Now published in Genetica 139:291-304 (2011)

Vertebrate Endothelial Lipase: Comparative Studies of an Ancient Gene and Protein in Vertebrate Evolution

Roger S Holmes[1-4], John L VandeBerg[1,2] and Laura A Cox[1,2]

[1]Department of Genetics and [2]Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA, and [3]School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD, Australia

[4]Corresponding Author:

Roger S Holmes, D.Sc.

Department of Genetics

Southwest National Primate Research Center

Southwest Foundation for Biomedical Research

San Antonio, TX, USA 78227

Email: rholmes@sfbrgenetics.org

Phone: 210-258-9687

Fax: 210-258-9600

Keywords: Vertebrates; amino acid sequence; endothelial lipase; evolution; gene duplication.

Running Head: Vertebrate Endothelial Lipase: comparative genomics and proteomics

Summary

Endothelial lipase (LIPG; E.C.3.1.1.3) is one of three members of the triglyceride lipase family that contributes to lipoprotein degradation within the circulation system and plays a major role in HDL metabolism in the body. In this study, *in silico* methods were used to predict the amino acid sequences, secondary and tertiary structures, and gene locations for *LIPG* genes and encoded proteins using data from several vertebrate genome projects. *LIPG* is located on human chromosome 18 and is distinct from 15 other human lipase genes examined. Vertebrate *LIPG* genes usually contained 10 coding exons located on the positive strand for most primates, as well as for horse, bovine, opossum, platypus and frog genomes. The rat *LIPG* gene however contained only 9 coding exons apparently due to the presence of a 'stop' codon' within exon 9. Vertebrate LIPG protein subunits shared 58-97% sequence identity as compared with 38-45% sequence identities with human LIPC (hepatic

lipase) and LIPL (lipoprotein lipase). Four previously reported human LIPG N-glycosylation sites were predominantly conserved among the 10 potential N-glycosylation sites observed for the vertebrate LIPG sequences examined. Sequence alignments and identities for key LIPG amino acid residues were observed as well as conservation of predicted secondary and tertiary structures with those previously reported for horse pancreatic lipase (LIPP) (Bourne et al., 1994). Several potential sites for regulating *LIPG* gene expression were observed including CpG islands near the 5'-untranslated regions of the human, mouse and rat *LIPG* genes; a predicted microRNA binding site near the 3'-untranslated region and several transcription factor binding sites within the human *LIPG* gene. Phylogenetic analyses examined the relationships and potential evolutionary origins of the vertebrate *LIPG* gene subfamily with other neutral triglyceride lipase gene families [*LIPC* and *LIPL*], other neutral lipase gene families [*LIPP, LIPR1, LIPR2, LIPR3*, *LIPI, LIPH* and *LIPS*], and the extended family of mammalian acid lipases (*LIPA, LIPF, LIPJ, LIPK, LIPM, LIPN* and *LIPO*). It is apparent that the triglyceride lipase ancestral gene for the vertebrate *LIPG* gene predated the appearance of fish during vertebrate evolution > 500 million years ago.

Introduction

Endothelial lipase (LIPG; E.C.3.1.1.3) is one of three members of the triglyceride lipase family that contributes to lipoprotein degradation within the circulation system and plays a major role in high-density lipoprotein cholesterol (HDL-C) metabolism in the body, catalyzing phospholipase and triglyceride lipase activities (Jaye et al., 1999; Hirata et al., 1999). Hepatic lipase (LIPC; E.C. 3.1.1.3) is another family member which serves a dual role in triglyceride hydrolysis and in ligand-binding for receptor-mediated lipoprotein uptake (Martin et al., 1988; Datta et al., 1988; Cai et al., 1989) while lipoprotein lipase (LIPL; E.C.3.1.1.34) functions in the hydrolysis of triglycerides of circulating chylomicrons and very low density lipoproteins (VLDL) (Wion et al., 1987; Dichek et al., 1991; Benlian et al., 1996). These enzymes share substantial sequence similarities (38-44% identities) and are commonly referred to as the vascular lipase gene family (Hirata et al., 1999; Ma et al., 2003; Brown & Rader, 2007).

The gene encoding LIPG (*LIPG*) is expressed in various cells and tissues of the body, including liver, lung, placenta, macrophages, smooth muscle and endothelial cells (Jaye et al., 1999; Hirata et al., 1999; Lindegaard et al., 2005) where the enzyme contributes significantly to the determination of HDL-C levels, structure and metabolism (see Ma et al., 2003). Studies of *Lipg*$^{-}$/*Lipg*$^{-}$ knock out mice have demonstrated multiple roles for LIPG in vascular lipoprotein metabolism, including serving potential roles in blood vessel inflammation (Kojima et al., 2004), promoting low-density lipoprotein cholesterol (LDL-C) uptake in

macrophages (Yasuda et al., 2007), modulating allergic asthma (Otera et al., 2009), HDL-C mediated repression of leukocyte adhesion (Ahmed et al., 2006) and influencing HDL particle size in the circulation (Jin et al., 2003). Clinical studies have also examined *LIPG* genetic variants in human populations with loss of function *LIPG* mutations leading to increased HDL-C levels and an associated protection from atherosclerotic cardiovascular disease (deLemos et al., 2002; Edmondson et al., 2009).

Structures for several human and animal *LIPG* genes have been determined, including human (Clark *et al*., 2003), mouse (Mouse Genome Sequencing Consortium 2002; The MGC Project Team, 2004) and rat (Bonne *et al*., 2001; Shimokawa et al., 2005). Mammalian *LIPG* genes usually contain 10 exons of DNA encoding LIPG sequences which may undergo exon shuffling generating several isoproteins in each case (Thierry-Mieg and Thierry-Mieg, 2006). Three dimensional studies of an analogous mammalian lipase (LIPP, pancreatic lipase) (Winkler et al., 1990; Bourne et al., 1994) have enabled the prediction of three major structural domains for the mammalian neutral lipase family, including an N-terminal domain with a catalytic triad of serine, aspartate and histidine residues; a 'lid' domain which covers the active site and contributes to the specificity for triglyceride and phosphoglyceride substrates; and a C-terminal or 'plat' domain, which contributes to lipid binding and specificity (Dugi et al., 1995; Broedl et al., 2004). Human LIPG is a heparin binding protein, behaves as a homodimer with a proposed head-to-tail conformation (Griffon et al., 2009) and is subject to proprotein convertase cleavage at a site in the 'hinge' region separating the N- and C-terminal enzyme domains (Jin et al., 2005).

This paper reports the predicted gene structures and amino acid sequences for several vertebrate *LIPG* genes and proteins, the predicted secondary and tertiary structures for vertebrate LIPG enzymes, and the structural, phylogenetic and evolutionary relationships for these genes and enzymes with those for human and mouse lipase gene families.

Methods

In silico vertebrate *LIPG* gene and protein identification.

BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al*, 1997). Protein BLAST analyses used vertebrate LIPG amino acid sequences previously described (Table 1). Non-redundant protein sequence databases for several mammalian genomes were examined using the blastp algorithm, including human (*Homo sapiens*) (International Human Genome Sequencing Consortium, 2001);

chimpanzee (*Pan troglodytes*) (Chimpanzee Genome Analysis Consortium, 2005); orangutan (*Pongo abelii*) (http://genome.wustl.edu) ; cow (*Bos Taurus*) (Bovine Genome Project, 2008); horse (*Equus caballus*) (Horse Genome Project, 2008); mouse (*Mus musculus*) (Mouse Sequencing Consortium, 2002); rat (*Rattus norvegicus*) (Rat Genome Sequencing Consortium, 2004); opossum (*Monodelphis domestica)* (Mikkelsen et al., 2007); platypus *(Ornithorhynchus anatinus)* (Warren et al., 2008); frog (*Xenopus tropicalis*) (http://genome.jgi-psf.org/Xentr3/Xentr3.home.html); and zebrafish (*Danio rerio*) (http://www.sanger.ac.uk/Projects/D_rerio/). This procedure produced multiple BLAST 'hits' for each of the protein databases which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded LIPG-like proteins . These records were derived from annotated genomic sequences using the gene prediction method: GNOMON and predicted sequences with high similarity scores for human LIPG. Predicted LIPG-like protein sequences were obtained in each case and subjected to *in silico* analyses of predicted protein and gene structures.

