review

Structure and evolution of the lipase superfamily

Winston A. Hide,* Lawrence Chan,* and Wen-Hsiung Li†

Departments of Cell Biology and Medicine,* Baylor College of Medicine, Houston, TX 77030, and Center for Demographic and Population Genetics,† University of Texas, P.O. Box 20344, Houston, TX 77225

Abstract The lipase superfamily includes three vertebrate and three invertebrate (dipteran) proteins that show significant amino acid sequence similarity to one another. The vertebrate proteins are lipoprotein lipase (LPL), hepatic lipase (HL), and pancreatic lipase (PL). The dipteran proteins are *Drosophila* yolk proteins 1, 2, and 3. We review the relationships among these proteins that have been established according to gene structural relatedness and introduce our findings on the phylogenetic relationships, distance relationships, and evolutionary history of the lipase gene superfamily. *Drosophila* yolk proteins contain a 104 amino acid residue segment that is conserved with respect to the lipases. We have used the yolk proteins as an outgroup to root a phylogeny of the lipase family. Our phylogenetic reconstruction suggests that ancestral PL diverged earlier than HL and LPL, which share a more recent root. Human and bovine LPL are shown to be more closely related to murine LPL than to guinea pig LPL. A comparison of the distance (a measure of the number of substitutions between sequences) between mammalian and avian LPL reveals that guinea pig LPL has the largest distance from the other mammals. Human, rodent, and rabbit HL show marked divergence from one another, although they have similar relative rates of amino acid substitution when compared to human LPL as an outgroup. Human and porcine PL are not as divergent as human and rat HL, suggesting that PL is more conserved than HL. However, canine PL demonstrates an unusually rapid rate of substitution with respect to the other pancreatic lipases. The lipases share several structurally conserved features. One highly conserved sequence (Gly-Xaa-Ser-Xaa-Gly) contains the active site serine. This feature, which agrees with that found in serine esterases and proteases, is found within the entire spectrum of lipases, including the evolutionarily unrelated prokaryotic lipases. We review the location and possible activity of putative lipid binding domains. We have constructed a conservation index (CI) to display conserved structural features within the lipase gene family, a CI of 1.0 signifying perfect conservation. We have found a correlation between a high CI and the position of conserved functional structures. The putative lipid-binding domains of LPL and HL, the disulfide-bridging cysteine residues, catalytic residues, and N-linked glycosylation sites of LPL, HL, and PL all lie within regions having a CI of 0.8 or higher. A number of amino acid substitutions have been identified in familial hyperchylomicronemia which result in loss of LPL function. These mutations are located at residues that have a CI of between 0.8 and 1.0. We have compared the positions of known in vivo substitutions with the degree of conservation of residues in human LPL. We discuss the conserved structures that appear to have a role in the conformational changes during catalytic activity of the lipases. Although exon/intron boundary positions are largely conserved within the lipases, the positions of boundaries are not coincident with low CI values. HL and LPL apparently have lost introns that still remain in PL.—**Hide, W. A., L. Chan, and W-H. Li.** Structure and evolution of the lipase superfamily. *J. Lipid Res.* 1992. **33:** 167–178.

Supplementary key words lipoprotein lipase • pancreatic lipase • hepatic lipase • *Drosophila* yolk protein • active site serine

The lipase gene family is made up of three vertebrate genes that share structural similarities and are derived from a common ancestral gene. The family includes lipoprotein lipase (LPL), hepatic lipase (HL), and pancreatic lipase (PL). The known function of these proteins is hydrolysis of circulating and dietary triglycerides, allowing subsequent assimilation and distribution to central and peripheral tissues. Three *Drosophila* yolk proteins also show limited sequence similarity to the lipases and will also be examined in the review.

LPL and HL are distinct enzymes but have very similar hydrolytic functions. They both hydrolyze lipoprotein triglycerides. LPL is essential for chylomicron and very low density lipoprotein catabolism and also for the transfer of cholesterol, phospholipids, and apolipoproteins among lipoprotein particles (1). It is dependent on apolipoprotein C-II for activation (2, 3). HL is also important in lipoprotein and phospholipid metabolism (4) and, through its actions on high density lipoproteins, may mediate delivery of cholesterol from peripheral tissues to the liver (5–7). It may also be involved in the metabolism of intermediate density lipoprotein to low density lipoprotein in the liver (4).

Abbreviations: LPL, lipoprotein lipase; HL, hepatic lipase; PL, pancreatic lipase; CI, conservation index.

ASBMB JOURNAL OF LIPID RESEARCH JLR

Downloaded from www.jlr.org by guest, on June 18, 2012

PL hydrolyzes dietary lipid in the intestine. It binds to the lipid interface of luminal contents emulsified in the presence of bile salts, colipase, and calcium and acts to hydrolyze ester bonds in dietary triglyceride (8).

Drosophila yolk proteins lack lipolytic activity and show no obvious functional similarity to the lipases. However, a 104 residue segment of *Drosophila* yolk protein 1 (YP1), 2 (YP2), and 3 (YP3) shows significant similarity at the amino acid level to the lipase family (9–11), including a region that has sequence similarity to the substrate-binding site of porcine PL (12).

Despite their disparate anatomical location and distinct physiological functions, lipases share several common structural features. The proteins in the lipase superfamily have significant amino acid sequence similarity and can be readily aligned. Detailed comparisons of structural similarities, sequence similarities, and exon/intron boundary distribution have led to the proposal of different models for evolutionary relationships of LPL with HL and PL and *Drosophila* yolk proteins (10, 13). However, none of these models are based on comprehensive phylogenetic or genetic distance comparison among members of the complete family.

By examination and comparison of structure at the amino acid sequence level, we have set out to establish the relationship of conserved structural to functional features. We review previous findings together with our analysis of evolutionary relationships between members of the lipase gene family.

EVOLUTIONARY RELATIONSHIPS

The amino acid sequences of all the published LPL, HL, and PL from different species can be readily aligned (**Fig. 1**). Based on this alignment, the relatedness among the three lipases has been deduced, and a phylogenetic reconstruction has been made by using the maximum parsimony method (Hide, W. A., L. Chan, and W-H. Li, unpublished results). The model presented in **Fig. 2** suggests that LPL and HL share a more recent ancestor and PL has branched off earlier than LPL and HL. This model agrees with that proposed by Persson et al. (10), who based their conclusions on the position of Cys residues, N-linked glycosylation sites, and polyanion-binding sites among the three lipases.

In addition to the study of Persson et al. (10), two other analyses of relationships among the proteins of the lipase family have been presented. Datta et al. (14) performed a study using distance data and comparative rates to determine the relationships of human and rat HL, human, mouse, and bovine LPL, and dog and pig PL. Times of divergence estimated from fossil evidence were used to derive rates of evolution, which were used to predict the likely relationships of the PL, HL, and LPL. Even though distance data suggested otherwise, a prediction was made that HL was more closely related to PL than to LPL. The rate of evolution of PL was found to be twice that of HL and about seven times greater than that of LPL. The gene relationships were estimated by use of differential rates of evolution but lacked the benefit of an outgroup against which they could be measured. In the present reanalysis, considerably more data have become available. Since it was pointed out that *Drosophila* YP genes have a region of significant sequence similarity to the lipases, these sequence data provide a reference for a more plausible phylogenetic reconstruction (Fig. 2) based on parsimony analysis rather than on simple rates and known divergence times as was previously done (14). Kirchgessner et al. (13) speculated on the evolutionary relationships of mammalian LPL, HL, PL, and *Drosophila* yolk protein 1 based on the structural organization of these genes and on exon shuffling and intron loss. Their model suggests that a primordial gene gave rise through an initial gene duplication, in-

Fig. 1. Alignment of amino acid sequences of mature lipase proteins and conserved region of *Drosophila* YP1, YP2, and YP3. Alignment of human (47), bovine (48), mouse (49), guinea pig (50), and chicken (43) LPL; human (14), rat (33), and rabbit (23) HL; human (51), pig (52), and dog (8) PL; and a conserved region of *Drosophila* yolk proteins 1 (53), 2 (54), and 3 (55). Alignment was performed using the algorithm of the computer program CLUSTAL (56) to produce an alignment that required the minimum number of gaps to achieve a parsimonious alignment of amino acids. Each gap was given a penalty of ten. The penalty specifies the number of exactly matching residues that must be found by introducing a gap. After computer alignment, careful visual adjustments were made. Several small gaps were removed. The N-terminal and C-terminal regions of the proteins are difficult to align and should be interpreted with caution. These regions were not used for phylogenetic comparison. The regions used for phylogenetic comparison were residues 68–170 and 180–359. Shaded-underlined residues represent the hydrophobic flanking "wings" surrounding the catalytic site in the PL three-dimensional structure. Shaded-boxed residues represent predicted N-linked glycosylation sites. Double bars represent putative lipid-binding domains. The α-helical "flap" is represented by a hatched bar. Conserved cysteine residues are shaded and are highlighted by solid circles. The disulfide-bridging residues in PL are shown connected by solid lines. The location of the three residues in the catalytic triad: serine (S), aspartic acid (D), and histidine (H) are marked by ⊕. The consensus sequence is marked by a dash at positions where there is no consensus, a capital letter where there is an identity, and a "+" where there is an evolutionarily conserved residue. Our alignment differs from a previously published alignment in the N-terminal region from consensus residues 1–47 and in the C-terminal region from consensus residues 420–513 (14). The previous study did not include guinea pig and chicken LPL, rabbit HL, and human PL and was performed manually, without bias for alignment of evolutionary conserved residues and without sufficient weighting to reduce the number of gaps. We have now used a multiple alignment algorithm for alignment (56). For phylogenetic comparison, we have excluded regions that produced ambiguous alignments, and restricted our comparison to those regions that are most highly conserved.

