41 | 38 | 42 | 44 | 53 | 53 |
| opMGAT2 | 5 | 5 | 47 | 47 | 45 | 45 | 39 | 40 | 44 | 44 | 55 | 52 |
| huMGAT3 | 12 | 12 | 48 | 48 | 44 | 45 | 39 | 41 | 45 | 44 | 43 | 40 |
| opMGAT3 | 8 | 5 | 52 | 52 | 47 | 48 | 42 | 45 | 48 | 47 | 44 | 43 |

| Gene | huDGAT1 | moDGAT1 | opDGAT1 | zfDGAT1A | zfDGAT1B |
|---|---|---|---|---|---|
| huDGAT1 | 100 | **85** | **78** | **61** | **62** |
| moDGAT1 | **85** | 100 | **74** | **59** | **60** |
| opDGAT1 | 78 | 74 | 100 | **60** | **61** |
| zfDGAT1A | **61** | **59** | **60** | 100 | **63** |
| zfDGAT1B | **62** | **60** | **61** | **63** | 100 |

| Gene | huDGAT2 | moDGAT2 | opDGAT2 | zfDGAT2 |
|---|---|---|---|---|
| huDGAT2 | 100 | **95** | **92** | **69** |
| moDGAT2 | **95** | 100 | **92** | **70** |
| opDGAT2 | **92** | **92** | 100 | **70** |
| zfDGAT2 | **69** | **70** | **70** | 100 |

| Gene | huMGAT1 | moMGAT1 | opMGAT1 | zfMGAT1 | huMGAT2 | moMGA |
|---|---|---|---|---|---|---|
| huMGAT1 | 100 | **72** | **69** | **56** | 52 | 52 |
| moMGAT1 | **72** | 100 | **67** | **55** | 52 | 52 |
| opMGAT1 | **69** | **67** | 100 | 59 | 53 | 54 |
| zfMGAT1 | **56** | **55** | **59** | 100 | 58 | 59 |
| huMGAT2 | 52 | 52 | 53 | 58 | 100 | **81** |
| moMGAT2 | 52 | 52 | 54 | 59 | **81** | 100 |
| opMGAT2 | 54 | 52 | 52 | 59 | **72** | **69** |
| zfMGAT2 | 54 | 52 | 53 | 62 | **63** | **65** |
| huMGAT3 | 43 | 40 | 40 | 43 | 46 | 46 |
| opMGAT3 | 45 | 42 | 44 | 49 | 46 | 47 |
| zfMGAT3 | 43 | 41 | 43 | 44 | 42 | 43 |

**Supplementary Table 2: Percentage identities for human, rat, opossum and zebrafish DGAT1 and DGAT2-like amino acid sequences**

Numbers show the percentage of amino acid sequence identities. Numbers in **bold** show enzymes of high sequence identity and identify members of the same family or sub-family; huDGAT1: human DGAT1; opDGAT1: opossum DGAT1; huDGAT2: human DGAT2; opDGAT2: opossum DGAT2; huAWAT1: human AWAT1; opAWAT1: opossum AWAT1; huAWAT2: human AWAT2; opAWAT2: opossum AWAT2; huDGAT2L6: human DGAT2L6; opDGAT2L6: opossum DGAT2L6; huMGAT1: human MGAT1; opMGAT1: opossum MGAT1; huMGAT2: human MGAT2; opMGAT2: opossum MGAT2; huMGAT3: human MGAT3; opMGAT3: opossum MGAT3; moDGAT1: mouse DGAT1; zfDGAT1A: zebrafish DGAT1A; zfDGAT1B: zebrafish DGAT1B; moDGAT2: mouse DGAT2; zfDGAT2: zebrafish DGAT2; moMGAT1: mouse MGAT1; moMGAT2: mouse MGAT2; moMGAT3: mouse MGAT3; zfMGAT1: zebrafish MGAT1; zfMGAT2: zebrafish MGAT2; zfMGAT3: zebrafish MGAT3.