BLAT analyses were subsequently undertaken for each of the predicted LIPG amino acid sequences using the UC Santa Cruz genome browser [http://genome.ucsc.edu/cgi-bin/hgBlat] (Kent *et al.* 2003) with the default settings to obtain the predicted locations for each of the mammalian *LIPG* genes, including predicted exon boundary locations and gene sizes. BLAT analyses were similarly undertaken for other human lipase genes using previously reported sequences for encoded lipases in each case (see Table 1 and Supplementary Table). Structures for human, mouse and rat isoforms (splicing variants) were obtained using the AceView website to examine predicted gene and protein structures (Thierry-Mieg and Thierry-Mieg, 2006) (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human) .

Predicted Structures and Properties of Vertebrate LIPG Proteins.

Predicted secondary and tertiary structures for human and other vertebrate LIPG-like proteins were obtained using the PSIPRED v2.5 web site tools provided by Brunel University (McGuffin *et al.* 2000) [http://bioinf.cs.ucl.ac.uk/psipred/psiform.html] and the SWISS MODEL web tools [http://swissmodel.expasy.org/], respectively (Guex & Peitsch 1997; Kopp & Schwede 2004). The reported tertiary structure for human pancreatic lipase related protein 1 (LIPR1) (Walker et al., 2010) served as the reference for the predicted human and opossum LIPG tertiary structures, with modeling ranges of residues 47 to 483 and 43 to 483, respectively; while the tertiary structure reported for horse pancreatic lipase (LIPP) (Bourne et al., 1994) served as a reference for frog and zebrafish LIPG tertiary structures, with modeling ranges from 53 to 486 and 70 to 486, respectively. Theoretical isoelectric

points and molecular weights for vertebrate LIPG proteins were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html). SignalP 3.0 web tools were used to predict the presence and location of signal peptide cleavage sites (http://www.cbs.dtu.dk/services/SignalP/) for each of the predicted vertebrate LIPG sequences (Emanuelsson et al 2007). The NetNGlyc 1.0 Server was used to predict potential N-glycosylation sites for vertebrate LIPG proteins (http://www.cbs.dtu.dk/services/NetNGlyc/).

Phylogenetic Studies and Sequence Divergence

Alignments of vertebrate LIPG and human lipase protein sequences were assembled using BioEdit v.5.0.1 and the default settings (Hall, 1999). Alignment ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis yielding alignments of 365 residues for comparisons of human lipase (LIPG, LIPC, LIPL, LIPP, LIPR1, LIPR2, LIPR3, LIPS, LIPH, LIPI, LIPA, LIPF, LIPJ, LIPK, LIPM and LIPN) and mouse acid lipase (LIPO) sequences with the fruit fly (*Drosophila melanogaster*) lipase (LIP3) sequence (Table 1 and Supplementary Table 1). Evolutionary distances were calculated using the Kimura option (Kimura, 1983) in TREECON (Van de Peer & de Wachter, 1994). Phylogenetic trees were constructed from evolutionary distances using the neighbor-joining method (Saitou & Nei, 1987) and rooted with the fruit fly lipase sequence (LIP3). Tree topology was reexamined by the boot-strap method (100 bootstraps were applied) of resampling and only values that were highly significant (≥90) are shown (Felsenstein, 1985).

Results and Discussion

Alignments of Vertebrate LIPG Amino Acid Sequences with Horse LIPP Pancreatic Lipase

The deduced amino acid sequences for opossum, frog and zebrafish LIPG are shown in Figure 1 together with previously reported sequences for human LIPG (Hirata et al., 1999; Jaye et al., 1999), mouse LIPG (The MGC Project Team, 2004), rat LIPG (Bonne et al., 2001) and horse pancreatic lipase (LIPP) (Bourne et al., 1994) (Table 1). Alignments of human and other vertebrate LIPG sequences examined in this figure showed between 58-80% identities, suggesting that these are products of the same family of genes, whereas comparisons of sequence identities of vertebrate LIPG proteins with human and mouse LIPC and LIPL and horse LIPP exhibited lower levels of sequence identities: LIPC (38% and 42% respectively); LIPL (44% and 45% respectively) and LIPP (25%), indicating that these are members of distinct lipase families (Table 2).

The amino acid sequences for human, mouse, opossum and zebrafish LIPG contained 500 residues whereas rat and frog LIPG contained 493 and 497 amino acids, respectively (Figure 1). Previous three dimensional studies on horse pancreatic lipase (LIPP) (Bourne et al., 1994) and modeling studies of human LIPG (Griffon et al., 2009) have

enabled predictions of key residues for these vertebrate LIPG proteins (sequence numbers refer to human LIPG). These included the catalytic triad for the active site (Ser169; Asp193; His274); the hydrophobic N-terminus signal peptides (see also Table 1) which facilitate enzyme secretion into the circulation system (Jin et al., 2003); disulfide bond forming residues (Cys64/Cys78; Cys252/Cys273; Cys297/Cys309; Cys312/Cys317; Cys463/Cys483); the predicted 'lid' region (253-271) which covers the active site and participates in lipid substrate binding in analogous lipases (Winkler et al., 1990; Bourne et al., 1994); a predicted 'hinge' region for vertebrate LIPG containing a proteolytic cleavage site for proprotein convertase (327Arg-328Asn-329Lys-330Arg) (Jin et al., 2005; 2007); and a heparin binding site [a 13 amino acid sequence (321Lys-→333Lys) high in basic amino acid content] which binds LIPG to heparan sulfate proteoglycans on the luminal side of endothelial cells (Hill et al., 1998; Sendak & Bensadoun, 1998). With the exception of the N-terminus signal peptides, all of these sequences were strictly conserved or underwent conservative substitutions which may reflect the essential nature of these residues in contributing to LIPG structure and function. The N-terminal region (residues 1-63) however underwent major changes in the number and sequence of amino acid residues but retained a predicted signal peptide property in each case (Fig. 1; Table 1). The horse LIPP sequence shared the catalytic triad residues, four of the five disulfide bonds predicted for the vertebrate LIPG sequences and an N-signal peptide sequence property however other sequences were distinct with only 25% identical residues observed for horse LIPP and human LIPG.

Four N-glycosylation sites have been previously reported for human LIPG at 80Asn-81Met-82Thr, 136Asn-137Asn-138Thr, 393Asn-394Ala-395Thr and 491Asn-492Glu-493Thr although a further potential site was also reported at 469Asn-470Thr-471Ser (Miller et al., 2004; Skropeta et al., 2007). A comparative analysis of potential N-glycosylation sites for vertebrate LIPG has shown that there are 10 sites overall although only four of these have been predominantly retained for the 16 vertebrate LIPG sequences examined (designated as sites 2, 3, 7 and 10) (Table 3). These corresponded to those previously reported for human LIPG for which specific roles in LIPG-mediated phospholipid hydrolysis in apo-E and apoA-1-containing high density lipoproteins have been reported (Skropeta et al., 2007). It is apparent from this site-directed mutagenesis study of mammalian LIPG that these N-glycosylation sites play important roles in contributing to catalytic efficiency and substrate specificity of LIPG mediated phospholipid hydrolysis. Rat LIPG is also of interest in this regard because it lacks N-glycosylation site 3 due to a 136Asn→Ser substitution (in comparison with mouse LIPG) but showed a new site at 67Asn-68Leu-69Ser which is N-glycosylated but acts as an inhibitor of rat LIPG activity (Skropeta et al., 2007). A single N-glycosylation site was observed for horse LIPP at 45Asn-46Leu-47Thr which is consistent with a previous report (Bourne et al.,

1994) but in contrast with the multiple N-glycosylation sites observed for human LIPG (Miller et al., 2004; Skropeta et al., 2007)

Predicted Secondary and Tertiary Structures for Vertebrate LIPG.