Downloaded from www.jlr.org by guest, on June 18, 2012

ASBMB JOURNAL OF LIPID RESEARCH JLR

ASBMB
JOURNAL OF LIPID RESEARCH
JLR

```
          1                                                                                              100
humlpl    ....ADQRRDFIDIESKFALRTPEDT......................AEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGMYESWVPKLVAALY
Bovlpl    ..DRITGGKDFRDIESKFALRTPEDT......................AEDTCHLIPGVTESVANCHFNHSSKTFVVIHGWTVTGMYESWVPKLVAALY
muslpl    ...AADAGRDFSDIESKFALRTPEDT......................AEDTCHLIPGLADSVSNCHFNHSSKTFVVIHGWTVTGMYESWVPKLVAALY
Gplpl     ....ANCQKDYTDIESKFARRTPENT......................VEDTCHLIPGVTESVANCHFNHSSKTFMVIHGWTVTGMYESWVPKLVAALY
cknlpl    ..SDPEAEMNFEGIESKFSLRTPAEP......................DEDVCYLVPGQMDSLAQCNFNHTSKTFVVIHGWTVTGMYESWVPKLVDALY
humhl     .......LGQSLKPEPFGRRAQAVETNKTLHEMK....TRFLLGETNQ....GCQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGVLENWIWQMVAALK
rabhl     .......HGQSLRPEPFGRRARVTATKKTLLETE....TRFLLFKDKA..NKGCQIRLHHADTLQECGFNSSLPLVMIVHGWSVDGLLESWIWQMVAALK
rathl     .......CGQGVGTEPFGRNLGATEERKPLQKPE....IRFLLFKDES.DRLGCQLRPQHPETLQECGFNSSHPLVMIIHGWSVDGLLETWIWKIVGALK
humpl     KEVCYERLGCFSDDSPWSGITE..RPLHILPWSPKDVNTRFLLY.TNENPNNFQEVAA.DSSSISGSNFKTNRKTRFIIHGFIDKGE.ENWLANVCKNLF
pigpl     SEVCFPRLGCFSDDAPWAGIVQ..RPLKILPPDK.DVDTRFLLY.TNQNQNNYQELVA.DPSTITNSNFRMDRKTRFIIHGFIDKGE.EDWLSNICKNLF
dogpl     KEVCYEQIGCFSDAEPWAGTAI..RPLKVLPWSPERIGTRFLLY.TKNKPNNFQTLLPSDPSTIEASNFQTDKKTRFTIHGFINKGE.ENWLLDMCKNMF
droypl    ...................................................................................................
droyp2    ...................................................................................................
droyp3    ...................................................................................................
Consensus -------------------------------------------------------------+++--+-F---------+HG+--+G--E+W+-++--++-

          101                                                                                            200
humlpl    KREPDS.NVIVVDWLSRAQEHYPVSAGYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAGIAGS....LTNKKVNRITGLDPAGPNFEYAEAP
Bovlpl    KREPDS.NVIVVDWLSRAQQHYPVSAGYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAGIAGS....LTNKKVNRITGLDPAGPNFEYAEAP
muslpl    KREPDS.NVIVVDWLYRAQQHYPVSAGYTKLVGNDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAGVAGS....LTNKKVNRITGLDPAGPNFEYAEAP
Gplpl     KREPDS.NVIVVDWLRRAQHHYPESADYTKLVGEDVARFINWMEDEFKYSVDNVHLLGYSLGAHAAGVAGS....RTNTKVSRITGLDPAGPNFEYAEAT
cknlpl    KREPDS.NVIVVDWLVRAQQHYPVSAAYTKLVGKDVAMFIDWMEEKFNYPLNNVHLLGYSLGAHAAGIAGS....LTKKKVNRITGLDPAGPTFEYADAP
humhl     SQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVGKEVAALLRWLEESVQLSRSHVHLIGYSLGAHVSGFAGSSIGG..THKIGRITGLDAAGPLFEGSAPS
rabhl     SQPARPVNVGLVDWISLAHSHYAVAVRNARLVGQEVAALLQWLEESAPFSRSNVHLIGYSLGAHVSGFAGSYISG..KHKIGRITGLDPAGPMFEGTSAS
rathl     SRQSQPVNVGLVDWISLAYQHYAIAVRNTRVVGQEVAALLLWLEESMKFSRSKVHLIGYSLGAHVSGFAGSSMGG..KRKIGRITGLDPAGPMFEGTSPN
humpl     KVE..SVNCICVDWKGGSRTGYTQASQNIRIVGAEVAYFVEFLQSAFGYSPSNVHVIGHSLGAHAAGEAGRRTNG....TIGRITGLDPAEPCFQGTPEL
pigpl     KVE..SVNCICVDWKGGSRTGYTQASQNIRIVGAEVAYFVEVLKSSLGYSPSNVHVIGHSLGSHAAGEAGRRTNG....TIERITGLDPAEPCFQGTPEL
dogpl     KVE..EVNCICVDWKKGSQTSYTQAANNVRVVGAQVAQMLSMLSANYSYSPSQVQLIGHSLGAHVAGEAGSRTPG.....LGRITGLDPVEASFQGTPEE
droypl    ..........................GAKIGKWIVQMVNELDMPFDTIHLIGQNVGAHVAGAAAQEFTRLTGHKLRRVTGLDPSKIVAKSKNTL
droyp2    ..........................GEIIGNRLVELTNTVNVPQEIIHLIGSGPAAHVAGVAGRQFTRQTGHKLRRITALDPTKIYGKPEERL
droyp3    ..........................GAMIGQTLIDLTN.KGVPQEIIHLIGQGISAHVAGAAGNKYTAQTGHKLRRITGLDPAKVLSKRPQIL
Consensus +------N---VDW---+---Y-+----++VG------+---+-+----++-+-++++G-+--++H++G-A+------------R+T+LD+-----+-----

          201                                                                                            300
humlpl    SRLSPDDADFVDVLHTFTRGSP.GRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVD.QLVKCSHERSIHLFIDSLLNEENPSKAYRCSS
bovlpl    SRLSPDDADFVDVLHTFTRGSP.GRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVD.QLVKCSHERSVHLFIDSLLNEENPSKAYRCNS
muslpl    SRLSPDDADFVDVLHTFTRGSP.GRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVD.QLVKCSHERSIHLFIDSLLNEENPSKAYRCNS
Gplpl     SRLSPDDAQFVDVLHTFTRGSP.GRSIGIQKPVGHVDIYPNGGSFQPGCNIQDALRVISQKGFGDMD.QLVKCSHERSIHLFIDSLLNEENPSKAYRCNS
cknlpl    IRLSPDDADFVDVLHTYTRGSP.DRSIGIQKPVGHIDIYPNGGGFQPGCNLGEALRLIAEKGFSDVD.QLVKCSHERSIHLFIDSLLYEEKPSMAYRCNT
humhl     NRLSPDDANFVDAIHTFTREHM.GLSVGIKQPIGHYDFYPNGGSFQPGCHSLELYRHIAQHGFNAIT.QTIKCSHERSVHLFIDSLLHAGTQSMAYPCGD
rabhl     DRLSPDDANFVDAIHTFTREHM.GLSVGIKQPVGHYDFYPNGGSFQPGCHFLELYKHIAQHGLNALS.QTIKCAHERSVHLFIDSLLHPSMQSTAYQCSD
rathl     ERLSPDDANFVDAIHTFTREHM.GLSVGIKQPIAHYDFYPNGGSFQPGCHFLELYKHIAEHGLNAIT.QTIKCAHERSVHLFIDSLQHSNLQNTGFHCSN
humpl     VRLDPSDAKFVDVIHTDGAPIVPNLGFGMSQVVGHLDFFPNGGVEMPGCKKNILSQIVDIDGIWEGTRDFAACNHLRSYKYYTDSIVNPDGF.AGFPCAS
pigpl     VRLDPSDAKFVDVIHTDAAPIIPNLGFGMSQTVGHLDFFPNGGKQMPGCQKNILSQIVDIDGIWEGTRDFVACNHLRSYKYYADSILNPDGF.AGFPCDS
dogpl     VRLDPTDADFVDVIHTDAAPLIPFLGFGTSQQMGHLDFFPNGGEEMPGCKKNALSQIVNLDGIWEGTRDFVACNHLRSYKYYSESILNPDGF.ASYPCAS
droypl    TGLARGDAEFVDAIHTSVY......GMGTPIRSGDVDFYPNG...........................................................
droyp2    TGLARGDADFVDAIHTSAY......GMGTSQRLANVDFFPNG...........................................................
droyp3    GGLSRGDADFVDAIHTSTF......AMGTPIRCGDVDFYPNG...........................................................
Consensus --L+++DA+FVD++HT---------+-G-----++-D++PNGG---PGC+------+-++--G+-++-+----C+H-RS-+-+-+S+--------+++C++

          301                                                                                            400
humlpl    KEAFEKGLCLSCRKNRCNNLGYEINKVRAKR...SSKMYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQAFEISLYGTVAESENIPFTLP..EVSTNKT
bovlpl    KEAFEKGLCLSCRKNRCNNMGYEINKVRAKR...SSKMYLKTRSQMPYKVFHYQVKIHFSGTESNTYTNQAFEISLYGTVAESENIPFTLP..EVSTNKT
muslpl    KEAFEKGLCLSCRKNRCNNLGYEINKVRAKR...SSKMYLKTRSQMPYKVFHYQVKIHFSGTENGKQHNQAFEISLYGTVAESENIPFTLP..EVSTNKT
gplpl     KEAFEKGLCLSCRKNRCNNVGYEINKVRAKR...SSKMYLKTRSQMPYKVFHYQVKIYFSGTETTTYTNQAFEISLYGTVAESENIPFTLP..EVSANNT
cknlpl    KEAFEKGLCLSCRKNRCNNLGYKVNRVRTKR...NTKMYLKTRAQMPYKVFHYQVKIHFFGKTNVTKVDQPFLISLYGTLDESENIPFTLP..EVSSNKT
humhl     MNSFSQGLCLSCKKGRCNTLGYHVRQEPRSK...SKRLFLVTRAQSPFKVYHYQLKIQF.INQTETPIQTTFTMSLLGTKEKMQKIPITLGKGIA.SNKT
rabhl     MDSFSQGLCLGCTKGRCNTLGYHIRQEPLSK...GKRLFLVTQAQSPFRVYHYQFKIQF.INQIEKPLEPTFTMSLLGTKEEMQKIPITLGEGIT.SNKT
rathl     MDTFSQGLCLNCKKGRCNSLGYDIRRIGHAK...SKTLFLITRAQSPFKVYHYQFKIQF.INQMEKPIEPTFTMTLLGTKEEIKKIPITLGEGIT.SNKT
humpl     YNVFTANKCFPCPSGGCPQMGHYADRYPGKTNDVGQKFYLDTGDASNFARWRYKVSVTLSGKKVTGHIL....VSLFGNKGNSKQYEIFKGTLKPDS..T
pigpl     YNVFTANKCFPCPSEGCPQMGHYADRFPGKTNGVSQVFYLNTGDASNFARWRYKVSVTLSGKKVTGHIL....VSLFGNEGNSRQYEIYKGTLQPDN..T
dogpl     YRAFESNKCFPCPDQGCPQMGHYADKFAVKTSDETQKYFLNTGDSSNFARWRYGVSITLSGKRATGQAK....VALFGSKGNTHQFNIFKGILKPGS..T
droypl    ...................................................................................................
droyp2    ...................................................................................................
droyp3    ...................................................................................................
Consensus ---F+-+-C+-C-+--C--+G+---+--------+---+L-T-+---+--++Y-++--+--+-----------++L+G+-++--+-------------+++T

          401                                                                                            500
humlpl    YSFLIYTEVDIGELLMLKLKWKS..........DSYFSWSDWWSSPGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKSG..
bovlpl    YSFLLYTEVDIGELLMLKLKWIS..........DSYFSWSNWWSSPGFDIGKIRVKAGETQKKVIFCSREKMSYLQKGKSPVIFVKCHDKSLNRKSG..
muslpl    YSFLIYTEVDIGELLMMKLKWMS..........DSYFSWPDWWSSPSFVIERIRVKAGETQKKVIFCAREKVSHLQKGKDSAVFVKCHDKSLK.KSG..
gplpl     YSFLIYTEVDIGELLMLKLKWIT..........ESYFSWSSWWGRPTFTIEKIRVKAGETQKKIVFCSREKVSKLQKGKEAPVFVKCHDKSLNKKSG..
cknlpl    FSFLIYTEVDIGDLLMLKLQWEK..........DTFFSWSDWWTPFAFTIQRVRVKSGETQKKVVFCSRDGSSRLGKGEEAAIFVKCLEQPVSRKRGGA
humhl     YSFLITLDVDIGELIMIKFKWENSAVWANVWDTVQTIIPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSKRKIR
rabhl     YSFLITLNLDIGELMVIKFKWENSAVWANVWNTVQTIIPWGIKPRNSGLILKTIRVKAGETQQRMTFCSENMDDLQLHPTQEKNFVRCEVNPKKLKLKIK
rathl     YSLLITLDKDIGELIMLKFKWENSAVWANVWNTVQTIMLWDTEPHYAGLILKTIWVKAGETQQRMTFCPDNVDDLQLHPTQEKVFVKCDLKSKD.....
humpl     HSNEFDSDVDVGDLQMVKFIWYNNV.......INPTL.........PRVGASKIIVET.NVGKQFNFCSPETV....REEVLLTLTPC..........
pigpl     HSDEFDSDVEVGDLQKVKFIWYNNNV.......INPTL.........PRVGASKITVEA.NDGKVYDFCSQETV....REEVLLTLNPC..........
dogpl     HSNEFDAKLDVGTIEKVKFLWNNNNV.......NPTF.........PKVGAAKITVQKGEEKTVHSFCSESTV....REDVLLTLTEC..........
droypl    ...................................................................................................
droyp2    ...................................................................................................
droyp3    ...................................................................................................
Consensus -S-------++G++--+K+-W---------------------------------+-V+---------FC-V---------+----+++--C-----------

          501     513
humlpl    .............
bovlpl    .............
muslpl    .............
gplpl     .............
cknlpl    KKASKENSAHESA
humhl     .............
rabhl     .............
rathl     .............
humpl     .............
pigpl     .............
droypl    .............
droyp2    .............
droyp3    .............
```

