Comparative studies of vertebrate lipoprotein lipase: A key enzyme of very low density lipoprotein metabolism

Author
Holmes, Roger, L. VandeBerg, John, A. Cox, Laura

Published
2011

Journal Title
Comparative Biochemistry and Physiology Part D: Genomics and Proteomics

DOI
https://doi.org/10.1016/j.cbd.2011.04.003

Copyright Statement
Copyright 2011 Elsevier. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. Please refer to the journal's website for access to the definitive, published version.

Downloaded from
http://hdl.handle.net/10072/42619
Comparative Studies of Vertebrate Lipoprotein Lipase: A Key Enzyme of Very Low Density Lipoprotein Metabolism

Roger S Holmes1,4, John L Vandeberg1,2, and Laura A Cox1,2

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

4Corresponding 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; lipoprotein lipase; evolution; gene duplication.

Running Head: Vertebrate lipoprotein lipase: comparative studies and evolution

Summary
Lipoprotein lipase (LIPL or LPL; E.C.3.1.1.34) serves a dual function as a triglyceride lipase of circulating chylomicrons and very-low-density lipoproteins (VLDL) and facilitates receptor-mediated lipoprotein uptake into heart, muscle and adipose tissue. Comparative LIPL amino acid sequences and protein structures and LIPL gene locations were examined using data from several vertebrate genome projects. Mammalian LIPL genes usually contained 9 coding exons on the positive strand. Vertebrate LIPL sequences shared 58-99% identity as compared with 33-49% sequence identities with other vascular triglyceride lipases, hepatic lipase (LIPC) and endothelial lipase (LIPE). Two human LIPL N-glycosylation sites were conserved among seven predicted sites for the vertebrate LIPL sequences examined. Sequence alignments, key amino acid residues and conserved predicted secondary and tertiary structures were also studied. Potential sites for regulating LIPL gene expression identified were a CpG island within the 5′untranslated region; transcription factor binding sites in the 5′flanking region and the 5′-untranslated region of the human LIPL gene; and a microRNA binding site within the 3′-
untranslated region of the human LIPL gene. Phylogenetic analyses examined the relationships and potential evolutionary origins of vertebrate lipase genes, LIPL, LIPG and LIPC which suggested that these have been derived from gene duplication events of an ancestral neutral lipase gene, prior to the appearance of fish during vertebrate evolution > 500 million years ago. Comparative divergence rates for these vertebrate sequences indicated that LIPL is evolving more slowly (2-3 times) than for LIPC and LIPG genes and proteins.

Introduction

Lipoprotein lipase (LIPL or LPL; E.C.3.1.1.34) is one of three members of the triglyceride lipase family that contributes to vascular lipoprotein degradation and plays major roles in hydrolyzing circulating chylomicrons and very-low-density lipoproteins (VLDL) and in facilitating receptor-mediated lipoprotein uptake into heart, muscle and adipose tissue of the body (Wion et al., 1987; Dichek et al., 1991; Benlian et al., 1996). Hepatic lipase (LIPC; E.C. 3.1.1.3) also serves a dual role in triglyceride hydrolysis and in ligand-binding for receptor-mediated lipoprotein uptake into the liver (Martin et al., 1988; Datta et al., 1988; Cai et al., 1989) whereas endothelial lipase (LIPE; E.C.3.1.1.3) functions in high density lipoprotein (HDL) hydrolysis in the body (Jaye et al., 1999; Hirata et al., 1999). These enzymes are members of the vascular lipase gene family which have significant sequence similarities (Hirata et al., 1999; Ma et al., 2003; Brown & Rader, 2007).