Predicted secondary structures for vertebrate LIPG sequences were compared with the previously reported secondary structure for horse LIPP (Bourne et al., 1994) (Figure 1). α-Helix and β-sheet structures for the vertebrate LIPG protein sequences were examined and found to be similar for several regions with the horse LIPP secondary structures. Consistent structures were predicted near key residues or functional domains including the β-sheet and α-helix structures near the active site residues (human LIPG numbers used) Ser169, Asp 193 and His274; the 'lid' domain (residues 253-271); and the 'hinge' region, which commences with an α-helix (Ala323-Lys324-Lys325), followed by a random coil region. Figure 2 describes predicted tertiary structures for human, opossum and zebrafish LIPG protein sequences and shows significant similarities for these polypeptides with horse pancreatic lipase (LIPP) (Bourne et al., 1994). The three LIPP and LIPG domains were readily apparent, including the N-terminal 'lipase' domain with the active site triad residues buried under the 'lid' domain observed for horse LIPP. The 'lid' has been previously shown to contribute to the preference for triglyceride and phopholipid substrates of other vascular lipases (LIPC and LIPL) (Dugi et al., 1995; Kobayashi et al., 1996) and may play a major role in determining the preference for phospholipid substrates for LIPG. A 'hinge' region was also observed for these vertebrate LIPG proteins, separating the 'lipase' and 'plat' domains, with the latter having a 'sandwich-like' β-pleated sheet structure. The 'plat' domain for LIPC and LIPL has been shown to be essential for binding these enzymes to lipoprotein micelles and also contributes to preferences in lipoprotein binding (Wong et al., 1991; reviewed in Griffon et al., 2009). Recent studies have also shown that LIPG behaves as a dimer with a proposed head-to-tail configuration (Griffon et al., 2009). In addition,a proprotein convertase proteolytic cleavage site has been demonstrated at the 'hinge' region, resulting in partially cleaved dimeric LIPG forms with reduced activities and unknown biochemical roles. These comparative studies for other vertebrate LIPG proteins suggest that these properties and key sequences are substantially retained for all of the sequences examined.

Predicted Gene Locations and Exonic Structures for Vertebrate *LIPG* Genes.

Table 1 summarizes the predicted chromosomal locations for vertebrate *LIPG* genes based upon BLAT interrogations of several vertebrate genomes using the reported sequences for human (Hirata et al., 1999; Jaye et al., 1999), mouse (The MGC Project Team, 2004) and rat LIPG (Rat Genome Sequencing Project Consortium, 2004) and the predicted sequences for horse LIPG and the UC Santa Cruz genome browser (Kent et al. 2003).

The predicted vertebrate *LIPG* genes were predominantly transcribed on the positive strand, with the exception of the marmoset, mouse, rat, guinea pig, dog and zebrafish genes, which were transcribed on the negative strand. Figure 1 summarizes the predicted exonic start sites for human, mouse, rat, opossum, frog and zebrafish *LIPG* genes with each having 10 coding exons, in identical or similar positions to those predicted for the human *LIPG* gene (Clark et al., 2003). In contrast, rat and frog *LIPG* genes contained only 9 coding exons with exon 10 being apparently missing for these 2 genes.

Figure 3 illustrates the predicted structures of mRNAs for human, mouse and rat *LIPG* transcripts for the major transcript isoform in each case (Theirry-Mieg & Thierry-Mieg, 2006). The transcripts were 22-32 kbs in length with 9 introns present for these *LIPG* mRNA transcripts, although in the case of the rat *LIPG* transcript, the coding component of exon 10 was missing. Figure 4 examines the predicted amino acid and nucleotide sequence for the C-terminus end of exon 9 for both mouse and rat *LIPG* sequences. It is proposed that exon 9 of the rat *LIPG* gene has undergone an A→T nucleotide substitution which introduces a terminating codon which may result in an incomplete C-terminus for rat LIPG and an absence of exon 10 for this gene. This reduction of 7 amino acids at the C-terminus for rat LIPG also results in the loss of an N-glycosylation site observed for mouse LIPG at 491Asn-492Lys-493Thr (see Table 3). The significance of these differences in rodent LIPG structure remains to be determined.

The human *LIPG* genome sequence contained several predicted transcription factor binding sites (TFBS), a microRNA site (miR214) located near the 3'-untranslated region and two CpG islands (CpG34 and CpG47) which were located upstream or within the 5'-untranslated region of human *LIPG* on chromosome 18. The occurrence of two CpG islands within the *LIPG* gene promoters may reflect their roles in regulating gene expression (Saxanov et al., 2006). Potentially significant transcriptional factor binding sites for human *LIPG* include the Forkhead box (FoxL1) site in the 5' LIPG region because of its established role as a transcriptional regulator (Myatt & Lamb, 2007) and the PPARalpha binding site, which is a candidate for high-density lipoprotein-mediated repression of leukocyte adhesion (Ahmed et al., 2006) . The prediction of a microRNA (miRNA; miR214) binding site near the 3'untranslated region of the human LIPG gene is potentially of major significance for the regulation of this gene. MiRNAs are small noncoding RNAs that regulate mRNA and protein expression and have been implicated in regulating gene expression during embryonic development (Stephani & Slack, 2008).

Phylogeny and Divergence of *LIPG* and Other Vertebrate Lipase Sequences.

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 16 human lipase amino acid sequences and the mouse LIPO sequence which was 'rooted' with the *Drosophila melanogaster* LIP3 sequence (Pistillo et al., 1998). The phylogram showed clustering of the 'neutral' lipases into 4 distinct groups: Neutral Lipase Group 1 (the 3 vascular lipases): LIPG, LIPC and LIPL; Neutral Lipase Group 2: LIPI and LIPH; Neutral Lipase Group 3 (pancreatic lipases): LIPP, LIPR1, LIPR2 and LIPR3; and Neutral Lipase Group 4: LIPS (steroid-sensitive lipase). These neutral lipase groups were significantly different from each other (with bootstrap values of ~ 100) and from the cluster of 7 distinct acid lipases, which were separated into 3 groups with significant bootstrap values of >90: Group 1 Acid Lipase: LIPA (encoding lysosomal lipase A) and LIPM (an epidermal lipase); Group 2 Acid Lipase: LIPF (gastric lipase), LIPJ, LIPK and LIPN (epidermal lipases); and Group 3 Acid Lipase: LIPO (mouse salivary gland lipase), which was recently reported in mouse and rat genomes (unpublished results). It is apparent from this study of vertebrate LIPG genes and proteins that this is an ancient protein for which the proposed common ancestor for the LIPG, LIPC and LIPL genes must have predated the appearance of bonyfish, which occurred > 500 million years ago (Donoghue & Benton, 2007). This proposal is consistent with a previous finding from Cohen (2003) who has reported the predicted amino acid sequences for human and pufferfish (*Takifugu rubripes*) LIPG, LIPL and LIPC.

Conclusions

The results of the present study indicate that vertebrate *LIPG* genes and encoded enzymes represent a distinct gene and enzyme family of neutral lipases which share key conserved sequences that have been reported for other neutral lipases previously studied (Datta et al., 1988; Cai et al., 1989; Bourne et al., 1994). This enzyme has a unique property among the neutral lipases studied in having a preference for phospholipid substrates and being a major determinant for HDL levels in the circulation system (see Ma et al., 2003; Brown & Rader, 2007). LIPG is encoded by a single gene among the vertebrate genomes studied and usually contained 10 coding exons. The rat *LIPG* gene however encoded a shorter form of this enzyme (493 residues compared with 500 amino acids for most mammalian LIPG sequences) due to the presence of a termination codon predicted in exon 9. Predicted secondary structures and tertiary structures for vertebrate LIPG proteins showed a strong similarity with human and horse LIPP (Winkler et al., 1990; Bourne et al., 1994). Three major structural domains were apparent for vertebrate LIPG, including the 'lipase' domain containing the catalytic triad residues; the 'lid' which covers the active site and may contribute to the substrate specificities of neutral lipases (Dugi et al., 1995; Kobayashi et al., 1996); and the 'plat' domain which contributes to lipoprotein binding (Wong et al., 1991). Phylogenetic studies using 16 distinct human

and mouse lipase sequences indicated that the *LIPG* gene has appeared early in vertebrate evolution, prior to the appearance of bony fish more that 500 million years ago.

Acknowledgements:

This project was supported by NIH Grants P01 HL028972 and P51 RR013986. In addition, this investigation was conducted in facilities constructed with support from Research Facilities Improvement Program Grant Numbers 1 C06 RR13556, 1 C06 RR15456, 1 C06 RR017515.

REFERENCES

Ahmed W, Orasanu G, Nehra V, Asatryan L, Rader DJ, Ziouzenkova O, Plutzky J. (2006) High-density lipoprotein hydrolysis by endothelial lipase activates PPARalpha: a candidate mechanism for high-density lipoprotein-mediated repression of leukocyte adhesion. Circ. Res. 98: 490-498.