Downloaded from www.jlr.org by guest, on June 18, 2012

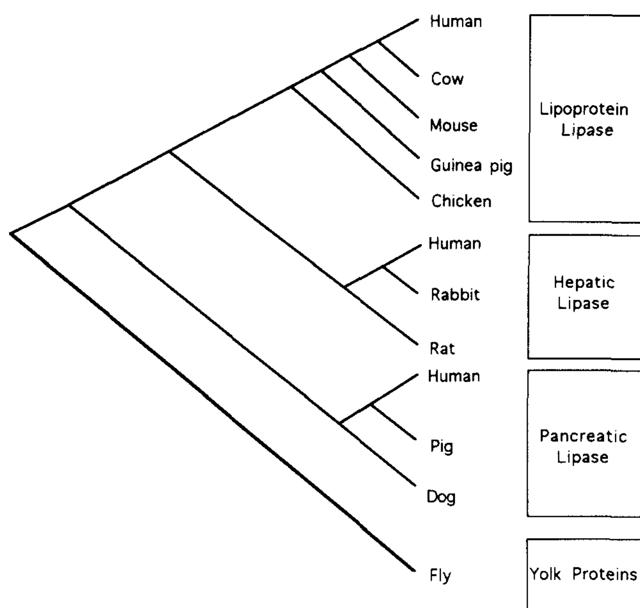

Fig. 2. Phylogenetic relationships of the lipase superfamily and *Drosophila* YP1. Maximum parsimony analysis was performed using PROTPARS from the phylogenetic computer analysis package PHYLIP (57). This method finds a tree that requires the minimum number of amino acid substitutions to explain the observed differences among sequences. Comparisons with *Drosophila* yolk proteins were made from alignment residue positions 132–242. In all cases gapped residues were not included in the comparison. The method of parsimony analysis is broadly used to resolve relationships at the sequence level between genes of different species. We have chosen this method in order to present a quantifiable relationship between the proteins making up the lipase superfamily. A maximum parsimony analysis of the aligned lipase sequences produced a consensus phylogeny shown in Fig. 2. In order to locate the root of the tree, we used the distant but phylogenetically useful *Drosophila* yolk proteins as an outgroup. Four trees of equal length (equal number of substitutions) resulted, each having the same position for the root and the same branching order for the three lipases. The trees differed only in the order of the branches within the LPLs. One of the trees was identical in branching order to a tree produced by analyzing the lipases alone.

tron loss, and another duplication to mammalian LPL and HL. After the initial duplication event a different path gave rise via intron loss to mammalian PL. Yet another path after the initial duplication event resulted in the gain of an exon, loss of introns, and gene duplication, leading to the three *Drosophila* yolk proteins. The theoretical basis for the model (13) was not presented.

Relationships within lipoprotein lipases

Lipoprotein lipase is a highly conserved protein and its phylogeny demonstrates several features that are of interest with respect to the evolution of mammals.

Guinea pig LPL appears to have diverged prior to mouse LPL (Fig. 2). Although this finding is not compatible with the traditional view that the guinea pig is a rodent, it is consistent with the recent analysis of the relationship of guinea pigs (a caviomorph) to the myomorph rodents (mice and rats) using available sequences from these species (15), which suggests that the guinea pig does not belong to the same order as the mouse and rat (myomorpha) but represents a separate evolutionary lineage (caviomorpha) that had diverged very early in the evolution of the eutherian mammals. This conclusion is consistent with the analysis of rates of nucleotide substitution among the LPLs of different species published by Semenkovich et al. (16). It was noted that the rate of nonsynonymous substitutions between human and murine is lower than between human and guinea pig. Also, the rate of nonsynonymous substitutions between bovine and guinea pig appeared higher than the rate of nonsynonymous substitutions between bovine and murine. Although it was determined that the rate of nonsynonymous substitutions is much higher in guinea pigs than in other rodent lineages, this observation fits well with the recently proposed phylogeny of the myomorpha and caviomorpha (15).