The gene encoding LIPL (LIPL or LPL) is expressed in various cells and tissues of the body, including heart, muscle, adipose tissue, brain, macrophages, lung, lactating mammary gland and endothelial cells where the enzyme hydrolyzes triglycerides from chylomicrons and very-low-density lipoproteins (VLDL) (Wion et al., 1987; Dichek et al., 1991; Benlian et al., 1996; Su et al., 2004). Studies of Lip⁺L⁻/Lip⁻L⁻ knock out mice have shown that LIPL-deficiency causes severe hypertriglyceridemia, reduced high-density lipoprotein (HDL) levels and death within 18 hours of birth (Weinstock et al., 1995). Human clinical studies have also examined loss of function LIPL mutations leading to familial chylomicronemia or hyperlipoproteinemia type I, a rare recessive disorder appearing in children and characterized by dramatically reduced HDL-cholesterol ratios and very high blood triglyceride levels (Amies et al., 1991; Faustinella et al., 1991; Mead et al., 2002). In addition, human LIPL polymorphisms influence significantly a number of major diseases, including atherosclerosis (Reymer et al., 1995; Shimo-Nakanishi et al., 2001; Tsutsumi, 2003), atherosclerotic cerebral infarction (Xu et al., 2008), ischemic stroke (Zhao et al., 2003), coronary artery disease (Zhang et al., 1998; Spence et al., 2003), pre-eclampsia (Hubel et al., 1999; Zhang et al., 2006), Alzheimer's disease (Papassotiropoulos et al., 2005; Blain et al., 2006), ulcerative colitis (Kosaka et al., 2006), hypertension (Chen et al., 2005), diabetes (Ukkola et al., 2005) and obesity (Huang et al., 2006; Radha et al., 2007).
Structures of several vertebrate LIPL genes have been determined, including human (Wion et al., 1987; Chuat et al., 1992), mouse (Zechner et al., 1991), rat (Brault et al., 1992; The MGC Project Team, 2004) and chicken (Cooper et al., 1992). Several LIPL cDNA and amino acid sequences have also been reported for other vertebrates including gorilla (Gorilla gorilla) and rhesus monkey (Macaca mulatta) (Martinez et al., 2001), baboon (Papio anubis) (Cole & Hixson, 1995), pig (Sus scrofa) (Harbitz et al., 1991), cow (Bos taurus) (Senda et al., 1987), sheep (Ovis aries) (Edwards et al., 1993), cat (Felis catus) (Ginzinger et al., 1996), goat (Capra hircus) (Badaoui et al., 2007) and guinea pig (Cavia porcellus) (Enerbaeck et al., 1987) and fish species, sea bass (Dicentrarchus labrax) (Jose Ilbanez et al., 2008) and bream (Sparus aurata; Pagrus major) (Saera-Vila et al., 2005; Oku et al., 2006). LIPL genes usually contain 9 exons of DNA encoding LIPL sequences which may undergo exon shuffling generating several isoproteins in each case (Thierry-Mieg and Thierry-Mieg, 2006). Three dimensional studies of pancreatic lipase (LIPP) (Winkler et al., 1990; Bourne et al., 1994) and molecular modeling of human LPL (van Tilbeurgh et al., 1994) have enabled identification 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. LIPL is synthesized by the endoplasmic reticulum (ER) of parenchymal cells and sequentially processed by the Golgi and ER with the addition of carbohydrate (Ailhaud, 1990; Stins et al., 1993; Hata et al., 1993). LIPL is also subject to proprotein convertase cleavage at a site in the ‘hinge’ region separating the N- and C-terminal enzyme domains (Jin et al., 2005) and behaves as a homodimer with a proposed head-to-tail conformation (Murthy et al., 1996; Wong et al., 1997; Kobayashi et al., 2002). Following secretion, LIPL binds to heparan sulfate proteoglycans on the endothelial surface by electrostatic charge effects onto the luminal surface of capillary endothelial cells and macrophages (reviewed by Tsutsumi, 2003).

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

Methods

Vertebrate LIPL 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 LIPL 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); stickleback (http://www.broadinstitute.org/models/stickleback) (Gasterosteus aculeatus); and seasquirt (Ciona intestinalis) (http://genome.jgi-psf.org/ciona4/ciona4.info.html). 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 LIPL-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 LIPL. Predicted LIPL-like protein sequences were obtained in each case and subjected to analyses of predicted protein and gene structures.

BLAT analyses were subsequently undertaken for each of the predicted LIPL 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 LIPL 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). Structures for human and mouse 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 Lipoprotein Lipases

Predicted secondary and tertiary structures for human and other vertebrate LIPL 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, opossum and frog LIPL tertiary structures, with modeling ranges of residues 36 to 450, 38 to 453 and 23 to 438 respectively. Theoretical isoelectric points and molecular weights for vertebrate LIPL 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 LIPL sequences (Emanuelsson et al 2007). The NetNGlyc 1.0 Server was used to predict potential N-glycosylation sites for vertebrate LIPL proteins (http://www.cbs.dtu.dk/services/NetNGlyc/).