Altschul F, Vyas V, Cornfield A, Goodin S, Ravikumar TS, Rubin EH, Gupta E (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

Benlian P., De Gennes J.L., Foubert L., Zhang H., Gagne S.E., Hayden M. (1996) Premature atherosclerosis in patients with familial chylomicronemia caused by mutations in the lipoprotein lipase gene. N. Engl. J. Med. 335:848-854.

Broedl UC, Jin W, Fuki IV, Glick JM, Rader DJ. (2004) Structural basis for endothelial lipase tropism for HDL. FASEB J. 18: 1891-1893.

Brown RJ, Rader DJ (2007) Lipases as modulators of atherosclerosis in murine models. Curr. Drug Targets 8: 1307-1309.

Bonne A.C.M., den Bieman M.G., van Lith H., van Zutphen B.F.M. (2001) Sequencing and chromosomal assignment of the rat endothelial-derived lipase gene (*Lipg*). DNA Seq. 12:285-287.

Bourne Y., Martinez C., Kerfelec B., Lombardo D., Chapus C., Cambillau C. (1994) Horse pancreatic lipase. The crystal structure refined at 2.3-A resolution. J. Mol. Biol. 238:709-732.

Bovine Genome Project (2008) http://hgsc.bcm.tmc.edu/projects/bovine

Cai S.J., Wong D.M., Chen S.H., Chan L. (1989) Structure of the human hepatic triglyceride lipase gene. Biochem. 28: 8966-8971.

Chenna R., Sugawara H., Koike T, Lopez R, Gibson TJ, Higgins DJ, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497-3500.

Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69-87.

Clark HF, Gurney AL, Abaya E, Baker K, Baldwin D, Brush J, Chen J, Chow B, Chui C, Crowley C, Currell B, Deuel B, Dowd P, Eaton D, Foster J, Grimaldi C, Gu Q, Hass PE, Heldens S, Huang A, Kim HS, Klimowski L, Jin Y, Johnson S, Lee J, Lewis L, Liao D, Mark M, Robbie E, Sanchez C, Schoenfeld J, Seshagiri S, Simmons L, Singh J, Smith V, Stinson J, Vagts A, Vandlen R, Watanabe C, Wieand D, Woods K, Xie MH, Yansura D, Yi S, Yu G, Yuan J, Zhang M, Zhang Z, Goddard A, Wood WI, Godowski P, Gray A (2003) The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res 13: 226-2270.

Cohen JC (2003) Endothelial lipase: direct evidence for a role in HDL metabolism. J. Clin. Invest. 111: 1-3.

Cox LA, Birnbaum S, Mahaney MC, Rainwater DL, Williams JT, VandeBerg JL (2007) Identification of promoter variants for baboon endothelial lipase that regulate baboon HDL-cholesterol levels. Circulation 116: 1185-1195.

Datta S., Luo C.C., Li W.H., VanTuinen P., Ledbetter D.H., Brown M.A., Chen S.H., Liu S., Chan L. (1988) Human hepatic lipase. Cloned cDNA sequence, restriction fragment length polymorphisms, chromosomal localization, and evolutionary relationships with lipoprotein lipase and pancreatic lipase. J. Biol. Chem. 263:1107-1110.

deLemos AS, Wolfe ML, Long CJ, Sivapackianathan R, Rader DJ. (2002) Identification of genetic variants in endothelial lipase in persons with elevated high-density lipoprotein cholesterol. Circ. 106: 1321-1326.

Dichek H.L., Fojo S.S., Beg O.U., Skarlatos S.I., Brunzell J.D., Cutler G.B. Jr., Brewer H.B. Jr. (1991) Identification of two separate allelic mutations in the lipoprotein lipase gene of a patient with the familial hyperchylomicronemia syndrome. J. Biol. Chem. 266:473-477.

Donoghue PCJ, Benton MJ (2007) Rocks and clocks: calibrating the tree of life using fossils and molecules. Trends Genet 22: 424-431.

Dugi KA, Dichek HL, Santamarina-Fojo S. (1995) Human hepatic and lipoprotein lipase: the loop covering the catalytic site mediates lipase substrate specificity. J. Biol. Chem. 270: 25396-25401.

Edmondson AC, Brown RJ, Kathiresan S, Cupples LA, demissie S, Manning AK, Jensen MK, Rimm EB, Wang J, Rodrigues A, Bamba V, Khetarpal SA, Wolfe ML, Derohannessian S, Li M, reilly MP, Aberle J, Evans D, Hegele RA, Rader DJ. (2009) Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. J. Clin. Invest. 119: 1042-1050.

Emmanuelsson O, Brunak S, von Heijne G, Nielson H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protocols 2: 953-971.

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18: 2714-2723.

Hirata K., Dichek H.L., Cioffi J.A., Choi S.Y., Leeper N.J., Quintana L., Kronmal G.S., Cooper A.D., Quertermous T. (1999) Cloning of a unique lipase from endothelial cells extends the lipase gene family. J. Biol. Chem. 274:14170-14175.

Horse Genome Project (2008) http://www.uky.edu/Ag/Horsemap/

Horton P, Nakai K (1997) Better prediction of cellular localization sites with the k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol 5: 147-152.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Jaye M., Lynch K.J., Krawiec J., Marchadier D., Maugeais C., Doan K., South V., Amin D., Perrone M., Rader D.J. (1999) A novel endothelial-derived lipase that modulates HDL metabolism. Nat. Genet. 21:424-428.

Jin W, Sun G-S, Marchadier D, Octtaviani E, Glick JM, Rader DJ. (2003) Endothelial cells secrete triglyceride lipase and phospholipase activities in response to cytokines as a result of endothelial lipase. Circ. Res. 92: 644-650.

Jin W., Fuki IV, Seidah NG, Benjannet S, Glick JM, Rader DJ. (2005) Proprotein convertases are responsible for proteolysis and inactivation of endothelial lipase. J. Biol. Chem. 280: 36551-36559.

Kent WJ, Sugnet CW, Furey TS et al (2003) The human genome browser at UCSC. Genome Res 12: 994-1006.

Kobayashi J, Applebaum-Bowden D, Dugi KA, Brown DR, Kashyap VS, Parrott C, Duarte C, Maeda N, Santamarina-Fojo S. (1996) Analysis of protein structure-function in vivo. Adenovirus-mediated transfer of lipase lid mutants in hepatic lipase-deficient mice. J. Biol. Chem 271: 26296-262301.

Kojima Y, Hirata K, Ishida T, Shimokawa Y, Inoue N, Kawashima S, Quertermous T, Yokoyama M. (2004) Endothelial lipase modulates monocyte adhesion to the vessel wall. A potential role in inflammation. J. Biol. Chem 279: 54032-54038.

Kopp J, Schwede T (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res 32: D230-D234.

Lindegaard MLS, Olivcrona G, Christofferson C, Kratky D, Hannibal J, Petersen BL, Zechner R, Damm P, Nielson LB (2005) Endothelial and lipoprotein lipases in human and mouse placenta. J. Lipid Res. 46: 2339-2346.

Ma K, Cilingiroglu M, Otvos JD, Ballantyne CM, Marian AJ, Chan L (2003) Endothelial lipase is a major genetic determinant for high-density lipoprotein concentration, structure and metabolism. Proc Natl Acad. Sci USA 100: 2748-2753.

Martin G.A., Busch S.J., Meredith G.D., Cardin A.D., Blankenship D.T., Mao S.J.T., Rechtin A.E., Woods C.W., Racke M.M., Schafer M.P., Fitzgerald M.C., Burke D.M., Flanagan M.A., Jackson R.L. (1988) Isolation and cDNA sequence of human postheparin plasma hepatic triglyceride lipase. J. Biol. Chem. 263:10907-10914.

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404-405.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.