The eutherian radiation is commonly thought to have occurred about 80 million years ago. The bovine lineage separated from other lineages at about this time (17, 18). Molecular data suggests that rodents may have separated from other mammals considerably earlier (19). Parsimony analysis of the lipoprotein lipases shows that the caviomorph lineage and then the rodent lineage separated from the other lineages before the bovine lineage (Fig. 2). Thus the branching order of mammals represented in the lipoprotein lipase phylogeny is in agreement with the order proposed by Li et al. (19), in which rodents were one of the earliest groups to separate off from other mammals.

The distance between two proteins can be determined by aligning them and comparing the rates of substitution of amino acid residues at each site. Differences in amino acids can be quantified as a percent difference or can be expressed according to empirically corrected (20) or theoretically predicted (21) patterns of substitution. Empirical methods such as the Dayhoff correction (20) make an allowance for more than one substitution to have occurred at a particular site during the course of the evolution of the protein. When the average corrected Dayhoff distances (d) between an outgroup (chicken LPL) and the mammalian LPLs are compared (see Table 1), it appears that guinea pig LPL has a relatively higher rate of substitution than other mammals. Human, murine, and bovine LPLs have similar distances from chicken LPL (15.5%, 14.6%, and 16.0%, respectively) while in contrast, guinea pig LPL has the highest amount of sequence divergence (20.7%).

Downloaded from www.jlr.org by guest, on June 18, 2012

Relationships within hepatic lipases

Three mammalian HL sequences have been published: human, rabbit, and rat (14, 22–24). The rabbit has traditionally been regarded as an animal essentially devoid of HL activity. Interestingly, HL mRNA is present in the liver of the rabbit. The low HL activity in this species is apparently the result of a low level expression of HL mRNA in the liver of the rabbit (23). When compared to human LPL as an outgroup, human (d = 72.3%), rodent (d = 74.2%), and rabbit (d = 69.7%) hepatic lipase show similar rates of amino acid substitution (**Table 1**).

Hepatic lipase is much less conservative than LPL (13, 14). This is demonstrated by the divergence between human and mouse (or rat), which is 23.6% for HL but is only 2.2% for LPL (see also Fig. 3).

TABLE 1. Genetic distances between lipase sequences

Species	Proportion of Differences (%)	Dayhoff Distance (%)
	Lipoprotein Lipase	
Human vs. mouse	2.2	2.2
Human vs. cow	4.0	4.2
Human vs. guinea pig	8.8	9.4
Human vs. chicken	14.0	15.5
Mouse vs. cow	4.0	4.2
Mouse vs. guinea pig	8.5	9.0
Mouse vs. chicken	13.2	14.6
Cow vs. chicken	14.3	16.0
Cow vs. guinea pig	9.9	10.7
Guinea pig vs. chicken	18.0	20.7
	Hepatic Lipase	
Human vs. rat	20.2	23.6
Human vs. rabbit	15.4	17.3
Rabbit vs. rat	19.1	22.1
	Pancreatic Lipase	
Human vs. pig	9.2	9.8
Human vs. dog	26.8	33.2
Pig vs. dog	26.8	33.2
	Gene Families	
Human LPL vs. human HL	47.1	72.3
Human LPL vs. rat HL	47.8	74.2
Human LPL vs. rabbit HL	46.0	69.7
Human LPL vs. human PL	64.7	131.2
Human HL vs. human PL	64.3	129.5

Protein sequences were aligned and compared to determine the number of amino acid differences between them. The proportion of differences between two amino acid sequences is represented in the table as the number of differences per 100 residues. The distance matrix was constructed correcting for multiple substitutions at the same residue position using the method of Dayhoff (20). The Dayhoff method uses an empirical matrix made by determining the substitution patterns in 30 mammalian proteins. It is thus possible to estimate the most likely number of substitutions that have occurred at a particular amino acid residue position.

Relationships within pancreatic lipases

The phylogeny of the PLs in Fig. 2 is not in agreement with the commonly accepted phylogeny of the mammals, in which the Carnivora are usually grouped with the Artiodactyla (18). We used HL as an outgroup for the phylogenetic reconstruction of the PLs. As the HLs have a large number of amino acid substitutions when compared to the PLs, they do not represent an ideal outgroup. In our reconstruction, porcine and human PL share a root, separate from canine. Without a suitable outgroup, we are unable to obtain the correct order of branching within the PLs.

The rate of amino acid substitution in canine PL is distinctly faster than in either porcine or human PL (human PL vs. pig HL d = 9.8%; pig PL vs. dog PL d = 33.2%; dog PL vs. human PL d = 33.2%).

A previous analysis has suggested that PLs are structurally much less conserved and demonstrate a much higher rate of amino acid substitution than the other lipases (14). We note, however, that human and porcine (artiodactyl) PL have not diverged very fast (d = 9.8%) in comparison with human and bovine (artiodactyl) LPL (d = 4.2%) or when compared to the divergence between human and rat HL (d = 23.6%). The human PL sequence was not available for the previous analysis of Datta et al. (14), which was influenced by the unusually rapid rate of substitution in canine PL.

Between family distance comparisons

Human LPL and HL are less distant from each other (d = 72.3%) than either is from human PL (human LPL vs. human PL d = 131.2%; human HL vs. human PL d = 129.5%). This is consistent with the closer relationship between LPL and HL (Fig. 2). LPL has not always been a slowly evolving gene. As noted above, LPL has evolved much slower than HL within the higher vertebrates. However, both LPL and HL have a similar distance from PL (Table 1). This suggests that prior to the time at which LPL and HL evolved within the higher vertebrates, LPL evolved at a faster rate than HL.

Exon/intron structure of the lipases

The exon/intron structure of human LPL and HL and canine PL has recently been analyzed in detail (8, 13, 25–27). We compared the exon/intron boundaries against the conservation profile as shown in **Fig. 3** (see below). Exon/intron boundary positions are not consistently coincident with peaks or valleys of conservation. Human LPL has 10 exons, human HL has 9 exons, and dog PL has 13 exons. Exon sizes and boundary positions are very similar in HL and LPL, although LPL has an extra exon comprised exclusively