**Comparative Human and Mouse Lipoprotein Lipase Gene (LIPL) Expression**

The genome browser (http://genome.ucsc.edu) (Kent et al. 2003) was used to examine GNF Expression Atlas 2 data using various expression chips for human and mouse LIPL genes (http://biogps.gnf.org) (Su et al, 2004). Gene array expression ‘heat maps’ were examined for comparative gene expression levels among human and mouse tissues showing high (red); intermediate (black); and low (green) expression levels.

**Phylogeny Studies and Sequence Divergence**

Alignments of vertebrate lipoprotein lipase (LIPL), hepatic lipase (LIPC) and endothelial (LIPE) 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 429 residues for comparisons of vertebrate LIPL sequences with human and mouse LIPC and LIPE sequences with the seasquirt (Ciona intestinalis) lipase sequence (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 seasquirt lipase sequence. 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 Lipoprotein Lipase (LIPL) Amino Acid Sequences with Horse Pancreatic Lipase (LIPP)**

The deduced amino acid sequences for opossum (Monodelphis domestica), frog (Xenopus tropicalis) and stickleback (Gasterosteus aculeatus) LIPL are shown in Figure 1 together with previously reported sequences for
human (Wion et al., 1987; Dichek et al., 1991), mouse (Zechner et al., 1991), rat (Brault et al., 1992), chicken LIPL (Cooper et al., 1992; Raisonnier et al., 1995) and horse pancreatic lipase (LIPP) (Bourne et al., 1994) (Table 1).

Alignments of human and other vertebrate LIPL sequences examined showed between 58-99% identities, suggesting that these are products of the same family of genes, whereas comparisons of sequence identities of vertebrate LIPL proteins with human and mouse LIPC and LIPE and horse LIPP exhibited lower levels of sequence identities: LIPC (41% and 44% respectively); LIPG (44% and 45% respectively) and LIPP (24%), indicating that these are members of distinct lipase families (Table 2).

The amino acid sequences for mammalian and chicken LIPL contained 474-478 residues whereas frog (Xenopus tropicalis) and stickleback (Gasterosteus aculeatus) LIPL contained 466 and 514 amino acids, respectively, with the latter having extended N- and C-terminal sequences (Figure 1). Previous three dimensional studies of horse pancreatic lipase (LIPP) (Bourne et al., 1994) and modeling studies of human LIPL (van Tilbeurgh et al., 1994) have enabled predictions of key residues for these vertebrate LIPL proteins (sequence numbers refer to human LIPL). These included the catalytic triad for the active site (Ser159; Asp183; His266); 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 (Cys54/Cys68; Cys243/Cys266; Cys291/Cys302; Cys305/Cys310; Cys445/Cys466) (the latter disulfide bond is apparently absent in the stickleback LIPL sequence); the predicted ‘lid’ region (244-265) which covers the active site and participates in lipid substrate binding in analogous lipases (Winkler et al., 1990; Bourne et al., 1994); and a predicted ‘hinge’ region for vertebrate LIPL, containing a proteolytic cleavage site for proprotein convertase (320Arg-321Ala-322Lys-323Arg) (Jin et al., 2003; 2005) (see Figure 2). Specific tyrosine residues predicted for nitration following lipopolysaccharide (LPS) challenging were identified (tyrosines 121, 127 and 314) which down-regulate LIPL activity and reduce triglyceride clearance from the body (Casanovas et al, 2009). These residues were conserved for all of the vertebrate LIPL sequences examined (Figure 1) possibly because of this role in reducing LIPL activity following LPS administration, resulting in ‘lipemia’ and increased binding of triglyceride-rich lipoproteins with LPS which undergo clearance by the liver (Gouni et al, 1993). 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 LIPL structure and function. The N-terminal region (residues 1-32 for human LIPL) 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 LIPL.