Myatt SS, & Eric W. -F. Lam EW-F (2007) The emerging roles of forkhead box (Fox) proteins in cancer. Nature Revs Cancer 7: 847-859.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,Wakamatsu, A., Hayashi, K., Sato, H., Nagai,K., Kimura, K.,Makita, H.,Sekine,M., Obayashi,M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T.,Wagatsuma,M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara,M., Takahashi,M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki,M., Ninomiya,K., Ishibashi, T., Yamashita, H.,Murakawa,K., Fujimori,K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida,M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T.O., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose,N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama,M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y.,Mizuno, T.,Morinaga,M., Sasaki, M., Togashi, T.,Oyama, M., Hata,H.,Watanabe,M.,Komatsu, T.,Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y.,Nakagawa, K.,Okumura,K.,Nagase, T.,Nomura,N., Kikuchi, H.,Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., Sugano, S., 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat. Genet. 36, 40–45.

Otera H, Ishida T, Nishiuma T, Kobayashi K, Kotani Y, Yasuda T, Kundu RK, Quertermous T, Hirata K, Nishimura Y. (2009) Targeted inactivation of endothelial lipase attenuates lung allergic inflammation through raising plasma HDL levels and inhibiting eosinophil infiltration. Am. J. Physiol. Lung Cell Mol. Physiol. 296: L594-602.

D. Pistillo D, Manzi A, Tino A, Boyl PP, Graziani F, Malva C. (1998) The *Drosophila melanogaster* lipase homologs: a gene family with tissue and developmental specific expression. J. Mol. Biol. 276: 877-885.

Rat Genome Sequencing Project Consortium. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428, 493-521.

Saxonov S, Berg P, and Brutlag DL. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters, Proc. Natl. Acad. Sci. USA 103 :1412–1417.

Schwede T, Kopp J, Guex N, Peitsch MC. (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 31: 3381-3385.

Shimokawa Y., Hirata K., Ishida T., Kojima Y., Inoue N., Quertermous T., Yokoyama M. (2005) Increased expression of endothelial lipase in rat models of hypertension. Cardiovasc. Res. 66:594-600.

Stefani G, Slack FJ (2008) Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol 9: 219–30.

The MGC Project Team. (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Res. 14:2121-2127.

Thierry-Mieg D, Thierry-Mieg J (2006) AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biology 7, S12 http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. 133: 17-21.

Walker JR, Davis T, Seitova A, Butler-Cole C, Weigelt J, Sundstrom M, Arrowsmith CH, Edwards AM, Bochkarev A, Dhe-Paganon S (2010) Protein data bank entry for human pancreatic lipase-related protein 1 (2PPL) http://www.rcsb.org/pdb/explore/explore.do?structureId=2PPL

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K (2008) Genome analysis of the platypus reveals unique signatures of evolution. Nature. 453:175-83.

Winkler F.K., D'Arcy A., Hunziker W. (1990) Structure of human pancreatic lipase. Nature 343:771-774.

Wion K.L, Kirchgessner T.G, Lusis A.J, Schotz MC, Lawn RM. (1987) Human lipoprotein lipase complementary DNA sequence. Science 235:1638-1641.

Wong H, Davis RC, Nikazy J, Seebart KE, Schotz MC (1991) Domain exchange: characterization of a chimeric lipase od hepatic lipase and lipoprotein lipase. Proc. Natl. Acad. Sci. USA 88: 11290-11294.

Woodburne MO, Rich TH, Springer MS (2003) The evolution of tribospheny and the antiquity of mammalian clades. Mol Phylogenet Evol 28: 360-385.

Yasuda T, Hirata K, Ishida T, Kojima Y, Tanaka H, Okada T, Quertermous T, Yokoyama M. (2007) Endothelial lipase is increased by inflammation and promotes LDL uptake in macrophages. J. Atheroscler. Thromb. 14: 192-201.

Figure 1: Amino Acid Sequence Alignments for Vertebrate EL and Horse LIPP Sequences

See Table 1 for sources of EL and LIPP sequences: Hu-human EL; Mo-mouse EL; Ra-rat EL; Op-opossum EL; Fr-frog EL; Zf-zebrafish EL; PL-horse pancreatic lipase (LIPP); * shows identical residues for lipase subunits; : similar alternate residues; . dissimilar alternate residues; residues involved in N-signal peptide formation are shown in red; N-glycosylated (marked as & for human LIPG) and potential N-glycosylated Asn sites are in **green bold**; active site (AS) triad residues Ser (S); Asp (D); and His (H) are in **pink bold**; predicted disulfide bond Cys residues are shown in **blue bold (•);** α-helix for horse LIPP or predicted for vertebrate LIPG is in shaded yellow; β-sheet for horse LIPP or predicted for vertebrate LIPG is in shaded grey; **bold** underlined font shows residues corresponding to known or predicted exon start sites; exon numbers

refer to human *LIPG* gene exons; #### refers to residues which correspond to the horse LIPP 'lid' region; **xxxxx** refers to the 'hinge' region for horse LIPP' ^ refers to hydrophobic amino acids in the 'plat' domain which are located near to the active site triad in the EL dimer model reported by Griffon et al (2009) .

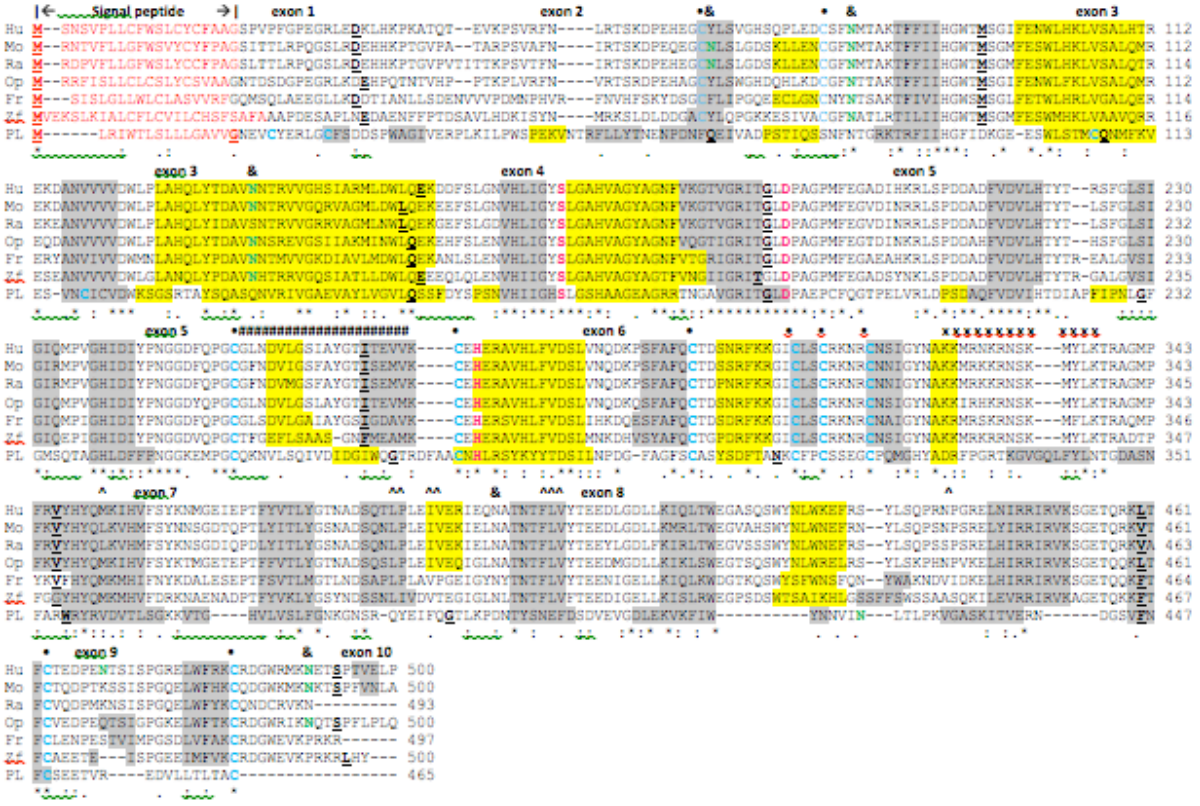


Figure 2: Tertiary Structure for Horse LIPP and Predicted Tertiary Structures for Human, Opossum and Zebrafish LIPG Proteins

The structure for horse LIPP is taken from Bourne et al, 1994; predicted human, opossum and zebrafish LIPG tertiary structures were obtained using SWISS MODEL methods; the rainbow color code describes the tertiary structures from the N- (blue) to C-termini (red color) for human, opossum and zebrafish LIPG; the horse LIPP tertiary structure shows the N- and C-termini, the 'lipase', 'lid' (in yellow) and 'plat' domains which are separated by a 'hinge' region; and the active site triad residues for horse LIPP which are shown in red.