ASBMB JOURNAL OF LIPID RESEARCH JLR

Downloaded from www.jlr.org by guest, on June 18, 2012

ASBMB
JOURNAL OF LIPID RESEARCH

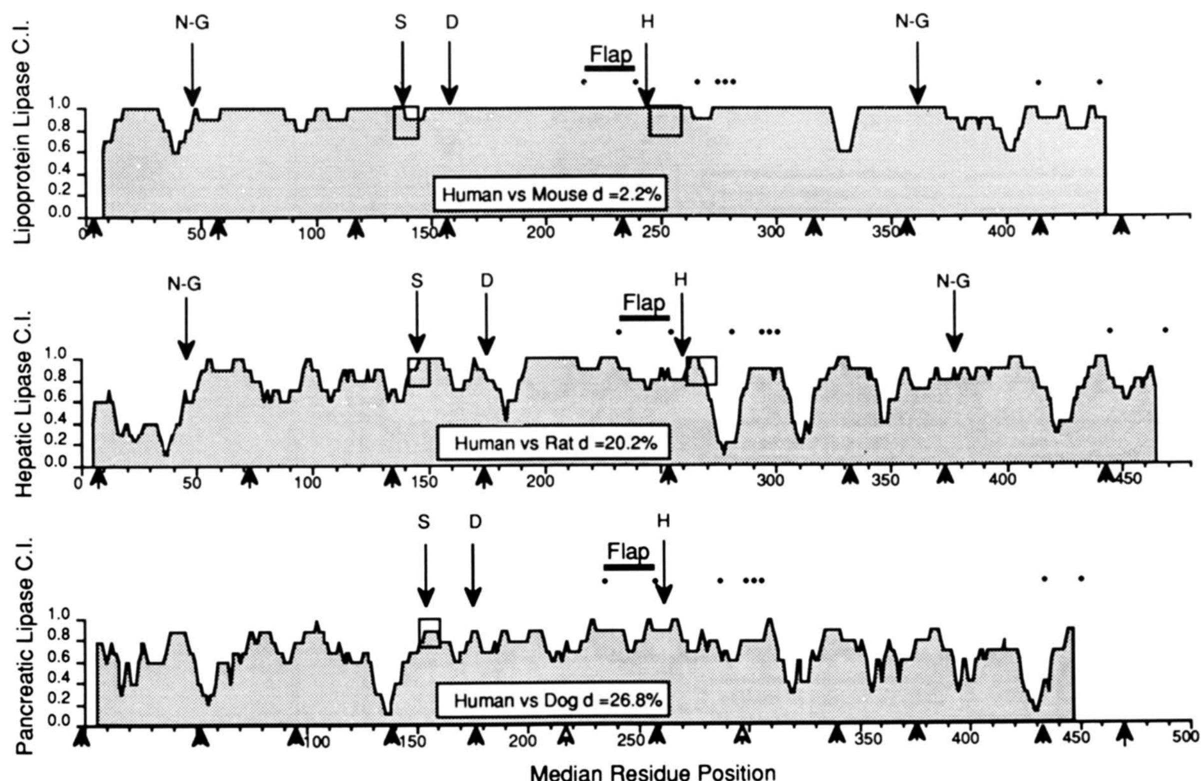

Fig. 3. Conserved structural features of lipase sequences. Pairs of proteins from within each family were aligned and given a consensus score for conserved residues at each site. A score of 1 was given when residues were the same at a particular position. A score of 0 was given when residues differed at the same position. A window size of 9 residues was found to give the most useful resolution for comparison of conservation. The average identity (conservation index) for each 9-residue window was determined, and the index value was plotted against the median residue position. The conservation indexes (CI) for two species of LPL (human and mouse), HL (human and rat), and PL (human and dog) are shown plotted against median residue position. N-linked glycosylation sites (N-G) that are conserved in the two species compared are marked with arrows; sites that are unique to only one of the two species are not shown. The locations of the three catalytic residues serine (S), aspartic acid (D), and histidine (H) are marked with arrows. Conserved cysteine residues are represented by solid circles above each plot. The α-helical "flap" is represented by a short bar. An outline box encloses the putative lipid-binding domain. Percent differences between each protein pair are shown as a value of d in each figure. Solid arrows below each plot represent the positions of exon/intron boundaries. Open arrows represent exon/intron boundaries in the PL gene that are absent in HL and LPL.

of the 3′ untranslated region of the gene. The similarity in exon/intron boundary distribution between LPL and HL suggests they have diverged more recently than PL (13, 25–27).

PL has a distinctly different organization of exon/intron boundaries with respect to the other lipases. PL has extra introns, "splitting" the exon organization. For example, exons 4 and 5 of PL are analogous to exon 3 of HL and LPL and exons 7 and 8 of PL are analogous to exon 5 of HL and LPL.

HL and LPL may have lost several introns after divergence from a common ancestor of the lipases. As it is likely that PL has greater similarity to an ancestral gene, it is therefore likely that the ancestral lipase from which the lipase gene family arose contained the same number of, or more, introns than modern PL (13).

Drosophila yolk proteins

By screening the Protein Identification Resource (PIR) and National Biomedical Research Foundation (NBRF) data banks, defined regions of *Drosophila* yolk proteins have been shown by many workers to have sequence similarity with the vertebrate lipases (9–11). Bownes et al. (12) have also independently detected a similarity between the yolk proteins and porcine PL. Persson et al. (10) determined that the similarity extends even to large parts of the hydrophobicity profiles. The three yolk protein genes are a distinct group of proteins that occur only in the higher diptera (28) and have 60% protein sequence similarity to one another. The best conserved regions in *Drosophila* yolk proteins are those which show similarity to the vertebrate lipases (10).

The region in the lipases that shows similarity to YP structure is also a well-conserved region containing the active site and proposed to be involved in lipid binding (10, 29, 30). *Drosophila* yolk proteins are generally termed vitellogenins and are analogous in function to the vitellogenins of other species. Vitellogenins are produced in the insect fat body as a food source, and are transported in the hemolymph to be taken up by the developing oocyte. The process involves receptor-

Downloaded from www.jlr.org by guest, on June 18, 2012

mediated endocytosis (31). Although the YP proteins show significant sequence similarity to the lipases, the catalytic Ser-Asp-His triad present in all lipases (8, 32) is not conserved in YP. The catalytic serine residue in the triad is replaced by an asparagine in YP1 and a glycine in YP2, which suggests that YP is not catalytically active (9). Catalytic activity has not been demonstrated for the YP protein family (11). Interestingly, the histidine in the triad is still conserved. A serine or another amino acid from a different part of the primary structure could possibly act as a catalytic residue if properly positioned, or different conditions may be necessary for catalytic activity (9). It is highly unlikely, however, that YP proteins show the same catalytic activity as the lipases. Yolk proteins have been demonstrated to bind fatty acid ecdysteroid conjugates (12). The tightly bound ecdysteroids are released as yolk proteins are proteolytically degraded. Embryogenesis could thus be influenced by timed release of bound ecdysteroids. Fatty acid ecdysteroid binding in *Drosophila* and lipid binding in vertebrates are functions that may be reflected in the sequence similarity of the conserved region of the lipase gene superfamily.

CONSERVATION AND STRUCTURAL SIMILARITY

Conservation and structure

Regions of a protein that serve an essential function are usually evolutionarily better conserved in structure than regions that have a less important function. For example, the catalytic triad residues, Ser-Asp-His, are essential to the catalytic activities of LPL, HL, and PL. Therefore, all three residues are strictly conserved in all species (Fig. 1). However, inspection of Fig. 1 reveals that many other residues in addition to the catalytic triad are completely conserved in all the available sequences. Furthermore, there are clusters of these residues in each of the lipases, suggesting that some regions of each protein are conserved as structural domains. An objective method for inferring the relative conservation of structural domains across the linear sequence within each protein is to calculate the relative proportion of conserved residues in a window that scans the whole sequence from the N-terminus to the C-terminus. When a moving window of nine amino acid residues is used for analysis of each lipase (designated a conservation index, or CI), the variation in structural conservation spanning the entire lipase protein becomes apparent (Fig. 3 and **Fig. 4**). A number of previously identified putative functional residues and domains can be re-examined in the light of the degree of conservation of local protein structure. In such comparisons the genetic distance between each protein pair is important. Since the three lipases have not been sequenced from the identical species, the CI analyses of LPL and HL have been performed on pairs of species (human/mouse for LPL, human/rat for HL) that have identical distances, and on PL in a pair (human/dog) that has a distance more similar to the human/rodent pair than the other available pairs (human/pig or pig/dog). It is clear from inspection of Fig. 3 that LPL is the most highly conserved of the three lipases whereas HL and PL are roughly similar in their degree of conservation. Therefore, the CI index analysis of sequence domain conservation is in agreement with that revealed in Table 1.

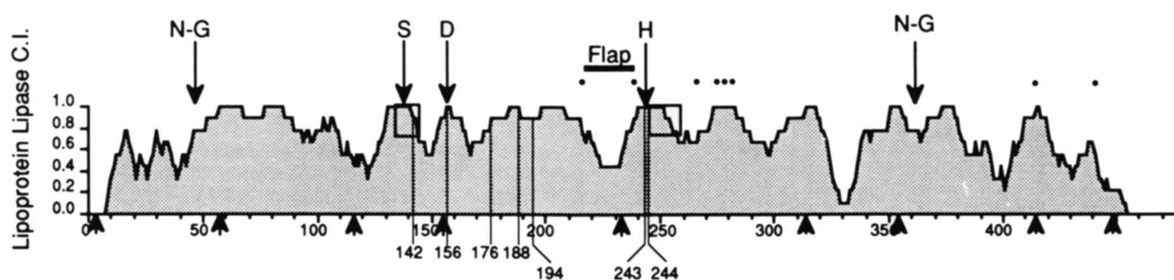

Fig. 4. Conserved structural features of lipoprotein lipases. Protein sequences of the LPLs were aligned and given a consensus score for conserved residues at each site. A score of 1 was given when all residues were the same at a particular position. A score of 0 was given when any differences in residues occurred at the same position. A window size of 9 residues was used. The average identity (conservation index) for each 9-residue window was determined, and the index value was plotted against the median residue position. The conservation indexes (CI) for complete, mature proteins of five species of LPL (human, mouse, cow, guinea pig, and chicken) are shown plotted against median residue position. N-linked glycosylation sites (N-G) that are conserved in all five species are marked with arrows; sites that are not present in all five species are not shown. The locations of the three putative catalytic residues serine (S), aspartic acid (D), and histidine (H) are marked with arrows. Conserved cysteine residues are represented by solid circles above the plot. The α-helical "flap" is represented by a short bar. Outline boxes enclose the putative lipid-binding domains. Solid arrows below each plot indicate the position of exon/intron boundaries. Vertical bars represent the location of residues at which natural amino acid substitution mutations have resulted in impaired function of LPL in patients with familial LPL deficiency. Amino acid substitutions are: 142, Gly→Glu (58); 156, Asp→Gly (59); 176, Ala→Thr (60); 188, Gly→Glu (61); 194, Ile→Thr (62); 207, Pro→Leu (63); 243, His→Arg (62); 244, Ser→Thr (64).