Heparin binding sites have been previously shown to play key roles in binding LIPL and related neutral vascular lipases, LIPC and LIPE, to heparin sulfate proteoglycans on the luminal side of endothelial cells (Hill et al., 1998; Sendak & Bensadoun, 1998). Figure 2 summarizes the comparative amino acid sequences for vertebrate LIPL for several sites previously investigated by Beg and coworkers (1998). A nonbasic amino acid region similar to type 1 repeats of thrombospondin (TSP) (Prater et al., 1991) and four basic amino acid clusters are compared for 14 vertebrate LIPL sequences, including 2 fish species, zebrafish (Danio rerio) and stickleback (Gasterosteus aculeatus). Human LIPL contains a TSP sequence (residues 414-421: Phe-Ser-Trp-Ser-Asp-Trp-Trp-Ser) similar to the repeats found in thrombospondin 1 (TSP1) that mediates cell-to-cell and cell-to-matrix interactions (see Wolf et al., 1990). The first and second Trp residues in this sequence were retained for all vertebrate LIPL sequences examined which is consistent with in vitro studies of Beg and coworkers (1998) for synthetic LIPL peptides. Comparisons of the four basic amino acid clusters showed conservation for these sequences (human LIPL numbers used): cluster 1 (residues 287-312) retained Arg288, Lys292, Lys297, Arg304, Lys305 and Arg307 for all of the vertebrate LIPL sequences examined with the exception of Lys292, which was substituted by Arg for the fish sequences, and Arg304, substituted by Lys for opossum LIPL; cluster 2 (residues 315-332) retained Lys317, Arg319, Lys321, Arg322, Lys325, Lys330 and Arg332, with the exception of chicken LIPL (first Lys substituted by Arg) and the fish LIPL sequences examined for the second Lys, which was substituted by Arg (Danio rerio) or by Thr (Gasterosteus aculeatus); cluster 3 (residues 173-177) retained Lys173, Lys174 and Arg177 with the exception of the fish LIPL sequences for which the first Lys was substituted by His; and cluster 4 (residues 329-441) which retained Lys330, Arg332, Lys334, Lys440 and Lys441, with the exception of Lys330 which was substituted by Arg for mouse and chicken LIPL sequences and of Lys440, which was substituted by Gln for the fish LIPL sequences. Synthetic peptide heparin binding properties for these clusters (Beg et al, 1998) is consistent with these results with clusters 1, 3 and 2, showing the strongest binding in vitro, whereas cluster 4 did not bind to heparin, under the conditions used in their study.

One N-glycosylation site has been previously reported for human LIPL at 70Asn-71His-72Thr and (Kobayashi et al, 1994) and a second site predicted at 386Asn-387-Lys-388Thr (van Tilbeurgh et al, 1994). A comparative analysis of potential N-glycosylation sites for vertebrate LIPL (Table 3) has shown that there are 7 sites overall although only two of these have been predominantly retained for the 20 vertebrate LIPL sequences examined (designated as sites 2 and 4) (with the exception of stickleback LIPL which has lost site 2 but gained site 1 at 34Asn-
35Thr-36Thr) (Table 3). It is apparent from the study by Kobayashi and coworkers (1994) of a human LIPL Asn70 variant that this N-glycosylation site is essential for catalysis and secretion. A key role for the second predicted N-glycosylation site is also likely given the conservation of this site for all vertebrate LIPL sequences examined (Table 3). Other vascular neutral lipases contained four N-glycosylation sites which play key roles and contribute to enzyme stability, secretion and catalytic activity: LIPC (hepatic lipase) (Wolle et al. 1993; Ben-Zeev et al. 1994) and LIPE (endothelial lipase) (Miller et al. 2004; 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).

**Predicted Secondary and Tertiary Structures for Vertebrate LIPL**

Predicted secondary structures for vertebrate LIPL 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 LIPL sequences were 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 LIPL numbers used) Ser159, Asp 183 and His269; the ‘lid’ domain (residues 243-265); and the ‘hinge’ region (residues 317-330) which concludes with a β-sheet: Lys326-Met327-Tyr328-Leu329-Lys330. Figure 3 describes predicted tertiary structures for human, opossum (Monodelphis domestica) and frog (Xenopus tropicalis) LIPL protein sequences and shows significant similarities for these polypeptides with horse pancreatic lipase (LIPP) (Bourne et al., 1994). The three LIPL and LIPP 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 LIPE) (Dugi et al., 1995; Kobayashi et al., 1996) and may play a major role in determining the preference for triglyceride rich lipoprotein LIPL substrates. A ‘hinge’ region was also observed for vertebrate LIPL proteins, separating the ‘lipase’ and ‘plat’ domains, with the latter having a ‘sandwich-like’ β-pleated sheet structure. The ‘plat’ domain for LIPC and LIPE 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). Biochemical studies have also shown that LIPL behaves as a dimer (Olivecrona & Bengtsson-Olivecrona, 1983). In addition, a proprotein convertase proteolytic cleavage site was observed at the ‘hinge’ region (Arg319-X320-Lys321-Arg322) (Figure 2), which may result in partially cleaved dimeric LIPL forms with reduced activities and unknown biochemical roles, similar to those observed for LIPE (Griffon et al, 2009). Comparisons of amino acid sequences studies for other vertebrate LIPL proteins suggest that these properties and key sequences are substantially retained for all of the sequences examined.
Gene Locations and Exonic Structures for Vertebrate LIPL Genes