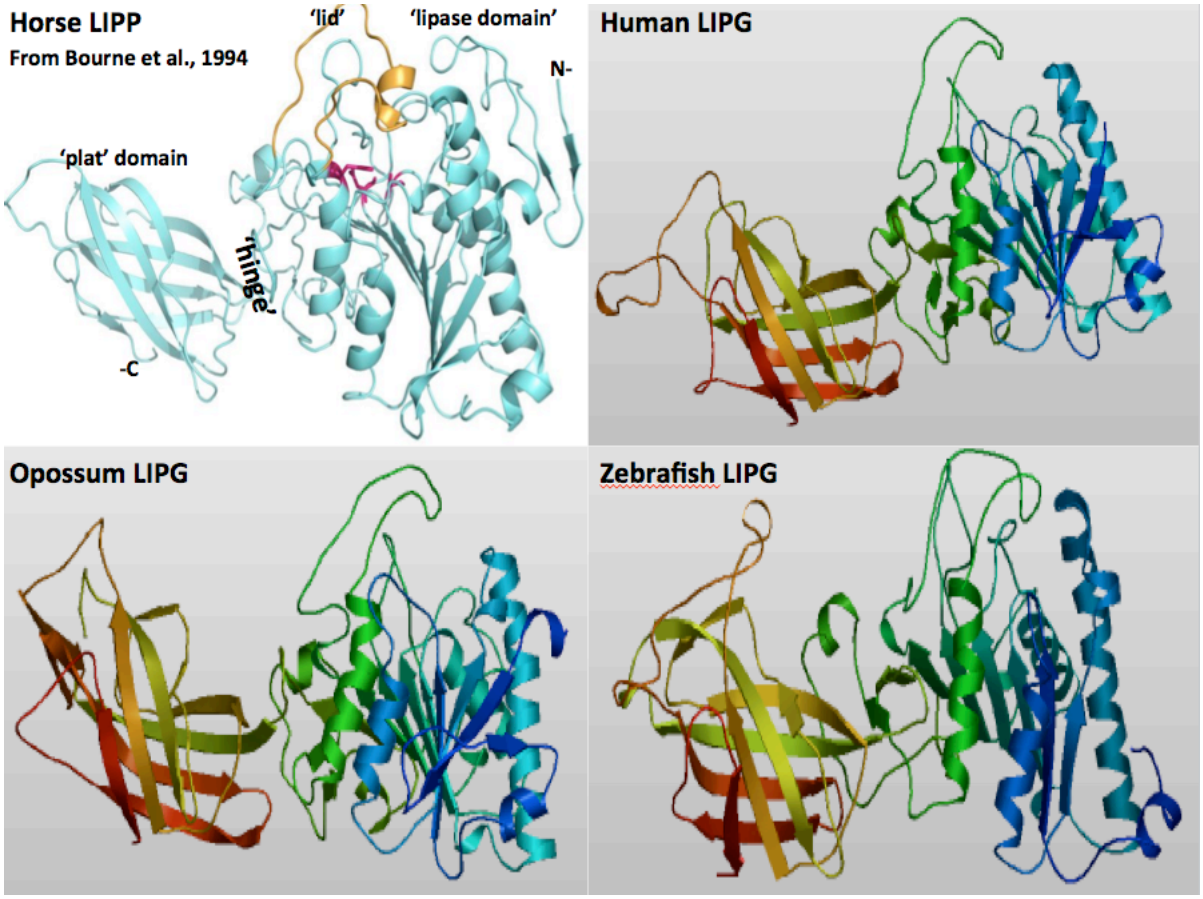
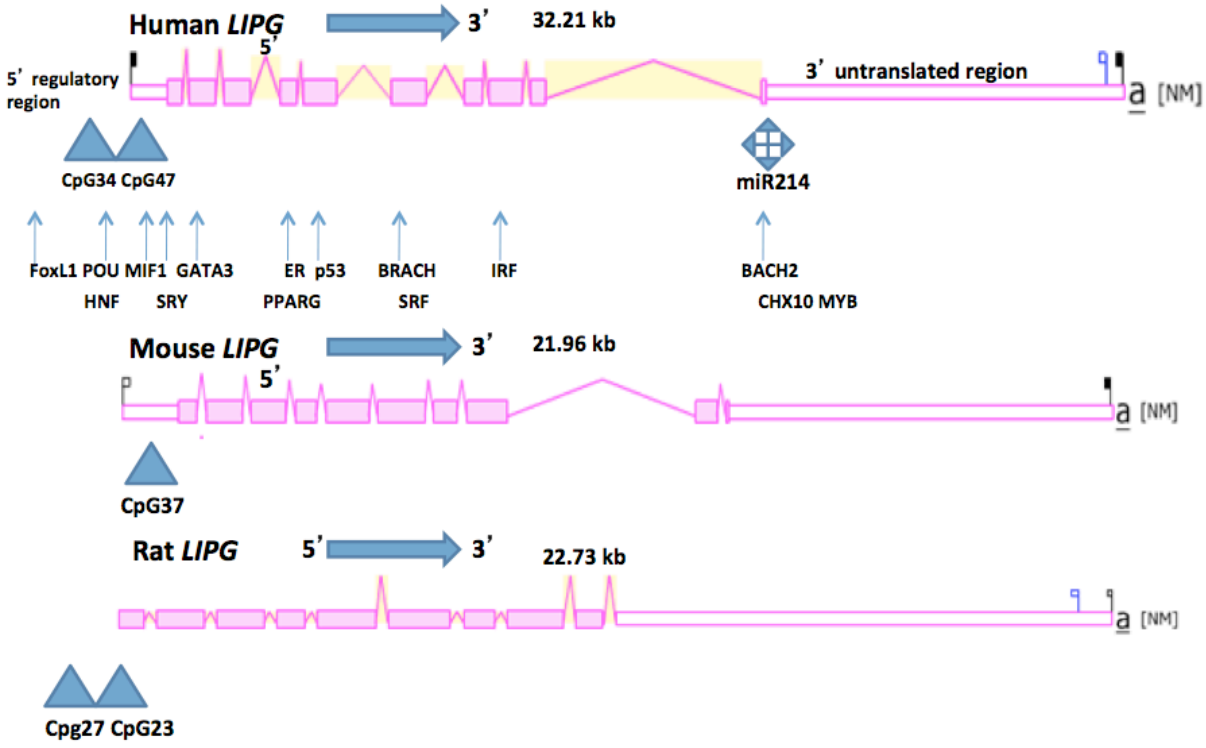
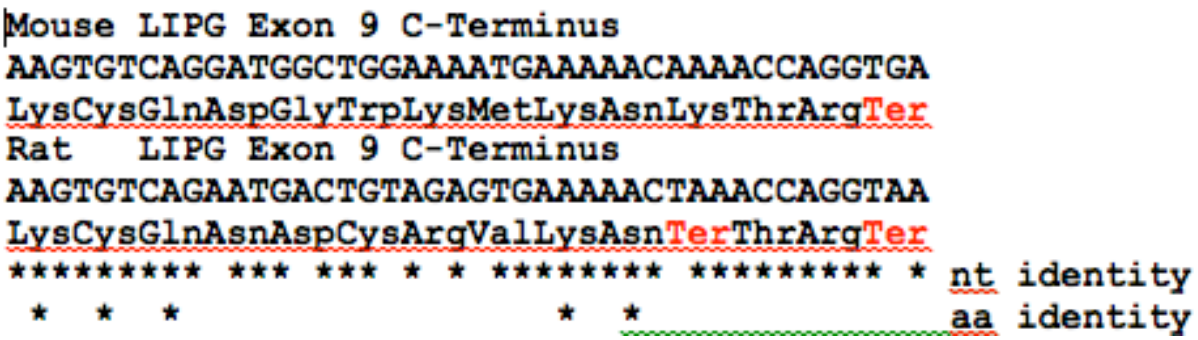


Figure 3: Gene Structures and Major Splicing Variant for the Human, Mouse and Rat *LIPG* Genes.