ASBMB JOURNAL OF LIPID RESEARCH JLR

Downloaded from www.jlr.org by guest, on June 18, 2012

Catalytic residues and lipid binding regions

The most highly conserved feature in all lipases is a 9-amino acid segment containing some hydrophobic side-chains ranging in hydropathy indices from 0.7 to 14.3 (8) with a consensus sequence Gly-Xaa-Ser-Xaa-Gly (Fig. 1), which agrees with analogous sequence surrounding the active site serines in serine esterases and serine proteases (33). This segment is very similar in vertebrate lingual/gastric lipase, prokaryotic lipases, and vertebrate lecithin:cholesterol acyltransferase (8, 10). This 9-residue sequence has been described by some investigators as part of a lipid-binding segment (29, 30), although this is not agreed upon by other workers. Mickel et al. (8) determined that the residues in this region are hydrophobic and that this may facilitate the action of hydrolysis at the aqueous lipid interface. Winkler, D'Arcy, and Hunziker (32) have shown that the hydrophobicity can be explained by packing of these residues with other hydrophobic side-chains and point out that they are not likely to take part in lipid-binding in PL. In fact, another putative lipid-binding region in rat and human HL has been predicted on the basis of similarity to the conserved hydrophobic region (24). We have annotated both regions in our alignment (see Fig. 1) and note that they are highly conserved (Figs. 3 and 4) and are similar in terms of hydrophobicity profile (data not shown). PL does not show a second putative lipid-binding region. We have found that the predicted probability of the second putative lipid-binding domain being present on the surface of the folded structures of LPL and HL is low (34).

X-ray crystallography of human PL has shown that the catalytic serine residue of PL is located in the N-terminal domain of the edge of a doubly wound parallel β-sheet, and is part of a Ser-Asp-His triad (32). The crystalline structure of human PL suggests that a substantial conformational change occurs before it can bind substrate in this postulated active site. There is a surface loop between disulfide-bridged residues (237 to 261) that covers the active site with a short one-turn α-helix. This "flap" has to be repositioned before the site can become accessible to substrate. Each amino acid of the three-part catalytic triad lies in a segment showing a high degree of conservation (CI > 0.8 for PL and CI = 1.0 for HL and LPL).

The PL crystal structure is likely to be analogous to that of LPL and HL in the region in which they share high amino acid sequence similarity (consensus alignment residues 108–320). We have performed surface prediction analyses that show that PL and LPL have marked similarities in predicted structural surface features in this region (data not shown). A recent study of the in vitro expression and activity of substitutions of highly conserved serine residues in LPL (35) demonstrated that single amino acid substitution mutants involving conserved Ser residues likely to be on the surface of the LPL molecule show the least impairment in LPL enzyme activity in comparison with the native enzyme. In contrast, the mutant LPL is totally inactive when the putative catalytic Ser^{132} residue in LPL is substituted by Ala, Thr, or Asp (35). Similarly, replacement of the analogous Ser^{147} with glycine in rat HL abolished its catalytic activity, suggesting that this residue may also be an active-site serine (36). Furthermore, a detailed comparison of the enzymatic activities of a large number of site-specific mutants of LPL produced in vitro indicates that LPL has a three-dimensional structure very similar to that of PL (35).

It is evident from Figs. 1, 3, and 4 that the amino acid sequence of the PL flap region is not highly conserved among the different lipases. There are, however, consistently aligned cysteine residues in all the lipases bounding the flap region. Reference to Figs. 3 and 4 indicates that the flap region has varying degrees of conservation within each of the lipases. In this region then, it is possible that the functional site of lipases is covered in each protein by a loop of amino acid residues serving only as structural components of a short α-helical flap which may be quite flexible in primary structure. Predicted α-helical moment data for the region in each of the lipases shows that despite the highly dissimilar amino acid sequences, it is probable that an α-helix is present in the "flap" region in all of the lipases (data not shown).

Brozozowski et al. (37) have modeled interfacial activation, using the X-ray crystal structure of the fungal *Rhizomucor miehei* triacylglycerol lipase. They have demonstrated that a helical flap structure is displaced during the conformational change associated with lipid binding and catalytic activity. The fungal lipase has the structurally analogous Ser-Asp-His catalytic triad present in human PL (32). A hydrophobic side of the flap becomes exposed, thus expanding the amount of non-polar surface around the active site. This effectively represents a "seal" against solvent (37, 38) allowing access of the lipid interface to the active site. Brozozowski et al. (37) have postulated that a similar mechanism may exist in human PL, but point out that other structural modifications are probably needed for substrate access in this enzyme.

Peptides in the postulated helical flap region in human LPL and human HL show similar Eisenberg hydrophobic moment profiles. In contrast, PL shows a substantially different hydrophobic moment profile in its helical flap region (data not shown).

In PL there are two sets of four hydrophobic residues that flank the α-helical flap region (32).

Downloaded from www.jlr.org by guest, on June 18, 2012

These residues are annotated on our alignment. We note that the first flanking region is similar in hydrophobicity profile with aligned amino acid residues in LPL and HL (see Fig. 1 for location). However, the similarity between PL, and HL and LPL does not extend to the second set of hydrophobic PL flanking residues.

N-linked glycosylation sites

Two putative N-linked glycosylation sites are conserved in the mammalian lipoprotein and hepatic lipases (the second site differs in position in chicken LPL). Both show significant surface probability values in human LPL (data not shown). The regions in which they lie are not as highly conserved as the lipid-binding domains but show conservation at or above 0.8 (Fig. 3). The positions are conserved in HL and LPL for both sites. There is a single N-linked glycosylation consensus sequence in human and pig PL and also one at a different location in dog PL (Fig. 1). These potential sites, which also appear to be on the surface of the enzyme (32), have no homologs at similar locations in LPL or HL. It is interesting that human (39) and probably mouse (40) and rat (41) LPL require proper N-linked glycosylation for activity.

Distribution of cysteine residues

There are eight cysteine residues that can be aligned to the same place in all members of the lipase family (Fig. 1). The conserved position of these residues probably reflects their role in the formation of disulfide bridges required for maintenance of enzyme structure and function. This is supported by the known disulfide-bridging demonstrated by porcine PL (30) and bovine LPL (42). Six of the eight residues are found in the central very highly conserved region of the family. One of the residues, at median residue position 268 in LPL, lies in a region that has a CI $\leq$ 0.8 for the equivalent position in LPL, HL, and PL. HL has a more varied CI profile than LPL, and lower CI values at its conserved cysteine residues. This reflects the greater structural variation within the HLs, and the greater genetic distance of the two proteins being compared. The final cysteine residue at median residue position 470 in HL is very near the end of the protein, has a low CI, but may have a structural role in all the lipases. It apparently has a bridging function in PL. Because the window of comparison between proteins is nine residues in size, it is not possible to perform adequate comparisons of conservation when within nine residues of the end of proteins being compared.

Structural conservation in LPL

As the LPLs are highly conserved, we found it informative to perform an additional CI analysis on the data from all the available LPL sequences (Fig. 4). There is a marked degree of conservation at regions sharing known functions. CI values of 1.0 are demonstrated at the putative lipid-binding site and catalytic triad residues Ser^{132}, Asp^{156}, and His^{243} (values refer to the corresponding position in human LPL). Possibly structurally important conserved cysteine residues also demonstrate interesting CI values. Of the eight cysteines conserved throughout the lipase superfamily, the first (CI = 0.9) and second (CI = 1.0) are bridging residues at the conserved boundaries of the putative flap region involved in conformational change of LPL induced by its substrate (Fig. 1) analogous to the situation in human PL (32). The third and fourth cysteines of LPL correspond to bridging residues in PL yet the third cysteine lies in a relatively poorly conserved region (CI = 0.7). The fourth, fifth, and sixth conserved cysteines all lie in a region of absolute conservation (CI = 1.0). The seventh (CI = 0.9) and eighth (CI = 0.6) cysteine residues lie at or very near peaks of CI values, surrounded by poorly conserved regions. This raises an interesting possibility that functional disulfide bonding with cysteine residues may require the conservation of secondary structural characters in the surrounding residues on either side of the cysteine. This provides a possible explanation as to why the cysteine residues in LPL appear to be associated with groups of conserved amino acids.

N-linked glycosylation sites of LPL, common to all known species, are placed within conserved regions. The first and second N-linked glycosylation sites have CI values of 0.8. In vitro expression of site-specific mutants of human LPL indicates that the first but not the second N-linked glycosylation site is required for LPL activity (35, 39). Chicken LPL has an additional predicted C-terminal N-linked glycosylation site at consensus residues 360–362 (Fig. 1) (43).

Mutant variants of human LPL

A number of single amino acid changes have been reported in human LPL mutants who present with Type I hyperlipoproteinemia (Fig. 4). Substitutions affecting in vivo function of LPL are located at residues that have a CI of at least 0.8. Clearly, substitutions in highly conserved regions of the protein can cause loss of function. The substitutions that have been characterized to date are localized in exons 4, 5, and 6 of human LPL. This gives further evidence that the crucial residues involved in the catalytic function of LPL may be contained within exons 4, 5, and 6.

Domains and exons

Exon shuffling has been proposed as a mechanism for the evolution of multidomain proteins. Separate

ASBMB
JOURNAL OF LIPID RESEARCH
Downloaded from www.jlr.org by guest, on June 18, 2012

domains are confined to separate exons, and combination (or shuffling) of exons results in the emergence of discrete multidomain proteins. Loss of introns results in exons "fusing" to form structures that contain more than one domain. This process has been inferred for the evolution of low density lipoprotein receptor (44) and serine proteases (45).