Table 1 summarizes the predicted locations for vertebrate LIPL genes based upon BLAT interrogations of several vertebrate genomes using the reported sequences for human (Wion et al., 1987; Chuat et al., 1992), mouse (Zechner et al., 1991), rat (Brault et al., 1992; The MGC Project Team, 2004) and chicken (Cooper et al., 1992) and the predicted sequences for other vertebrate LIPL genes and the UC Santa Cruz genome browser (Kent et al. 2003). The predicted vertebrate LIPL genes were predominantly transcribed on the positive strand, with the exception of the marmoset, rat, dog, platypus, chicken and frog genes, which were transcribed on the negative strand. Figure 1 summarizes the predicted exonic start sites for human, mouse, rat, opossum, chicken, frog and stickleback LIPL genes with each having 9 coding exons, in identical or similar positions to those predicted for the human LIPL gene (Wion et al., 1987; Chuat et al., 1992), with the exception of stickleback LIPL, which contained an additional exon encoding an extended C-terminal sequence.

Figure 4 shows the predicted structures of mRNAs for human and mouse LIPL transcripts for the major isoform in each case (Theirry-Mieg & Thierry-Mieg, 2006). The transcripts were 27-28.5 kbs in length with 9 introns present for these LIPL mRNA transcripts and in each case, an extended 3'-untranslated region (UTR) was observed. The human LIPL genome sequence contained several predicted transcription factor binding sites (TFBS), a microRNA site (miR-29) located in the 3'-untranslated region and a CpG island (CpG112) which includes the 5'-untranslated region of human LIPL on chromosome 8. This CpG island within the LIPL gene promoter may play a role in maintaining a very high level of gene expression (4.5 times the average for human genes) (Theirry-Mieg & Thierry-Mieg, 2006) which is similar to the CpG islands within housekeeping gene promoters expressed in most tissues (Saxonov et al., 2006). Of particular significance among the TFBS sites observed is the PPARγ (peroxisome proliferator-activated receptor-γ) site which participates in regulating cholesterol uptake and efflux from macrophages, promotes uptake of oxidized LDL and the subsequent differentiation of monocytes into foam cells (Merkel et al, 2002; Stein & Stein, 2003; AshokKumar et al, 2010). In addition, an extensively studied LIPL genetic variant (Pro12Ala) is associated with lowered transcription of PPARγ2 and a decreased risk for type 2 diabetes in human populations (Nagy & Szanto, 2005). The prediction of a microRNA (MiR-29) binding site within the 3’ untranslated region of human LIPL is also potentially of major significance because small noncoding RNAs regulate mRNA and protein expression and have been implicated in regulating gene expression during embryonic development (Stephani & Slack, 2008).

Comparative Human and Mouse LIPL Tissue Expression
Figure 5 presents ‘heat maps’ showing comparative gene expression for various human and mouse tissues obtained from GNF Expression Atlas Data using the U133A and GNF1H (human) and GNF1M (mouse) chips (http://genome.ucsc.edu; http://biogps.gnf.org) (Su et al, 2004). These data supported a broad and high level tissue expression for human and mouse LIPL, particularly for heart, skeletal muscle, adipose tissue and lung, which is consistent with previous reports for these genes (Levak-Frank et al, 1999; Mead et al, 2002). Other comparisons of human and mouse LIPL tissue expression indicated significant species differences, with higher levels of gene expression observed in human placenta and nerve tissues but lower expression levels in human liver, kidney and ovary than for the corresponding mouse tissues. Overall however, human and mouse LIPL tissue expressions levels were >4 times the average level of gene expression which supports the key role played by this enzyme in lipid metabolism, especially in heart, skeletal muscle and adipose tissue (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) (Theirry-Mieg & Thierry-Mieg, 2006).