Derived from the AceView website http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/ (Thierry-Mieg and Thierry-Mieg,2006); mature isoform variants (a) are shown with capped 5'- and 3'- ends for the predicted mRNA sequences; NM refers to the NCBI reference sequence; exons are in pink; the directions for transcription are shown as 5' → 3'; blue triangles show predicted CpG island sites at or near the 5'untranslated regions of the gene; the blue square shows a predicted microRNA binding site (miR214) observed at or near the human *LIPG* 3'untranslated region; sizes of mRNA sequences are shown in kilobases (kb); diagram includes 2kb upstream 5'-regulatory region for human LIPG; blue arrows show predicted transcription factor binding sites for human *LIPG*: FoxL1-forkhead protein L1; HNF hepatocyte nuclear factor; POU6-novel DNA-binding motif; MIF1-Krüppel-associated box domain protein; SRY-sex determining region Y gene binding protein; GATA3-zinc finger-like transcription factor; ER-estrogen receptor DNA binding; PPARG-peroxisome proliferator-activated receptor γ; p53-tumor suppressor gene transcription factor; BACH-B cell specific transcription factor; SRF-serum response factor; IRF-interferon regulatory factor; CHX10-homeobox gene transcription factor; MYB-hematopoietic cell proliferating regulator.

Figure 4: Nucleotide and Amino Acid Sequence Alignments for Mouse and Rat *LIPG* Genes: Predicted C-termini and Exon 9 Sequences.



Identical nucleotide (nt) and amino acid (aa) sequences are shown (*). Ter (in red) refers to predicted terminating codons.

Figure 5: Phylogenetic Tree of Human and Mouse Lipase Amino Acid Sequences.

The tree is labeled with the lipase name and the name of the animal and is 'rooted' with the *Drosophila melanogaster* lipase 3 sequence (*Lip3*). Note the 2 major clusters for human and mouse (for LIPO) sequences corresponding to the 'neutral' (human lipase L, G etc) and 'acid' (human lipase A, M etc) lipase groups. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates are shown. Only replicate values of 90 or more which are highly significant are shown with 100 bootstrap replicates performed in each case. Note the significant separation of clades for the human vascular lipases (LIPL, LIPG and LIPC) from the other neutral lipases.

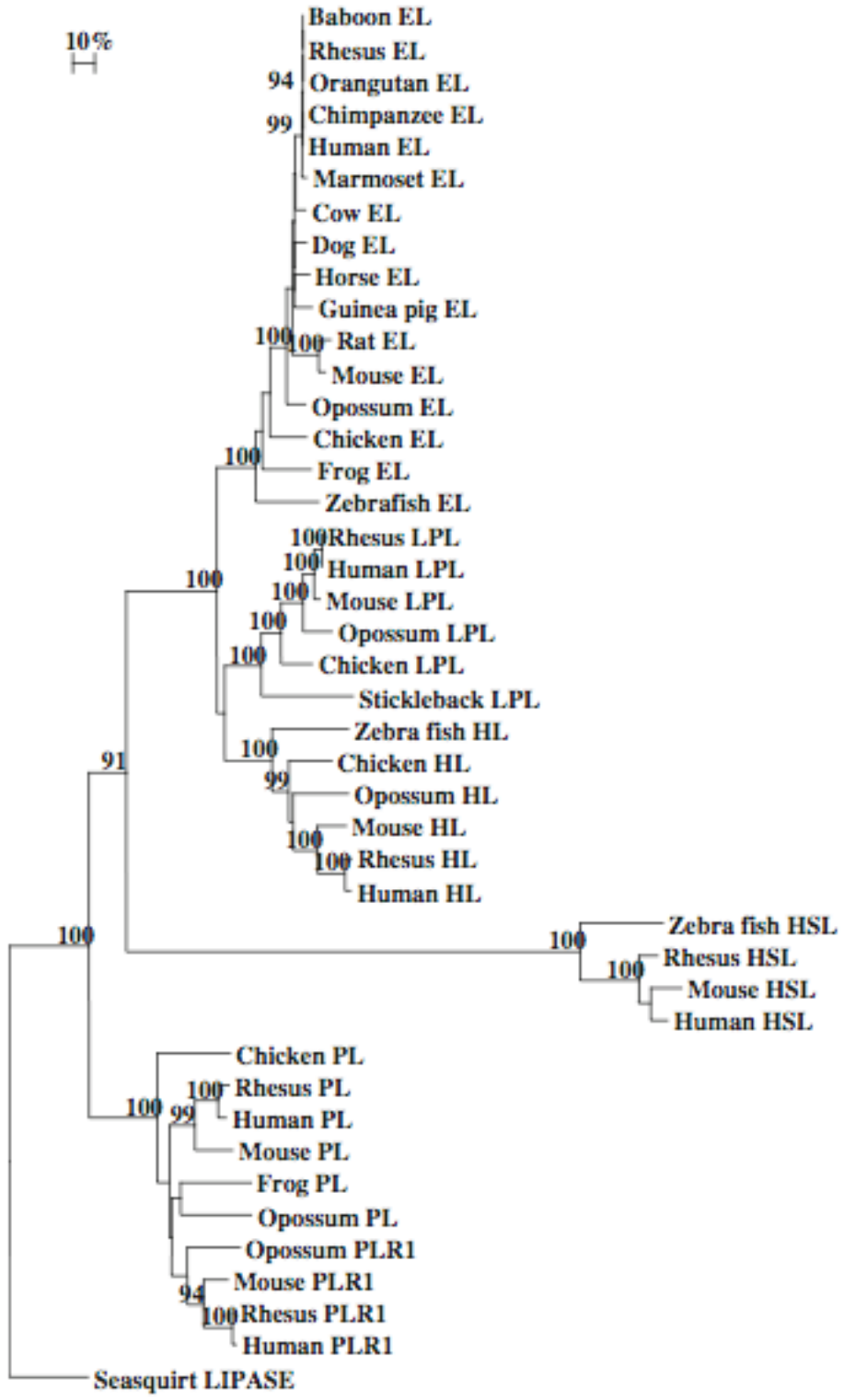

10%
Baboon EL
Rhesus EL
Orangutan EL
Chimpanzee EL
Human EL
Marmoset EL
Cow EL
Dog EL
Horse EL
Guinea pig EL
Rat EL
Mouse EL
Opossum EL
Chicken EL
Frog EL
Zebrafish EL
Rhesus LPL
Human LPL
Mouse LPL
Opossum LPL
Chicken LPL
Stickleback LPL
Zebra fish HL
Chicken HL
Opossum HL
Mouse HL
Rhesus HL
Human HL
Zebra fish HSL
Rhesus HSL
Mouse HSL
Human HSL
Chicken PL
Rhesus PL
Human PL
Mouse PL
Frog PL
Opossum PL
Opossum PLR1
Mouse PLR1
Rhesus PLR1
Human PLR1
Seasquirt LIPASE

Table Legends

Table 1: Vertebrate endothelial lipase (*LIPG*), fruit fly (*Drosophila melanogaster*) lipase 3 *(Lip3)* and horse pancreatic lipase (*LIPP*) genes and proteins.