Exon shuffling has been put forward as one mechanism in the evolution of the lipases (13) and we discuss here pertinent points of that hypothesis. For example, the putative lipid-binding domain common to members of the lipase gene family is contained within a distinct exon. The equivalent lipid-binding domain located in exon 4 of HL and LPL and exon 6 of PL is bounded on either side by an exon/intron boundary. A major domain border, determined by sensitivity to proteolytic cleavage, has been suggested around residue 228 in human LPL (46). An exon/intron boundary lies just downstream of this residue in the sequence.

Pancreatic lipase has a marked degree of similarity in CI profile with HL between median residues 140–350 (Fig. 3). In particular, the putative lipid binding domain is near a peak of high CI value, and cysteine residues and the residues of the catalytic triad are at conserved peaks. The conserved region is represented by exons 4, 5, and 6 in LPL and HL. The same region is represented by exons 6, 7, 8, 9, and 10 in PL. Thus it appears that HL and LPL have lost two introns from the conserved region, resulting in the formation of multidomain exons after divergence from a common ancestral gene to the PL.

Drosophila vitellogenins do not have the same intron structure as the vertebrate lipases. In particular, the conserved putative lipid-binding domain present in the vitellogenins does not have a comparable intron distribution. This is in agreement with the considerable divergence between the vitellogenins and a common ancestor to the lipases. However, it is likely that the exon containing the putative lipid-binding domain was present in a common ancestor. Exon shuffling could thus have been a mechanism for the accumulation of other exons, resulting in proteins as different as the vitellogenins and the lipases.

Catalytic function

The catalytic function of the lipases may have arisen after the lipid-binding function. The region of similarity with the *Drosophila* vitellogenins contains a putative lipid binding region (Fig. 1). Although highly similar to lipase sequences, this region in the vitellogenins misses the conserved "catalytic" serine residue, which is replaced by an asparagine in YP1 and a glycine in YP2, suggesting that YP is not catalytically active. It has been postulated that the serine residue in porcine PL that is equivalent to human PL Ser^{152} may not be an active-site residue (29). However, that is not consistent with results of site-directed mutagenesis experiments on human LPL (35), rat HL (36), mutational screening, and structural evidence from human PL (32).

If the YP proteins contain domains that share ancestry with the lipases, and if exon shuffling was one of the processes which shaped the structure of the modern lipases, then the catalytic function of lipases has probably arisen after the lipid-binding function. The putative lipid-binding domains are contained within discrete exons, but the three-part putative catalytic site is made up of residues in discrete exons (which are contiguous in HL and LPL, but noncontiguous in PL [Fig. 3]). Thus, the catalytic function could not have been directly derived from an exon-shuffling event.

We wish to thank Dr. Lori Sadler for help in design of the conservation index. This work was supported by National Institutes of Health grants HL-16512 (to L. C.) and GM-39927 (to W-H. L.) and a grant from the Juvenile Diabetes Foundation International (to L. C.).

Manuscript received 31 January 1991 and in revised form 31 October 1991.

REFERENCES

1. Goldberg, I. J., N-A. Le, H. N. Ginsberg, R. M. Krauss, and F. T. Lindgren. 1988. Lipoprotein metabolism during acute inhibition of lipoprotein lipase in the cynomolgous monkey. *J. Clin. Invest.* **81**: 561–568.
2. Garfinkel, A. S., and M. C. Schotz. 1987. Lipoprotein lipase. *In* Plasma Lipoproteins. A. M. Gotto, Jr., editor. Elsevier, Holland. 335–357.
3. Olivecrona, T., and G. Bengtsson-Olivecrona. 1987. Lipoprotein lipase from milk—the model enzyme in lipoprotein lipase research. *In* Lipoprotein Lipase. J. Borensztajn, editor. Evener Press, Chicago. 15–58.
4. Goldberg, I. J., N. Le, J. R. Paterniti, H. N. Ginsberg, F. T. Lindgren, and W. V. Brown. 1982. Lipoprotein metabolism during acute inhibition of hepatic triglyceride lipase in the cynomolgous monkey. *J. Clin. Invest.* **70**: 1184–1192.
5. Kuusi, T., E. A. Nikkila. 1979. Hepatic endothelial lipase antiserum influences rat plasma low and high density lipoproteins in vivo. *FEBS Lett.* **104**: 384–388.
6. Jansen, H., A. van Tol, and W. C. Hulsmann. 1980. On the metabolic function of heparin-releasable liver lipase. *Biochem. Biophys. Res. Commun.* **92**: 53–59.
7. Bamberger, M., S. Lund-Katz, M. C. Phillips, and G. H. Rothblat. 1985. Mechanism of the hepatic lipase-induced accumulation of high-density lipoprotein cholesterol by cells in culture. *Biochemistry.* **24**: 3693–3701.
8. Mickel, S. F., F. Weidenbach, B. Swarovsky, K. S. LaForge, and G. A. Scheele. 1989. Structure of the canine pancreatic lipase gene. *J. Biol. Chem.* **264**: 12895–12901.

Downloaded from www.jlr.org by guest, on June 18, 2012

9. Baker, M. 1988. Is vitellogenin an ancestor of apolipoprotein B-100 of human LDL and human lipoprotein lipase? *Biochem. J.* **255**: 1057–1060.
10. Persson, B., G. Bengtsson-Olivecrona, S. Enerbäck, T. Olivecrona, and H. Jörnvall. 1989. Structural features of lipoprotein lipase. *Eur. J. Biochem.* **179**: 39–45.
11. Terpstra, P., and A. B. Geert. 1988. Homology of *Drosophila* yolk proteins and the triacylglycerol lipase family. *J. Mol. Biol.* **22**: 663–665.
12. Bownes, M., A. Shirras, M. Blair, J. Collins, and A. Coulson. 1988. Evidence that insect embryogenesis is regulated by ecdysteroids released from yolk proteins. *Proc. Natl. Acad. Sci. USA.* **85**: 1554–1557.
13. Kirchgessner, T., J-C. Chuat, C. Heinzmann, J. Etienne, S. Guilhot, K. Svenson, A. Ameis, C. Pilon, L. Auriol, A. Andalibi, M. Schotz, F. Galibert, and A. Lusis. 1989. Organization of the human lipoprotein lipase gene and evolution of the lipase gene family. *Proc. Natl. Acad. Sci. USA.* **86**: 9647–9651.
14. Datta, S., C-C. Luo, W-H. Li, P. van Tuinen, D. H. Ledbetter, M. A. Brown, S-H. Chen, W-H. Liu, and L. Chan. 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.
15. Graur, D., W. A. Hide, W-H. Li. 1991. Is the guinea pig a rodent? *Nature (London).* **351**: 649–652.
16. Semenkovich, C. F., S-H. Chen, M. Wims, C-C. Luo, W-H. Li, and L. Chan. 1989. Lipoprotein lipase and hepatic lipase mRNA tissue specific expression, developmental regulation, and evolution. *J. Lipid Res.* **30**: 423–432.
17. Romer, A. S. 1966. Vertebrate Paleontology. University of Chicago Press, Chicago.
18. Szalay, F. S. 1977. Phylogenetic relationships and a classification of the eutherian mammalia. *In* Major Patterns in Vertebrate Evolution. M. Hecht. P. C. Goody, and B. M. Hecht, editors. Plenum Press, New York. 315–374.
19. Li, W-H, M. Gouy, P. Sharp, C. O'hUigin, and Y. Yang. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla and Carnivora and molecular clocks. *Proc. Natl. Acad. Sci. USA.* **87**: 6703–6707.
20. Dayhoff, M. O. 1978. Atlas of Protein Sequence and Structure (Vol 5). National Biomedical Research Foundation, Silver Spring, MD.
21. Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, Great Britain.
22. Martin, G. A., S. J. Busch, G. D. Meredith, A. D. Cardin, D. T. Blankenship, S. J. T. Mao, A. E. Richtin, C. W. Woods, M. M. Racke, M. P. Schafer, M. C. Fitzgerald, D. M. Burke, M. A. Flanagan, and R. L. Jackson. 1988. Isolation and cDNA sequence of human postheparin plasma hepatic triglyceride lipase. *J. Biol Chem.* **263**: 10907–10914.
23. Warren, J. R., D. Ebert, A. Mitchell, and P. Barter. 1991. Rabbit hepatic lipase cDNA sequence: low activity is associated with low messenger RNA levels. *J. Lipid Res.* **32**: 1333–1339.
24. Komaromy, M. C., and M. C. Schotz. 1987. Cloning of rat hepatic lipase cDNA: evidence for a lipase gene family. *Proc. Natl. Acad. Sci. USA.* **84**: 1526–1530.
25. Deeb, S., and R. Peng. 1989. Structure of the human lipoprotein lipase gene. *Biochemistry.* **28**: 4131–4135.
26. Cai, S. J., D. M. Wong, S-H. Chen, and L. Chan. 1989. Structure of the human hepatic triglyceride lipase gene. *Biochemistry.* **28**: 8966–8971.
27. Ameis, D., G. Stahnke, J. Kobayashi, J. McLean, G. Lee, M Büscher, M. Schotz, and H. Will. 1990. Isolation and characterization of the human hepatic lipase gene. *J. Biol. Chem.* **265**: 6652–6555.
28. Harnish, D. G., and B. N. White. 1982. Insect vitellins, identification, purification and characterization from 8 orders. *J. Exp. Zool.* **1**: 405–413.
29. Guidoni, A., F. Benkouka, J. De Caro, and M. Rovery. 1981. Characterization of the serine reacting with diethyl *p*-nitrophenol phosphate in porcine pancreatic lipase. *Biochim. Biophys. Acta.* **660**: 148–150.
30. Verger, R. 1984. Pancreatic Lipases. *In* Lipases. B. Borgström, and H. L. Brockman, editors. Elsevier, Amsterdam. 83–150.
31. Bownes, M. 1986. Expression of the genes coding for vitellogenin (Y protein). *Annu. Rev. Entomol.* **31**: 507–531.
32. Winkler, F. K., A. D'Arcy, and W. Hunziker. 1990. Structure of human pancreatic lipase. *Nature (London).* **343**: 771-774.
33. Brenner, S. 1988. The molecular evolution of genes and proteins: a tale of two serines. *Nature (London).* **334**: 528–530.
34. Janin, J., S. Wodak, M. Levitt, and B. Maigrett. 1978. Conformation of amino-acid side chains in proteins. *J. Mol. Biol.* **125**: 357–386.
35. Faustinella, F., L. Smith, C. Semenkovich, and L. Chan. 1991. Structural and functional roles of highly conserved serines in human lipoprotein lipase: evidence that serine 132 is essential for enzyme catalysis. *J. Biol Chem.* **266**: 9481–9485.
36. David, R. C., G. Stahnke, H. Wong, M. H. Doolittle, D. Ameis, H. Will, and M. C. Schotz. 1990. Hepatic lipase: site-directed mutagenesis of a serine residue important for catalytic activity. *J. Biol. Chem.* **265**: 6291–6295.
37. Brozozowski, A. M., U. Derewenda, Z. S. Derewenda, G. G. Dodson, D. M. Lawson, J. P. Turkenburg, F. Bjorkling, B. Huge-Jensen, S. A. Patkar, and L. Thim. 1991. A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature (London).* **351**: 491-494.
38. Blow, D. 1991. Lipases reach the surface. *Nature (London).* **351**: 444–445.
39. Semenkovich, C. F., C. Luo, M. K. Nakanishi, S-H. Chen, L. C. Smith, and L. Chan. 1990. In vitro expression and site-specific mutagenesis of the cloned human lipoprotein lipase gene: potential N-linked glycosylation site asparagine 43 is important for both enzyme activity and secretion. *J. Biol. Chem.* **265**: 5429–5433.
40. Olivecrona, T., S. S. Chernick, G. Bengtsson-Olivecrona, M. Garrison, and R. O. Scow. 1987. Synthesis and secretion of lipoprotein lipase in 3T3-L1 adipocytes: demonstration of inactive forms of lipase in cells. *J. Biol. Chem.* **262:** 10748–10759.
41. Chajek-Shaul, T., G. Friedman, H. Knobler, O. Stein, J. Etienne, and Y. Stein. 1985. Importance of the different steps of glycosylation for the activity and secretion of lipoprotein lipase in rat preadipocytes studied with monensin and tunicamycin. *Biochim. Biophys. Act.* **837**: 123–134.
42. Yang, C-Y., Z-W. Gu, H-X. Yang, M. F. Rohde, A. M. Gotto, Jr., and H. J. Pownall. 1989. Structure of bovine