Phylogeny and Divergence of LIPL and Other Vertebrate Lipase Sequences

A phylogenetic tree (Figure 6) was calculated by the progressive alignment of 20 vertebrate LIPL amino acid sequences with human and mouse LIPC and LIPE sequences which was ‘rooted’ with the Ciona intestinalis (sea squirt) lipase sequence (see Table 1). The phylogram showed clustering of the LIPL sequences into groups which were consistent with their evolutionary relatedness as well as groups for human and mouse LIPC and LIPE sequences, which were distinct from the sea squirt lipase sequence. These groups were significantly different from each other (with bootstrap values of ~ 100/100). It is apparent from this study of vertebrate LIPL genes and proteins that this is an ancient protein for which a proposed common ancestor for the LIPC, LIPG and LIPL neutral lipase genes may have predated the appearance of fish > 500 million years ago (Donoghue & Benton, 2007). This proposal is consistent with a previous report from Cohen (2003) which described predicted amino acid sequences for human and pufferfish (Takifugu rubripes) LIPE, LIPL and LIPC. Genetic distances for human, cow, mouse and rat LIPL, LIPG and LIPC sequences calculated from a mammalian common ancestor were as follows: 0.037±0.007, 0.086±0.007 and 0.125±0.011, respectively, which suggest that mammalian LIPL sequences are diverging ~ 2-3 times more slowly than for LIPC and LIPE sequences. This is indicative of a conservative LIPL protein during mammalian evolution.

Conclusions

The results of the present study indicate that vertebrate LIPL 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; Jaye et al., 1999; Hirata et al., 1999). This enzyme has a unique property among the neutral lipases studied in hydrolyzing circulating chylomicrons and very low density lipoproteins (VLDL) and in facilitating receptor-mediated lipoprotein uptake into heart, muscle and adipose tissue of the body (Wion et al., 1987; Dichek et al., 1991; Benlian et al., 1996). LIPL is encoded by a single gene among the vertebrate genomes studied which is highly expressed in human and mouse tissues, particularly in heart, adipose tissue and skeletal muscle, and usually contained 9 coding exons. Predicted secondary structures and tertiary structures for vertebrate LIPL proteins showed a strong similarity with horse pancreatic lipase (LIPP) (Bourne et al., 1994). Three major structural domains were apparent for vertebrate LIPL, 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 20 vertebrate LIPL with human and mouse LIPC and LIPE sequences indicated that the LIPL gene has appeared early in vertebrate evolution, prior to the appearance of bony fish more that 500 million years ago, and is evolving 2-3 times more slowly that the other neutral lipase genes and proteins during mammalian evolution.

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. We also acknowledge the expert assistance of Dr Bharet Patel of Griffith University with the phylogeny studies.

REFERENCES


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


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


Levak-Frank S, Hofmann W, Weinstock PH, Radner H, Sattler W, Breslow J.L, Zechner R (1999) Induced mutant mouse lines that express lipoprotein lipase in cardiac muscle, but not in skeletal muscle and adipose tissue, have normal plasma triglyceride and high-density lipoprotein-cholesterol levels. Proc Natl Acad Sci USA 96:3165-3170


Figure 1: Amino Acid Sequence Alignments for Vertebrate Lipoprotein Lipase and Horse Pancreatic Lipase Sequences

See Table 1 for sources of lipoprotein lipase (LIPL or LPL) and pancreatic lipase (LIPP) sequences: Hu-human LIPL; Mo-mouse LIPL; Ra-rat LIPL; Op-opossum LIPL; Fr-frog LIPL; St-stickleback LIPL; 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; human N-glycosylated and potential N-glycosylated Asn sites are in green bold; active site (A) 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 LIPL is in shaded yellow; β-sheet for horse LIPP or predicted for vertebrate LIPL is in shaded grey; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human LIP L 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 LIPE dimer model reported by Griffon et al (2009); Ø refers to amino acid substitutions reported for human LIPL that result in activity loss and associated lipoproteinemia.