Endothelial Lipase	Species	RefSeq ID	GenBank ID	UNIPROT	Amino	Chromosome	Exons	Gene Size	pI	Subunit	Signal Peptide
Gene *LIPG*		[1]Ensembl (predicted)		ID	acids	location	(strand)	bps		MW	(Cleavage site)
Human	*Homo sapiens*	NM_006033.2	BC060825	Q9Y5X9	500	18:45,342,677-45,367,216	10 (+ve)	24,540	8.13	56,795	1-20 [AG-SP]
Chimpanzee	*Pan troglodytes*	[1]XP_512126.2	[2]	[2]	500	18:45,797,471-45,819,707	10 (+ve)	22,237	8.13	56,765	1-20 [AG-SP]
Orangutan	*Pongo abelii*	[1]ENSPPYT00000010689	[2]	[2]	500	18:61,955,160-61,978,895	10 (+ve)	23,736	8.31	56,713	1-20 [AG-SP]
Rhesus	*Macaca mulatta*	[1]XP_001090086.1	[2]	[2]	500	18:42,603,162-42,624,867	10 (+ve)	21,706	7.25	56,582	1-20 [AG-SP]
Baboon	*Papio hamadryas*	[2]	ABS50584.1	A7LD44	500	[2]	10 (+ve)	25,970	6.81	56,783	1-20 [AG-SP]
Marmoset	*Callithrix jacchus*	[1]Contig2102.003.a	[2]	[2]	502	[4]Contig2102:222,399-245,468	10 (-ve)	23,070	8.29	56,905	1-20 [AG-SP]
Mouse	*Mus musculus*	NM_010720.3	BC020991	Q9WVG5	500	18:75,102,996-75,120,628	10 (-ve)	17,633	8.79	56,629	1-20 [AG-SI]
Rat	*Rattus norvegicus*	NM_001012741	AY916123	Q8VBX1	[6]493	18:71,797,144-71,816,056	[6]9 (-ve)	18,913	8.82	[6]55,924	1-20 [AG-SL]
Guinea Pig	*Cavia porcellus*	[1]ENSCPOT00000008970	[2]	[2]	498	14:15,934,328-15,953,431	10 (-ve)	19,104	7.27	56,624	1-20 [AG-SP]
Horse	*Equus caballus*	[1]XP_001499209.1	[2]	[2]	500	8:68,116,286-68,134,659	10 (+ve)	18,374	7.89	56,565	1-18 [FA-AG]
Cow	*Bos taurus*	[1]XP_586851.3	[2]	[2]	500	24:50,881,328-50,965,323	10 (+ve)	83,996	7.23	56,207	1-20 [TG-DP]
Dog	*Canis familaris*	[1]ENSCAFT00000030233	[2]	[2]	497	7:82,253,865-82,270,894	10 (-ve)	17,030	6.74	56,183	1-18 [LA-AG]
Opossum	*Monodelphis domestica*	[1]XP_001372654.1	[2]	[2]	500	3:69,419,824-69,449,888	10 (+ve)	30,065	8.14	56,816	1-19 [AA-GN]
Platypus	*Ornithorhynchus anatinus*	[1]ENSOANT00000001800	[2]	[2]	501	[4]Contig2412:6,656-24,362	10 (+ve)	17,707	6.89	56,859	1-23 [SA-QI]
Frog	*Xenopus tropicalis*	[1]ENSXETT00000048464	[2]	[2]	497	[7]sc97:2,765,234-2,771,503	9 (+ve)	6,270	6.36	56,117	1-18 [RF-GQ]
Zebrafish	*Danio rerio*	[1]ENSDART00000098850	BC044146	[2]	500	8:30,952,666-30,959,571	10 (-ve)	6,906	6.16	55,952	1-25 [FA-AA]
Fruit Fly *LIP*	*Drosophila melanogaster*	NM_137322	[8]CG6472	A1ZAL5	394	2R:12,767,778-12,772,002	6 (+ve)	4,228	8.79	43,512	Nil
Pancreatic Lipase											
Horse	*Equus caballus*	NM_001163949	X66218	P29183	465	1:15,534,773-15,551,621	12 (-ve)	16,849	5.46	54,435	1-16 [VG-NE]

RefSeq: the reference amino acid sequence; [1,3]predicted Ensembl amino acid sequence; [2]not available; [4] Contig refers to a DNA scaffold for sequencing analyses; [5]BLAT predicted amino acid sequences are shown; [6] rat LIPG lacks 7 residues at the C-terminus; [7]scaffold of DNA used in sequencing frog genome; [8] FlyBase ID; GenBank IDs are derived from NCBI sources http://www.ncbi.nlm.nih.gov/genbank/; Ensembl ID was derived from Ensembl genome database http://www.ensembl.org ; UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual acid lipases (see http://kr.expasy.org); bps refers to base pairs of nucleotide sequences; pI refers to theoretical isoelectric points; the number of coding exons are listed.

Table 2: Percentage identities for vertebrate LIPG and human and mouse LIPC and LIPL amino acid sequences.

Lipase Gene	Human LIPG	Baboon LIPG	Mouse LIPG	Rat LIPG	Horse LIPG	Cow LIPG	Opossum LIPG	Frog LIPG	ZebrafishLIPG	Human LIPC	Mouse LIPC	Human LIPL	Mouse LIPL
Human LIPG	100	**97**	**80**	**79**	**86**	**85**	**80**	**62**	**58**	38	42	44	45
Baboon LIPG	**97**	100	**80**	**80**	**86**	**85**	**81**	**63**	**59**	40	42	44	45
Mouse LIPG	**80**	**80**	100	**90**	**77**	**77**	**74**	**59**	**55**	37	40	45	46
Rat LIPG	**79**	**80**	**90**	100	**77**	**76**	**74**	**58**	**55**	37	38	44	45
Horse LIPG	**86**	**86**	**77**	**77**	100	**85**	**79**	**61**	**58**	39	41	46	46
Cow LIPG	**85**	**85**	**77**	**76**	**85**	100	**79**	**60**	**57**	39	41	44	45
Opossum LIPG	**80**	**81**	**74**	**74**	**79**	**79**	100	**62**	**57**	41	42	43	43
Frog LIPG	**62**	**63**	**59**	**58**	**61**	**60**	**62**	100	**58**	39	41	45	46
Zebrafish LIPG	**58**	**59**	**55**	**55**	**58**	**57**	**57**	**58**	100	36	38	42	44
Human LIPC	38	40	37	37	39	39	41	39	36	100	**74**	41	42
Mouse LIPC	42	42	40	38	41	41	42	41	38	**74**	100	44	43
Human LIPL	44	44	45	44	46	44	43	45	42	41	44	100	**92**
Mouse LIPL	45	45	46	45	46	45	43	46	44	42	43	**92**	100

Numbers show the percentage of amino acid sequence identities. Numbers in **bold** show higher sequence identities for lipases from the same gene family.

Table 3: Predicted N-glycosylation sites for vertebrate LIPG proteins

Numbers refer to amino acids in the acid sequences, including N-asparagine; K-lysine; I-isoleucine; M-methionine; H-histidine; S-serine; R-arginine; T-threonine; Q-glutamine; and V-valine. Note that there are 10 potential sites identified, including 4 confirmed sites for human LIPG (*sites 2, 3, 7 and 10) (see Skropeta et al., 2007). Note that Asn→Ala changes for 3 of these sites (sites 3, 7 and 10 in green) results in a significant decrease in LIPG activity whereas a similar change for another site results in enhancement of LIPG activity (site 2 in red). High (**in bold**) and lower probability N-glycosylation sites were identified using the NetNGlyc 1.0 web server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Vertebrate	Species	Site 1	Site 2*	Site 3*	Site 4	Site 5	Site 6	Site 7*	Site 8	Site 9	Site 10*
Human	*Homo sapiens*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Chimpanzee	*Pan troglodytes*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Orangutan	*Pongo abelii*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Rhesus	*Macaca mulatta*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Baboon	*Papio hamadryas*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Marmoset	*Callithrix jacchus*		80 NMT (-)	136 NNT (+)				393 NAT (+)		469 NTS (-)	491 NET (+)
Mouse	*Mus musculus*	65NLS (+)	80 NMT (-)	136 NNT (++)		359 NNS (-)		393 NAT (++)	439 NPS (-)		491 NKT (++)
Rat	*Rattus norvegicus*	67NLS (++)	82 NMT (-)					395 NAT (+)			
Guinea Pig	*Cavia porcellus*		78 NVT (+)	134 NNS (++)				391 NAT (+)		467 NSS (-)	489 NET (+)
Horse	*Equus caballus*		80 NMT (-)	136 NNT (++)				393 NAT (+)		469 NTS (-)	491 NET (+)
Cow	*Bos taurus*		80 NVT (+)	136 NNT (+)				393 NAT (+)			491 NQT (+)
Dog	*Canis familaris*		77 NVT (+)	136 NNT (++)				390 NAT (+)		466 NTS (-)	488 NET (+)
Opossum	*Monodelphis domestica*		80 NTT (-)	136 NNS (++)				393 NAT (+)			491 NQT (+)
Platypus	*Ornithorhynchus anatinus*		81 NMT (-)	137 NHT (++)				394 NAT (+)			492 NQT (-)
Frog	*Xenopus tropicalis*		82 NTS (+)	138 NNT (++)	159 NLS (+)		380 NDS (+)	396 NYT (+++)			
Zebrafish	*Danio rerio*		84 NAT (+)	140 NHT (++)			381 NDS (+)	397 NLT (++)			