Downloaded from www.jlr.org by guest, on June 18, 2012

ASBMB JOURNAL OF LIPID RESEARCH

milk lipoprotein lipase. *J. Biol. Chem.* **264**: 16822–16827.
43. Cooper, D., J. Stein, P. Strieleman, and A. Bensadoun. 1989. Avian adipose lipoprotein lipase: cDNA sequence and reciprocal regulation of mRNA levels in adipose and heart. *Biochim. Biophys. Acta.* **1008**: 92–101.
44. Südhof, T. C., D. W. Russel, J. L. Goldstein, M. S. Brown, R. Sanchez-Pescador, and G. Bell. 1985. Cassette of 8 exons shared by genes for LDL receptor and EGF precursor. *Science.* **228**: 893–895.
45. Rogers, J. 1985. Exon shuffling and intron insertion in serine protease genes. *Nature (London).* **315**: 458–459.
46. Bengtsson-Olivecrona, G., T. Olivecrona, and H. Jörnvall. 1986. Lipoprotein lipase from cow, guinea pig and man. *Eur. J. Biochem.* **161**: 281–288.
47. Wion, K. L., T. G. Kirchgessner, A. J. Lusis, M. C. Schotz, and R. M. Lawn. 1987. Human lipoprotein lipase complementary DNA sequence. *Science.* **235**: 1638–1641.
48. Senda, M., K. Oka, W. V. Brown, P. K. Qasba, and Y. Furuichi. 1987. Molecular cloning and sequence of a cDNA coding for bovine lipoprotein lipase. *Proc. Natl. Acad. Sci. USA.* **84**: 4369–4373.
49. Kirchgessner, T. G., K. L. Svenson, A. J. Lusis, and M. C. Schotz. 1987. The sequence of cDNA encoding mouse lipoprotein lipase. A member of a lipase gene family. *J. Biol. Chem.* **262**: 8463–8466.
50. Enerback, S., H. Semb, G. Bengtsson-Olivecrona, P. Carlsson, M. L. Hermansson, T. Olivecrona, and G. Bjursell. 1987. Molecular cloning and sequence analysis of cDNA encoding lipoprotein lipase of guinea pig. *Gene.* **262**: 8463–8466.
51. Lowe, M. E., J. L. Rosenblum, and A. W. Strauss. 1989. Cloning and characterization of human pancreatic lipase cDNA. *J. Biol. Chem.* **264**: 20042–20048.
52. De Caro, J., M. Boudouard, J. Bonicel, A. Guidoni, P. Desnuelle, and M. Rovery. 1981. Porcine pancreatic lipase: completion of the primary structure. *Biochim. Biophys. Acta.* **671**: 129–138.
53. Hung, M-C., and P. C. Wensink. 1981. The sequence of the *Drosophila melanogaster* gene for yolk protein 1. *Nucleic Acids Res.* **9**: 6407–6419.
54. Hovemann, B., and R. Galler. 1982. Vitellogenin in *Drosophila melanogaster*: a comparison of the YPI and YPII genes and their transcription products. *Nucleic Acids Res.* **10**: 2261–2274.
55. Garabedian, M. J., A. D. Shirras, M. Bownes, and P. C. Wensink. 1987. Nucleotide sequence of the gene coding for *Drosophila melanogaster* yolk protein 3. *Gene.* **55**: 1–8.
56. Higgins, D. G., and P. M. Sharp. 1989. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene.* **73**: 237–244.
57. Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. *Q. Rev. Biol.* **57**: 379–404.
58. Ameis, D., J. Kobayishi, R. C. Davis, O. Ben-Zeev, M. J. Malloy, J. P. Kane, G. Lee, H. Wong, R. J. Havel, and M. C. Schotz. 1991. Familial chylomicronemia (Type I hyperlipoproteinemia) due to a single missense mutation in the lipoprotein lipase gene. *J. Clin. Invest.* **87**: 1165–1170.
59. Faustinella, F., A. Chang, J. P. van Biervliet, M. Rosseneu, N. Vinaimont, L. C. Smith. S-H. Chen, and L. Chan. 1991. Catalytic triad residue mutation (Asp^{156}→Gly) causing familial lipoprotein deficiency: coinheritance with a nonsense mutation (Ser^{447}→Ter) in a Turkish family. *J. Biol Chem.* **266**: 14418–14424.
60. Beg, O., M. Meng, S. Skarlatos, L. Previato, J. Brunzell, B. Brewer, and S. Fojo. 1990. Lipoprotein $Lipase_{Bethesda}$: a single amino acid substitution (Ala^{176}→Thr) leads to abnormal heparin binding and loss of enzymatic activity. *Proc. Natl. Acad. Sci. USA.* **57**: 3474–3478.
61. Emi, M., A. Hata, M. Robertson, P. H. Iverius, R. Hegele, and J. M. Lalouel. 1990. Lipoprotein lipase deficiency resulting from a nonsense mutation in exon 3 of the lipoprotein lipase gene. *Am. J. Hum. Genet.* **47**: 107–111.
62. Dichek, H. L., S. S. Fojo, O. U. Beg, S. I. Skarlatos, J. D. Brunzell, G. B. Cutter, Jr., and H. B. Brewer, 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.
63. Henderson, H. E., Y. Ma, M. F. Hassan, M. V. Monsalve, A. D. Marais, F. Winkler, K. Gubernator, J. Peterson, J. D. Brunzell, and M. R. Hayden. 1991. Amino acid substitution (Ile^{194}→Thr) in exon 5 of the lipoprotein gene causes lipoprotein lipase deficiency in three unrelated probands. Support for a multicentric origin. *J. Clin. Invest.* **87**: 2005–2011.
64. Hata, A., E. Mitsuru, G. Luc, A. Basvedant, P. Gambert, P. Iverus, and J. M. Lalouel. 1990. Compound heterozygote for lipoprotein lipase deficiency: $Ser→Thr^{244}$ and transition in 3′ splice site of intron 2 (AG→AA) in the lipoprotein lipase gene. *Am. J. Hum. Genet.* **47**: 721–726.

Downloaded from www.jlr.org by guest, on June 18, 2012