Figure 2: Amino Acid Sequence Alignments for Predicted Heparin Binding Sites for Vertebrate Lipoprotein Lipases

Basic amino acid and thrombospondin (TSP)-like clusters reported by Beg et al (1998) for human LIPL; K-lysine residue; R-arginine residue; W-tryptophan residue; numbers refer to C-terminal residue for each of the clusters examined for the 14 vertebrate LIPL sequences examined; for consensus sequences, X refers to any amino acid.
The structure for horse pancreatic lipase (LIPP) is taken from Bourne et al., 1994; predicted human, opossum and frog lipoprotein lipase (LIPL) 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 LIPL; 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: Tertiary Structure for Horse Lipoprotein Lipase and Predicted Tertiary Structures for Human, Opossum and Frog Lipoprotein Lipases**

<table>
<thead>
<tr>
<th>Figure 4: Gene Structures and Major Splicing Variant for the Human and Mouse Lipoprotein Lipase (LIPL) Genes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derived from the AceView website <a href="http://www.ncbi.nlm.nih.gov/IEB/Research/Assembly/">http://www.ncbi.nlm.nih.gov/IEB/Research/Assembly/</a> (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 (miR29) observed at or near the human LIPL 3’ untranslated region; sizes of mRNA sequences are</td>
</tr>
</tbody>
</table>
shown in kilobases (kb); predicted transcription factor binding sites (TFBS) for human LIPL are shown: FoxJ2-Foxhead J2 protein; Zic1-Zinc finger protein Zic1; E2F-transcription factor E2F2; HNF4 hepatocyte nuclear factor; ER-estrogen receptor DNA binding; PPARG-peroxisome proliferator-activated receptor γ.

Figure 5: Comparative Tissue Expression for Human and Mouse Lipoprotein Lipase Genes (LIPL)

Expression ‘heat maps’ (GNF Expression Atlas 2 data) ([http://biogps.gnf.org](http://biogps.gnf.org)) (Su et al, 2004) were examined for comparative gene expression levels among human and mouse tissues for LIPL genes showing high (red); intermediate (black); and low (green) expression levels. Derived from human and mouse genome browsers ([http://genome.ucsc.edu](http://genome.ucsc.edu)) (Kent et al. 2003).
Figure 6: Phylogenetic Tree of Vertebrate Lipoprotein Lipase (LIPL) and Human and Mouse Hepatic Lipase (LIPC) and Endothelial Lipase (LIPE) Amino Acid Sequences

The tree is labeled with the lipase name and the name of the animal and is ‘rooted’ with the seasquirt lipase sequence (*Ciona intestinalis*). Note the 3 major clusters corresponding to the LIPL, LIPG and LIPC neutral lipase gene families, encoding lipoprotein lipase (LIPL), endothelial lipase (LIPE) and hepatic lipase (LIPC), respectively. 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.

Table 1: Vertebrate lipoprotein lipase (LIPL), seasquirt (*Ciona intestinalis*) lipase, human and mouse hepatic lipase (LIPC) and endothelial lipase (LIPG) and horse pancreatic lipase (LIPP) genes and proteins

RefSeq: the reference amino acid sequence; *¹* predicted Ensembl amino acid sequence; *²* not available; *³* Contig refers to a DNA scaffold for sequencing analyses; GenBank IDs are derived NCBI [http://www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/); Ensembl ID was derived from Ensembl genome database [http://www.ensembl.org](http://www.ensembl.org); UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual acid lipases (see [http://kr.expasy.org](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 lipoprotein lipase (LIPL), human and mouse hepatic lipase (LIPC) and endothelial lipase (LIPE), horse pancreatic lipase (LIPP) and seasquirt lipase amino acid sequences

<table>
<thead>
<tr>
<th>Gene LIP</th>
<th>Homo sapiens</th>
<th>Rhesus</th>
<th>Mouse</th>
<th>Rat</th>
<th>Horse</th>
<th>Opossum</th>
<th>Chicken</th>
<th>Stickleback</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Human</th>
<th>Horse</th>
<th>LIPP</th>
<th>SEASQUIRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIPL</td>
<td>100</td>
<td>99</td>
<td>92</td>
<td>91</td>
<td>94</td>
<td>92</td>
<td>81</td>
<td>71</td>
<td>72</td>
<td>58</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>LIPC</td>
<td>99</td>
<td>100</td>
<td>96</td>
<td>91</td>
<td>80</td>
<td>89</td>
<td>80</td>
<td>71</td>
<td>71</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPE</td>
<td>91</td>
<td>100</td>
<td>96</td>
<td>90</td>
<td>89</td>
<td>80</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPE</td>
<td>94</td>
<td>100</td>
<td>96</td>
<td>91</td>
<td>100</td>
<td>91</td>
<td>86</td>
<td>81</td>
<td>81</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPE</td>
<td>99</td>
<td>100</td>
<td>96</td>
<td>90</td>
<td>89</td>
<td>80</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPE</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>91</td>
<td>100</td>
<td>91</td>
<td>86</td>
<td>81</td>
<td>81</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPP</td>
<td>81</td>
<td>82</td>
<td>80</td>
<td>90</td>
<td>79</td>
<td>85</td>
<td>81</td>
<td>71</td>
<td>71</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPP</td>
<td>82</td>
<td>84</td>
<td>80</td>
<td>90</td>
<td>79</td>
<td>85</td>
<td>81</td>
<td>71</td>
<td>71</td>
<td>57</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPP</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>40</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPP</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>40</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Numbers show the percentage of amino acid sequence identities. Numbers in **bold** show higher sequence identities for lipases from the same gene family. LIPL, LIPC, LIPE and LIPP refer to lipoprotein lipase, hepatic lipase, endothelial lipase and pancreatic lipase amino acid sequences, respectively.

Table 3: Predicted N-glycosylation sites for vertebrate lipoprotein lipases

<table>
<thead>
<tr>
<th>Genes</th>
<th>Site</th>
<th>Homo sapiens</th>
<th>Mouse</th>
<th>Rat</th>
<th>Horse</th>
<th>Opossum</th>
<th>Chicken</th>
<th>Stickleback</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Human</th>
<th>Horse</th>
<th>LIPP</th>
<th>SEASQUIRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIPL</td>
<td>7</td>
<td>60</td>
<td>60</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>47</td>
<td>47</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPL</td>
<td>6</td>
<td>60</td>
<td>60</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>47</td>
<td>47</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>LIPL</td>
<td>5</td>
<td>60</td>
<td>60</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>47</td>
<td>47</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Numbers refer to amino acids in the acid sequences, including N-asparagine; K-lysine; I-isoleucine; H-histidine; S-serine; T-threonine; Q-glutamine; D-aspartate; Y-tyrosine; and V-valine. Note that there are 7 potential sites identified, including 2 sites for human LIPL (sites 2 and 4) (see Kobayashi et al, 1994; van Tilbeurgh et al, 1994). N-glycosylation sites were identified using the NetNGlyc 1.0 web server (http://www.cbs.dtu.dk/services/NetNGlyc/).
<table>
<thead>
<tr>
<th>Vertebrate</th>
<th>Species</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>Site 5</th>
<th>Site 6</th>
<th>Site 7</th>
<th>No of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo abelii</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Rhesus</td>
<td>Macaca mulatta</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Baboon</td>
<td>Papio anubis</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Callithrix jacchus</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>411NDS</td>
<td>3</td>
</tr>
<tr>
<td>Guinea Pig</td>
<td>Cavia porcellus</td>
<td>70NHS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Horse</td>
<td>Equus caballus</td>
<td>63NQS</td>
<td>379NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Cow</td>
<td>Bos taurus</td>
<td>73NHS</td>
<td>389NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis familiaris</td>
<td>66NHT</td>
<td>382NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
<td>69NHS</td>
<td>385NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>471NKS</td>
<td>3</td>
</tr>
<tr>
<td>Pig</td>
<td>Sus scrofa</td>
<td>73NHS</td>
<td>389NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Elephant</td>
<td>Loxodonta africana</td>
<td>70NYS</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Opossum</td>
<td>Monodelphis domestic</td>
<td>73NHS</td>
<td>389NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>452NIS</td>
<td>3</td>
</tr>
<tr>
<td>Platypus</td>
<td>Ornithorhynchus anatinus</td>
<td>71NHT</td>
<td>387NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>70NHT</td>
<td>354NVT</td>
<td>386NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Freg</td>
<td>Xenopus tropicalis</td>
<td>60NHT</td>
<td>344NLT</td>
<td>376NKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Stickleback</td>
<td>Gasterosteus aculeatus</td>
<td>34NTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>407NTT</td>
<td>2</td>
</tr>
</tbody>
</table>